Stochastic Environmental Research and Risk Assessment
https://doi.org/10.1007/s00477-019-01709-8

ORIGINAL PAPER

Gaussian process for estimating parameters of partial differential equations and its application to the Richards equation

Pankaj Kumar Rai[1] · Shivam Tripathi[1]

© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

This paper proposes a new collocation method for estimating parameters of a partial differential equation (PDE), which uses Gaussian process (GP) as a basis function and is termed as Gaussian process for partial differential equation (GPPDE). The conventional method of estimating parameters of a differential equation is to minimize the error between observations and their estimates. The estimates are produced from the forward solution (numerical or analytical) of the differential equation. The conventional approach requires initial and boundary conditions, and discretization of differential equations if the forward solution is obtained numerically. The proposed method requires fitting a GP regression model to the observations of the state variable, then obtaining derivatives of the state variable using the property that derivative of a GP is also a GP, and finally adjusting the PDE parameters so that the GP derived partial derivatives satisfy the PDE. The method does not require initial and boundary conditions, however if these conditions are available (exactly or with measurement errors), they can be easily incorporated. The GPPDE method is evaluated by applying it on the diffusion and the Richards equations. The results suggest that GPPDE can correctly estimate parameters of the two equations. For the Richards equation, GPPDE performs well in the presence of noise. A comparison of GPPDE with HYDRUS-1D software showed that their performances are comparable, though GPPDE has significant advantages in terms of computational time. GPPDE could be an effective alternative to conventional approaches for finding parameters of high-dimensional PDEs where large datasets are available.

Keywords Gaussian process · Partial differential equation · Inverse problem · Diffusion equation · Richards equation · Parameter estimation

Abbreviations

AGM	Adaptive gradient matching
GM	Gradient matching
ASMO	Adaptive surrogate modeling based optimization
ASMO-PODE	ASMO-parameter optimization and distribution estimation
GP	Gaussian process
GPODE	Gaussian process ordinary differential equation
GPPDE	Gaussian process for partial differential equation
ODE	Ordinary differential equation
PDE	Partial differential equation

List of symbols

α	Air-entry pressure head soil hydraulic parameter
λ	Diffusivity
$\boldsymbol{\lambda}$	Set of parameters of PDE
$\boldsymbol{\phi}$	$[l, \sigma_s]$
$\boldsymbol{\Phi}$	$[l, \sigma_s, \sigma_y]$
σ_s	Hyperparameter of GP (scale parameter of kernel function)
σ_y	Hyperparameter of GP (noise in GP model)
θ	Volumetric soil moisture
θ_r	Residual soil moisture
θ_s	Saturated soil moisture
Θ	Relative saturation; $\frac{\theta - \theta_r}{\theta_s - \theta_r}$
$\mathbf{I}$	Identity matrix
$I(\boldsymbol{\lambda})$	Fisher information matrix

✉ Shivam Tripathi
shiva@iitk.ac.in

Pankaj Kumar Rai
pankajrai87@gmail.com

[1] Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur, India

Published online: 26 July 2019

k	Soil hydraulic conductivity
k_s	Saturated soil hydraulic conductivity
$k(\cdot)$	Covariance/kernel function
$\mathbf{K}$	Covariance matrix
l	Hyperparameter of GP (length scale of kernel function)
m	Number of independent variables
m_y	Mean of Gaussian process
n	Pore-size distribution soil hydraulic parameter
N	Number of observations
$\mathbb{R}$	Real space
t	Temporal dimension
x	Independent variable/a spatial dimension
$\mathbf{x}$	Vector of training independent variables
$\mathbf{x}'$	Vector of testing independent variables
$\mathbf{X}$	Design matrix of dimension $N \times m$
y	State variable
$\mathbf{y}$	Column vector of state variable
$\tilde{y}$	Observation
$\tilde{\mathbf{y}}$	Column vector of N observations
z	A spatial dimension

1 Introduction

A differential equation relates a function with its derivatives (Newton 1736). The function and its derivatives represent a state variable and its rate of change, respectively. The differential equations have important application in science, engineering, economics and biology (Strang 2014; Vyshemirsky and Girolami 2007). Ordinary Differential Equations (ODEs) contain one or more functions of an independent variable and their derivatives, whereas Partial Differential Equations (PDEs) contain multivariate functions of independent variables and their partial derivatives.

In hydraulics and hydrology, differential equations are often used to represent conservation equations. The conventional method for estimating parameters of a differential equation is to minimize an error function [e.g. L1-Norm, L2-Norm or L-infinity-Norm of the error; though use of L2-Norm is more common (Kool et al. 1985)], where error is the difference between the observations and their estimates. The estimates are obtained from the forward solution (numerical or analytical) of the differential equation. This method of parameter estimation is referred to as inverse problem. There are many methods of error minimization (Rao 2009; Deb 2012), but gradient based methods are usually preferred (McLaughlin and Townley 1996; Simunek et al. 2005).

The conventional approach of parameter estimation requires initial and boundary conditions (which are not readily available and difficult to impose in forward solutions), and discretization of differential equations if the forward solution is obtained numerically. Further, the computational time of the conventional approach increases significantly with increase in the dimensionality of the state variable and the number of unknown parameters. If the initial and boundary conditions are not available, they are considered as parameters in the conventional approach, augmenting the parameter set, which further increases their computational burden. Moreover, sometimes the data may not be sufficient to estimate the augmented parameter set.

A number of studies have attempted to estimate the parameters without explicitly solving differential equations. These methods belong to the general class of collocation methods (Varah 1982; Poyton et al. 2006; Ramsay et al. 2007). The collocation methods for parameter estimation of differential equation require fitting of either a polynomial or a spline to the observations of the state variable, and adjusting the differential equation parameters such that the polynomial spline interpolated values of the state variable and its derivatives satisfy the differential equation relationship. Since the collocation methods do not solve the forward problem, the initial and boundary conditions are not required (Ramsay et al. 2007). The Gaussian process (GP) based methods, which also fall under the category of collocation method, have been developed and successfully applied to estimate the parameters of ODEs (Calderhead et al. 2008; Dondelinger et al. 2013; Wang and Barber 2014; Macdonald and Husmeier 2015). These methods require fitting a GP regression model to the observations of the state variable and estimating their derivatives using the property that derivative of a GP is also a GP, and finally adjusting the ODE parameters so that the GP derived derivatives satisfy the differential equations.

Calderhead et al. (2008) proposed the first GP based method for ODE parameter estimation and termed it gradient matching (GM) method. The GM method consists of two steps (1) fitting a GP regression model to the observed values of the state variable and evaluating the state variable and its derivatives using the fitted GP model, and (2) matching the estimates of the state derivative obtained from the GP model and obtained analytically from the given ODE relation by using the Product of Experts approach. A Product of Experts combines probability distributions of individual experts by taking their products (Mayraz and Hinton 2001; Hinton 2002). Thus in the GM method, there are two distributions for the state derivative, one from derivative of the fitted GP model and another from the ODE relation. Using these two distributions, the Product of Experts forms a new distribution. This new distribution is used to sample the parameters of the ODE.

The GM method lacks any feedback mechanism from the ODE back to the GP model. Dondelinger et al. (2013) proposed a sampling modification for the GM method, and

termed their proposed method as Adaptive Gradient Method (AGM). In the AGM method, sampling of the GP regression model parameters and the ODE parameters are not done in two stages, rather these parameters are simultaneously sampled. The posterior distribution for the GP model and the ODE parameters are obtained by marginalizing over the state derivative and conditioning on the state variable. Sampling starts by initializing the GP parameters and then the ODE parameters are sampled using the Product of Experts. Next, fixing the ODE parameters to their estimated values, GP parameters are sampled using the Product of Experts. The sampling is repeated until the GP and the ODE parameters converge.

Wang and Barber (2014) proposed an alternate approach to AGM which they termed Gaussian process ordinary differential equation (GPODE). The GM and AGM methods were based on the idea that derivative of a GP is a GP, but the GPODE approach harnesses the fact that the state variable can be represented as a GP if the state derivative is a GP (in other words, the integral of a GP is also a GP). In the GPODE method, the posterior distribution of the unknown parameters is obtained by marginalizing over the state variable and conditioning on the state derivatives. Similar to the AGM method, the GPODE method also samples the GP and the ODE parameters simultaneously, rather than in two steps. A critical review of GPODE by Macdonald et al. (2015) points out an anomaly in the graphical representation of the method and its logical construct. Specifically, they demonstrated that GPODE suffers from intrinsic identifiability problem. Readers are referred to Macdonald et al. (2015) for further details.

The GP based gradient matching methods discussed above are applicable only to ODEs, where the state derivatives can be expressed as an explicit function of the state variable and the independent variables. However, for many ODEs, the state derivatives can not be expressed as an explicit function and it would not be possible to calculate the state derivative analytically as done in the GP based gradient matching methods (GM, AGM and GPODE have been demonstrated on the ODEs: Lotka–Volterra and Signal Transduction Cascade; wherein the gradients are explicit functions of state variables). Similarly for the PDEs, the partial derivatives of the state variable are not explicit functions and hence these GP based methods can not be applied to find PDE parameters. This study proposes Gaussian process for partial differential equation (GPPDE) method which addresses the above mentioned limitation.

Recently Raissi et al. (2017b) have proposed a GP based method for estimating the parameters of linear non-homogeneous PDEs. Their method is based on the ideas presented by Graepel (2003) and Särkkä (2011) to encode physics of a differential equation into the covariance function of the GP regression model. Raissi et al. (2017b) method assumes that the state variable and the forcing term in a differential equation are GPs. Since a linear operation on a GP results in another GP (Solak et al. 2003), the cross-covariance function of the state variable and the forcing term can be expressed in terms of the parameters of the differential equation. Thus, the parameters of the differential equation are turned into parameters of the GP which are estimated when fitting the GP regression model to the observations of the state variable and the forcing term. A limitation of this method is that it cannot be applied to nonlinear differential equations and to equations in which the forcing term is constant. Raissi et al. (2017a, 2018) used the same method for solving differential equations using noisy multi-fidelity data for linearized PDE. In addition, Cockayne et al. (2016) and Bilionis (2016) have also discussed the possibility of approximating the solutions of linear PDEs using GP.

In the context of Bayesian parameter estimation, GP regression model has been used as a surrogate to dynamical systems described by differential equations. In hydrology, Wang et al. (2014) proposed an Adaptive Surrogate Modeling based Optimization (ASMO) method that uses GP regression as a surrogate model and demonstrated its applicability by estimating parameters of Hartman Function and Sacramento Soil Moisture Accounting hydrological model. Gong and Duan (2017) proposed ASMO-Parameter Optimization and Distribution Estimation (ASMO-PODE) method as an improvement over ASMO method. The ASMO-PODE method uses an adaptive resampling strategy that significantly reduces the computation cost of estimating parameters. They demonstrated the effectiveness of ASMO-PODE by estimating parameters of a land surface model. Zhang et al. (2016) used GP regression to construct a locally accurate surrogate model and showed that it significantly reduces the computational cost for groundwater contaminant source identification problem.

The use of GP regression as a surrogate model in the above cited studies is different from the way it is used in the proposed GPPDE method. The surrogate models are trained on the forward simulations of the differential equations, and hence a method for forward solution is required. The surrogate GP models do not explicitly satisfy initial and boundary conditions, and it is not straight forward to account for measurement errors and mismatch between forward solution and GP regression in the parameter estimation. GPPDE, on the other hand, uses the gradient matching approach in which the estimates of the state variable and its derivatives are directly used to approximate parameters of the dynamical systems described by PDEs. It relies on the key observation that any linear transformation of a GP, such as differentiation and integration is still a GP. Since GP regression model in

GPPDE is trained on observations, a forward solution of PDE is not required. The method can work even when initial and boundary conditions are not available, however, if these conditions are available, GPPDE can incorporate them. Further, if measurement errors are known they can be used in training the GP regression model.

To the best of our knowledge gradient matching methods for estimating parameters of PDE have never been used in water resources. The purpose of this paper is to propose a new GP based method for estimating parameters of both linear and non-linear PDEs and demonstrate its application. The application of the new GPPDE method is demonstrated on two PDEs (the groundwater diffusion and the Richards equations) using synthetic data. The diffusion equation is a parabolic PDE that describes the spatio-temporal variation of groundwater level in homogeneous aquifers. The Richards equation is a nonlinear parabolic PDE that describes the movement of water in unsaturated zone. Flow in unsaturated zone is one of the most complex flows in nature. The numerical solution of the Richards equation is computationally expensive, and a universally robust and accurate solution methodology is yet to be discovered (Farthing and Ogden 2017).

This paper is organized in six sections: the first section discusses differential equations and the conventional approach to estimate their parameters. The section also reviews the GP based parameter estimation methods for differential equations, presents their advantages over the conventional approach and discusses their limitations. The second section provides mathematical formulation of the proposed GPPDE method for estimating parameters of PDEs. The third section briefly explains the groundwater diffusion equation and the Richards equation. The fourth section describes the data used to demonstrate the GPPDE method. The fifth section presents and discusses the results from the GPPDE method. The sixth section summarizes and concludes the paper.

2 Mathematical formulation

The objective of the proposed GPPDE method is to estimate parameters of a PDE, given the observations of state variable. GP, like other regression methods, does not rely on physics but on the data to fit, which it tends to statistically fit the response surface. That's why GPPDE method is able to estimate parameters of a PDE if the initial and boundary conditions are not available or available with measurement error. This section describes the mathematical formulation of the GPPDE method. In the following subsections, we will discuss the formulation and training of the GP regression model, provide expressions for its partial derivatives, explain the procedure for incorporating derivative information in the GP regression model, and present the methodology for estimating PDE parameters.

2.1 Gaussian process

Gaussian process (GP) is a kernel based non-parametric approach for supervised (regression and classification problem) and unsupervised (clustering problem) learning (Rasmussen and Williams 2006). In the field of geostatistics, GP was independently derived and termed as kriging (Krige 1951; Matheron 1963). In water resources, GP has been used as a regression model (Marrel et al. 2009; Sun et al. 2014; Holman et al. 2014; Zhang et al. 2016). A brief mathematical formulation of GP (using the properties of conditional Gaussian distribution) for a regression problem is provided below.

In this paper, the following notations are used. All vectors are column vectors and are represented by lower case bold letters. Matrices are represented by upper case bold letters. Italic letters represent scalars. Identity matrix is represented by $\mathbf{I}$.

GP assumes that the data points follow multivariate Gaussian distribution in which the covariance function is homogeneous and stationary, and depends only on the distance between the two points. A covariance function, also known as kernel, provides a measure of closeness between a training point ($\mathbf{x}$) and a test point ($\mathbf{x}'$), notationally given by $k(\mathbf{x}, \mathbf{x}')$. In the present study, Gaussian covariance function is used, which is given by

$$k(\mathbf{x}, \mathbf{x}') = {\sigma_s}^2 \exp\left(-\frac{1}{2} \sum_{j=1}^{m} \frac{\left(x_j - x_j'\right)^2}{l_j^2} \right); \tag{1}$$

where m is the dimension of the input space or the number of independent variables in $\mathbf{x}$, $x_j \in \mathbf{x}$, l_j is the length scale for the jth independent variable x_j, and σ_s is the scale parameter of the covariance function. Together, they are referred as hyperparameters and are represented by $\boldsymbol{\phi} = [l_1, \ldots, l_m, \sigma_s]$. The Gaussian covariance function, also known as squared exponential covariance function, is infinitely differentiable and hence very smooth. It is probably the most widely used covariance function in GP regression models. The choice of covariance function depends on the physical process being modelled. Readers are referred to Chapter 4 of Rasmussen and Williams (2006) for a detailed discussion on other covariance functions and their properties.

For the formulation of GP, the conditional property of the Gaussian distribution is used, i.e., if two sets of random variables are jointly Gaussian, then the distribution of one set conditioned on the other set is also Gaussian

(section 2.3 of Bishop (2006); Appendix A.3 of Rasmussen and Williams (2006)). Let the training dataset be given by, $\mathcal{D} = \{(\mathbf{x}_i, \tilde{y}_i) \mid i = 1, \ldots, N; \mathbf{x}_i \in \mathbb{R}^m; \tilde{y}_i \in \mathbb{R}\}$, where, $\mathbf{x}_i$ is an input vector (independent variable) in m-dimensional space, $\tilde{y}_i$ is the corresponding observed value of the dependent variable and N is the number of observations. All training input vectors are arranged in $N \times m$ training data matrix ($\mathbf{X}$) and the observations are collectively represented by $N \times 1$ vector, $\tilde{\mathbf{y}}$. The observations have independent and identically distributed noise, which is assumed to be normally distributed with mean zero and standard deviation σ_y

$$\tilde{y}_i = y_i + \mathcal{N}\left(0, \sigma_y^2\right) \tag{2}$$

where y_i is the true value of the state variable at location $\mathbf{x}_i$ and all y_i's are collected in vector $\mathbf{y}$. The vector of the state variable $\mathbf{y}$, follows a Gaussian distribution with mean $m_y(\mathbf{X})$ and covariance matrix $\mathbf{K}$, represented as

$$\mathbf{y} \sim \mathcal{N}\left(m_y(\mathbf{X}), \mathbf{K}\right) \tag{3}$$

where $\mathbf{y}$ is $N \times 1$ vector, $\mathbf{K} = k(\mathbf{X}, \mathbf{X})$ is a $N \times N$ positive definite covariance matrix and $m_y(\mathbf{X})$ is $N \times 1$ vector of the mean of GP at training data points $\mathbf{X}$ (also referred as a priori mean function). In most cases the a priori mean may not be available, but the GP model is powerful enough to capture trends in data even if the mean of GP is taken as zero or a constant value (Rasmussen and Williams 2006).

The GP model prediction ($\mathbf{y}_*$) at new test points ($\mathbf{X}_*$) is given by

$$\mathbf{y}_* \sim \mathcal{N}\left(m_y(\mathbf{X}_*), \mathbf{K}_{**}\right) \tag{4}$$

where $\mathbf{K}_{**} = k(\mathbf{X}_*, \mathbf{X}_*)$ and $\mathbf{X}_*$ is the testing data matrix, in which, testing points are arranged in rows. The joint distribution of $\tilde{\mathbf{y}}$ and $\mathbf{y}_*$ is given by Eq. (5), in which, $m_y(\mathbf{X})$ and $m_y(\mathbf{X}_*)$ are the vectors of mean values, evaluated for design matrix $\mathbf{X}$ and test matrix $\mathbf{X}_*$, respectively, and $\mathbf{K}_* = k(\mathbf{X}, \mathbf{X}_*)$ (The elements of $\mathbf{K}_*, k_{ij}$ are covariance between the ith training point and the jth test point)

$$\begin{bmatrix} \tilde{\mathbf{y}} \\ \mathbf{y}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_y(\mathbf{X}) \\ m_y(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K} + \sigma_y^2 \mathbf{I} & \mathbf{K}_* \\ \mathbf{K}_*^{\mathrm{T}} & \mathbf{K}_{**} \end{bmatrix} \right). \tag{5}$$

The predictive distribution $p\left(\mathbf{y}_* | \mathbf{X}_*, \boldsymbol{\phi}, \sigma_y, \mathbf{X}, \tilde{\mathbf{y}}\right)$, based on the conditional property of the Gaussian distribution, is given by

$$p\left(\mathbf{y}_* | \mathbf{X}_*, \boldsymbol{\phi}, \sigma_y, \mathbf{X}, \tilde{\mathbf{y}}\right) = \mathcal{N}(\mathbf{y}_* | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{6}$$

where

$$\begin{aligned} \boldsymbol{\mu} &= \mathbb{E}\left[\mathbf{y}_* | \mathbf{X}_*, \boldsymbol{\phi}, \sigma_y, \mathbf{X}, \tilde{\mathbf{y}}\right] \\ &= m_y(\mathbf{X}_*) + \mathbf{K}_*^{\mathrm{T}} \left(\mathbf{K} + \sigma_y^2 \mathbf{I}\right)^{-1} \left(\tilde{\mathbf{y}} - m_y(\mathbf{X})\right) \end{aligned} \tag{7a}$$

$$\begin{aligned} \boldsymbol{\Sigma} &= \mathrm{cov}\left(\mathbf{y}_* | \mathbf{X}_*, \boldsymbol{\phi}, \sigma_y, \mathbf{X}, \tilde{\mathbf{y}}\right) \\ &= \mathbf{K}_{**} - \mathbf{K}_*^{\mathrm{T}} \left(\mathbf{K} + \sigma_y^2 \mathbf{I}\right)^{-1} \mathbf{K}_* + \sigma_y^2 \mathbf{I} \end{aligned} \tag{7b}$$

2.2 Estimation of hyperparameters

Estimation of hyperparameters is still an open problem and there are many approaches for it (Rasmussen and Williams 2006). Generally, the nomenclature hyperparameter is used for the parameters of a covariance function, but in the present study, the noise σ_y has also been considered as one of the hyperparameters. Collectively they are represented as, $\Phi = \left[l_1, \ldots, l_m, \sigma_s, \sigma_y\right]$ and termed as hyperparameters of GP. In the present study, the approach of maximum marginal likelihood (Rasmussen and Williams 2006) is used for estimating the hyperparameters. A probabilistic graph for the GP regression model is given in Fig. 1a. The marginal likelihood function is obtained by marginalizing (integrating out) the state variable $\mathbf{y}$ from the joint distribution of $\tilde{\mathbf{y}}$ and $\mathbf{y}$, given the hyperparameters

$$\begin{aligned} p\left(\tilde{\mathbf{y}} | \mathbf{X}, \boldsymbol{\phi}, \sigma_y\right) &= \int p\left(\tilde{\mathbf{y}}, \mathbf{y} | \mathbf{X}, \boldsymbol{\phi}, \sigma_y\right) d\mathbf{y} \\ &= \int p\left(\tilde{\mathbf{y}} | \mathbf{y}, \mathbf{X}, \sigma_y\right) p(\mathbf{y} | \mathbf{X}, \boldsymbol{\phi}) d\mathbf{y} \end{aligned} \tag{8}$$

where the likelihood $p\left(\tilde{\mathbf{y}} | \mathbf{y}, \mathbf{X}, \sigma_y\right)$ in above equation using Eq. (2) is given by

$$\begin{aligned} p\left(\tilde{\mathbf{y}} | \mathbf{y}, \mathbf{X}, \sigma_y\right) &= \mathcal{N}\left(\tilde{\mathbf{y}} | \mathbf{y}, \sigma_y^2 \mathbf{I}\right) \\ &= \frac{1}{(2\pi)^{N/2} \sigma_y^N} \exp\left\{ -\frac{1}{2\sigma_y^2} (\tilde{\mathbf{y}} - \mathbf{y})^{\mathrm{T}} (\tilde{\mathbf{y}} - \mathbf{y}) \right\} \end{aligned} \tag{9}$$

and $p(\mathbf{y} | \mathbf{X}, \boldsymbol{\phi})$ is a GP and from Eq. (3) it is given by

$$\begin{aligned} p(\mathbf{y} | \mathbf{X}, \boldsymbol{\phi}) = \mathcal{N}\left(\mathbf{y} | m_y(\mathbf{X}), \mathbf{K}\right) &= \frac{1}{(2\pi)^{N/2} |\mathbf{K}|^{1/2}} \\ &\times \exp\left\{ -\frac{1}{2} \left(\mathbf{y} - m_y(\mathbf{X})\right)^{\mathrm{T}} \mathbf{K}^{-1} \left(\mathbf{y} - m_y(\mathbf{X})\right) \right\}. \end{aligned} \tag{10}$$

Combining Eqs. (8)–(10) and taking logarithm of $p\left(\tilde{\mathbf{y}} | \mathbf{X}, \boldsymbol{\phi}, \sigma_y\right)$, gives

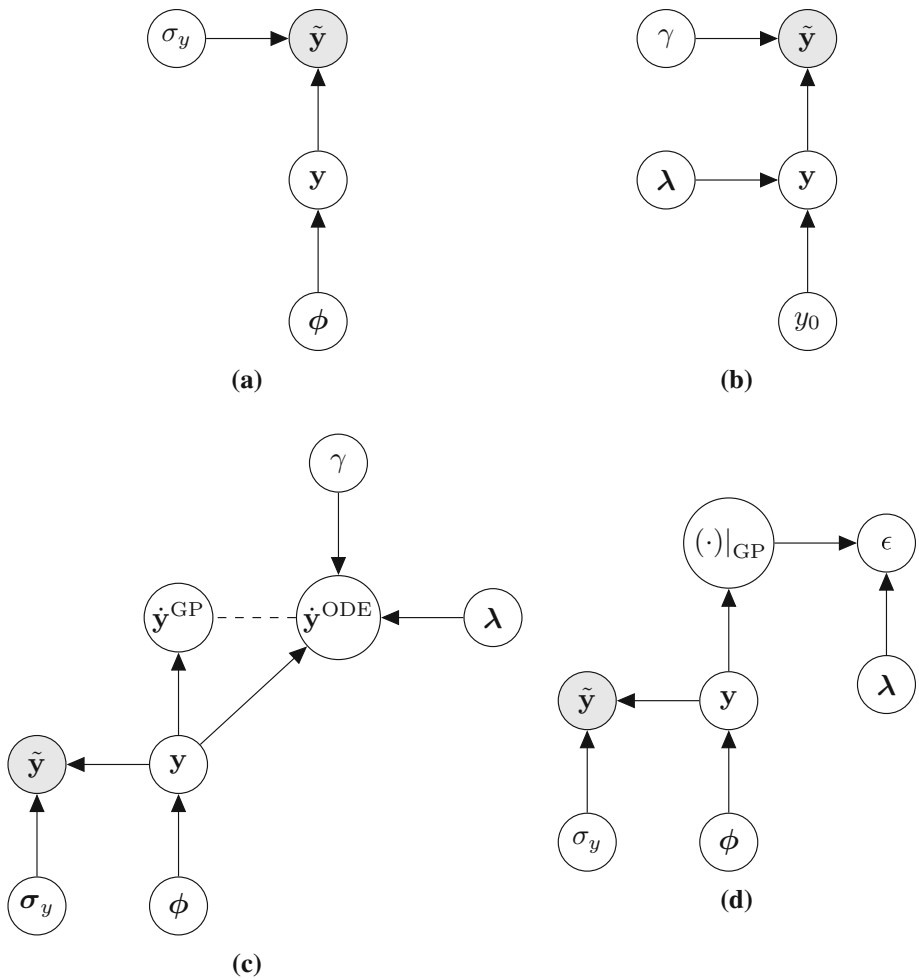

Fig. 1 Probabilistic graphs for **a** GP, **b** ODE, **c** the gradient matching (GM) model as proposed by Calderhead et al. (2008), where, γ is variance of the state derivative and $\boldsymbol{\lambda}$ represents parameters of ODE, **d** proposed GPPDE model. The shaded nodes represent observations of the state variable, the non-shaded nodes represent random variables, the directed links represent dependencies and the undirected dashed link represent Product of Experts between the two variables connected by the link

$$\log p(\tilde{\mathbf{y}}|\mathbf{X},\Phi) = -\frac{1}{2}\left(\tilde{\mathbf{y}} - m_y(\mathbf{X})\right)^{\mathrm{T}}\mathbf{K}_y^{-1}\left(\tilde{\mathbf{y}} - m_y(\mathbf{X})\right) - \frac{1}{2}\log|\mathbf{K}_y| - \frac{N}{2}\log 2\pi \tag{11}$$

where $\mathbf{K}_y = \mathbf{K} + \sigma_y^2\mathbf{I}$. The hyperparameters of GP, Φ, are estimated by maximizing the log marginal likelihood Eq. (11). Maximization using a gradient based approach requires the partial derivatives of Eq. (11) with respect to hyperparameters, which are given by

$$\begin{aligned}\frac{\partial}{\partial\Phi_i}\log p(\tilde{\mathbf{y}}|\mathbf{X},\Phi) &= \frac{1}{2}\left(\tilde{\mathbf{y}} - m_y(\mathbf{X})\right)^{\mathrm{T}}\mathbf{K}_y^{-1}\frac{\partial\mathbf{K}_y}{\partial\Phi_i}\mathbf{K}_y^{-1}\left(\tilde{\mathbf{y}} - m_y(\mathbf{X})\right) \\ &- \frac{1}{2}\mathrm{tr}\left(\mathbf{K}_y^{-1}\frac{\partial\mathbf{K}_y}{\partial\Phi_i}\right); i = 1,2,\ldots,m+2\end{aligned} \tag{12}$$

The derivatives of $\mathbf{K}_y$ with respect to the hyperparameters l_i, σ_s and σ_y are given in "Appendix 1" section. In this study, the log marginal likelihood function is maximized using Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Fletcher 2013; Nocedal and Wright 1999). The initial estimates of the length scales $(l_1,\ldots,l_m)$ is taken as the standard deviations of the predictors $(\mathbf{X})$, and the initial values for the scale hyperparameter (σ_s) and noise parameter (σ_y) are taken as the standard deviation of the state variable (y) divided by $\sqrt{2}$.

2.3 Derivatives of Gaussian process

The proposed GPPDE method requires derivatives of the state variable (fitted by a GP model) with respect to independent variables. Since derivative is a linear operator, the derivative of a GP is another GP (Solak et al. 2003). This section provides formulation of partial derivatives of GP at test points $(\mathbf{X}_*)$.

The GP relation Eq. (7a) could be straightforwardly differentiated with respect to test input variable (see "Appendix 2" section). For a constant mean GP, the first order derivative is given by $p(\dot{\mathbf{y}}^{\mathrm{GP}}|\mathbf{y},\Phi) = p(\dot{\mathbf{y}}^{\mathrm{GP}}|m_y,\Phi) = \mathcal{N}(\dot{\boldsymbol{\mu}},\dot{\boldsymbol{\Sigma}}^{\mathrm{GP}})$

$$\dot{\boldsymbol{\mu}} = \frac{\partial}{\partial x_i^*}\mathbb{E}\left[\mathbf{y}_*|\mathbf{X}_*, \boldsymbol{\phi}, \sigma_y, \mathbf{X}, \tilde{\mathbf{y}}\right] = \dot{\mathbf{K}}_*^{\mathrm{T}}\left(\mathbf{K}+\sigma_y^2\mathbf{I}\right)^{-1}\left(\tilde{\mathbf{y}} - m_y(\mathbf{X})\right) \tag{13a}$$

$$\dot{\boldsymbol{\Sigma}}^{\mathrm{GP}} = \dot{\dot{\mathbf{K}}}_{**} - \dot{\mathbf{K}}_*^{\mathrm{T}}\left(\mathbf{K} + \sigma_y^2\mathbf{I}\right)^{-1}\dot{\mathbf{K}}_* \tag{13b}$$

where $\dot{\boldsymbol{\Sigma}}^{\mathrm{GP}}$ is the covariance matrix of $\dot{\mathbf{y}}^{\mathrm{GP}}$ (Calderhead et al. 2008). Matrix $\dot{\mathbf{K}}_*$ is populated by mixed covariance function between the state variable and their first order partial derivatives, $\mathrm{cov}\left(y, \frac{\partial}{\partial x_{*_i}} y_*\right) = \frac{\partial}{\partial x_{*_i}} k(\mathbf{x}, \mathbf{x}_*)$, also known as cross-covariance between the state variable and its derivative. The elements of matrix $\dot{\dot{\mathbf{K}}}_{**}$ are mixed covariance function between partial derivatives $\mathrm{cov}\left(\frac{\partial}{\partial x_i} y, \frac{\partial}{\partial x_{*_j}} y_*\right) = \frac{\partial^2}{\partial x_i \partial x_{*_j}} k(\mathbf{x}, \mathbf{x}_*)$, also known as auto covariance of the state derivative.

The second order derivative of GP is given by (see "Appendix 2")

$$\ddot{\mathbf{y}}^{\mathrm{GP}} = \frac{\partial^2}{\partial x_i^{*2}}\mathbb{E}\left[\mathbf{y}_*|\mathbf{X}_*, \boldsymbol{\phi}, \sigma_y, \mathbf{X}, \tilde{\mathbf{y}}\right] = \ddot{\mathbf{K}}_*^{\mathrm{T}}\left(\mathbf{K}+\sigma_y^2\mathbf{I}\right)^{-1}\left(\tilde{\mathbf{y}} - m_y(\mathbf{X})\right) \tag{14}$$

where the matrix $\ddot{\mathbf{K}}_*$ is populated by mixed covariance function between the state variable and its second order partial derivative, $\mathrm{cov}\left(y, \frac{\partial^2}{\partial x_i^{*2}} y^*\right) = \frac{\partial^2}{\partial x_i^{*2}} k(\mathbf{x}, \mathbf{x}^*)$.

2.4 Incorporating derivatives information in GP

In some applications, along with the observations of the state variable, information about the derivatives of the state variable may also be available. To incorporate derivative information in the GP regression model, Eq. 3 can be changed to

$$\begin{pmatrix} \mathbf{y}_{N\times 1} \\ \dot{\mathbf{y}}_{M\times 1} \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_y \\ m_{\dot{y}} \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{N\times N} & \dot{\mathbf{K}}_{N\times M} \\ \dot{\mathbf{K}}'_{M\times N} & \ddot{\mathbf{K}}_{M\times M} \end{bmatrix}\right) \tag{15}$$

where $\dot{y} = \frac{\partial y}{\partial x_j}$ is the derivation of the state variable with respect to the jth independent variable and M represents number of derivative observations. The a priori mean of the derivatives is represented by $M \times 1$ vector $m_{\dot{y}}$. The components of the covariance matrix are $\dot{K} = \left(\frac{\partial k(\cdot)}{\partial x_j}\right)_{N\times M}$ and $\ddot{K} = \left(\frac{\partial^2 k(\cdot)}{\partial x_j \partial x'_j}\right)_{M\times M}$.

To predict the state variable or its derivatives at a new test point ($\mathbf{X}_*$), the conditional property of the Gaussian distribution as used in Eq. (5) can be applied to Eq. (15). The expressions for the predictive distributions are given in Solak et al. (2003) and Wu et al. (2017). The methodology presented above can be easily extended to incorporate higher order derivatives in the GP regression model.

2.5 PDE parameter estimation

Let the PDE be given by an implicit function $f\left(x_1, \ldots, x_m, y, \frac{\partial y}{\partial x_1}, \ldots, \frac{\partial y}{\partial x_m}, \frac{\partial^2 y}{\partial x_1 \partial x_1}, \ldots, \frac{\partial^2 y}{\partial x_1 \partial x_m}, \ldots, \boldsymbol{\lambda}\right) = 0$, where, y is the state variable, x_j $(j = 1, \ldots m)$ are independent variables and $\boldsymbol{\lambda}$ is a set of PDE parameters. The GPPDE method uses a GP to model the observations of the state variable and estimate its derivatives involved in the PDE. The residual error in the PDE at an observed point is given by

$$\epsilon = f\left(x_1, \ldots, x_m, y^{\mathrm{GP}}, \left(\frac{\partial y}{\partial x_1}, \ldots \frac{\partial y}{\partial x_m}, \frac{\partial^2 y}{\partial x_1 \partial x_1}, \ldots, \frac{\partial^2 y}{\partial x_1 \partial x_m}, \ldots\right)\Big|_{\mathrm{GP}}, \boldsymbol{\lambda}\right) \tag{16}$$

where the notation $(\cdot)\big|_{\mathrm{GP}}$, incorporates all the partial derivatives of the state variable, means that the arguments are evaluated using the GP relation given in Sect. 2.3. Figure 1 uses probabilistic graphs to represent a GP regression model, an ODE, the GM model of Calderhead et al. (2008) and the proposed GPPDE model.

The parameter set $\boldsymbol{\lambda}$ can be obtained by assuming a distribution for the residual errors and then applying either the method of maximum likelihood or the Bayesian estimation method using MCMC sampling. These methods not only provide a point estimate of $\boldsymbol{\lambda}$ but also an estimation interval. However, to keep the parameter estimation simple, we have obtained $\boldsymbol{\lambda}$ by minimizing the sum of square of residual error (SSRE), given as

$$SSRE = \boldsymbol{\epsilon}^{\mathrm{T}}\boldsymbol{\epsilon} = \sum_{\forall \mathrm{obs.}} f\left(\left(x_1, \ldots, x_m, y^{\mathrm{GP}}, \left(\frac{\partial y}{\partial x_1}, \ldots, \frac{\partial y}{\partial x_m}, \frac{\partial^2 y}{\partial x_1 \partial x_1}, \ldots, \frac{\partial^2 y}{\partial x_1 \partial x_m}, \ldots\right)\Big|_{\mathrm{GP}}, \boldsymbol{\lambda}\right)^2\right) \tag{17}$$

where obs. denotes observations. The minimization is performed using Nelder–Mead simplex algorithm (Lagarias et al. 1998). The GP regression model provides standard error for the estimated state variables and its derivatives. This uncertainty information is propagated to SSRE by weighting each term in the Eq. (17), where weights are inversely proportional to the variance of the terms.

The confidence interval of the estimated parameters is obtained by Cram-rRao bound, which provides a lower bound on the variance of any unbiased estimator (Rao 2008). The confidence interval corresponding to estimated parameter set $\hat{\boldsymbol{\lambda}}$ is given by $\hat{\boldsymbol{\lambda}} \pm c\sqrt{\mathrm{diag}\left(I(\hat{\boldsymbol{\lambda}})\right)}$, where $I(\hat{\boldsymbol{\lambda}})$ is the Fisher information matrix evaluated at estimated values of parameters ($\hat{\boldsymbol{\lambda}}$) (in the present case it is taken as the Hessian matrix of SSRE) and c is z-critical value for a standard normal distribution. In the present study, 95% confidence interval has been estimated, for which the corresponding critical value is 1.96.

Since the GPPDE method does not solve the PDE (forward problem), the initial and boundary conditions are not required. However, if the initial and boundary conditions are available, exactly or with measurement errors, they could be easily incorporated in the GPPDE model. To impose Dirichlet boundary conditions, the values of the state variable at the boundaries are considered as observations with no measurement errors. In the GPPDE model, it is achieved by assigning a very small value to the noise term (σ_y) for those boundary observations. If the measurement errors at the boundaries are known, the same can be assigned to σ_y. Similarly, to impose Neumann boundary conditions, the values of the state derivatives at the boundaries are considered as observations and the methodology given in Sect. 2.4 is used to learn the GP model. The same approach can also be used to incorporate any combinations of Dirichlet and Neumann boundary conditions.

3 Application of GPPDE method

The proposed GPPDE method is applicable to PDEs and system of PDEs. In this study, we demonstrate its utility by applying it to one-dimensional groundwater diffusion equation and one-dimensional Richards equation. The mathematical representation of these equations is given in the following subsections.

3.1 Groundwater diffusion equation

The groundwater diffusion equation is a second order parabolic PDE that describes the spatio-temporal variation of groundwater level in homogeneous aquifers. It requires one initial and two boundary conditions for forward solution. One-dimensional form of the groundwater equation is given as

$$\frac{\partial h}{\partial t} = \lambda \frac{\partial^2 h}{\partial x^2} \tag{18}$$

where h is the groundwater level or head (L; the state variable), t is time (T), x is distance from the origin of the coordinate axis (L) and λ is the parameter of the equation, known as aquifer diffusivity ($\mathrm{L}^2 \cdot \mathrm{T}^{-1}$). The inverse problem is to find λ given observed values of head at different location and time. The GPPDE method estimates λ by minimizing the sum of the square of residual errors (SSRE), defined as

$$SSRE = \boldsymbol{\epsilon}^{\mathrm{T}}\boldsymbol{\epsilon} = \sum_{\forall x}\sum_{\forall t}\left(\left.\frac{\partial h}{\partial t}\right|_{\mathrm{GP}} - \lambda\left.\frac{\partial^2 h}{\partial x^2}\right|_{\mathrm{GP}}\right)^2 \tag{19}$$

where $\left.\frac{\partial h}{\partial t}\right|_{\mathrm{GP}}$ and $\left.\frac{\partial^2 h}{\partial x^2}\right|_{\mathrm{GP}}$ are obtained from a GP regression model fitted to observed head values.

3.2 Richards equation

The Richards equation is a nonlinear parabolic PDE that describes the movement of water in unsaturated zone (Richards 1931; Celia et al. 1990). Flow in unsaturated zone is one of the most complex flows in nature. One-dimensional form of the Richards equation is given by

$$\begin{aligned}\frac{\partial \theta}{\partial t} &= \frac{\partial}{\partial z}\left[k\frac{\partial h}{\partial z}\right] + \frac{\partial k}{\partial z} \\ &= \underbrace{\frac{\partial k}{\partial z}\frac{\partial h}{\partial z}}_{\text{Term1}} + \underbrace{k\frac{\partial^2 h}{\partial z^2}}_{\text{Term2}} + \underbrace{\frac{\partial k}{\partial z}}_{\text{Term3}}\end{aligned} \tag{20}$$

where θ is volumetric soil moisture ($\mathrm{L}^3 \cdot \mathrm{L}^{-3}$; the state variable), t is time (T), z is vertical dimension (L; positive upward), k is hydraulic conductivity ($\mathrm{L} \cdot \mathrm{T}^{-1}$) and h is pressure head (L). The hydraulic conductivity (k) and the pressure head (h) can be related to the soil moisture by van Genuchten model (van Genuchten 1980), given by

$$k = k_s \Theta^{1/2}\left[1 - \left(1 - \Theta^{1/m}\right)^m\right]^2 \tag{21a}$$

$$\Theta = \begin{cases} [1 + |\alpha h|^n]^{-m} & h < 0 \\ 1, & h \geq 0 \end{cases} \tag{21b}$$

$$\Theta = \frac{\theta - \theta_r}{\theta_s - \theta_r} \tag{21c}$$

where k_s is saturated hydraulic conductivity, Θ is relative saturation, and θ_r and θ_s are residual and saturated soil moisture content, respectively. The terms α (L^{-1}), m and n represent model parameters where, m and n are related as $n(1-m) = 1$. The parameters α and n are related to air-entry pressure head and pore-size distribution, respectively. Eq. (21a) is known as hydraulic conductivity function and Eq. (21b) is known as soil water retention function.

In the literature, field observations have been used to estimate soil hydraulic parameters using inverse solution of the Richards equation which requires initial and boundary

conditions (Ritter et al. 2003; Simunek et al. 2005). Since, initial and boundary condition in real world are not easily available, the conventional approach becomes cumbersome (Farthing and Ogden 2017). The proposed GPPDE overcomes this limitation. In the GPPDE method, the estimation of parameters is done by minimizing SSRE, given by

$$SSRE = \boldsymbol{\epsilon}^{\mathrm{T}}\boldsymbol{\epsilon} = \sum_{\forall t's}\sum_{\forall z's}\left[\left.\frac{\partial\theta}{\partial t}\right|_{\mathrm{GP}} - \left(\underbrace{\frac{\partial k}{\partial z}\frac{\partial h}{\partial z}}_{\mathrm{Term1}} + \underbrace{k\frac{\partial^2 h}{\partial z^2}}_{\mathrm{Term2}} + \underbrace{\frac{\partial k}{\partial z}}_{\mathrm{Term3}}\right)\Bigg|_{\mathrm{GP}}\right]^2. \tag{22}$$

In the present work, we have estimated only α and n, but the method can be easily extended to estimate any desired number of parameters. To apply GPPDE, the terms of $\boldsymbol{\epsilon}^{\mathrm{T}}\boldsymbol{\epsilon}$ should be expressed in the terms of the state variable (θ), independent variables (z and t) and derivatives of the state variable with respect to the independent variables $\left(\frac{\partial\theta}{\partial z}, \frac{\partial^2\theta}{\partial z^2}\right.$ and $\left.\frac{\partial\theta}{\partial t}\right)$. This factorization of $\boldsymbol{\epsilon}^{\mathrm{T}}\boldsymbol{\epsilon}$ is provided in "Appendix 2" section.

We compared the GPPDE estimates of hydraulic parameters with that obtained from conventional approach by using HYDRUS-1D software (Simunek et al. 2005). The software first solves the forward problem and then estimates hydraulic parameters by minimizing the difference between the simulated and observed values. In the present paper z is discretized at 1 cm interval. For each iteration HYDRUS automatically adjusts the time step to achieve fast convergence and the time step varies in the range of 0.00024–0.024 h. The maximum number of iterations allowed for a given time step is set to 10 for forward solution of Richards equation using a modified Picard method. The maximum necessary absolute change in the value of the water content and pressure head between two successive iterations during a particular time step are set to 0.001 and 1 cm, respectively. Levenberg–Marquardt algorithm is used for minimization. If the observations have measurement error, HYDRUS-1D uses a weighted error function where weights are inversely proportional to the variance of the measurement error. However, unlike GPPDE, it cannot account for measurement error in initial or boundary conditions. Further minute details of implementation of HYDRUS model is available in the HYDRUS' user manual.

4 Data used in the study

This section describes the data used for testing the GPPDE method on the diffusion equation and the Richards equation. Newly developed parameter estimation methods are usually first tested on synthetic data because the true parameters are known and sensitivity of these methods to measurement errors can be evaluated by introducing error of different levels. Hence, in this work we have tested the GPPDE method on synthetic data.

4.1 Data used for the groundwater diffusion equation

Reynolds (1987) reported the water level fluctuations for three sites in a glacial-outwash valley aquifer near Cortland, New York, USA. The data reported by Reynolds for Site-1 have been used in many studies to demonstrate the performance of parameter estimation methods in estimating aquifer diffusivity of the site (Srivastava 2006; Swamee and Singh 2003). In Reynolds' study for Site-1, variation of water level in two different wells were observed, one close to a stream and other 152 m from the stream. Reynolds assumed that the water level in the well closer to the stream represents the water level of the stream. The data of water level fluctuations are provided in figures of the original study, which we extracted using WebPlotDigitizer software (Rohatgi 2011). The variation of water level of the wells are plotted in Fig. 2. The reported value of the aquifer diffusivity varies in range of 550–2200 $\mathrm{m}^2 \cdot \mathrm{h}^{-1}$, depending on the assumptions and methodologies. The performance of GPPDE is evaluated for the following two cases.

4.1.1 Synthetic data

Synthetic data are generated by assuming observed water level of the stream for Site-1 as the left boundary condition and for the right boundary, it is assumed that there is no effect of variation in the stream water level on the

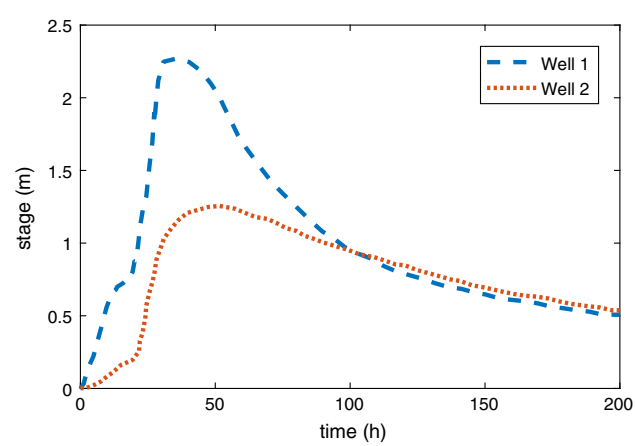

Fig. 2 Observed water levels in two wells at Cortland, New York based on the data provided by Reynolds (1987). Dashed curve for well near the stream and the dotted curve for well 152 m from the stream

groundwater level at $x = 500\,\text{m}$, $h(x = 500\,\text{m}, t) = 0\,\text{m}$. Initial condition is assumed as $h(x, t = 0) = 0\,\text{m}$. To generate water level data, an implicit finite difference scheme was applied with $\lambda = 1000\,\text{m}^2 \cdot \text{h}^{-1}$. The synthetic data are plotted in Fig. 3a. The GPPDE method is demonstrated on the data generated for five intervals of x, $\Delta x = [10, 25, 50, 75, 150]\,\text{m}$ and for time $t \in [0, 200]$ h. Simulations in the range $x \in$ [1 150] m are used as observations for GPPDE analysis. For example, if $\Delta x = 25$ m, simulations at $x = [25, 50, 75, 100, 125, 150]\,\text{m}$ are treated as observations for GPPDE parameter estimation.

4.1.2 Field data

For estimating aquifer diffusivity at the study site, the field observations made at the second well ($x = 152$ m) is considered as a response to fluctuations in the stream water level. For this analysis, we again made the assumption that the groundwater level at $x = 500$ m is unaffected by the variations in the stream water level, i.e. $h(x = 500\,\text{m}, t) = 0\,\text{m}$.

4.2 Data used for the Richards equation

The applicability of GPPDE on the Richards equation is tested using synthetic data described in this subsection. Four cases with different combinations of boundary conditions are studied. Within these four cases, we also studied the effect of different number of observations and error levels on GPPDE parameter estimates.

The finite difference scheme for mixed-form of the Richards equation provided by Celia et al. (1990) is adopted to generate synthetic data. The synthetic data are generated with five hydraulic parameters $\theta_r = 0.075$, $\theta_s = 0.287$, $\alpha = 0.05\,\text{cm}^{-1}$, $n = 2$, and $k_s = 1.5\,\text{cm}\cdot\text{h}^{-1}$ with initial condition $h(z, t = 0) = -65\,\text{cm}$, having boundary conditions as $h(z = 0, t) = -65\,\text{cm}$ at the bottom and $h(z = 60, t) = -20.7\,\text{cm}$ at the top. The above initial and boundary conditions are in terms of pressure head, which translates to volumetric water content as $\theta(z, t = 0) = 0.137$, and $\theta(z = 0, t) = 0.137$ and $\theta(z = 60, t) = 0.222$, respectively, using the van Genuchten (1980) model. The soil moisture is estimated using the finite difference scheme at internal nodes with $\Delta z = 1\,\text{cm}$ and $\Delta t = 1\,\text{h}$ over the interval [0, 50] h.

Using the synthetic data described above, the following four cases are studied:

1. Case 1 (Top: Dirichlet and Bottom: Dirichlet): The boundary condition described in previous paragraph corresponds to Case-1. Figure 5a shows the 3-D plot of the synthetic data generated for this case.
2. Case 2 (Top: Dirichlet and Bottom: Free drainage): The top boundary in this case is as described in Case-1, and the bottom boundary is free drainage which is a Neumann boundary with $\frac{\partial \theta}{\partial z} = 0$.
3. Case 3 (Top: Time Varying Dirichlet and Bottom: Dirichlet): The top boundary in this case represents diurnal variation of soil moisture at the soil surface which is obtained by superimposing the top boundary in Case-1 with a sinusoidal wave of amplitude 3 cm and periodicity of 24 h, mathematically $h(z = 60, t) = -20.7 + 3\sin\left(\frac{2\pi t}{24}\right)$ cm, where

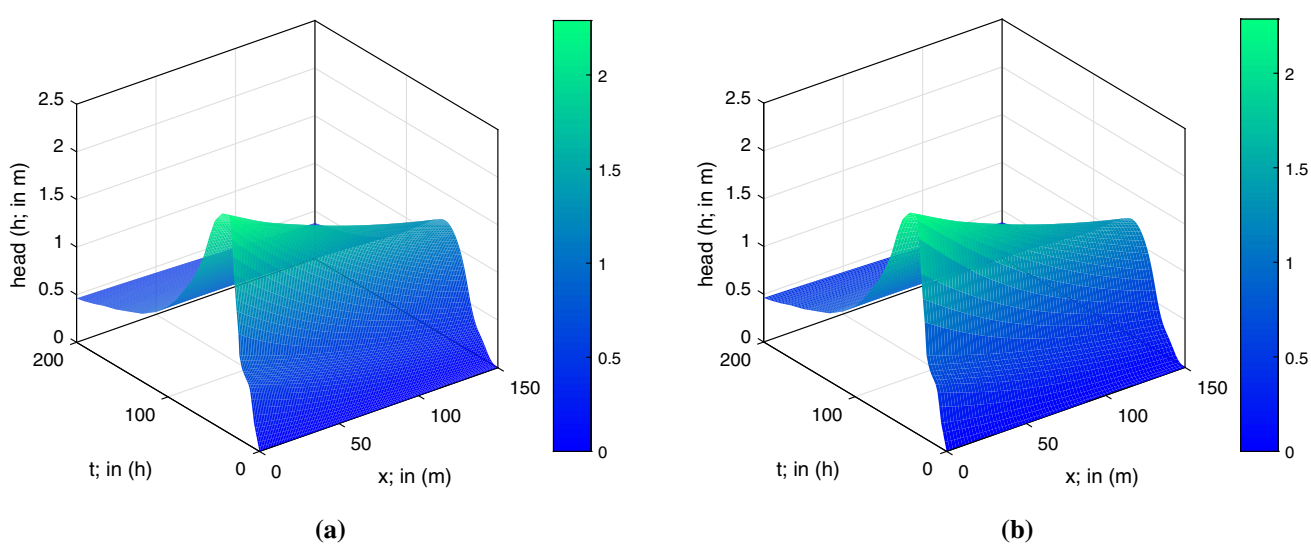

Fig. 3 **a** Synthetically generated observations of groundwater level plotted at a resolution of 1 m, **b** mean of GP regression fit to the groundwater profile, trained using data with $\Delta x = 10\,\text{m}$

$t \in (0, 50]$ h. The bottom boundary is fixed as in Case-1 ($h(z = 0, t) = -65$ cm).

4. Case 4 (Top: Time Varying Dirichlet and Bottom: Free drainage): Time varying top boundary condition is adopted from Case-3 and a bottom Neumann condition is adopted from Case-2.

The boundary conditions for these four cases are summarized in Table 1. To study the performance of the GPPDE method for coarser data, the observations at five intervals of z ($\Delta z = 1, 2, 3, 4, 5$ and 10 cm) are used. Next, to study the performance of GPPDE on noisy data, zero mean Gaussian noise of standard deviations 1, 2, 3, 4 and 5% of the soil moisture observation is added to the synthetic data.

5 Results and discussion

This section presents the results obtained from the proposed GPPDE method for the diffusion and the Richards equations. In the case of the Richards equation, the GPPDE parameter estimates are compared with the HYDRUS estimates.

Coefficient of determination (R^2) and Root Mean Square Error (RMSE) are calculated by the following equations for evaluating the performance of GP regression, and parameter estimates by GPPDE and HYDRUS.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (\tilde{y}_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (\tilde{y}_i - \bar{y})^2} \tag{23}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\tilde{y}_i - \hat{y}_i)^2}{N}} \tag{24}$$

where $\tilde{y}_i$ are observations, $\hat{y}_i$ are predicted values, $\bar{y}$ is the mean of the observations, and N is the number of observations.

5.1 Groundwater diffusion equation

5.1.1 Synthetic data

The synthetic water level data used in this section are described in Sect. 4.1 and shown in Fig. 3a. The data with $\Delta x = 1$ m is used to train GP regression model and the hyperparameters of GP, $\Phi = [l_x, l_t, \sigma_s, \sigma_y]$ are estimated using the method explained in Sect. 2.2, where l_x and l_t are length scale parameters in spatial (x) and temporal (t) dimensions, respectively. The estimated values of the hyperparameters are $\widehat{\Phi} = [130\,\text{m}, 1.7\,\text{h}, 0.875, 5.3\text{E}{-}04\,\text{m}]$. The GPPDE method requires estimation of the state variable and its partial derivatives involved in the PDE. Using the obtained hyperparameters, the state variable (h) is calculated using Eq. (7a) and is plotted in Fig. 3b, which suggests a good fit. The state derivatives $\left(\frac{\partial h}{\partial t} \text{ and } \frac{\partial^2 h}{\partial x^2}\right)$ are obtained from the methodology given in Sect. 2.3.

The PDE parameter λ is obtained by incorporating the estimates of the state variable and its derivative in SSRE Eq. 19, and then minimizing the SSRE function. For $\Delta x = 10$ m, the GPPDE method estimates $\hat{\lambda} = 999.6 \pm 2.41\,\text{m}^2\cdot\text{h}^{-1}$ which is close to the true value of diffusivity ($1000\ \text{m}^2\cdot\text{h}^{-1}$), suggesting that the GPPDE method can successfully estimate the aquifer diffusivity (λ).

To study the performance of the GPPDE method for coarse resolution data, the method is applied on sub-samples of the synthetic data, i.e. observations at intervals $\Delta x = [25, 50, 75, 150]$ m are used for learning the parameters. The estimates of λ for different number of observations along with their 95% confidence interval are given in Table 2. It is concluded from the table that with the decrease in number of observations, the uncertainty in the estimate of λ increases. The length scale parameters (l_x and l_t) can be interpreted as characteristic length, which is the distance required in the input space (along a particular axis) for the function values to become uncorrelated (Rasmussen and Williams 2006). Thus, with decrease in the number of observations in the x-dimension, the l_x increases. It is observed that learning length-scale parameters from sparse data may sometimes result in an overly complicated function because of local minima in the marginal log likelihood function Eq. (11). This is a well known problem in GP models (Rasmussen and Williams 2006; Tripathi and Govindaraju 2007). We found in our experiments that if such a problem occurs, it can be easily detected by visualizing the fitted surface. Further, we realized that with

Table 1 The boundary conditions used to demonstrate the application of GPPDE on the Richards equation

Case	Top boundary ($z = 60$ cm)	Bottom boundary ($z = 0$ cm)
1	Dirichlet $h = -20.7$ cm	Dirichlet $h = -65$ cm
2	Dirichlet $h = -20.7$ cm	Neumann $\frac{\partial \theta}{\partial z} = 0$
3	Dirichlet $h = -20.7 + 3\sin\left(\frac{2\pi t}{24}\right)$ cm; $t \in (0, 50]$ h	Dirichlet $h(z = 0, t) = -65$ cm
4	Dirichlet $h = -20.7 + 3\sin\left(\frac{2\pi t}{24}\right)$ cm; $t \in (0, 50]$ h	Neumann $\frac{\partial \theta}{\partial z} = 0$

Table 2 The estimates of aquifer diffusivity, λ, for different intervals of x, along with their 95% confidence interval

S. no.	Δx (m)	No. of obs	Kenrel paramter			Noise	$\hat{\lambda}$ ($m^2 \cdot h^{-1}$)
			l_x (m)	l_t (h)	σ_s	σ_y (m)	
1	10	15	130	1.7	0.875	5.3E−04	999.6 ± 2.41
2	25	6	180	1.7	0.662	9.5E−04	998.9 ± 6.20
3	50	3	230	1.7	0.627	2.6E−03	993.5 ± 13.38
4	75	2	280	1.7	0.660	3.3E−03	993.0 ± 19.30
5	150	1	330	1.7	0.620	3.8E−03	1005.0 ± 58.23

some familiarity with the data, length scale parameters can be easily specified and the results are not sensitive to small changes in their values. The logarithm of SSRE for $\Delta x = [10, 50, 150]$ m versus λ is plotted in Fig. 4 (the logarithmic transformation of SSRE is chosen for better visualization). From the figure and the table, we realize that the GPPDE method can estimate λ with fewer observations, however it increases estimation uncertainty.

5.1.2 Field data

The GP model is trained with observations at $x = [0, 152, 500]$ m. For $x = 0$ m, the time series observations from the well closer to the stream is used; for $x = 152$ m, the time series observations from the well away from the stream is used; and for $x = 500$ m, it is assumed that the groundwater level is unaffected by any variations in the stream water level, $(h(x = 500\,\text{m}, t) = 0\,\text{m})$. The water level values at x equal to 0 and 500 m are treated as Dirichlet boundary conditions. The hyperparameters for the fitted GP model are $\widehat{\Phi} = [330\,\text{m}, 1.7\,\text{h}, 0.610, 2.8\text{E}{-}03\,\text{m}]$. The estimated aquifer diffusivity is $\lambda = 564.8 \pm 29.3\,\text{m}^2 \cdot \text{h}^{-1}$, which is within the limits of reported values (550–2200 $\text{m}^2 \cdot \text{h}^{-1}$) in the literature (Reynolds 1987; Swamee and Singh 2003; Srivastava 2006).

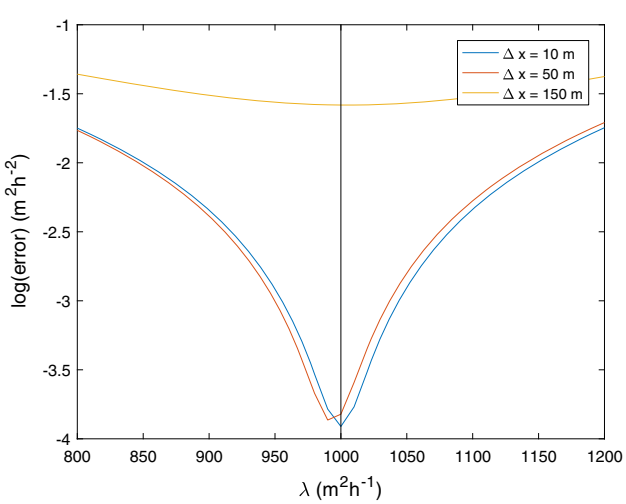

Fig. 4 The logarithmic transformed values of SSRE for different intervals of x as a function of aquifer diffusivity (λ). The vertical black line represents the true value of λ

5.2 Richards equation

The following subsections demonstrate the application of the GPPDE method to estimate soil hydraulic parameters and compare those estimates with HYDRUS-1D estimates. To evaluate the performance of the GPPDE method for coarse resolution data, the method is trained using observations at six intervals of z ($\Delta z = [1, 2, 3, 4, 5, 10]$ cm), as described in Sect. 4.2. The estimated hyperparameters and the estimated soil hydraulic parameters of the Richards equation for the four cases (different boundary conditions) are presented in Table 3.

5.2.1 Parameter estimation

The first step in the GPPDE method is to fit a GP regression model to the soil moisture data and obtain the hyperparameters, $\Phi = [l_z, l_t, \sigma_s, \sigma_y]$ using the methodology given in Sect. 2.2, where l_z and l_t are length scale parameters of the covariance function in spatial (z) and temporal (t) dimensions, respectively. Next, the derivatives of the state variable $\left(\frac{\partial \theta}{\partial t}, \frac{\partial \theta}{\partial z} \text{ and } \frac{\partial^2 \theta}{\partial z^2}\right)$ are obtained using the methodology given in Sect. 2.3. As the final step, the estimates of the state variable and its partial derivatives (obtained from the GP modelling) are incorporated in SSRE Eq. (22) to estimate soil hydraulic parameters $\boldsymbol{\lambda} = [\alpha, n]$.

The synthetic soil moisture data generated using the Richards equation as described in Sect. 4.2 are plotted in Fig. 5a. This dataset consists of two Dirichlet boundaries and intermediate observations at 59 locations in the range $z \in [1, 59]$ cm with $\Delta z = 1$ cm and $t \in [1, 50]$ h with $\Delta t = 1$ h. The GP regression model is forced to predict precise Dirichlet boundary conditions using the methodology presented in the last paragraph of Sect. 2.5. The curve obtained from the GP regression model for the learned hyperparameters $\widehat{\Phi} = [5\,\text{cm}, 3\,\text{h}, 0.188, 9.3\text{E}{-}05]$

Table 3 Estimates of GP hyperparameters and soil hydraulic parameters obtained from the GPPDE method. The true values of α and n are $0.05\,\mathrm{cm}^{-1}$ and $n = 2$, respectively

Cases	S. no.	Δz (m)	No. of obs	Kenrel paramter			Noise	van Genuchten	
				l_z (cm)	l_t (h)	σ_s	σ_y (units of θ; $\mathrm{L}^3\cdot\mathrm{L}^{-3}$)	α (cm^{-1})	n
Case-1;	1	1	59	5	3	0.188	9.3E−05	0.049 ± 1.0E−03	2.008 ± 8.0E−03
Top: Dirichlet and bottom: Dirichlet	2	2	29	6	3	0.085	2.1E−04	0.052 ± 2.0E−03	2.037 ± 1.9E−02
	3	3	19	7	3	0.077	3.0E−04	0.052 ± 3.0E−03	2.048 ± 3.3E−02
	4	4	14	8	3	0.074	4.6E−04	0.053 ± 5.0E−03	2.058 ± 5.0E−02
	5	5	11	9	3	0.073	5.4E−04	0.054 ± 7.0E−03	2.081 ± 7.2E−02
	6	10	5	14	3	0.075	1.8E−04	0.050 ± 2.5E−02	2.127 ± 2.94E−01
Case-2;	1	1	59	5	3	0.178	9.4E−05	0.050 ± 4.15E−06	2.017 ± 4.88E−05
Top: Dirichlet and bottom: free drainage	2	2	29	6	3	0.084	2.1E−04	0.052 ± 2.29E−05	2.040 ± 2.60E−04
	3	3	19	7	3	0.078	2.9E−04	0.053 ± 4.84E−05	2.055 ± 5.34E−04
	4	4	14	8	3	0.076	4.4E−04	0.053 ± 8.99E−05	2.065 ± 9.73E−04
	5	5	11	9	3	0.075	5.1E−04	0.055 ± 1.53E−04	2.087 ± 1.60E−03
	6	10	5	14	3	0.081	1.6E−04	0.050 ± 7.67E−04	2.126 ± 8.75E−03
Case-3;	1	1	59	5	3	0.180	1.4E−04	0.050 ± 5.86E−04	2.035 ± 7.30E−03
Top: time varying Dirichlet and bottom: Dirichlet	2	2	29	6	3	0.088	2.9E−04	0.054 ± 1.54E−03	2.083 ± 1.81E−02
	3	3	19	7	3	0.079	4.1E−04	0.055 ± 2.75E−03	2.101 ± 3.15E−02
	4	4	14	8	3	0.076	6.1E−04	0.056 ± 4.35E−03	2.120 ± 4.82E−02
	5	5	11	9	3	0.074	7.3E−04	0.059 ± 6.72E−03	2.158 ± 7.05E−02
	6	10	5	14	3	0.078	2.3E−04	0.053 ± 2.32E−02	2.223 ± 2.82E−01
Case-4;	1	1	59	5	3	0.170	1.4E−04	0.052 ± 4.14E−06	2.053 ± 5.04E−05
Top: Time varying Dirichlet and bottom: free drainage	2	2	29	6	3	0.086	2.9E−04	0.054 ± 2.22E−05	2.088 ± 2.59E−04
	3	3	19	7	3	0.079	4.0E−04	0.056 ± 4.86E−05	2.113 ± 5.46E−04
	4	4	14	8	3	0.077	5.8E−04	0.057 ± 9.15E−05	2.130 ± 9.99E−04
	5	5	11	9	3	0.076	6.9E−04	0.059 ± 1.58E−04	2.166 ± 1.63E−03
	6	10	5	14	3	0.083	2.0E−04	0.054 ± 7.23E−04	2.234 ± 8.51E−03

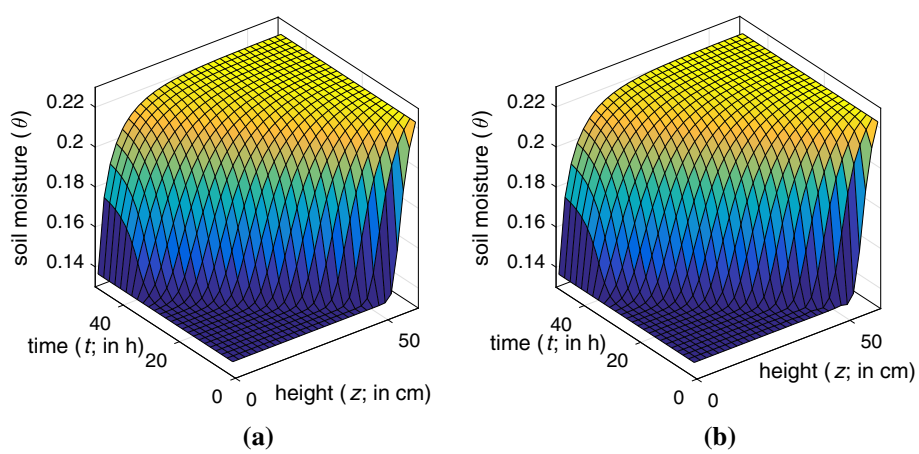

Fig. 5 **a** 3-D plot of the synthetic soil moisture data for Case 1, **b** mean of GP regression fit to the soil moisture data

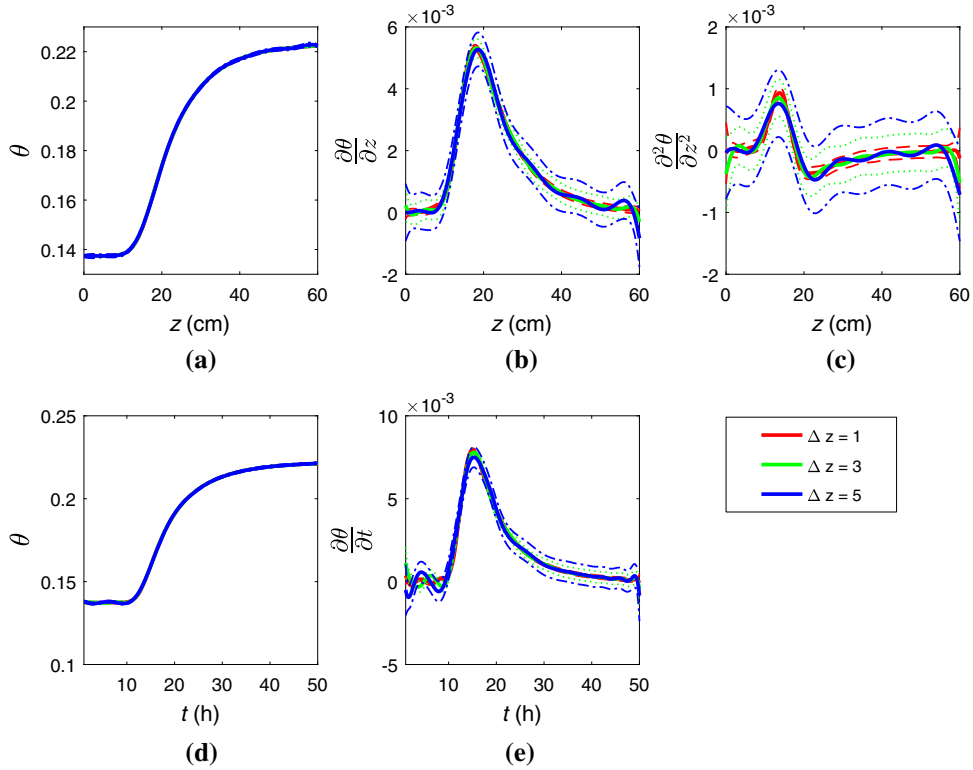

Fig. 6 GP regression fit to the synthetic soil moisture data for Case 1 and its partial derivatives. The first row shows the variations along the height (z) of soil column at $t = 25$ h—**a** soil moisture observations and GP fit, **b** first derivative of θ w.r.t. z, **c**) second derivative of θ w.r.t. z. The second row shows the variations along the time (t) at $z = 30$ cm—**d** soil moisture observations and GP fit, **e** derivative of θ w.r.t. t. The vertical dashed lines represent the standard error of the fitted curves

is plotted in Fig. 5b, which is very similar to the true data ($R^2 = 1, \text{RMSE} = 8.12\text{E}{-}05$).

The top row in Fig. 6a–c shows the variation of the estimated soil moisture with z, and its first and second order derivatives with respect to z at time $t = 25$ h. The bottom row of Fig. 6d, e shows the variation of the estimated soil moisture and its derivative with respect to t at height $z = 30$ cm. Like the state variable, the partial derivatives are also GP, which are smooth functions. Figure 6 shows that the GP estimates have fluctuations present near the boundaries, which increase for higher order derivatives. The standard errors of the fitted curves (shown as dashed lines) also increase near the boundaries. These fluctuations and increase in the standard error are reflections of higher uncertainty in the GP fitting near the boundaries.

The estimated values of the hydraulic parameters are $\hat{\boldsymbol{\lambda}} = [0.049 \pm 0.001\,\text{cm}^{-1}, 2.008 \pm 0.008]$, which are very close to the true value $\boldsymbol{\lambda}_0 = [0.05\,\text{cm}^{-1}, 2]$. Figure 8a shows the contours of the logarithmic transformed values of SSRE as a function of α and n. Again, the logarithmic transformation is used to aid visualization. The contours are closely spaced near the vicinity of the true solution and their intervals are wider for the cases with sparser data. The estimate of parameters, α and n, for Cases 2, 3 and 4 from GPPDE method are provided in Table 3. Like Case 1, the GPPDE model could reasonably estimate $\boldsymbol{\lambda}$ for these cases as well. The parameter uncertainties for both α and n are higher for Cases 1 and 3, where bottom boundary condition imposes a fixed soil moisture (Dirichlet) compared to Cases 2 and 4, where bottom boundary condition is free drainage (Neumann). This could be attributed to the abrupt change in soil moisture near the bottom boundary for Dirichlet conditions (see Figs. 5, 7).

5.2.2 Sparse observations

To study the performance of GPPDE for coarse resolution data, it is trained with data obtained at intervals, $\Delta z = [1, 2, 3, 4, 5, 10]$ cm. Figure 6 compares the estimates of the state variable and its derivatives for different number of observations ($\Delta z = [1, 3, 5]$ m). With increase in Δz (decrease in number of observations), the standard error of the estimated state variable and its derivatives increase. The estimated values are however very close for different Δz, except near the boundaries. These differences among the curves are relatively large for space derivatives that the time derivative because the number of observations have reduced in the space dimension.

The number of observational depths, the estimates of hyperparameters and estimated soil hydraulic parameters for different Δz and boundary conditions are shown in Table 3. The logarithm of SSRE as a function of α and n for different z intervals are plotted in subfigures of Figs. 8 and 9 for Cases 1 and 2, respectively.

The training of GPPDE for different Δz showed that the time length scale hyperparameter (l_t) do not change

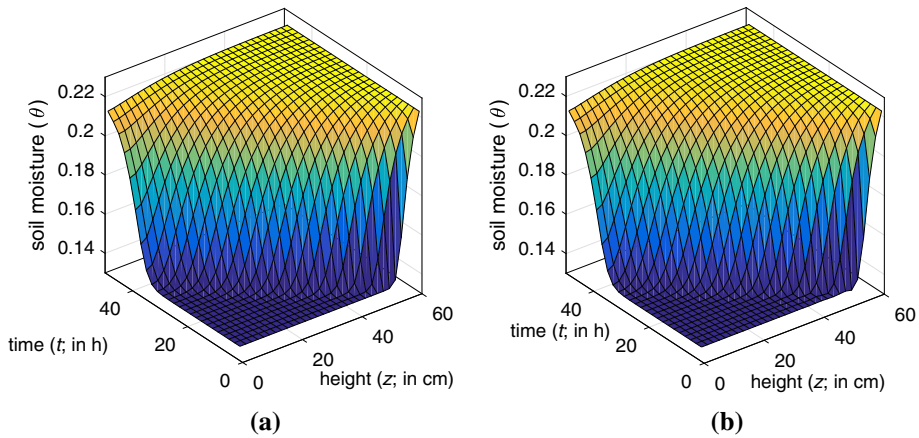

Fig. 7 **a** 3-D plot of the synthetic soil moisture data for Case 2, **b** mean of GP regression fit to the soil moisture data

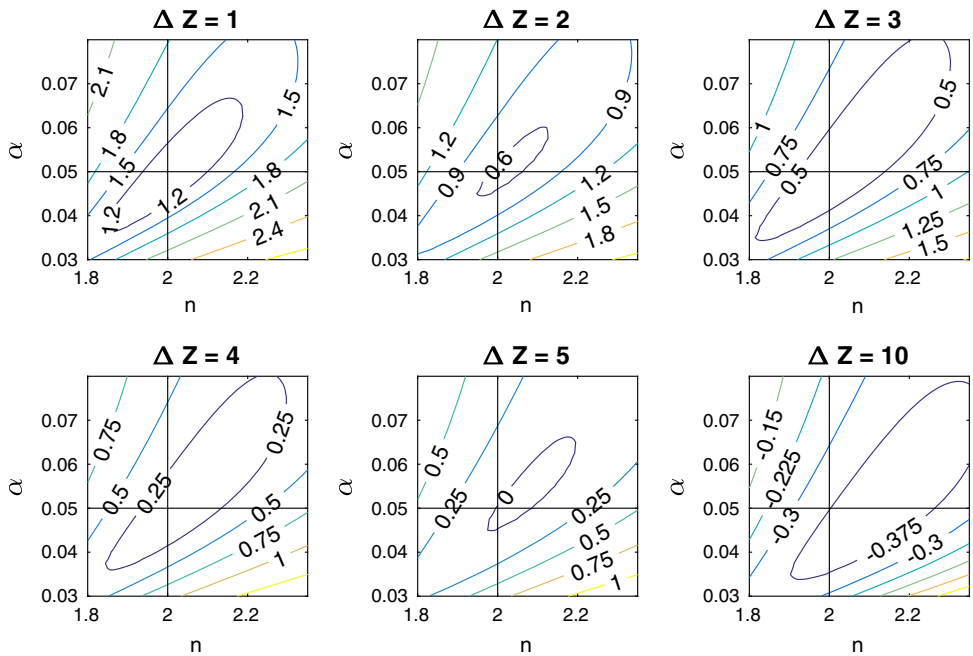

Fig. 8 Contours of logarithm of SSRE for Case 1 for different intervals of z. The intersection of the horizontal and the vertical lines represents the true parameters

because the temporal resolution of data remains the same. The length scale (l_z) increases with increase in Δz, resulting in smoother curve for coarse resolution data. The contour plots of logarithmic SSRE (Fig. 8) show that even for the sparse data set, the true solution lies in the vicinity of the lowest contour. However, the differences between the true and the estimated values of the hydraulic parameters increase with decrease in number of observations (Table 3). The value of contours decreases with increase in Δz which happens due to smaller number of terms in SSRE estimates. Further, the contour intervals are wider for smaller number of observations, suggesting increased uncertainty in the estimated parameters (also see Table 3).

5.2.3 Measurement errors

Errors of five different levels are added to the soil moisture observations for investigating the performance of GPPDE method in the presence of noise. The additive errors are normally distributed noise with mean zero and standard deviation 1, 2, 3, 4 and 5% of the soil moisture observations. Figures 10 and 11 provide the estimates of the parameters (α and n) for Cases 1 and 2, respectively, obtained from 100 Monte Carlo simulations. As expected, higher error levels increase the variability in the estimated parameters. However, no significant systematic bias is evident in the estimated parameters with increased error levels. These figures do not plot HYDRUS estimates, because HYDRUS consistently estimated α and n close to the true value for all the error levels.

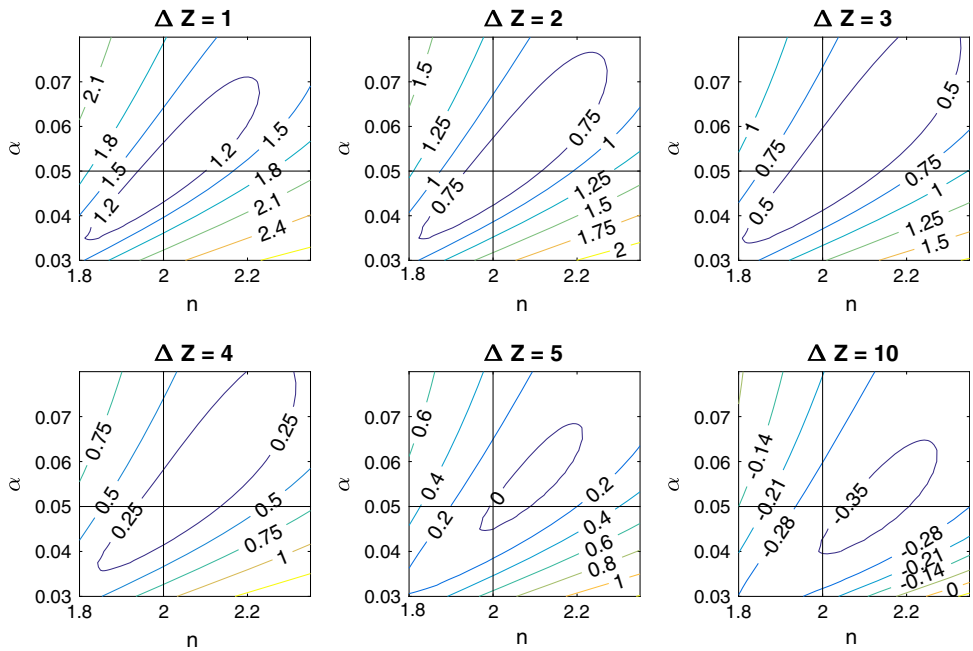

Fig. 9 Contours of logarithm of SSRE for Case 2 for different intervals of z. The intersection of the horizontal and the vertical lines represents the true parameters

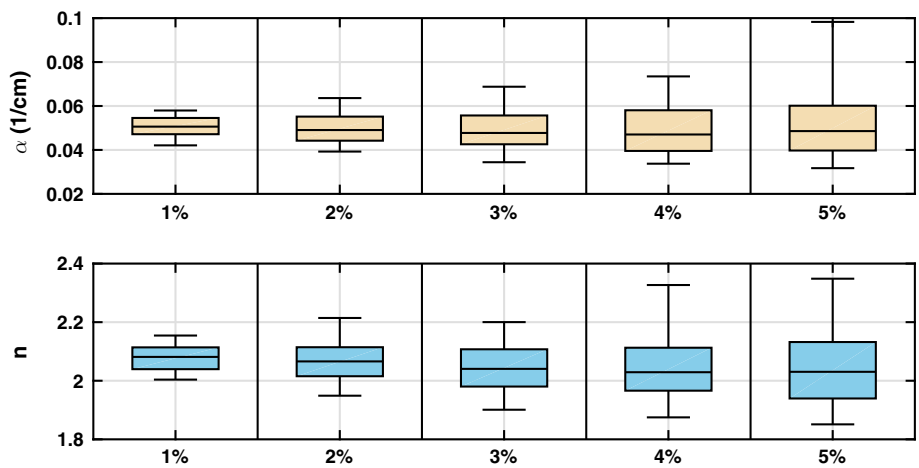

Fig. 10 Box plots for the estimates of α and n corresponding to different levels of Gaussian noise for $\Delta z = 5$ cm and boundary conditions as specified for Case 1

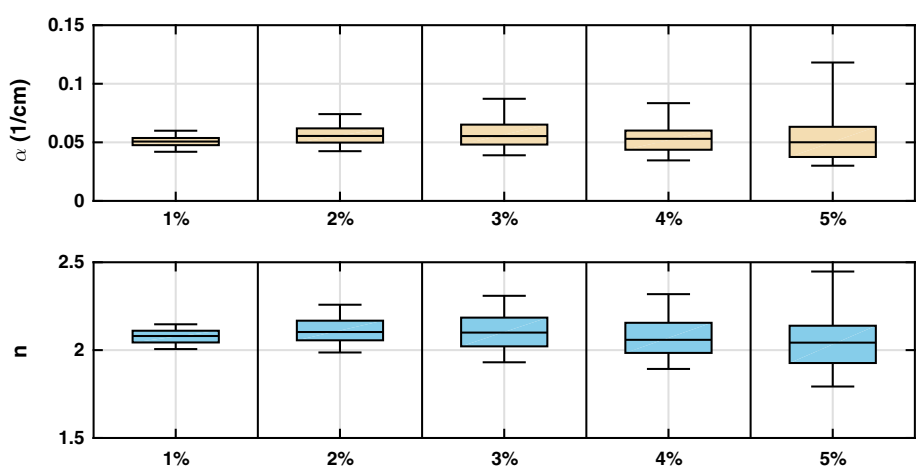

Fig. 11 Box plots for the estimates of α and n corresponding to different levels of Gaussian noise for $\Delta z = 5$ cm and boundary conditions as specified for Case 2

5.2.4 Perturbed boundary conditions

In applications where boundary conditions are obtained from observations, the measurement errors may induce uncertainty in boundary conditions. To test the performance of GPPDE and HYDRUS under uncertain boundary conditions, the Dirichlet boundaries for Cases 1 and 2 are perturbed with zero mean Gaussian noise and standard deviations ranging from 1 to 5% of the true soil moisture value.

Figures 12 and 13 show the box plots of the estimated α and n for GPPDE and HYDRUS (inverse modelling) for Cases 1 and 2, respectively. Since HYDRUS simulations were computationally expensive, results obtained from 60 Monte Carlo simulations are shown.

For Case 1, HYDRUS marginally overestimates α compared to GPPDE estimates, though it has relatively smaller variability. For n, the variability is smaller for GPPDE estimates at higher error levels. The results for Cases 2 are similar. Overall, the results suggest that GPPDE and HYDRUS estimates are only marginally different. However, GPPDE has significant advantages over HYDRUS in terms of computational time. For example, the computation time for one run of HYDRUS is 5.787 ± 0.004 s (based on 100 simulations). Whereas for a MATLAB implementation of GPPDE on the same sample, the computational time for one run is $(5.002 \pm 0.065) \times 10^{-3}$ s.

6 Summary and concluding remarks

A Gaussian process (GP) based method for estimating parameters of PDEs is proposed and termed as Gaussian process for Partial Differential Equation (GPPDE) method. Unlike the conventional method, the GPPDE method does not require setting-up of initial and boundary conditions explicitly, which is often difficult in real world applications. However, if the initial and boundary conditions are known (exactly or with measurement errors), they could be incorporated in the method.

The new GPPDE method is evaluated by applying it on the diffusion and the Richards equations. The results

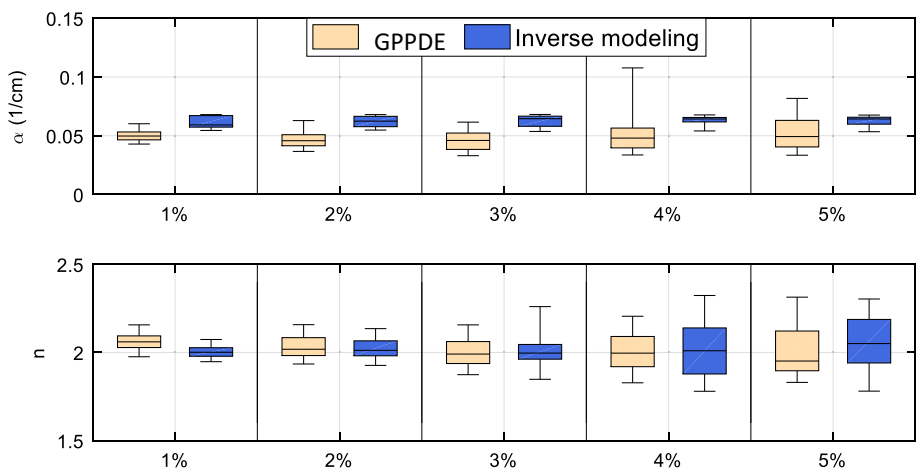

Fig. 12 Figure compares the performance of GPPDE and HYDRUS under boundary conditions perturbed with zero mean Gaussian noise (standard deviation equal to 1, 2, 3, 4 and 5% of the boundary values) for Case 1. The estimates of α and n obtained from 60 Monte Carlo simulations are shown as box plots. The parameters are obtained for the synthetic soil moisture data with $\Delta z = 5$ cm

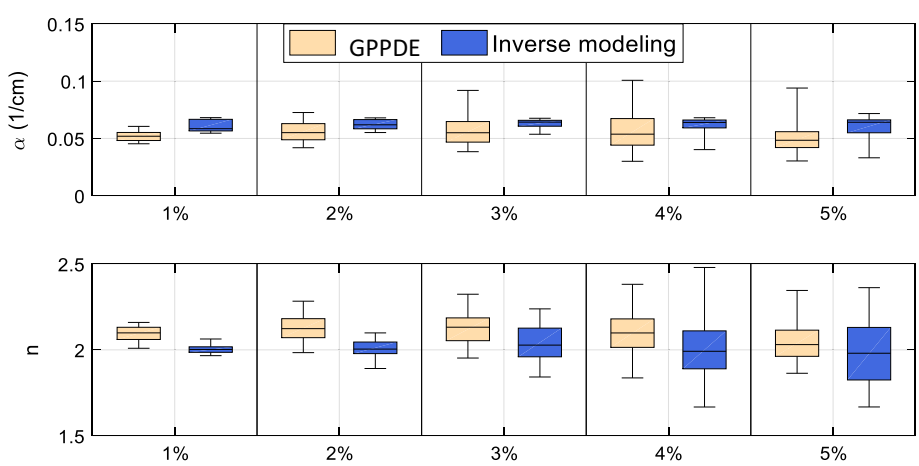

Fig. 13 Figure compares the performance of GPPDE and HYDRUS under boundary conditions perturbed with zero mean Gaussian noise (standard deviation equal to 1, 2, 3, 4 and 5% of the boundary values) for Case 2. The estimates of α and n obtained from 60 Monte Carlo simulations are shown as box plots. The parameters are obtained for the synthetic soil moisture data with $\Delta z = 5$ cm

suggest that the GPPDE method can correctly estimate parameters of the diffusion equation for the synthetic data. It provided a reasonable estimate for the diffusivity of glacial-outwash valley aquifer, Cortland, New York. The GPPDE method has also estimated parameters of the Richards equation satisfactorily for all the four cases of different boundary conditions investigated in the study.

A comparison of the GPPDE and HYDRUS (a conventional inverse modelling approach) suggested that both the methods could reasonably estimate soil hydraulic parameters in the presence of Gaussian noise in the range 1–5%. However, GPPDE has significant advantages in terms of computational time. We found that for the same dataset and computer, GPPDE was three orders of magnitude faster than HYDRUS. However, to reliably estimate PDE parameters, GPPDE requires large dataset. Its performance, to a large extent, depends on the accuracy of the curve fitted to the observations of the state variable. Accuracy of GP regression fit is also limited by difficulty in selection of a suitable covariance function.

Several avenues should be explored to further refine the proposed GPPDE method for PDE parameter estimation. The GPPDE method can be formulated in the Bayesian framework that will provide credible intervals for the estimated values of the parameters. Also, methods should be developed to relate kernel length-scale parameters to the characteristic length-scales of the input dimensions based on the physics of the problem.

Comprehensively, the results suggest that the proposed method can be an effective alternate to the conventional PDE parameter estimation methods. It has advantages for large observational datasets in which the state variable has high dimensionality. Overall, the approach of interpreting numerical problems of differential equations as probabilistic inference problems, and solving them using machine learning methods like GP is an exciting new area of research. It has potential to provide novel methods for handling observational uncertainties, which are ubiquitous in water resources problems but often ignored when used with differential equations.

Acknowledgements Pankaj Kumar Rai received scholarship from Ministry of Human Resource Development, Government of India, for his Ph.D. program. Authors are thankful of Mr. Hemanta Medhi and Mr. Pramod Soni for their help in preparing the manuscript.

Appendix 1: Derivative of Gaussian covariance function $k(\mathbf{x}, \mathbf{x}')$

Gaussian covariance function is given by relation

$$k(\mathbf{x}, \mathbf{x}') = {\sigma_s}^2 \exp\left\{ -\frac{1}{2} \sum_{i=1}^{m} \frac{\left(x_i - x_i'\right)^2}{l_i^2} \right\} \tag{25}$$

where l_i's and σ_s are the length and the scale parameters, respectively.

Derivative of covariance function with respect to hyperparameters

The derivatives of $k_y(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') + \sigma_y^2$, where σ_y is the noise term, are given by.

$$k_y(\mathbf{x}, \mathbf{x}') = {\sigma_s}^2 \exp\left\{ -\frac{1}{2} \sum_{i=1}^{m} \frac{\left(x_i - x_i'\right)^2}{l_i^2} \right\} + \sigma_y^2 \tag{26}$$

$$\frac{\partial k_y}{\partial l_j} = {\sigma_s}^2 \exp\left\{ -\frac{1}{2} \sum_{i=1}^{m} \frac{\left(x_i - x_i'\right)^2}{l_i^2} \right\} \frac{\left(x_i - x_i'\right)^2}{l_j^3}; \quad 1 \leqslant j \leqslant m \tag{27}$$

$$\frac{\partial k_y}{\partial \sigma_s} = 2\sigma_s \exp\left\{ -\frac{1}{2} \sum_{i=1}^{m} \frac{\left(x_i - x_i'\right)^2}{l_i^2} \right\} \tag{28}$$

$$\frac{\partial k_y}{\partial \sigma_y} = 2\sigma_y \tag{29}$$

Derivatives of covariance function with respect to arguments

First derivative

$$\begin{aligned} \frac{\partial k(\mathbf{x}, \mathbf{x}')}{\partial x_j'} &= \frac{\partial}{\partial x_j'} \left[{\sigma_s}^2 \exp\left\{ -\frac{1}{2} \sum_{i=1}^{m} \frac{(x_i - x'_i)^2}{l_i^2} \right\} \right]; 1 \leqslant j \leqslant m \\ &= k(\mathbf{x}, \mathbf{x}') \frac{\partial}{\partial x_j'} \left[-\frac{1}{2} \sum_{i=1}^{m} \frac{(x_i - x'_i)^2}{l_i^2} \right] \\ &= k(\mathbf{x}, \mathbf{x}') \frac{(x_j - x'_j)}{l_j^2} \end{aligned} \tag{30}$$

Second derivative

$$\begin{aligned} \frac{\partial^2 k(\mathbf{x}, \mathbf{x}')}{\partial x_j'^2} &= \frac{\partial}{\partial x_j'} \left[k(\mathbf{x}, \mathbf{x}') \frac{(x_j - x'_j)}{l_j^2} \right]; 1 \leqslant j \leqslant m \\ \frac{\partial^2 k(\mathbf{x}, \mathbf{x}')}{\partial x_j'^2} &= k(\mathbf{x}, \mathbf{x}') \frac{(x_j - x'_j)}{l_j^2} \frac{(x_j - x'_j)}{l_j^2} \\ &\quad + k(\mathbf{x}, \mathbf{x}') \left(\frac{-1}{l_j^2} \right) \\ &= \frac{k(\mathbf{x}, \mathbf{x}')}{l_j^2} \left[\frac{(x_j - x'_j)^2}{l_j^2} - 1 \right] \end{aligned} \tag{31}$$

Auto covariance

$$\frac{\partial^2 k(\mathbf{x},\mathbf{x}')}{\partial x_j \partial x'_j} = \frac{\partial}{\partial x_j}\left[k(\mathbf{x},\mathbf{x}')\frac{(x_j - x'_j)}{l_j^2}\right] = \frac{k(\mathbf{x},\mathbf{x}')}{l_j^2}\left[1 - \frac{(x_j - x'_j)^2}{l_j^2}\right] \tag{32}$$

Appendix 2: Factorization of the Richards equation

This appendix provides the representation of the Richards equation in terms of state variable (θ) and its partial derivatives $\left(\frac{\partial\theta}{\partial t}, \frac{\partial\theta}{\partial x}\text{ and }\frac{\partial^2\theta}{\partial x^2}\right)$.

The van Genuchten (1980) model is used to relate hydraulic conductivity (k), pressure head (h) and relative saturation (Θ) as

$$\Theta = \frac{1}{[1 + (\alpha|h|)^n]^m} \tag{33}$$

$$k = k_s \Theta^{1/2}\left[1 - \left(1 - \Theta^{1/m}\right)^m\right]^2 \tag{34}$$

where k_s is saturated hydraulic conductivity, and θ_r and θ_s are residual and saturated soil moisture content, respectively. The relative saturation terms α (L^{-1}), m and n represent model parameters where, m and n are related as $n(1-m) = 1$. Term (Θ) is defined as

$$\Theta = \frac{\theta - \theta_r}{\theta_s - \theta_r} \tag{35}$$

To write the Richards equation in terms of soil moisture and its partial derivatives, we need to estimate $\frac{\partial\Theta}{\partial\theta}, \frac{\partial k}{\partial\Theta}, \frac{\partial\Theta}{\partial h}$ and $\frac{\partial h}{\partial\theta}$. These partial derivatives were estimated as follows. Differentiating Θ with respect to θ

$$\frac{\partial\Theta}{\partial\theta} = \frac{1}{\theta_s - \theta_r} \tag{36}$$

Differentiating (34) with respect to Θ

$$\frac{\partial k}{\partial\Theta} = 2k_s\Theta^{1/2}\left[1 - \left(1 - \Theta^{1/m}\right)^m\right]\left[\left(1 - \Theta^{1/m}\right)^{m-1}\left(\Theta^{1/m-1}\right)\right] + \frac{k_s}{2}\Theta^{-1/2}\left[1 - \left(1 - \Theta^{1/m}\right)^m\right]^2 \tag{37}$$

Rearranging Eq. (34)

$$1 + (\alpha|h|)^n = \left(\frac{1}{\Theta}\right)^{1/m}$$
$$(\alpha|h|) = \left[\left(\frac{1}{\Theta}\right)^{1/m} - 1\right]^{1/n} \tag{38}$$

Again rearranging (33) and differentiating it with respect to h

$$\frac{1}{\Theta} = [1 + (\alpha|h|)^n]^m \tag{39}$$

$$\frac{\partial}{\partial h}\left(\frac{1}{\Theta}\right) = m\{1 + (\alpha|h|)^n\}^{m-1}\left\{n(\alpha|h|)^{n-1}\right\}\alpha\mathrm{sign}(h) \tag{40}$$

$$\frac{-1}{\Theta^2}\frac{\partial\Theta}{\partial h} = \alpha mn\mathrm{sign}(h)\{1 + (\alpha|h|)^n\}^{m-1}\left\{(\alpha|h|)^{n-1}\right\} \tag{41}$$

Substituting (38) into (41) (to represent $\frac{\partial\Theta}{\partial h}$ as a function of Θ)

$$\begin{aligned}\frac{\partial\Theta}{\partial h} &= -\frac{\alpha m\mathrm{sign}(h)}{1-m}\Theta^2\left\{\left(\frac{1}{\Theta}\right)^{1/m}\right\}^{m-1}\left\{\left(\frac{1}{\Theta}\right)^{1/m} - 1\right\}^m \\ &= -\frac{\alpha m\mathrm{sign}(h)}{1-m}\Theta^{(-m+1)/m+2}\left\{\left(\frac{1}{\Theta}\right)^{1/m} - 1\right\}^m \\ &= -\frac{\alpha m\mathrm{sign}(h)}{1-m}\Theta^{(m+1)/m}\left\{\left(\frac{1}{\Theta}\right)^{1/m} - 1\right\}^m \\ &= B_0\Theta^{(1+m)/m}\left\{\Theta^{-1/m} - 1\right\}^m\end{aligned} \tag{42}$$

where $B_0 = -\frac{\alpha m\mathrm{sign}(h)}{1-m}$

Using Eq. (42), $\frac{\partial h}{\partial\theta}$ could be obtained as

$$\frac{\partial h}{\partial\theta} = \left(\frac{\partial\Theta}{\partial h}\right)^{-1}\frac{1}{\theta_s - \theta_r} = \frac{1}{B_0}\Theta^{-(1+m)/m}\left\{\Theta^{-1/m} - 1\right\}^{-m} \tag{43}$$

The Richards equation can be factorized into three terms as

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial z}\left[k\frac{\partial h}{\partial z}\right] + \frac{\partial k}{\partial z} \tag{44}$$

$$= \underbrace{\frac{\partial k}{\partial z}\frac{\partial h}{\partial z}}_{\text{Term1}} + \underbrace{k\frac{\partial^2 h}{\partial z^2}}_{\text{Term2}} + \underbrace{\frac{\partial k}{\partial z}}_{\text{Term3}} \tag{45}$$

Term1:

$$\frac{\partial k}{\partial z}\frac{\partial h}{\partial z} = \text{Term3}\left[\frac{\partial h}{\partial \theta}\frac{\partial \theta}{\partial z}\right] \tag{46}$$

$$= \text{Term3}\left[\left\{\left(\frac{\partial \Theta}{\partial h}\right)^{-1}\frac{1}{\theta_s - \theta_r}\right\}\frac{\partial \theta}{\partial z}\right] \tag{47}$$

Term1 is obtained by substituting (42) in (47).

Term2:

$$\begin{aligned}\frac{\partial^2 h}{\partial z^2} &= \frac{\partial}{\partial z}\left(\frac{\partial h}{\partial z}\right)\\ &= \frac{\partial}{\partial z}\left(\frac{\partial h}{\partial \theta}\frac{\partial \theta}{\partial z}\right)\\ &= \frac{\partial^2 h}{\partial z \partial \theta}\frac{\partial \theta}{\partial z} + \left(\frac{\partial h}{\partial \theta}\right)\frac{\partial^2 \theta}{\partial z^2}\end{aligned} \tag{48}$$

Deriving $\frac{\partial^2 h}{\partial z \partial \theta}$ by differentiating (43) with respect to z.

$$\begin{aligned}\frac{\partial}{\partial z}\left(\frac{\partial h}{\partial \theta}\right) = &-\frac{1}{B_0}\left(\frac{1+m}{m}\right)\Theta^{-(1+2m)/m}\left\{\Theta^{-1/m}-1\right\}^{-m}\frac{\partial \Theta}{\partial z}\\ &+\frac{1}{B_0}\Theta^{-(1+m)/m}\left\{\Theta^{-1/m}-1\right\}^{-m-1}\left(\Theta^{-(1+m)/m}\right)\frac{\partial \Theta}{\partial z}\end{aligned} \tag{49}$$

$$\frac{\partial}{\partial z}\left(\frac{\partial h}{\partial \theta}\right) = \frac{1}{B_0}\frac{\partial \Theta}{\partial z}\times \left[\begin{aligned}&-\left(\frac{1+m}{m}\right)\Theta^{-(1+2m)/m}\left(\Theta^{-1/m}-1\right)^{-m}\\ &+\Theta^{-(2+2m)/m}\left(\Theta^{-1/m}-1\right)^{-m-1}\end{aligned}\right] \tag{50}$$

$$\frac{\partial^2 h}{\partial z \partial \theta} = \frac{1}{B_0}\frac{1}{\theta_s - \theta_r}\frac{\partial \theta}{\partial z}\times \left[\begin{aligned}&-\left(\frac{1+m}{m}\right)\Theta^{-(1+2m)/m}\left\{\Theta^{-1/m}-1\right\}^{-m}\ldots\\ &+\Theta^{-(2+2m)/m}\left(\Theta^{-1/m}-1\right)^{-m-1}\end{aligned}\right] \tag{51}$$

Term2 is obtained by substituting (43) and (51) in (48)

Term3:

$$\frac{\partial k}{\partial z} = \frac{\partial k}{\partial \theta}\frac{\partial \theta}{\partial z} \tag{52}$$

$$\frac{\partial k}{\partial \theta} = \frac{\partial k}{\partial \Theta}\frac{\partial \Theta}{\partial \theta} \tag{53}$$

Using (36), (53) and (37), Eq. (52) is expressed as

$$\frac{\partial k}{\partial z} = \frac{\partial k}{\partial \theta}\frac{\partial \theta}{\partial z} = \frac{1}{\theta_s - \theta_r}k_s\frac{\partial \theta}{\partial z}\times \left[\begin{aligned}&2\Theta^{1/2}\left\{1-\left(1-\Theta^{1/m}\right)^m\right\}\left\{\left(1-\Theta^{1/m}\right)^{m-1}\left(\Theta^{1/m-1}\right)\right\}\\ &+\frac{1}{2}\Theta^{-1/2}\left\{1-\left(1-\Theta^{1/m}\right)^m\right\}^2\end{aligned}\right] \tag{54}$$

References

Bilionis I (2016) Probabilistic solvers for partial differential equations. arXiv preprint arXiv:160703526

Bishop CM (2006) Pattern recognition and machine learning. Springer, Berlin

Calderhead B, Girolami M, Lawrence ND (2008) Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes. In: Advances in neural information processing systems, pp 217–224

Celia MA, Bouloutas ET, Zarba RL (1990) A general mass-conservative numerical solution for the unsaturated flow equation. Water Resour Res 26(7):1483–1496

Cockayne J, Oates CJ, Sullivan T, Girolami M (2016) Probabilistic meshless methods for partial differential equations and Bayesian inverse problems. arXiv preprint arXiv:160507811

Deb K (2012) Optimization for engineering design: algorithms and examples. PHI Learning Pvt. Ltd., New Delhi

Dondelinger F, Husmeier D, Rogers S, Filippone M (2013) ODE parameter inference using adaptive gradient matching with Gaussian processes. In: Artificial intelligence and statistics, pp 216–228

Farthing MW, Ogden FL (2017) Numerical solution of richards equation: a review of advances and challenges. Soil Sci Soc Am J 81:1257–1269

Fletcher R (2013) Practical methods of optimization. Wiley, Hoboken

Gong W, Duan Q (2017) An adaptive surrogate modeling-based sampling strategy for parameter optimization and distribution estimation (ASMO-PODE). Enviro Model Softw 95:61–75

Graepel T (2003) Solving noisy linear operator equations by Gaussian processes: application to ordinary and partial differential equations. In: ICML, pp 234–241

Hinton GE (2002) Training products of experts by minimizing contrastive divergence. Neural Comput 14(8):1771–1800

Holman D, Sridharan M, Gowda P, Porter D, Marek T, Howell T, Moorhead J (2014) Gaussian process models for reference ET estimation from alternative meteorological data sources. J Hydrol 517:28–35

Kool J, Parker J, van Genuchten MT (1985) Determining soil hydraulic properties from one-step outflow experiments by parameter estimation: I. Theory and numerical studies. Soil Sci Soc Am J 49(6):1348–1354

Krige DG (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. J South Afr Inst Min Metall 52(6):119–139

Lagarias JC, Reeds JA, Wright MH, Wright PE (1998) Convergence properties of the nelder-mead simplex method in low dimensions. SIAM J Optim 9(1):112–147

Macdonald B, Husmeier D (2015) Gradient matching methods for computational inference in mechanistic models for systems biology: a review and comparative analysis. Front Bioeng Biotechnol 3:180

Macdonald B, Higham C, Husmeier D (2015) Controversy in mechanistic modelling with Gaussian processes. In: International conference on machine learning, pp 1539–1547

Marrel A, Iooss B, Laurent B, Roustant O (2009) Calculations of sobol indices for the gaussian process metamodel. Reliab Eng Syst Saf 94(3):742–751

Matheron G (1963) Principles of geostatistics. Econ Geol 58(8):1246–1266

Mayraz G, Hinton GE (2001) Recognizing hand-written digits using hierarchical products of experts. In: Advances in neural information processing systems, pp 953–959

McLaughlin D, Townley LR (1996) A reassessment of the groundwater inverse problem. Water Resour Res 32(5):1131–1161

Newton I (1736) The method of fluxions and infinite series. Henry Woodfall, London

Nocedal J, Wright S (1999) Numerical optimization. Springer, New York

Poyton A, Varziri MS, McAuley KB, McLellan P, Ramsay JO (2006) Parameter estimation in continuous-time dynamic models using principal differential analysis. Comput Chem Eng 30(4):698–708

Raissi M, Perdikaris P, Karniadakis GE (2017a) Inferring solutions of differential equations using noisy multi-fidelity data. J Comput Phys 335:736–746

Raissi M, Perdikaris P, Karniadakis GE (2017b) Machine learning of linear differential equations using Gaussian processes. J Comput Phys 348:683–693

Raissi M, Perdikaris P, Karniadakis GE (2018) Numerical Gaussian processes for time-dependent and nonlinear partial differential equations. SIAM J Sci Comput 40(1):A172–A198

Ramsay JO, Hooker G, Campbell D, Cao J (2007) Parameter estimation for differential equations: a generalized smoothing approach. J R Stat Soc Ser B (Stat Methodol) 69(5):741–796

Rao CR (2008) Cramr-Rao bound. Scholarpedia 3(8):6533. https://doi.org/10.4249/scholarpedia.6533

Rao SS (2009) Engineering optimization: theory and practice. Wiley, Hoboken

Rasmussen CE, Williams CK (2006) Gaussian processes for machine learning, vol 1. MIT Press, Cambridge

Reynolds RJ (1987) Diffusivity of glacial-outwash aquifer by the floodwave-response technique. Groundwater 25(3):290–299

Richards LA (1931) Capillary conduction of liquids through porous mediums. Physics 1(5):318–333

Ritter A, Hupet F, Muñoz-Carpena R, Lambot S, Vanclooster M (2003) Using inverse methods for estimating soil hydraulic properties from field data as an alternative to direct methods. Agric Water Manag 59(2):77–96

Rohatgi A (2011) WebPlotDigitizer. http://arohatgi.info/WebPlotDigitizer/app. Accessed 15 Feb 2018

Särkkä S (2011) Linear operators and stochastic partial differential equations in Gaussian process regression. In: International conference on artificial neural networks. Springer, pp 151–158

Simunek J, van Genuchten MT, Sejna M (2005) The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. Univ Calif-Riverside Res Rep 3:1–240

Solak E, Murray-Smith R, Leithead WE, Leith DJ, Rasmussen CE (2003) Derivative observations in Gaussian process models of dynamic systems. In: Advances in neural information processing systems, pp 1057–1064

Srivastava R (2006) Aquifer diffusivity estimation from response to stream stage variation. J Hydrol Eng 11(3):273–277

Strang G (2014) Differential equations and linear algebra. Wellesley-Cambridge Press, Wellesley

Sun AY, Wang D, Xu X (2014) Monthly streamflow forecasting using Gaussian process regression. J Hydrol 511:72–81

Swamee PK, Singh SK (2003) Estimation of aquifer diffusivity from stream stage variation. J Hydrol Eng 8(1):20–24

Tripathi S, Govindaraju RS (2007) On selection of kernel parametes in relevance vector machines for hydrologic applications. Stoch Environ Res Risk Assess 21(6):747–764

van Genuchten MT (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci Soc Am J 44(5):892–898

Varah JM (1982) A spline least squares method for numerical parameter estimation in differential equations. SIAM J Sci Stat Comput 3(1):28–46

Vyshemirsky V, Girolami MA (2007) Bayesian ranking of biochemical system models. Bioinformatics 24(6):833–839

Wang C, Duan Q, Gong W, Ye A, Di Z, Miao C (2014) An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. Environ Model Softw 60:167–179

Wang Y, Barber D (2014) Gaussian processes for Bayesian estimation in ordinary differential equations. In: International conference on machine learning, pp 1485–1493

Wu A, Aoi MC, Pillow JW (2017) Exploiting gradients and hessians in Bayesian optimization and Bayesian quadrature. arXiv preprint arXiv:170400060

Zhang J, Li W, Zeng L, Wu L (2016) An adaptive Gaussian process-based method for efficient Bayesian experimental design in groundwater contaminant source identification problems. Water Resour Res 52(8):5971–5984

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.