Flow, Turbulence and Combustion (2024) 112:191–216
https://doi.org/10.1007/s10494-023-00467-6

RESEARCH

CFD Uncertainty Quantification using PCE–HDMR: Exemplary Application to a Buoyancy-Driven Mixing Process

Philipp J. Wenig[1] · Stephan Kelm[2] · Markus Klein[1]

Received: 1 May 2023 / Accepted: 27 July 2023 / Published online: 18 August 2023

Abstract
For the investigation of uncertainties in high dimensional spaces of computationally expensive engineering applications, reliable Uncertainty Quantification (UQ) methods are needed. These methods should provide accurate and efficient High-Dimensional Model Representations of stochastic results using a reasonable number of calculations. Therefore, the PCE–HDMR approach (Polynomial Chaos Expansion–High-Dimensional Model Representation) is utilized to qualify appropriate UQ methods for large-scale computations in the field of Computational Fluid Dynamics. This technique is a combination of Cut-HDMR, a hierarchical decomposition modeling approach, with PCE. To demonstrate its effectiveness, the PCE–HDMR methodology in conjunction with complementary modeling techniques is applied for the UQ analysis of a buoyancy-driven mixing process between two miscible fluids within the Differentially Heated Cavity of aspect ratio 4. The results include a thorough probabilistic representation of time-dependent response quantities that comprehensively describe the mixing process. The stochastic models are derived from Large Eddy Simulations using PCE–HDMR and the Sparse Grid Method, which serves as a reference for the results from PCE–HDMR. The results show that PCE–HDMR provides accurate statistics of the modeled time-dependent stochastic processes and shows good agreement with the reference results. Thus, PCE–HDMR indicates great potential for UQ of technical-scale computations due to its efficiency and flexibility in the construction of stochastic models.

Keywords Uncertainty Quantification · PCE–HDMR · Buoyancy-driven transient mixing process · Error estimation · Variance-based decomposition

✉ Philipp J. Wenig
philipp.wenig@unibw.de

Stephan Kelm
s.kelm@fz-juelich.de

Markus Klein
markus.klein@unibw.de

[1] Institute of Applied Mathematics and Scientific Computing, University of the Bundeswehr Munich, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany

[2] Institute of Energy and Climate Research, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

1 Introduction

For the application of Uncertainty Quantification (UQ) to complex engineering applications with high-dimensional and computationally expensive character, it is important to employ reliable and efficient methods for uncertainty analysis. Therefore, as a basis for the application to a technical scale experiment (Freitag and Schmidt 2022), the methodology is established using a generic test case from literature. This test case reflects similar physics to those anticipated for the application case. The utilization of a simplified configuration in the initial application of existing methods provides an opportunity to assess and enhance these methods at a reasonable computational cost. As a test case, the Differentially Heated Cavity (DHC) with aspect ratio 4 (Trias et al. 2007, 2010a, b) was chosen and extended to a superimposed mixing process. It is a tall cavity with hot left wall and cold right wall, in which natural convection flow and buoyancy-induced mixing processes occur in the presence of two gas mixtures with different densities. In the present work these transient mixing processes are modeled with Large Eddy Simulation (LES). The definition of uncertain input parameters within the Computational Fluid Dynamics (CFD) simulation leads to uncertainty in results or responses, which need to be quantified with UQ methods.

In the following, a brief summary of related UQ studies, which are reported in the literature, is provided. The investigation of single-phase mixing in a flow channel has been explored in several research studies (Badillo and Kapulla 2013; Fokken et al. 2019; Cutrono et al. 2018, 2019, 2020). The CFD simulation involved the turbulent mixing of two parallel streams with different densities that were initially separated by a splitter plate. Polynomial Chaos Expansion (PCE) was applied to analyze random fields and the stochastic results were compared to experimental data. Le Maître et al. (2002, 2005) investigated the thermo-fluid flow in the DHC by solving low-Mach-number equations in both Boussinesq and non-Boussinesq limits. They studied uncertainties in the mean velocity field caused by the uncertain cold wall temperature using PCE. In Le Maître et al. (2004), the use of PCE was employed to examine the effects of uncertainty in the temperature of a heated bottom wall on Rayleigh–Bénard convection (RBC). In the field of CFD for aerospace engineering, Huan et al. (2019) utilized Karhunen-Loève Expansion (KLE) to investigate uncertain spatially dependent fields. The objective of this research was to contribute to the development of scramjet engines, with a focus on ensuring efficient and stable propulsion during hypersonic flight conditions. An advanced UQ methodology, which involves the combination of PCE and KLE, was applied for the analysis of uncertainties of a turbulent round jet by Jivani et al. (2021). Furthermore, different methods have been presented in literature to facilitate stochastic modeling of intricate random fields or stochastic processes through the change of variables and realignment techniques. Liao and Zhang (2016) applied a change of variables to approximate time-dependent 1-D and 2-D solute transport problems. First, stochastic models for time were built as a function of the response quantity. The inverse transform of model samples yielded the response as a function of time. Giraldi et al. (2017) modeled the sea surface anomaly, which takes place after an earthquake. The arrival time of the wave anomaly was approximated with PCE and the time-dependent wave height evolution was modeled with a second PCE model whose start time corresponds to the arrival time. Subsequently, both models were combined for stochastic predictions. Colombo et al. (2018) utilized two-step surrogate models to approximate porosity in multi-layered sedimentary basins as a function of depth. They applied a change of the coordinate system to align discontinuities of the target function and transformed it back via piecewise linear mapping. Bonnaire et al. (2021) employed linear time scaling to

phase and then approximate oscillatory responses in the Kármán vortex street problem. Subsequently, inverse time scaling is used to restore the original oscillatory response from model predictions. To further increase the efficiency of available modeling approaches, one promising metamodeling approach was developed by Yue et al. (2021) and involves the approximation of a response function with PCE–HDMR (Polynomial Chaos Expansion–High-Dimensional Model Representation). The methodology has been successfully tested with functions in analytic form and several engineering examples.

The present work aims for the description of stochastic processes with accurate and efficient HDMR models that can represent stochastic results effectively while minimizing the number of required computations. Therefore, the PCE–HDMR approach, which combines Cut-HDMR with PCE, is adopted and applied to an intricate dynamic CFD system in conjunction with complementary modeling strategies for the first time and continues the work of Wenig et al. (2021a, b, 2022, 2023) in the field of uncertainty quantification for buoyancy-driven mixing processes between two miscible fluids. In Wenig et al. (2021a), preliminary work including a mesh convergence study, global sensitivity analysis and UQ of scalar integral quantities, which provide a plain description of the mixing process, are presented. Stochastic models for the scalar integral quantities from different CFD models with different fidelities such as URANS, LES and DNS were combined through multifidelity modeling in Wenig et al. (2021b). Subsequently, in Wenig et al. (2023) stochastic modeling was extended to the approximation of random fields and the two-step model approach called Stochastic Model Composition (SMC) was proposed for the approximation of highly nonlinear time-dependent stochastic processes. Using the SMC approach, stochastic models of the transient responses were built that describe their chronological progress as a function of the mixing state. This state-dependent perspective aligned strongly non-linear phenomena and reduced the modeling complexity. By means of an additional stochastic model, which describes the relationship between the mixing state and time, time-dependent model realizations could be derived. The basic idea behind the SMC approach is similar to realignment techniques pursued in Liao and Zhang (2016), Giraldi et al. (2017), Colombo et al. (2018) and Bonnaire et al. (2021). Building on the methods already implemented, first results from stochastic modeling with the PCE–HDMR approach were presented in Wenig et al. (2022).

The PCE–HDMR approach basically involves the construction of high-dimensional stochastic models through a number of low-dimensional submodels, which are built by PCE. The obtained results provide a probabilistic representation of time-dependent response quantities that comprehensively describe the mixing process. Furthermore, the variance of responses was decomposed into fractions, which can be attributed to individual inputs. This was achieved through the determination of total-order Sobol indices. Error estimates were computed for the assessment of the stochastic model accuracy. The stochastic models are derived using PCE–HDMR and the Sparse Grid Method (SGM), which serves as a reference to evaluate the accuracy of PCE–HDMR. The UQ results show that PCE–HDMR provides accurate statistics of the modeled time-dependent stochastic processes and it is in good agreement with the SGM reference. Therefore, PCE–HDMR has the potential to be an efficient method for performing UQ analysis of large-scale applications, which is the long-term goal of this research.

The paper is organized as follows. Section 2 details the methods, which were used for the prediction and stochastic representation of the underlying flow phenomena. The obtained results are presented and discussed in Sect. 3. In Sect. 4, the results are summarized and conclusions are drawn.

2 Methods

Section 2.1 outlines the mathematical models and numerical methods used to predict the fluid flow dynamics. A comprehensive description of the UQ techniques is detailed in Sect. 2.2.

2.1 Computational Fluid Dynamics

The governing equations are presented in Sect. 2.1.1. Subsequently, Sect. 2.1.2 details the case setup and the definition of uncertainties, while Sect. 2.1.3 provides information on the numerical framework and the discretization technique.

2.1.1 Governing Equations

The study at hand pertains to the analysis of a low Mach number flow involving two Newtonian viscous fluids. The LES governing equations for continuity, momentum, energy, and species transport read as follows:

$$\begin{aligned}
&\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \\
&\frac{\partial \rho \mathbf{u}}{\partial t} + \nabla \cdot (\rho \mathbf{u}\mathbf{u}) = -\nabla p + \rho \mathbf{g} + \nabla \cdot \left(2\mu_{eff}\, \tau(\mathbf{u})\right) - \nabla\Big(\frac{2}{3}\mu_{eff}(\nabla \cdot \mathbf{u})\Big), \\
&\frac{\partial \rho h}{\partial t} + \frac{\partial \rho K}{\partial t} + \nabla \cdot (\rho \mathbf{u} h) + \nabla \cdot (\rho \mathbf{u} K) - \frac{\partial p}{\partial t} = \nabla \cdot \left(\rho \alpha_{eff} (\nabla h)\right) + \rho \mathbf{u} \cdot \mathbf{g} + S_D, \\
&\frac{\partial \rho Y_i}{\partial t} + \nabla \cdot \left(\rho \mathbf{u} Y_i\right) = \nabla \cdot \left(\rho D_{eff} \left(\nabla Y_i\right)\right).
\end{aligned} \tag{1}$$

For the sake of simplicity, the LES filtering operations were omitted. $\mathbf{u}$ is the velocity vector field, ρ is the density field, p is the static pressure field, $\mathbf{g} = (0, g, 0)$ is the gravitational acceleration vector, h is the enthalpy, $K = \frac{1}{2}|\mathbf{u}|^2$ is the kinetic energy of the system, Y_i is the mass fraction of the ith species from the set of gas species indices given by $N = \{1, 2\}$, and the rate of strain tensor is defined as $\tau(\mathbf{u}) = \frac{1}{2}\left(\nabla \mathbf{u} + (\nabla \mathbf{u})^T\right)$. S_D accounts for the enthalpy transport due to diffusive mass transport and the associated correction of the heat conduction (Wenig et al. 2023). The effective dynamic viscosity μ_{eff} is the sum of the molecular and subgrid-scale viscosity. h is the sum of the internal energy per unit mass e and the kinematic pressure $h = e + \frac{p}{\rho}$. According to the gradient flux approach, the effective thermal diffusivity results from $\alpha_{eff} = \frac{\mu}{\rho \cdot Pr} + \frac{\nu_{sgs}}{Pr_t}$ with the kinematic subgrid-scale viscosity ν_{sgs} and the turbulent Prandtl number $Pr_t = 0.85$. The effective molecular diffusivity results from $D_{eff} = D + \frac{\nu_{sgs}}{Sc_t}$ with the turbulent Schmidt number $Sc_t = 0.85$. The molecular diffusivity D is assumed to be constant. Mixture properties Φ_m are computed from the individual species properties Φ_i and species mass fractions Y_i.

2.1.2 Case Setup and Uncertainties

The investigations are conducted using the DHC with aspect ratio 4, which is filled with air and 40 vol% of helium in the upper third, as shown in Fig. 1. The DHC is characterized by a hot left wall and a cold right wall. The resulting temperature difference between both

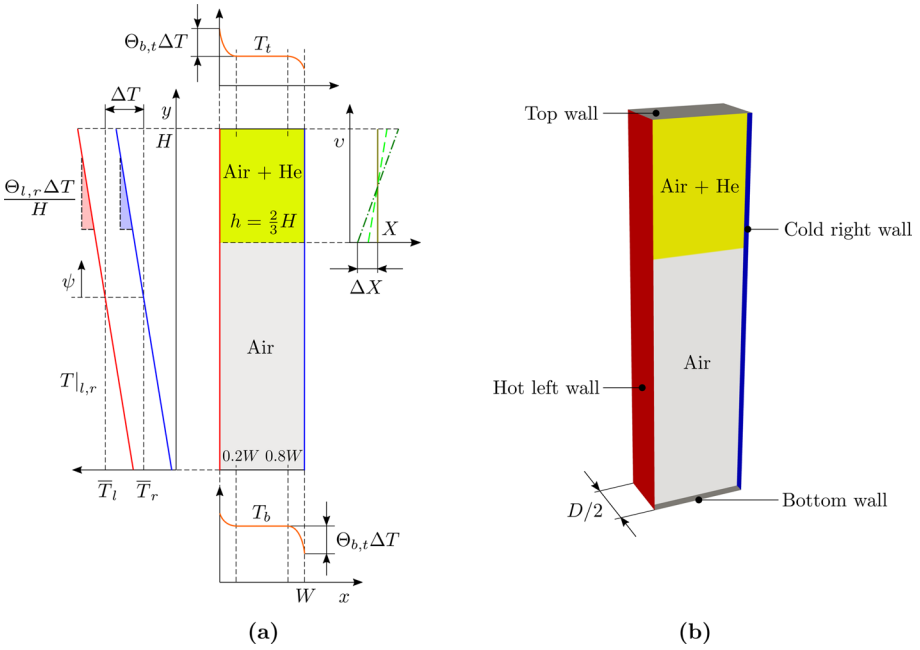

Fig. 1 Case setup of the DHC with superimposed mixing process: **a** Schematic sketch of the Differentially Heated Cavity with illustration of uncertain parameters, such as the wall temperature difference, the wall-tangential temperature gradient, the top and bottom wall temperature and the initial helium stratification, **b** 3D representation of the DHC in full section view

walls generates a natural convection flow within the cavity. The geometrical dimensions of the DHC are the height H, width W, and depth D. The height-to-width aspect ratio and the width-to-depth aspect ratio are $\varphi_{HW} = H/W = 4$ and $\varphi_{WD} = W/D = 1$, respectively. The Prandtl numbers $Pr = 0.71$ and $Pr = 0.66$ used correspond to air and helium, respectively. The material properties of air are used to define the Rayleigh number $Ra = \frac{g\beta\Delta T H^3 \rho}{\mu\alpha} = 2 \times 10^9$. The non-slip boundary condition is imposed on the velocity at the four enclosing walls. The cavity is subject to a temperature difference $\Delta T = \overline{T}_l - \overline{T}_r$. The nominal value of the wall temperature difference is $\Delta T^* = 21.431\,\mathrm{K}$. Thermal radiation is neglected. The third spatial dimension is taken into account due to the three-dimensional character of turbulent flow. Hence, the flow field is assumed to be periodic in the z-direction. For the mass fraction, a zero-gradient at the enclosing walls is defined: $\left(\nabla Y \cdot \vec{n}\right)\Big|_{\partial\Omega} = 0$, where $\vec{n}$ denotes the wall-normal unit vector. A zero gradient is defined for the kinematic subgrid-scale viscosity ν_{sgs} and the subgrid-scale thermal diffusivity α_{sgs}: $\left(\nabla \nu_{sgs} \cdot \vec{n}\right)\Big|_{\partial\Omega} = 0$, $\left(\nabla \alpha_{sgs} \cdot \vec{n}\right)\Big|_{\partial\Omega} = 0$. The Wall-Adapting Local Eddy-viscosity (WALE) model with $C_w = 0.5$ is applied to modelling viscous subgrid-scale effects (Nicoud and Ducros 1999). The initial conditions of the system are the temperature of $T_0 = \overline{T} = 298.15\,\mathrm{K}$ and pressure of $p_0 = 1\,\mathrm{bar}$. The molecular diffusion coefficient is assumed to be constant and is derived using Fuller's method (Kabelac et al. 2013), with a nominal value of $D^* = 6.904 \times 10^{-5}\,\mathrm{m^2/s}$. The other properties of the system are calculated using the ideal gas law at $\overline{T} = 298.15\,\mathrm{K}$.

The aim of the study was to investigate the propagation of uncertainties in the initial and boundary conditions, as well as material properties. Several uncertain input parameters were defined and are treated as random variables. These were chosen to be representative of possible uncertainties in the application case (Freitag and Schmidt 2022) for which the UQ methods are developed. Figure 1a provides an illustration of these uncertain input parameters. The thermal boundary conditions are subject to uncertainties that include the temperature difference between the left and right wall ΔT , the vertical temperature gradient at the left and right wall defined by $\Theta_{l,r}$ and the temperatures at the top and bottom wall of the cavity, which are determined by the uncertain parameter $\Theta_{b,t}$. Additionally, uncertainty is also imposed on the initial distribution of the helium mole fraction X through the uncertain mole fraction difference ΔX. The molecular diffusion coefficient D is also considered as an uncertain parameter.

The average temperature at the left and right wall is given by

$$\overline{T}_{l/r} = \overline{T} \pm \frac{\Delta T}{2} \,. \tag{2}$$

The average temperature of the boundary $\overline{T}$ is kept constant. The linear profile of the temperature at the left and right wall $T|_{l/r}$ is defined with the vertical coordinate $\psi = y - H/2$ according to the expression

$$T|_{l/r}(\psi) = \frac{\Theta_{l,r}\,\Delta T}{H}\,\psi + \overline{T}_{l/r} \,, \tag{3}$$

where $\Theta_{l,r}$ indicates the relative temperature change due to the temperature gradient over the entire cavity height with respect to the characteristic temperature difference ΔT. The bottom and top wall temperature is specified by the expression

$$T_{b/t} = \overline{T}_{r/l} \pm \Theta_{b,t}\,\Delta T \,, \tag{4}$$

where $\Theta_{b,t}$ is the relative temperature at the bottom and top wall with respect to the characteristic temperature difference ΔT. To maintain consistency in the temperature field, a parabolic profile towards the corners and edges is applied, see (Wenig et al. 2021a). The initial helium stratification is changed by variation of the mole fraction difference ΔX with the coordinate $\upsilon = y - 2H/3$:

$$X(\upsilon) = \frac{2\Delta X}{H-h}\upsilon + (X^* - \Delta X) \,, \tag{5}$$

where $X^* = 0.40$ is the constant nominal mole fraction of the helium layer.

Model distributions for the uncertain input variables $\mathcal{Q}_i$ with corresponding realizations q_i were defined and are listed in Table 1. $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{LN}(\mu, \sigma^2)$ denote a

Table 1 Definition of mutually independent random input variables

q_i	Probability distribution for $\mathcal{Q}_i$
ΔT	$\mathcal{N}(0,\, [0.1\Delta T^*]^2)$
$\Theta_{l,r}$	$\mathcal{LN}(0.1,\, 0.1^2)$
$\Theta_{b,t}$	$\mathcal{LN}(0.2,\, 0.1^2)$
ΔX	$\mathcal{TN}(0,\, [0.2X^*]^2, 0, X^*)$
D	$\mathcal{N}(D^*,\, [0.1D^*]^2)$

normal distribution and a log-normal distribution with expectation μ and variance σ^2. $\mathcal{TN}(\mu, \sigma^2, a, b)$ denotes a truncated normal distribution with a and b as the lower and upper bounds. Further information about the case setup and regarding the definition of uncertain parameters can be found in Wenig et al. (2023).

2.1.3 Framework and Discretization

The nonlinear set of governing equations was solved in a finite-volume framework using the open-source C++ toolbox `OpenFOAM v.2006` (Weller et al. 1998; OpenFOAM documentation v2006 2019). The pressure–velocity coupling was addressed by using the PIMPLE algorithm. It is ensured that the normalized residuals of the pressure–velocity coupling fall below the value 10^{-4}, while for the matrix solvers below 10^{-6}. The convective momentum flux was evaluated using the second-order linear upwind scheme, while the remaining convective and diffusive fluxes were evaluated using the limited linear scheme. The convective flux of the helium mass fraction was discretized using the limited linear scheme, which is bounded between 0 and 1. Temporal advancement was achieved by blending 10% implicit Euler and 90% Crank–Nicolson schemes. Noting that Uncertainty Quantification requires a large number of simulation runs to be performed, the numerical schemes were deliberately chosen as a compromise between accuracy and robustness. Furthermore, it has been ensured that the CFL number is always below the value of 0.5. A sufficiently fine spatial grid resolution was used to resolve most of the turbulent fluctuations for LES. The mesh refinement strategy used in Wenig et al. (2021a) was adopted to ensure appropriate resolution of the wall-boundary layer. The dimensionless horizontal and vertical wall normal distances of the wall adjacent cell centroids were defined at $y^+ \leq 1$. The mesh was refined linearly from the central planes of the cavity towards the walls with constant expansion factors of 1.156 and 1.020 for the horizontal and vertical directions, respectively. The number of cells is 44 in the horizontal direction, 220 in the vertical direction, and 34 in the depth direction. The maximum dimensionless wall tangential cell sizes in streamwise direction were set with $\Delta x^+ \approx 30$. A sufficient length in the periodic direction $\varphi_{WD} = W/D = 1$ was applied to ensure that turbulence fluctuations are uncorrelated at a separation of one half-period (Trias et al. 2007). The mesh in periodic direction was uniformly distributed with $\Delta z^+ \approx 20$. A detailed description of the mesh with a supplementary grid convergence study can be found in Wenig et al. (2021a).

2.2 Uncertainty Quantification

The construction of stochastic models via Stochastic Spectral Methods is presented in Sect. 2.2.1. The PCE–HDMR methodology is introduced in Sect. 2.2.2. Section 2.2.3 provides an explanation of the Stochastic Model Composition technique. Furthermore, Sects. 2.2.4 and 2.2.5 provide principles for variance-based decomposition and error estimation, respectively.

2.2.1 Stochastic Spectral Methods

In this section, PCEs and KLEs are introduced as two applied representatives of Stochastic Spectral Methods. The use of PCE (Wiener 1938; Ghanem and Spanos 1991) was motivated by its capability to approximate stochastic results in an computationally efficient manner due to high convergence rates as the number of simulation runs increases. Let $\boldsymbol{\mathcal{Q}}$ be random input variables $\boldsymbol{\mathcal{Q}} : \Omega \rightarrow \boldsymbol{\Upsilon} \subset \mathbb{R}^n$ that map events $\omega \in \Omega$ from sample space Ω to

realizations $\mathbf{q} \in \Upsilon$. In the employment of PCE, random response functions $\mathcal{R}(\omega) = R(\boldsymbol{\mathcal{Q}})$ are approximated through a series of suitable multidimensional orthogonal polynomials $\Psi_{\boldsymbol{\beta}}(\boldsymbol{\mathcal{Q}})$ with corresponding expansion coefficients $\alpha_{\boldsymbol{\beta}}$. The expression for the truncated PCE is given in the following:

$$\mathcal{R}(\omega) \approx R^{\mathcal{B}}(\boldsymbol{\mathcal{Q}}) = \sum_{\boldsymbol{\beta} \in \mathcal{B}^{n,d}} \alpha_{\boldsymbol{\beta}} \Psi_{\boldsymbol{\beta}}(\boldsymbol{\mathcal{Q}}) \,, \tag{6}$$

where $\boldsymbol{\beta} = \left(\beta_1, \cdots, \beta_n\right)$ with $\beta_i \geq 0$ is a n-dimensional multi-index from the finite index set $\mathcal{B}^{n,d} = \{\boldsymbol{\beta} \in \mathbb{N}^n : |\boldsymbol{\beta}| \leq d\}$, which specifies the univariate polynomial degree tuples of each polynomial basis element. Therefore, Ψ_{β_i} denote orthogonal polynomials of order β_i, which yield $\Psi_{\boldsymbol{\beta}} = \prod_{i=1}^{n} \Psi_{\beta_i}\left(\mathcal{Q}_i\right)$. For computational purposes, the PCE is truncated by retaining a total polynomial degree of $|\boldsymbol{\beta}| = \sum_{i=1}^{n} \beta_i \leq d$. This truncation leads to a total number of expansion terms P. Thus, PCE can also be formulated with a term-based index $p \in [1, P]$, which indicates the p-th multi-index $\boldsymbol{\beta}^{(p)}$ in the polynomial basis $\mathcal{B}^{n,d}$:

$$\mathcal{R}(\omega) \approx R^{P}(\boldsymbol{\mathcal{Q}}) = \sum_{p=1}^{P} \alpha_{\boldsymbol{\beta}^{(p)}} \Psi_{\boldsymbol{\beta}^{(p)}}(\boldsymbol{\mathcal{Q}}) \,. \tag{7}$$

The orthogonal polynomials were numerically generated by using the Gramm–Schmidt (Witteveen and Bijl 2006) approach. The expansion coefficients $\alpha_{\boldsymbol{\beta}^{(p)}}$ were estimated by using the Smolyak sparse grid method (Smolyak 1963; Xiu 2007; Constantine et al. 2012) for the reference and tensor quadrature for the PCE–HDMR model. The sparse grid quadrature rule is defined by

$$\mathcal{A}(m,n) = \sum_{m+1 \leq |\mathbf{l}| \leq m+n} (-1)^{m+n-|\mathbf{l}|} \begin{pmatrix} n-1 \\ m+n-|\mathbf{l}| \end{pmatrix} \cdot \left(\mathcal{U}_{l_1}^{(1)} \otimes \cdots \otimes \mathcal{U}_{l_n}^{(1)}\right), \tag{8}$$

where $\mathcal{U}_{l_i}^{(1)}$ denotes one-dimensional quadrature operators with the level $l_i \in \mathbb{N}_+$ and $\mathbf{l}$ denotes the multi-index $\mathbf{l} = \left(l_1, \ldots, l_n\right) \in \mathbb{N}_+^n$. The dimension independent maximum sparse grid level m controls the number of function evaluations and the associated accuracy of the PCE. Regarding the reference model, an isotropic sparse grid with level 2 and Gaussian quadrature rules are applied. Further details are provided in previous work (Wenig et al. 2023). The PCE coefficients for the PCE–HDMR approach were computed by using discrete projection through Gaussian quadrature with quadrature order of 5. The Gauss points and weights are computed by the Golub–Welsch (Golub and Welsch 1969) tridiagonal eigensolution. The open-source software Dakota 6.10 (Adams et al. 2019) was used as framework for the determination of the PCEs.

The PCE approach was further complemented by the application of KLE (Karhunen 1947; Loève 1978). For this purpose, a stochastic model of random fields or stochastic processes was constructed through a finite linear combination of orthogonal deterministic basis functions, or principal components, multiplied by uncorrelated random variables. With this approach, the variance in the random response data is captured through its most important modes of variability. Let $\mathcal{R}(\boldsymbol{x}, \omega)$ be a random field (RF), which is dependent on the deterministic field variable $\boldsymbol{x} \in \mathcal{X}$ and random event $\omega \in \Omega$, where $\mathcal{X}$ denotes the field domain and Ω is the sample space. The RF can be divided into its mean $\mu_{\mathcal{R}}(\boldsymbol{x})$ and centered field $\mathcal{R}_0(\boldsymbol{x}, \omega)$ in the following way:

$$\mathcal{R}(\boldsymbol{x},\omega) = \mu_{\mathcal{R}}(\boldsymbol{x}) + \mathcal{R}_0(\boldsymbol{x},\omega)\ . \tag{9}$$

Subsequently, a truncated KLE is applied to approximate the centered field $\mathcal{R}_0(\boldsymbol{x},\omega)$

$$\mathcal{R}_0(\boldsymbol{x},\omega) \approx R_0^K(\boldsymbol{x},\boldsymbol{\zeta}) = \sum_{k=1}^{K} \sqrt{\lambda_k}\varphi_k(\boldsymbol{x})\zeta_k\,, \tag{10}$$

where $\boldsymbol{\zeta} = \zeta_k(\omega)$ are mutually uncorrelated random variables with zero mean and unit variance. The scalars λ_k and field-dependent deterministic functions $\varphi_k(\boldsymbol{x})$ are, respectively, the eigenvalues and orthogonal eigenfunctions to the homogeneous Fredholm equation of second kind. For solving the integral within the Fredholm equation, the discrete Karhunen-Loève method (Schenk and Schuëller 2005), also known as Principal Component Analysis (PCA), was employed along with a uniform discretization of the field domain with B points. An estimation of the expectation of the random field $\hat{\mu}_{\mathcal{R}}(\boldsymbol{x})$ was determined by utilizing the underlying quadrature rule according to the PCE coefficient estimation method. Due to the sparsity of simulation results in the PCE–HDMR approach, the covariance matrix was constructed in an analogous manner through the estimation of individual entries of the covariance matrix $\mathbb{E}\left[R_0(x_a,\boldsymbol{\mathcal{Q}})R_0(x_b,\boldsymbol{\mathcal{Q}})\right]$ based on the underlying quadrature rule. The covariance matrix for the reference model was constructed by taking the sample covariance (Wenig et al. 2023). The truncation of the KLE was accomplished through a practical rule based on the eigenvalue decay rate, i.e. terms with eigenvalues λ_k, which decayed to some fraction (e.g. 10%) of the largest eigenvalue λ_1, are omitted. In the present work, the decay rate $\lambda_{K+1}/\lambda_1 \leq 0.02$ was applied. The mutually uncorrelated random variables ζ_k were approximated with PCE (Huan et al. 2019; Jivani et al. 2021), to establish the relationship between ζ_k and $\mathcal{Q}_i$: $\zeta_k^P(\omega) = \sum_{p=1}^{P} \gamma_{k,\boldsymbol{\beta}^{(p)}}\Psi_{\boldsymbol{\beta}^{(p)}}(\boldsymbol{\mathcal{Q}})$, where $\gamma_{k,\boldsymbol{\beta}^{(p)}}$ are the PCE coefficients for the approximation of ζ_k. This eventually yields a response approximation $\mathcal{R}^{P,K}(\boldsymbol{x},\omega)$ based on PCE in conjunction with KLE with field dependent expansion coefficients $\alpha^K_{\boldsymbol{\beta}^{(p)}}(\boldsymbol{x})$:

$$\begin{aligned}
\mathcal{R}^{P,K}(\boldsymbol{x},\omega) &= \hat{\mu}_{\mathcal{R}}(\boldsymbol{x}) + \sum_{p=1}^{P}\left[\sum_{k=1}^{K}\sqrt{\lambda_k}\varphi_k(\boldsymbol{x})\gamma_{k,\boldsymbol{\beta}^{(p)}}\right]\Psi_{\boldsymbol{\beta}^{(p)}}(\boldsymbol{\mathcal{Q}}) \\
&= \hat{\mu}_{\mathcal{R}}(\boldsymbol{x}) + \sum_{p=1}^{P}\left[\alpha^K_{0,\boldsymbol{\beta}^{(p)}}(\boldsymbol{x})\right]\Psi_{\boldsymbol{\beta}^{(p)}}(\boldsymbol{\mathcal{Q}}) \\
&= \sum_{p=1}^{P}\left[\alpha^K_{\boldsymbol{\beta}^{(p)}}(\boldsymbol{x})\right]\Psi_{\boldsymbol{\beta}^{(p)}}(\boldsymbol{\mathcal{Q}})\,,
\end{aligned} \tag{11}$$

where $\alpha^K_{0,\boldsymbol{\beta}^{(p)}}(\boldsymbol{x})$ are the field-dependent PCE coefficients of the centered field. The estimation of the mean $\hat{\mu}_{\mathcal{R}}(\boldsymbol{x})$ can be added to the first expansion coefficients $\alpha^K_{0,\boldsymbol{\beta}^{(1)}}(\boldsymbol{x})$ and this yields the coefficients $\alpha^K_{\boldsymbol{\beta}^{(p)}}(\boldsymbol{x})$ for the full random field approximation, which is based on both PCE and KLE.

The KLE provides a representation of a random field that minimizes the mean squared error (MSE) between the original field and its approximation using a given number of terms. This property might facilitate an efficient construction of the random field approximation while preserving its essential characteristics. KLE also incorporates covariances that can enhance the accuracy of physical model predictions beyond those obtained by directly applying PCE to responses at individual points in time. Furthermore, the long-time

integration problem (Gerritsma et al. 2010) is mitigated, since KLE describes stochastic processes coherently with its modes of variability, which can be attributed to the variability of the random input parameters via PCE approximation of the scalar random variables ζ_k. In this way, the approximation of highly dynamic stochastic processes with a reasonable number of PCE terms can be facilitated.

2.2.2 PCE–HDMR

The PCE–HDMR approach is explained in more detail in this Section. Response functions $R(\boldsymbol{x}, \boldsymbol{\mathcal{Q}})$ can be represented by a unique HDMR or Sobol decomposition, which in practice is commonly truncated after the second-order term (Sobol' 1993):

$$R(\boldsymbol{x}, \boldsymbol{\mathcal{Q}}) \approx R_0(\boldsymbol{x}) + \sum_{i=1}^{n} R_i\left(\boldsymbol{x}, \mathcal{Q}_i\right) + \sum_{1 \le i < j \le n} R_{ij}\left(\boldsymbol{x}, \mathcal{Q}_i, \mathcal{Q}_j\right) . \tag{12}$$

Cut-HDMR or anchored HDMR proposes the determination of the series component functions of the HDMR through the evaluation of the response function along cut lines, planes, and hyperplanes through a reference point $\mathbf{q}^*$. The evaluation up to the second order term is conducted according to the following equations:

$$\begin{aligned} R_0(\boldsymbol{x}) &= R(\boldsymbol{x}, \mathbf{q}^*) , \\ R_i\left(\boldsymbol{x}, \mathcal{Q}_i\right) &= R\left(\boldsymbol{x}, \mathcal{Q}_i , \mathbf{q}^*_{\sim i}\right) - R_0(\boldsymbol{x}) , \\ R_{ij}\left(\boldsymbol{x}, \mathcal{Q}_i, \mathcal{Q}_j\right) &= R\Big(\boldsymbol{x}, \mathcal{Q}_i , \mathcal{Q}_j , \mathbf{q}^*_{\sim ij}\Big) - R_i\left(\boldsymbol{x}, \mathcal{Q}_i\right) - R_j\left(\boldsymbol{x}, \mathcal{Q}_j\right) - R_0(\boldsymbol{x}) , \end{aligned} \tag{13}$$

where $\mathbf{q}^*_{\sim i}$ indicates that all input parameters except q_i are at the reference point, and $\mathbf{q}^*_{\sim ij}$ analogously indicates that all input parameters except q_i and q_j are at the reference point. $\mathcal{Q}_i$ or $\mathcal{Q}_j$ denote the random input variables under current consideration. The superposition of the individual terms of the Cut-HDMR in Eq. 13 yields an approximation of the response $R(\boldsymbol{x}, \boldsymbol{\mathcal{Q}})$. Figure 2 illustrates the Cut-HDMR approach through geometric interpretation of the first Sobol decomposition terms of a two-dimensional response function. The first term refers to the reference point result, which is depicted in Fig. 2a. The univariate effects by individual parameters are considered in the first-order terms. Corresponding functions are

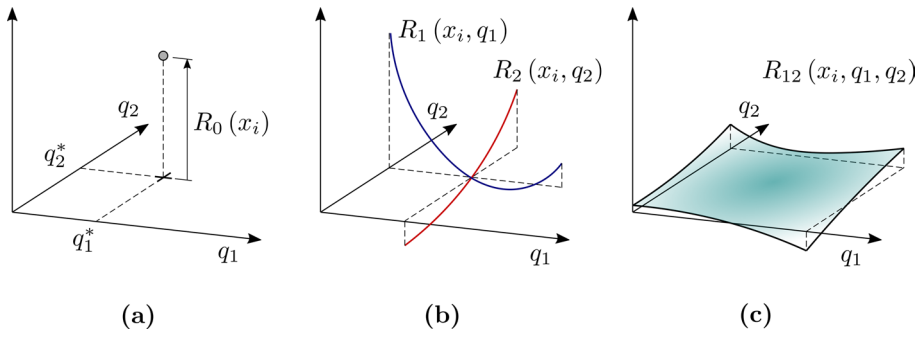

Fig. 2 Schematic sketch of the first three Sobol decomposition terms of a two-dimensional function: **a** Reference point result, **b** univariate result functions, and **c** bivariate result function

schematically shown in Fig. 2b. The second-order term includes bivariate or interaction effects with regard to the response. Figure 2c schematically shows the two-dimensional plane of this term. When using PCE–HDMR, metamodels of these component functions within the Cut-HDMR are simultaneously approximated through PCEs (Yue et al. 2021):

$$\begin{aligned} R_0(\boldsymbol{x}) &= R(\boldsymbol{x}, \mathbf{q}^*)\,, \\ R_i\left(\boldsymbol{x}, \mathcal{Q}_i\right) &= \sum_{\boldsymbol{\beta} \in \mathcal{B}^{1,d}} \alpha_{\boldsymbol{\beta}}(\boldsymbol{x}, \mathbf{q}^*) \Psi_{\boldsymbol{\beta}}\left(\mathcal{Q}_i\right)\,, \\ R_{ij}\left(\boldsymbol{x}, \mathcal{Q}_i, \mathcal{Q}_j\right) &= \sum_{\boldsymbol{\beta} \in \mathcal{B}^{2,d}} \alpha_{\boldsymbol{\beta}}(\boldsymbol{x}, \mathbf{q}^*) \Psi_{\boldsymbol{\beta}}\left(\mathcal{Q}_i\,, \mathcal{Q}_j\right)\,. \end{aligned} \tag{14}$$

The summation of these terms yields the PCE–HDMR approximation of the response:

$$\begin{aligned} R(\boldsymbol{x}, \boldsymbol{\mathcal{Q}}) \approx R(\boldsymbol{x}, \mathbf{q}^*) &+ \sum_{1 \le i \le n} \sum_{\boldsymbol{\beta} \in \mathcal{B}^{1,d}} \alpha_{\boldsymbol{\beta}}(\boldsymbol{x}, \mathbf{q}^*) \Psi_{\boldsymbol{\beta}}\left(\mathcal{Q}_i\right) \\ &+ \sum_{1 \le i < j \le n} \sum_{\boldsymbol{\beta} \in \mathcal{B}^{2,d}} \alpha_{\boldsymbol{\beta}}(\boldsymbol{x}, \mathbf{q}^*) \Psi_{\boldsymbol{\beta}}\left(\mathcal{Q}_i\,, \mathcal{Q}_j\right)\,. \end{aligned} \tag{15}$$

In the process, a suitable reference point $\mathbf{q}^*$ has to be selected. Its choice is crucial and it was defined at the center point of sparse grid quadrature according to recommendations in Gao and Hesthaven (2010). Prior analysis by means of total-order Sobol indices revealed that the uncertain input parameters show negligible interaction behavior with respect to considered random responses in Wenig et al. (2023). For this reason, the second-order terms in Eq. (15) were omitted and solely univariate effects were taken into account. The PCE approximations were computed according to Sect. 2.2.1. Gaussian quadrature with quadrature order of 5 was applied for discrete projection and responses are approximated by polynomials of order $d = 4$.

PCE–HDMR enables the gradual construction of high-dimensional stochastic models by combining multiple low-dimensional models and can thus increase computational efficiency.

2.2.3 Stochastic Model Composition

The field of CFD often involves the prediction of transient processes that are characterized by highly dynamic behavior. As a result, the stochastic approximation of random responses in the presence of uncertainties might be impeded since the occurrence of nonlinear effects can exhibit temporal variability due to its dependence on the random input variables. In the mixing process under consideration, such nonlinear temporal effects occur and their practicable approximation can be facilitated by realignment and using an intermediate variable. Therefore, Stochastic Model Composition (SMC), or Random Field Composition (RFC), was introduced in Wenig et al. (2023), as a novel approach for modeling the transient behavior of time-dependent stochastic processes with function composition. Similar realignment techniques, which utilize linear scaling or transform, are reported in literature (Liao and Zhang 2016; Giraldi et al. 2017; Colombo et al. 2018; Bonnaire et al. 2021). In the present work, the SMC methodology is employed to perform UQ of the buoyancy-induced mixing process. The realignment technique is distinguished through the concept of introducing an additional intermediate field variable with which the chronological process of transient responses can be represented and nonlinear effects can be aligned to similar

locations with respect to the intermediate variable. This intermediate variable should accurately capture and represent the essential characteristics of the underlying physical process and thereby provide a measure to which the stochastic processes at hand exhibit similar transient behavior. Therefore, the realignment procedure, which is determined by the intermediate variable, is derived based on physical principles. For instance, entropy could be a suitable measure for realignment in the context of a thermodynamic process as entropy increases monotonically and characterizes the progress of a process. In the context of a chemical process, for example, random fields or stochastic processes could be realigned by means of the reaction progress variable, which takes on values from 0 to 1. When examining the mixing processes in this work, the primary focus is on the temporal mixing behavior. Therefore, for the description of the progress of desired responses, it is beneficial to choose an intermediate variable indicating the mixing state, which starts from an inhomogeneous state and ends with a fully mixed homogeneous state. The present study accomplishes the description of the mixing state through the normalized mixture uniformity Σ_X:

$$z = \Sigma_X = \frac{\sigma_X}{\sigma_{X_0}} = \frac{\sqrt{V^{-1}\int_V \left(X - \bar{X}\right)^2 \mathrm{d}V}}{\sqrt{V^{-1}\int_V \left(X_0 - \bar{X}\right)^2 \mathrm{d}V}}, \tag{16}$$

where X is the helium mole fraction, X_0 is the initial helium mole fraction, $\bar{X}$ is the helium mole fraction of the homogeneous mixture and V is the total fluid volume. Σ_X is a continuous and monotonic function, which starts with a value of $\Sigma_X = 1$, indicating the inhomogeneous state and monotonically decreases to the value $\Sigma_X = 0$, characterizing the homogeneous state of the mixture. These properties allows to represent the responses as functions of the mixing state $G(z, \boldsymbol{\mathcal{Q}})$. Given that the considered mixing processes exhibit similar transient behavior with respect to the mixing state, this facilitates a reduction in the complexity of the response function. To establish the relationship between the mixing state with physical time, an additional stochastic model $Z(t, \boldsymbol{\mathcal{Q}})$ is constructed, which represents the mixing state as a function of time. Subsequently, the state-dependent stochastic models of responses are combined with the time-dependent stochastic model of the mixing state through function composition $G(Z(t, \boldsymbol{\mathcal{Q}}), \boldsymbol{\mathcal{Q}})$, which is the mathematical concept behind the SMC methodology. Figure 3a illustrates this operation with a domain diagram. The function Z maps times to different states, and subsequently, the function G maps each of these states to their respective responses. This allows to establish the relationship sequentially between time and a response quantity. In practice, tabular data of realizations of quantities like time t, state z and responses r is determined from stochastic models, as shown in Fig. 3b. Finally, the relationship between a response and time can be derived by calculation of the time instants for the corresponding states through linear interpolation. The algorithm can be summarized as follows:

1. Run simulations for stochastic model construction and collect data of responses R, the state variable z and time t;
2. Compute state-dependent stochastic models of the response $G(z, \boldsymbol{\mathcal{Q}})$ and a stochastic model $Z(t, \boldsymbol{\mathcal{Q}})$, which describes the relationship between the state variable z and time t;
3. Derive model realizations for the state-dependent response $G(z, \mathbf{q})$ and for the state-time relationship $Z(t, \mathbf{q})$, which are respectively based on identical input parameter samples;

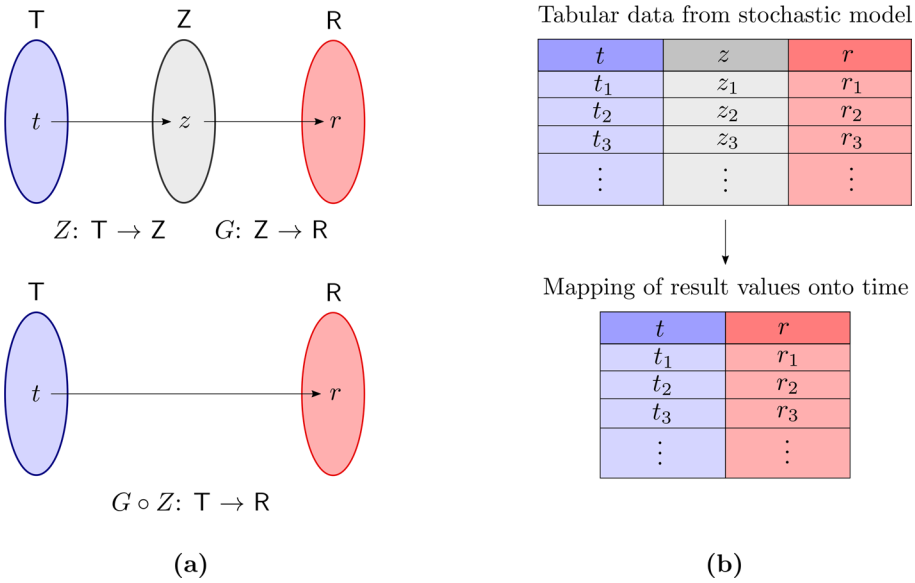

Fig. 3 Illustration of the SMC method: **a** Domain diagram and **b** practical implementation of the methodology with tabulated data

4. Derive time-dependent realizations by calculating the time instants, which correspond to the respective states, via interpolation and allocation of the time instants to the individual response values.

In this way, SMC enables the approximation of intricate stochastic processes $\mathcal{R}(t, \omega)$ by realignment of nonlinear temporal effects and by partitioning the stochasticity into two stochastic models.

2.2.4 Variance-Based decomposition

Through the determination of variance-based measures of sensitivity (Sobol 2001), it becomes possible to identify the sources of response uncertainties and attribute them to uncertainties in input parameters. In this approach, the variance is split into partial variances that stem from individual parameters or from their interaction behavior:

$$\mathbb{V}\mathrm{ar}\,(\mathcal{R}) = \sum_{i=1}^{n} V_{\mathcal{Q}_i} + \sum_{i<j}^{n} V_{\mathcal{Q}_i\mathcal{Q}_j} + \cdots + V_{\mathcal{Q}_i\mathcal{Q}_j\ldots\mathcal{Q}_n}\,. \tag{17}$$

From this decomposition, variance-based measures of sensitivity can be inferred. In the present work total-order Sobol indices $S_{T_{\mathcal{Q}_i}}$ are considered, which characterize the overall variance caused by an individual parameter through its univariate effect and interaction with other parameters:

$$S_{T_{\mathcal{Q}_i}} = \frac{V_{T_{\mathcal{Q}_i}}}{\mathbb{V}\mathrm{ar}\,(\mathcal{R})}\,,\; V_{T_{\mathcal{Q}_i}} = V_{\mathcal{Q}_i} + \sum_{\substack{j=1 \\ j \neq i}}^{n} V_{\mathcal{Q}_i\mathcal{Q}_j} + \cdots + V_{\mathcal{Q}_i\mathcal{Q}_j\ldots\mathcal{Q}_n}\,. \tag{18}$$

A Monte Carlo estimator was used for the derivation of the total-order indices (Saltelli et al. 2010):

$$S_{T_i} \approx \frac{N-1}{2N} \frac{\sum_{j=1}^{N}\left(R(\mathbf{A}_j) - R\left(\left[\mathbf{A}_{\mathbf{B}}^{i}\right]_j\right)\right)^2}{\sum_{j=1}^{N}\left(R(\mathbf{A}_j) - \left[N^{-1}\sum_{k=1}^{N} R(\mathbf{A}_k)\right]\right)^2}, \tag{19}$$

where $\mathbf{A}$ and $\mathbf{B}$ are two independent sample matrices. $\left[\mathbf{A}_{\mathbf{B}}^{i}\right]$ contains the elements for the ith input variable from the second matrix $\mathbf{B}$.

The variance-based decomposition enables a thorough evaluation of the stochastic processes by identifying the most significant input parameters that contribute to the response variance. In this way, it is feasible to attain informed decision-making and interpretation of results is facilitated.

2.2.5 Model Error Estimation

The assessment of the accuracy of stochastic models and UQ results is accomplished through the estimation of the root-mean-square error (RMSE), which is equivalent to the square root of the generalization error (Vapnik 2000; Blatman and Sudret 2011). The RMSE is defined as:

$$\mathrm{RMSE} = \left(\mathbb{E}\left[\left(\hat{R}(t,\boldsymbol{\mathcal{Q}}) - R(t,\boldsymbol{\mathcal{Q}})\right)^2\right]\right)^{1/2}, \tag{20}$$

where $\hat{R}(t,\boldsymbol{\mathcal{Q}})$ denotes the stochastic model, which approximates the stochastic process of the response $R(t,\boldsymbol{\mathcal{Q}})$. The RMSE is the square root of the expectation of the squared differences between the predicted and actual process. In this way, the accuracy of the stochastic model predictions can be assessed with respect to the underlying computational model. In the present work, the RMSE was determined using the following estimator:

$$\mathrm{RMSE} = \left(\frac{1}{N}\sum_{i=1}^{N}\left(\hat{R}\left(t,\mathbf{q}^{(i)}\right) - R\left(t,\mathbf{q}^{(i)}\right)\right)^2\right)^{1/2} \tag{21}$$

where $\hat{R}\left(t,\mathbf{q}^{(i)}\right)$ and $R\left(t,\mathbf{q}^{(i)}\right)$ are respectively the realization transients predicted by the stochastic model and realization transients computed by the CFD simulation, which were used for stochastic model construction. For the estimation of the RMSE with regard to the SGM and PCE–HDMR approach, the realization transients from the simulation runs for the construction of the SGM reference model are taken into account. This ensures comparability between both approaches.

As global measure, the temporal mean RMSE, denoted by $\langle\mathrm{RMSE}\rangle$, is determined through the square root of the integral mean of the mean square error $\mathrm{MSE} = (\mathrm{RMSE})^2$ over the whole time span of the model $\Delta t = t_{\mathrm{end}} - t_{\mathrm{start}}$: It is determined through the expression

$$\langle \mathrm{RMSE} \rangle = \left(\Delta t^{-1} \int_{t_{\mathrm{start}}}^{t_{\mathrm{end}}} \mathrm{MSE}\, \mathrm{d}t \right)^{1/2}, \tag{22}$$

where the integral is evaluated through the trapezoidal rule.

The aforementioned error metrics facilitate the assessment of the accuracy of the stochastic models and provide a measure for identifying potential model improvements.

3 Results and Discussion

In Sect. 3.1, the physical phenomenology of the mixing process within the DHC is explained. Subsequently, the analyzed Quantities of Interest are introduced in Sect. 3.2, which is followed by the presentation and discussion of results for time-dependent stochastic processes in Sect. 3.3.

3.1 Phenomenology

This section presents the physical phenomena that occur during the mixing process within the DHC, with the aim of facilitating result interpretation. The mixing process comprises multiple consecutive phases. It starts with a quiescent stratification of helium and air, as shown by the initial fluid property fields in Fig. 4. A buoyancy-driven mixing process is initiated due to the temperature difference between the left and right walls, which results in the establishment of two distinct circulating natural convection flows in the upper helium-rich and lower air-rich regions. During this stage, mixing occurs mainly by diffusive mass transport, which is evident from the fluid property fields in Fig. 5. Diffusive mixing persists until the density difference between the upper and lower regions of the DHC becomes small enough for buoyancy forces to erode the remaining helium layer, leading to complete

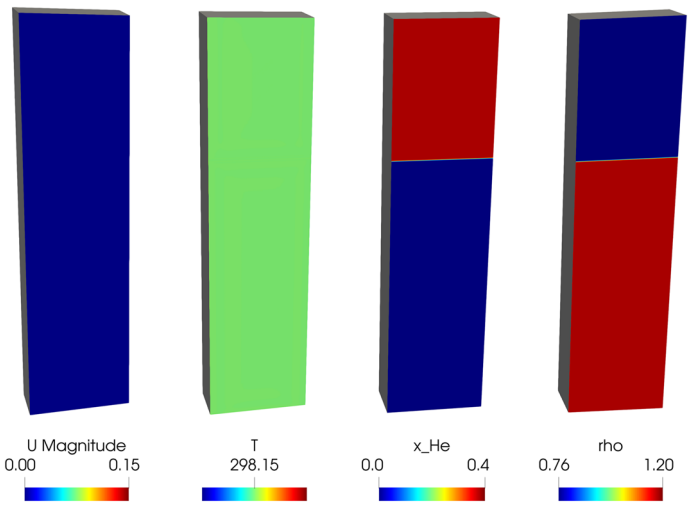

Fig. 4 3D full section view of DHC with fluid property fields for the initial state

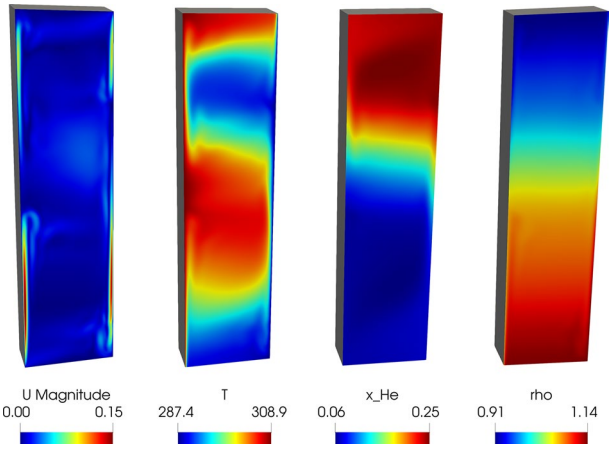

Fig. 5 3D full section view of DHC with fluid property fields during the diffusive phase

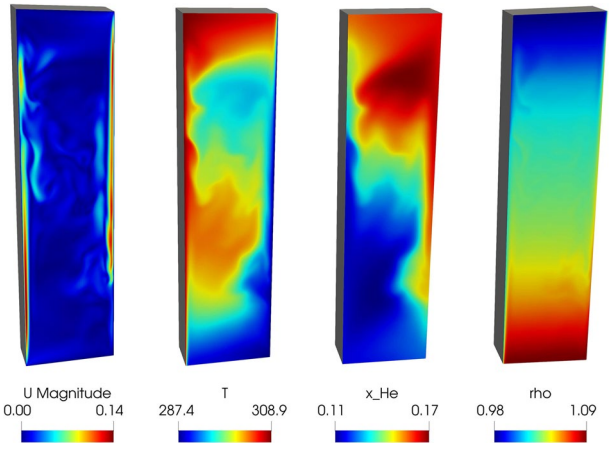

Fig. 6 3D full section view of DHC with fluid property fields during the erosion phase

mixing of air and helium. The fluid property fields during the convective erosion are shown in Fig. 6.

Dynamic transition stages like the convective erosion show significant impact on stochastic results and pose considerable challenges in terms of their approximation using stochastic models. This is caused by the fact that these dynamic transitions occur at different points in time depending on the random input variables, inducing strong non-linearities in the model to be approximated. However, with the methods described in the previous sections, a reliable prediction of such stochastic processes is accomplished.

3.2 Quantities of Interest

For the UQ study, relevant result quantities, which characterize the buoyancy-driven mixing process, need to be analyzed. These result quantities are commonly referred to as Quantities of Interest (QoI) in the field of uncertainty quantification. In natural convection flows, convective heat transfer plays a crucial role. To assess this phenomenon, the spatially averaged Nusselt number Nu_l over the left wall is considered. This dimensionless

quantity provides valuable information about convective heat transfer and is determined according to the following expression

$$Nu_l = \frac{1}{A_w} \int\limits_{A_w} \frac{\left(\nabla T \cdot \vec{n}\right)\Big|_w H}{\Delta T} \mathrm{d}A \,, \tag{23}$$

where $\vec{n}$ denotes the wall-normal unit vector and A_w denotes the face area of the left wall. Furthermore, the global kinetic energy serves as an indicator of the convection mechanisms that are taking place. The corresponding quantity E_k is the ratio of the global kinetic energy divided by the reference kinetic energy $\alpha^2 Ra/H^2$, which is defined with material properties of air. The definition of E_k reads as follows:

$$E_k = \left(\frac{\alpha^2}{H^2} Ra\right)^{-1} \cdot \frac{1}{M} \int\limits_M \frac{1}{2} \mathbf{u}^2 \mathrm{d}m \,. \tag{24}$$

M denotes the total mass in the fluid domain and dm denotes the mass in a cell. Natural convection occurs inside the cavity, resulting in the gradual mixing of air and helium through both diffusion and convection until an homogeneous mixture forms. The temporal progress of the mixing process is measured by a quantity called the mixture uniformity σ_X. This quantity represents the volume-weighted standard deviation of the mole fraction X from the mole fraction of the homogeneous state $\bar{X}$ over the entire fluid domain. It is defined by the expression

$$\sigma_X = \left(V^{-1} \int\limits_V (X - \bar{X})^2 \mathrm{d}V \right)^{1/2} . \tag{25}$$

The mixing process starts from the inhomogeneous state characterized by the temporal maximum values for σ_X. When the mixing proceeds, the normalized mixture uniformity finally takes on the value $\sigma_X = 0$, which describes the homogeneous state. The response variables described above are time-dependent stochastic processes and are considered in the next section over dimensionless time represented by the Fourier number $Fo = D^* t / H^2$, where D^* is the nominal molecular diffusion coefficient.

3.3 Time-Dependent Stochastic Processes

The state-dependent stochastic models of the aforementioned random responses together with the stochastic model, that describes the relationship between the mixing state and time, were created using PCE in conjunction with KLE based on the PCE–HDMR and SGM methods. Subsequently, time-dependent model realizations were determined from the two-step models using SMC and numerous statistics were computed. Results, which were derived from PCE–HDMR and SGM, are depicted in Figs. 7, 8 and 9. As a reference for the PCE–HDMR method, results from SGM are employed. Comprehensive stochastic representations of the mixture uniformity σ_X are given in Fig. 7. The sub figures at the top show the results obtained from the PCE–HDMR approach, while the reference results derived by SGM are analogously presented at the bottom. When considering the respective

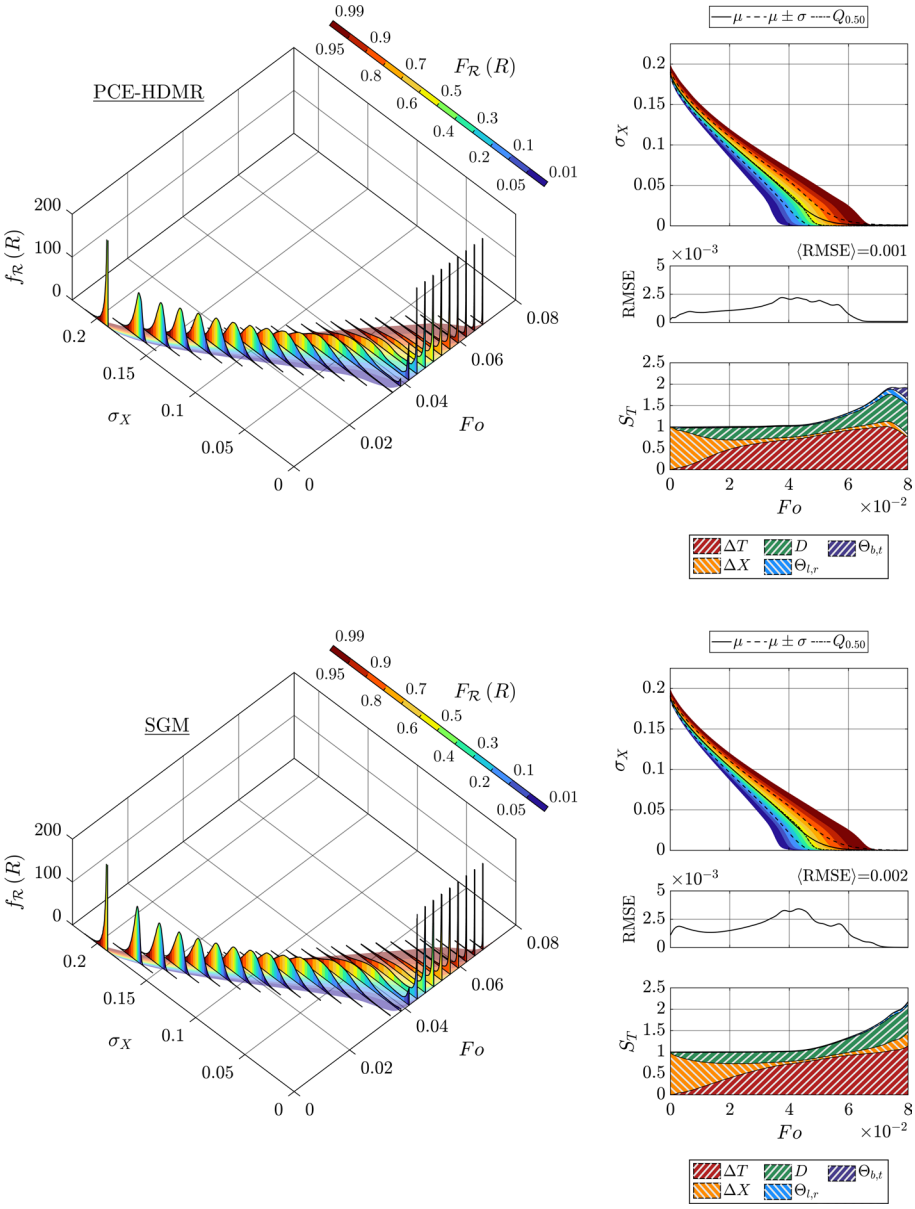

Fig. 7 Probabilistic characterization of σ_X derived from PCE–HDMR and SGM with PDFs $f_{\mathcal{R}}(R)$, discrete levels of the CDFs $F_{\mathcal{R}}(R)$, expectation μ, standard deviation σ, quantiles Q, RMSE estimation and stacked total-order Sobol indices S_T

results for PCE–HDMR and SGM, the left subplots show the probability density functions (PDFs) $f_{\mathcal{R}}(R)$ at different points in time. In addition, discrete levels of the cumulative distribution functions (CDFs) $F_{\mathcal{R}}(R)$ are indicated with colors according to a colormap. The PDFs were estimated with a kernel density estimator based on a normal kernel function

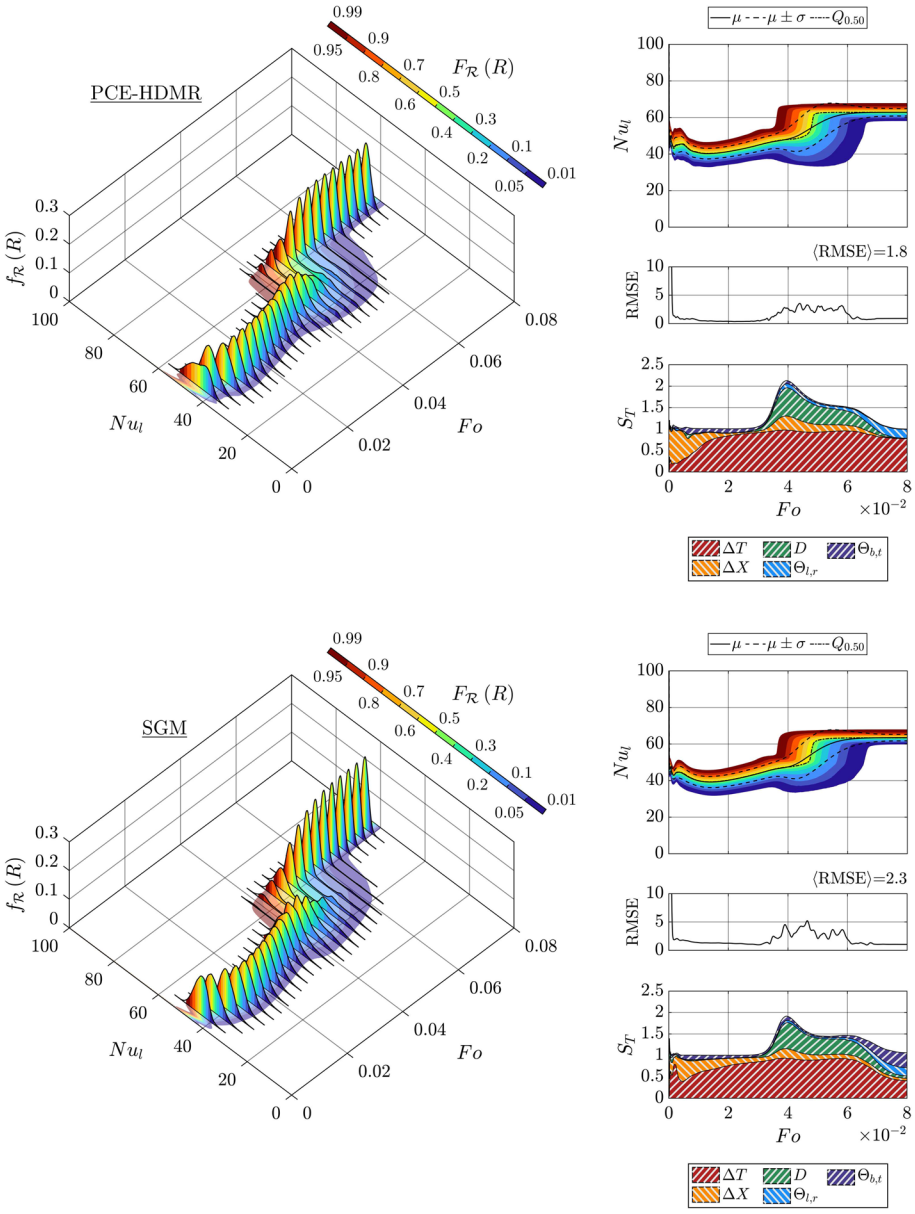

Fig. 8 Probabilistic characterization of Nu_l derived from PCE–HDMR and SGM with PDFs $f_{\mathcal{R}}(R)$, discrete levels of the CDFs $F_{\mathcal{R}}(R)$, expectation μ, standard deviation σ, quantiles Q, RMSE estimation and stacked total-order Sobol indices S_T

(The MathWorks Inc 2019a; Bowman and Azzalini 1997). The quantiles corresponding to the discrete levels of the CDFs were computed by interpolating realization values from the sorted sample data (The MathWorks Inc 2019b; Langford 2006). The sample size utilized for the estimations was $N_s = 10^5$. Furthermore, the top right subplots shows further

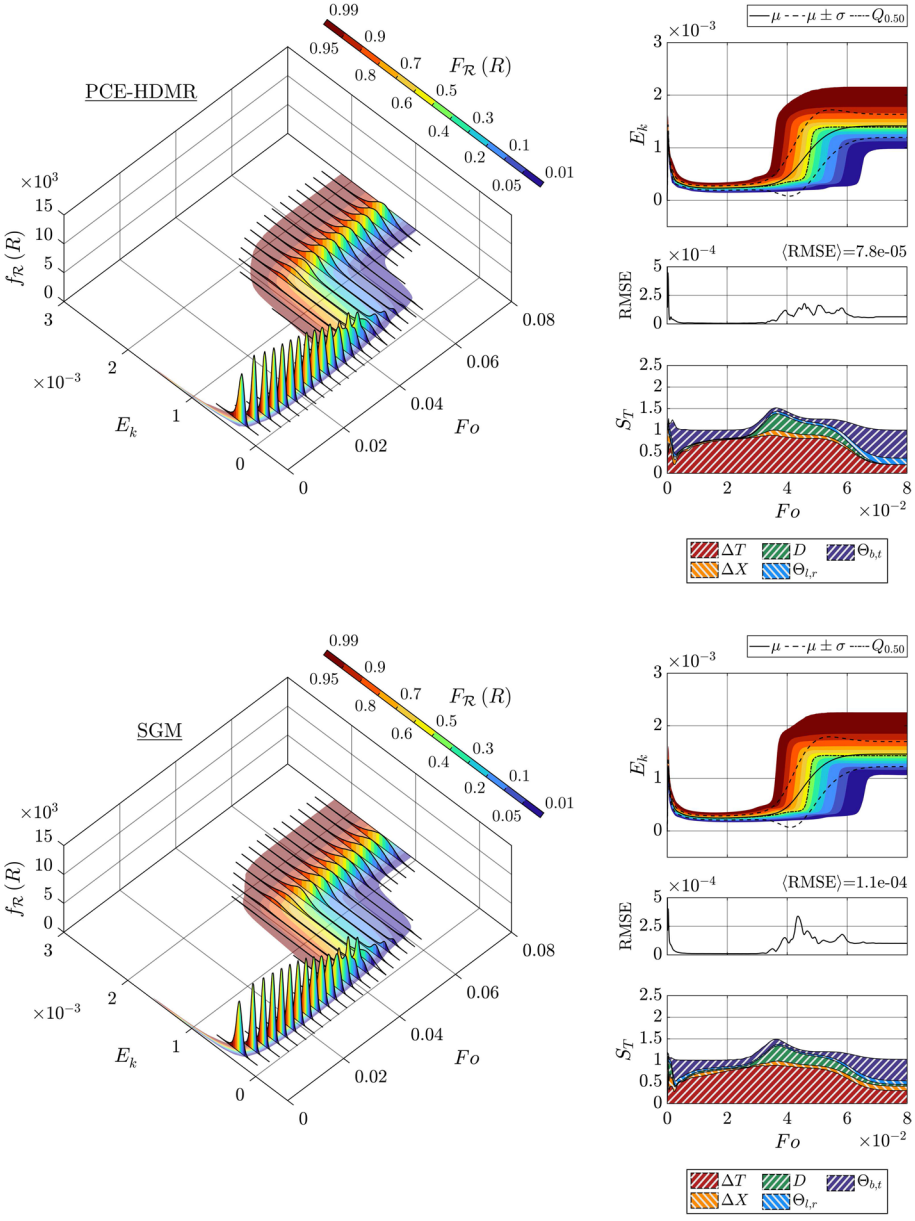

Fig. 9 Probabilistic characterization of E_k derived from PCE–HDMR and SGM with PDFs $f_{\mathcal{R}}(R)$, discrete levels of the CDFs $F_{\mathcal{R}}(R)$, expectation μ, standard deviation σ, quantiles Q, RMSE estimation and stacked total-order Sobol indices S_T

statistics like expectation μ, standard deviation σ, median $Q_{0.50}$ along with quantiles Q, which are located at the discrete color transitions according to the values of $F_{\mathcal{R}}(R)$. The center right subplot shows the error estimates of the underlying stochastic model with respect to the simulation model by means of the RMSE. Furthermore, stacked total-order

Sobol indices are shown at the bottom right subplot and enable the assignment of result uncertainties to uncertain input parameters. The values for S_T were computed according to Eq. 19 with a sample size of 10^6.

From the left and top right plot one can see that the mixture uniformity σ_X starts with high values for σ_X, which represent the initial inhomogeneous state and ends with $\sigma_X = 0$, which represents the homogeneous state. At the beginning of the mixing process a distribution of σ_X is recognizable, which aligns with the definition of the uncertain initial distribution of the helium. This initial distribution turns into a normal distribution, which gradually widens during the mixing process. In this phase the mixing is mostly driven by diffusive mass transport. At a certain point in time, the remaining helium layer is eroded by the establishment of the circulation flow over the whole cavity height. This eventually leads to the homogeneous state of the mixture. Therefore, the probability accumulates around $\sigma_X = 0$, which is indicated by high values for the PDFs $f_{\mathcal{R}}(R)$. In order to ensure the visibility of the initial distributions, the PDFs were truncated at $f_{\mathcal{R}}(R) = 200$. Since σ_X reaches zero at different points in time, it is evident, that the duration of the mixing process is subject to uncertainty. This is due to the definition of uncertain input parameters that slow down or accelerate the mixing process. The comparison between the statistical results obtained from SGM and PCE–HDMR reveals that PCE–HDMR demonstrates a high level of accuracy in replicating the reference results achieved by SGM. Additionally, PCE–HDMR achieves this level of precision while incurring a significantly reduced computational cost of 75%, since PCE–HDMR merely required 26 simulations, in contrast to 105 simulation runs necessitated by the SGM approach. The 105 simulations of the SGM approach are defined by the sparse grid topology of level $m = 2$. The number of calculations for the PCE–HDMR approach originate from five simulations per random input variable and the calculation at the reference point. The efficiency of PCE–HDMR is also confirmed by the RMSE estimates, which are shown in the center right subplots. It exhibits low magnitudes for both models and indicates that the difference between the predictions of the stochastic models and the actual values of the simulation runs is relatively small. The stochastic models thus provide an accurate approximation of the result quantity and reliable statistical information can be derived from the individual models. The analysis revealed that the error estimate of the PCE–HDMR approach is lower compared to the RMSE estimate from the SGM reference. This disparity can be attributed to the fact that the PCE–HDMR methodology in this work relies on the utilization of univariate Gaussian integration, which is known for its high precision and accuracy. In the stacked total-order Sobol indices subplot at the bottom right, the orange area indicates that at the beginning of the mixing process the variance is mostly caused by the initial structure of the helium stratification, which changes due to the uncertain parameter ΔX. Subsequently, the variance of the mixing state measured by σ_X is mainly based on the uncertain temperature difference between the left and right wall ΔT and the uncertain molecular diffusion coefficient D, which is indicated by the red and green areas, respectively. Consequently, these parameters have a decisive influence on the progress of the mixing process.

Statistics and complementary results regarding the left wall Nusselt number Nu_l are depicted in Fig. 8. The results are also derived from the PCE–HDMR and SGM approach. In both cases, the left and top right plot indicates a stagnant phase of Nu_l at the beginning of the mixing process, which is also characterized by a normal distribution. This is followed by a transient transition to the new quasi steady state, which takes place at different points in time and thus entails a larger uncertainty band for the values of Nu_l. This transition, which refers to the convective erosion of the remaining helium layer, is evident from the gradual decrease in probability at low Nu_l and the gradual increase at higher

Nu_l over time. After the transition, the dynamic system attains a new equilibrium state, which is characterized by a Gaussian distribution once again. Minor differences between PCE–HDMR and SGM can be observed in the region, where the transient transition to the quasi steady state occurs. In the remaining temporal domain both methods provide good agreement in the derived statistics. From the center right subplots one can see that the RMSE estimate takes on larger values at the beginning of the mixing process. These inaccuracies arise, because the initial transient processes proceed similar in terms of time and, consequently, the approximation is impeded by the state-dependent perspective of the SMC approach (Wenig et al. 2023). After the initial process, the RMSE estimates indicate negligible magnitudes. However, at the onset of the aforementioned transition process deviations between the model predictions and simulation runs are present, but the RMSE estimate is still reasonably low. The bottom right subplot, which depicts the stacked total-order Sobol indices, shows that the uncertain temperature difference between the left and right wall ΔT has major influence on Nu_l. Since transient effects, such as the convective erosion process during the transition stage, depend on the mixing progress and thus on time, the uncertain input parameter D, which significantly contributes to the progress of the mixing process next to ΔT, leads to considerable proportion of the variance for Nu_l after the start of the erosion.

Results for the global kinetic energy E_k are given in Fig. 9. From the left and top right subplot, it can be seen that the statistics of the global kinetic energy indicate a narrow uncertainty band during the diffusive phase. The convective erosion of the remaining helium layer leads to the new equilibrium state. This transition stage takes place at different points in time and is therefore accompanied by a wide uncertainty band. In analogy to Nu_l, the PDFs regarding E_k also show a gradual decrease in probability while they increase at higher values during transition. The statistics from PCE–HDMR and SGM show good agreement, whereby small differences in the outer quantiles can be discerned. Furthermore, the error estimates in the center right subplot show that the error exhibits larger magnitudes at the beginning, which is due to the state-dependent perspective of the SMC approach (Wenig et al. 2023). The diffusive phase, which is characterized by the narrow uncertainty band, is precisely predicted. During the transition stage, error peaks are still evident due to small inadequacies in the model prediction of the erosion time. Once again it becomes clear that the error estimate from PCE–HDMR provides smaller values than the SGM reference. The stacked total-order Sobol indices at the bottom right reveal that the temperature differences between the left and right wall ΔT and the top and bottom wall temperature definition via $\Theta_{b,t}$ have significant impact on the variance of E_k. During the erosion phase, the diffusion coefficient D also shows a considerable effect.

Based on the findings, it is evident that the PCE–HDMR method can yield results that are comparable to those obtained from the SGM reference. Consequently, this methodology serves as an additional foundation for the efficient construction of stochastic models in the context of uncertainty quantification.

4 Summary and Conclusion

For the implementation of UQ on technical scale applications in the field of CFD, efficient methods are required. In the present work, a suitable method was qualified using a generic test case, which corresponds to the Differentially Heated Cavity. An initial stratification with 40 vol% helium next to air was defined in the upper third of the enclosure.

As a result of natural convection, a buoyancy-driven mixing process occurs within the enclosure, which is aimed to be predicted by LES. The definition of uncertain input parameters within the CFD simulation leads to uncertainty in results or responses, which were approximated via Stochastic Spectral Methods in conjunction with PCE–HDMR and SGM. The responses include time-dependent stochastic processes of quantities, which characterize the mixing behavior, heat transfer and convection mechanisms. The corresponding UQ results include a thorough probabilistic representation of these responses that comprehensively describe the mixing process. From the underlying stochastic models, total-order Sobol indices were computed, which allowed for the assignment of variance proportions of the responses to individual uncertain input parameters. In addition, supplementary RMSE error estimates were determined for the assessment of the respective model accuracy.

The stochastic model statistics and complementary results from PCE–HDMR and the SGM reference revealed very good agreement. Therefore, PCE–HDMR is proven a promising UQ-method for large-scale CFD applications, since it enables the construction of high-dimensional stochastic models through low-dimensional submodels and in this way allows for the investigation of a large number of random inputs. Moreover, the stochastic model construction by means of PCE–HDMR is distinguished through its efficiency by taking advantage of the high convergence rate of numerical integration techniques like Gaussian quadrature for individual low-dimensional random input spaces. This reduced the number of 105 simulation runs for SGM for the reference UQ results to 26 simulation runs for PCE–HDMR, which originate from five simulations per random input variable and the calculation at the reference point. PCE–HDMR facilitates tremendous savings of computing resources while preserving high accuracy and is therefore suitable for the application to industrial scale CFD applications.

Acknowledgements Parts of the work presented in this paper are funded by the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV, Grant No. 1501595) on the basis of a decision by the German Bundestag. The computational resources for this project have been provided by the Gauss Centre for Supercomputing/Leibniz Supercomputing Centre under the Project Name pn29ce.

Author Contributions PJW: Methodology, Software, Formal Analysis, Investigation, Visualization, Writing—Original draft, Writing—Review and Editing, Funding acquisition. SK: Conceptualization, Writing–Review and Editing, Project administration, Funding acquisition. MK: Conceptualization, Writing—Review and Editing, Supervision, Project administration, Funding acquisition.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The data that support the findings of this study are available on request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval N/A.

Informed Consent N/A.

References

Adams, B.M., Bohnhoff, W.J., Dalbey, K.R., Ebeida, M.S., Eddy, J.P., Eldred, M.S., Hooper, R.W., Hough, P.D., Hu, K.T., Jakeman, J.D., Khalil, M., Maupin, K.A., Monschke, J.A., Ridgway, E.M., Rushdi, A.A., Seidl, D.T., Stephens, J.A., Swiler, L.P., Winokur, J.G.: Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.10 User's Manual. Sandia Technical Report SAND2014-4633, (May 2019)

Badillo, A., Kapulla, R.: Uncertainty quantification in CFD simulations of Isokinetic Turbulent mixing layers. In: NURETH-15, Pisa, Italy (2013)

Blatman, G., Sudret, B.: Adaptive sparse polynomial chaos expansion based on least angle regression. J. Comput. Phys. **230**(6), 2345–2367 (2011). https://doi.org/10.1016/j.jcp.2010.12.021

Bonnaire, P., Pettersson, P., Silva, C.F.: Intrusive generalized polynomial chaos with asynchronous time integration for the solution of the unsteady Navier–Stokes equations. Comput. Fluids **223**, 104952 (2021). https://doi.org/10.1016/j.compfluid.2021.104952

Bowman, A.W., Azzalini, A.: Applied Smoothing Techniques for Data Analysis. Oxford Universtiy Press Inc., New York (1997)

Colombo, I., Nobile, F., Porta, G., Scotti, A., Tamellini, L.: Uncertainty quantification of geochemical and mechanical compaction in layered sedimentary basins. Comput. Methods Appl. Mech. Eng. **328**, 122–146 (2018). https://doi.org/10.1016/j.cma.2017.08.049

Constantine, P.G., Eldred, M.S., Phipps, E.T.: Sparse pseudospectral approximation method. Comput. Methods Appl. Mech. Eng. **229**, 1–12 (2012). https://doi.org/10.1016/j.cma.2012.03.019

Cutrono, Rakhimov A., Visser, D.C., Komen, E.M.J.: Uncertainty quantification method for CFD applied to the turbulent mixing of two water layers. Nucl. Eng. Des **333**, 1–15 (2018). https://doi.org/10.1016/j.nucengdes.2018.04.004

Cutrono, Rakhimov, A., Visser, D.C., Komen, E.M.J.: Uncertainty quantification method for CFD applied to the turbulent mixing of two water layers—II: deterministic sampling for input uncertainty. Nucl. Eng. Des. **348**, 146–158 (2019). https://doi.org/10.1016/j.nucengdes.2019.04.016

Fokken, J., Krohn, B., Kapulla, R., Niceno, B., Prasser, H.-M., Badillo, A.: NEA benchmark exercise : computational fluid dynamic prediction and uncertainty quantification of a GEMIX mixing layer test. Technical report, OECD/NEA/CSNI (2019)

Freitag, M., Schmidt, E.: Simulation benchmark based on THAI—experiment on generation and dissolution of a light gas stratification by natural convection. In: NURETH-19, pp. 1–17 (2022)

Gao, Z., Hesthaven, J.S.: On ANOVA expansions and strategies for choosing the anchor point. Appl. Math. Comput. (2010). https://doi.org/10.1016/j.amc.2010.08.061

Gerritsma, M., Steen, J.B., Vos, P., Karniadakis, G.: Time-dependent generalized polynomial chaos. J. Comput. Phys. **229**(22), 8333–8363 (2010). https://doi.org/10.1016/j.jcp.2010.07.020

Ghanem, R.G., Spanos, P.D.: Stochastic Finite Elements: A Spectral Approach. Springer, New York, NY (1991). https://doi.org/10.1007/978-1-4612-3094-6

Giraldi, L., Le Maître, O.P., Mandli, K.T., Dawson, C.N., Hoteit, I., Knio, O.M.: Bayesian inference of earthquake parameters from buoy data using a polynomial chaos-based surrogate. Comput. Geosci. **21**(4), 683–699 (2017). https://doi.org/10.1007/s10596-017-9646-z

Golub, G.H., Welsch, J.H.: Calculation of gauss quadrature rules. Math. Comput. **23**, 221–230 (1969). https://doi.org/10.2307/2004418

Huan, X., Safta, C., Vane, Z.P., Lacaze, G., Oefelein, J.C., Najm, H.N.: Uncertainty propagation using conditional random fields in large-eddy simulations of scramjet computations. In: AIAA Scitech 2019 Forum, pp. 1–17 (2019). https://doi.org/10.2514/6.2019-0724

Jivani, A., Huan, X., Safta, C., Zhou, B.Y., Gauger, N.R.: Uncertainty quantification for a turbulent round jet using multifidelity karhunen-loève expansions. In: AIAA Scitech 2021 Forum, pp. 1–16 (2021). https://doi.org/10.2514/6.2021-1367

Kabelac, S., Kind, M., Martin, H., Mewes, D., Schaber, K., Stephan, P.: VDI-Wärmeatlas, 11 edn. (2013)

Karhunen, K.: Über lineare Methoden in der Wahrscheinlichkeitsrechnung. Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys. **37**, 1–79 (1947)

Langford, E.: Quartiles in elementary statistics. J. Stat. Educ. (2006). https://doi.org/10.1080/10691898.2006.11910589

Le Maître, O.P., Reagan, M.T., Najm, H.N., Ghanem, R.G., Knio, O.M.: A stochastic projection method for fluid flow. II. Random process. J. Comput. Phys. **181**(1), 9–44 (2002). https://doi.org/10.1006/jcph.2002.7104

Le Maître, O.P., Knio, O.M., Najm, H.N., Ghanem, R.G.: Uncertainty propagation using Wiener–Haar expansions. J. Comput. Phys. **197**(1), 28–57 (2004). https://doi.org/10.1016/j.jcp.2003.11.033

Le Maître, O.P., Reagan, M.T., Debusschere, B., Najm, H.N., Ghanem, R.G., Knio, O.M.: Natural convection in a closed cavity under stochastic non-boussinesq conditions. SIAM J. Sci. Comput. **26**(2), 375–394 (2005). https://doi.org/10.1137/S1064827503422853

Liao, Q., Zhang, D.: Probabilistic collocation method for strongly nonlinear problems: 3. Transform by time. Water Resour. Res. **52**(3), 2366–2375 (2016). https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Loève, M.: Probability Theory. Vol. II, 4th Ed. Graduate Texts in Mathematics., 4th edn. Springer, New York, NY (1978)

Nicoud, F., Ducros, F.: Subgrid-scale stress modelling based on the square of the velocity. Flow Turbul. Combust. **62**, 183–200 (1999). https://doi.org/10.1023/A:1009995426001

OpenFOAM documentation v2006 (2019)

Rakhimov, A.C., Visser, D.C., Komen, E.M.J.: Uncertainty quantification method for CFD validated for turbulent mixing experiments from GEMIX. Nucl. Eng. Des. **358**, 110444 (2020). https://doi.org/10.1016/j.nucengdes.2019.110444

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S.: Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. **181**(2), 259–270 (2010). https://doi.org/10.1016/j.cpc.2009.09.018

Schenk, C.A., Schuëller, G.I.: Uncertainty Assessment of Large Finite Element Systems, 1st edn Springer, Berlin Heidelberg (2005). https://doi.org/10.1007/11673941

Smolyak, S.A.: Quadrature and interpolation formulas for tensor products of certain classes of functions. Dokl. Akad. Nauk SSSR **4**, 240–243 (1963)

Sobol', I.M.: Sensitivity estimates for nonlinear mathematical models. Math. Model. Comput. Exp. **1**, 407–414 (1993)

Sobol, I.M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simul. **55**, 271–280 (2001). https://doi.org/10.1016/S0378-4754(00)00270-6

The MathWorks Inc.: MATLAB version: 9.7.0 (R2019b)—ksdensity function, Natick, Massachusetts, US (2019). https://de.mathworks.com/help/stats/ksdensity.html

The MathWorks Inc.: MATLAB version: 9.7.0 (R2019b)—quantile function, Natick, Massachusetts, US (2019). https://de.mathworks.com/help/matlab/ref/quantile.html

Trias, F.X., Soria, M., Oliva, A., Pérez-segarra, C.D.: Direct numerical simulations of two- and three-dimensional turbulent natural convection flows in a differentially heated cavity of aspect ratio 4. J. Fluid Mech. **586**, 259–293 (2007). https://doi.org/10.1017/S0022112007006908

Trias, F.X., Gorobets, A., Soria, M., Oliva, A.: Direct numerical simulation of a differentially heated cavity of aspect ratio 4 with Rayleigh numbers up to 1011—Part I: numerical methods and time-averaged flow. Int. J. Heat Mass Transf. **53**(4), 665–673 (2010). https://doi.org/10.1016/j.ijheatmasstransfer.2009.10.026

Trias, F.X., Gorobets, A., Soria, M., Oliva, A.: Direct numerical simulation of a differentially heated cavity of aspect ratio 4 with Rayleigh numbers up to 1011—Part II: Heat transfer and flow dynamics. Int. J. Heat Mass Transf. **53**(4), 674–683 (2010). https://doi.org/10.1016/j.ijheatmasstransfer.2009.10.027

Vapnik, V.N.: The Nature of Statistical Learning Theory, 2nd edn. Springer, New York, NY (2000). https://doi.org/10.1007/978-1-4757-3264-1

Weller, H.G., Tabor, G., Jasak, H., Fureby, C.: A tensorial approach to computational continuum mechanics using object-oriented techniques. Comput. Phys. **12**(6), 620 (1998). https://doi.org/10.1063/1.168744

Wenig, P.J., Ji, R., Kelm, S., Klein, M.: Uncertainty quantification for the buoyancy-driven mixing process between two miscible fluids using multifidelity polynomial chaos expansions. In: ETMM-13 (2021)

Wenig, P.J., Kelm, S., Klein, M.: Uncertainty Quantification of LES for buoyancy-driven mixing processes using PCE-HDMR. In: Proceedings of DLES-13 (2022)

Wenig, P.J., Ji, R., Kelm, S., Klein, M.: Towards uncertainty quantification of LES and URANS for the buoyancy-driven mixing process between two miscible fluids—differentially heated cavity of aspect ratio 4. Fluids **6**, 161 (2021). https://doi.org/10.3390/FLUIDS6040161

Wenig, P.J., Kelm, S., Klein, M.: CFD uncertainty quantification using stochastic spectral methods–exemplary application to a buoyancy-driven mixing process. Nucl. Eng. Des. **409**, 112317 (2023). https://doi.org/10.1016/j.nucengdes.2023.112317

Wiener, N.: The homogeneous chaos. Am. J. Math. **60**(4), 897–936 (1938). https://doi.org/10.2307/2371268

Witteveen, J.A.S., Bijl, H.: Modeling arbitrary uncertainties using gram-schmidt polynomial chaos. In: Collection of Technical Papers—44th AIAA Aerospace Sciences Meeting **14** (2006). https://doi.org/10.2514/6.2006-896

Xiu, D.: Efficient collocational approach for parametric uncertainty analysis. Commun. Comput. Phys. **2**(2), 293–309 (2007)

Yue, X., Zhang, J., Gong, W., Luo, M., Duan, L.: An adaptive PCE–HDMR metamodeling approach for high-dimensional problems. Struct. Multidiscip. Optim. (2021). https://doi.org/10.1007/s00158-021-02866-7