From Data to Uncertainty: An Efficient Integrated Data-Driven Sparse Grid Approach to Propagate Uncertainty

Fabian Franzelin and Dirk Pflüger

Abstract We present a novel data-driven approach to propagate uncertainty. It consists of a highly efficient integrated adaptive sparse grid approach. We remove the gap between the subjective assumptions of the input's uncertainty and the unknown real distribution by applying sparse grid density estimation on given measurements. We link the estimation to the adaptive sparse grid collocation method for the propagation of uncertainty. This integrated approach gives us two main advantages: First, the linkage of the density estimation and the stochastic collocation method is straightforward as they use the same fundamental principles. Second, we can efficiently estimate moments for the quantity of interest without any additional approximation errors. This includes the challenging task of solving higher-dimensional integrals. We applied this new approach to a complex subsurface flow problem and showed that it can compete with state-of-the-art methods. Our sparse grid approach excels by efficiency, accuracy and flexibility and thus can be applied in many fields from financial to environmental sciences.

1 Introduction

There are different types of uncertainty [33] that influence the outcome of large systems that support risk assessment, planning, decision making, validation, etc. Uncertainties can enter the system due to missing knowledge about the physical domain, think of subsurface flow simulations, or there are inherent stochastic processes driving the system, such as Brownian motion. The quantification of the influence of such stochastic components on some quantity of interest is the task of forward propagation of uncertainty in the field of uncertainty quantification (UQ). This is challenging since the statistical characteristics of the uncertainties can be unknown or don't have an analytic representation. Furthermore, the systems or models we are interested in can be arbitrarily complex (highly nonlinear, discontinuous, etc.).

F. Franzelin (✉) • D. Pflüger

Department of Simulation Software Engineering, IPVS, University of Stuttgart, Stuttgart, Germany

e-mail: fabian.franzelin@ipvs.uni-stuttgart.de; dirk.pflueger@ipvs.uni-stuttgart.de

Therefore, we need efficient and reliable algorithms and software that can make expensive statistical analysis feasible.

In this paper we want to focus on data-driven quantification of uncertainties in numerical simulations. There are two main problems we face in that context: First, the uncertainty of the quantity of interest depends strongly on the uncertainty in the input. Therefore, one needs objective measures to get an unbiased representation of the uncertainty of the input. Second, to quantify the uncertainties of some quantity of interest, we need to evaluate the model, which can be very costly and involves to run a whole simulation. However, the accuracy of the quantities we compute should be very high, which means in general that we need to evaluate the model often, and the main challenge is to balance costs and accuracy.

The first problem of obtaining objective measures for the input's uncertainties has been assessed in various articles in the past. One idea is to use data and estimate the stochastic properties using density estimation techniques. The authors of [7] used kernel density estimators, and [1] proposed to use kernel moment matching methods, for example. A comparison between data-driven approaches and approaches based supervised estimation by experts can be found in [21]. However, often the combination of expert knowledge and data is essential if the reliability in the data is low. A very popular approach to combine them is Bayesian inference [35, 39].

The incorporation of data or estimated densities into a UQ forward problem depends on the method that is required by the application to propagate the uncertainty. For non-intrusive methods, for example, there has been done a lot of work in the field of polynomial chaos expansion (PCE) [37, 38]. The generalized PCE, however, is defined for analytic, independent marginal distributions. It has therefore been extended to the arbitrary PCE [34] that supports also dependent marginal distributions [19] and data [23]. However, global polynomials are not always the best choice to propagate uncertainty [6]. Stochastic collocation methods [2, 36] became popular in the last years, especially due to sparse grids [15, 20]. They are used to obtain a surrogate model of the expensive model function. They can overcome the curse of dimensionality to some extent [3], can handle large gradients [8] or even discontinuities in the response functions [13, 29].

In this paper we present a new approach to incorporate data into the UQ forward propagation pipeline. We propose an integrated data-driven sparse grid method, where we estimate the unknown density of the input using the sparse grid density estimation (SGDE) [24, 26] method and propagate the uncertainty using sparse grid collocation (SGC) with adaptively refined grids. The SGDE method has been widely used for Data Mining problems and can be applied for either small or large data sets. It is highly efficient with respect to learning, evaluating and sampling. It interacts seamlessly with SGC since both are based on the same fundamental principles, which can be exploited to reduce the numerical errors in the forward propagation pipeline.

This paper is structured as follows: First, we give a formal definition of a data-driven UQ forward-propagation problem in Sect. 2. In Sect. 3 we describe the methods we use for density estimation and for uncertainty propagation. Then we compare the performance of our approach with other techniques in Sect. 4. We present a lower-dimensional analytic example and a higher-dimensional subsurface flow problem in porous media. In Sect. 5 we summarize the paper and give an outlook to future work.

2 Problem Formulation

We define (Ω, Σ, P) being a complete probability space with Ω being the set of outcomes, $\Sigma \subset 2^{\Omega}$ the σ-algebra of events and $P \colon \Sigma \to [0, 1]$ a probability measure. Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d) \in \Omega$ be a random sample and let the probability law of $\boldsymbol{\xi}$ be completely defined by the probability density function $f \colon \Omega \to \mathbb{R}^+$ with $\int_{\Omega} f(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} = 1$. Consider a model $\mathcal{M}$ defined on a bounded physical domain $\mathbf{x} \in D \subset \mathbb{R}^{d_s}$ with $1 \le d_s \le 3$, a temporal domain $t \in T \subset \mathbb{R}$ and the probability space (Ω, Σ, P) describing the uncertainty in the model inputs as

$$u(\mathbf{x}, t, \boldsymbol{\xi}) = \mathcal{M}(\mathbf{x}, t, \boldsymbol{\xi}) \colon D \times T \times \Omega \to \mathbb{R}^{d_r} , \tag{1}$$

with $0 < d_r \in \mathbb{N}$. We restrict ourselves without loss of generality to scalar quantities of u and define an operator Q, which extracts the quantity we are interested in, i.e.

$$Q[u(\mathbf{x}, t, \boldsymbol{\xi})] \colon \mathbb{R}^{d_r} \to \mathbb{R}. \tag{2}$$

The outcome of u becomes uncertain due to its uncertain inputs $\boldsymbol{\xi}$. This uncertainty is what we want to quantify. The probability law of $\boldsymbol{\xi}$, of course, influences heavily the probability law of u. Therefore, in data-driven UQ one assumes to have a set of samples $\mathcal{D} := \{\boldsymbol{\xi}^{(k)}\}_{k=1}^{n}$, with $\boldsymbol{\xi}^{(k)} = (\xi_1^{(k)}, \ldots, \xi_d^{(k)}) \in \Omega$, which are drawn from the unknown probability density f. $\mathcal{D}$ is an objective measure describing the uncertainty we want to propagate. A schematic representation of the data-driven UQ pipeline is given in Fig. 1.
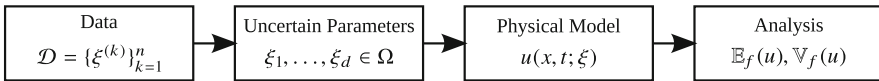
Data $\mathcal{D} = \{\xi^{(k)}\}_{k=1}^{n}$	Uncertain Parameters $\xi_1, \ldots, \xi_d \in \Omega$	Physical Model $u(x, t; \xi)$	Analysis $\mathbb{E}_f(u), \mathbb{V}_f(u)$

Fig. 1 Data-driven UQ forward pipeline. The data set $\mathcal{D}$ describes the stochastic characteristics of the uncertain parameters $\boldsymbol{\xi}$ for some physical model u. The underlying probability density function f is unknown. The stochastic analysis of the uncertain outcome of u depends strongly on f

3 Methodology

In this section we introduce the methods we use to propagate uncertainty. We formally introduce stochastic collocation and the concept of sparse grids based on an interpolation problem. We present the sparse grid density estimation method and describe how it can be used to estimate efficiently moments of some quantity of interest.

3.1 *Stochastic Collocation*

In stochastic collocation we search for a function g that approximates the unknown model function u. We solve N deterministic problems of u at a set of collocation points $\Xi_N := \{\boldsymbol{\xi}^{(k)}\}_{k=1}^N \subset \Omega$ and impose

$$g^{(i)}(\boldsymbol{\xi}^{(k)}) := Q[u(x_i, t_i, \boldsymbol{\xi}^{(k)})], \ \forall \boldsymbol{\xi}^{(k)} \in \Xi_N \ . \tag{3}$$

at a selected point in space $x_i \in D$ and time $t_i \in T$. This is, of course, nothing else than an interpolation problem. A common choice for $g^{(i)}$ is to use a sum of ansatz functions on some mesh with either global [17] or local support [8, 12, 15]. The expensive stochastic analysis is then done on the cheap surrogate $g^{(i)}$.

For simplicity in the notation we omit in the following the index i on g and focus on the approximation in Ω. Without loss of generality, we assume furthermore that there exists a bijective transformation from Ω to the unit hypercube and assume the collocation nodes $\boldsymbol{\xi}^{(k)}$ in the following to stem from $[0, 1]^d$.

3.2 *Sparse Grids*

We introduce here the most important properties of sparse grids in the context of interpolation problems. The general idea of sparse grids is based on a hierarchical definition of a one-dimensional basis. This means that the basis is inherently incremental. We exploit this property in higher-dimensional settings to overcome the curse of dimensionality to some extent. For details and further reading we recommend [3, 25]. For adaptive sparse grids and efficient implementations of sparse grid methods we refer to [27], for suitable refinement criteria for UQ problems you may read [8, 15, 16].

Suppose we are searching for an interpolant g of an unknown multivariate function $u(\boldsymbol{\xi}) \in \mathbb{R}$ on the unit hypercube, i.e. $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_d) \in [0, 1]^d$. For g we restrict ourselves to the space of piecewise d-linear functions V_ℓ with ℓ being the maximum discretization level in each dimension.

Let $\mathbf{l} = \{l_1, \ldots, l_d\}$ and $\mathbf{i} = \{i_1, \ldots, i_d\}$ with $l_k > 0$ and $i_k > 0$ be multi-indices. We define a nested index set

$$\mathcal{I}_\mathbf{l} := \{(\mathbf{l}, \mathbf{i}) : 1 \leq i_k < 2^{l_k}, i_k \text{ odd}, k = 1, \ldots, d\} \tag{4}$$

of level-index vectors defining the grid points

$$\boldsymbol{\xi}_{\mathbf{l},\mathbf{i}} := (2^{-l_1} i_1, \ldots, 2^{-l_d} i_d) . \tag{5}$$

For each grid point we use the general one-dimensional reference hat function $\psi(\xi) := \max(1 - |\xi|, 0)$ to obtain the linear one-dimensional hierarchical hat functions $\psi_{l,i}(\xi)$ centered at the grid points by scaling and translation according to level l and index i as $\psi_{l,i}(\xi) := \psi(2^l \xi - i)$, see Fig. 2 (left). We obtain the higher-dimensional basis via a tensor-product approach,

$$\psi_{\mathbf{l},\mathbf{i}}(\boldsymbol{\xi}) := \prod_{k=1}^{d} \psi_{l_k,i_k}(\xi_k) . \tag{6}$$

Note that the level-index vectors $(\mathbf{l}, \mathbf{i}) \in \mathcal{I}_\mathbf{l}$ define a unique set of hierarchical increment spaces $W_\mathbf{l} := \text{span}(\{\psi_{\mathbf{l},\mathbf{i}} : (\mathbf{l}, \mathbf{i}) \in \mathcal{I}_\mathbf{l}\})$, which are shown in the center of Fig. 2. All increment spaces up to $|\mathbf{l}|_\infty = \max_i l_i \leq \ell$ span the space of piecewise d-linear functions on a full grid.

Now we take advantage of the hierarchical definition of the basis and reduce the number of grid points by choosing just those spaces $W_\mathbf{l}$ that contribute most to our approximation. An optimal choice is possible a priori if u is sufficiently smooth, i.e. if u is a function of the mixed Sobolev space H^2_{mix} where the mixed, weak derivatives
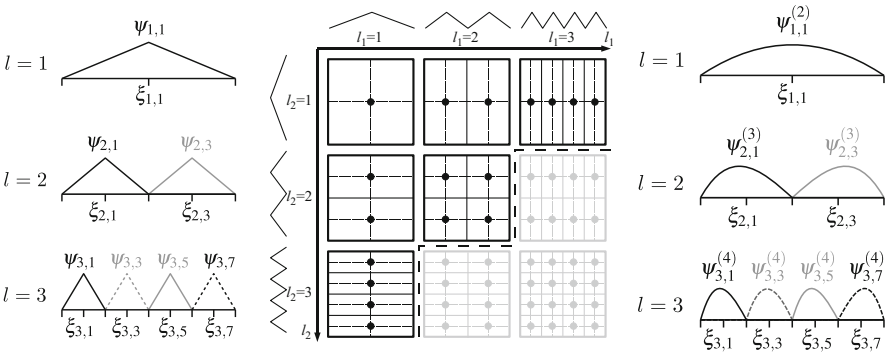

Fig. 2 One-dimensional piecewise linear basis functions up to level 3 (*left*), polynomial ones (*right*), and the tableau of hierarchical increments $W_\mathbf{l}$ up to level 3 in two dimensions (*center*)

up to order 2 are bounded. We define the sparse grid space $V_\ell^{(1)}$ as

$$V_\ell^{(1)} := \bigoplus_{|\mathbf{l}|_1 \leq \ell + d - 1} W_\mathbf{l} \ , \tag{7}$$

where we select just those subspaces that fulfill $|\mathbf{l}|_1 = \sum_{k=1}^d l_k \leq \ell + d - 1$, see the upper triangle in the center of Fig. 2. We define a sparse grid function $g_{\mathcal{I}_1} \in V_\ell^{(1)}$ as

$$g_{\mathcal{I}_1}(\boldsymbol{\xi}) = \sum_{(\mathbf{l},\mathbf{i}) \in \mathcal{I}_1, |\mathbf{l}|_1 \leq \ell + d - 1} v_{\mathbf{l},\mathbf{i}} \psi_{\mathbf{l},\mathbf{i}}(\boldsymbol{\xi}) \ , \tag{8}$$

where $v_{\mathbf{l},\mathbf{i}}$ are the so-called hierarchical coefficients. Note that we omit the index $\mathbf{l}$ when we refer to $g_{\mathcal{I}}$ being a sparse grid function defined on an adaptively refined grid.

The sparse grid space $V_\ell^{(1)}$ has one main advantage over the full tensor space V_ℓ: The number of grid points is reduced significantly from $O((2^\ell)^d)$ for a full grid to $O(2^\ell \ell^{d-1})$ while the interpolation accuracy is of order $O((2^{-\ell})^2 \ell^{d-1})$, which is just slightly worse than the accuracy of a full grid $O((2^{-\ell})^2)$ [3].

If we can impose a higher smoothness for u in a sense that all the weak mixed derivatives up to order $p + 1$ are bounded, it makes sense to employ a higher-order piecewise polynomial basis $\psi_{\mathbf{l},\mathbf{i}}^{(p)}$ with maximum degree $1 \leq p \in \mathbb{N}$ in each dimension. Note that these polynomials are defined locally, see Fig. 2 (right). Therefore, we don't suffer Runge's phenomenon even though we use equidistant grid spacing in each dimension. For details about the construction of the basis we refer to [3]. The number of grid points, of course, is the same as for the piecewise linear case. However, the interpolation accuracy is now of order $O((2^{-\ell})^{p+1} \ell^{d-1})$.

3.3 Sparse Grid Density Estimation Method

The sparse grid density estimation (SGDE) method is based on a variational problem presented in [11], first mentioned in the context of sparse grids in [9] and first developed in [26].

We want to estimate some unknown but existing probability density function f from which a set of observations/samples are available, $\mathcal{D} := \{\boldsymbol{\xi}^{(k)}\}_{k=1}^n \subset \Omega$. The SGDE method can be interpreted as a histogram approach with piecewise linear ansatz functions. We search for a sparse grid function $\hat{f}_\mathcal{K} \in V_\ell^{(1)}$ with $|\mathcal{K}| = M$ grid points that minimizes the following functional [9]

$$R(\hat{f}_\mathcal{K}) = \left\| \hat{f}_\mathcal{K} \right\|_{L_2}^2 - \frac{1}{n} \sum_{\boldsymbol{\xi}^{(k)} \in \mathcal{D}} \hat{f}_\mathcal{K}(\boldsymbol{\xi}^{(k)}) + \lambda \left\| S\hat{f}_\mathcal{K} \right\|^2 \ , \tag{9}$$

where S is some regularization operator and $0 \leq \lambda \in \mathbb{R}$ a regularization parameter.

In the unit domain the second term is the discrete version of the first one given a set of observations $\mathcal{D}$. $\mathcal{D}$ is a realization of f, so the second term implicitly has larger weights where the probability is larger. This is done explicitly in the first term by the multiplication of the pay-off function with the density $\hat{f}_{\mathcal{K}}$.

By minimizing $R(\hat{f}_{\mathcal{K}})$ we therefore search for a piecewise continuous density $\hat{f}_{\mathcal{K}}$ for which the first two terms are equal. Note that the first term is equal to the definition of the expectation value where the pay off function is the probability density function itself.

From the point of view of histograms, we can say that the sparse grid discretization defines the (overlapping) buckets. The first term in R collects the density mass in all the buckets while the second term does the same for the observations available for each hierarchical ansatz function. The penalty term balances fidelity in the data and smoothness of $\hat{f}_{\mathcal{K}}$ via the regularization parameter λ and the regularization operator S.

Solving $R(\hat{f}_{\mathcal{K}})$ leads to a system of linear equations [11]. The system matrix is the mass matrix of $\hat{f}_{\mathcal{K}}$, which depends only on the number of grid points and is therefore independent of the number of samples n. We obtain the regularization parameter via cross validation: we split $\mathcal{D}$ in a training and a test set, solve the optimization problem on the training set and compute the L_2-norm of the residual of the system of linear equations applied on the test set. For details see [26]. The estimated density function has unit integrand if we choose $S = \nabla$ [26].

Positivity, however, is not guaranteed with this approach. For the numerical examples in this paper we forced $\hat{f}_{\mathcal{K}}$ to be positive by employing a local full grid search on $\hat{f}_{\mathcal{K}}$. For piecewise linear ansatz functions there exists a simple algorithm, see Algorithm 1. A sparse grid function can locally be negative if the coefficient of an arbitrary level-index vector $(\mathbf{l}, \mathbf{i})$ is negative. If this is the case, then, due to monotony of $\hat{f}_{\mathcal{K}}$ between grid points, it is sufficient to apply a full grid search on the support of $(\mathbf{l}, \mathbf{i})$. We add grid points whenever its function value is negative and

---

Algorithm 1: Forcing the sparse grid density to be positive everywhere

Data: training sample set $\mathcal{D}$ and sparse grid $\mathcal{I}$
Result: positive sparse grid function $(\mathcal{I}, \mathbf{v})$ with unit integrand
done $\leftarrow$ *False*;
while *not done* **do**
    $\mathbf{v} \leftarrow$ *doSparseGridDensityEstimation*$(\mathcal{D}, \mathcal{I})$;
    newGridPoints $\leftarrow$ list();
    for $(\mathbf{l}, \mathbf{i}) \in \{(\mathbf{l}, \mathbf{i}) \in \mathcal{I} : v_{\mathbf{l},\mathbf{i}} < 0\}$ **do**
        negativeGridPoints $\leftarrow$ *findNegativeFullGridPointsLocally*$(\mathcal{I}, \mathbf{v}, \mathbf{l}, \mathbf{i})$;
        newGridPoints $\leftarrow$ *append*(newGridPoints, negativeGridPoints);
    if *newGridPoints is not empty* **then**
        $\mathcal{I} \leftarrow$ *addGridPoints*$(\mathcal{I}, \text{newGridPoints})$;
    else
        done $\leftarrow$ *True*
return $\mathbf{v}, \mathcal{I}$;

obtain the hierarchical coefficients for these grid points by learning the density on the extended grid. We repeat this process until we don't find any negative function value. This algorithm is, of course, just feasible if the number of local full grid points to be examined is small. Note that for a piecewise polynomial basis the algorithm doesn't work because the maximum of each ansatz function is not at a grid point.

3.4   Sparse Grid Collocation

The sparse grid collocation method (SGC) is based on the sparse grid discretization scheme (see Sect. 3.2) of the stochastic input space. The level-index vectors $(\mathbf{l}, \mathbf{i}) \in \mathcal{I}$ of some sparse grid with $|\mathcal{I}| = N$ define our set of collocation nodes as

$$\Xi_N := \{\boldsymbol{\xi}_{\mathbf{l},\mathbf{i}}\}_{(\mathbf{l},\mathbf{i}) \in \mathcal{I}} \tag{10}$$

with $\boldsymbol{\xi}_{\mathbf{l},\mathbf{i}}$ being the grid points, see Eq. (5). We evaluate u at every collocation node of Ξ_N and solve the interpolation problem

$$g_{\mathcal{I}}(\boldsymbol{\xi}_{\mathbf{l},\mathbf{i}}) := Q[u(\cdot, \boldsymbol{\xi}_{\mathbf{l},\mathbf{i}})], \forall \boldsymbol{\xi}_{\mathbf{l},\mathbf{i}} \in \Xi_N , \tag{11}$$

by a basis transformation from the nodal basis to the hierarchical basis. Efficient transformation algorithms for both linear and polynomial bases can be found in [3, 27]. We can furthermore employ adaptive refinement to consider local features in u. Suitable refinement criteria can be found in [8, 12, 13, 15, 28].

3.5   Moment Estimation

Let $\mu = \mathbb{E}_f(u)$ and $\sigma^2 = \mathbb{V}_f(u)$ be the unknown stochastic quantities of u for the true density f we want to estimate, $g_{\mathcal{I}}$ be a sparse grid surrogate model for u and therefore $\mu \approx \mathbb{E}_f(g_{\mathcal{I}})$ and $\sigma^2 \approx \mathbb{V}_f(g_{\mathcal{I}})$. Let $\hat{f}$ be an estimated density for the unknown probability density f obtained by a set of samples $\mathcal{D} := \{\boldsymbol{\xi}^{(k)}\}_{k=1}^n$ drawn from f.

To estimate the expectation value and the variance we need to solve integrals, which can be higher-dimensional, depending on the correlations of $\boldsymbol{\xi}$ and the density estimation technique we use. An easy method that can be applied to any estimated density we can sample or evaluate is vanilla Monte Carlo quadrature. We generate a new set of samples $\hat{\mathcal{D}} = \{\boldsymbol{\xi}^{(k)}\}_{k=1}^{\hat{n}}$ drawn from $\hat{f}$ with $\hat{n} \gg n$. We can now substitute $\hat{f}$ by the discrete density

$$\hat{f}_\delta(\xi) = \frac{1}{\hat{n}} \sum_{\boldsymbol{\xi}^{(k)} \in \hat{\mathcal{D}}} \delta(\boldsymbol{\xi} - \boldsymbol{\xi}^{(k)}) , \tag{12}$$

with δ being the Dirac delta function and estimate the expectation value as

$$\mathbb{E}_{\hat{f}}(g_I) = \int_\Omega g_I(\boldsymbol{\xi})\hat{f}(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} \approx \int_\Omega g_I(\boldsymbol{\xi})\hat{f}_\delta(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} = \frac{1}{\hat{n}} \sum_{\boldsymbol{\xi}^{(k)} \in \hat{\mathcal{D}}} g_I(\boldsymbol{\xi}^{(k)}) \,. \tag{13}$$

We obtain the result for the sample variance using the same approach for the numerically stable two-pass algorithm [4]

$$\mathbb{V}_{\hat{f}}(g_I) \approx \frac{1}{\hat{n}-1} \sum_{\boldsymbol{\xi}^{(k)} \in \hat{\mathcal{D}}} (g_I(\boldsymbol{\xi}^{(k)}) - \mathbb{E}_{\hat{f}}(g_I))^2 \,. \tag{14}$$

Due to the substitution of $\hat{f}$ by $\hat{f}_\delta$ we can solve the higher-dimensional integrals easily but we introduce a new numerical error, which converges slowly with respect to $\hat{n}$.

However, this substitution is not necessary if the estimated density $\hat{f}$ is a sparse grid function $\hat{f}_\mathcal{K}$. Due to the tensor-product approach we can decompose the higher-dimensional integral into one-dimensional ones and solve them separately without a numerical error larger than the machine precision ϵ. Let us additionally define some arbitrary order on the collocation points so that we can iterate over them in a predefined order. The expectation value of g_I with respect to $\hat{f}_\mathcal{K}$ can be computed as

$$\begin{aligned}
\mathbb{E}_{\hat{f}_\mathcal{K}}(g_I) &= \int_\Omega g_I(\boldsymbol{\xi})\hat{f}_\mathcal{K}(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} \\
&= \int_\Omega \sum_{(\mathbf{l},\mathbf{i})\in\mathcal{I}} v_{\mathbf{l},\mathbf{i}}\psi_{\mathbf{l},\mathbf{i}}^{(p)}(\boldsymbol{\xi}) \sum_{(\mathbf{k},\mathbf{j})\in\mathcal{K}} w_{\mathbf{k},\mathbf{j}}\varphi_{\mathbf{k},\mathbf{j}}^{(q)}(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} \\
&= \sum_{(\mathbf{l},\mathbf{i})\in\mathcal{I}} v_{\mathbf{l},\mathbf{i}} \sum_{(\mathbf{k},\mathbf{j})\in\mathcal{K}} w_{\mathbf{k},\mathbf{j}} \int_\Omega \psi_{\mathbf{l},\mathbf{i}}^{(p)}(\boldsymbol{\xi})\varphi_{\mathbf{k},\mathbf{j}}^{(q)}(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} \\
&= \sum_{(\mathbf{l},\mathbf{i})\in\mathcal{I}} v_{\mathbf{l},\mathbf{i}} \sum_{(\mathbf{k},\mathbf{j})\in\mathcal{K}} w_{\mathbf{k},\mathbf{j}} \underbrace{\int_{\Omega_1} \psi_{l_1,i_1}^{(p)}\varphi_{k_1,j_1}^{(q)}\mathrm{d}\boldsymbol{\xi}_1 \cdot \ldots \cdot \int_{\Omega_d} \psi_{l_d,i_d}^{(p)}\varphi_{k_d,j_d}^{(q)}\mathrm{d}\boldsymbol{\xi}_d}_{=:A_{(\mathbf{l},\mathbf{i}),(\mathbf{k},\mathbf{j})}} \\
&= \mathbf{v}^T A \mathbf{w} \,. \tag{15}
\end{aligned}$$

The same holds for the variance for which we use Steiners translation theorem

$$\mathbb{V}_{\hat{f}_\mathcal{K}}(g_I) = \mathbb{E}_{\hat{f}_\mathcal{K}}(g_I^2) - \mathbb{E}_{\hat{f}_\mathcal{K}}(g_I)^2 \tag{16}$$

and compute

$$
\begin{aligned}
\mathbb{E}_{\hat{f}_{\mathcal{K}}}(g_I^2) &= \int_{\Omega} g_I^2(\boldsymbol{\xi}) \hat{f}_{\mathcal{K}}(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} \\
&= \int_{\Omega} \left[\sum_{(\mathbf{l},\mathbf{i}) \in \mathcal{I}} v_{\mathbf{l},\mathbf{i}} \psi_{\mathbf{l},\mathbf{i}}^{(p)}(\boldsymbol{\xi}) \right] \left[\sum_{(\tilde{\mathbf{l}},\tilde{\mathbf{i}}) \in \mathcal{I}} v_{\tilde{\mathbf{l}},\tilde{\mathbf{i}}} \psi_{\tilde{\mathbf{l}},\tilde{\mathbf{i}}}^{(p)}(\boldsymbol{\xi}) \right] \hat{f}_{\mathcal{K}}(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} \\
&= \sum_{(\mathbf{l},\mathbf{i}) \in \mathcal{I}} v_{\mathbf{l},\mathbf{i}} \sum_{(\tilde{\mathbf{l}},\tilde{\mathbf{i}}) \in \mathcal{I}} v_{\tilde{\mathbf{l}},\tilde{\mathbf{i}}} \underbrace{\int_{\Omega} \psi_{\mathbf{l},\mathbf{i}}^{(p)}(\boldsymbol{\xi}) \psi_{\tilde{\mathbf{l}},\tilde{\mathbf{i}}}^{(p)}(\boldsymbol{\xi}) \hat{f}_{\mathcal{K}}(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}}_{=:B_{(\mathbf{l},\mathbf{i}),(\tilde{\mathbf{l}},\tilde{\mathbf{i}})}} \\
&= \mathbf{v}^T B \mathbf{v} ,
\end{aligned}
\tag{17}
$$

where the matrix entries $B_{(\mathbf{l},\mathbf{i}),(\tilde{\mathbf{l}},\tilde{\mathbf{i}})}$ are

$$
\begin{aligned}
B_{(\mathbf{l},\mathbf{i}),(\tilde{\mathbf{l}},\tilde{\mathbf{i}})} &= \sum_{(\mathbf{k},\mathbf{j}) \in \mathcal{K}} w_{\mathbf{k},\mathbf{j}} \int_{\Omega} \psi_{\mathbf{l},\mathbf{i}}^{(p)}(\boldsymbol{\xi}) \psi_{\tilde{\mathbf{l}},\tilde{\mathbf{i}}}^{(p)}(\boldsymbol{\xi}) \varphi_{\mathbf{k},\mathbf{j}}^{(q)}(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} \\
&= \sum_{(\mathbf{k},\mathbf{j}) \in \mathcal{K}} w_{\mathbf{k},\mathbf{j}} \underbrace{\int_{\Omega_1} \psi_{l_1,i_1}^{(p)} \psi_{\tilde{l}_1,\tilde{i}_1}^{(p)} \varphi_{k_1,j_1}^{(q)} \mathrm{d}\boldsymbol{\xi}_1 \cdot \ldots \cdot \int_{\Omega_d} \psi_{l_d,i_d}^{(p)} \psi_{\tilde{l}_d,\tilde{i}_d}^{(p)} \varphi_{k_d,j_d}^{(q)} \mathrm{d}\boldsymbol{\xi}_d}_{=:b_{(\mathbf{k},\mathbf{j})}} \\
&= \mathbf{w}^T \mathbf{b} .
\end{aligned}
\tag{18}
$$

The runtime $O(N \cdot M)$ for a naive implementation for the expectation value is determined by the matrix vector product. For the variance it holds $O(N^2 \cdot M)$ accordingly. In the inner loop we need to compute the scalar products of the one-dimensional basis functions, which can be done a-priori using Gauss-Legendre quadrature for the corresponding polynomial degree. However, we can reduce the quadratic dependency of the runtime on the number of sparse grid points to just a linear dependency by employing the UpDown-scheme. In the UpDown-scheme we exploit the tree-structure of the grid and apply the uni-directional principle to compute the inner-products [27]. This reduces the quadratic run time for the expectation value to be just linear in the number of grid points, i.e. $O(N + M)$.

3.6   The Sparse Grid Data-Driven UQ Forward Pipeline

Here we want to discuss the numerical properties of the sparse grid based data-driven UQ pipeline for forward problems (see Fig. 3). It consists of four steps: (1) The data set $\mathcal{D}$ is a randomly chosen set of n samples drawn from f. The quality of the set, how good it represents the moments of f, can vary significantly depending on it's size. Furthermore, it makes the estimated density $\hat{f}_{\mathcal{K}}$ (2) to be a random variable.
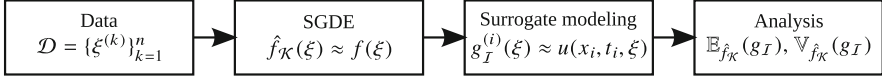
Data $\mathcal{D} = \{\xi^{(k)}\}_{k=1}^n$	SGDE $\hat{f}_{\mathcal{K}}(\xi) \approx f(\xi)$	Surrogate modeling $g_I^{(i)}(\xi) \approx u(x_i, t_i, \xi)$	Analysis $\mathbb{E}_{\hat{f}_{\mathcal{K}}}(g_I), \mathbb{V}_{\hat{f}_{\mathcal{K}}}(g_I)$

Fig. 3 Data-driven sparse grid UQ forward pipeline

Peherstorfer et al. [26] showed in this context consistency for the SGDE method, i.e.

$$P\left(\left\|f - \hat{f}_{\mathcal{K}}\right\|_{L_2}^2 = 0\right) = 1 \ , \tag{19}$$

for $|\mathcal{K}| \to \infty$ and $n \to \infty$. The accuracy of the sparse grid surrogate model g_I (3) depends on the smoothness of u. Bungartz and Griebel [3] showed that with a piecewise polynomial basis of degree p for regular sparse grids it holds

$$\|u - g_I\|_{L_2} \in O((2^{-\ell})^{p+1} \cdot \ell^{d-1}) \ . \tag{20}$$

For very specific adaptively refined sparse grids we refer to the results in [15] for a convergence proof. The estimated moments of u in the last step become random variables due to $\hat{f}_{\mathcal{K}}$ being a random variable. Therefore, when we talk about convergence of the moments of u we need to consider the mean accuracy with respect to density estimations based on different realizations of f. We define the mean relative error for the expectation value as

$$\mathbb{E}||(\mu - \mathbb{E}_{\hat{f}_{\mathcal{K}}}(g_I))/\mu|| \ , \tag{21}$$

and for the variance as

$$\mathbb{E}||(\sigma^2 - \mathbb{V}_{\hat{f}_{\mathcal{K}}}(g_I))/\sigma^2|| \ , \tag{22}$$

where μ and σ^2 are the true solutions. The quadrature step itself adds numerical errors in the order of the machine precision $O(\epsilon)$ per operation since it consists just of computing scalar products of one-dimensional polynomials.

4 Numerical Examples

In this section we discuss an analytic example, consisting of independent marginal Beta-distributions, and a three-dimensional subsurface flow problem with borehole data.

4.1 Analytic Example

In preparation of the subsurface problem we consider a two-dimensional analytic scenario with ξ_1, ξ_2 being two independent Beta-distributed random variables,

$$\xi_1 \sim \mathcal{B}(\alpha_1, \beta_1); \xi_2 \sim \mathcal{B}(\alpha_2, \beta_2) , \tag{23}$$

with shape parameters $\alpha_1 = 5$, $\beta_1 = 4$, $\alpha_2 = 3$ and $\beta_2 = 2$ and sample space $\Omega_1 = \Omega_2 = [0, 1]$. The corresponding density functions are defined as

$$f_k(\xi_k) = c_k \xi_k^{\alpha_k-1}(1 - \xi_k)^{\beta_k-1}, \ k \in \{1, 2\} , \tag{24}$$

where $c_k = \Gamma(\alpha_k + \beta_k)/(\Gamma(\alpha_k)\Gamma(\beta_k))$ with Γ being the gamma function. Let $f(\xi_1, \xi_2) = f_1(\xi_1)f_2(\xi_2)$ be the joint probability density function defined on $\Omega = \Omega_1 \times \Omega_2$. One realization $\mathcal{D}$ of f of size n was created in two steps: first, we generated n uniformly distributed samples in each dimension. Second, we applied the inverse cumulative distribution function of f_k and obtained samples from the beta space Ω. As a model function u we use a simple parabola with $u(\partial\Omega) = 0$ and $u(0.5, 0.5) = 1$

$$u(\boldsymbol{\xi}) = \prod_{k=1}^{2} 4\xi_k(1 - \xi_k) . \tag{25}$$

This model function has two main advantages: First, we can compute analytic solutions for the expectation value μ and the variance σ^2 of u as

$$\mu = \frac{c_1 c_2}{4725} \approx 0.71111 \tag{26}$$

$$\sigma_2 = \frac{c_1^3 c_2^3}{75{,}014{,}100{,}000} - \frac{2c_1^2 c_2^2}{22{,}325{,}625} + \frac{4c_1 c_2}{24{,}255} \approx 0.04843 , \tag{27}$$

Second, u can be approximated perfectly with a sparse grid function g_I of level 1 with a piecewise quadratic basis. This means that the numerical error in the surrogate model vanishes. The only two errors remaining are the approximation error estimating the density $\hat{f}$ and the quadrature error in moment estimation if we use Monte Carlo. The second one can be minimized by increasing the sample size; the first one, however, is limited to the amount of information there is about f, which is encoded in the data $\mathcal{D}$ we use for density estimation. Due to the randomness in the data we measure the error in the expectation value and the variance according to Eqs. (21) and (22). We compare these two errors for (1) vanilla Monte Carlo ($k = 200$ realizations of f), omitting the density estimation step, (2) a kernel-density estimator using the libagf library [18] ($k = 20$ realizations), (3) density trees (dtrees) [30] ($k = 50$ realizations) and (4) SGDE using the SG++ library [27] ($k = 20$ realizations). For the SGDE approach we used regular sparse grids with different levels, estimated the density according to [26] and chose the best approximation

Table 1 KL-divergence (KL) and cross entropy (L) for different density estimation methods and sizes for the training data sets $\mathcal{D}$. The test data set $\mathcal{T}$ to compute the measure had size $m = 10^4$

# samples	libagf		dtrees		SGDE	
	KL	L	KL	L	KL	L
50	0.2655	−0.7046	2.829	1.85	0.2157	−0.7491
75	0.2387	−0.7314	1.837	0.858	0.1838	−0.781
100	0.213	−0.7571	1.424	0.446	0.1533	−0.8115
500	0.1081	−0.8655	0.4294	−0.5488	0.1157	−0.849
1000	0.07851	−0.8951	0.2744	−0.7027	0.06948	−0.8953
5000	0.03964	−0.9356	0.1185	−0.8598	0.02778	−0.937
10,000	0.03001	−0.9352	0.09217	−0.8847	0.02014	−0.9446

with respect to the minimal cross entropy L for a test set $\mathcal{T}$ of size $m = 10^4$.

$$L_{\mathcal{T}}(\hat{f}) = -\frac{1}{m} \sum_{\xi^{(k)} \in \mathcal{T}} \log_2(\hat{f}(\xi^{(k)})) . \tag{28}$$

The test set $\mathcal{T}$ is generated analogous to $\mathcal{D}$ with a different seed. This measure is known to minimize the Kullback-Leibler-divergence (KL) and is therefore a suitable criterion [32].

Table 1 shows the KL-divergence and the cross entropy for the different density estimation methods and different sizes of training sets. We can see three main aspects: First, as expected, the cross entropy minimizes the KL-divergence. Second, the cross entropy decreases monotonically with the sample size for all density estimation methods. This means that the density estimation methods are able to capture the increasing information they get from the larger sample sets. Third, while SGDE and libagf have very similar results for all number of samples, the density trees have a poor performance especially for smaller sample sizes. The KL-divergence of the density trees is 10 times larger for 50–100 training samples than the ones of SGDE and libagf. This is a significant drawback for applications where the real costs lie in obtaining the samples. Think of boreholes that need to be drilled into the ground to obtain information about the physical domain of subsurface flow problems. If we look at the convergence of the expected error in the expectation value, see Fig. 4 (left), SGDE performs almost one order of magnitude better than libagf. Both methods converge with $n^{-1/2}$, which is basically the Monte Carlo convergence for the quadrature problem. The convergent phase for density trees starts later at a size of 1000 samples. We see basically the same picture if we look at the variance, see Fig. 4 (right). SGDE performs best compared to the other density estimation methods. However, it seems that the density estimation methods can not outperform Monte Carlo. The reason is that density estimation is based on Monte Carlo estimates for the moments of the distribution, see Eq. (9). It would pay off if there could be gained additional information from the data by extrapolation or regularization. There is no extrapolation in this case due to the definition of the
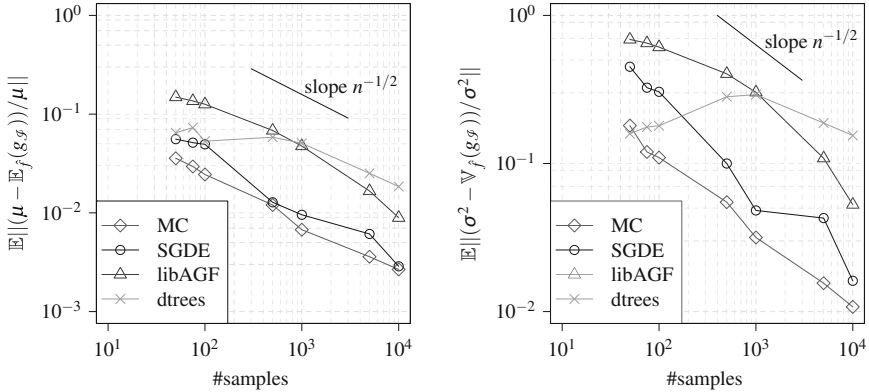
Fig. 4 Decay of the average error in the expectation value (*left*) and the variance (*right*) for Monte Carlo (MC), the sparse grid density estimation (SGDE) method, the kernel density estimator (libagf), and the density trees (dtrees)

joint probability density function f, which is zero at the boundaries. Neither does regularization, since there is no noise in the data and we measured the mean error over several realizations of f and made it therefore independent of single realizations where regularization could pay off. However, if the number of samples is limited, as it is in the CO_2 benchmark problem, the data-driven forward propagation approach with density estimation will reduce the error compared to Monte Carlo.

4.2 Multivariate Stochastic Application

In this application we simulate carbon sequestration based on the CO_2 benchmark model defined by [5]. We will not introduce here the modeling of this highly non-linear multiphase flow in a porous media problem but rather refer to [5] and focus on the stochastic part. The basic setting however is the following: We inject CO_2 through an injection well into a subterranean aquiferous reservoir. The CO_2 starts spreading according to the geological characteristics of the reservoir until it reaches a leaky well where the CO_2 rises up again to a shallower aquifer. A schematic view on the problem is shown in Fig. 5. The CO_2 leakage at the leaky well is the quantity we are interested in. It depends on the plume development in the aquifer and the pressure built up in the aquifer due to injection. While we can control the injection pressure, we cannot control the geological properties of the reservoir like porosity, permeability, etc. This is where the uncertainty comes in and for which we use data to describe it.
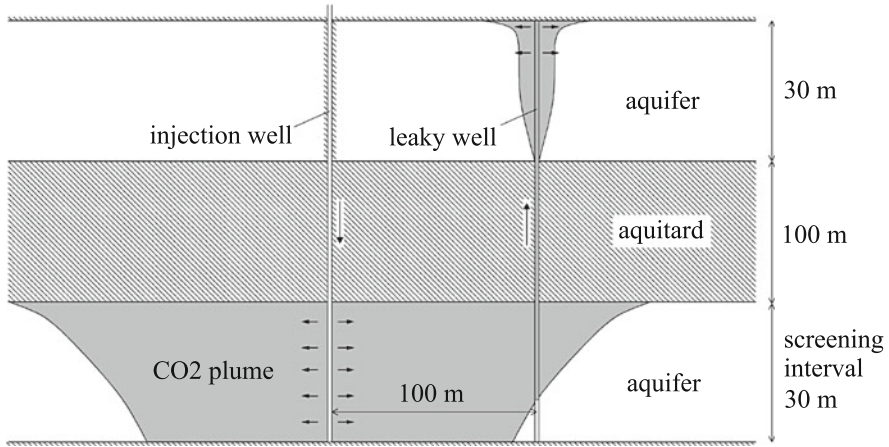
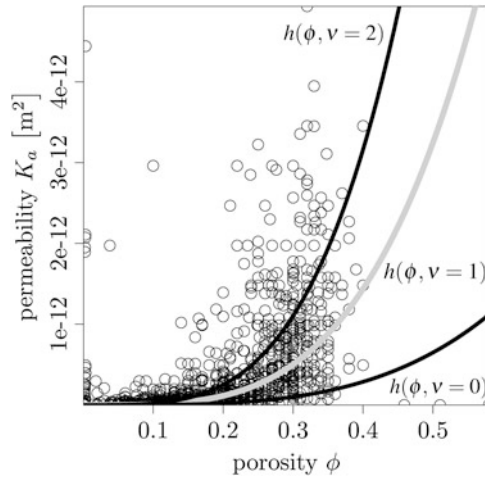**Fig. 5** Cross-section through the subterranean reservoir [22]

Fig. 6 Raw data for porosity ϕ and permeability K_a from the U.S. National Petroleum Council Public Database including 1270 data points. The *curves* show the upper and the lower bound of the transformed and truncated sample space with respect to the variation v

4.2.1 Stochastic Formulation

The CO_2 benchmark model has three stochastic parameters with respect to the geological properties of the reservoir and the leaky well: (1) the reservoir porosity ϕ, (2) the reservoir permeability K_a, and (3) the leaky well permeability K_L. For the description of the reservoir, i.e. K_a and ϕ, we use a raw data set from the U.S. National Petroleum Council Public Database including 1270 reservoirs, shown in Fig. 6, see also [14]. The data set is assumed to be one realization of the unknown

Table 2 Marginal densities and ranges used in the *analytic* approach to describe the uncertainty of the CO_2 benchmark model. They are fitted to the decorrelated samples according to [22]

Uncertain parameter	Probability density function	Range Ω_k
Porosity ϕ	$f_1(\phi) = \frac{1}{\phi\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(\phi)-\mu)^2}{2\sigma^2}\right)$ with $\mu = -27.6310, \sigma = 0.3579$	$[0.0896, 0.2511]$
Permeability K_a	$f_2(K_L) = \frac{1}{K_L\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(K_L)-\mu)^2}{2\sigma^2}\right)$ with $\mu = -1.8971, \sigma = 0.2$	$[3.88, 25.8] \times 10^{-13}$
Variation ν	$f_3(\nu) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \nu^{\alpha-1}(1-\nu)^{\beta-1}$ with $\alpha = 3, \beta = 3$	$[0, 2]$

probability density function f that defines the uncertainty of the problem. For the leaky well permeability there is no data available, which makes it necessary to make further assumptions. We make here the same assumptions as in Oladyshkin et al. in [22], see permeability in Table 2. Furthermore, Oladyshkin et al. presented in the same paper an integrative approach to quantify the stochastic outcome of the CO_2 benchmark model using analytic densities based on the data at hand, see Table 2. They defined a sample space that is different to the parameter space of the simulation. They substituted $K_a \in \Omega_{K_a}$ by a variation parameter $\nu \in \Omega_\nu$ and encoded the correlation in a transformation function from the new sample space to the parameter space, i.e.

$$
\begin{aligned}
h\colon \Omega_\phi \times \Omega_\nu &\to \Omega_{K_a} \\
\phi, \nu &\mapsto c_1 \phi^{c_2} \left[1 + c_3 \nu\right] ,
\end{aligned}
\tag{29}
$$

with parameters $c_1 = 4.0929 \times 10^{-11}, c_2 = 3.6555, c_3 = -2$. In the new sample space $\Omega := \Omega_\phi \times \Omega_\nu \times \Omega_{K_L}$ all variables are decorrelated. Therefore we can define analytic independent marginal densities and can propagate the resulting uncertainty directly through the model function using polynomial chaos expansion, for example. We refer to this approach as the *analytic* approach. To apply stochastic collocation to this problem, we truncated the infinite sample space Ω such that in directions of ϕ and K_a we collect 99.99 % of their mass around the corresponding mean, see Table 2. We denote the samples that lie within this truncated space as $\mathcal{D} := \{(K_a^{(k)}, \phi^{(k)})\}_{k=1}^{413}$, see Fig. 7. We use $\mathcal{D}$ for density estimation to obtain objective measures of the input's uncertainty.

4.2.2 Results

In this section we want to illustrate the ability of the integrated sparse grid approach to predict accurate expectation values and variances for the CO_2 benchmark model with the given input data set. We assume that a surrogate model using adaptive sparse grid collocation is available and approximates the unknown model function so well that the expectation value and the variance can be estimated accurately with

Fig. 7 The plot shows the transformed raw data points $h^{-1}(\phi, K_a)$ from Fig. 6 that lie within the parameter ranges given in Table 2. They are additionally linearly scaled to $[0, 1]^2$. The remaining number of samples is 413
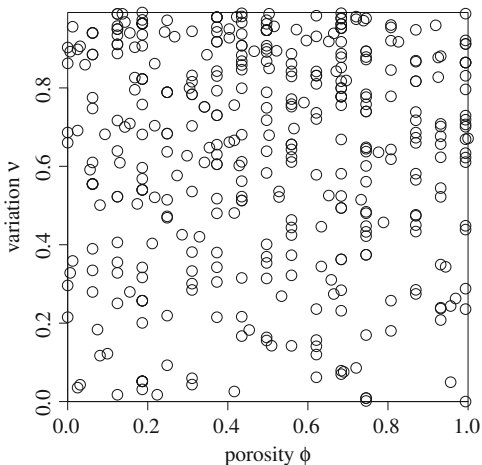

Table 3 Mean cross entropy (L) for different density estimation methods and for 10 randomly chosen training data sets of size 363 and test data sets of size 50

# training	# test	analytic (L)	libagf (L)	dtrees (L)	SGDE (L)
363	50	0.7279	0.00278	−0.1314	−0.3042

respect to the analytic approach ($N = 114$, refinement according to the *variance surplus refinement*, see [8] for details). We compare these results of the (1) analytic approach with (2) SGDE, (3) kernel-densities, (4) density trees, and (5) Monte Carlo with bootstrapping on the available data $\mathcal{D}$.

The results of the density estimations with respect to the cross entropy for a test set $\mathcal{T}$ with $m = 50$ are listed in Table 3. The analytic approach has by far the largest cross entropy, which suggests that it doesn't capture the underlying data as good as the others do. The SGDE method performs best as it has the lowest cross entropy. Note that some samples we use for training are located close to the boundary of the domain (see Fig. 7). This affects the SGDE method since we need to consider the boundary in the discretization of the sample space. For this problem we used trapezoidal boundary sparse grids [27] where each inner grid point has two ancestors in each dimension that lie on the corresponding boundary of the domain.

If we look at the results of the estimation of the expectation value, see Fig. 8 (left), and the variance, see Fig. 8 (right), we obtain surprising results. Let us assume that the results of the Monte Carlo quadrature approach using bootstrapping on the available data is our ground truth. By this we say indirectly that we have large confidence in the available data. Compared to this ground truth, the analytic approach overestimates the expectation value and underestimates the variance significantly. We call this difference the "subjective gap", which has been introduced by expert knowledge. The other density estimation methods lie in between these approaches. The density trees match the expectation value of the data almost exactly. The SGDE method overestimates the expectation value slightly, the kernel density
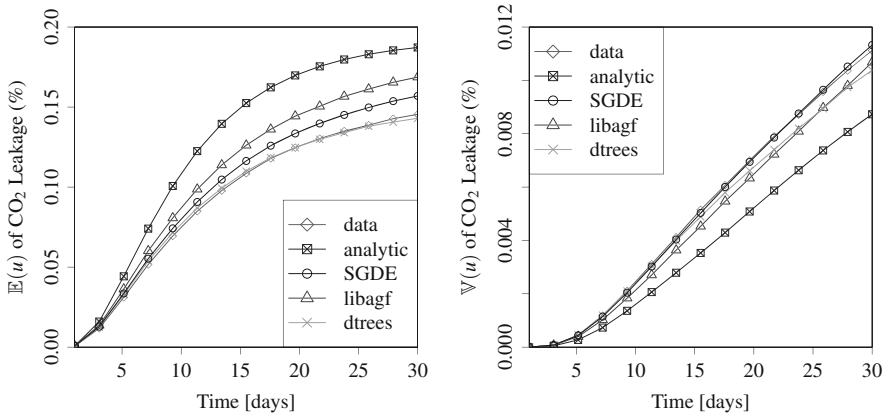
Fig. 8 Expectation value (*left*) and variance (*right*) over simulation time for the bootstrapping method with the raw data (data), the analytic approach as in [22] (analytic), the sparse grid density estimation (SGDE) method, the kernel density estimator (libagf), and the density trees (dtrees)

does as well. But if we look at the variance then the SGDE method gives the best results compared to the data, while all others underestimate it.

The overestimation of the expectation value for SGDE with respect to the data can be explained by extrapolation: We used a sparse grid with trapezoidal boundary for SGDE because some samples are located close to the boundary of the domain. Furthermore, we impose smoothness on the unknown density. These two facts let the SGDE method extrapolate towards the boundary resulting in a larger density than there is in the data. This is not even wrong since the boundaries of our transformed and truncated domain Ω are located in the middle of the parametric domain in which the raw data lies, see Fig. 6 (left). This leads then to the higher expectation values we see in Fig. 8 (left). In fact, if we use a sparse grid without boundary points for density estimation we match the expectation value of the data as well as the density trees do. However, the cross entropy for this sparse grid density is larger compared to the others ($L = 0.2632$) indicating a worse estimation. And indeed, with this estimation we overestimate now the variance significantly.

Due to these arguments, we question the ground truth, which in this application was based on a very limited data set. Of course, this makes the comparison of different methods difficult. But, since the balancing between fidelity in the data and smoothness of the density function is defined clearly for the SGDE approach, we consider it a reliable and robust approach in the context of UQ.

5 Conclusions

In this paper we presented a new integrated sparse grid approach for data-driven uncertainty quantification forward problems. It has two main advantages over common approaches for such problems: First, it is an unsupervised approach that relies on the data at hand. It is not influenced by expert knowledge. Second, the integrated sparse grid approach allows a seamless interaction with the stochastic collocation method with respect to adaptive refinement and quadrature.

Furthermore, the numerical experiments showed that the SGDE method gives good approximations of the unknown probability density function already for small sample sets. It did better than the very popular kernel density. Even newer approaches such as density trees showed worse results compared to SGDE.

However, the SGDE method has drawbacks when it comes to statistical applications for which we need to assure unit integrand and positivity. We presented one way to overcome these problems by suitable discretization and using appropriate regularization. This approach however is limited in terms of the problem's dimensionality since it implies a local full grid search. There is ongoing work in this field, see [10], to overcome these problems.

In this paper we focused on small sample sets. When it comes to large data sets one can speed up this process significantly using fast algorithms such as SGDE. For example, when one uses Markov chain Monte Carlo to obtain a discrete posterior density in an inverse UQ setting. Such densities are often correlated and pose problems to established forward propagation methods such as the generalized polynomial chaos expansion. The Rosenblatt transformation [31] and the inverse Rosenblatt transformation can play an important role in this context. They can be computed very efficiently without additional numeric errors using sparse grids.

References

1. A. Alwan, N. Aluru, Improved statistical models for limited datasets in uncertainty quantification using stochastic collocation. J. Comput. Phys. **255**, 521–539 (2013)
2. I. Babuška, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data. SIAM J. Numer. Anal. **45**(3), 1005–1034 (2007)
3. H.-J. Bungartz, M. Griebel, Sparse grids. Acta Numer. **13**, 147–269 (2004)
4. T.F. Chan, G.H. Golub, R.J. LeVeque, Algorithms for computing the sample variance: analysis and recommendations. Am. Stat. **37**(3), 242–247 (1983)
5. H. Class, A. Ebigbo, R. Helmig, H. Dahle, J. Nordbotten, M. Celia, P. Audigane, M. Darcis, J. Ennis-King, Y. Fan, B. Flemisch, S. Gasda, M. Jin, S. Krug, D. Labregere, A. Naderi Beni, R. Pawar, A. Sbai, S. Thomas, L. Trenty, L. Wei, A benchmark study on problems related to CO_2 storage in geologic formations. Comput. Geosci. **13**(4), 409–434 (2009)

6. M. Eldred, J. Burkardt, Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification, in *Aerospace Sciences Meetings* (American Institute of Aeronautics and Astronautics, Reston, 2009)

7. H.C. Elman, C.W. Miller, Stochastic collocation with kernel density estimation. Comput. Methods Appl. Mech. Eng. **245–246**(0), 36–46 (2012)

8. F. Franzelin, P. Diehl, D. Pflüger, Non-intrusive uncertainty quantification with sparse grids for multivariate peridynamic simulations, in *Meshfree Methods for Partial Differential Equations VII*, ed. by M. Griebel, M.A. Schweitzer. Volume 100 of Lecture Notes in Computational Science and Engineering (Springer, Cham, 2015), pp. 115–143

9. J. Garcke, Maschinelles Lernen durch Funktionsrekonstruktion mit verallgemeinerten dünnen Gittern, Ph.D. thesis, Universität Bonn, Institut für Numerische Simulation, 2004

10. M. Griebel, M. Hegland. A finite element method for density estimation with gaussian process priors. SIAM J. Numer. Anal. **47**(6), 4759–4792 (2010)

11. M. Hegland, G. Hooker, S. Roberts, Finite element thin plate splines in density estimation. ANZIAM J. **42**, 712–734 (2000)

12. J. Jakeman, S. Roberts, Local and dimension adaptive stochastic collocation for uncertainty quantification, in *Sparse Grids and Applications*, ed. by J. Garcke, M. Griebel. Volume 88 of Lecture Notes in Computational Science and Engineering (Springer, Berlin/Heidelberg, 2013), pp. 181–203

13. J.D. Jakeman, R. Archibald, D. Xiu, Characterization of discontinuities in high-dimensional stochastic problems on adaptive sparse grids. J. Comput. Phys. **230**(10), 3977–3997 (2011)

14. A. Kopp, H. Class, H. Helmig, Investigations on CO_2 storage capacity in saline aquifers – part 1: dimensional analysis of flow processes and reservoir characteristics. Int. J. Greenh. Gas Control **3**, 263–276 (2009)

15. X. Ma, N. Zabaras, An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. J. Comput. Phys. **228**(8), 3084–3113 (2009)

16. X. Ma, N. Zabaras, An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations. J. Comput. Phys. **229**(10), 3884–3915 (2010)

17. O.P.L. Maître, O.M. Knio, *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Scientific Computation (Springer, Dordrecht/New York, 2010)

18. P. Mills, Efficient statistical classification of satellite measurements. Int. J. Remote Sens. **32**(21), 6109–6132 (2011)

19. M. Navarro, J. Witteveen, J. Blom, Polynomial chaos expansion for general multivariate distributions with correlated variables, Technical report, Centrum Wiskunde & Informatica, June 2014

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

21. S. Oladyshkin, H. Class, R. Helmig, W. Nowak, A concept for data-driven uncertainty quantification and its application to carbon dioxide storage in geological formations. Adv. Water Resour. **34**(11), 1508–1518 (2011)

22. S. Oladyshkin, H. Class, R. Helmig, W. Nowak, An integrative approach to robust design and probabilistic risk assessment for CO_2 storage in geological formations. Comput. Geosci. **15**(3), 565–577 (2011)

23. S. Oladyshkin, W. Nowak, Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. Reliab. Eng. Syst. Saf. **106**(0), 179–190 (2012)

24. B. Peherstorfer, Model order reduction of parametrized systems with sparse grid learning techniques. Ph.D. thesis, Technical University of Munich, Aug 2013

25. B. Peherstorfer, C. Kowitz, D. Pflüger, H.-J. Bungartz, Selected recent applications of sparse grids. Numer. Math. Theory Methods Appl. **8**(01), 47–77, Feb 2015

26. B. Peherstorfer, D. Pflüger, H.-J. Bungartz, Density estimation with adaptive sparse grids for large data sets, in *Proceedings of the 2014 SIAM International Conference on Data Mining*, Philadelphia, 2014, pp. 443–451

27. D. Pflüger, *Spatially Adaptive Sparse Grids for High-Dimensional Problems* (Verlag Dr. Hut, Munich, 2010)
28. D. Pflüger, Spatially adaptive refinement, in *Sparse Grids and Applications*, ed. by J. Garcke, M. Griebel. Lecture Notes in Computational Science and Engineering (Springer, Berlin/Heidelberg, 2012), pp. 243–262
29. D. Pflüger, B. Peherstorfer, H.-J. Bungartz, Spatially adaptive sparse grids for high-dimensional data-driven problems. J. Complex. **26**(5), 508–522 (2010). Published online Apr 2010
30. P. Ram, A.G. Gray, Density estimation trees, in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, 21–24 Aug 2011, pp. 627–635
31. M. Rosenblatt, Remarks on a multivariate transformation. Ann. Math. Stat. **23**(1952), 470–472 (1952)
32. B. Silverman, *Density Estimation for Statistics and Data Analysis*, 1st edn. (Chapman & Hall, London/New York, 1986)
33. W. Walker, P. Harremoes, J. Rotmans, J. van der Sluijs, M. van Asselt, P. Janssen, M.K. von Krauss, Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integr. Assess. **4**(1), 5–17 (2005)
34. X. Wan, E. Karniadakis, Multi-element generalized polynomial chaos for arbitrary probability measures. SIAM J. Sci. Comput. **28**(3), 901–928 (2006)
35. M. Xiang, N. Zabaras, An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method. Inverse Probl. **25**(3), 035013 (2009)
36. D. Xiu, J.S. Hesthaven, High-order collocation methods for differential equations with random inputs. SIAM J. Sci. Comput. **27**(3), 1118–1139 (2005)
37. D. Xiu, G. Karniadakis, The Wiener–askey polynomial chaos for stochastic differential equations. SIAM J. Sci. Comput. **24**(2), 619–644 (2002)
38. D. Xiu, G.E. Karniadakis, Modeling uncertainty in flow simulations via generalized polynomial chaos. J. Comput. Phys. **187**(1), 137–167 (2003)
39. G. Zhang, D. Lu, M. Ye, M.D. Gunzburger, C.G. Webster, An adaptive sparse-grid high-order stochastic collocation method for Bayesian inference in groundwater reactive transport modeling. Water Resour. Res. **49**(10), 6871–6892 (2013)