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_17-002-cd

ISGSR 2022

Hydraulic Conductivity Field Marginalization in HMC Based Estimation of Piping Zone Boundary

Kazunori Fujisawa[1], Michael C. Koch[2], Misato Osugi[3] and Akira Murakami[4]

[1]Graduate School of Agriculture, Kyoto University, Kyoto, Japan.
E-mail: fujisawa.kazunori.2s@kyoto-u.ac.jp
[2]Graduate School of Agriculture, Kyoto University, Kyoto, Japan.
E-mail: koch.michaelconrad.5w@kyoto-u.ac.jp
[3]Nippon Koei Co., Ltd., Tokyo, Japan.
E-mail: a9471@n-koei.co.jp
[4]Kyoto University, Kyoto, Japan.
E-mail: murakami.akira.5u@kyoto-u.ac.jp

Abstract: Probabilistic inversion is conducted for the identification of the boundary configuration of a piping zone, which usually develops in soil structures, such as levees. However, the spatially distributing hydraulic conductivity in the soil structure is also unknown and has a significant influence on the estimation of the piping zone. Bayesian inference of both the spatially distributing property, hydraulic conductivity, and the boundary geometry of the piping region become computationally expensive. It is the case even if modern and statistically efficient gradient-based Markov Chain Monte Carlo (MCMC) algorithms called Hamiltonian Monte Carlo (HMC) are implemented. To circumvent this problem, the Bayesian Approximation Error (BAE) approach is employed, by which only the (interesting) piping zone boundary is considered as the target of inversion with the (uninteresting) hydraulic conductivity field fixed. The BAE approach is incorporated into numerical procedures of HMC for boundary estimation in a synthetic steady seepage flow field. Observation data of the hydraulic head and the total outward normal flux at discrete instances in time are numerically prepared. Inversion results are presented with the comparisons of the posterior marginal distributions for the cases with/without the approximation error considered.

Keywords: Inverse problem; Hamiltonian Monte Carlo; Bayesian approximation error; Hydraulic conductivity.

1 Introduction

Maintenance of embankments, such as dams, levees and irrigation tanks (relatively small embankment for irrigation purpose), continues to be an important task for prevention and mitigation of natural disasters in Japan, which is one of the most earthquake-prone country in the world. Piping can be a threat for the maintenance of the embankment because it unknowingly develops within the earthen structure and significantly affect its mechanical stability. According to the statistical investigation of Foster et al. (2000), piping accounted for approximately 40% of failures and accidents of world-wide embankment dams. However, at present, it is difficult to detect the configuration of the piping zone accurately, because of a lack of sophisticated and accessible techniques to identify complicated shapes of subsurface cavities. The identification of the subsurface cavity can be regarded as a geometrical inverse problem.

This article presents a method for solving above-mentioned geometrical inverse problems. The method is based on the Bayesian approach and enables the configuration of a subsurface cavity, such as piping zone, to be identified through the observation of hydraulic head and discharge rate of seepage water. Because seepage water flow concentrates into the piping zone due to its higher hydraulic conductivity and the configuration of the highly permeable region changes the spatial distribution of the hydraulic head, the measurement of the hydraulic head and the discharge rate provides informative observation data for this inverse problem. Although the seepage flow problem is considered in this article, it should be noted that the method proposed herein is not limited to the specific problem.

Figure 1 illustrates the target inverse problem, in which there is a developing cavity of the piping zone with unknown length and thickness to be identified. The hydraulic head spatially distributes under boundary conditions as well as spatially varying hydraulic conductivity. Since the spatial distribution of the hydraulic conductivity is usually unknown, estimation of the configuration of the piping zone requires a geometrical inverse analysis identifying the boundary shape of the seepage flow domain under the unknown hydraulic conductivity. This means that simultaneous identification of the seepage flow boundary and the spatial field of the hydraulic conductivity is needed, but it is a computationally expensive task even though a powerful technique for such simultaneous inverse analysis has been proposed by Koch et al. (2021) with the aid of statistically efficient gradient based Markov Chain Monte Carlo (MCMC) algorithms called Hamiltonian Monte Carlo (HMC).

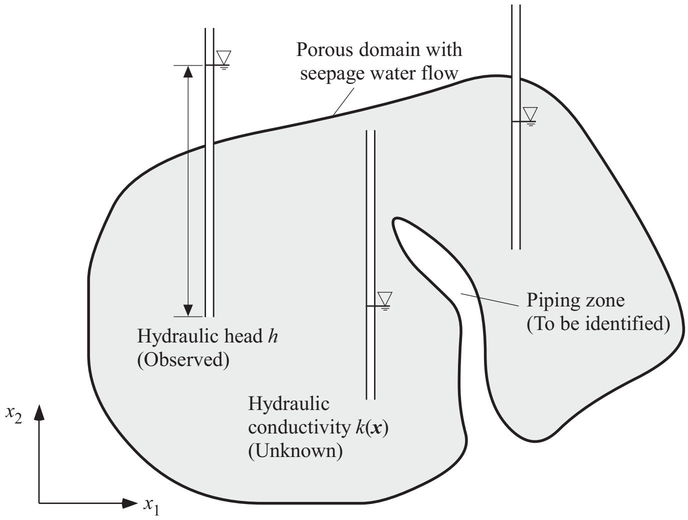

Figure 1. Domain of seepage water flow with a piping zone

The Bayesian Approximation Error (BAE) (Kaipio and Somersalo, 2007; Kolehmainen et al., 2011) approach is employed wherein the (uninteresting) hydraulic conductivity is not to be identified and only the (interesting) piping zone boundary becomes the target of inverse analysis. The ensuing approximation error is accounted for in the inversion process and is premarginalized in terms of a Gaussian approximation for the joint probability density of the parameters describing the boundary and the approximation error. This article presents the results of the inverse analysis identifying the boundary of the piping zone with or without the approximation error considered.

2 Forward Problem, Hamiltonian Monte Carlo (HMC) and Bayesian Approximation Error (BAE)

The well-known steady seepage flow problem in a saturated porous domain is adopted as the forward problem in the Bayesian approach and its governing equation is given as

$$\frac{\partial}{\partial x_i}\left(k(\boldsymbol{x})\frac{\partial h}{\partial x_i} \right) = 0 \tag{1}$$

where$h$ and k denote the hydraulic head and conductivity, respectively, and x_i(or $\boldsymbol{x}$) means the Cartesian coordinates.The discretized form of Eq. (1) by FEM is written in the following form with the observation equation.

$$K_s\boldsymbol{h} = \boldsymbol{q} \quad , \quad \boldsymbol{y} = H_s\boldsymbol{m} + \boldsymbol{e} \quad , \quad \boldsymbol{m}^T = (\boldsymbol{h}^T, \boldsymbol{q}^T) \tag{2}$$

where K_s, $\boldsymbol{h}$, $\boldsymbol{q}$, $\boldsymbol{y}$, H_s, $\boldsymbol{m}$ and $\boldsymbol{e}$ denote the global hydraulic conductivity matrix, the nodal hydraulic head and flux vectors, the vector of observation data, the measurement model matrix, the state vector and the observation noise, respectively. The state vector $\boldsymbol{m}$is obtained by solving the first equation in Eq. (2) and the matrix H_s works as an operator which extracts the values of the hydraulic head or the flux at the observation points from the state vector $\boldsymbol{m}$.

The boundary configuration of the piping zone is parametrized with a parameter vector $\boldsymbol{\theta}$, such as length l and width w. Then, the parameter $\boldsymbol{\theta}$ is the probabilistic variable for Bayesian inversion, which obeys the following posterior distribution.

$$p(\boldsymbol{\theta}|\boldsymbol{y}_{1:n}) = \frac{p(\boldsymbol{y}_{1:n}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\boldsymbol{y}_{1:n}|\boldsymbol{\theta})p(\boldsymbol{\theta})\,d\boldsymbol{\theta}} \propto p(\boldsymbol{y}_{1:n}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{3}$$

where $\boldsymbol{y}_{1:n}$denotes the observations $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, …,$\boldsymbol{y}_n$at different discrete times points. HMC is a variant of MCMC, which can efficiently generate samples (realizations)obeying a certain probability density distribution with an inherent structure similar to the Hamiltonian dynamics. In HMC, the Hamiltonian H is defined as

$$H(\boldsymbol{\theta}, \boldsymbol{p}) = \varphi(\boldsymbol{\theta}) + K(\boldsymbol{p}), \quad \varphi(\boldsymbol{\theta}) = -\log p(\boldsymbol{\theta}|\boldsymbol{y}_{1:n}), \quad K(\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^T M^{-1}\boldsymbol{p} \tag{4}$$

where $\boldsymbol{p}$, φ, K and M are called momentum, potential energy, kinematic energy and mass matrix, respectively. The algorithm of HMC is summarized as below.

Algorithm of HMC

Give $\boldsymbol{\theta}^0$ as the initial value of $\boldsymbol{\theta}$
Sample $\boldsymbol{p}^0 \sim \mathrm{N}(\mathbf{0}, M)$
For j =1 to J
  Update $\boldsymbol{\theta}^{j-1}$ and $\boldsymbol{p}^{j-1}$ into $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{p}}$, respectively by solving
  $$\frac{d\boldsymbol{p}}{dt} = -\frac{\partial H(\boldsymbol{\theta},\boldsymbol{p})}{\partial \boldsymbol{\theta}} = -\frac{\partial \varphi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \quad \frac{d\boldsymbol{p}}{dt} = \frac{\partial H(\boldsymbol{\theta},\boldsymbol{p})}{\partial \boldsymbol{p}} = M^{-1}\boldsymbol{p}$$
  Calculate acceptance probability α
  $$\alpha = \min\left\{1, \exp\left(-H(\tilde{\boldsymbol{\theta}}, -\tilde{\boldsymbol{p}}) + H(\boldsymbol{\theta}^{j-1}, \boldsymbol{p}^{j-1})\right)\right\}$$
  Sample $u \sim \mathrm{U}(0,1)$
  Set $\boldsymbol{\theta}^j \leftarrow \tilde{\boldsymbol{\theta}}$ when $u \le \alpha$, $\boldsymbol{\theta}^j \leftarrow \boldsymbol{\theta}^{j-1}$ when $u > \alpha$
End

In the above algorithm, the gradient of the potential energy, $\partial\varphi/\partial\boldsymbol{\theta}$ is required for proposing the next sample. The calculation of the gradient is special to HMC, which is not include in typical MCMC methods, andthe detailed numerical procedure for it can be found in Koch (2021).

The BAE approach makes the following change in the error part of the observation equation which originally includes the observation noise $\boldsymbol{e}$.

$$\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{\theta}, k(\boldsymbol{x})) + \boldsymbol{e} = \boldsymbol{A}(\boldsymbol{\theta}, k_0) + \boldsymbol{\varepsilon} + \boldsymbol{e} \tag{5}$$

$$\boldsymbol{\varepsilon} = \boldsymbol{A}(\boldsymbol{\theta}, k(\boldsymbol{x})) - \boldsymbol{A}(\boldsymbol{\theta}, k_0) \tag{6}$$

In Eq. (5),$A(\theta, k(x))$ denotes the value of the state variable computed at the observation points with the boundary parameterθ and the spatially changing hydraulic conductivity $k(x)$, while $A(\theta, k_0)$ means the value of the state variable computed with the parameter θ and a constant k_0, i.e., spatially uniform hydraulic conductivity. The error ε in Eq. (6) is involved with the approximation of the original prediction model $A(\theta, k(x))$ by the simpler model $A(\theta, k_0)$, and it can be evaluated in a statistical manner. It should be noted that an arbitrary value of k_0 can be given by those who conduct the inverse analysis. Assuming the approximation error εandthe prior distribution of $\theta$to be Gaussian, $p(\varepsilon|\theta)$canbe obtained in a statistical manner throughevaluation of the joint probability density $p(\varepsilon, \theta)$ as follows.

$$\boldsymbol{\varepsilon}\big|\boldsymbol{\theta} \sim \mathrm{N}\left(\boldsymbol{\varepsilon}_{*,\theta}, \Gamma_{\varepsilon|\theta}\right) \tag{7}$$

$$\boldsymbol{\varepsilon}_{*,\theta} = \boldsymbol{\varepsilon}_* + \Gamma_{\varepsilon\theta}\Gamma_\theta^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_*\right), \quad \Gamma_{\varepsilon|\theta} = \Gamma_\varepsilon + \Gamma_{\varepsilon\theta}\Gamma_\theta^{-1}\Gamma_{\varepsilon\theta} \tag{8}$$

where ε_*, θ_*, Γ_θ and $\Gamma_{\varepsilon\theta}$denote the means of ε and θ, the variance matrix of θ, and the covariance between ε and θ, respectively. The derivation of Eqs. (7)and (8)is referred toin Kolehmainen et al. (2011). Adopting Eq. (5) as the observation equation, the arbitrary selection of the uniform hydraulic conductivity $k_0$becomes possible and the estimation of $k(x)$ can be avoided.

$p(\boldsymbol{\varepsilon}|\boldsymbol{\theta})$ given in Eq. (7) is needed when the likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$relatedto Eq.(3) is evaluated as

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \int p(\boldsymbol{y}|\boldsymbol{\theta}, \varepsilon)\, p(\boldsymbol{\varepsilon}|\boldsymbol{\theta})\, d\boldsymbol{\varepsilon} \tag{9}$$

Substituting Eqs. (5) and (7) to Eq. (9), the following relationship is obtained, which is directly used for computing the posterior distribution, $p(\boldsymbol{\theta}|\boldsymbol{y})$ in Eq. (3).

$$\boldsymbol{y}\big|\boldsymbol{\theta} \sim \mathrm{N}\left(A(\boldsymbol{\theta}, k_0) + \boldsymbol{v}_{*|\theta}, \Gamma_{v|\theta}\right) \tag{10}$$

$$\boldsymbol{v}_{*|\theta} = \boldsymbol{e}_* + \boldsymbol{\varepsilon}_{*,\theta} = \boldsymbol{e}_* + \boldsymbol{\varepsilon}_* + \Gamma_{\varepsilon\theta}\Gamma_\theta^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_*\right), \quad \Gamma_{y|\theta} = \Gamma_e + \Gamma_{\varepsilon|\theta} = \Gamma_e + \Gamma_\varepsilon + \Gamma_{\varepsilon\theta}\Gamma_\theta^{-1}\Gamma_{\varepsilon\theta} \tag{11}$$

where $\boldsymbol{e}_*$ and Γ_e denote the mean and the variance matrix of the observation noise $\boldsymbol{e}$.

3 Numerical Analysis and Results

A numerical analysis for identifying a simple rectangular shape of the piping zone is presented in this section. Figure 2 shows the finite element mesh and the boundary conditions of the problem considered here. The computational domain covers 0.4 m in length and 0.32 m in width. The top right part has the piping zone with the dimensions of *l* in length and *w* in width. The computational domain includes 5 points for the observation of the hydraulic head and 6 sections for the measurement of the discharge rate (corresponding to q_1 to $q_6$in Figure 2)at its right side. As for the boundary conditions, both the top and bottom sides do not allow inflow or outflow of seepage water, the hydraulic head is imposed onto the left side, and the hydraulic head at the right side and the boundary of the piping zone was kept at zero, which means the free inflow/outflow boundary.

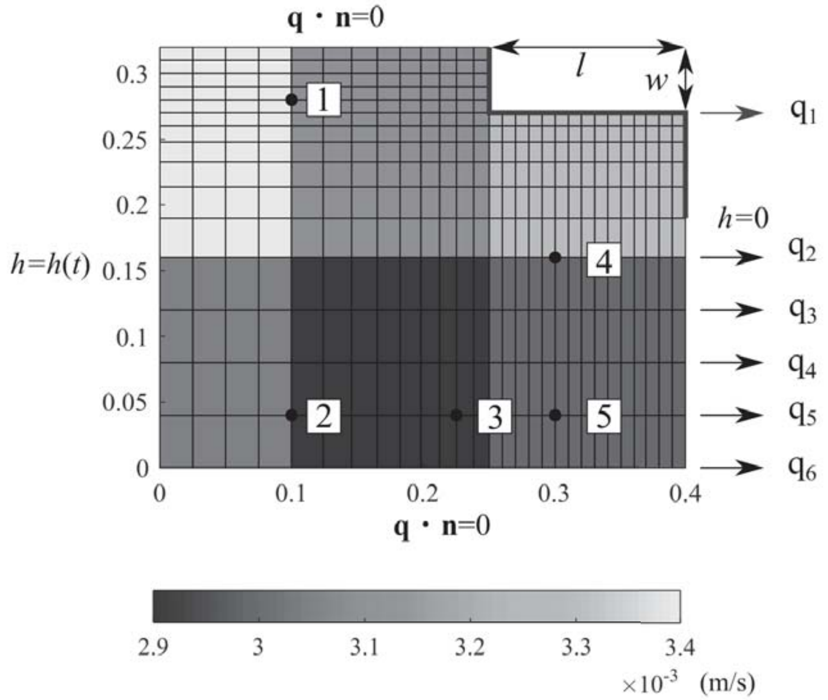

Figure 2. Finite element mesh and boundary conditions

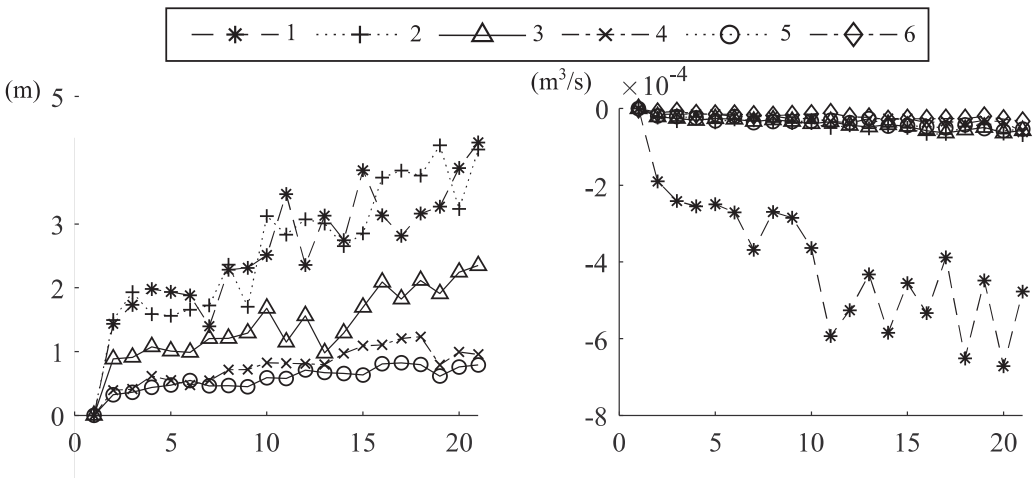

Figure 3. Observation data (Left: Hydraulic head, Right: Discharge rate)

The observation data used for the inverse analysis are shown in Figure 3. Assuming the spatial distribution of the hydraulic conductivity shown in Figure 2, l=0.15 m and w=0.05 m, and imposing the predetermined hydraulic head onto the left side, the synthetic observation data were prepared (The average of the hydraulic conductivity is almost 0.003 m/s). The horizontal axes in Figure 3 implies the discrete time points when the observation or the measurement was conducted. The left plot in the figure corresponds to the observation data of the hydraulic head, while the right plot exhibits those of the discharge rate. These data are produced by adding a certain amount of noise to numerical results of the forward problem.

In order to construct the approximation error ε, the following probability densities for the hydraulic conductivity k (m/s) and the boundary parameter $\theta = (w, l)^T$ (m) were assumed.

$$k \sim N(0.004, 0.002^2)\ ,\ w \sim N(0.06, 0.039^2)\ ,\ l \sim N(0.12, 0.072^2) \tag{12}$$

Before the HMC computation starts, the forward analysis of Eq. (6) enables $p(\varepsilon|\theta)$ to be constructed with the given probability densities shown in Eq. (12). For the sampling process by HMC, the probability density of the observation noise *e* becomes necessary. For simplicity, the normal distribution with the mean of 0 and the standard deviation of 15% of the observed value was assumed for each component of the observation noise ***e***.

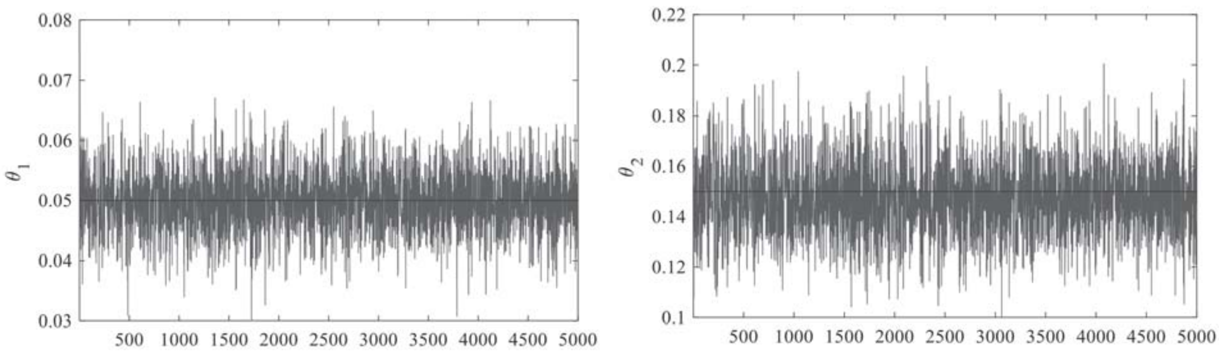

Figure 3. Markov chains for θ_1 and θ_2 corresponding to width*w*and length *l* (HMC with BAE)

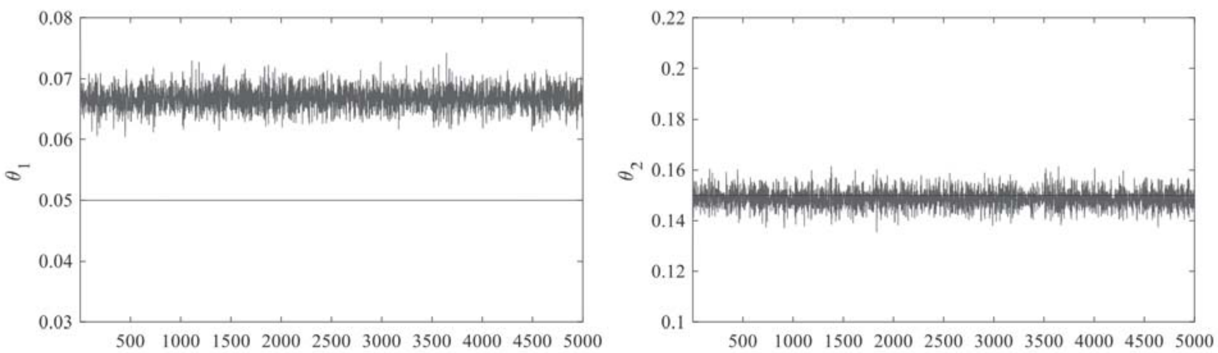

Figure 4. Markov chains for θ_1 and θ_2 corresponding to to width *w* and length *l* (HMC without BAE)

The inverse analysis for identifying the width *w* and the length *l* of the piping zone was carried out, giving k_0=0.004 m/s to the spatially uniform field of hydraulic conductivity. Figure 3 shows the 5,000 Markov chain samples for θ_1 (=*w*) and θ_2(=*l*) computed by HMC with the approximation error considered. It is seen in the figure that the width *w* and the length *l* are nicely identified. On the other hand, Figure 4 provides the Markov chains for θ_1 (=*w*) and θ_2(=*l*) without consideration of the approximation error, which implies that $\boldsymbol{\varepsilon}$ was removed from Eq. (5). As shown in the figure, the width θ_1 (=*w*) is not accurately estimated though the accuracy of the length θ_2(=*l*) is sufficient. The above results indicate that the parameter estimation can be successfully conducted if the approximation error is properly considered, even though unknown material properties (hydraulic conductivity in this case) are included in the forward model.

Acknowledgments

The authors gratefully acknowledge the support provided by JSPS KAKENHI, Grant Number 21H02304.

References

Foster, M., Fell, R. and Spannagle, M. (2000). The statistics of embankment dam failures and accidents.*Canadian Geotechnical Journal*, 37(5), 1000-1024.

Kaipio, J., and Somersalo, E. (2007). Statistical inverse problems: Discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics*, 198(2), 493-504.

Koch, M.C., Osugi, M., Fujisawa, K., and Murakami, A. (2021). Hamiltonian Monte Carlo for Simultaneous Interface and Spatial Field Detection (HMCSISFD) and its application to a piping zone interface detection problem. *International Journal for Numerical and Analytical Methods in Geomechanics*, 45(17), 2602-2626.

Kolehmainen, V., Tarvainen, T., Arridge, S.R., and Kaipio, J.P. (2011). Marginalization of uninteresting distributed parameters in inverse problems- Application to diffuse optical tomography. *International Journal for Uncertainty Quantification*, 1(1), 1-17.