Version of Record: https://www.sciencedirect.com/science/article/pii/S0021999120306781
Manuscript_1bd860e9e738357cce76ebaec330378e

Physics-Informed Machine Learning with Conditional Karhunen-Loève Expansions

A.M. Tartakovsky[a,*], D.A. Barajas-Solano[a], Q. He[a]

[a] *Pacific Northwest National Laboratory, Richland, WA 99354*

Abstract

We present a new physics-informed machine learning approach for the inversion of partial differential equation (PDE) models with heterogeneous parameters. In our approach, the space-dependent partially observed parameters and states are approximated via Karhunen-Loève expansions (KLEs). Each of these KLEs is then conditioned on their corresponding measurements, resulting in low-dimensional models of the parameters and states that resolve observed data. Finally, the coefficients of the KLEs are estimated by minimizing the norm of the residual of the PDE model evaluated at a finite set of points in the computational domain, ensuring that the reconstructed parameters and states are consistent with both the observations and the PDE model to an arbitrary level of accuracy.

In our approach, KLEs are constructed using the eigendecomposition of covariance models of spatial variability. For the model parameters, we employ a parameterized covariance model calibrated on parameter observations; for the model states, the covariance is estimated from a number of forward

*Corresponding author
Email addresses: Alexandre.Tartakovsky@pnnl.gov (A.M. Tartakovsky), David.Barajas-Solano@pnnl.gov (D.A. Barajas-Solano), Qizhi.He@pnnl.gov (Q. He)

simulations of the PDE model corresponding to realizations of the parameters drawn from their KLE. We apply the proposed approach to identifying heterogeneous log-diffusion coefficients in diffusion equations from spatially sparse measurements of the log-diffusion coefficient and the solution of the diffusion equation. We find that the proposed approach compares favorably against state-of-the-art point estimates such as maximum a posteriori estimation and physics-informed neural networks.

Keywords: Conditional Karhunen-Loéve expansions, Parameter estimation, Model inversion, Machine learning.

1. Introduction

Parameter estimation is a critical step in modeling natural and engineered systems [1]. Here, we propose a new physics-informed machine learning method for estimating both parameters and states in systems described by differential equations. We consider the behavior of stationary physical systems modeled by PDEs over the simulation domain $D \subset \mathbb{R}^d$, $d \in [1, 3]$. For simplicity, we assume that the system can be described by a single spatially heterogeneous scalar parameter $y \colon D \to \mathbb{R}$, one state variable $u \colon D \to \mathbb{R}$, and the stationary PDE problem $\mathcal{L}(u, y) = 0$, where $\mathcal{L}(\cdot, \cdot)$ denotes the governing equation and boundary conditions. In this context, the "forward" problem is the problem of computing u given y, and the "inverse" problem is the problem of estimating both y and u given measurements of y and u. In this work, we focus on the inverse problem with spatially sparse measurements of y and u.

We assume that N_{s}^u measurements of u, $\{u_i\}_{i=1}^{N_{\mathrm{s}}^u}$, are collected at spatial

locations $\{x_i^u\}_{i=1}^{N_\mathrm{s}^u}$. Similarly, N_s^y measurements of y, $\{y_i\}^{N_\mathrm{s}^y}$ are collected at locations $\{x_i^y\}_{i=1}^{N_\mathrm{s}^y}$. The observations are organized into the vector of observations $\mathbf{u}_\mathrm{s} = (u_1, \ldots, u_{N_\mathrm{s}^u})^\top$ and $\mathbf{y}_\mathrm{s} = (y_1, \ldots, y_{N_\mathrm{s}^y})^\top$, while the observation locations are organized into the observation matrices $X_\mathrm{s}^u = (x_1^u, \ldots, x_{N_\mathrm{s}^u}^u)$ and $X_\mathrm{s}^y = (x_1^y, \ldots, x_{N_\mathrm{s}^y}^y)$. Finally, we assume that the observations are contaminated by normally distributed observation error, and we denote by Σ_u and Σ_y the error covariance matrices of the u and y observations, respectively.

The inverse problem can be defined as finding the functions u and y that minimize the discrepancy with respect to the observed data while satisfying the governing equations and boundary conditions [2, 3], that is,

$$\begin{aligned} \min_{u,y} \quad & \|u(X_\mathrm{s}^u) - \mathbf{u}_\mathrm{s}\|_{\Sigma_u}^2 + \|y(X_\mathrm{s}^y) - \mathbf{y}_\mathrm{s}\|_{\Sigma_y}^2, \\ \text{s.t.} \quad & \mathcal{L}(u, y) = 0, \end{aligned} \tag{1}$$

where $\|\mathbf{v}\|_\Sigma \coloneqq \mathbf{v}^\top \Sigma^{-1} \mathbf{v}$ denotes the ℓ_2 norm of the vector $\mathbf{v}$ weighted by the inverse of the covariance matrix Σ.

The minimization problem (1) is often solved numerically by discretizing the fields y and u and replacing the PDE constraint with its weak form corresponding to the discretization scheme. Let N denote the number of degrees of freedom of the discretization of the PDE problem. In practice, $N \gg N_s^u + N_s^y$ and, therefore, the minimization problem (1) requires regularization to have a unique solution [4]. The regularized problem reads

$$\begin{aligned} \min_{u,y} \quad & \|u(X_\mathrm{s}^u) - \mathbf{u}_\mathrm{s}\|_{\Sigma_u}^2 + \|y(X_\mathrm{s}^y) - \mathbf{y}_\mathrm{s}\|_{\Sigma_y}^2 + \gamma \mathcal{R}(y), \\ \text{s.t.} \quad & \mathcal{L}(u, y) = 0, \end{aligned} \tag{2}$$

where $\mathcal{R}(\cdot)$ is a regularization penalty encoding regularization assumptions. The regularization parameter $\gamma > 0$ controls the degree to which the dis-

crepancy terms are minimized versus how much the regularization term is minimized. In the context of Bayesian inference [5, 6, 7, 8], up to additive constants, the discrepancy terms are equivalent to the negative log likelihood of the observations, and $\gamma\mathcal{R}(y)$ is equivalent to the negative log prior density of y. Therefore, the solution for y of Eq. (2) is equivalent to the so-called *maximum a posteriori* (MAP) estimate, a Bayesian point estimate defined as the largest mode of the posterior density of y conditional on the observations. Common choices for $\mathcal{R}(\cdot)$ include the H_1 norm, $\|\nabla(\cdot)\|_2^2$, and total variation denoising (TVD), $\|\nabla(\cdot)\|_1$ [9].

Another approach to regularize the minimization problem (1) is the pilot point method [10, 11, 12]. This method consists of parametrizing y in terms of its value at a set of so-called "pilot points." Everywhere else in D, y is evaluated by regressing y measurements and the pilot point values using, e.g., Gaussian Process regression (also known as "kriging") [13, 14, 15, 16, 17]. The value of y at the pilot point locations are estimated from the minimization problem (1).

Bayesian methods, such as Ensemble Kalman Filter (EnKF) [18, 19, 20, 21, 22] and cokriging [23, 24], are commonly used for approximately solving the inversion problem (1). Following stochastic approach to modeling flow and transport [25], EnKF and cokriging treat y and u as the random fields $y(x,\omega) = \mathbb{E}[y(x,\omega)] + y'(x,\omega)$ and $u(x,\omega) = \mathbb{E}[u(x,\omega)] + u'(x,\omega)$ with expectations $\bar{y}(x) \coloneqq \mathbb{E}[y(x,\omega)] =$ and $\bar{u}(x) \coloneqq \mathbb{E}[u(x,\omega)]$, and zero-mean fluctuations $u'(x,\omega)$ and $y'(x,\omega)$. The parameter estimate is computed us-

ing a cokriging update rule of the form

$$y^{\mathrm{EnKF}}(x) = \bar{y}(x) + C_{yu}\left(x, X_{\mathrm{s}}^{u}\right)\left[C_u(X_{\mathrm{s}}^{u}, X_{\mathrm{s}}^{u})\right]^{-1}\left[\mathbf{u}_{\mathrm{s}} - \bar{u}(X_{\mathrm{s}}^{u})\right] + C_y\left(x, X_{\mathrm{s}}^{y}\right)\left[C_y(X_{\mathrm{s}}^{y}, X_{\mathrm{s}}^{y})\right]^{-1}\left[\mathbf{y}_{\mathrm{s}} - \bar{y}(X_{\mathrm{s}}^{y})\right], \quad (3)$$

where C_y and C_u denote the covariances of the y and u random fields, respectively, and C_{yu} denotes the y-u cross-covariance. These covariances are evaluated in practice using sample-based estimates. Inversion schemes of the form of Eq. (3) are straightforward to implement and do not require directly solving a minimization problem. Nevertheless, the resulting estimate y^{EnKF} is not consistent with both data and physics; that is, the solution u of $\mathcal{L}(u, y^{\mathrm{EnKF}}) = 0$ does not match the observed at $\mathbf{u}_s$. Fully Bayesian methods [5] address this inconsistency but often incur significant computational effort, although significant advances have been made in recent years to address computational cost [26, 27].

Machine Learning (ML) methods have arisen in recent years as popular approaches for scientific applications. In general, ML methods require a large amount of data and, therefore, are not feasible for parameter estimation with sparse measurements. To address this challenge, the physics-informed neural network (PINN) method [28, 29, 30] was extended for solving the inverse problem (1) [31]. In this method, both y and u are represented by two separate feed-forward deep neural networks as $u(x) \approx \hat{u}(x;\boldsymbol{\theta})$ and $y(x) \approx \hat{y}(x;\boldsymbol{\gamma})$, where $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ denote the vectors of neural network weights. Next, a "residual" network is defined as

$$\hat{r}(x;\boldsymbol{\theta},\boldsymbol{\gamma}) = \mathcal{L}\left(\hat{u}(x;\boldsymbol{\theta}), \hat{y}(x;\boldsymbol{\gamma})\right), \quad (4)$$

where differentiation with respect to x is performed using automatic differ-

entiation. These three networks are trained jointly by minimizing the loss function

$$\min_{\boldsymbol{\theta},\boldsymbol{\gamma}} \|\hat{u}(X_{\mathrm{s}}^{u};\boldsymbol{\theta}) - \mathbf{u}_{\mathrm{s}}\|_{\Sigma_u}^2 + \|\hat{y}(X_{\mathrm{s}}^{y};\boldsymbol{\gamma}) - \mathbf{y}_{\mathrm{s}}\|_{\Sigma_y}^2 + \rho\,\|\hat{r}(X^{r};\boldsymbol{\theta},\boldsymbol{\gamma})\|_2^2\,, \tag{5}$$

where the residual network is evaluated at certain "residual" points $\{x_i^r \in D\}_{i=1}^{N_r}$, organized into the matrix $X^r = (x_1^r, \ldots, x_{N_r}^r)$. In this approach, the PDE constraint in Eq. (1) is replaced with a weaker constraint on the residuals of $\mathcal{L}(u, y)$; therefore, the estimated fields only approximately satisfy the physics. The advantage of using PINNs is that it does not require discretizing the governing PDE for solving inverse problems.

Here, we propose a new physics-informed ML method for inverse problems based on conditional Karhunen Loève expansions (cKLEs) [32]. In our approach, we model the fields y and u as realizations of Gaussian random fields $\hat{y}^c$ and $\hat{u}^c$ conditioned on observed data. These Gaussian random fields encode the spatial correlation structure of the fields y and u; for the u variable, the corresponding random field satisfies both the data and governing PDE problem $\mathcal{L}(u, y) = 0$. For both random fields, we compute their cKLEs, which allow us to parametrize their realizations in terms of so-called KL coefficients. These KL coefficients are then estimated by solving a regularized form of Eq. (1). We refer to cKLEs trained in this manner as "physics-informed cKLEs," or PICKLEs.

Similarly to PINNs, PICKLEs are trained to satisfy the governing equation $\mathcal{L}(u, y) = 0$ by penalizing the norm of a vector of residuals. Unlike deep neural networks, the KLE of a field enforces its spatial correlation structure and acts as a regularizer. Our results indicate that if the correlation structure of the underlying fields to be estimated is known or can be well estimated

from observation data, then the PICKLE method for inverse problems leads to more accurate parameter estimates than such state-of-the-art inversion approaches as MAP estimation or PINNs.

The remainder of this manuscript is structured as follows. In Section 2, we introduce cKLEs. We describe our algorithm for inverse problems based on PICKLEs in Section 3. Finally, in Section 4, we apply the PICKLE method for the inverse problem of estimating the heterogeneous log-diffusion coefficient of the diffusion equation from sparse measurements of the log-diffusion coefficient and the solution of the diffusion equation. PICKLE estimates are found to compare favorably against MAP and PINNs estimates.

2. Conditional Karhunen Loève expansions

Karhunen Loève expansions (KLEs) [33] are used for representing random fields in terms of linear combinations of uncorrelated random variables. In this work, we employ KLEs as parameterized, deterministic representations of u and y. Specifically, we treat partially known u and y as realizations random fields $\hat{u}^c : D \times \Omega \to \mathbb{R}$ and $\hat{y}^c : D \times \Omega \to \mathbb{R}$ (where Ω is the corresponding random outcome space) conditioned on observed data. Next, we compute the KLEs of these fields, which we use to parametrize their realizations. We refer to these KLEs as *conditional* KLEs, or cKLEs, as by construction they resolve observed data; i.e., at the observation locations the cKLE mean is equal to the field's observation and the cKLE variance is equal to the observation error variance.

In this section, we discuss the construction of the cKLEs. The selection of the Gaussian random field models $\hat{u}^c$ and $\hat{y}^c$ is discussed in Section 3.1.

To introduce cKLEs, we consider a Gaussian random field $z\colon D\times\Omega\to\mathbb{R}$ with the expectation and covariance function, respectively,

$$\bar{z}(x)\coloneqq\mathbb{E}\left[z(x,\omega)\right],\quad C(x,x')\coloneqq\mathbb{E}\left\{[z(x,\omega)-\bar{z}(x)][z(x',\omega)-\bar{z}(x')]\right\}.$$

Next, we assume that a number of noisy spatial observations of z are available, and similar to Section 1, these observations and the observation locations are organized into the vector $\mathbf{z}_\mathrm{s}$ and the matrix X_s, respectively. Furthermore, we denote by $C_s\coloneqq C(X_\mathrm{s},X_\mathrm{s})+\Sigma$ the covariance matrix of the observations, where Σ is the covariance matrix of observation errors. Employing Gaussian process regression (GPR) [14], we find that the conditional Gaussian process (GP) $z^c(x,\omega)\coloneqq z(x,\omega)\mid(\mathbf{z}_\mathrm{s},X_\mathrm{s})$ has the conditional mean and covariance kernels

$$\bar{z}^c(x)=\bar{z}(x)+C(x,X)C_\mathrm{s}^{-1}\left[\mathbf{z}_\mathrm{s}-\bar{z}(X_\mathrm{s})\right],\tag{6}$$

$$C^c(x,x')=C(x,x')-C(x,X)C_\mathrm{s}^{-1}C(X,x'),\tag{7}$$

where the superindex c stands for "conditional" on observations.

The cKLE of z, $z^c(x,\boldsymbol{\xi}(\omega))$, reads

$$z^c\left(x,\boldsymbol{\xi}(\omega)\right)=\bar{z}^c(x)+\sum_{i=1}^{\infty}\phi_i(x)\sqrt{\lambda_i}\xi_i(\omega),\tag{8}$$

where $\boldsymbol{\xi}(\omega)=(\xi_1(\omega),\xi_2(\omega),\cdots)^\top$ is a vector of zero-mean, independent, identically-distributed standard Gaussian random variables, and the eigenpairs $\{\phi_i(x),\lambda_i\}_{i=1}^{\infty}$ are the solutions to the eigenvalue problem

$$\int_D C^c(x,x')\phi(x')\,\mathrm{d}x'=\lambda\phi(x).$$

The sequence of eigenfunctions forms an orthonormal basis on $L_2(D)$.

As the sum in Eq. (8) is infinite, the cKLE in this form is not directly amenable to numerical calculations. Instead, in this work, we will truncate cKLEs to a finite number of terms. For random fields with non-trivial correlation structures (i.e., the non-zero correlation length), the eigenspectrum (i.e., the sequence of eigenvalues λ_i) decays towards zero for increasing i. This, together with the Mercer theorem, justifies the truncation of the KLE to a finite number of terms [34]. Specifically, let z^c_M denote KLE truncated to M terms, given by

$$z^c_M\left(x, \boldsymbol{\xi}_M(\omega)\right) = \bar{z}^c(x) + \sum_{i=1}^{M} \phi_i(x)\sqrt{\lambda_i}\xi_i(\omega). \tag{9}$$

By Mercer's theorem, the (pointwise) variance of the difference between z^c and z^c_M converges uniformly to zero with increasing M, that is,

$$\lim_{M\to\infty} \mathbb{E}\left\{\left[z^c(x,\omega) - z^c_M\left(x, \boldsymbol{\xi}(\omega)\right)\right]^2\right\} = \lim_{M\to\infty} \left\{ C^c(x,x) - \sum_{i=1}^{M} \lambda_i \phi_i^2(x) \right\} = 0$$

uniformly.

This statement of convergence provides a means for selecting M a priori in the context of uncertainty quantification. By the orthonormality of the basis, it follows that the bulk variance and the mean-square truncation error are given by

$$\int_D \operatorname{Var} z^c(x)\, \mathrm{d}x = \sum_{i=1}^{\infty} \lambda_i$$

and

$$\int_D \mathbb{E}\left\{\left[z^c(x,\omega) - z^c_M(x, \boldsymbol{\xi}(\omega))\right]^2\right\} \mathrm{d}x = \sum_{i=M+1}^{\infty} \lambda_i, \tag{10}$$

respectively. Therefore, M is commonly chosen based on either of the fol-

lowing relative and absolute conditions

$$\sum_{i=M+1}^{\infty} \lambda_i \leq \text{rtol} \int_D \text{Var}\, z(x)\, \mathrm{d}x, \qquad \sum_{i=M+1}^{\infty} \lambda_i \leq \text{atol}, \tag{11}$$

for certain relative and absolute tolerances rtol and atol, respectively. We must note that Eq. (10) is a statement about the bulk squared truncation error averaged over all realizations of z^c, and does not provide a bound for the bulk squared truncation error for any given realization.

For the sake of brevity, we organize the sequences of eigenvalues, eigenvectors, and random variables into the vector of functions

$$\boldsymbol{\psi}(x) = \left[\sqrt{\lambda_1}\psi_1(x), \cdots, \sqrt{\lambda_M}\psi_M(x)\right]^\top, \quad \boldsymbol{\xi}(\omega) = (\xi_1(\omega), \cdots, \xi_M(\omega))^\top \tag{12}$$

so that the truncated cKLE (9) can be rewritten in dot product form as

$$z_M^c\left(x, \boldsymbol{\xi}(\omega)\right) = \bar{z}(x) + \boldsymbol{\psi}^\top(x)\boldsymbol{\xi}(\omega). \tag{13}$$

If we treat the ξ_is in the sequence $\{\xi_i\}_{i=1}^M$ not as random variables but as expansion coefficients, we can understand the cKLE as a parameterized representation of functions that satisfy up to measurement error the observed data $(\mathbf{z}_\text{s}, X_\text{s})$. In this context, we refer to the ξ_is as the cKLE "coefficients." Estimating a certain function that satisfies the observed data is then a matter of estimating the cKLE coefficients. We will employ this interpretation of cKLEs to construct our parameter estimation approach in the following section.

3. cKLE-based inversion

In this section, we present the PICKLE method for parameter estimation. In Section 3.1, we describe the selection of the Gaussian random fields used

to construct the cKLEs of y and u. In Section 3.2, we describe how we train these cKLEs subject to a PDE constraint.

3.1. Constructing cKLEs of y and u

To construct the cKLEs of y and u, we first construct conditional GPs $\hat{y}^c\colon D \times \Omega \to \mathbb{R}$ and $\hat{u}^c\colon D \to \Omega \to \mathbb{R}$. Specifically, we select the (unconditional) mean and covariance kernel of these GPs so that they encode the spatial correlation structure of the fields to be estimated. Once the unconditional mean and covariance kernel are selected, the conditional random fields $\hat{y}^c$ and $\hat{u}^c$ are obtained by conditioning on observation data using Eqs. (6) and (7).

3.1.1. cKLE of y

For $\hat{y}^c$, we set the unconditional mean to zero and select the unconditional covariance kernel from a parameterized family of covariance kernels $C^y(\cdot,\cdot \mid \boldsymbol{\theta})$ such as the Matérn, exponential, or square exponential (i.e., Gaussian) kernels. The parameters $\boldsymbol{\theta}$ of the kernel are estimated from the observation data $(\mathbf{y}_\mathrm{s}, X^y_\mathrm{s})$ via marginal likelihood maximization or leave-one-out cross-validation as is commonly done in GPR. We justify this GPR-based approach by noting that it is commonly used in geophysics, under the name of kriging, for estimating spatially heterogeneous geophysical parameters from sparse observations.

Once the unconditional covariance kernel is selected and the conditional mean and covariance of $\hat{y}^c$ are evaluated (via Eqs. (6) and (7)), we construct the cKLE of $\hat{y}^c$ truncated to N_ξ terms of the form of Eq. (13), namely,

$$\hat{y}^c(x,\boldsymbol{\xi}) = \bar{y}^c(x) + \boldsymbol{\psi}_y^\top(x)\boldsymbol{\xi}, \quad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_\xi}), \tag{14}$$

where $\mathcal{N}(0, \mathbf{I}_{N_\xi})$ is the multivariate normal distribution, and $\mathbf{I}_{N_\xi}$ is the $N_\xi \times N_\xi$ identity matrix.

3.1.2. cKLE of u

For $\hat{u}^c$, the data-driven GPR-based strategy is inadequate, as samples from common parameterized Gaussian process models are not guaranteed to satisfy the governing equations and boundary conditions. Therefore, in this work we employ a Monte Carlo simulation-based method for computing the unconditional mean and covariance for u. We construct an ensemble of N_{ens} realizations of $\hat{y}^c$, $\{y^{(i)}\}_{i=1}^{N_{\text{ens}}}$ by sampling $\boldsymbol{\xi}^{(i)}$ from $\mathcal{N}(0, \mathbf{I}_{N_\xi})$ and, then, evaluating the cKLE model of $\hat{y}^c$, Eq. (14), with $\boldsymbol{\xi} = \boldsymbol{\xi}^{(i)}$, that is,

$$y^{(i)}(x) = \bar{y}^c(x) + \boldsymbol{\psi}_y^\top(x)\boldsymbol{\xi}^{(i)}, \quad \boldsymbol{\xi}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_\xi}). \tag{15}$$

For each member of the ensemble $\{y^{(i)}\}_{i=1}^{N_{\text{ens}}}$, we calculate $u^{(i)}$ by solving the PDE problem $\mathcal{L}(u^{(i)}, y^{(i)}) = 0$, thus obtaining the ensemble of u fields, $\{u^{(i)}\}_{i=1}^{N_{\text{ens}}}$. The unconditional mean and covariance of u ($\overline{u}$ and $C_u(x, x')$) are then computed as the ensemble estimates

$$\overline{u}(x) = \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} u^{(i)}(x), \tag{16}$$

$$C_u(x, x') = \frac{1}{N_{\text{ens}} - 1} \sum_{i=1}^{N_{\text{ens}}} \left[u^{(i)}(x) - \overline{u}(x)\right] \left[u^{(i)}(x') - \overline{u}(x')\right]. \tag{17}$$

This procedure is summarized in Algorithm 1.

Once the unconditional mean and covariance kernels are estimated, the conditional mean and covariance of $\hat{u}^c$ are calculated using Eqs. (6) and (7), and the cKLE model for $\hat{u}^c$ is constructed in the form of Eq. (13), namely,

$$\hat{u}^c(x, \boldsymbol{\eta}) = \bar{u}^c(x) + \boldsymbol{\psi}_u^\top(x)\boldsymbol{\eta}. \tag{18}$$

Algorithm 1 Sampling-based covariance model for u

Require: X_{s}^{u}, $\mathbf{u}_{\mathrm{s}}$, N_{ens}

1: **for** $i \leftarrow 1, N_{\mathrm{ens}}$ **do**
2: Generate $y^{(i)}$ via Eq. (15)
3: Compute $u^{(i)}$ by solving $\mathcal{L}(u^{(i)}, y^{(i)}) = 0$
4: **end for**
5: Compute ensemble mean and covariance of $\{u^{(i)}\}$, $\overline{u}$ and $C_u(x, x')$, using Eqs. (16) and (17)
6: Compute conditional mean and covariance of $\hat{u}^c$, $\overline{u}^c$ and $C_u^c(x, x')$, using Eqs. (6) and (7)

The ensemble covariance estimate requires some discussion. We require $N_{\mathrm{ens}} > N_{\mathrm{s}}^{u}$ so that the rank of the unconditional covariance is larger than N_{s}^{u} and the conditional covariance is not trivial. This limitation can be avoided by instead employing shrinkage estimators, which are known to be robust for a small number of ensemble elements [35]. In [24], the computational cost of the Monte Carlo simulations for estimating the unconditional covariance of u was reduced by using the Multilevel Monte Carlo method [36]. For small unconditional variance of y, the moment equation method can be used to derive a system of deterministic equations for the unconditional covariance of u [37, 38]. Then, the unconditional covariance of u can be found by solving these equations numerically.

Finally, we point out that the accuracy of cKLE-based inversion depends on the expressivity of the truncated cKLE of u, i.e., its capacity of approximately representing the solutions of the PDE problem $\mathcal{L}(u, y) = 0$. Due to the Gibbs phenomenon, the accuracy of cKLEs is limited for fields with sharp

gradients or discontinuities [39]. If the PDE problem and the Gaussian ansatz for y, Eq. (15), leads to a solution field u that is uni-modal non-Gaussian (such as in the numerical experiments presented in Section 4) then we find that cKLEs lead to adequate inversion results. On the other hand, the PDE problem may lead to a solution field u with a multi-modal distribution even for a Gaussian y [40]. For such problems, cKLE-based inversion as formulated in this work may not be adequate. We will consider such problems in future work.

3.2. The PICKLE method for inverse problems

In this section we describe the proposed PICKLE method for inverse problems. Similarly to PINN, in PICKLE we replace the PDE constraint in Eq. (1) with a penalty on the norm of the vector of residuals,

$$\mathbf{r}[u,y] \coloneqq \left[\mathcal{L}(u,y)\mid_{x=x_1^r},\ldots,\mathcal{L}(u,y)\mid_{x=x_{N_r}^r}\right]^\top,$$

where each component of the vector corresponds to the residual of the PDE problem evaluated at the ith "residual" point of the sequence $\{x_i^r \in D\}_{i=1}^{N_r}$.

The constraint on the residuals is added as a penalty term into the objective function, leading to the minimization problem

$$\min_{u,y} \quad \|u(X_\mathrm{s}^u)-\mathbf{u}_\mathrm{s}\|_{\Sigma_u}^2+\|y(X_\mathrm{s}^y)-\mathbf{y}_\mathrm{s}\|_{\Sigma_y}^2+\rho\|\mathbf{r}[u,y]\|_2^2, \tag{19}$$

where $\rho>0$ is a penalty parameter.

We now proceed to introduce the cKLE models for y and u. Namely, we interpret the cKLEs Eq. (14) and Eq. (18) as representations of functions parameterized by the vectors of cKLE coefficients $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, leading to the

deterministic cKLE models

$$y^c(x, \boldsymbol{\eta}) = \bar{y}^c(x) + \boldsymbol{\psi}_y^\top(x)\boldsymbol{\xi}, \tag{20}$$

$$u^c(x, \boldsymbol{\eta}) = \bar{u}^c(x) + \boldsymbol{\psi}_u^\top(x)\boldsymbol{\eta}. \tag{21}$$

Substituting Eqs. (20) and (21) into Eq. (19), we obtain the following minimization problem in terms of the cKLE parameters:

$$\min_{\boldsymbol{\xi},\boldsymbol{\eta}} \quad \|u^c(X_{\mathrm{s}}^u, \boldsymbol{\eta}) - \mathbf{u}_{\mathrm{s}}\|_{\Sigma_u}^2 + \|y^c(X_{\mathrm{s}}^y, \boldsymbol{\xi}) - \mathbf{y}_{\mathrm{s}}\|_{\Sigma_y}^2 + \rho \left\|\mathbf{r}[u^c(\cdot, \boldsymbol{\eta}), y^c(\cdot, \boldsymbol{\xi})]\right\|_2^2.$$

By construction, the cKLE models minimize the discrepancy terms. This leaves only the penalty term, so that the coefficient ρ can be dropped.

It remains to regularize the problem. In this work, we choose to penalize the ℓ_2-norm of the vectors of cKLE parameters, resulting in the final PICKLE minimization problem

$$\min_{\boldsymbol{\xi},\boldsymbol{\eta}} \quad \left\|\mathbf{r}[u^c(\cdot, \boldsymbol{\eta}), y^c(\cdot, \boldsymbol{\xi})]\right\|_2^2 + \gamma \left(\|\boldsymbol{\xi}\|_2^2 + \|\boldsymbol{\eta}\|_2^2\right), \tag{22}$$

where $\gamma > 0$ is a regularization penalty. Substituting $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, estimated from Eq. (22), into Eq. (20) and Eq. (21) provides the PICKLE estimates of the y and u fields. The proposed model inversion algorithm is summarized in Algorithm 2.

3.3. Computational cost

Common iterative, gradient-based approaches to the solution of the PDE-constrained minimization problem of (1) aim to minimize the objective function with respect to y, with u given explicitly at every iteration of the procedure as the solution of the PDE constraint, $\mathcal{L}(u, y) = 0$, for given y. The gradient of the objective function with respect to y is then found by the

Algorithm 2 cKLE-based inversion

Require: X_{s}^{y}, $\mathbf{y}_{\mathrm{s}}$, X_{s}^{u}, $\mathbf{u}_{\mathrm{s}}$, $C^{y}(\cdot,\cdot \mid \boldsymbol{\theta})$, N_{ξ}, N_{η}, N_{ens}

1: Estimate $\boldsymbol{\theta}$ via GPR model selection
2: Compute conditional mean and covariance of y^{c} using Eqs. (6) and (7)
3: Calculate KLE of y^{c}
4: Calculate cKLE model, Eq. (20), truncated to N_{ξ} terms
5: Compute conditional mean and covariance of u^{c} using Algorithm 1
6: Calculate KLE of u^{c}
7: Calculate cKLE model, Eq. (21), truncated to N_{η} terms
8: Estimate $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ via Eq. (22)
9: Compute y and u from estimated $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ using Eqs. (20) and (21)

application of the chain rule and the adjoint method, e.g., see [41, 42]. Such approaches require solving the PDE constraint at every step of the iteration process. In contrast, in PICKLE, there is no need to solve the governing PDE. Instead, our approach requires only evaluating the norm of the vector of residuals and its gradient with respect to the cKLE coefficients.

The calculation of the residuals' norm gradient deserves special consideration. One can consider a strong or weak form of the PDE residual. The strong form of the PDE residual requires evaluating the spatial derivatives of the cKLEs of y and u, which in turn, requires obtaining the cKLE in terms of closed form functions. While the eigenproblem for the cKLE cannot be exactly solved in closed form in general, closed-form approximations in terms of orthogonal polynomials (e.g., Chebyshev polynomials) can be obtained (e.g., [43, 44]). The benefit of having the closed-form cKLEs is that the norm of residuals and its gradients can be evaluated using automatic differentiation

of the composition of the residual and the cKLEs. In this work, we consider the residual of a weak form of the PDE constraint. In this case, it suffices to solve the eigenproblems and compute the cKLEs of y and u on the discretized grids corresponding to the weak form of the PDE problem. In Section 4, we discuss the finite volume (FV) scheme we employ to discretize PDEs in our numerical experiments, and how the gradients of the residual are computed using this scheme.

The three factors that chiefly control the computational cost of the PICKLE approach are (i) the number of samples in the ensemble $\{u^{(i)}\}$, N_{ens}, (ii) the number of cKLE parameters, and (iii) the size of the vector of residuals. In the following, we discuss these sources of computational cost one-by-one.

(i) In PICKLE, the governing PDE is solved N_{ens} times, a number specified a priori. In comparison, traditional gradient-based approaches to the solution of the PDE-constrained optimization problem require a number of solutions of the PDE constraint that cannot be controlled a priori. Therefore, the computational cost of solving complex physics problems cannot, in general, be controlled a priori for such approaches. Also, in PICKLE, each of N_{ens} realizations can be run independently. Therefore, PICKLE is trivially parallelizable, which can dramatically reduce the computational time associated with this cost.

(ii) As we show in Section 4, the cKLEs allow us to represent the y and u fields in terms of a relatively low number of KLE parameters, which makes it possible to tackle high-dimensional problems once the cKLE eigenfunctions and eigenvalues have been computed. Specifically, accurate solutions can be obtained with a number of KLE parameters significantly less than the

number necessary to represent the unconditional y and u fields. The accuracy of PICKLE strongly depends on the expressive capacity of truncated cKLEs, which is known to be limited for fields with sharp gradients or discontinuities due to the Gibbs phenomenon. Nevertheless, piecewise-continuous y fields can be treated with our approach by introducing latent fields, as described in Section 4.2.

The computational cost of solving the eigenvalue problem must be taken into account when comparing the cost of PICKLE against that of alternative approaches to inversion. In this work, we compute cKLEs by solving the dense eigenvalue problems corresponding to evaluating conditional covariances on a FV grid. For the problems considered in Section 4, the associated cost is negligible. Nevertheless, discretizations with a large number of degrees of freedom would lead to large eigenvalue problems of significant computational cost. For such problems, more efficient algorithms for computing cKLEs are necessary, e.g., algorithms based on fast multipole methods [45], hierarchical matrices [46], or Chebyshev polynomials [43, 44].

(iii) With regard to the vector of residuals, the proposed inversion approach provides significant flexibility for the choice of residuals. For the numerical experiments shown in Section 4, we employ the residuals of the FV discretization of the PDE constraint evaluated at a subset of the FV elements of the discretization. The residuals' vector size can be adjusted to reduce the computational cost of the inverse problem solution.

4. Numerical experiments

In this section, we use PICKLE for estimating the heterogeneous diffusion coefficient of the elliptic diffusion equation. Specifically, we consider the PDE problem

$$\begin{aligned} \nabla \cdot \left[e^{y(x)} \nabla u(x) \right] &= 0, \quad x \in D := [0,1]^2, \\ u(x) &= 1, \quad x_1 = 0, \\ u(x) &= 0, \quad x_1 = 1, \\ e^{y(x)} \frac{\partial u(x)}{\partial x_2} &= 0, \quad x_2 = \{0,1\}, \end{aligned} \tag{23}$$

where $u(x)$ is the PDE solution and $y(x)$ is the log-diffusion coefficient. Among other problems, this equation describes saturated flow in heterogeneous porous media [47]. Our goal is to estimate the spatial distribution of y from noiseless sparse observations of y and u.

In our numerical experiments, we discretize the simulation domain into a uniform grid of $m \times m$ rectangular elements, for a total of $n = m^2$ cells. The PDE problem Eq. (23) is then discretized, employing a cell-centered FV scheme and the two-point flux approximation [48]. After discretization, the problem (23) is translated into the system of nonlinear algebraic equations

$$\mathbf{r}[\mathbf{u}, \mathbf{y}] \equiv \mathbf{A}(\mathbf{y})\mathbf{u} - \mathbf{b}(\mathbf{y}) = 0, \tag{24}$$

where $\mathbf{y} \in \mathbb{R}^n$ and $\mathbf{u} \in \mathbb{R}^n$ are the cell-centered discretizations of the y and u fields, respectively, and $\mathbf{A}\colon \mathbb{R}^n \to \mathbb{R}^{n \times n}$ and $\mathbf{b}\colon \mathbb{R}^n \to \mathbb{R}^n$ are the stiffness matrix and right-hand side vector, respectively. We use the chain rule to compute the gradient of $\mathbf{r}[\cdot, \cdot]$ with respect to cKLE coefficients. The numerical implementation of the FV scheme is extended to compute the gradients $\partial \mathbf{A}/\partial \mathbf{y}$ and $\partial \mathbf{b}/\partial \mathbf{y}$ using closed-form expressions for the derivatives

of the entries of $\mathbf{A}$ and $\mathbf{b}$ with respect to the entries of $\mathbf{y}$. Finally, the gradients $\partial\mathbf{y}/\partial\boldsymbol{\xi}$ and $\partial\mathbf{u}/\partial\boldsymbol{\eta}$ are computed from Eqs. (20) and (21).

To evaluate the accuracy of the PICKLE method, we compare the reconstructed log-diffusion field to the reference field employed to generate the synthetic observations. Furthermore, we compare the reconstructed field against the MAP estimate with H_1 regularization, a commonly used PDE-constrained optimization-based inversion approach, and the PINN method [31]. For both MAP and PICKLE, we use the regularization parameter $\gamma = 1 \times 10^{-6}$. The optimization problems for MAP, Eq. (2), and PICKLE, Eq. (22), are solved using the Levenberg-Marquard algorithm [49] as implemented in Scientific Python (SciPy) [50]. For MAP, the gradient of the solution $\mathbf{u}$ of Eq. (24) with respect to $\mathbf{y}$ is computed using the discrete adjoint method [41, 42, 26]. The PINN method does not employ regularization other than the regularization provided by physics constraints. The PINN implementation details are given in Section 4.1.

The PICKLE method and MAP estimation were implemented in SciPy. PICKLE and MAP estimates were computed using a Intel Xeon W-2135 workstation and GNU Parallel [51]. The PINN method was implemented in TensorFlow [52], and PINN estimates were computed using a Intel Xeon ES-1620 workstation.

4.1. Continuous diffusion field

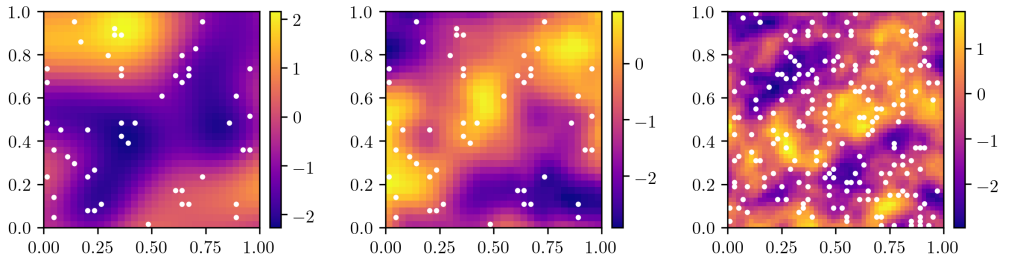
We first consider continuous reference y fields of various degrees of smoothness. Three reference fields are generated as realizations of zero-mean Gaussian processes with the isotropic Matérn (with $\nu = \{3/2, 5/2\}$) and Gaussian kernels for the values of the kernel hyperparameters (namely the correlation

Table 1: Properties of synthetic reference log-diffusion fields and PICKLE estimation parameters.

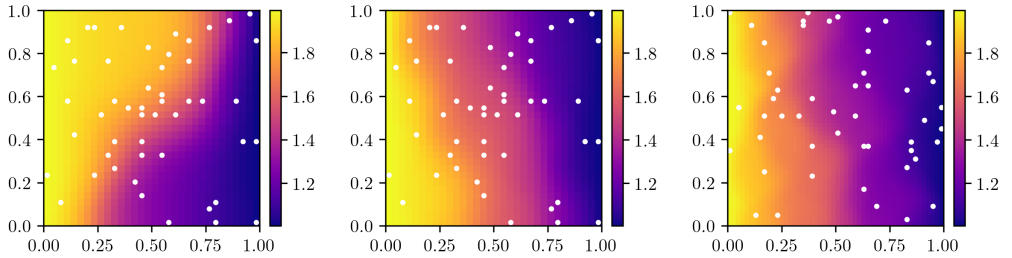
	λ	σ	m	N_ξ	N_η	N_s^y	N_s^u
Gaussian	0.2	1.0	32	100	100	50	50
Matérn $\nu = 5/2$	0.2	1.0	32	100	100	50	50
Matérn $\nu = 3/2$	0.1	1.0	50	300	200	200	50

length λ and standard deviation σ), listed in Table 1. The corresponding reference u fields are computed by solving the FV discretization of the PDE of Eq. (23) using a $m \times m$ uniform grid with m given in Table 1. Finally, observation locations for y and u are chosen randomly from the set of cell centers of the FV scheme. The number of observations are listed in Table 1. For the reference fields with the Gaussian and Matérn $\nu = 5/2$ kernels, we assume that 50 observations for both y and u are available. For the Matérn $\nu = 3/2$ case, we use 200 y observations and 50 u observations to estimate the field. The larger number of observations is necessary for the latter case as this problem is more challenging due to its short correlation length and lower smoothness. The number of KLE terms for this kernel given by the condition Eq. (11) with rtol = 99% is 511. The reference fields and observation locations are shown in Fig. 1.

As described in Section 3.1.1, we construct the cKLE model for y by training a GPR model to the observation data $(\mathbf{y}_\mathrm{s}, X_\mathrm{s}^y)$. We consider two scenarios: (i) the data-generating kernel and its hyperparameters are known, and (ii) the data-generating kernel is known but the hyperparameters are unknown. We will refer to the first scenario as "cKLI" and the second scenario



(a) $y(x)$, Gaussian kernel (b) $y(x)$, Matérn $\nu = 5/2$ (c) $y(x)$, Matérn $\nu = 3/2$



(d) $u(x)$, Gaussian kernel (e) $u(x)$, Matérn $\nu = 5/2$ (f) $u(x)$, Matérn $\nu = 3/2$

Figure 1: Sample reference y (above) and u (below) fields for the parameters of Table 1.

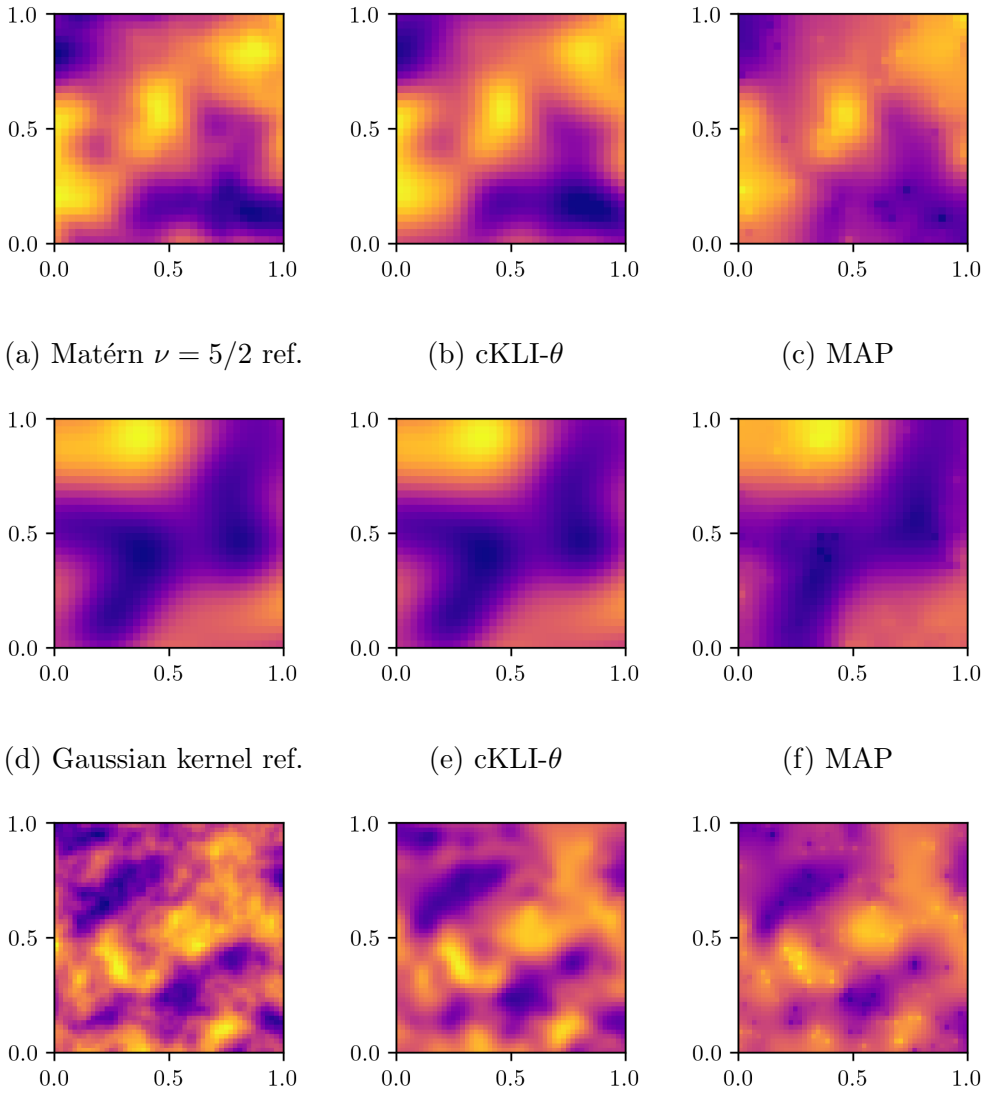
as "cKLI-θ." For cKLI-θ, the kernel hyperparameters are estimated using the observations y_s via marginal likelihood estimation, which is performed using the library GPy [53].

Fig. 2 presents the reference y fields and the cKLI-θ and MAP estimates of these fields. The cKLI-θ estimates are computed using N_ξ and N_η listed in Table 1. For all cases, we used $N_\mathrm{ens} = 5 \times 10^3$. It can be seen that the cKLI-θ estimates of y are more accurate than the MAP estimates for all considered cases; this advantage is more noticeable for the Matérn cases, which is less smooth than the Gaussian case. Furthermore, the MAP estimates exhibit peaks at the y observation locations, a phenomenon typical to H_1 regularization, whereas the cKLI-θ estimates are smooth and, thus, better approximate the reference fields.

We define the "relative ℓ_p error" as the ℓ_p-norm of the estimation error with respect to the ℓ_p-norm of the reference field, that is,

$$\text{relative } \ell_p \text{ error} \coloneqq \|y_\mathrm{ref} - y_\mathrm{est}\|_p / \|y_\mathrm{ref}\|_p.$$

In Table 2, we present the relative ℓ_2 error of the estimates shown in Fig. 2. We also present the relative error of the cKLI-θ estimate computed using subsampled residuals with a subsampling factor of 2 in each direction (resulting in a reduction in the dimension of the vector of residuals by a factor of 4). Subsampling reduces the dimension of the vector of residuals and, therefore, reduces the computational effort of computing the cKLI-θ estimate. For comparison, we also present the ℓ_2 error in the estimates obtained with the cKLI method. As expected, the accuracy of the cKLI estimated y is the same or better than that of the cKLI-θ estimate for all considered cases. This is because the accuracy of PICKLE estimation depends on the



(a) Matérn $\nu = 5/2$ ref. (b) cKLI-θ (c) MAP

(d) Gaussian kernel ref. (e) cKLI-θ (f) MAP

(g) Matérn $\nu = 3/2$ ref. (h) cKLI-θ (i) MAP

Figure 2: cKLI-θ estimate (middle) and MAP estimate (right) of the reference log-diffusion fields of Fig. 1.

accuracy of the estimated y kernel, and in the cKLI case we assume that the y kernel is known exactly. It can be seen that for the cases considered so far, the PICKLE for estimating y is more accurate in the ℓ_2 sense than MAP, and that subsampling by a factor of 2 of the vector of residuals does not significantly increase the PICKLE estimation error.

For comparison, we also estimate y using the PINN method for parameter estimation [31]. Here, we represent the y and u fields using deep feed-forward neural networks with three hidden layers and 30 neurons per layer. The residual is estimated at 1024 points in the interior of the simulation domain. Boundary condition residuals are evaluated at m points along each of the four segments of the boundary, where m is given in Table 1. We conduct ten simulations with different initializations using the Xavier's initialization scheme and train the PINN networks by using the L-BFGS-B method. The mean and standard deviation across initializations of the relative ℓ_2 error is reported in Table 2. The relative ℓ_2 error of PICKLE cKLI-θ estimation (with residual subsampling) is approximately 414% smaller than of PINN estimation for the Gaussian kernel reference, 52% smaller for the Matérn $\nu = 3/2$ reference, and 49% smaller for the Matérn $\nu = 5/2$ reference.

In terms of computational effort, the MAP estimate was computed in ~ 24 s for the Gaussian and Matérn $\nu = 5/2$ cases, and ~ 450 s seconds for Matérn $\nu = 3/2$ case. The PICKLE estimate (with residual subsampling) was computed in ~ 10 s for all cases. Finally, the PINN estimate was computed in ~ 300 s for each initialization and all cases. Note that the PINN estimate was computed using a different computer than the MAP and PICKLE estimates, so no direct comparison of wall clock times is possible.

Table 2: Relative ℓ_2 error of the y estimates shown in Fig. 2 and obtained with PINNs. For cKLI-θ, "Full" indicates the estimate computed using the full vector of FV residuals, and "Subsampled" indicates the estimate computed using a subsampling of the vector of residuals by a factor of 2 in each spatial direction.

	cKLI		cKLI-θ		MAP	PINN
	Full	Subsampled	Full	Subsampled		
Gaussian	0.010	0.015	0.009	0.014	0.150	0.072(27)
Matérn $\nu = 5/2$	0.099	0.109	0.099	0.111	0.224	0.169(11)
Matérn $\nu = 3/2$	0.257	0.254	0.263	0.261	0.419	0.388(16)

To evaluate the robustness of PICKLE, we calculate the relative ℓ_2 estimation error for different reference fields (generated as realizations of the random fields with the Matérn ($\nu = 5/2$) and Gaussian kernels with $\sigma = 1.0$ and the correlation lengths $\lambda = 0.2$ and 0.5) and choices of observation locations. Furthermore, we study how the relative ℓ_2 error depends on the number of cKLE terms in the expansions of the y and u fields. For each combination of kernel and correlation length, we generate 10 reference y fields and the corresponding u fields. For each reference field, we randomly generate observation locations and compute the cKLI and cKLI-θ estimates of y. We do this for $N_{\mathrm{s}}^u = 50$, $N_\eta = 100$, two values of N_{s}^y, 10 and 50, and various values of N_ξ.

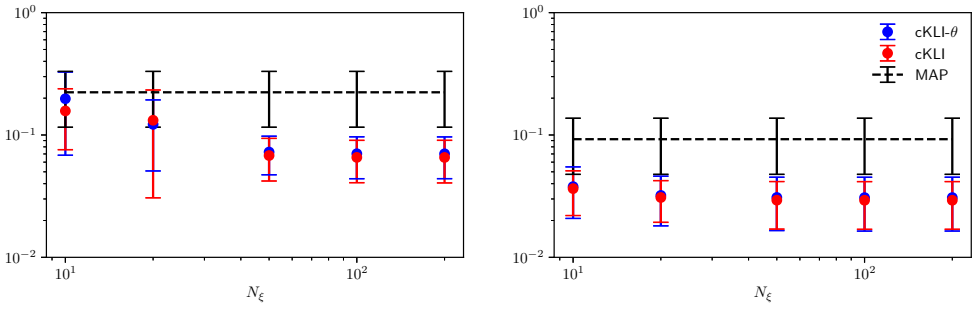
The relative ℓ_2 error as a function of N_ξ is shown in Figs. 3 and 4. As in the previous example, we can see that the accuracy of the cKLI estimated y is the same or better than that of the cKLI-θ estimate for all considered cases. The cKLI estimates are consistently more accurate in the ℓ_2 sense

than the MAP estimate for sufficiently large N_ξ. In particular, we note that a good rule of thumb for N_ξ is to be larger than the number of KLE terms of the reference kernel for rtol = 99% minus the number of observations.

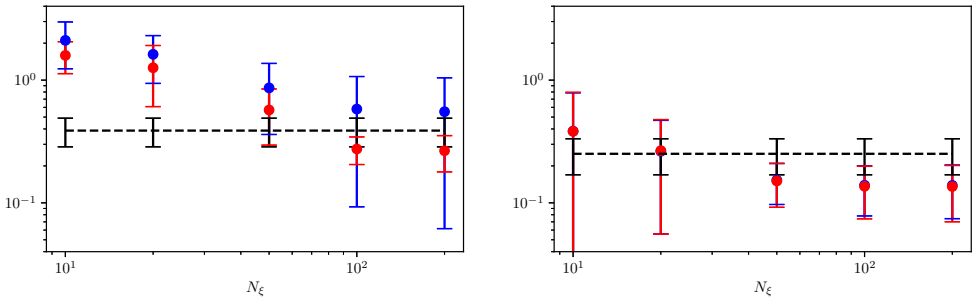
As expected, for the rougher Matérn kernel, more KL terms are needed to obtain an accurate y estimate than for the smoother Gaussian kernel. The same observation is true with respect to the correlation length: the smaller is the correlation length, the more KL terms are needed to obtain an accurate y estimate.

The cKLI-θ estimates require additional discussion. For all considered fields except one, the cKLI-θ estimate of y is more accurate than the MAP estimate for sufficiently large N_ξ. For $N_\mathrm{s}^y = 10$ and the rough (Matérn) kernel with small correlation length, the cKLI-θ is worse than the MAP estimate for all considered N_ξ (Fig. 3c). This is because 10 y observations are not sufficient to obtain adequate estimates of the hyperparameters of the kernel for such a rough y field. Fig. 3d shows that, for the same Matérn kernel, a very accurate estimate of hyperparameters is obtained with 50 y measurements, and the cKLI-θ estimates of y are as accurate as the cKLI estimates and more accurate than MAP estimation for sufficiently large values of N_ξ.

The comparison of the cKLI and cKLI-θ results show that the more accurate the estimate of the y kernel is, the less terms in the cKLE model of y are needed to obtain an accurate estimate of y. These results also indicate that it is not necessary to know the y kernel exactly for PICKLE estimation to produce an accurate estimate of y, given that N_ξ is sufficiently large.

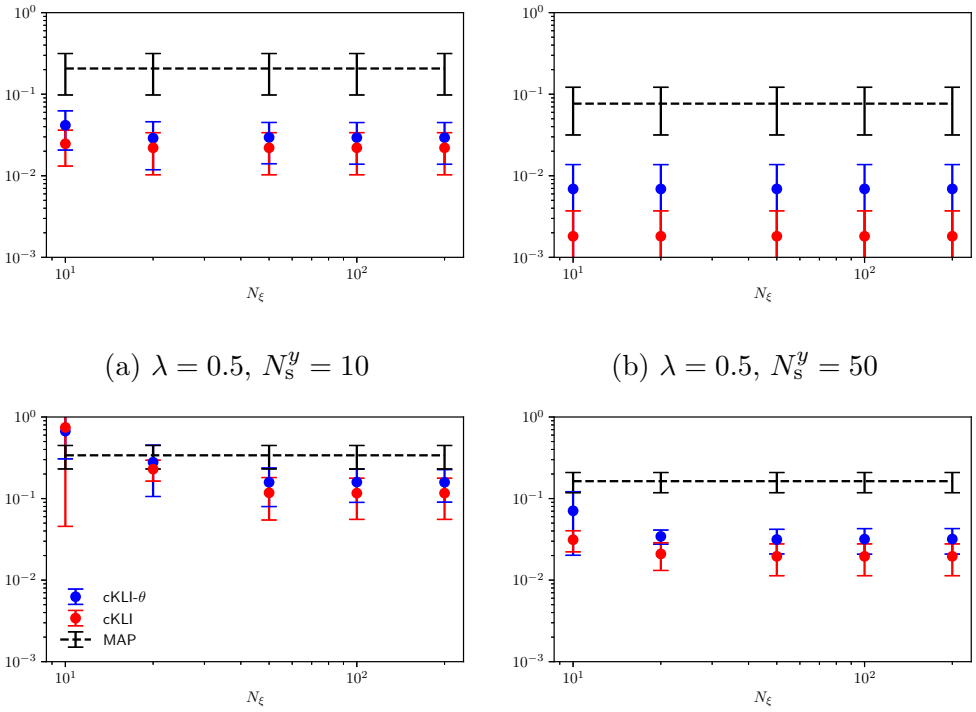


(a) $\lambda = 0.5$, $N_{\mathrm{s}}^y = 10$ (b) $\lambda = 0.5$, $N_{\mathrm{s}}^y = 50$



(c) $\lambda = 0.2$, $N_{\mathrm{s}}^y = 10$ (d) $\lambda = 0.2$, $N_{\mathrm{s}}^y = 50$

Figure 3: Relative ℓ_2 error for Matérn covariance kernel with $\nu = 5/2$ and with different values of λ and N_{s}^y.



(a) $\lambda = 0.5$, $N_{\mathrm{s}}^y = 10$

(b) $\lambda = 0.5$, $N_{\mathrm{s}}^y = 50$

(c) $\lambda = 0.2$, $N_{\mathrm{s}}^y = 10$

(d) $\lambda = 0.2$, $N_{\mathrm{s}}^y = 50$

Figure 4: Relative ℓ_2 error for the Gaussian covariance kernel with different values of λ and N_{s}^y.

4.2. Piecewise-constant diffusion field

The problem of reconstructing piecewise-constant and piecewise-continuous fields arise naturally in geophysical applications, where spatially distributed geophysical properties are arranged into facies [54, 55]. In this section we apply PICKLE to estimating a piecewise-constant diffusion field k, shown in Fig. 5.

Due to the Gibbs phenomenon, a large number of KLE terms would be necessary to accurately represent either k or $y \coloneqq \log k$ directly. Therefore, to treat this case with our proposed method, we introduce a latent field $f(x)$ that can be represented accurately using a finite-dimensional representation. For this application, we assume that the log-diffusion field consists of two facies, with constant log-diffusion values of y_1 and y_2, with $y_1 > y_2$. The log-diffusion coefficient is then approximated in terms of the latent field $f(x)$ as

$$y(x) = (y_1 - y_2)\,\mathrm{expit}\left(\varepsilon^{-1} f(x)\right) + y_2, \tag{25}$$

where $\mathrm{expit} \coloneqq 1/[1 + \exp(-x)]$ is the logistic function, and $\varepsilon > 0$ is a small constant [56]. In the limit $\varepsilon \to 0$, $\mathrm{expit}\,(\varepsilon^{-1}(\cdot))$ approximates the step function from 0 to 1. The latent function is similar to the level set function in level set inversion [57], but instead of using thresholding to obtain the piecewise-constant field, here we pass the latent function through the logistic function.

To compute the PICKLE cKLI-θ estimate of y, we first construct a cKLE for the latent field f from sparse measurements of y, which we accomplish via GP classification [14]. As the latent field is not observed directly, we cannot use the GPR Eqs. (6) and (7) to construct the conditional GP model $\hat{f}^c(x, \omega)$.

Instead, we proceed as follows. The observations $\mathbf{y}_\mathrm{s} = (y_1, \ldots, y_{N_\mathrm{s}^y})^\top$ are translated into a vector of binary values $\mathbf{b}_\mathrm{s} = (b_i, \ldots, b_{N_\mathrm{s}^y})^\top$, where $b_i = 0$ and $b_i = 1$ indicate $y_i = y_2$ and $y_i = y_1$, respectively. These binary observations are employed to construct the logistic GP classifier $\hat{f}^c(x, \omega)$, corresponding to the random field $\hat{f}$ conditioned on the outcomes b_i of the Bernoulli random variables (e.g., random variables with binary outcome) with probability of $b = 1$ given by $\mathrm{expit}(\hat{f}(X_i^y))$[1], that is,

$$b_i \sim \mathrm{Bernoulli}(\mathrm{expit}(\hat{f}(X_i^y))), \quad i \in [1, N_\mathrm{s}^y].$$

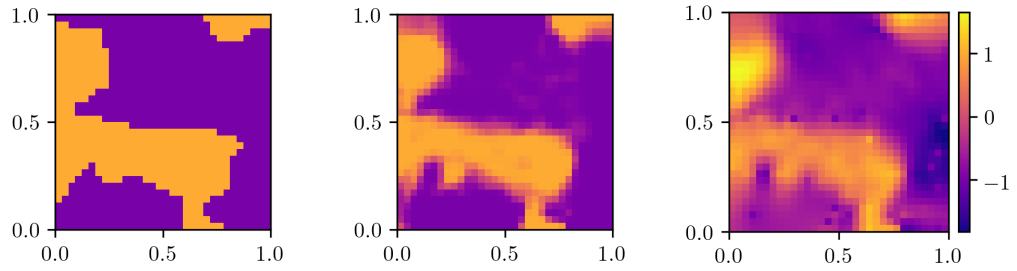
The conditioning is performed using the expectation propagation algorithm as implemented by the library GPy [53]. Once the conditional mean and covariance have been estimated, we then compute the cKLE of $\hat{f}^c$.

Next, we construct the sampling-based covariance model for u by using Algorithm 1. The realizations $\{y^{(i)}\}$ are generated by sampling fields $\{f^{(i)}\}$ from the cKLE of $\hat{f}^c$, which are then substituted into Eq. (25). Once the conditional covariance of u is found, the PICKLE estimate of f (and of y through Eq. (25)) is computed using Algorithm 2.

Fig. 5 shows the reference binary y field and the PICKLE cKLI-θ and MAP estimates of y using 25 measurements of y and 100 measurements of u. The reference f field is generated as a realization of the zero-mean Gaussian process with the isotropic Matérn ($\nu = 5/2$) kernel, $\sigma = 1.0$, and $\lambda = 0.2$. The reference y field is generated by substituting the reference f

[1]Note that GPR, Eqs. (6) and (7), can be understood in the same terms. Specifically, $\hat{y}^c$ is equivalent to the random field $\hat{f}$ conditioned on the outcomes $\mathbf{y}_\mathrm{s}$ of the random variable $\mathcal{N}(\hat{f}(X_\mathrm{s}^y), \Sigma)$.

field into Eq. (25) with $\epsilon = 100$. As before, the reference u field is generated by solving Eq. (23) for the reference y field. It can be seen that the PICKLE cKLI-θ estimate of y is closer to the reference y and has a significantly sharper boundary between the "y_1" and "y_2" regions than the MAP estimated y. The relative l_1 error of the PICKE cKLI-θ estimate of y is 0.179, more than two times smaller than the MAP estimation error of 0.380. These results indicate that PICKLE estimation can be employed to estimate discontinuous fields by expressing these fields in terms of cKLEs of continuous latent fields.



(a) Ref. (b) cKLI-θ (c) MAP

Figure 5: Reference piecewise-continuous log-diffusion field, and estimates computed using cKLI-θ and MAP. Relative ℓ_1 error of cKLI-θ is 0.179 and of MAP is 0.380.

5. Conclusions

We presented a new physics-informed machine learning approach, termed PICKLE, for learning parameters and states of stationary physical systems from sparse measurements constrained by the stationary PDE models governing the behavior of said systems. In PICKLE, parameters and states are approximated using cKLEs, i.e., KLEs conditioned on measurements, resulting in low-dimensional models of spatial fields that honor observed data.

Finally, the coefficients in the cKLEs are estimated by minimizing the norm of the residual of the PDE model evaluated at a finite set of points in the computational domain, ensuring that the reconstructed parameters and states are consistent with both the observations and the PDE model to an arbitrary level of accuracy.

The cKLEs are constructed using the eigendecomposition of covariance models of spatial variability. For the model parameter (space-dependent diffusion coefficient), we employed a parameterized covariance model calibrated on parameter observations; for the model state, the covariance was estimated from a number of forward simulations of the PDE model corresponding to realizations of the parameter drawn from its cKLE. We demonstrated that the accuracy of the PICKLE method depends on the accuracy of the estimated parameter covariance, which in turn depends on the number of measurements. It is important to note that transfer learning could be used to estimate the covariance of parameters, e.g., measurements collected in other systems with statistically similar properties can be used to estimate the covariance function of the model parameters.

We applied PICKLE to solve an inverse problem associated with the steady-state diffusion equation with unknown space-dependent diffusion coefficient. Specifically, we used PICKLE to estimate the log-diffusion coefficient from sparse measurements of the log-diffusion coefficient and the state of the system. We considered continuous and discontinuous diffusion coefficients. For continuous diffusion coefficients with different degrees of roughness (corresponding to different covariance kernels and correlation lengths), we demonstrated that the PICKLE estimates of the diffusion coefficient are

more accurate than those of the MAP and PINN methods. The comparison with the PINN method suggests that cKLEs are better representations of sparsely measured and spatially correlated fields than neural networks. We also found that PICKLE provides a better estimate of the discontinuous conductivity field than the MAP method.

Our results indicate that the PICKLE method can be used for estimating space-dependent parameters and states regardless of their underlying statistical distribution. Even though the cKLE expansion in PICKLE is constructed using the GPR estimates of the mean and covariance functions, we demonstrated that accurate estimates can be obtained when cKLE is used to model fields with highly non-Gaussian statistics, including the solution of the diffusion equation on the bounded domain and the discontinuous diffusion coefficient.

Acknowledgments

This work was supported by the Applied Mathematics Program within the U.S. Department of Energy Office of Advanced Scientific Computing Research. Pacific Northwest National Laboratory is operated by Battelle for the DOE under Contract DE-AC05-76RL01830.

[1] J. A. Vrugt, P. H. Stauffer, T. Wöhling, B. A. Robinson, V. V. Vesselinov, Inverse modeling of subsurface flow and transport properties: A review with new developments, Vadose Zone Journal 7 (2) (2008) 843–864.

[2] A. Golmohammadi, M.-R. M. Khaninezhad, B. Jafarpour, Exploiting sparsity in solving pde-constrained inverse problems: Application to

subsurface flow model calibration, in: Frontiers in PDE-Constrained Optimization, Springer, 2018, pp. 399–434.

[3] A. H. Elsheikh, I. Hoteit, M. F. Wheeler, Efficient bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates, Computer Methods in Applied Mechanics and Engineering 269 (2014) 515–537.

[4] H. W. Engl, K. Kunisch, A. Neubauer, Convergence rates for tikhonov regularisation of non-linear ill-posed problems, Inverse problems 5 (4) (1989) 523.

[5] A. M. Stuart, Inverse problems: A bayesian perspective, Acta Numerica 19 (2010) 451559. doi:10.1017/S0962492910000061.

[6] X. Ma, N. Zabaras, An efficient bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method, Inverse Problems 25 (3) (2009) 035013.

[7] Y. M. Marzouk, H. N. Najm, Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems, Journal of Computational Physics 228 (6) (2009) 1862–1902.

[8] M. Burger, F. Lucka, Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper bayes estimators, Inverse Problems 30 (11) (2014) 114004. doi:10.1088/0266-5611/30/11/114004.

[9] D. A. Barajas-Solano, B. E. Wohlberg, V. V. Vesselinov, D. M. Tartakovsky, Linear functional minimization for inverse modeling, Water Resour. Res. 51 (2014) 4516–4531. doi:10.1002/2014WR016179.

[10] J. Doherty, Ground water model calibration using pilot points and regularization, Groundwater 41 (2) (2003) 170–177.

[11] A. Alcolea, J. Carrera, A. Medina, Pilot points method incorporating prior information for solving the groundwater flow inverse problem, Advances in water resources 29 (11) (2006) 1678–1689.

[12] B. S. RamaRao, A. M. LaVenue, G. De Marsily, M. G. Marietta, Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. theory and computational experiments, Water Resources Research 31 (3) (1995) 475–493.

[13] M. L. Stein, Interpolation of spatial data: some theory for kriging, Springer Series in Statistics, Springer-Verlag, New York, 1999.

[14] C. K. Williams, C. E. Rasmussen, Gaussian processes for machine learning, Vol. 2, MIT press Cambridge, MA, 2006.

[15] N. Cressie, The origins of kriging, Mathematical geology 22 (3) (1990) 239–252.

[16] H. Gunes, S. Sirisup, G. E. Karniadakis, Gappy data: To krig or not to krig?, Journal of Computational Physics 212 (1) (2006) 358–382.

[17] N. A. C. Cressie, Geostatistics, John Wiley & Sons, Inc., 2015, pp. 27–104. doi:10.1002/9781119115151.ch2.

[18] X. Chen, G. E. Hammond, C. J. Murray, M. L. Rockhold, V. R. Vermeul, J. M. Zachara, Application of ensemble-based data assimilation techniques for aquifer characterization using tracer data at hanford 300 area, Water Resources Research 49 (10) (2013) 7064–7076.

[19] G. Evensen, Data assimilation: the ensemble Kalman filter, Springer Science & Business Media, 2009.

[20] M. Camporese, C. Paniconi, M. Putti, P. Salandin, Ensemble kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow, Water Resources Research 45 (10).

[21] C. Schillings, A. M. Stuart, Analysis of the ensemble kalman filter for inverse problems, SIAM Journal on Numerical Analysis 55 (3) (2017) 1264–1290. arXiv:https://doi.org/10.1137/16M105959X, doi:10.1137/16M105959X.
URL https://doi.org/10.1137/16M105959X

[22] T. Xu, J. J. Gmez-Hernndez, Simultaneous identification of a contaminant source and hydraulic conductivity via the restart normal-score ensemble kalman filter, Advances in Water Resources 112 (2018) 106 – 123. doi:https://doi.org/10.1016/j.advwatres.2017.12.011.
URL http://www.sciencedirect.com/science/article/pii/S030917081730756X

[23] D. McLaughlin, Recent developments in hydrologic data assimilation, Reviews of Geophysics 33 (S2) (1995) 977–984.

[24] X. Yang, D. A. Barajas-Solano, G. Tartakovsky, A. M. Tartakovsky, Physics-informed cokriging: A gaussian-process-regression-based multifidelity method for data-model convergence, J. Comput. Phys. 395 (2019) 410–431. doi:10.1016/j.jcp.2019.06.041.

[25] G. Dagan, S. P. Neuman, Subsurface flow and transport: a stochastic approach, Cambridge University Press, 2005.

[26] D. A. Barajas-Solano, A. M. Tartakovsky, Approximate bayesian model inversion for pdes with heterogeneous and state-dependent coefficients, J. Comput. Phys. 395 (2019) 247–262. doi:10.1016/j.jcp.2019.06.010.

[27] A. Beskos, M. Girolami, S. Lan, P. E. Farrell, A. M. Stuart, Geometric mcmc for infinite-dimensional inverse problems, Journal of Computational Physics 335 (2017) 327 – 351. doi:https://doi.org/10.1016/j.jcp.2016.12.041.
URL http://www.sciencedirect.com/science/article/pii/S0021999116307033

[28] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics informed deep learning (part ii): Data-driven discovery of nonlinear partial differential equations, arXiv preprint arXiv:1711.10566.

[29] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations, arXiv preprint arXiv:1711.10561.

[30] M. Raissi, Deep hidden physics models: Deep learning of nonlinear partial differential equations, arXiv preprint arXiv:1801.06637.

[31] A. M. Tartakovsky, C. O. Marrero, P. Perdikaris, G. D. Tartakovsky, D. Barajas-Solano, Learning parameters and constitutive relationships with physics informed deep neural networks, arXiv preprint arXiv:1808.03398v2.

[32] R. Tipireddy, D. A. Barajas-Solano, A. M. Tartakovsky, Conditional karhunen-loève expansion for uncertainty quantification and active learning in partial differential equation models, arXiv preprint arXiv:1904.08069.

[33] S. Huang, S. Quek, K. Phoon, Convergence study of the truncated karhunen–loeve expansion for simulation of stochastic processes, International journal for numerical methods in engineering 52 (9) (2001) 1029–1043.

[34] P. D. Spanos, R. Ghanem, Stochastic finite element expansion for random media, Journal of engineering mechanics 115 (5) (1989) 1035–1053.

[35] Y. Chen, A. Wiesel, A. O. Hero, Shrinkage estimation of high dimensional covariance matrices, in: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 2937–2940. doi:10.1109/ICASSP.2009.4960239.

[36] M. B. Giles, Multilevel monte carlo methods, Acta Numerica 24 (2015) 259328. doi:10.1017/S096249291500001X.

[37] K. D. Jarman, A. M. Tartakovsky, A comparison of closures for stochastic advection-diffusion equations, SIAM/ASA Journal on Uncertainty Quantification 1 (1) (2013) 319–347. doi:10.1137/120897419.

[38] D. M. Tartakovsky, Z. Lu, A. Guadagnini, A. M. Tartakovsky, Unsaturated flow in heterogeneous soils with spatially distributed uncertain hydraulic parameters, Journal of Hydrology 275 (3) (2003) 182 – 193.

[39] P. M. Tagade, H.-L. Choi, Mitigating gibbs phenomena in uncertainty quantification with a stochastic spectral method, Journal of Verification, Validation and Uncertainty Quantification 2 (1) (2017) 011003.

[40] J. Nolen, G. Papanicolaou, Fine scale uncertainty in parameter estimation for elliptic equations, Inverse Problems 25 (11) (2009) 115021. doi:10.1088/0266-5611/25/11/115021.
URL https://doi.org/10.1088%2F0266-5611%2F25%2F11%2F115021

[41] H. Zhang, S. Abhyankar, E. Constantinescu, M. Anitescu, Discrete adjoint sensitivity analysis of hybrid dynamical systems with switching, IEEE Transactions on Circuits and Systems I: Regular Papersdoi:10.1109/TCSI.2017.2651683.

[42] M. B. Giles, M. C. Duta, J.-D. Müller, N. A. Pierce, Algorithm developments for discrete adjoint methods, AIAA Journal 41 (2) (2003) 198–205.

[43] I. Sraj, O. P. L. Matre, O. M. Knio, I. Hoteit, Coordinate transformation and polynomial chaos for the bayesian inference of a gaussian

process with parametrized prior covariance function, Computer Methods in Applied Mechanics and Engineering 298 (2016) 205 – 228. doi:https://doi.org/10.1016/j.cma.2015.10.002.
URL http://www.sciencedirect.com/science/article/pii/S0045782515003217

[44] Q. Liu, X. Zhang, A chebyshev polynomial-based galerkin method for the discretization of spatially varying random properties, Acta Mechanica 228 (6) (2017) 2063–2081. doi:10.1007/s00707-017-1819-2.
URL https://doi.org/10.1007/s00707-017-1819-2

[45] C. Schwab, R. A. Todor, Karhunenlove approximation of random fields by generalized fast multipole methods, Journal of Computational Physics 217 (1) (2006) 100 – 122, uncertainty Quantification in Simulation Science. doi:https://doi.org/10.1016/j.jcp.2006.01.048.
URL http://www.sciencedirect.com/science/article/pii/S0021999106000349

[46] B. N. Khoromskij, A. Litvinenko, H. G. Matthies, Application of hierarchical matrices for computing the karhunen–loève expansion, Computing 84 (1) (2009) 49–67.

[47] J. Bear, Dynamics of fluids in porous media, Courier Corporation, 2013.

[48] K.-A. Lie, An Introduction to Reservoir Simulation Using MATLAB/GNU Octave: User Guide for the MATLAB Reservoir Simulation Toolbox (MRST), Cambridge University Press, 2019. doi:10.1017/9781108591416.

[49] J. J. Moré, The levenberg-marquardt algorithm: Implementation and theory, in: G. A. Watson (Ed.), Numerical Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg, 1978, pp. 105–116.

[50] T. E. Oliphant, Python for scientific computing, Computing in Science Engineering 9 (3) (2007) 10–20. doi:10.1109/MCSE.2007.58.

[51] O. Tange, Gnu parallel - the command-line power tool, ;login: The USENIX Magazine 36 (1) (2011) 42–47. doi:10.5281/zenodo.16303.
URL http://www.gnu.org/s/parallel

[52] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, software available from tensorflow.org (2015).
URL https://www.tensorflow.org/

[53] GPy, GPy: A gaussian process framework in python, http://github.com/SheffieldML/GPy (since 2012).

[54] M. Armstrong, A. Galli, H. Beucher, G. Loc'h, D. Renard, B. Doligez, R. Eschard, F. Geffroy, Plurigaussian Simulations in Geosciences, Springer-Verlag Berlin Heidelberg, 2011.

[55] A. Astrakova, D. S. Oliver, Conditioning truncated pluri-gaussian models to facies observations in ensemble-kalman-based data assimilation, Mathematical Geosciences 47 (3) (2015) 345–367.

[56] S. Menard, Applied logistic regression analysis, Vol. 106, Sage, 2002.

[57] M. M. Dunlop, M. A. Iglesias, A. M. Stuart, Hierarchical bayesian level set inversion, Statistics and Computing 27 (6) (2017) 1555–1584.