GEM - International Journal on Geomathematics (2024) 15:11
https://doi.org/10.1007/s13137-024-00250-3

ORIGINAL PAPER

Uncertainty modeling and propagation for groundwater flow: a comparative study of surrogates

Oliver G. Ernst[1] · Björn Sprungk[2] · Chao Zhang[3]

Received: 25 March 2023 / Accepted: 13 June 2024

Abstract

We compare sparse grid stochastic collocation and Gaussian process emulation as surrogates for the parameter-to-observation map of a groundwater flow problem related to the Waste Isolation Pilot Plant in Carlsbad, NM. The goal is the computation of the probability distribution of a contaminant particle travel time resulting from uncertain knowledge about the transmissivity field. The latter is modelled as a lognormal random field which is fitted by restricted maximum likelihood estimation and universal kriging to observational data as well as geological information including site-specific trend regression functions obtained from technical documentation. The resulting random transmissivity field leads to a random groundwater flow and particle transport problem which is solved realization-wise using a mixed finite element discretization. Computational surrogates, once constructed, allow sampling the quantities of interest in the uncertainty analysis at substantially reduced computational cost. Special emphasis is placed on explaining the differences between the two surrogates in terms of computational realization and interpretation of the results. Numerical experiments are given for illustration.

Keywords Sparse grid stochastic collocation · Gaussian process emulation · Uncertainty propagation · Kriging · Darcy flow · Mixed finite elements

Dedicated to the memory of K. Andrew Cliffe (1953–2014).

Björn Sprungk and Chao Zhang contributed equally to this work.

✉ Oliver G. Ernst
oernst@math.tu-chemnitz.de

Björn Sprungk
bjoern.sprungk@math.tu-freiberg.de

Chao Zhang
chaz@dtu.dk

[1] Department of Mathematics, TU Chemnitz, Chemnitz, Germany

[2] Faculty of Mathematics and Computer Science, TU Bergakademie Freiberg, Freiberg, Germany

[3] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark

Published online: 16 July 2024

Mathematics Subject Classification 60G60 · 60H35 · 62P12 · 62M30 · 65C05 · 65D12 · 65C30 · 65N75

1 Introduction

By their very nature, the earth sciences have had to cope with uncertainty from early on, and scientists from this field such as Harold Jeffreys and Albert Tarantola have had foundational and lasting impact on how uncertainty is modeled and merged with physical models in the interdisciplinary field now known as *uncertainty quantification (UQ)*. A current account of uncertainty quantification in subsurface hydrology can be found in Linde et al. (2017). Many UQ studies involve a system governed by a partial differential equation (PDE) in which one or more input quantities are uncertain. When this uncertainty is described in probabilistic terms we arrive at a PDE with random data, or *random PDE* for short. Such random data may be modeled by one or more scalar random variables or, in case of distributed quantities, random functions which in mathematical terms are stochastic processes indexed by space and/or time and in this context usually referred to as *random fields*. In all these cases the solution of the random PDE is also a random field. The task of determining the probability distribution of the solution of a random PDE, or of *quantities of interest* derived from such solutions, is known as *uncertainty propagation* or *forward UQ* (cf. Ernst et al. 2022). Approximation methods for random fields and their incorporation into computational solution methods for random PDEs have been actively developed in the engineering and numerical analysis communities in the past two decades, and excellent surveys can be found in Ghanem and Spanos (1991); Babuška et al. (2010); Schwab and Gittelson (2011); Gunzburger et al. (2014). The distinguishing feature of these approaches is that they parameterize the approximate random PDE solution or functionals thereof as functions—typically polynomials—of a set of independent reference random variables whose number can be large or even countably infinite. Reflecting the construction principles on which these approximations are based, the approaches are called *stochastic Galerkin* or *stochastic collocation* methods. At the same time, sampling-based simulation techniques known as *Gaussian process emulators* have gained popularity in the statistics community for solving similar problems, cf. Sacks et al. (1989); Currin et al. (1991); Kennedy and O'Hagan (2001); O'Hagan (2006). Here the random solution is modeled as a Gaussian process conditioned on realizations of the solutions obtained for certain realizations of the random inputs.

Our objective in this work is the direct comparison of these two approaches using Monte Carlo sampling as a reference in a case study on the hydrogeological transport of radionuclides within the site assessment for a nuclear waste repository. In doing so, we place particular emphasis on the careful construction of a stochastic model of the random PDE data—in this case a lognormal random field modeling the uncertain hydraulic transmissivity—using geostatistical techniques based on observational data of transmissivity and hydraulic head as well as additional geological background information. Besides the computational efficiency and approximation qualities of the two approaches, we provide an introduction to both methods highlighting the assumptions on which they are based and consequences for interpreting the results obtained with each.

The uncertainty propagation techniques we shall consider are based on generating realizations (samples) of the uncertain input parameters, solving the PDE for each realization and then determining the statistical properties of the quantities of interest in a post-processing step. As each PDE solution typically requires considerable computational resources, the mapping of random input parameters to quantities of interest is often substituted by *surrogate models*, which are considerably less costly to evaluate, thus speeding up the uncertainty propagation analysis. The two surrogates we shall compare, *sparse polynomial collocation* and Gaussian process emulation are interesting in that they were developed in different fields (numerical analysis and statistics), display different performance characteristics, and also differ in the interpretations of the surrogates they produce. Our work is closest in spirit to Owen et al. (2017), where Gaussian process emulation is compared with polynomial chaos expansion surrogates for two black-box computer simulators. Although different in construction, polynomial chaos surrogates yield a multivariate polynomial approximation of the input–output map realized by the computer simulator as does stochastic collocation, whereas the latter is considerably easier to integrate into PDE solvers. In place of a small number of discrete parameters in the models considered in Owen et al. (2017), the random input in our groundwater model is a random field, i.e., its realizations are functions, which can be considered as parameterized by a countably infinite number of parameters. The propagation of geometry-induced uncertainties in aerodynamic modeling using surrogate models based on quasi-Monte-Carlo quadrature as well as kriging and radial basis techniques is compared in Liu et al. (2017). An overview of surrogate models for uncertainty quantification can be found in Sudret et al. (2017).

The remainder of the paper is organized as follows: Section 2 presents the problem of predicting the travel or *exit time* of radionuclides transported by groundwater flow through a horizontal layer above the Waste Isolation Pilot Plant, an operational underground disposal site for nuclear waste, in a scenario where a hypothetical future accidental breach leads to the release of radioactive material. The physical as well as the probabilistic model are presented as well as how observational data of hydraulic transmissivity is incorporated, leading to the generation of samples of the exit time quantity of interest. Section 3 describes the computational realization for solving the Darcy flow equations, the construction of the truncated Karhunen-Loève representation of the random transmissivity field as well as the estimation of the cumulative distribution function of the exit time quantity of interest. Section 4 gives detailed description of the two surrogate types to be compared, Gaussian process emulation and sparse polynomial collocation, emphasising their differences with respect to construction, computation and interpretation. In Sect. 5, we present the results of numerical computations with both surrogates using original data from the WIPP site, and present our conclusions in Sect. 6.

2 Uncertainty propagation for a groundwater flow problem

In this section we introduce the application setting, physical model, UQ task as well as the probabilistic model with which this is addressed.

2.1 The waste isolation pilot plant (WIPP)

The Waste Isolation Pilot Plant (WIPP) in Carlsbad, NM, is a long-term deep geologic storage facility for transuranic waste operated by the U.S. Department of Energy since 1999. One of the issues investigated in the course of an extensive performance assessment for WIPP was the risk of hazardous materials escaping to the biosphere in the event of a future accidental breach of the enclosure system. As the most likely pathway for such contaminants is transport through the subsurface via groundwater, we are led to the objective of predicting the groundwater flow and transport of contaminants released from the storage site. The WIPP disposal area lies within in the *Salado* bedded salt formation. The Salado itself as well as the overlying formations are essentially impermeable to groundwater with the exception of a laterally extensive but narrow layer of rock known as the *Culebra Dolomite*. Details of the geological site characterization can be found in the extensive documentation[1] in the WIPP certification and recertification applications (U.S. Department of Energy (DOE) 2004, 2014) which are produced every five years. Figure 1, taken from (U.S. Department of Energy (DOE) 2014), shows the location of the WIPP site within the UTM coordinate system, the location of boreholes where measurements of transmissivity and hydraulic head were obtained as well as the boundaries of areas with distinct geological features.

A highly relevant quantity of interest in this context is the travel or *exit time* of radionuclides after release from a point within the Culebra layer above the site to reach the boundary of the repository area, the computation of which requires simulating the groundwater flow and transport in the Culebra. As the precise transmissivity properties of the rock are uncertain, the same applies to the exit time. In the remainder of this section we describe a model for groundwater flow and contaminant transport in which the uncertain transmissivity is modeled stochastically, incorporating geological background information, standard geostatistical assumptions as well as available measurement data.

2.2 Darcy flow and particle transport

We model the flow of groundwater through the Culebra dolomite geological unit by stationary single-phase Darcy flow. Denoting by p the *hydraulic head* (pressure) and by K the (scalar) *hydraulic conductivity*, the *volumetric flux* (Darcy flux) $\boldsymbol{q}$ is given by

$$\boldsymbol{q} = -K \nabla p. \tag{1}$$

[1] These can be found at https://wipp.energy.gov/epa-certification-documents.asp.

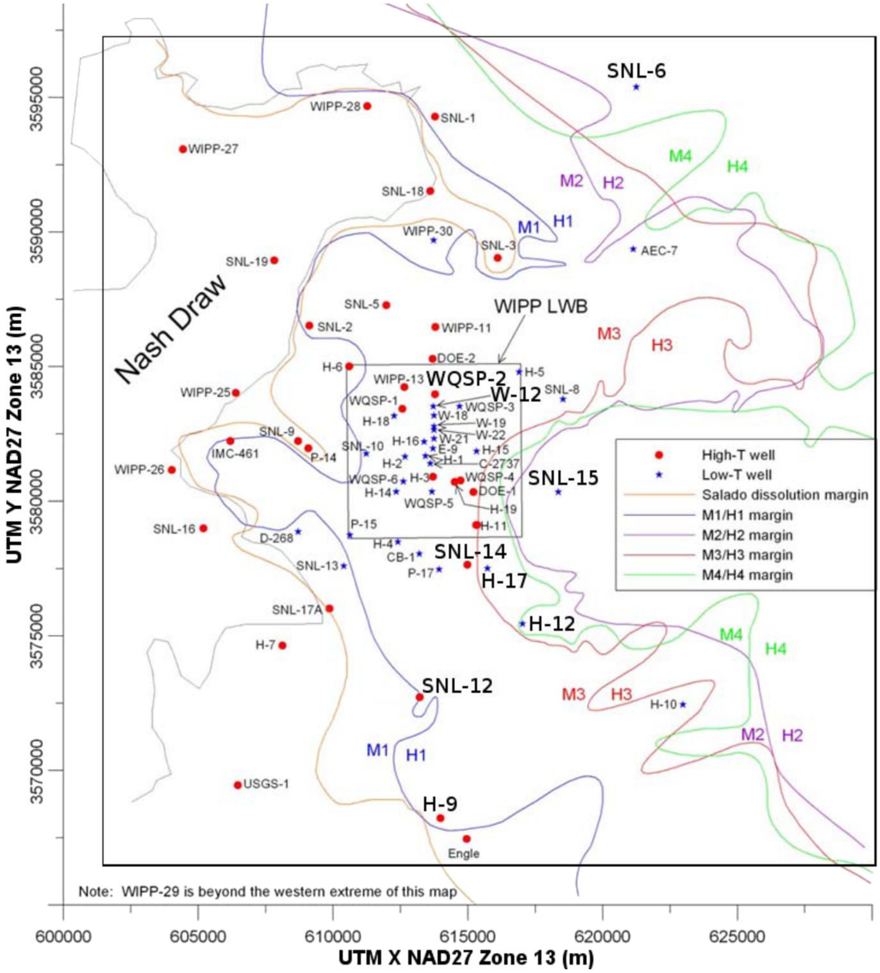


Fig. 1 Horizontal location of WIPP repository (small black square, land withdrawal boundary LWB), observation boreholes with markers indicating low and high transmissivity values as well as boundaries of distinct geological features; these are accounted for in the trend model of the transmissivity field in Sect. 2.4.1. Source: (U.S. Department of Energy (DOE) 2014)

If $\boldsymbol{u}$ denotes the pore velocity of the groundwater, which is related to the Darcy flux in terms of the *porosity* ϕ as $\boldsymbol{q} = \phi\boldsymbol{u}$, conservation of mass in the absence of sources and sinks leads to the divergence-free condition

$$\nabla\cdot \boldsymbol{u} = 0. \tag{2}$$

Since the aquifer under consideration is essentially horizontal with a much larger lateral than vertical extent, we model the flow as two-dimensional and consider the hydraulic *transmissivity* $T = bK$ in place of conductivity, where b denotes the aquifer thickness.

On the boundary ∂D of the bounded computational domain D, we distinguish impermeable segments Γ_N along which the normal flux vanishes and their complement $\Gamma_D = \partial D \backslash \Gamma_N$, where we prescribe the value of the hydraulic head p. Denoting by $\boldsymbol{n}$ the exterior unit normal vector along Γ_N and by g the prescribed head data along Γ_D, this leads to the boundary conditions

$$\boldsymbol{n} \cdot \boldsymbol{u} = 0 \ \text{ on } \Gamma_N, \qquad p = g \ \text{ on } \Gamma_D. \tag{3}$$

The computational domain D as well as the boundary segments Γ_N and Γ_D are displayed in the left panel in Fig. 2. The Dirichlet data g is obtained by evaluating a kriging interpolant (cf. Sect. 2.4.4) of observational hydraulic head data taken from (U.S. Department of Energy (DOE) 2014). As the flux variable $\boldsymbol{u}$ is of primary interest in view of the subsequent transport calculation we employ the usual mixed formulation of the boundary value problem presented by (1), (2) and (3). The associated variational formulation consists in finding the pair $(\boldsymbol{u}, p) \in \mathcal{V} \times \mathcal{W}$ such that

$$\left(\frac{\phi b}{T}\boldsymbol{u}, \boldsymbol{v}\right) - (p, \nabla\cdot \boldsymbol{v}) = -\langle g, \boldsymbol{n} \cdot \boldsymbol{v}\rangle_{\Gamma_D} \qquad \forall \boldsymbol{v} \in \mathcal{V}, \tag{4a}$$

$$(\nabla\cdot \boldsymbol{u}, q) = 0 \qquad \forall q \in \mathcal{W} \tag{4b}$$

with suitable boundary data $g \in H^{1/2}(\Gamma_D)$. Here $(\cdot, \cdot)$ denotes the $L^2(D)$ inner product, the variational spaces are given by

$$\mathcal{V} = \{\boldsymbol{v} \in \boldsymbol{H}(\text{div}; D), \boldsymbol{n} \cdot \boldsymbol{v}|_{\Gamma_N} = 0\}, \qquad \mathcal{W} = L^2(D)$$

and $\langle \cdot, \cdot \rangle_{\Gamma_D}$ denotes the duality pairing $H^{1/2}(\Gamma_D) \times H^{-1/2}(\Gamma_D)$. Given the flux solution $\boldsymbol{u}$ of (4), the trajectory of a particle from a release point $\boldsymbol{x}_0 \in D$ neglecting hydraulic dispersion is found as the solution of the initial value problem

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{u}(\boldsymbol{x}(t)), \qquad t \geq 0, \quad \boldsymbol{x}(0) = \boldsymbol{x}_0. \tag{5}$$

A discussion of the regularity requirements for the Darcy flow problem (4) needed to ensure existence and uniqueness of the particle trajectory (5) can be found in (Graham et al. (2016), Section 5.3). As we shall see below, for the probabilistic model of transmissivity with finite-dimensional noise, which we shall employ in our calculations, these requirements are satisfied. As a *quantity of interest* derived from the solution of the random Darcy flow equations, we choose the logarithm of the travel or exit time of a particle released at a location $\boldsymbol{x}_0$ inside the Culebra layer above the WIPP repository until it reaches the boundary of the subdomain $D_0 \subset D$ marking the edge of the WIPP site projected vertically up to the Culebra layer within the surrounding computational domain D,

$$f_{\text{exit}} := \log \min\{t > 0 : \boldsymbol{x}(t) \notin D_0, \ \boldsymbol{x}_0 \in D_0\}.$$

The location of the release point $\boldsymbol{x}_0$, the perimeter of the WIPP site D_0 as well as a number of particle trajectory realizations from $\boldsymbol{x}_0$ to ∂D_0 are displayed in Fig. 2.

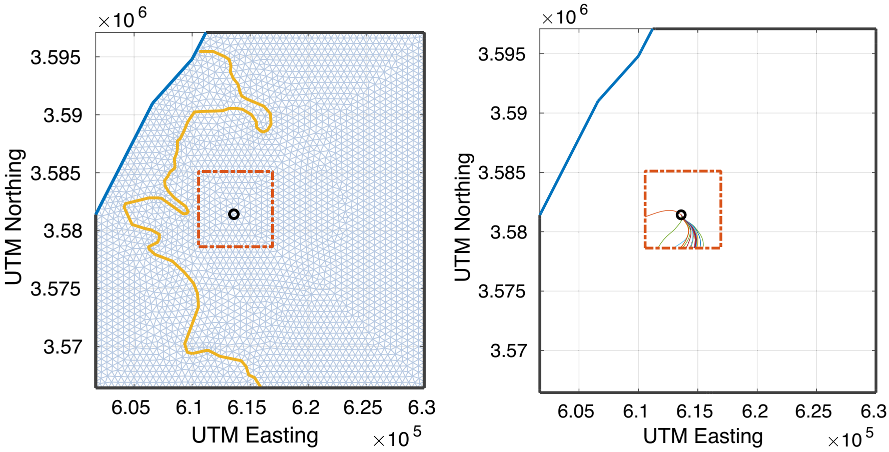


Fig. 2 Left: Computational domain D with Neumann boundary Γ_N (blue) and Dirichlet boundary Γ_D (black) as well as the perimeter of the WIPP site D_0 (red dashed), location of particle release point $\boldsymbol{x}_0$ (black circle), and boundary of the Salado dissolution zone D_1 (yellow), cf. Section 2.4.1 below, respected by the triangular finite element mesh. Right: Simulation of several realizations of random particle trajectories from $\boldsymbol{x}_0$ to ∂D_0 (colour figure online)

2.3 Probabilistic modeling of uncertain transmissivity

The primary source of uncertainty in the modeling of flow and transport in the Culebra dolomite is the spatial variation of hydraulic conductivity, or, in our horizontal two-dimensional setting, transmissivity T. The prevailing mathematical description of uncertainty is probabilistic, i.e., the quantities in question are modeled as random variables following a given probability distribution. The randomness thus introduced is an expression of uncertainty due to lack of knowledge of the precise spatial variation of transmissivity throughout the domain D in the sense that some realizations of transmissivity across the domain are more likely than others. Rather than a deterministic value $T = T(\boldsymbol{x})$, transmissivity at a point $\boldsymbol{x} \in D$ (scaled by porosity and thickness) is thus expressed as a random variable $T(\boldsymbol{x}, \omega)$ governed by a probability measure $\mathbf{P}$ defined on a probability space $(\Omega, \mathfrak{A}, \mathbf{P})$ with elementary outcome set Ω carrying a σ-algebra $\mathfrak{A}$ on which a probability measure $\mathbf{P}$ is defined. The collection of all such random variables $\{T(\boldsymbol{x}, \omega) : \boldsymbol{x} \in D\}$ is known as a *random field*, i.e., a stochastic process for which the index variable $\boldsymbol{x}$ is a spatial coordinate.[2] The most well-established probabilistic model for transmissivity in the hydrology literature assumes that $T(\boldsymbol{x}, \cdot)$ follows a *lognormal* distribution, i.e., that $Z(\boldsymbol{x}, \cdot) := \log T(\boldsymbol{x}, \cdot)$ is a Gaussian random field (cf. Freeze (1975), Hoeksema and Kitanidis (1985) and de Marsily (1986), Chapter 11)). By consequence, realizations of $T = \exp(Z)$ are always positive. Such a Gaussian random field Z is completely specified by its mean and covariance function

[2] We will, following statistical convention, omit the random field argument ω (or dot) denoting the elementary event for typographical convenience except when we wish to emphasize its random nature.

$$\begin{aligned} \overline{Z}(\boldsymbol{x}) &= \mathbf{E}\left[Z(\boldsymbol{x})\right], & \boldsymbol{x} \in D, \\ \text{and} \quad c(\boldsymbol{x}, \boldsymbol{y}) &= \mathbf{E}\left[(Z(\boldsymbol{x}) - \overline{Z}(\boldsymbol{x}))(Z(\boldsymbol{y}) - \overline{Z}(\boldsymbol{y}))\right], & \boldsymbol{x}, \boldsymbol{y} \in D, \end{aligned}$$

respectively, where $\mathbf{E}[\cdot]$ denotes mathematical expectation with respect to $\mathbf{P}$.

We assume throughout that the covariance function of $Z = \log T$ is *isotropic* and that the fluctuation $Z - \overline{Z}$ is *wide-sense stationary* such that we have $c(\boldsymbol{x}, \boldsymbol{y}) = c(|\boldsymbol{x} - \boldsymbol{y}|)$, i.e., the covariance depends only on the (Euclidean) separation distance $r = |\boldsymbol{x} - \boldsymbol{y}|$. Moreover, we assume $c(r)$ to belong to the *Matérn* family of covariance models

$$c(r) = \frac{\sigma^2}{2^{\nu-1}\,\Gamma(\nu)} \left(\frac{2\sqrt{\nu}\,r}{\rho}\right)^{\nu} K_\nu\left(\frac{2\sqrt{\nu}\,r}{\rho}\right), \qquad r = |\boldsymbol{x} - \boldsymbol{y}|, \tag{6}$$

where K_ν denotes the modified Bessel function of order $\nu > 0$. The quantity ν is called the *smoothness parameter*, $\sigma^2 = c(0) = \mathbf{Var}\, Z(\boldsymbol{x})$ is the (marginal) *variance* (constant in $\boldsymbol{x}$) and the parameter $\rho > 0$ is called the *correlation length*, a measure of how quickly the covariance decays with separation distance. A detailed justification for using the Matérn model as well as a discussion of its properties and scaling variants can be found in Stein (1999, pp. 48).

For the particular scaling (6), the Matérn covariance coincides with the exponential covariance for $\nu = \frac{1}{2}$, the Bessel covariance for $\nu = 1$ and the squared exponential covariance in the limit $\nu \to \infty$. The smoothness of the realizations of Z increases with ν, and the spatial scale of variation is described by ρ. We determine the values of the *hyperparameters* (σ, ρ, ν) by statistical estimation based on data published in the WIPP Compliance Recertification Assessment U.S. Department of Energy (DOE) (2014) documents, which contain measurements of transmissivity in the Culebra dolomite at 62 boreholes throughout the assessment site (cf. Fig. 1). Figure 3 displays realizations of a Gaussian random field describing $Z = \log T$ throughout the computational domain D representing the Culebra flow domain. It can be seen that larger values of ν result in realizations that are smoother, and smaller values of ρ lead to structures which decorrelate faster with separation distance.

2.4 Statistical estimation of transmissivity field

As described in Sect. 2.3, we model the uncertain hydraulic transmissivity T as a lognormal random field on the bounded simulation domain $D \subset \mathbb{R}^2$, so that the random field

$$Z := \log T = \overline{Z}(\boldsymbol{x}) + \tilde{Z}(\boldsymbol{x}, \omega) \tag{7}$$

is Gaussian with (deterministic) mean $\overline{Z}$ and (centered) residual field $\tilde{Z}$. Due to the complexity and irregular features of geological structures, it is crucial to merge the stochastic model with available measurement data in a transparent fashion. Below we summarize the statistical techniques by which available data is incorporated into the stochastic model of uncertain transmissivity. Its construction proceeds in three steps:

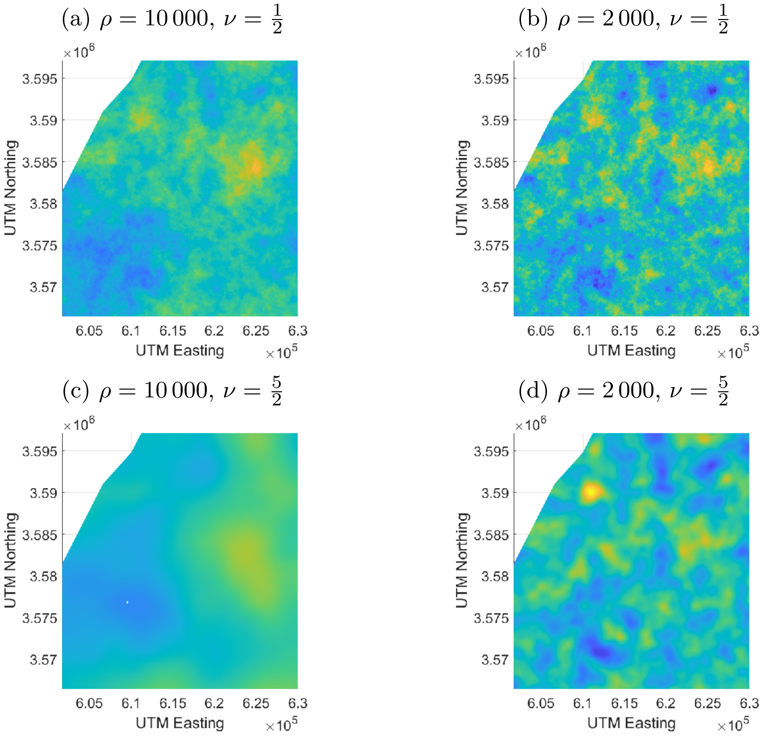


Fig. 3 Realizations of mean-zero Gaussian random fields with Matérn covariance function for different values of ρ and ν. All plots use the same color map and σ^2 was set to 1 in each case

(1) The assumptions that T follows a lognormal distribution and that the covariance function of $\log T$ belongs to the Matérn class;
(2) The parameters σ, ν and ρ in the Matérn covariance function (6) are determined by *restricted maximum likelihood estimation (RML)*;
(3) The lognormal field thus obtained is then further conditioned on the available observations of transmissivity at the WIPP site.

We present some background on these techniques and how they are applied to our model of WIPP transmissivity in the following subsections.

2.4.1 Regression model of mean transmissivity

The deterministic mean $\overline{Z}$ of the log-transmissivity field is constructed as a linear regression model

$$\overline{Z}(\boldsymbol{x}) = \sum_{j=1}^{k} \beta_j h_j(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x})^\top \boldsymbol{\beta}, \qquad \boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} h_1(\boldsymbol{x}) \\ \vdots \\ h_k(\boldsymbol{x}) \end{bmatrix}, \tag{8}$$

in which the k components of $\boldsymbol{h}$ consist of *regression functions* from which an approximate trend behavior of Z can be obtained by linear combination. Known geological features of the area under study can be incorporated by choosing the regression functions as, e.g., indicator functions of subdomains possessing distinguishing characteristics, linear or polynomial trends to be fitted as well as the variation of available quantities known or believed to affect the transmissivity field. Based on the available WIPP technical documents, a model comparison was made using the five regression functions

$$\begin{aligned} h_1(\boldsymbol{x}) &\equiv 1 && \text{(constant)}, & h_4(\boldsymbol{x}) &= d(\boldsymbol{x}) && \text{(overburden)},\\ h_2(\boldsymbol{x}) &= x_1 && \text{(linear in } x_1), & h_5(\boldsymbol{x}) &= \mathbb{1}_{D_1}(\boldsymbol{x}) && \text{(zone indicator)}.\\ h_3(\boldsymbol{x}) &= x_2 && \text{(linear in } x_2), \end{aligned} \tag{9}$$

The first three regression functions allow to fit a basic affine trend. The *overburden* $d(\boldsymbol{x})$ denotes the vertical distance between the ground surface and the top of the Culebra layer above location $\boldsymbol{x}$. This is an indication of the extent to which erosion has led to stress relief on the underlying Culebra layer, possibly causing new fracturing or the opening of pre-existing fractures and thereby enhancing transmissivity. Regression function h_5 is the indicator function of a subdomain $D_1 \subset D$ to the north, south and west of the WIPP site, where dissolution of the upper Salado formation has led to strain in the overlying rock, including the Culebra, leading to larger apertures in existing fractures, collapse and brecciation and thus to a generally higher transmissivity (cf. U.S. Department of Energy (DOE) 2004).

2.4.2 Restricted maximum likelihood estimation

Under the models for the mean (8) and covariance structure (6), the Gaussian log-transmissivity field (7) has the covariance function $c_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y})$, where $\boldsymbol{\theta} = (\sigma^2, \rho, \nu)$ denotes the triplet of parameters consisting of variance σ^2, correlation length ρ and smoothness parameter ν. The specification of the probabilistic model for the random field Z consists in determining the vector $\boldsymbol{\beta}$ of regression coefficients and the covariance parameter vector $\boldsymbol{\theta}$. It is desired that estimation techniques for these based on observations be *unbiased*, i.e., that the average estimation error is zero, and that this error be optimal in a least squares sense. Another desirable property is *consistency*, whereby the estimates converge to the true values as more and more observations are added.

The restriction of Z to a finite set of observation points $\{\boldsymbol{x}_j\}_{j=1}^n \subset D$ forms a multivariate Gaussian random vector, which we denote by

$$\boldsymbol{Z}\colon \Omega \to \mathbb{R}^n, \qquad \omega \mapsto \boldsymbol{Z}(\omega) = \begin{bmatrix} Z(\boldsymbol{x}_1, \omega) \\ \vdots \\ Z(\boldsymbol{x}_n, \omega) \end{bmatrix}. \tag{10}$$

In view of (7), its expectation is

$$\mathsf{E}[\boldsymbol{Z}] = \boldsymbol{H\beta}, \qquad [\boldsymbol{H}]_{i,j} = h_j(\boldsymbol{x}_i), \quad i = 1, \ldots, n, \quad j = 1, \ldots, k,$$

and its joint probability density function given for $\boldsymbol{\xi} \in \mathbb{R}^n$ by

$$p(\boldsymbol{\xi}; \boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^n \det \boldsymbol{C_\theta}}} \exp\left(-\frac{1}{2}(\boldsymbol{\xi} - \boldsymbol{H\beta})^\top \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\xi} - \boldsymbol{H\beta})\right), \tag{11}$$

in which $\boldsymbol{C_\theta}$ denotes the covariance matrix

$$\boldsymbol{C_\theta} = \mathsf{E}\left[(\boldsymbol{Z} - \boldsymbol{H\beta})(\boldsymbol{Z} - \boldsymbol{H\beta})^\top\right] = \left[c_{\boldsymbol{\theta}}(\boldsymbol{x}_i, \boldsymbol{x}_j)\right]_{i,j=1}^n \in \mathbb{R}^{n\times n}$$

of the random vector $\boldsymbol{Z}$.

When the covariance parameters $\boldsymbol{\theta}$ are known, an unbiased, consistent and optimal estimate of $\boldsymbol{\beta}$, given a vector of observations $\boldsymbol{\zeta} \in \mathbb{R}^n$, is obtained by minimizing the (generalized) least squares functional

$$\|\boldsymbol{\zeta} - \boldsymbol{H\beta}\|^2_{\boldsymbol{C}_{\boldsymbol{\theta}}^{-1}} := (\boldsymbol{\zeta} - \boldsymbol{H\beta})^\top \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\zeta} - \boldsymbol{H\beta}),$$

resulting in the estimate

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{H}^\top \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^\top \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\zeta}. \tag{12}$$

If, by contrast, the covariance parameters $\boldsymbol{\theta}$ are not known, one approach is to estimate them from the data along with $\boldsymbol{\beta}$ by *maximum likelihood (ML)* estimation, where the joint probability density function (11) is maximized for the given observation vector $\boldsymbol{\xi} = \boldsymbol{\zeta}$ as a function of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. To solve this nonlinear optimization problem one usually *minimizes* the negative logarithm $\ell(\boldsymbol{\zeta}; \boldsymbol{\beta}, \boldsymbol{\theta}) := -\log p(\boldsymbol{\zeta}; \boldsymbol{\beta}, \boldsymbol{\theta})$ of the likelihood given by

$$\ell(\boldsymbol{\zeta}; \boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{2}\left[n\log(2\pi) + \log\det \boldsymbol{C_\theta} + (\boldsymbol{\zeta} - \boldsymbol{H\beta})^\top \boldsymbol{C}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\zeta} - \boldsymbol{H\beta})\right]. \tag{13}$$

As is argued, e.g., in Kitanidis (1987), when random field hydrogeological parameters are estimated based on data from a finite region where the separation distance of the measurements is of the same order as the correlation length, the use of fitted means may introduce a bias in the estimation of the covariance parameters, resulting typically in an underestimation of both the variance and correlation length parameters. This bias is the result of strong correlations in the observations, preventing the estimation error from entering the asymptotic regime as more observations are added, since the number of independent measurements does not increase due to these strong correlations.

A remedy known as *restricted maximum likelihood estimation (RML)* (cf. Harville 1977; Stein 1999, p. 170) is to apply a transformation to the data which filters out the mean. In the case of the linear model (8) for the mean, we consider the random vector $\boldsymbol{Z}'$ obtained by projecting $\boldsymbol{Z}$ orthogonally onto the orthogonal complement of the range of $\boldsymbol{H}$, hence removing any effect of the estimated regression coefficients $\boldsymbol{\beta}$ on the estimation of the covariance parameters. Indeed, if the columns of $\boldsymbol{Q} \in \mathbb{R}^{n\times(n-k)}$ form an orthonormal basis of range$(\boldsymbol{H})^{\perp}$, then $\boldsymbol{Q}^{\top}\boldsymbol{H} = \boldsymbol{O}$ and therefore the random vector

$$\boldsymbol{Z}' := \boldsymbol{Q}\boldsymbol{Q}^{\top}\boldsymbol{Z}$$

has expectation

$$\mathbf{E}\left[\boldsymbol{Z}'\right] = \mathbf{E}\left[\boldsymbol{Q}\boldsymbol{Q}^{\top}(\boldsymbol{H}\boldsymbol{\beta} + \widetilde{\boldsymbol{Z}})\right] = \mathbf{E}\left[\widetilde{\boldsymbol{Z}}\right] = \boldsymbol{0}$$

regardless of the value of $\boldsymbol{\beta}$. Here $\widetilde{\boldsymbol{Z}}$ denotes the random vector obtained by restricting the residual random field $\widetilde{Z}$ to the observation points. RML now maximizes the likelihood of the transformed random vector $\boldsymbol{Z}'$, which has an $(n-k)$-dimensional multivariate normal distribution with zero mean and covariance matrix $\boldsymbol{Q}^{\top}\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{Q} \in \mathbb{R}^{(n-k)\times(n-k)}$. The minimizing $\boldsymbol{\theta}$ can then be inserted into (12) to obtain $\boldsymbol{\beta}$.

2.4.3 Hyperparameter estimation and model selection

For all combinations of the regression functions (9), a *restricted maximum likelihood* (RML) estimation procedure detailed in Sect. 2.4.2 was used to determine the hyperparameters σ^2, ρ and ν of the Matérn covariance model (6) based on the 62 transmissivity observations published in U.S. Department of Energy (DOE) (2014). Based on this calibrated covariance structure, a model comparison was carried out following a procedure proposed in Kitanidis (1997b), in which a significance test is used to determine whether adding further regression functions to a model better explains the data. The test computes the sums of the decorrelated squared errors of both regression models at the observation locations and compares their normalized relative difference. If the the ratio exceeds a chosen quantile of a suitable F distribution, the smaller regression model is not considered sufficient, i.e., it is a classical variance ratio test.

In this way, we arrived at a trend model (8) consisting of the regression functions $\{h_1, h_2, h_5\}$ from (9). In the following we refer to this parametrization of the mean as the *best model* and to that containing only the constant trend function h_1 as the *constant model*. The resulting estimates of the hyperparameters σ, ρ and ν for both models are given in Table 1. Note that we have fixed $\nu = 0.5$ in both cases since the estimates for ν were sufficiently close to this value,[3] which also allows a more efficient evaluation of the associated covariance function. The regression model estimated by the (generalized) least squares method for the mean is then

$$\overline{Z}(\boldsymbol{x}) = 143.98 - 2.55\cdot 10^{-4}x_1 + 3.31\mathbb{1}_{D_1}(\boldsymbol{x}).$$

Note that the values for x_1 (UTM Easting coordinates) are of order $6\cdot 10^5$ for the WIPP computational domain D.

[3] If we do not fix $\nu = 0.5$ but estimate it as well the RML results are $\hat{\sigma}^2 = 6.14$, $\hat{\rho} = 2005.2$, and $\hat{\nu} = 0.48$.

Table 1 Restricted maximum likelihood estimation of hyperparameters σ^2 (variance or *sill*) and ρ (correlation length or *range*) for two trend models based on the 64 observations of transmissivity

Trend model	Sillσ^2	Range ρ	Smoothness ν
h_1	17.12	6509.8	0.5
h_1, h_2, h_5	6.15	1948.0	0.5

The smoothness parameter was fixed at $\nu = 1/2$, which corresponds to the exponential covariance kernel

2.4.4 Conditioning on transmissivity data

Once the mean and covariance functions of the Gaussian random field $Z = \log T$ have been determined, the log transmissivity measurements $\{z(\boldsymbol{x}_j)\}_{j=1}^N$ may be used to further calibrate the stochastic model to fit the observations in a statistical sense using the technique known as *kriging* (cf. Cressie 1991; Kitanidis 1997a; Stein 1999). Kriging refers to *best linear unbiased prediction* (BLUP) in which the value of the random field Z at an arbitrary location $\boldsymbol{x} \in D$ is estimated as an affine combination

$$\hat{Z} = \hat{Z}(\boldsymbol{x}, \omega) = \lambda_0(\boldsymbol{x}) + \boldsymbol{\lambda}(\boldsymbol{x})^\top \boldsymbol{Z}(\omega) \tag{14}$$

of the (random) realizations $\boldsymbol{Z} = (Z(\boldsymbol{x}_1), \dots, Z(\boldsymbol{x}_N))^\top$, with spatially varying coefficients $\lambda_0 : D \to \mathbb{R}$ and $\boldsymbol{\lambda} = (\lambda_1(\boldsymbol{x}), \dots, \lambda_N(\boldsymbol{x})) : D \to \mathbb{R}^N$ chosen to make the estimator *unbiased* and *mean square optimal*, which requires that, for all $\boldsymbol{x} \in D$, we have

$$\mathbf{E}\left[\hat{Z}(\boldsymbol{x})\right] = \mathbf{E}\left[Z(\boldsymbol{x})\right] \quad \text{and} \quad \mathbf{E}\left[|Z(\boldsymbol{x}) - \hat{Z}(\boldsymbol{x})|^2\right] \to \min_{\lambda_0, \boldsymbol{\lambda}}!.$$

For a known mean function $\overline{Z}$ the solution is given by the *(simple) kriging prediction* or *interpolation*

$$\hat{Z}(\boldsymbol{x}) = \hat{Z}(\boldsymbol{x}, \omega) = \overline{Z}(\boldsymbol{x}) + \boldsymbol{c}(\boldsymbol{x})^\top \boldsymbol{C}^{-1}\left(\boldsymbol{Z}(\omega) - \overline{\boldsymbol{Z}}\right),$$

where $\overline{\boldsymbol{Z}} := [\overline{Z}(\boldsymbol{x}_1), \dots, \overline{Z}(\boldsymbol{x}_N)]^\top$, $\boldsymbol{c}(\boldsymbol{x}) := (c(\boldsymbol{x}, \boldsymbol{x}_1), \dots, c(\boldsymbol{x}, \boldsymbol{x}_N))^\top$ and $\boldsymbol{C} := (c(\boldsymbol{x}_i, \boldsymbol{x}_j))_{i,j=1,\dots,N} \in \mathbb{R}^{N \times N}$, with mean square error given via the *kriging (error) covariance*

$$\mathbf{E}\left[\left|Z(\boldsymbol{x}) - \hat{Z}(\boldsymbol{x})\right|^2\right] = \hat{c}(\boldsymbol{x}, \boldsymbol{x}), \qquad \hat{c}(\boldsymbol{x}, \boldsymbol{y}) := c(\boldsymbol{x}, \boldsymbol{y}) - \boldsymbol{c}(\boldsymbol{x})^\top \boldsymbol{C}^{-1} \boldsymbol{c}(\boldsymbol{y}).$$

Note that for a Gaussian random field Z the kriging prediction $\hat{Z}$ is again Gaussian and coincides with the conditioned random field $Z(\boldsymbol{x})|\boldsymbol{Z} = \boldsymbol{z}$, where $\boldsymbol{z} = (z_1, \dots, z_N)^\top$ with $z_i = z(\boldsymbol{x}_i)$ for $i = 1, \dots, N$, so that $\hat{Z}(\boldsymbol{x}) \sim \mathsf{N}\left(\overline{Z}(\boldsymbol{x}) + \boldsymbol{c}(\boldsymbol{x})^\top \boldsymbol{C}^{-1}\left(\boldsymbol{z} - \overline{\boldsymbol{Z}}\right), \hat{c}(\boldsymbol{x}, \cdot)\right)$. It is easily verified that at the observation sites $\{\boldsymbol{x}_j\}_{j=1}^N$ we have $\hat{Z}(\boldsymbol{x}_j) = z(\boldsymbol{x}_j)$ and $\hat{c}(\boldsymbol{x}_j, \boldsymbol{x}_j) = 0$, hence the kriging estimate $\hat{Z}$ of the random field Z interpolates the measurements.

In the variant called *universal kriging*, the mean $\overline{Z}$ is not assumed known and instead modelled as in (8). Forming the least squares estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and proceeding as above with $\overline{Z}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x})^\top \hat{\boldsymbol{\beta}}$ would fail to account for uncertainty in this estimate. Instead, we require that unbiasedness of the kriging estimate (14) hold for all $\boldsymbol{\beta} \in \mathbb{R}^k$, resp. for all possible mean functions. Applying unbiasedness as a constraint in the pointwise minimization over $\lambda_0, \boldsymbol{\lambda}$ via Lagrange multipliers yields the *universal kriging prediction* or interpolation

$$\hat{Z}(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{c}(\boldsymbol{x}) \\ \boldsymbol{h}(\boldsymbol{x}) \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{C} & \boldsymbol{H} \\ \boldsymbol{H}^\top & 0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{Z} \\ 0 \end{bmatrix}, \tag{15}$$

where

$$\boldsymbol{H} = \begin{bmatrix} h_1(\boldsymbol{x}_1) & \dots & h_k(\boldsymbol{x}_1) \\ \vdots & & \vdots \\ h_1(\boldsymbol{x}_N) & \dots & h_k(\boldsymbol{x}_N) \end{bmatrix} \in \mathbb{R}^{N \times k},$$

or, equivalently,

$$\hat{Z}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x})^\top \hat{\boldsymbol{\beta}} + \boldsymbol{c}(\boldsymbol{x})^\top \boldsymbol{C}^{-1} \left(\boldsymbol{Z} - \boldsymbol{H}\hat{\boldsymbol{\beta}} \right), \tag{16}$$

where $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{H}^\top \boldsymbol{C}^{-1} \boldsymbol{H}\right)^{-1} \boldsymbol{H}^\top \boldsymbol{C}^{-1} \boldsymbol{Z}$, with mean square error $\mathbf{E}\left[|Z(\boldsymbol{x}) - \hat{Z}(\boldsymbol{x})|^2\right] = \hat{c}(\boldsymbol{x}, \boldsymbol{x})$ given in this case by the *universal kriging (error) covariance*

$$\hat{c}(\boldsymbol{x}, \boldsymbol{y}) := c(\boldsymbol{x}, \boldsymbol{y}) - \boldsymbol{c}(\boldsymbol{x})^\top \boldsymbol{C}^{-1} \boldsymbol{c}(\boldsymbol{y}) + \boldsymbol{\gamma}(\boldsymbol{x})^\top \boldsymbol{V} \boldsymbol{\gamma}(\boldsymbol{y}), \tag{17}$$

where $\boldsymbol{\gamma} = \boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{H}^\top \boldsymbol{C}^{-1} \boldsymbol{c}(\boldsymbol{x})$ and $\boldsymbol{V} = (\boldsymbol{H}^T \boldsymbol{C}^{-1} \boldsymbol{H})^{-1}$. Thus, the universal kriging prediction (16) consists in obtaining the mean as the least squares estimate $\boldsymbol{h}(\boldsymbol{x})^\top \hat{\boldsymbol{\beta}}$ and proceeding as in simple kriging. However, the universal kriging mean square error contains the additional term $\boldsymbol{\gamma}(\boldsymbol{x})^\top \boldsymbol{V} \boldsymbol{\gamma}(\boldsymbol{x}) \geq 0$ compared to that of simple kriging, which accounts for the additional uncertainty present in the estimated mean and $\boldsymbol{\beta}$. Note further that, even for Gaussian Z, the universal kriging mean and (co)variance do not, in general, possess an interpretation as those of a conditioned Gaussian random field as is the case with simple kriging.

We now use the *universal kriged Gaussian random field* $\hat{Z}$ obtained from the available log transmissivity measurements $\boldsymbol{z} = \{z(\boldsymbol{x}_j)\}_{j=1}^N$ as our final stochastic model for the uncertain transmissivity field, i.e.,

$$\hat{Z}(\boldsymbol{x}) \sim \mathsf{N}\left(\hat{z}(\boldsymbol{x}), \hat{c}(\boldsymbol{x}, \cdot)\right)$$

with $\hat{c}$ given in (17) and $\hat{z}$ resulting by inserting the realization $\boldsymbol{Z} = \boldsymbol{z}$ in (15). The resulting kriged mean $\hat{z}$ and pointwise variance $\hat{c}$ are displayed in Fig. 4.

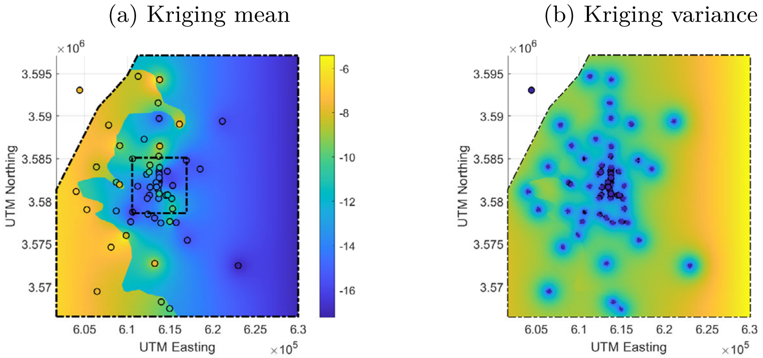


Fig. 4 Universal kriging prediction of $Z = \log T$ based on 62 available transmissivity observations. Left: kriged mean field $\hat{z}(\boldsymbol{x})$. Right: pointwise kriging variance $\hat{c}(\boldsymbol{x}, \boldsymbol{x})$. The circular markers indicate the locations (and values) of the observational log transmissivity data. The interpolation property of $\hat{z}(\boldsymbol{x})$ is apparent

2.5 Uncertainty propagation for the quantity of interest

For a random transmissivity field $T(\omega) = T(\cdot, \omega)$, $\omega \in \Omega$, we consider individual realizations of the associated random boundary value problem in its mixed formulation (4), i.e.,

$$\left(\frac{\phi b}{T(\omega)}\boldsymbol{u}(\omega), \boldsymbol{v}\right) - (p(\omega), \nabla\cdot \boldsymbol{v}) = -\langle g, \boldsymbol{n}\cdot\boldsymbol{v}\rangle_{\Gamma_D} \qquad \forall \boldsymbol{v} \in \mathcal{V}, \tag{18a}$$

$$(\nabla\cdot \boldsymbol{u}(\omega), q) = 0 \qquad \forall q \in \mathcal{W}, \tag{18b}$$

with random solution pair $(\boldsymbol{u}(\omega), p(\omega)) \in \mathcal{V} \times \mathcal{W}$. The Eq. (18) are now understood as holding $\mathbf{P}$-almost surely. Under suitable assumptions (cf. Babuška et al. 2007) we have $(\boldsymbol{u}, p) \in L^2_{\mathbf{P}}(\mathcal{V} \times \mathcal{W})$, i.e., the norm of the solution is square integrable against the probability measure $\mathbf{P}$.

For the quantity of interest under consideration, the exit time for particle trajectories, each realization of the random flux yields a realization of the associated random initial value problem

$$\dot{\boldsymbol{x}}(t, \omega) = \boldsymbol{u}(\boldsymbol{x}(t, \omega), \omega), \quad t \geq 0, \qquad \boldsymbol{x}(0, \omega) = \boldsymbol{x}_0. \tag{19}$$

$\mathbf{P}$-almost surely, and hence, the quantity of interest becomes a random variable

$$f_{\text{exit}}(\omega) := \log \min\{t > 0 : \boldsymbol{x}(t, \omega) \notin D_0,\ \boldsymbol{x}_0 \in D_0\}. \tag{20}$$

A complete characterization of the uncertainty in f_{exit} is given by its cumulative distribution function (CDF)

$$F(s) := \mathbf{P}(f_{\text{exit}} \leq s), \qquad F : \mathbb{R} \to [0, 1].$$

Due to the complexity of the problem, F cannot be given in analytic form and has to be approximated. We comment on the computational aspects in the next section.

3 Computational realization

In this section we describe (i) the spatial discretization used for solving the Darcy flow Eq. (4) or (18), respectively, given a realization of the transmissivity field T, (ii) a discrete representation of the random model for the transmissivity field T as well as (iii) a Monte Carlo approach for approximating the CDF of the quantity of interest.

3.1 Finite element solution of Darcy flow problem

We solve the Darcy flow Eq. (4a) – or individual realizations of their random form (18) – using a mixed finite element discretization consisting of the lowest order Raviart-Thomas space $\mathcal{V}_h \subset \mathcal{V}$ for the flux variable and piecewise constant space $\mathcal{W}_h \subset \mathcal{W}$ for the hydraulic head with respect to a triangulation $\mathcal{T}_h$ of the domain D, where $h > 0$ is a measure of mesh resolution. This discretization is known to be inf-sup-stable (cf. Boffi et al. 2013, Chapter 7; Ern and Guermond 2021, Chapter 51).

We choose a fixed triangulation of the two-dimensional computational domain with mesh width h chosen such that at least 10 elements correspond to the correlation length of the random transmissivity field, resulting in a mesh consisting of 28 993 triangles with the associated finite element spaces containing 72 705 degrees of freedom (43 712 for flux and 28 993 for hydraulic head). Note that a coarser mesh is depicted in Fig. 2 for illustration purposes. The particle tracking is performed by solving the ordinary differential Eq. (19) for the given realization. For the lowest-order Raviart-Thomas discretization, the constraint of zero divergence results in an elementwise constant flux, making this computation trivial and incurring no additional discretization error.

3.2 Conditioned Karhunen–Loève expansion

Various methods exist to generate realizations of random fields, among these the turning bands method, circulant embedding and Karhunen–Loève expansion, see (cf. Lord et al. 2014,). In this work, we generate approximate realizations of the Gaussian log transmissivity field by truncating its Karhunen–Loève expansion, an orthogonal expansion of a random field based on the spectral decomposition of its covariance operator

$$C : L^2(D) \to L^2(D), \qquad u \mapsto Cu, \qquad (Cu)(\boldsymbol{x}) = \int_D c(\boldsymbol{x}, \boldsymbol{y}) u(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}, \tag{21}$$

which for continuous covariance functions is compact and selfadjoint, positive definite and hence possesses a system of orthonormal eigenfunctions $(z_m)_{m=1}^{\infty}$ which are complete in $L^2(D)$. Denoting by $\lambda_m \geq 0$ the eigenvalue (ordered descending) associated with eigenfunction z_m, a second-order random field Z on D with mean $\overline{Z}$ possesses the expansion

$$Z(\boldsymbol{x}) = \overline{Z}(\boldsymbol{x}) + \sum_{m=1}^{\infty} \sqrt{\lambda_m}\, z_m(\boldsymbol{x})\, \xi_m, \qquad \boldsymbol{x} \in D, \tag{22}$$

converging in L^2, where $(\xi_m)_{m\in\mathbb{N}}$ is a sequence of pairwise uncorrelated random variables and $(\lambda_m)_{m\in\mathbb{N}}$ is square summable. In the present setting, the log transmissivity field Z is Gaussian, as stated in Sect. 2.3, therefore we have $\xi_m \sim \mathsf{N}(0, 1)$ for all m.

An approximation suited for computation is obtained by truncating the infinite expansion in (22) after a finite number M of terms, hence the accuracy of the resulting approximation

$$Z(\boldsymbol{x}) \approx \overline{Z}(\boldsymbol{x}) + \sum_{m=1}^{M} \sqrt{\lambda_m} z_m(\boldsymbol{x})\xi_m \tag{23}$$

for fixed M will depend on the decay rate of the eigenvalues.

Once a truncation index M has been fixed, the random field can be regarded as parameterized by the uncorrelated M-variate normal random vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_M)^\top \sim \mathsf{N}(\boldsymbol{0}, \boldsymbol{I})$, which takes values in $\mathbb{R}^M$. We may thus consider all random quantities in (18), i.e., the transmissivity field T and the solution $(\boldsymbol{u}, p)$ of the Darcy flow equations as well as the particle trajectories (19) and exit time f_{exit} in (20) as parameterized by realizations of this single random vector.

Explicit closed-form solutions to the eigenvalue problem (21) are known only for a small number of special cases, hence we approximate the eigenpairs numerically. We approximate the covariance operator C, where the covariance kernel is obtained from the universal kriging covariance $\hat{c}$ in (17), by Galerkin projection into a finite-dimensional subspace $\mathcal{W}_h$ of $L^2(D)$ consisting of piecewise constant functions with respect to a triangulation of the domain D, which we assume to be polygonal for simplicity.[4] Denoting by $\{\phi_1, \ldots, \phi_N\}$ a basis of $\mathcal{W}_h$, we represent functions in $\mathcal{W}_h$ as

$$u(\boldsymbol{x}) = \sum_{i=1}^{N} u_i \phi_i(\boldsymbol{x}) \tag{24}$$

with coefficient vector $\boldsymbol{u} = (u_1, \ldots, u_N)^\top$. Substituting (24) into (21), multiplying it by test functions ϕ_j and integrating over D we arrive at the discrete generalized eigenvalue problem

$$\boldsymbol{C}\boldsymbol{u} = \lambda \boldsymbol{M}\boldsymbol{u}, \tag{25}$$

where $\boldsymbol{C}$ is a symmetric positive semi-definite matrix with entries

$$[\boldsymbol{C}]_{i,j} = (C\phi_i, \phi_j)_{L^2(D)} = \int_D \phi_j(\boldsymbol{x}) \int_D c(\boldsymbol{x}, \boldsymbol{y})\phi_i(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\boldsymbol{x} \tag{26}$$

[4] We use the same finite element space as for the piecewise constant discretization of the hydraulic head p for convenience.

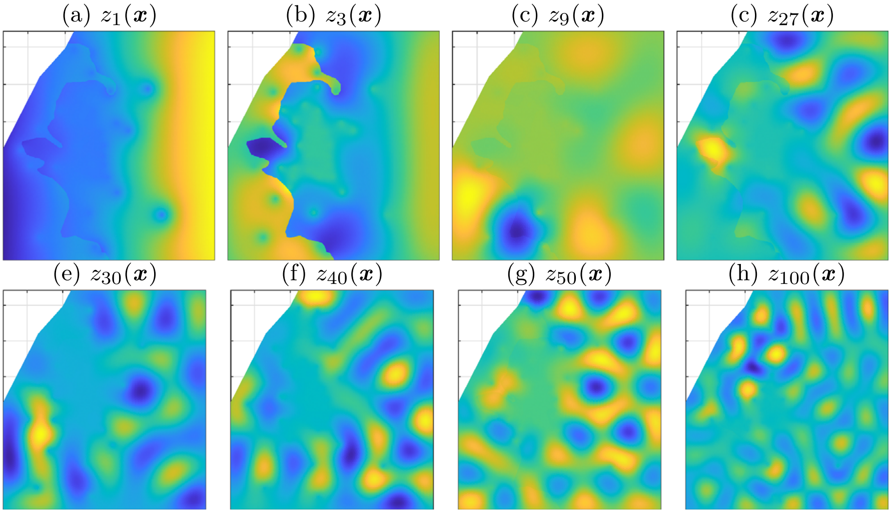


Fig. 5 Computed eigenfunctions of the kriging covariance function $\hat{c}$ in (17), cf. Fig. 4

and $\boldsymbol{M}$ is the symmetric positive definite Gram matrix of the piecewise constant basis with entries

$$[\boldsymbol{M}]_{i,j} = (\phi_j, \phi_i)_{L^2(D)} = \int_D \phi_j(\boldsymbol{x})\phi_i(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \tag{27}$$

An immediate difficulty with solving (25) is that $\boldsymbol{C}$ is a dense matrix due to the nonlocal nature of the integral operator C, hence generating and storing $\boldsymbol{C}$ is computationally expensive already for problems on two-dimensional domains, and even more so in three dimensions. Note that $\boldsymbol{M}$ is diagonal due to the disjoint supports of the ϕ_i. Moreover, even if generating and storing $\boldsymbol{C}$ were feasible, solving a dense eigenvalue problem by the standard symmetric QR algorithm results in excessive computation costs. We address this problem by first using an iterative method for approximating only the dominant M eigenvalues of $\boldsymbol{C}$ using a variant of the *thick-restart-Lanczos method* of Wu and Simon (2000), which requires only matrix vector products with $\boldsymbol{C}$ in the course of the iteration. Second, we represent $\boldsymbol{C}$ in *hierarchical matrix format* (cf. Hackbusch 2015), which brings the cost of generating, storing and multuplying $\boldsymbol{C}$ by a vector from $\mathcal{O}(N^2)$ to a complexity $\mathcal{O}(N \log N)$. Further details on using hierarchical matrices in the context of random field generation with the Galerkin method can be found in Eiermann et al. (2007) and Khoromskij et al. (2009).

Figure 5 shows a few computed eigenfunctions z_m for the kriging covariance function $\hat{c}$ in (17) displayed in Fig. 4.

3.3 Empirical estimation of the CDF

A common and straightforward way to approximate the CDF F of the random quantity of interest $f_{\text{exit}}(\boldsymbol{\xi}) := \log \min\{t > 0 : \boldsymbol{x}(t, \boldsymbol{\xi}) \notin D_0,\ \boldsymbol{x}_0 \in D_0\}$ is by generating n

samples $f_1, \ldots, f_n$ of the random f_{exit} by sampling n different realizations $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ of the random coefficient vector $\boldsymbol{\xi}$ in the KL expansion of $\log T$ and solving the corresponding n boundary and initial value problems to obtain $f_i = f_{\text{exit}}(\boldsymbol{\xi}_i)$. The empirical CDF (ECDF) of $f_{\text{exit}}(\boldsymbol{\xi})$ is then given by

$$F_n(s) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{(-\infty, f_j]}(s).$$

The ECDF F_n is a random approximation to the CDF F of the quantity of interest f_{exit} due to the randomly drawn samples $f_1, \ldots, f_n$. We denote the error between the (random) ECDF and the true CDF by

$$D_n := \sup_{s \in \mathbb{R}} |F(s) - F_n(s)|\,. \tag{28}$$

For i.i.d. samples a classical result known as Donsker's theorem (Athreya and Lahiri (2006), Corollary 11.4.13) states

$$\sqrt{n} D_n \xrightarrow[n\to\infty]{d} \sup_{t \in [0,1]} |B(t)|,$$

where B denotes a standard Brownian bridge on the unit interval $[0, 1]$. This theoretical result can be employed to compute the necessary minimal sample size n for a desired error criterion, which we fix here by requiring

$$\mathbf{P}\,(D_n > 0.01) \leq 0.05. \tag{29}$$

Using the asymptotic result provided by Donsker's theorem as well as $\mathbf{P}\left(\|B\|_{C[0,1]} > 1.36\right) \approx 0.05$, see Williams (2004, p. 343), we obtain for $n \approx 20\,000$ that $\mathbf{P}\,(D_n > 0.01) \approx 0.05$. Hence, in the present setting this means that, for this level of accuracy in approximating the CDF of the quantity of interest, we need to solve $n = 20\,000$ Darcy flow equations and compute the associated particle trajectories. Thus, the question arises whether we could save computational work by employing surrogates for the mapping from the random parameter vector $\boldsymbol{\xi}$ to the solution of the random PDE or the quantity of interest f_{exit} itself.

Estimation of CDF based on surrogates Assuming now that we have an approximation $\hat{f}_{\text{exit}} \colon \mathbb{R}^M \to \mathbb{R}$ to the quantity of interest f seen as mapping from $\boldsymbol{\xi} \in \mathbb{R}^M \to \mathbb{R}$, the resulting approximate ECDF $\hat{F}_n(s)$ based on n samples $\hat{f}_1, \ldots, \hat{f}_n$ of $\hat{f}_{\text{exit}}$ resulting from n samples $\boldsymbol{\xi}_i$ of the random KL parameter $\boldsymbol{\xi}$, where $\hat{f}_i = \hat{f}_{\text{exit}}(\boldsymbol{\xi}_i)$ is given by

$$\hat{F}_n(s) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{(-\infty, \hat{f}_j]}(s).$$

The question we investigate in this work is whether, for common surrogate constructions such as stochastic collocation and Gaussian process emulators, the approximation error $\|f_{\text{exit}} - \hat{f}_{\text{exit}}\|$ (measured in a suitable norm) can be made smaller than the sampling error D_n in the empirical estimation of the CDF. To this end, we evaluate the quality of the surrogate $\hat{f}_{\text{exit}}$ by a two-sample *Kolmogorow-Smirnov (KS)* test which is a well-known hypothesis test for checking whether sets of two samples—in our case $\hat{f}_1, \dots, \hat{f}_n$ and $f_1, \dots, f_n$—are likely to have been drawn from the same distribution. Specifically, in our case the KS test is passed at significance level $\alpha = 0.05$ if the KS-statistic K satisfies

$$K := \sup_{s \in \mathbb{R}} \left| \hat{F}_n(s) - F_n(s) \right| \leq 1.36 \frac{\sqrt{2}}{n},$$

cf. Williams (2004).

4 Propagation surrogates

In the following, we recall *sparse grid polynomial collocation* and *Gaussian process emulators (GPE)* as surrogate techniques for approximating a function $f : \boldsymbol{\Xi} \to \mathcal{Y}$ of M (random or parametric) variables $\boldsymbol{\xi} \in \mathbb{R}^M$ taking values either in $\mathcal{Y} = \mathbb{R}$, as for scalar quantities of interest such as the exit time, or a function space, e.g., $\mathcal{Y} = \mathcal{V} \times \mathcal{W}$, as for the solution of the mixed formulation (18) of the Darcy flow equations with random conductivity.

We begin by illustrating the basic principles of polynomial collocation and Gaussian process emulation for approximating a function of a single variable, i.e., $\boldsymbol{\Xi} \subseteq \mathbb{R}$, before proceeding to the technical details for the multivariate case $\boldsymbol{\Xi} \subseteq \mathbb{R}^M$, where we assume $\boldsymbol{\Xi}$ to be of product form $\boldsymbol{\Xi} = \Xi^M$ with $\Xi \subseteq \mathbb{R}$.

4.1 Univariate collocation and emulation

As a simple example in the style of the GPE tutorial O'Hagan (2006), consider the function

$$y = f(\xi) := \xi + 3 \sin \frac{3\xi}{4}, \qquad \xi \in \Xi := [0, 6].$$

The presence of *input uncertainty*, i.e., uncertainty with regard to the precise value of the independent variable ξ, is accounted for by modeling it as a random variable $\xi \sim \mathrm{U}[0, 6]$. Suppose further that f is only accessible in the form of a finite number of point evaluations $f(\xi)$, as is the case for the exit time in our WIPP case study, where each evaluation of the former requires solving the Darcy flow problem followed by particle tracking up to the exit boundary. The task is to construct a computationally inexpensive approximation $\hat{f} : \Xi \to \mathbb{R}$ of f given n evaluations

$$y_j = f(\xi_j), \qquad j = 1, \dots, n.$$

The points of evaluation ξ_j are often called *design points* in the emulator literature and *nodes* or *knots* in the context of collocation. Their choice depends on the type of surrogate being constructed. We begin with an elementary numerical analysis procedure and then contrast this with an approach rooted in the statistics community.

Polynomial Collocation In the univariate case polynomial collocation simplifies to Lagrange interpolation by global polynomials, and the surrogate $\hat{f}$ for f takes the familiar form

$$\hat{f}_n(\xi) := \sum_{j=1}^{n} f(\xi_j)\ell_j(\xi), \qquad \ell_j(\xi) = \prod_{k \neq j} \frac{\xi - \xi_k}{\xi_j - \xi_k}$$

with $\{\ell_j\}_{j=1}^n$ the Lagrange fundamental polynomials associated with the nodes $\{\xi_1, \ldots, \xi_n\}$. Although this expression is well-defined for any set of distinct nodes, good approximation quality is only achieved if the points are chosen with care. A classical choice for bounded intervals is the family of *Clenshaw–Curtis nodes* (also called *Chebyshev nodes*). Scaled to the interval [0, 6], the set of n Clenshaw–Curtis nodes is given by

$$\xi_j = 3 + 3\cos\left(\frac{j-1}{n-1}\pi\right), \qquad j = 1, \ldots, n.$$

Other common choices, particularly for UQ applications, are the roots of the n-th orthogonal polynomial associated with the probability density of ξ on Ξ, e.g., Gauss–Legendre nodes for the uniform distribution or Gauss–Hermite nodes for the normal distribution, cf. Babuška et al. (2010). For optimal convergence of the interpolants for smooth functions f it is well known that the spatial distribution of the nodes $\xi_j \in \Xi$ should follow the equilibrium distribution in the sense of logarithmic potential theory, which for the standard interval $\Xi = [-1, 1]$ is given by $\mathrm{d}\mu(\xi) = 1/\pi\sqrt{1-\xi^2}$, cf. Trefethen (2013, Chapter 12). In particular, the nodes should cluster near the interval endpoints. Figure 6 shows two polynomial interpolation surrogates for f as well as the CDF of the output $f(\xi)$.

The approximation quality of polynomial interpolation depends not only on the choice of interpolation nodes, but also on the *smoothness* of f. For example, we have for $f \in C^r(\Xi)$, $r \in \mathbb{N}$, that

$$\|f - \hat{f}_n\|_\infty \leq c_r(f)\, n^{-r}\, \left(1 + \Lambda_{\xi_1,\ldots,\xi_n}\right)$$

where $\|f - \hat{f}_n\|_\infty = \sup_{\xi\in\Xi}|f(\xi) - \hat{f}_n(\xi)|$, $c_r(f)$ is a constant depending only on r and f, and $\Lambda_{\xi_1,\ldots,\xi_n}$ denotes the *Lebesgue constant* of the nodes $\xi_1, \ldots, \xi_n$. Thus, we should choose nodes which have a small Lebesgue constant, and one which grows only slowly with n. This is the case for Chebyshev and Clenshaw–Curtis nodes, for which

$$\Lambda_{\xi_1,\ldots,\xi_n} \in \mathcal{O}(\log n).$$

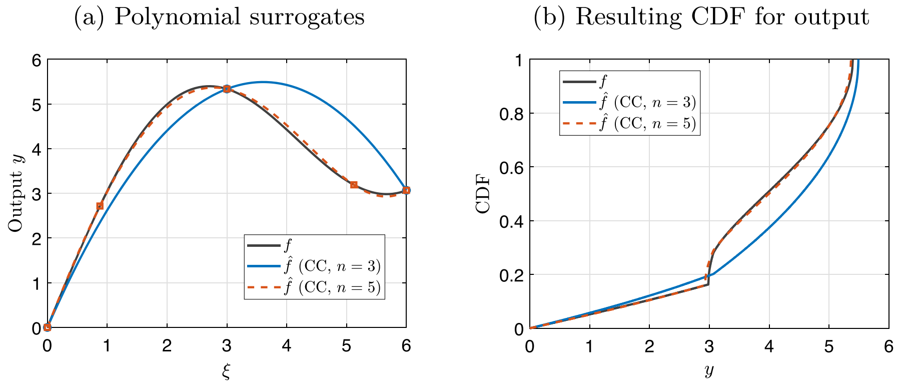


Fig. 6 The function $f(\xi) = \xi + 3\sin(3\xi/4)$ on $\Xi = [0, 6]$ and its Lagrange interpolation $\hat{f}_n$ based on $n = 3$ and $n = 5$ Clenshaw–Curtis nodes (left) and the resulting CDF for the output $y = f(\xi)$ and $\hat{y} = \hat{f}_n(\xi)$, resp., if $\xi \sim \mathsf{U}(\Xi)$

Beside uniform convergence there are also classical results on convergence in the L^p sense Nevai (1976; 1980; 1984), e.g., for Gauss–Legendre and Gauss–Hermite nodes

$$\lim_{n\to\infty} \|f - \hat{f}_n\|_{L^p_\mu} = 0, \qquad \|f - \hat{f}_n\|_{L^p_\mu} = \left(\int_\Xi |f(\xi) - \hat{f}_n(\xi)|^p \, \mu(\mathrm{d}x)\right)^{1/p},$$

where $\mu = \mathsf{U}(\Xi)$ or $\mu = \mathsf{N}(0, 1)$, respectively. However, if f has low regularity or is discontinuous, then convergence may fail or it may take a very large number of nodes to approximate f with sufficient accuracy.

In summary, polynomial collocation constructs a (deterministic) interpolating polynomial as a surrogate for f based on evaluations of f at n judiciously chosen nodes, for which the error decays with n at a rate depending on the smoothness of f.

Gaussian Process Emulation

The GPE approach consists in applying a method originating in geostatistics, namely the conditioning of Gaussian processes on observations (kriging), to the input–output map of a computer code. The latter is again represented by a scalar-valued function $f: \Xi \to \mathbb{R}$ for now. Again, we assume f is only accessible via selected point evaluations, i.e., a closed-form expression for f is not known. Thus, as for the transmissivity of subsurface layers known only at measurement sites, the function f is unknown but for selected evaluations $f(\xi)$. This initial uncertainty regarding f in the absence of point evaluations is modelled by a Gaussian process, i.e., a random function which follows a Gaussian distribution. Then, given finitely many evaluations $f(\xi_j)$ at design points $\xi_j \in \Xi$, we update our knowledge about f by conditioning the Gaussian process model on the observed data—analogous to the conditioning of the Gaussian log transmissivity on measurements in Sect. 2.4.4. The resulting conditioned mean function or *kriging prediction* is then employed as a (deterministic) surrogate $\hat{f}$ for f. As an additional feature, the GPE also provides a probabilistic quantification of the uncertainty in f which remains after conditioning, i.e., the deviation $\hat{f}(\xi) - f(\xi)$ of the conditioned Gaussian process mean at points $\xi \neq \xi_j$. This is called *code* or *output uncertainty* in the GPE literature, and is distinct from the *input uncertainty* modelled by *random* ξ: we have

$$\begin{array}{ll}\text{input uncertainty:} & \xi \text{ random and } \xi \mapsto f(\xi) \text{ fixed}\\ \text{output uncertainty:} & \xi \text{ fixed} \quad \text{and } \xi \mapsto f(\xi) \text{ random}\end{array}$$

Of course, both uncertainty types can be superposed, as we shall see later. Thus, an *emulator* provides in fact a random surrogate or statistical approximation of a function f which in this context is referred to as the *simulator* (cf. O'Hagan 2006). Before we provide a more detailed discussion of this form of *output uncertainty quantification*, we briefly describe how a GPE surrogate is constructed.

Analogously to Sect. 2.3 we first choose a Gaussian process model $G \sim \mathsf{N}(m, c)$ on Ξ with a (parametrized) mean function $m\colon \Xi \to \mathbb{R}$, e.g.,

$$m(\xi) = m(\xi; \boldsymbol{\beta}) = \sum_{k=1}^{p} \beta_k h_k(\xi), \qquad \boldsymbol{\beta} \in \mathbb{R}^p,$$

and a (parametrized) covariance function $c\colon \Xi \times \Xi \to \mathbb{R}$, e.g., a Matérn covariance (6) or squared exponential covariance

$$c(\xi, \xi') = c(\xi, \xi'; \sigma^2, \rho) = \sigma^2 \exp(-(\xi - \xi')^2/\rho), \qquad \xi, \xi' \in \Xi. \tag{30}$$

In a fully Bayesian approach, prior probability distributions are placed on the hyperparameters $\boldsymbol{\beta}, \sigma^2, \rho$ of m and c. For now, however, we assume the covariance c to be fixed and m to be given as linear regression model—in analogy to Sect. 2.3. Conceptually, the Gaussian process describes our "prior beliefs" about the unknown f in the form of, e.g., characteristic dependencies reflected in the regression functions h_k in the mean model or smoothness properties encoded in the choice of c. Given evaluations $f(\xi_j)$ of f at n design points ξ_j, we condition the Gaussian process G on this data and obtain $\hat{G}_n \sim \mathsf{N}(\hat{m}_n, \hat{c}_n)$ with $\hat{m}_n$ and $\hat{c}_n$ determined by the relations for (simple or universal) kriging, see Sect. 2.4.4. The resulting *surrogate* $\hat{f}_n$ is the conditional mean (or kriging prediction) of $\hat{G}_n$

$$\hat{f}_n(\xi) = \hat{m}_n(\xi) = \sum_{k=1}^{p} \hat{\beta}_k h_k(\xi) + \sum_{j=1}^{n} \hat{\gamma}_j c(\xi, \xi_j)$$

where the coefficients $\hat{\beta}_k$ and $\hat{\gamma}_k$ depend on ξ_j and linearly on the $f(\xi_j)$ and are computed via universal kriging, cf. (16). We illustrate the GPE mean/surrogate for f as above and the resulting CDF for the output $\hat{f}_n(\xi)$ if $\xi \sim \mathsf{U}[0, 6]$ in Fig. 7. Here we have used, similar to O'Hagan (2006),

$$m(\xi; \boldsymbol{\beta}) = \beta_1 + \beta_2 \xi, \qquad c(\xi, \xi') = \exp\left(-\frac{1}{4}(\xi - \xi')^2\right).$$

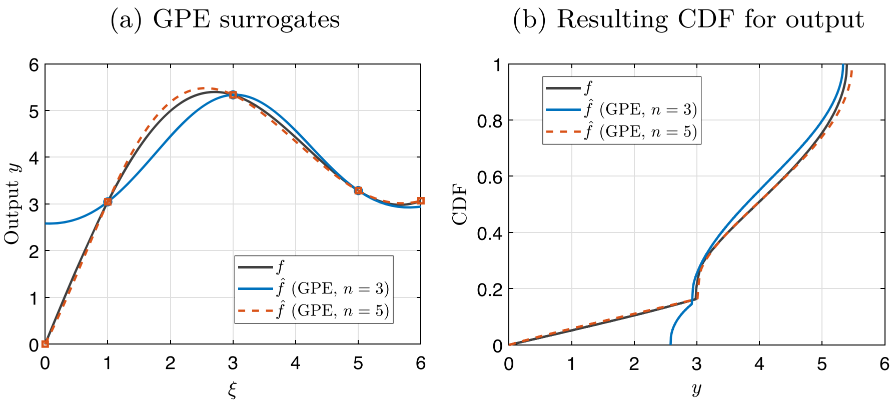


Fig. 7 The function $f(\xi) = \xi + 3\sin(3\xi/4)$ on $\Xi = [0, 6]$ and its GPE surrogates based on $n = 3$ and $n = 5$ design points $\xi_j \in \{1, 3, 5\}$ and $\xi_j \in \{0, 1, 3, 5, 6\}$ (left) and the resulting CDF for the output $y = f(\xi)$ and $\hat{y} = \hat{f}_n(\xi)$, resp., if $\xi \sim \mathsf{U}(\Xi)$

The choice of design points ξ_j for GPE follows different considerations than for polynomial interpolation. It is well known that kriging coincides with *kernel interpolation*, see Scheuerer et al. (2013). If we assume for simplicity that $m \equiv 0$ and c is given, then we can straightforwardly apply established approximation results from kernel interpolation theory by Narcowich et al. (2006, Proposition 3.2); Wendland (2004, Theorem 11.14), i.e., for $f \in H^r(\Xi)$ with $r \geq 1$ and suitable[5] covariance functions c such as Matérn kernels (6)

$$\|f - \hat{f}_n\|_\infty \leq C_r(f)\, \mathrm{D}_{\xi_1,\ldots,\xi_n}(\Xi)^{r-\frac{1}{2}}$$

where

$$\mathrm{D}_{\xi_1,\ldots,\xi_n}(\Xi) := \max_{\xi\in\Xi}\ \min_{j=1,\ldots,n} |\xi - \xi_j|$$

denotes the *fill distance* of the node set $\{\xi_1, \ldots, \xi_n\}$. For the Gaussian covariance function (30) we even obtain exponential convergence if the function f is *analytic*, see Wendland (2004),

$$\|f - \hat{f}_n\|_\infty \leq C(f)\, r^{\mathrm{D}_{\xi_1,\ldots,\xi_n}(\Xi)}, \qquad r < 1.$$

Thus, for good approximation properties, GPE requires a *space filling* strategy for choosing design points, i.e., one which minimizes fill distance. In the univariate case this is achieved by equispaced points, in stark contrast to the optimal equilibrium distribution for interpolation nodes.

As mentioned, a GPE not only provides a surrogate $\hat{f}_n$ but also a probabilistic quantification of the remaining pointwise error $f - \hat{f}_n$, which represents another

[5] "Suitable" means here, that the *native* or *reproducing kernel Hilbert space* of c coincides with $H^r(\Xi)$. For more details we refer to Scheuerer et al. (2013), Wendland (2004).

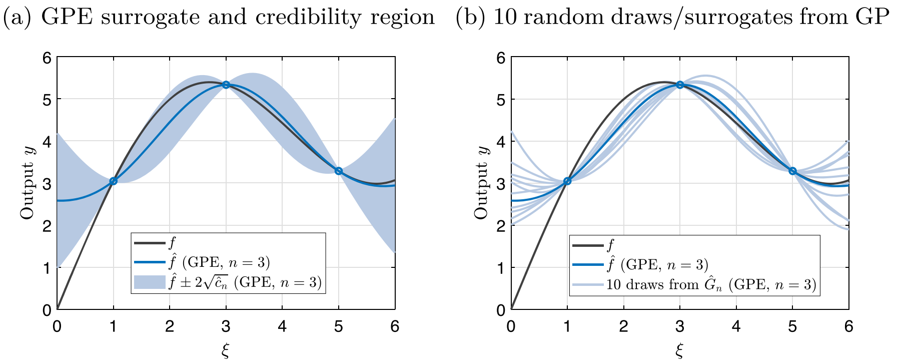


Fig. 8 The function $f(\xi) = \xi + 3\sin(3\xi/4)$ on $\Xi = [0, 6]$, its GPE surrogate and the related 95% credibility region for f (left) as well as 10 paths (or surrogates) drawn from the conditioned GP $\hat{G}_n$

important difference to (polynomial) collocation. In order to better understand this probabilistic error, recall that the conditioned Gaussian process $\hat{G}_n$ can be seen as our “posterior belief” about the unknown f given n evaluations $f(\xi_j)$. Thus, as for the transmissivity field in subsurface flow (which is deterministic but unknown) we model our *uncertainty about the true output* $f(\xi)$ at a *fixed input* $\xi \in \Xi$ by $\hat{G}_n(\xi) \sim \mathsf{N}(\hat{f}_n(\xi), \hat{c}_n(\xi))$. We illustrate the output uncertainty provided by the GPE in Fig. 8: the left panel shows f, $\hat{f}_n$ as well as pointwise error estimates for $f - \hat{f}_n$ given by two times the standard deviation of $\hat{G}_n(\xi)$, which can be also understood as the pointwise 95% credibility region for the unknown $f(\xi)$; the right panel shows 10 realizations of the Gaussian process $\hat{G}_n$. Each of these could equally well be used as a surrogate $\hat{f}_n$ in place of $\hat{m}_n$, since they are also valid (random) guesses for f. In this way, $\hat{G}_n$ provides a *random* surrogate for f.

Random draws from $\hat{G}_n$ can then be used to quantify the effect of the output uncertainty about the value $f(\xi) \neq \hat{f}_n(\xi)$ within an uncertainty analysis for varying ξ, e.g., for estimating the CDF of $f(\xi)$ when $\xi \sim \mathsf{U}(\Xi)$, see, e.g. Oakley and O’Hagan (2002). To explain this in more detail: Regarding the input uncertainty modelled by $\xi \sim \mathsf{U}(\Xi)$ we would like to quantify its effect on the outcome by the CDF

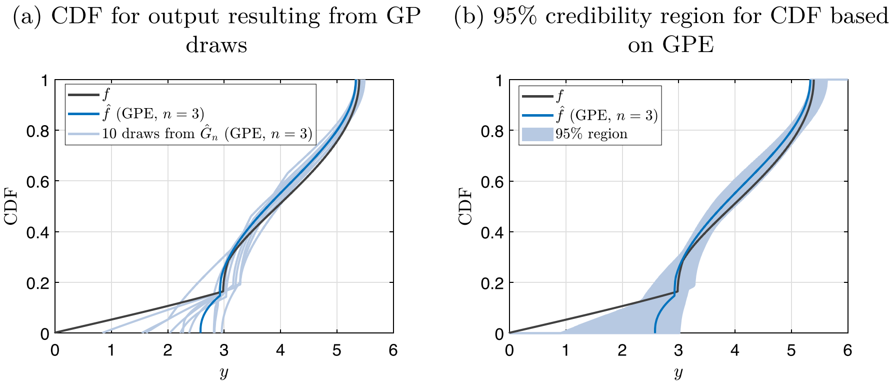
$$F(y) = \mathbf{P}(f(\xi) < y).$$

This is a deterministic function for uncertainty analysis for random ξ. However, if we are not able to use f itself to compute F but rather use a GPE $\hat{G}_n$ for f, we can, besides a deterministic approximation of F based on a deterministic surrogate $\hat{f}_n$ for f

$$F(y) \approx \mathbf{P}_\xi(\hat{f}_n(\xi) < y),$$

also incorporate our remaining output uncertainty about f via the conditioned Gaussian process $\hat{G}_n$ for f. This then yields a *random CDF*

$$\hat{F}_n(y) = \mathbf{P}_\xi(\hat{G}_n(\xi) < y),$$


(a) CDF for output resulting from GP draws

(b) 95% credibility region for CDF based on GPE

Fig. 9 Resulting CDFs for the output $\hat{y} = \hat{f}(\xi)$, $\xi \sim \mathsf{U}(\Xi)$, based on the mean and 10 random draws from the GPE $\hat{G}_n$ (left), and the resulting 95% credibility region for the CDF of $y = f(\xi)$ derived from the GPE (right)

due to the random $\hat{G}_n$ where we emphasize that the CDF is *only* w.r.t. randomness of the ξ. To illustrate this we show in Fig. 9 the resulting CDFs for $\hat{f}_n(\xi)$, $\xi \sim \mathsf{U}(\Xi)$ using $\hat{f}_n = \hat{m}_n$ as well as $\hat{f}_n$ set to be each of the 10 draws from $\hat{G}_n$ (left) as well as the 95% credibility region for the true (but unknown) CDF values $F(y) = \mathbf{P}(f(\xi) < y)$ based on 10,000 draws from $\hat{G}_n$. The credibility region thus quantifies our uncertainty about the true CDF resulting from using a (random) surrogate instead of the true quantity of interest f.

Discussion Polynomial collocation and Gaussian process emulators are well-established surrogate techniques based on point evaluations of the underlying quantity of interest f, and both approaches rely on a certain smoothness of f. However, they also differ in several aspects. These include the type of basis functions from which each surrogate is constructed (polynomials vs. kernel functions or radial basis functions) as well as the selection strategies for nodes ξ_j (potential-theoretic equilibrium distribution vs. space filling). Moreover, the GPE surrogate $\hat{f}_n = \hat{m}_n$ is based on minimizing the *average error* w.r.t. an assumed probability distribution over a function space, whereas interpolation error bounds are obtained from a *worst-case error* analysis over a function class. We refer to Ritter (2000) for more details on these two contrasting approaches. In particular, for GPE we explicitly assume a probability distribution for the unknown function f, given by the prior Gaussian process model G, whereas for collocation we simply assume that f is sufficiently smooth. This prior probability distribution for f is then updated given the data $f(\xi_j)$ in a Bayesian fashion. Thus, GPE can be related to *Bayesian numerical analysis*, see Diaconis (1988), or *probabilistic numerics*, see Hennig et al. (2022), respectively, and be seen as a Bayesian approach to kernel interpolation. In particular, the conditioned (posterior) distribution for the unknown f provided by $\hat{G}_n$ yields an indicator for the remaining (output) uncertainty about f after its evaluation at n nodes ξ_j. Of course, the assumption of Gaussianity for this computer output uncertainty is debatable. We refer to Bastos and O'Hagan (2009) for diagnostics to validate the GP ansatz as well as to Kracker et al. (2010) for a performance study of GPE for "Gaussian" as well as "non-Gaussian" f.

4.2 Polynomial sparse grid collocation

Polynomial collocation in the context of UQ or parametric problems can roughly be described as computing an M-variate polynomial approximation to $f\colon \boldsymbol{\Xi} \to \mathcal{Y}$, $\boldsymbol{\Xi} \subseteq \mathbb{R}^d$, based on multivariate Lagrange interpolation. Sparse grid collocation uses sparse grids as multivariate interpolation node sets in order to mitigate the curse of dimensionality associated with straightforward tensor-product interpolation for high-dimensional parameter spaces.

While more sophisticated sparse grid techniques have been developed in recent years, in this work we consider a basic and simple construction known as *(Smolyak) sparse grid collocation* introduced for UQ settings, e.g., in Xiu and Hesthaven (2005); Nobile et al. (2008). To this end, assume $f \in C(\boldsymbol{\Xi}; \mathcal{Y})$, i.e., the mapping f is continuous, and denote by

$$\mathscr{P}_n(\Xi; \mathcal{Y}) = \left\{\sum_{k=0}^n a_k \xi^k : a_k \in \mathcal{Y}\right\}$$

the space of all $\mathcal{Y}$-valued univariate polynomials of degree at most n. Then for a given sequence of univariate node sets $\Xi_k := \{\xi_1^{(k)}, \dots, \xi_{n_k}^{(k)}\} \subseteq \Xi, k \geq 1$, where we assume $n_1 = 1$ and $n_k < n_{k+1}$ throughout, we denote the associated univariate (Lagrange) interpolation operators by

$$\mathcal{I}_k\colon C(\Xi; \mathcal{Y}) \to \mathscr{P}_{n_k}(\Xi; \mathcal{Y}), \qquad (\mathcal{I}_k f)(\xi) := \sum_{j=1}^{n_k} f\left(\xi_j^{(k)}\right) \ell_j^{(k)}(\xi), \quad \xi \in \Xi,$$

with $\ell_j^{(k)} \in \mathscr{P}_{n_k}(\Xi; \mathbb{R})$ the Lagrange fundamental polynomials associated with Ξ_k. The most immediate extension of the interpolation operator to the M-dimensional parameter domain $\boldsymbol{\Xi}$ would be the multivariate interpolation operator $\mathcal{I}_{\boldsymbol{k}}\colon C(\boldsymbol{\Xi}; \mathcal{Y}) \to \mathscr{P}_{\boldsymbol{n_k}}(\boldsymbol{\Xi}; \mathcal{Y})$ obtained by tensorization

$$(\mathcal{I}_{\boldsymbol{k}} f)(\boldsymbol{\xi}) := \left(\mathcal{I}_{k_1} \otimes \cdots \otimes \mathcal{I}_{k_M}\right) f(\boldsymbol{\xi}) = \sum_{\boldsymbol{j} \leq \boldsymbol{n_k}} f\left(\boldsymbol{\xi}_{\boldsymbol{j}}^{(\boldsymbol{k})}\right) \ell_{\boldsymbol{j}}^{(\boldsymbol{k})}(\boldsymbol{\xi}),$$

with multi-indices $\boldsymbol{j} = (j_1, \dots, j_M), \boldsymbol{n_k} = (n_{k_1}, \dots, n_{k_M}) \in \mathbb{N}^M$, multivariate nodes $\boldsymbol{\xi}_{\boldsymbol{j}}^{(\boldsymbol{k})} = (\xi_{j_1}^{(k_1)}, \dots, \xi_{j_M}^{(k_M)}) \in \boldsymbol{\Xi_k} := \Xi_{k_1} \times \cdots \times \Xi_{k_M}$, and tensorized Lagrange fundamental polynomials $\ell_{\boldsymbol{j}}^{(\boldsymbol{k})}(\boldsymbol{\xi}) = \ell_{j_1}^{(k_1)}(\xi_1) \cdots \ell_{j_M}^{(k_M)}(\xi_M)$ for $\boldsymbol{\xi} = (\xi_1, \dots, \xi_M) \in \boldsymbol{\Xi}$. However, this construction suffers heavily from the curse of dimensionality since the computational work for evaluating f at all points in the Cartesian product grid $\boldsymbol{\Xi_k}$ grows exponentially with dimension M.

Sparse grid constructions, which improve this to polynomial complexity in M, are based on the univariate *detail operators*

$$\Delta_i = \mathcal{I}_i - \mathcal{I}_{i-1}, \quad i \geq 1, \qquad \mathcal{I}_0 \equiv 0,$$

so that $\mathcal{I}_k = \sum_{i=1}^{k} \Delta_i$, yielding the tensor product interpolation operator as

$$\mathcal{I}_{\boldsymbol{k}} f = \sum_{\boldsymbol{i} \leq \boldsymbol{k}} \Delta_{\boldsymbol{i}} f, \qquad \Delta_{\boldsymbol{i}} = \Delta_{i_1} \otimes \cdots \otimes \Delta_{i_M}.$$

By contrast, the *(Smolyak) sparse grid collocation operator* is defined by

$$\mathcal{S}_{\ell,M} f := \sum_{|\boldsymbol{i}-\mathbf{1}|_1 \leq \ell} \Delta_{\boldsymbol{i}} f, \qquad |\boldsymbol{i} - \mathbf{1}|_1 := \sum_{j=1}^{M} |i_j - 1|, \qquad \ell \geq 0.$$

By combinatorical arguments, one can obtain the equivalent *combination technique* representation

$$\mathcal{S}_{\ell,M} f = \sum_{\ell-M+1 \leq |\boldsymbol{i}-\mathbf{1}| \leq \ell} (-1)^{\ell+M-|\boldsymbol{i}|} \binom{M-1}{\ell+M-|\boldsymbol{i}|} \mathcal{I}_{\boldsymbol{i}} f,$$

which expresses the Smolyak operator as a linear combination of selected M-variate tensor product interpolation operators. For the associated *sparse grid*

$$\boldsymbol{\Xi}_{\ell,M} := \bigcup_{\ell-M+1 \leq |\boldsymbol{i}-\mathbf{1}| \leq \ell} \boldsymbol{\Xi}_{\boldsymbol{i}}$$

consisting of all multivariate nodes occurring in these representations, the cardinality $|\boldsymbol{\Xi}_{\ell,M}|$ grows only polynomially w.r.t. M (cf. Novak and Ritter 1999), while the overall order of accuracy remains close to that of the full tensor product $\mathcal{I}_{(\ell+1,\ldots,\ell+1)}$. In particular, it can be shown Baeck et al. (2011, Proposition 1) that $\mathcal{S}_{\ell,M}$ is a projection on

$$\mathscr{P}_{\ell,M}(\boldsymbol{\Xi}; \mathcal{Y}) := \sum_{|\boldsymbol{i}-\mathbf{1}| \leq \ell} \mathscr{P}_{n_{i_1}}(\Xi; \mathcal{Y}) \otimes \cdots \otimes \mathscr{P}_{n_{i_M}}(\Xi; \mathcal{Y}).$$

Note, however, that in general $\mathcal{S}_{\ell,M}$ is *not* interpolatory unless the univariate nodes sets are *nested* $\Xi_k \subset \Xi_{k+1}$ Barthelmann et al. (2000, Proposition 6). The latter is the case for Clenshaw–Curtis nodes with the "doubling sequence" $n_k = 2^k - 1$ $(k \geq 1)$, or (weighted) Leja nodes with linear growth $n_k = k$ (Ernst et al. 2021). In the following, we shall use the non-nested nodal sequence of *Gauss-Hermite* nodes, i.e., the roots of Hermite polynomials. This choice is common for collocation applied to functions of Gaussian random variables, see Babuška et al. (2007), Nobile et al. (2008), Ernst and Sprungk (2014).

Convergence and Application If f is sufficiently smooth then $\mathcal{S}_{\ell,M} f$ can be shown to converge to f, specifically

$$\|f - \mathcal{S}_{\ell,M} f\|_{L^2_\mu} \in \mathcal{O}\left(|\boldsymbol{\Xi}_{\ell,M}|^{-r}\right),$$

for an $r < 1$ using Gauss-Hermite nodes $\xi_i^{(k)}$ with linear growth $n_k = k$ or doubling growth $n_k = 2^{k-1} + 1$ ($k \geq 1$), see, e.g., Ernst and Sprungk (2014); Ernst et al. (2018). The rate of convergence r w.r.t. the number of collocation points depends, of course, on the smoothness class of f. In particular, it is well-known that sparse grid techniques such as Smolyak's construction above require a dominating mixed smoothness of f to work well, see, e.g., references Novak and Ritter (1999), Barthelmann et al. (2000), Sickel and Ullrich (2007), Ernst et al. (2018) for more details.

It was shown in Ernst and Sprungk (2014, Section 3) that the solution $(\boldsymbol{u}, p)$ of the random/parametric mixed variational problem (4) allows for a holomorphic extension into $\mathbb{C}^M$ under suitable assumptions, which are satisfied by truncated KL expansions (23) of a lognormal transmissivity field. Thus, applying $\mathcal{S}_{\ell,M}$ to approximate the solution map $(\boldsymbol{u}, p): \Xi \to \mathcal{V} \times \mathcal{W}$ is justified. By contrast, the quantity of interest given by the exit time f_{exit} may, in general, not even be a continuous function of the parameters $\boldsymbol{\xi}$, as is immediate from considering the case of a particle grazing the exit boundary and returning into the domain for a particular parameter setting. Thus, applying $\mathcal{S}_{\ell,M}$ to approximate f_{exit} directly may lead to inaccurate surrogate approximation or even divergence with increasing $|\Xi_{\ell,M}|$.

However, a simple remedy is to use the surrogate

$$\hat{f}_{\text{exit},\ell} = G_{\text{exit}}\left(\mathcal{S}_{\ell,M}\boldsymbol{u}\right)$$

where $G_{\text{exit}}: \mathcal{V} \to \mathbb{R}$ denotes the mapping from a velocity field on D to the log breakthrough time of a particle following this field released at $\boldsymbol{x}_0$ at time $t = 0$, which is inexpensive to evaluate compared to solving the Darcy flow equations. Then, since L^2-convergence implies convergence in distribution, assuming that the set of points of discontinuity of the mapping G_{exit} has probability measure zero, we have by the continuous mapping theorem of probability theory

$$\lim_{\ell\to\infty} \|F - \hat{F}_\ell\|_\infty = 0, \qquad \hat{F}_\ell(s) := \mathbf{P}_{\boldsymbol{\xi}\sim\mu}\left(G_{\text{exit}}\left(\mathcal{S}_{\ell,M}\boldsymbol{u}(\boldsymbol{\xi})\right) \leq s\right),$$

where F denotes the true CDF of f_{exit} Thus, we are assured convergence of the CDF based on the surrogate $\mathcal{S}_{\ell,M}\boldsymbol{u}$ for the true velocity $\boldsymbol{u}$ to the true CDF for the breakthrough time.

4.3 Gaussian process emulators

Having described basic GPE methodology in Sect. 4.1, we now turn to the construction of GPEs for multivariate scalar-valued functions $f: \Xi \to \mathbb{R}$. Again, the approach is similar to multivariate geostatistics. We shall consider the *full Bayesian* approach to GPE (cf. Kennedy and O'Hagan 2001, O'Hagan 2006), which also entails specifying prior distributions for the hyperparameters contained in the mean and covariance functions which are also conditioned on the evaluations of f at the design points $\boldsymbol{\xi}_j$.

As before, we start with a linear regression model for the mean

$$m \colon \Xi \to \mathbb{R}, \qquad m(\boldsymbol{\xi}) = m(\boldsymbol{\xi}; \boldsymbol{\beta}) = \sum_{k=1}^{p} \beta_k h_k(\boldsymbol{\xi}), \quad \boldsymbol{\beta} \in \mathbb{R}^p,$$

with known regression functions $\boldsymbol{h} = (h_1, \dots, h_p)$, $h_k \colon \Xi \to \mathbb{R}$ ($h_1 \equiv 1$ and $h_2(\boldsymbol{\xi}) = \boldsymbol{\xi}$ are common choices) and unknown coefficients $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$. For the emulator's covariance function $c \colon \Xi \times \Xi \to \mathbb{R}$ we fix the squared exponential kernel

$$c(\boldsymbol{\xi}, \boldsymbol{\xi}') = c(\boldsymbol{\xi}, \boldsymbol{\xi}'; \sigma^2, B) = \sigma^2 \exp(-(\boldsymbol{\xi} - \boldsymbol{\xi}')^\top B(\boldsymbol{\xi} - \boldsymbol{\xi}')), \qquad \boldsymbol{\xi}, \boldsymbol{\xi}' \in \Xi, \quad (31)$$

where $\sigma^2 > 0$ is the marginal variance and $B = \operatorname{diag}(b_1, \dots, b_M) \in \mathbb{R}^{M \times M}$, $b_i > 0$ is a matrix of so-called *smoothness parameters*. For the squared exponential covariance (31) and choices for h_1 and h_2 mentioned above, it is known that the realizations of the Gaussian process are almost surely analytic w.r.t. $\boldsymbol{\xi}$. For other covariance functions, such as the family of Matérn kernels, one obtains Gaussian processes with realizations of different smoothness orders.[6]

Thus, for fixed given $\boldsymbol{\beta}$, σ^2, and B, the (prior) Gaussian process model for the output of f for an arbitrary input $\boldsymbol{\xi} \in \Xi$ is

$$f(\boldsymbol{\xi}) \sim \mathsf{N}(m(\boldsymbol{\xi}; \boldsymbol{\beta}), c(\boldsymbol{\xi}, \boldsymbol{\xi}; \sigma^2, B)).$$

Similarly, for fixed $\boldsymbol{\beta}$, σ^2, and B, the vector $\boldsymbol{f} = \big(f(\boldsymbol{\xi}_1), \dots, f(\boldsymbol{\xi}_n)\big)^\top$ of values of the Gaussian process at a set of design points $\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n\}$ has the n-variate Gaussian distribution

$$\boldsymbol{f} = \big(f(\boldsymbol{\xi}_1), \dots, f(\boldsymbol{\xi}_n)\big)^\top \sim \mathsf{N}(\boldsymbol{H}\boldsymbol{\beta}, \boldsymbol{C}_{\sigma^2, B})$$

where $\boldsymbol{H} = (h_k(\xi_j)) \in \mathbb{R}^{n \times p}$ and $\boldsymbol{C}_{\sigma^2, B} = (c(\xi_i, \xi_j; \sigma^2, B)) \in \mathbb{R}^{n \times n}$. We denote the probability density of this random vector $\boldsymbol{f} \in \mathbb{R}^n$ by

$$p(\boldsymbol{f} \mid \boldsymbol{\beta}, \sigma^2, B) \propto \exp\left(-\frac{1}{2}(\boldsymbol{f} - \boldsymbol{H}\boldsymbol{\beta})^\top \boldsymbol{C}_{\sigma^2, B}^{-1}(\boldsymbol{f} - \boldsymbol{H}\boldsymbol{\beta})\right).$$

Suitable values for the parameters $\boldsymbol{\beta}$, σ^2, and B are usually not known a priori and should be inferred based on the evaluations $\boldsymbol{f}$. This is typically done in a Bayesian fashion, i.e., we choose hyperpriors for these parameters which are then conditioned on the data $\boldsymbol{f} = \big(f(\boldsymbol{\xi}_1), \dots, f(\boldsymbol{\xi}_n)\big)^\top$. Common choices for $(\boldsymbol{\beta}, \sigma^2)$ are a normal-inverse-gamma prior or a Jeffreys prior with density $p(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2}$ (cf. Oakley and O'Hagan 2002, Stone 2011) since these allow for closed-form expressions for the

[6] We have also explored other covariance models such as the Matérn kernels for GPE surrogates; however, the overall conclusions in the numerical experiments were about the same as for the squared exponential (31).

resulting (marginal) posteriors. Given evaluations $\boldsymbol{f}$, the resulting posterior for the parameters $(\boldsymbol{\beta}, \sigma^2)$ is then

$$p(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{f}, B) \propto p(\boldsymbol{f} \mid \boldsymbol{\beta}, \sigma^2, B)\, p(\boldsymbol{\beta}, \sigma^2).$$

For the estimation of the smoothness parameters B a "full" Bayesian inference based on data $\boldsymbol{f}$ would require Markov chain Monte Carlo simulations. Instead, one often simply computes a point estimate based on maximizing the marginal likelihood $p(\boldsymbol{f} \mid B) \propto \int p(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{f}, B) p(\boldsymbol{\beta}, \sigma^2)\, \mathrm{d}\boldsymbol{\beta}\, \mathrm{d}\sigma^2$ for which analytic formulas are available Stone (2011, Section 2.3.4). This often yields competitive results to a full Bayesian inference Kracker et al. (2010).

Given $\boldsymbol{f}$, the posterior density for the output $f(\boldsymbol{\xi})$ at new location $\boldsymbol{\xi}$ is then

$$p(f(\boldsymbol{\xi}) \mid \boldsymbol{f}, \boldsymbol{\beta}, \sigma^2, B) \propto p(\boldsymbol{f} \mid \boldsymbol{\beta}, \sigma^2, B) p(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{f}, B).$$

Marginalization by integrating out $\boldsymbol{\beta}$ and σ^2 can be done analytically for a normal-inverse-gamma or Jeffreys prior $p(\boldsymbol{\beta}, \sigma^2)$ and results in a *Student-t process* (cf. Shah 2014) for the prediction of the output of f, i.e.,

$$f(\boldsymbol{\xi}) \mid \boldsymbol{f} \;\sim\; \mathrm{t}_{n-p}\left(\hat{m}_n(\boldsymbol{\xi}), \hat{\sigma}^2 \hat{c}_n(\boldsymbol{\xi}, \boldsymbol{\xi})\right), \tag{32}$$

where $\hat{m}_n$ and $\hat{c}_n$ are the mean and covariance obtained by universal kriging applied to f given the observations $\boldsymbol{f}$ (see (16) and (17)) with $\sigma^2 = 1$, respectively, and where $\hat{\sigma}^2$ is given by

$$\hat{\sigma}^2 = \frac{1}{n-p} \boldsymbol{f}^\top \boldsymbol{C}^{-1/2} \left(\boldsymbol{I} - \boldsymbol{C}^{-1/2} \boldsymbol{H} \left(\boldsymbol{H}^\top \boldsymbol{C}^{-1} \boldsymbol{H}\right)^{-1} \boldsymbol{H}^\top \boldsymbol{C}^{-1/2}\right) \boldsymbol{C}^{-1/2} \boldsymbol{f}.$$

For the prediction of f at multiple new points we obtain a multivariate Student-t-distribution with mean vector given by the evaluation of $\hat{m}_n$ at those points and covariance matrix given by evaluating $\hat{\sigma}^2 \hat{c}_n$.

Regarding the choice of the design points for multivariate GPE we require again *space filling* designs. For compact $\Xi \subset \mathbb{R}^M$ these are, e.g., Sobol' points (Owen et al. 2017) or Latin hypercube designs (Viana 2015). The latter extend also to $\Xi = \mathbb{R}^M$ w.r.t. $\mu = \mathrm{N}(0, \boldsymbol{I})$ as we require for the WIPP problem. As for the appropriate number $n \in \mathbb{N}$ of training points $\Xi_n = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n\} \subset \Xi$, a common rule of thumb calls for $n = cM$ (Loeppky et al. 2009) with a factor $c \geq 10$.

Convergence and Application Since the GPE surrogate $\hat{f}_n = \hat{m}_n$ and its covariance $\hat{c}_n$ are derived by universal kriging, we can again exploit the relation between kriging and kernel interpolation (Scheuerer et al. 2013). Again, assume $m \equiv 0$ for simplicity and c fixed as in (31). Then for compact $\Xi \subset \mathbb{R}^M$ and analytic $f\colon \Xi \to \mathbb{R}$ we have

$$\|f - \hat{f}_n\|_\infty \leq C(f)\, r^{\mathrm{D}_{\xi_1,\ldots,\xi_n}(\Xi)},$$

for a $0 < r < 1$ as well as

$$\hat{c}_n(x, x) \leq C\, r^{2\mathrm{D}_{\xi_1,\ldots,\xi_n}(\Xi)}.$$

Thus, besides uniform convergence of the surrogate $\hat{f}_n \to f$, we also have vanishing output uncertainty regarding $f(\boldsymbol{\xi})$ as $n \to \infty$—which is a consistency statement for the posterior for f here given by the Gaussian or Student-t process $\widehat{G}_n$. However, to our knowledge, no L^2-convergence statements are available for the case of unbounded $\Xi = \mathbb{R}^M$, as the setting of the WIPP problem would require.

In the next section we will apply GPE to approximate the quantity of interest f_{exit} directly. Thus, for convergence with $n \to \infty$, we require f_{exit} to be sufficiently smooth (see above) which may not be the case in general. However, it may well be that the surrogate $\hat{f}_n$ and the related output uncertainty provided by the GPE for finite $n = cM$ design points is sufficiently accurate for CDF estimation. We note that also vector-valued GPE are available, see Álvarez et al. (2012), Bilionis et al. (2013), Cleary et al. (2021), Higdon et al. (2008). Hence, we could apply a GPE to approximate the FE solution of the random parametric variational problem (which depends analytically on $\boldsymbol{\xi}$, see comment above) and proceed as for polynomial collocation to provide approximate samples of $f_{\mathrm{exit}}(\boldsymbol{\xi})$. We do not consider this option in this work, since the FE space is very high dimensional (of order 10^4) and thus the GPE would involve too many parameters to estimate based on not more than 20, 000 design points.

5 Numerical results

We now perform a numerical study comparing sparse grid polynomial collocation and Gaussian process emulators as surrogates for the task of approximating the CDF of the exit time $f_{\mathrm{exit}}(\boldsymbol{\xi})$ using M terms and coefficients $\boldsymbol{\xi} \sim \mathsf{N}(0, \boldsymbol{I})$ in the truncated KL expansion of the log transmissivity field $Z = \log T$. We vary $M = 10, 20, 30$ and apply the following three surrogate approaches:

- *SGC-PDE* We apply Smolyak sparse grid polynomial collocation $\mathcal{S}_{\ell,M}$ to approximate the solution pair $(\boldsymbol{u}, p)$ of the mixed formulation and then obtain approximate samples $\hat{f}_{\mathrm{exit}}(\boldsymbol{\xi}_i)$ of the exit time by simulating the particle transport given the approximate velocity field $\mathcal{S}_{\ell,M}\boldsymbol{u}(\boldsymbol{\xi}_i)$, i.e., $\hat{f}_{\mathrm{exit}}(\boldsymbol{\xi}_i) = G_{\mathrm{exit}}(\mathcal{S}_{\ell,M}\boldsymbol{u}(\boldsymbol{\xi}_i))$ where $\boldsymbol{\xi}_i \sim \mathsf{N}(0, \boldsymbol{I})$, $i = 1, \ldots, N$ iid.
- *SGC-QoI* We apply Smolyak sparse grid polynomial collocation $\mathcal{S}_{\ell,M}$ directly to approximate the exit time $f_{\mathrm{exit}}(\boldsymbol{\xi}_i)$ and in this way obtain approximate samples via $\hat{f}_{\mathrm{exit}}(\boldsymbol{\xi}_i) = \mathcal{S}_{\ell,M} f_{\mathrm{exit}}(\boldsymbol{\xi}_i)$ where $\boldsymbol{\xi}_i \sim \mathsf{N}(0, \boldsymbol{I})$, $i = 1, \ldots, N$ iid.
- *GPE* We apply Gaussian process emulation to approximate the exit time $f_{\mathrm{exit}}(\boldsymbol{\xi}_i)$ and obtain approximate samples via $\hat{f}_{\mathrm{exit}}(\boldsymbol{\xi}_i) = \hat{m}_n(\boldsymbol{\xi}_i)$ where $\boldsymbol{\xi}_i \sim \mathsf{N}(0, \boldsymbol{I})$, $i = 1, \ldots, N$ iid and $\hat{m}_n$ denotes the GPE mean.

For each surrogate we generate $N = 20\,000$ approximate samples of the quantity of interest and compare these to $N = 20\,000$ samples of the "true" f_{exit} evaluated by solving the Darcy flow equations and particle transport problem each time (denoted **MC** for Monte Carlo in the following). The number $N = 20\,000$ of samples is derived from

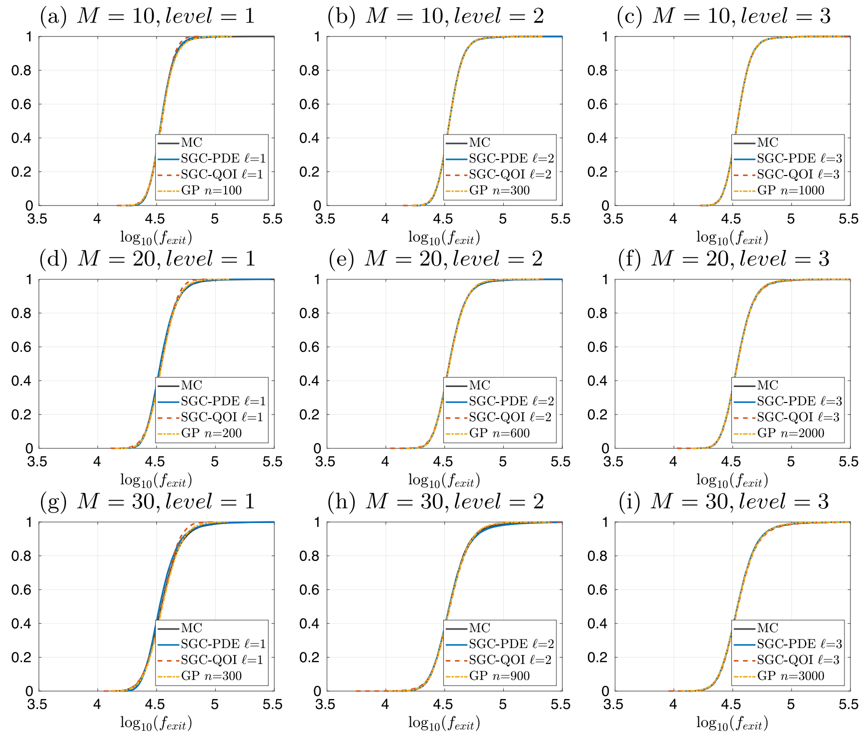


Fig. 10 Empirical CDFs obtained by Monte Carlo, SGC and GPE surrogates for different lengths M of the KLE

the error criterion outlined in Sect. 3.3. For SGC we use different levels $\ell = 1, 2, 3$, and for the GPE different numbers of design points $n = cM$ with $c = 10, 20, 30, 50, 100$. We show the resulting empirical CDFs for the log exit time in Fig. 10. It is apparent that, for each $M = 10, 20, 30$, all surrogate methods yield a very good fit to the reference ECDF obtained by the plain Monte Carlo approach. Slight deviations can be seen for the lowest level $\ell = 1$ for **SGC-QoI**, but, at least for $\ell \geq 2$, it is difficult to distinguish the four ECDFs. Therefore, we take a closer look at the performance of the surrogates in Table 2, where we report the resulting values of the KS statistic $K = \sup_{s\in\mathbb{R}} \left|\hat{F}_n(s) - F_n(s)\right|$ of the empirical CDF F_n obtained by Monte Carlo sampling of f_{exit} and the empirical CDF $\hat{F}_n$ obtained by Monte Carlo sampling of the surrogate $\hat{f}_{\text{exit}}$. Moreover, we indicate by an asterisk that the error K in the ECDFs is negligible, i.e., that the Kolmogorov–Smirnov test is passed (at significance level $\alpha = 0.05$), and hence there is no indication that the samples were drawn from different distributions. We make the following observations:

- For $M = 10, 20$ all three surrogates pass the KS-test at least for level $\ell \geq 2$ (SGC) or $n \geq 30M$ design points (GPE). For $M = 30$ this is also the case for SGC-PDE with $\ell \geq 2$ and GPE with $n = 100M$. Thus, by employing the considered surrogates we can obtain an ECDF for the exit time which is essentially

Table 2 Performance of the SGC and GPE surrogates for different lengths M of the KL expansion measured by the value of the resulting KS statistic K

Surrogate		M = 10		M = 20		M = 30	
		n	K	n	K	n	K
SGC-PDE	$\ell = 1$	21	0.0128*	41	0.0281	61	0.0495
SGC-PDE	$\ell = 2$	241	0.0028*	881	0.0045*	1921	0.0118∗
SGC-PDE	$\ell = 3$	2001	0.0019*	13,201	0.0023*	41,601	0.0052∗
SGC-QOI	$\ell = 1$	21	0.0271	41	0.0293	61	0.0435
SGC-QOI	$\ell = 2$	241	0.0065*	881	0.0088*	1921	0.0196
SGC-QOI	$\ell = 3$	2001	0.0048*	13,201	0.0089*	41,601	0.0138
GPE	$c = 10$	100	0.0136	200	0.0245	300	0.0309
GPE	$c = 20$	200	0.0092*	400	0.0191	600	0.0228
GPE	$c = 30$	300	0.0062*	600	0.0116*	900	0.0171
GPE	$c = 50$	500	0.0041*	1000	0.0070*	1500	0.0141
GPE	$c = 100$	1000	0.0031*	2000	0.0064*	3000	0.0087∗

Here, n refers to the number of PDEs to be solved for building the surrogate and an asterisk denotes that the KS-test was passed at significance level $\alpha = 0.05$

indistinguishable (for $\alpha = 0.05$) from the "true" ECDF but which required just a fraction of the computational cost of the latter. Indeed, compared to $N = 20\,000$ solutions of the Darcy flow equations, we require merely between ≈ 200 ($M = 10$) and ≈ 2000 ($M = 30$) PDE solves when a surrogate is used.

- For SGC-PDE as well as SGC-QOI we observe a steep increase in the number of PDE solves n with M but overall a robust and good performance.
- For the SGC-QoI approach we observe a significantly worse performance for $M = 30$ which may be due to insufficient (mixed) smoothness of f_{exit}.
- For the GPE approach we observe deteriorating performance for increasing M, i.e., we require a larger factor c for the number of design points $n = cM$ in order to pass the KS test and have small values of K ($c = 20$ for $M = 10$, $c = 30$ for $M = 20$ and $c = 100$ for $M = 30$). This may be due to the curse of dimensionality for kernel interpolation methods.

Changing the trend model for $\log T$

Despite the overall positive observations for the employed surrogates made so far we report how the outcome may change if we simply use a different trend model for the mean of the log transmissivity field $\log T$. Instead of using the constant, linear in x_1, and zone indicator regression functions h_1, h_2, and h_5, respectively, see (9), we only use the constant h_1. This leads to a different Matérn covariance function used for $\log T$, see Table 1 and thus also to different eigenvalues and eigenfunctions in the KL expansion. Moreover, the smoothness properties of the mapping $\boldsymbol{\xi} \mapsto f_{\text{exit}}(\boldsymbol{\xi})$ may change as well. In fact, in Table 3 we observe a much diminished performance of all three surrogate techniques: Now only SGC-PDE passes the KS test and only for the shorter KL truncation kength $M = 10, 20$. However, SGC-PDE and GPE provide a visually acceptable fit to the reference ECDF in Fig. 11, whereas we clearly

Table 3 Rerun of Table 2 but for constant mean for log T

Surrogate		M = 10		M = 20		M = 30	
		n	K	n	K	n	K
SGC-PDE	$\ell = 1$	21	0.0537	41	0.0653	61	0.0621
SGC-PDE	$\ell = 2$	241	0.0123*	881	0.0130*	1921	0.0146
SGC-PDE	$\ell = 3$	2001	0.0121*	13,201	0.0345	41,601	0.0387
SGC-QOI	$\ell = 1$	21	0.1099	41	0.1340	61	0.1301
SGC-QOI	$\ell = 2$	241	0.0485	881	0.0798	1921	0.0697
SGC-QOI	$\ell = 3$	2001	0.0369	13,201	0.0577	41,601	0.1711
GPE	$c = 10$	100	0.0366	200	0.0546	300	0.0815
GPE	$c = 20$	200	0.0373	400	0.0415	600	0.0591
GPE	$c = 30$	300	0.0153	600	0.0368	900	0.0615
GPE	$c = 50$	500	0.0188	1000	0.0405	1500	0.0415
GPE	$c = 100$	1000	0.0192	2000	0.0258	3000	0.0422

An asterisk denotes that the KS test was passed at significance level $\alpha = 0.05$

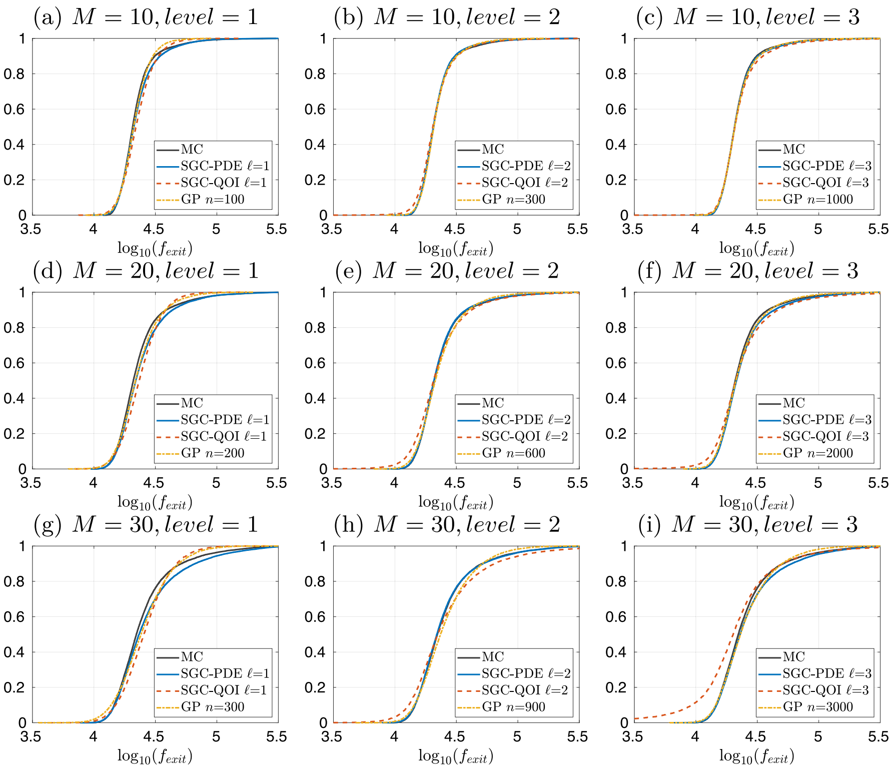


Fig. 11 Rerun of Fig. 10 but for constant mean for log T

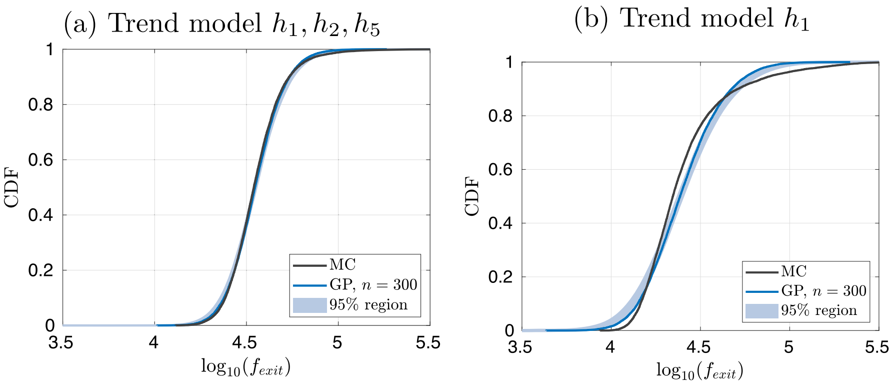


Fig. 12 95% credibility region for CDF of breakthrough time based on GPE with $n = 300$ for $M = 30$ KL terms for different trend models

see a deterioration for the SGC-QoI surrogate. This distinctly worse performance of SGC-QoI may be due to insufficient smoothness of $\boldsymbol{\xi} \mapsto f_{\text{exit}}(\boldsymbol{\xi})$ in this case.

For the GPE surrogate we also evaluate to what extent the accompanying Gaussian model for this surrogate's output uncertainty covers the deviation from the reference CDF. To this end, we focus on the setting where the GPE surrogate performs worst, i.e., $M = 30$ using $n = 300$ design points, and compute a 95% credibility region for the CDF based on 10 000 random draws of surrogates from the trained GPE. The results are reported in Fig. 12 for both trend models. We observe that the Gaussian output uncertainty model appears overconfident in the case of the constant trend model. Thus, this experiment indicates that a sufficiently good performance of the surrogates for CDF estimation of exit times may depend on various aspects of the problem—such as the choice of the trend model for the log transmissivity field.

Convergence Study The negative results for the constant trend model raise the question whether we simply did not use enough design points n or sufficiently high sparse collocation level ℓ for the GPE and SGC surrogates, respectively, or whether the quantity of interest is simply too rough to be approximated well by these methods. To this end, we perform a convergence study for both scenarios: constant trend model and "best" trend model using h_1, h_2, and h_5 in (9). We report the associated L^2_μ-errors of the SGC surrogates for the flux $\boldsymbol{u}$ and the quantity of interest in Tables 4 and 5, respectively. We notice significantly larger errors for the constant trend model. In order to allow for a sufficiently high polynomial degree for SGC to observe a significant error decay, we restrict ourselves to the low-dimensional case of $M = 2$ and $M = 5$ KL terms. We report the resulting errors of the velocity and the quantity of interest in Fig. 13. There we clearly observe a decaying error for increasing level ℓ and number of sparse grid nodes $|\Xi_{\ell,M}|$, respectively. Moreover, we observe that the rate of convergence for both quantities is affected by the larger number of KL terms and the choice of the trend model. The former was already observed in Ernst and Sprungk (2014). The latter is also related to an observation made in Ernst and Sprungk (2014): since the constant trend model yields a larger estimated value for the variance σ^2, this in turn leads to a slower convergence rate of SGC.

Table 4 $L^2_\mu(\Xi, \boldsymbol{H}(\mathrm{div}; D))$ error of SGC surrogates for the flux $\boldsymbol{u}$ for the two different trend models

Trend model	Surrogate		M = 10	M = 20	M = 30
h_1,h_2,h_5	SGC	$\ell = 1$	5.9897E-3	1.2933E-2	1.6810E-2
	SGC	$\ell = 2$	2.1354E-3	6.3868E-3	9.3400E-3
	SGC	$\ell = 3$	6.1168E-4	2.5686E-3	4.3738E-3
h_1	SGC	$\ell = 1$	4.0723E-2	1.1149E-1	1.7963E-1
	SGC	$\ell = 2$	4.0331E-2	1.1113E-1	1.7329E-1
	SGC	$\ell = 3$	3.9595E-2	1.0598E-1	1.6928E-1

Table 5 $L^2_\mu(\Xi, \mathbb{R})$ error of SGC surrogates for the exit time f_{exit} for the two different trend models

Trend model	Surrogate		M = 10	M = 20	M = 30
h_1, h_2, h_5	SGC	$\ell = 1$	1.2296E-3	2.7434E-3	6.0602E-3
	SGC	$\ell = 2$	1.2699E-4	4.8917E-4	2.1426E-3
	SGC	$\ell = 3$	2.0075E-5	9.6514E-5	4.4401E-4
h_1	SGC	$\ell = 1$	7.0990E-3	1.5259E-2	2.8396E-2
	SGC	$\ell = 2$	2.9464E-3	7.7314E-3	1.5502E-2
	SGC	$\ell = 3$	1.9730E-3	9.2632E-3	1.8001E-2

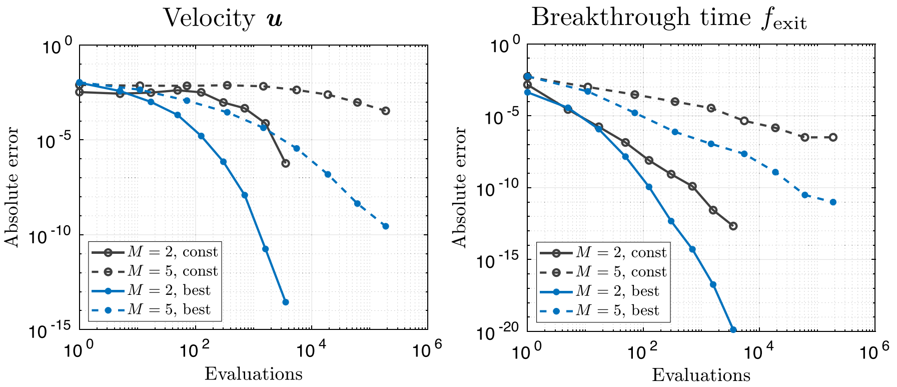
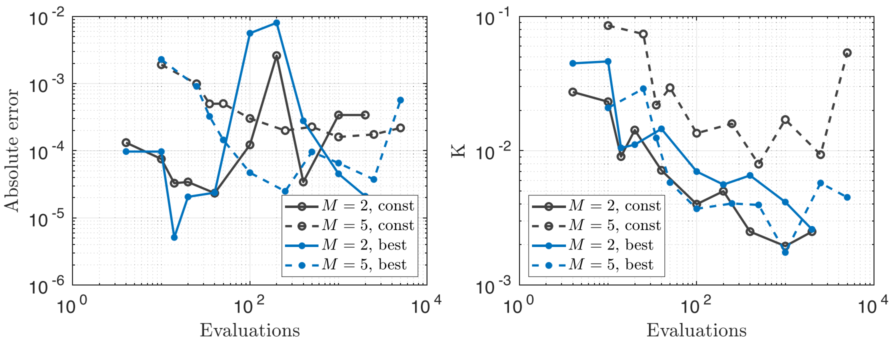


Fig. 13 L^2_μ-error of SGC surrogates for the velocity $\boldsymbol{u}$ (left) and exit time f_{exit} (right). For the flux we used the norm in $\boldsymbol{H}(\mathrm{div}; D)$ to quantify the difference between $\boldsymbol{u}(\xi)$ and $\mathcal{S}_{\ell,M}\boldsymbol{u}(\boldsymbol{\xi})$

Regarding the application of GPE to approximate the quantity of interest, we perform a similar study as for SGC using $M = 2$ and $M = 5$ KL terms. The results are displayed in Fig. 14. We observe that the L^2_μ-error (left panel) does not decay with increasing number of design points, at least not in the applied regime of up to $n = 1000M$ design points. Despite this, we observe a decay of the KS test statistic value K, i.e., the L^∞-error of the ECDF for the quantity of interest, except for $M = 5$ and the constant trend model.



2

Fig. 14 L^2_μ-error (left) and K-S test value K (right) of GPE surrogates for exit time f_{exit}

6 Conclusion

In this work we have presented a complete uncertainty propagation workflow for groundwater flow and particle transport simulations based on a real-world application related to the site performance assessment for a nuclear waste repository. We described in detail the construction of a stochastic model for an uncertain transmissivity field by geostatistical methods using the available observational data. Our main focus was the direct comparison of two established surrogate approaches for uncertainty propagation analysis: sparse grid stochastic collocation and Gaussian process emulation. Both methods originate from different communities, i.e., numerical analysis and computational statistics, respectively. Our purpose was to describe and contrast the fundamental ideas and principles underlying both methods and compare their performance for the UQ problem under consideration, specifically for CDF estimation for scalar quantities of interest, in this case the travel time of groundwater-borne radionuclides. The overall conclusion is that both methods can achieve significant reduction in computational cost over naive Monte Carlo simulation, reducing the computational burden by a factor of 10 to even 100 in some cases considerered. Moreover, we have observed that the GPE surrogate seems to be more adversely affected by the high dimensionality of the input space compared with sparse grid collocation, which is not surprising given the unfavorable scaling of the filling distance with dimension. On the other hand, stochastic collocation must also be applied with care, since the quantity to be approximated has to depend sufficiently smoothly on the random inputs—such as the solution of the random PDE. However, the remarkable performance of both surrogates seems to be affected by modelling choices for the random log transmissivity field such as choice of the trend or regression model for the mean. Although this effect could be explained mathematically in our case, it does place limitations on the practical benefits of UQ surrogate methods for CDF estimation in groundwater flow applications.

Acknowledgements The authors would like to thank Elisabeth Ullmann (TU München) as well as Gerald van den Boogaart and Silke Konsulke (HZDR Dresden-Rossendorf) for numerous contributions in an early phase of this work. The first two authors gratefully acknowledge their invitation to the 2018 Uncertainty Quantification Programme at the Isaac Newton Institute, where they participated in numerous clarifying discussions on surrogates.

Funding Open Access funding enabled and organized by Projekt DEAL. The first and third author gratefully acknowledge partial funding support from the German society for radioactive waste disposal BGE in the MeQUR project within the research cluster URS.

Declaration

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

References

Álvarez, M.A., Rosasco, L., Lawrence, N.D.: Kernels for vector-valued functions: a review. Found. Trends® Mach. Learn. **4**(3), 195–266 (2012). https://doi.org/10.1561/2200000036

Athreya, K.B., Lahiri, S.N.: Measure Theory and Probability Theory. Springer, Berlin (2006)

Babuška, I., Nobile, F., Tempone, R.: A stochastic collocation method for elliptical partial differential eqautions with random input data. SIAM J. Numer. Anal. **45**(3), 1005–1034 (2007)

Babuška, I., Nobile, F., Tempone, R.: A stochastic collocation method for elliptic partial differential equations with random input data. SIAM Rev. **52**(1), 317–355 (2010)

Bäck J, Nobile, F., Tamellini, L., et al.: Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: A numerical comparison. In: Spectral and High Order Methods for Partial Differential Equations. Springer Berlin Heidelberg, pp 43–62, (2011) https://doi.org/10.1007/978-3-642-15337-2_3

Barthelmann, V., Novak, E., Ritter, K.: High dimensional polynomial interpolation on sparse grids. Adv. Comput. Math. **12**, 273–288 (2000)

Bastos, L.S., O'Hagan, A.: Diagnostics for gaussian process emulators. Technometrics **51**(4), 438–524 (2009)

Bilionis, I., Zabaras, N., Konomi, B.A., et al.: Multi-output separable gaussian process: Towards an efficient, fully Bayesian paradigm for uncertainty quantification. J. Comput. Phys. **241**, 212–239 (2013). https://doi.org/10.1016/j.jcp.2013.01.011

Boffi, D., Brezzi, F., Fortin, M.: Mixed Finite Element Methods and Applications. Springer Series in Computational Mathematics. Springer Science & Business Media, Berlin (2013)

Cleary, E., Garbuno-Inigo, A., Lan, S., et al.: Calibrate, emulate, sample. J. Comput. Phys. **424**(109), 716 (2021). https://doi.org/10.1016/j.jcp.2020.109716

Cressie, N.A.: Statistics for Spatial Data. Wiley-Interscience, New York (1991)

Currin, C., Mitchell, T., Morris, M., et al.: Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. J. Am. Stat. Assoc. **86**(416), 953–963 (1991)

de Marsily, G.: Quantitative Hydrogeology: Groundwater Hydrology for Engineers. Academic Press (1986)

Diaconis, P.: Bayesian numerical analysis. In: Gupta, S.S., Berger, J.O. (eds.) Statistical Decision Theory and Related Topics IV. Springer, vol. 1, pp. 163–175. New York, NY (1988)

Eiermann, M., Ernst, O.G., Ullmann, E.: Computational aspects of the stochastic finite element method. Comput. Vis. Sci. **10**(1), 3–15 (2007)

Ern, A., Guermond, J.L.: Finite Elements II: Galerkin approximation, elliptic and mixed PDEs, texts in applied mathematics. Springer Nature, Switzerland (2021)

Ernst, O.G., Sprungk, B.: Stochastic collocation for elliptic PDEs with random data: The lognormal case. In: Garcke J, Pflüger D (eds) Sparse Grids and Applications – Munich 2012, LNCSE, vol 97. Springer International Publishing, p 29–53 (2014)

Ernst, O.G., Sprungk, B., Tamellini, L.: Convergence of sparse collocation for functions of countably many Gaussian random variables (with application to elliptic PDEs). SIAM J. Numer. Anal. **56**(2), 877–905 (2018). https://doi.org/10.1137/17m1123079

Ernst, O.G., Sprungk, B., Tamellini, L.: On expansions and nodes for sparse grid collocation of lognormal elliptic PDEs. In: Bungartz HJ, Garcke J, Pflüger D (eds) Sparse grids and applications – Munich 2018, LNCSE, vol 144. Springer International Publishing, pp 1–31 (2021) https://doi.org/10.1007/978-3-030-81362-8_1

Ernst, O.G., Pichler, A., Sprungk, B.: Wasserstein sensitivity of risk and uncertainty propagation. SIAM/ASA J. Uncertain. Quant. **10**(3), 915–948 (2022). https://doi.org/10.1137/20M1325459

Freeze, R.A.: A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. Water Resour. Res. **11**(5), 725–741 (1975). https://doi.org/10.1029/WR011i005p00725

Ghanem, R.G., Spanos, P.D.: Stochastic Finite Elements: A Spectral Approach. Springer, New York (1991)

Graham, I.G., Scheichl, R., Ullmann, E.: Mixed finite element analysis of lognormal diffusion and multilevel Monte Carlo methods. Stoch PDE Anal. Comp. **4**, 41–75 (2016)

Gunzburger, M.D., Webster, C.G., Zhang, G.: Stochastic finite element methods for partial differential equations with random input data. Acta Numer. **23**, 521–650 (2014). https://doi.org/10.1017/S0962492914000075

Hackbusch, W.: Hierarchical Matrices: Algorithms and Analysis, vol. 49. Springer, Berlin (2015)

Harville, D.A.: Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Stat. Assoc. **72**(358), 320–338 (1977). https://doi.org/10.1080/01621459.1977.10480998

Hennig, P., Osborne, M.A., Kersting, H.P.: Probabilistic Numerics - Computation as Machine Learning. Cambridge University Press (2022). https://doi.org/10.1017/9781316681411

Higdon, D., Gattiker, J., Williams, B., et al.: Computer model calibration using high-dimensional output. J. Am. Stat. Assoc. **103**(482), 570–583 (2008)

Hoeksema, R.J., Kitanidis, P.K.: Analysis of the spatial structure of properties of selected aquifers. Water Resour. Res. **21**(4), 563–572 (1985). https://doi.org/10.1029/WR021i004p00563

Kennedy, M.C., O'Hagan A.: Bayesian calibration of computer models. J. R. Stat. Soc. Part B 63(Part 3):425–464 (2001)

Khoromskij, B.N., Litvinenko, A., Matthies, H.G.: Application of hierarchical matrices for computing the Karhunen-loève expansion. Computing **84**(1–2), 49–67 (2009)

Kitanidis, P.K.: Parametric estimation of covariances of regionalized variables. Water Resour. Bull. **23**(4), 557–567 (1987). https://doi.org/10.1111/j.1752-1688.1987.tb00832.x

Kitanidis, P.K.: Introduction to Geostatistics: Applications to Hydrogeology. Cambridge University Press, Cambridge (1997a)

Kitanidis, P.K.: A variance-ratio test for supporting a variable mean in kriging. Math. Geol. **29**(3), 335–348 (1997). https://doi.org/10.1007/BF02769639

Kracker, H., Bornkamp, B., Kuhnt, S., et al.: Uncertainty in gaussian process interpolation. In: Devroye L, Karasözen B, Kohler M, et al (eds) Recent Developments in Applied Probability and Statistics: Dedicated to the Memory of Jürgen Lehn. Physica-Verlag HD, Heidelberg, pp 79–102, https://doi.org/10.1007/978-3-7908-2598-5_4 (2010)

Linde, N., Ginsbourger, D., Irving, J., et al.: On uncertainty quantification in hydrogeology and hydrogeophysics. Adv. Water Resour. **110**, 166–181 (2017). https://doi.org/10.1016/j.advwatres.2017.10.014

Liu, D., Litvinenko, A., Schillings, C., et al.: Quantification of airfoil geometry-induced aerodynamic uncertainties–comparison of approaches. SIAM/ASA J. Uncertain. Quant. **5**(1), 334–352 (2017). https://doi.org/10.1137/15M1050239

Loeppky, J.L., Sacks, J., Welch, W.J.: Choosing the sample size of a computer experiment: a practical guide. Technometrics **51**(4), 366–376 (2009)

Lord, G.J., Powell, C.E., Shardlow, T.: An Introduction to Computational Stochastic PDEs, Cambridge Texts in Applied Mathematics, vol. 50. Cambridge University Press, Cambridge (2014)

Narcowich, F., Ward, J., Wendland, H.: Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. Constr. Approx. **24**, 175–186 (2006). https://doi.org/10.1007/s00365-005-0624-7

Nevai, G.P.: Mean convergence of Lagrange interpolation, I. J. Approx. Theory **18**(4), 363–377 (1976). https://doi.org/10.1016/0021-9045(76)90008-3

Nevai, P.: Mean convergence of Lagrange interpolation III. Trans. Am. Math. Soc. **282**(2), 669–698 (1984). https://doi.org/10.2307/1999259

Nevai, P.G.: Mean convergence of Lagrange interpolation, ii. J. Approx. Theory **30**(4), 263–276 (1980). https://doi.org/10.1016/0021-9045(80)90030-1

Nobile, F., Tempone, R., Webster, C.G.: A sparse grid stochastic collocation method for partial differential equations with random input data. SIAM J. Numer. Anal. **46**(5), 2309–2345 (2008)

Novak, E., Ritter, K.: Simple cubature formulas with high polynomial exactness. Constr. Approx. **15**, 499–522 (1999)

Oakley, J., O'Hagan, A.: Bayesian inference for the uncertainty distribution of computer model outputs. Biometrika **89**(4), 769–784 (2002)

O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial. Reliab. Eng. Syst. Saf. **91**, 1290–1300 (2006). https://doi.org/10.1016/j.ress.2005.11.025

Owen, N.E., Challenor, P., Menon, P.P., et al.: Comparison of surrogate-based uncertainty quantification methods for computationally expensive simulators. SIAM/ASA J. Uncertain. Quant. **5**, 403–436 (2017). https://doi.org/10.1137/15M1046812

Ritter, K.: Average-Case Analysis of Numerical Problems. No. 1733 in Lecture Notes in Mathematics, Springer. https://doi.org/10.1007/BFb0103934 (2000)

Sacks, J., Welch, W.T., Mitchell, T.J., et al.: Design and analysis of computer experiments. Stat. Sci. **4**(4), 409–423 (1989)

Scheuerer, M., Schaback, R., Schlather, M.: Interpolation of spatial data: a stochastic or a deterministic problem? Eur. J. Appl. Math. **24**(4), 601–629 (2013). https://doi.org/10.1017/S0956792513000016

Schwab, C., Gittelson, C.J.: Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. Acta Numer. **20**, 291–467 (2011)

Shah, A., Wilson, A., Ghahramani Z.: Student-t processes as alternatives to gaussian processes. In: Kaski, S., Corander, J. (eds) Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, vol 33. PMLR, Reykjavik, Iceland, pp. 877–885, https://proceedings.mlr.press/v33/shah14.html (2014)

Sickel, W., Ullrich, T.: The Smolyak agorithm, sampling on sparse grids and functions spaces of dominated mixed smoothness. East J. Approx. **13**(4), 387–426 (2007)

Stein, M.: Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York (1999)

Stone, N.: Gaussian process emulators for uncertainty analysis in groundwater flow. PhD thesis, University of Nottingham (2011)

Sudret, B., Marelli, S., Wiart, J.: Surrogate models for uncertainty quantification: An overview. In: 2017 11th European Conference on Antennas and Propagation (EUCAP), pp 793–797,https://doi.org/10.23919/EuCAP.2017.7928679 (2017)

Trefethen, L.N.: Approximation Theory and Approximation Practice. SIAM, Philadelphia (2013)

U.S. Department of Energy (DOE) (2004) Title 40 CFR Part 191 Subparts B and C. Compliance Recertification Application 2004 for the Waste Isolation Pilot Plant Appendix TFIELD-2004 Transmissivity Fields. Tech. Rep. DOE/WIPP 2004/3231, Carlsbad Field Office, Carlsbad, NM

U.S. Department of Energy (DOE) (2014) Title 40 CFR Part 191 Subparts B and C. Compliance Recertification Application 2014 for the Waste Isolation Pilot Plant Appendix TFIELD-2014 Transmissivity Fields. Tech. Rep. DOE/WIPP 14-3503, Carlsbad Field Office, Carlsbad, NM

Viana, F.A.C.: A tutorial on latin hypercube design of experiments. Qual. Reliabail. Eng. Int. **32**, 1975–1985 (2015). https://doi.org/10.1002/qre.1924

Wendland, H.: Scattered Data Approximation. Cambridge University Press, Cambridge (2004)

Williams, D.: Weighing the Odds, 2nd edn. Cambridge University Press, Cambridge (2004)

Wu, K., Simon, H.: Thick-restart Lanczos method for large symmetric eigenvalue problems. SIAM J. Matrix Anal. Appl. **22**(2), 602–616 (2000)

Xiu, D., Hesthaven, J.S.: High-order collocation methods differential equations with random inputs. SIAM J. Sci. Comput. **37**(3), 1118–1139 (2005)