

A Review of Recent Advances in Surrogate Models for Uncertainty Quantification of High-Dimensional Engineering Applications

Zeynab Azarhoosh, Majid Ilchi Ghazaan *

School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran

HIGHLIGHTS

- The concept of high-dimensionality is discussed from different perspectives.
- Challenges in surrogate modeling for high-dimensional spaces are comprehended.
- A literature review is conducted addressing the challenges to quantify uncertainty.
- Solutions are defined into dimension reduction, multi-fidelity model, and sampling.
- High-dimensional benchmark functions assessing the surrogate models are provided.

ARTICLE INFO

Key words:

Uncertainty quantification
Uncertainty propagation
Surrogate modeling
High-dimensionality
Curse of dimensionality

ABSTRACT

In fields where predictions may have vital consequences, *uncertainty quantification* (UQ) plays a crucial role, as it enables more accurate forecasts and mitigates the potential risks associated with decision-making. However, performing uncertainty quantification in real-world scenarios necessitates multiple evaluations of complex computational models, which can be both costly and time-consuming. To address these challenges, surrogate models (also known as meta-models)—which are low-cost approximations of computational models—can be an influential tool. Nonetheless, as the complexity of the problem increases and the number of input variables grows, the computational burden of constructing an efficient surrogate model also rises, leading to the so-called *curse of dimensionality* in *uncertainty propagation* from inputs to outputs. Additionally, dealing with constraints, ensuring the robustness and generalization of surrogate models across different inputs, and interpreting the output results can present significant difficulties. Therefore, techniques must be implemented to enhance the performance of these models. This paper reviews the developments of the past years in surrogate modeling for *high-dimensional* inputs, with the goal of quantifying output uncertainty. It proposes general approaches, including dimension reduction techniques, multi-fidelity surrogate models, and advanced sampling schemes, to overcome challenges in various practical problems. This comprehensive study provides an initial guide for effective surrogate modeling in engineering practices by outlining key components of solving algorithms and screening mathematical benchmark functions, all while ensuring sufficient accuracy for overall predictions. Additionally, this study identifies research gaps, suggests future directions, and describes the applications of the proposed solutions.

* Corresponding author: M. Ilchi Ghazaan, Telephone: +98 21 73227133; Fax: +98 21 77240398
E-mail address: ilchi@iust.ac.ir (M. Ilchi Ghazaan).

1. Introduction

Discrepancies between intended and actual systems are always possible. This discrepancy, referred to as uncertainty, can arise from the inherent randomness of the system being measured or modeled, such as measurement errors or natural variability. Alternatively, it can result from limited knowledge about the system, including insufficient empirical data, simplifications, or a lack of understanding of complex physical phenomena [1]. In the literature, various probabilistic and non-probabilistic methods have been employed to model uncertainties in engineering problems [2,3] in order to measure the proximity between predictions and observed values through a logical process. This process, known as uncertainty quantification (UQ), serves multiple essential purposes. It is used to assess the reliability of computational predictions, validate and compare scientific and engineering models, optimize robust designs under uncertainty, and support technical decision-making through computational predictions. Furthermore, UQ is a desired tool for data assimilation, constructing surrogate models, and establishing correlations between multi-scale and multi-physics models [4]. It also plays a pivotal role in developing numerical modeling strategies that optimally utilize computational models while addressing variability and input uncertainties. This role is especially significant when employing a computational model to propagate information about high-dimensional input parameters to estimate the uncertainty of the system's response. Various numerical methods have been developed to propagate uncertainty as an important component of uncertainty analysis [2].

Meanwhile, the measurable response of a system as a function of input parameters can be predicted using computational models, which can be thought of as input-output mappings. For real-world systems, quantifying uncertainty is often costly and time-consuming because it typically involves numerous evaluations of complex computational models, even with recent advances in computer hardware and processing capacity [5]. Moreover, several engineering fields such as mechanical, civil, electrical, aerospace, material, chemical, petroleum, hydrologic, environmental, geological, marine, etc. frequently encounter high-dimensional problems that pose significant computational burdens due to numerical simulation or/and data processing. They often involve a large number of inputs and responses in complex systems with little or no information about their internal performance [6]. Advanced techniques are required to manage these challenges, as traditionally solving such problems has been difficult or impossible by repeatedly running computational models. These techniques are essential for effectively managing and interpreting large volumes of data to predict system behavior and make well-informed decisions regarding design, maintenance, and operation.

By employing surrogate models as effective approximations of computational models, predictions from simulations can be computed with a limited amount of data or fewer executions of the original model. Using these models instead of complex computational models significantly increases system processing speed. Accordingly, employing surrogate models in uncertainty quantification is one of its common applications [1]. The schematic diagram in Fig. 1 illustrates the role of surrogate modeling in uncertainty propagation.

Constructing surrogate models for accurate predictions typically requires smooth relationships between inputs and outputs, which may not be practical or accurately reflect the complexity of high-dimensional real-world applications [5]. Therefore, developing surrogate models for engineering problems where input parameters significantly influence computational complexity is a fundamental challenge. These problems require solutions for developing appropriate surrogate models and addressing various challenges.

On one hand, these challenges stem from the sparsity and noise in the dataset [7] and the difficulty in identifying meaningful patterns and relationships for surrogate modeling. On the other hand, issues such as difficulty in space-filling for sampling [8-10], the risk of overfitting and underfitting [11], and the high costs associated with processing, storing, and transferring data make these

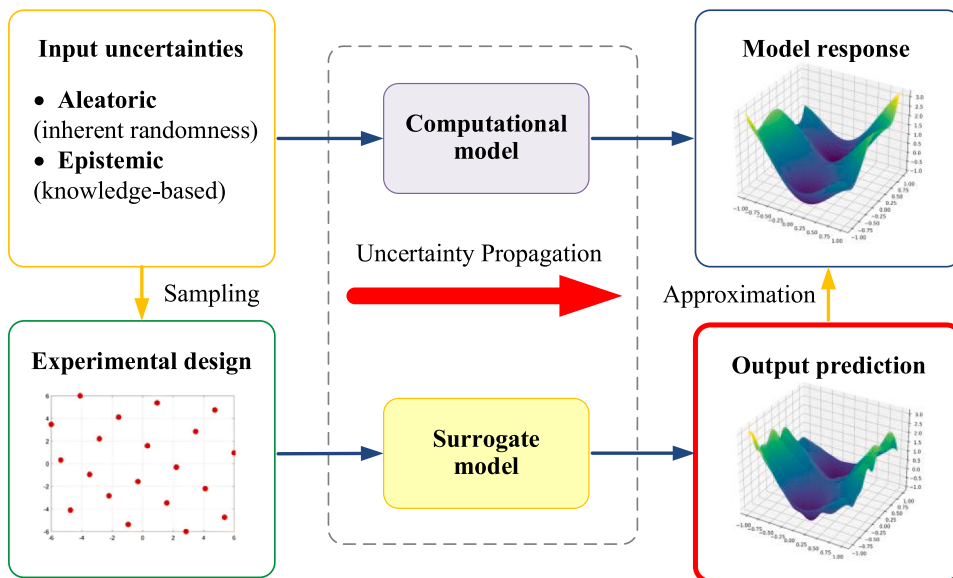


Fig. 1. - Schematic diagram of the surrogate modeling approach for uncertainty propagation

challenges inevitable. The challenges are broadly classified as follows:

- **The curse of dimensionality**, which refers to the extremely rapid growth in the complexity of solving problems as the number of variables increases [12]. This phenomenon creates a significant obstacle in the study of high-dimensional systems [13]. In such systems, the dimensionality results in an exponential increase in the input space volume needed to fit an accurate surrogate model [6]. As the number of input variables grows, especially in the context of uncertainty propagation, the demand for evaluations required to construct a surrogate model also increases, often making it practically impossible to solve these problems [14].
- **Dataset collection**: Due to the complexity of computational simulations in high-dimensional systems, collecting a comprehensive set of input-output data to construct accurate surrogate models can be costly and time-consuming [15]. This makes the development and validation of surrogates more intricate. The computational burden of creating efficient surrogate models in high-dimensional spaces can be so significant that it sometimes renders surrogate modeling unfeasible for these problems [12].
- **Model accuracy**: The accuracy of surrogate models depends on the quality and quantity of the dataset [16-20]. In high-dimensional systems, the computational costs of collecting sufficient data can make evaluating the surrogate model challenging [7]. Furthermore, the sparsity or outliers of some samples from the main part of the dataset [21] and the presence of noise [22] can cause local errors to be overlooked.
- **Model robustness**: Surrogate models must be robust to variations in input parameters while accurately predicting responses across a wide range of inputs. Owing to the nonlinearity of high-dimensional systems, even small perturbations in input parameters can significantly affect the model response. Therefore, it is crucial to develop surrogate models that are not overly sensitive to changes in input data and noise [23].
- **Model interpretability**: Understanding the behavior of surrogate models in high dimensions can be highly complex [24], making model interpretability particularly challenging, especially when propagating uncertainties from input variables to outputs. This complexity significantly limits the utility of surrogate models in supporting decision-making [25].
- **Physical constraints**: Surrogate modeling in high-dimensional spaces, especially when subject to physical constraints, is crucial for achieving accurate predictions that closely align with real-world behaviors [26]. By increasing the search space, it is not easy to incorporate the equations that fulfill the physical constraints into the surrogate models in a way that makes the predictions precisely satisfy the constraints [27].
- **Invariance properties**: Surrogate models must preserve invariance properties, ensuring consistency and accurate predictions even when the search space undergoes scaling, transformation, or rotation. However, maintaining these properties in high-dimensional systems poses significant challenges [28].
- **Uncertainty criteria**: The high dimensionality of the input space poses a significant obstacle to uncertainty quantification analysis [13]. Surrogate models are tasked not only with predicting the system response but also with providing uncertainty criteria that demonstrate the reliability of these predictions. Consequently, accurately estimating uncertainties associated with predictions in high-dimensional spaces is inherently challenging.

Despite the lack of earlier attempts, several scholars have recently focused on constructing effective surrogate models to quantify uncertainties in high-dimensional systems. An overview of some studies conducted in the past years is presented in Table 1. The solutions proposed in these studies fall into four main categories: (1) dimension reduction techniques (using either a sequential or joint approach to surrogate modeling); (2) multi-fidelity surrogate models; (3) advanced sampling schemes; and (4) combined methods. In addition to these four general approaches, various strategies have been proposed to address additional challenges encountered during the development of surrogate models. These strategies aim not only to decrease computational costs and enhance prediction performance but also to facilitate pattern recognition and accurate response prediction. The integration of these strategies with machine learning algorithms in high-dimensional problems has made the results of uncertainty quantification easier to interpret. However, all of these approaches come with their own set of difficulties. As a result, ongoing research continues to develop more broadly applicable techniques and offers solutions to limitations related to a variety of issues.

In addition to effectively guiding future research, the comprehensive overview of research gaps and solution approaches provided in this review can serve as a foundation for developing algorithms to create efficient surrogate models for quantifying uncertainty in high-dimensional systems. Evaluating different approaches when applied to mathematical benchmark functions can help determine the optimal method for surrogate modeling in engineering applications.

This paper is organized as follows. Section 2 begins with a discussion of researchers' perspectives on the concept of high-dimensionality. Following that, Section 3 describes conventional surrogate models and their theoretical foundations, and discusses the challenges of constructing surrogate models for high-dimensional inputs. Section 4 explores the main approaches suggested in the literature to address these issues and reviews the state-of-the-art applications of these approaches for uncertainty quantification. Section 5 offers a classification of high-dimensional mathematical benchmark functions for quantifying response uncertainty. Finally, Section 6 concludes the study. Fig. 2 depicts an overview of this comprehensive study.

2. High dimensionality

Real-world systems often involve inputs that are random processes or require a large number of parameters to identify [106]. These systems are termed high-dimensional due to the significant number of input variables, which greatly impacts the computational complexity involved in addressing them. To develop surrogate models that accurately predict outputs and quantify response uncertainties, it is essential to understand the concept of high-dimensional spaces [107]. This section explores the concept of

Table 1

- Proposed techniques aiming to construct surrogate models for uncertainty quantification in high-dimensional problems

Reference	Year	Application/Dataset	Goal/Objective	Challenges	Techniques
Hombal and Mahadevan [10]	2013	3D crack growth in a mechanical component under various load histories	Assess performance of model for life prediction	Curse of dimensionality and model robustness	Principal component analysis (PCA) together Gaussian process regression
Doostan et al. [29]	2013	A manufactured function, elliptic stochastic equation, and hydrogen oxidation problem	Uncertainty quantification	Curse of dimensionality	Low-rank approximation
Kubicek et al. [30]	2015	Borehole problem	Sensitivity analysis	High dimensionality and sensitivity indices	Cut-HDMR (high dimension model representation) and sampling strategy
Chen and Quarteroni [15]	2015	Heat diffusion in thermal blocks and groundwater flow through porous medium	Uncertainty quantification	Curse of dimensionality	Adaptive and reduced algorithm in combination with sparse grid approximation
Kersaudy et al. [31]	2015	Three benchmark functions and the fetus-exposure problem	Global sensitivity analysis	High dimensionality and sensitivity indices	Least-angle regression (LARS)-Kriging-Polynomial chaos (PC)
Jakeman et al. [32]	2015	Corner-peak function, random oscillator, diffusion equation, and resistor network	Uncertainty quantification	High dimensionality and model accuracy	Adaptively find coefficients of polynomial chaos expansions
Konakli and Sudret [33, 34]	2015, 2016	Sobol function, beam deflection, 2D heat diffusion model, and 2D truss structure	Global sensitivity analysis	High dimensionality, model accuracy, and sensitivity indices	Low-rank tensor approximation
Konakli and Sudret [35]	2016	2D truss structure, 2D heat diffusion model, and a 3-span and 5-story frame structure	Reliability analysis	High dimensionality and model accuracy	Low-rank tensor approximation
Demian et al. [36]	2016	2D multi-layered hydrogeological model	Global sensitivity analysis	Sensitivity indices	Sparse polynomial chaos expansion
Liang and Mahadevan [37]	2016	A mathematical problem and an aircraft wing model	Uncertainty propagation in multidisciplinary analysis	High dimensionality and model accuracy	Bayesian network, copula-based sampling, and principal component analysis (PCA)
Papaioannou et al. [38]	2016	Convex, concave, noisy limit-state functions, series system reliability problem, and a linear function	Structural reliability analysis	High dimensionality	Sequential importance sampling
Shin and Xiu [39]	2016	Genz test functions and Stochastic diffusion equation	Stochastic collocation for uncertainty quantification	Data collection	Quasi-optimal sampling
Tripathy et al. [40]	2016	Projection matrix, elliptic partial differential equation, and 1D granular crystals	Uncertainty propagation	High dimensionality and model robustness	Gradient-free approach for dealing with noisy outputs, Gaussian process regression with dimensionality reduction
Schöbi et al. [41]	2016	Four-branch function, Borehole function, and 2D truss structure	Structural reliability analysis (rare event)	High dimensionality and model accuracy	Polynomial-Chaos Kriging approach coupled with an active learning algorithm (AK-MCS)
Perdikaris et al. [42]	2016	Borehole function, elliptic problem subjected to random forcing, and Sobol function	Uncertainty quantification of dynamical systems	Massive data and high dimensionality	Multi-fidelity model, dimensionality reduction, and Gaussian process
Rushdi et al. [43]	2017	Herbie, Smooth Herbie, circular cone, planar cross, and trench functions, and an integral of a high-dimensional function	Numerical integration and failure probability estimation	Curse of dimensionality and model accuracy	Voronoi Piecewise Surrogate (VPS) model
Fajraoui et al. [14,44]	2017	Ishigami function, Sobol function, 2D truss structure, and 1D diffusion problem	Uncertainty quantification	High dimensionality and model accuracy	Adaptive sequential sampling to generate sparse polynomial chaos expansions
Marelli and Sudret [45]	2017	Four-branch function and slope stability problem	Structural reliability analysis	High dimensionality and small initial data	Active polynomial-chaos-bootstrap Monte Carlo simulation (APCB-MCS)
Soize and Ghanem [46]	2017	Two datasets and a petro-physics database	Uncertainty quantification	High dimensionality, large dataset, and model robustness	Generating samples to construct polynomial chaos expansion on manifold
Hadidi et al. [47]	2017	Three 2D non-structural examples, a 23-rod truss and a 2-span 6-story frame structure	Structural reliability analysis	High dimensionality and model accuracy	Selection of sample point for exponential response surface method
Mai [48]	2018	Dynamical systems including Duffing oscillator, Bouc-Wen oscillator, and steel frame	Uncertainty quantification in earthquake engineering (seismic fragility curves)	High dimensionality and model accuracy	Polynomial chaos - nonlinear autoregressive with exogenous input (PC-NARX) model
Xiao et al. [49]	2018	A system with four performance functions, nonlinear oscillator, and a cantilever tube	Structural reliability analysis	High dimensionality, data availability, and model accuracy	Neural network models and adaptive sequential sampling

(continued on next page)

Table 1 (continued)

Reference	Year	Application/Dataset	Goal/Objective	Challenges	Techniques
Hadigola and Doostan [50]	2018	Manufactured functions, nonlinear duffing oscillator, and life of batteries	Uncertainty propagation	High dimensionality	Optimal sampling for least-squares polynomial chaos expansion
Sadoughi et al. [51]	2018	Four mathematical functions and piezoelectric energy harvester	Reliability analysis	High dimensionality and model accuracy	Kriging-based sequential sampling
Thimmisetty et al. [52]	2018	Standard well log data	Geological (seismic wave propagation) properties	High dimensionality	Manifold learning, Gaussian process regression, and diffusion maps
Tripathy and Bilonis [53]	2018	Stochastic elliptic partial differential equation	Uncertainty propagation	High dimensionality	Deep neural network surrogate model
Bassamzadeh and Ghanem [54]	2018	A real wellbore dataset from the Gulf of Mexico	Uncertainty propagation through complex system	High dimensionality and limited data	Bayesian network surrogate model
Marelli and Sudret [55]	2018	Four-branch function, 2D truss structure, and a 3-span and 5-story frame structure	Structural reliability analysis	High dimensionality and small initial experimental design	An active-learning combined with sparse polynomial chaos expansions and bootstrap resampling
Yang and Perdikaris [56]	2019	Data including random noise, a synthetic example, and Burger's equation	Uncertainty propagation in dynamical systems	High dimensionality and model robustness	Conditional deep neural network surrogate model and multi-fidelity systems
Li et al. [57,58]	2019 2020	Coastal regions in San Francisco Bay and a complex analytical function	Sensitivity analysis	High dimensionality and sensitivity indices	Dimension reduction and Kriging based sensitivity analysis
Lataniotis [6]	2019	Sobol function, electrical resistor networks, 2D heat diffusion, and a wind turbine	Approximate entire response PDF	High dimensionality	Dimensionality reduction together with surrogate modeling (DRSM)
Sheikholeslami et al. [59]	2019	Sobol function and hydrologic model (MESH)	Global sensitivity analysis	High dimensionality, model robustness, and sensitivity indices	Grouping input factors
Zhu et al. [26]	2019	Steady-state flow in random heterogeneous media	Uncertainty propagation	Physical constraints	Convolutional encoder-decoder neural network physics-constrained models
Cheng and Lu [60]	2020	Series system with four branches, 2-bar supporting structure, 25-bar truss, and heat conduction problem	Structural reliability analysis	Complex systems and model accuracy	Ensemble of surrogate models
Xu and Wang [61, 62]	2020	A linear function of independent random variables	Reliability analysis (rare event)	High dimensionality and model accuracy	Sequential importance Sampling
Ehre et al. [63]	2020	A 23-rod truss and 2D steel plate	Global sensitivity analysis	High dimensionality and sensitivity indices	Partial least squares-driven polynomial chaos expansion (PLS-PCE)
Vohra et al. [7]	2020	Residual stress in a manufactured component	Global sensitivity analysis and reliability analysis	High dimensionality in inputs and field output and model accuracy	Combination of principal component analysis and active subspaces (PCAS)
Zhou et al. [64]	2020	A cantilever tube structure, heat diffusion problem, and an isotropic plate	Approximate model response	Curse of dimensionality	Embed dimension reduction into polynomial chaos expansion
Savin and Hantrais-Gervois [65]	2020	An aircraft wing model	Approximate response PDF and sensitivity indices	Curse of dimensionality and sensitivity indices	Non-intrusive technique to construct sparse polynomial surrogates
Zhou and Lu [66]	2020	Rackwitz function, Morris' function, and an aero-engine compressor disc	Uncertainty quantification	High dimensionality, model accuracy and robustness	Kriging model combined with dimension reduction
Uribe et al. [67]	2021	Linear limit-state function, quadratic limit-state function, and 2D steel plate	Reliability analysis (rare event)	High dimensionality and model accuracy	Cross-entropy-based importance sampling
Xu and Wang [68]	2021	Griewank function	Uncertainty analysis	High dimensionality and training sample size	Kriging model setting with squared exponential Kernel and distance metrics
Peng et al. [69]	2021	30D and 60D analytical functions as well as 25-DOF lumped-mass nonlinear frame structure under seismic excitations	Structural reliability analysis	Curse of dimensionality and model accuracy	Active learning sampling combined with kernel principal component analysis, and Gaussian process regression (AL-KPCA-GPR)
Guo et al. [17]	2021	An aircraft fuselage panel	Uncertainty quantification	High dimensionality and model accuracy	Dimension reduction of the input and output space to construct Gaussian surrogate model
Kontolati et al. [5]	2022	Dielectric cylinder in homogeneous electric field, Lotka-Volterra dynamical system, and	Uncertainty quantification	High dimensionality and out-of-sample predictions	Grassmannian diffusion maps and manifold learning-based

(continued on next page)

Table 1 (continued)

Reference	Year	Application/Dataset	Goal/Objective	Challenges	Techniques
Wycoff et al. [70]	2022	advection-diffusion-reaction equations Test functions, high-dimensional observational datasets, and General Motors problem	Sensitivity analysis	High dimensionality	polynomial chaos expansions (GD Maps PCE) Input warping and local Gaussian process modeling
Kapusuzoglu et al. [18]	2022	Seven benchmark functions and a gas turbine engine blade	Spatio-temporal multi-physics dynamic system analysis	High dimensionality and model accuracy	Dimension reduction of outputs and subsequent adaptive sampling
Kapusuzoglu et al. [16]	2022	An aircraft fuselage panel and a gas turbine engine blade	Spatio-temporal multi-physics dynamic system analysis	High dimensionality and model accuracy	Dimension reduction of both the input and output space to construct Kriging
Kontolati et al. [71]	2022	1D stochastic Poisson's equation, 2D stochastic steady-state heat equation, and 2D time-dependent stochastic Brusselator model	Uncertainty propagation	Curse of dimensionality	Dimension reduction and polynomial chaos expansions
Shustin et al. [72]	2022	Three datasets in the context of machine learning algorithms and two datasets for solving PDEs (Pyro and Chaospy packages)	Uncertainty propagation	Curse of dimensionality	A neural network (autoencoder) dimensionality reduction combined with polynomial chaos expansion
Hong et al. [73]	2022	A four-branch function, a truss structure, a cantilever tube, and a journal bearing of gear pump	Structural reliability analysis	Model accuracy	Surrogate models (Kriging and radial basis function) based active learning
Zhou et al. [21]	2022	Three numerical examples, a simplified wing box structure, and a structure in composite material	Reliability analysis	Model accuracy	Adaptive ensemble of surrogate models (Kriging, support vector machine, and radial basis function) based on hybrid measure (AESMHM)
Liu et al. [74]	2022	75D time-variant system, a cantilever tube model, a heat conduction model, and an oscillator model	Uncertainty propagation	High dimensionality	Kriging combined with partial least squares
Zhou and Peng [8]	2022	A 3-bay, 10-story planner steel frame structure	Reliability analysis	Curse of dimensionality and model accuracy	Active learning-based sampling combined with deep learning (autoencoder) and Gaussian process regression (AL-DLGPR)
Navaneeth and Souvik [75]	2022	Sobol function, a composite beam, and 25-element space truss	Structural reliability analysis	High dimensionality	Active subspace combined with sparse polynomial chaos expansion to construct hybrid polynomial correlated function expansion (SAS-HPCFE)
Xu and Wang [76]	2022	A mathematical example and a multi-output with multi-dimensional inputs problem	Uncertainty quantification	Dimensionality and dataset containing missing value	Bayesian Gaussian process latent variable model (BGPLVM) with adaptive sampling
Yin [11]	2022	A parabolic function, a cantilever beam, a 52-bar truss system, a 25-story shear frame structure subjected to stochastic seismic excitation, MNIST dataset, and heat transfer problems	Uncertainty propagation	Curse of dimensionality and model accuracy	(1) Dimension reduction, (2) generalized sliced inverse regression (GSIR), active learning, and importance sampling, (3) convolutional neural network Gaussian process (CNN-GP) regression
Bigoni et al. [77]	2022	Isotropic function, Borehole function, deep composition of function, and a bridge	Uncertainty quantification	Curse of dimensionality	Dimension reduction of adaptive polynomial function space using gradient information
Dhulipala et al. [78]	2022	Four-branch function, Rastrigin function, Borehole function, and two finite element model case studies	Reliability analysis (rare event)	High dimensionality and model accuracy	Active learning with low-fidelity and high-fidelity model
Yildiz et al. [79]	2022	Park 4D function, Borehole Function, Sobol function, and supersonic aircraft design model	Approximate the response PDF	High dimensionality and model accuracy	Multi-fidelity low-rank approximation
Conti et al. [80]	2023	An electrical signal in excitable cells, a fluid flow around a cylindrical obstacle, and Lotka-Volterra system	Construct uncertainty bounds of multi-fidelity approximations	High dimensionality, model accuracy, and data scarcity	Multi-fidelity long short-term memory (LSTM) neural networks model
Ji et al. [81]	2023	A mathematical time-dependent model, a roof truss structure, a beam under stochastic loads, and a stone arch bridge under hurricane loads	Time-dependent reliability analysis	Curse of dimensionality and model accuracy	Adaptive dimension reduction strategy and Kriging

(continued on next page)

Table 1 (continued)

Reference	Year	Application/Dataset	Goal/Objective	Challenges	Techniques
Shang et al. [82]	2023	Ishigami and Borehole functions, stochastic partial differential equation and 2D airfoil problems	Global sensitivity analysis	High dimensionality and model accuracy	Multi-fidelity Kriging surrogate model
Sun et al. [83]	2023	Soil moisture over the Huaihe river simulation	Uncertainty quantification	High dimensionality and model accuracy	Deep autoregressive neural network
Guo et al. [20]	2023	An aircraft fuselage panel	Evaluate the quality of the surrogate model	High dimensionality in both inputs and outputs and model accuracy	Dimension reduction and Gaussian process model
Cheng and Zimmermann [84]	2023	1D and 2D analytical functions, Borehole, Rosenbrock, and Dixon-Price functions, and an aerodynamic model of airfoil	Global sensitivity analysis	Curse of dimensionality	Sliced gradient-enhanced Kriging (SGE-Kriging)
Liu and Jiang [85]	2023	150D, 200D and 300D mathematical functions and a plane heat conduction problem	Uncertainty propagation	Curse of dimensionality	Deep kernel polynomial chaos expansion (DKPCE)
Schar et al. [86]	2024	A coupled spring-mass system subjected to random excitation and a wind turbine	Approximate the response of complex dynamical systems	High dimensionality and model accuracy	Manifold nonlinear autoregressive with exogenous (mNARX) input
Guo et al. [87]	2024	McComick, Three-hump camel, and Griewank functions, a component with residual stress field	Improve surrogate models for iterative calculations	High dimensionality in both inputs and outputs	Dimension reduction and adaptive learning strategy
Mufti et al. [88]	2024	2D RAE 2822 airfoil and a 3D NASA CRM wing	Uncertainty quantification (error prediction)	Curse of dimensionality	Multi-fidelity methodology for reduced order models (ROM)
Bharti and Ghosh [89]	2024	A mistuned bladed disk problem and a beam on Winkler foundation	Approximate response PDF and failure probability	Stochastic dimensionality	Non-intrusive reduced order model (ROM)
Xian and Wang [90]	2024	A linear elastic cantilever beam, an oscillator with nonlinear viscous damper, and a multi-degree-of-freedom hysteretic system	Rare event simulation	Curse of dimensionality	Coupled physics-data-driven surrogate model
Kim et al. [91]	2024	A nonlinear mathematical function, a space truss structure, and a steel lattice transmission tower	Reliability estimation (rare events)	Curse of dimensionality	Adaptive active subspace-based heteroscedastic Gaussian process
Li and Montomoli [92]	2024	1D function with linear/non-linear correlation, 32D function, 100D function, and turbine nozzle flow	Approximate response PDF	Curse of dimensionality and limited amount of high-fidelity data	Multi-fidelity deep neural networks
Zhou et al. [93]	2024	Sobol function, steady-state diffusion problem, and lid driven cavity problem	Approximate response PDF	Curse of dimensionality	An active-subspace-enhanced support vector regression model
He et al. [94]	2024	Three synthetic functions, A cantilever tube structure, a 72-bar space truss, and a box structure	Approximate response PDF	High dimensionality	An adaptive data-driven subspace polynomial dimensional decomposition
Song et al. [95]	2024	Dixon-Price, Griewank, Ackley, Rosenbrock, and Morris functions, a cantilever beam, a truss bridge, and a hybrid joint structure	Approximate response PDF	Curse of dimensionality	Improved sufficient dimension reduction-based Kriging modeling method (KISDR)
Shi et al. [96]	2024	Messerschmitt-Bölkow-Blohm beam and compliant force inverter	Approximate response PDF	High dimensionality	Deep convolutional dimension reduction network (ConvDR)
Kim et al. [97]	2024	A uniaxial linear elastic bar and a 5-story shear building under stochastic excitation	Approximate response PDF	Curse of dimensionality	Dimensionality reduction-based surrogate modeling (DR-SM)
Wang et al. [98]	2024	Nonlinear piston simulation model and air-taxi trajectory optimization scenario	Uncertainty quantification (prediction of relative error)	Curse of dimensionality	Combining active subspace with non-intrusive polynomial chaos (NIPC) and accelerated model evaluations on Tensor grids using computational graph transformations (AMTC)
Zanoni et al. [99]	2024	A simple example and two high-dimensional functions	Uncertainty propagation (distribution of estimators)	Computationally expensive models	Multi-fidelity Monte Carlo with active subspace (MFMC-AS) and multi-fidelity Monte Carlo with autoencoder (MFMC-AE)
Gong and Pan [100]	2024	Forrester function, 2D stochastic oscillator, 8D borehole	Quantify rare event statistics	High dimensionality	Multi-fidelity Bayesian experimental design

(continued on next page)

Table 1 (continued)

Reference	Year	Application/Dataset	Goal/Objective	Challenges	Techniques
Loukrezis et al. [101]	2024	hydrological model, and ship motion in waves A simply supported beam, an induction motor, and an electrical power grid model	Model efficiency and accuracy, and uncertainty estimation accuracy	Curse of dimensionality	Multivariate sensitivity-adaptive polynomial chaos expansion (MVSA PCE)
Catalanotti [102]	2024	A 40-D problem with open hole tension and compression specimens	Approximate response PDF	Lack of knowledge on distribution, model efficiency and accuracy	Combining bootstrapping and Bayesian analysis
Baisthakur and Fitzgerald [103]	2024	A multi-body dynamic numerical model of the IEA-15MW wind turbine	Evaluations of response	High dimensionality and model accuracy	Physics-Informed Neural Network (PINN) surrogate model
Nguyen et al. [104]	2024	Three reliability benchmarks, a hyperstatic truss, a 1-story frame structure with a viscous damper under ground motions, and a post-tensioned reinforced concrete beam	Structural reliability analysis	High dimensionality	Active learning framework using uncertainty quantification and flexible meta-models (AUQ-meta)
Wan et al. [105]	2024	Multi-sharp corner example, a nonlinear exponential function, Rackwitz function, a classic oscillator problem, an arc-stiffened plate, and a carriage structure	Structural reliability analysis	High dimensionality and model accuracy	Combining subset simulation and multi-class adaptive support vector machine (SS-MASVM)

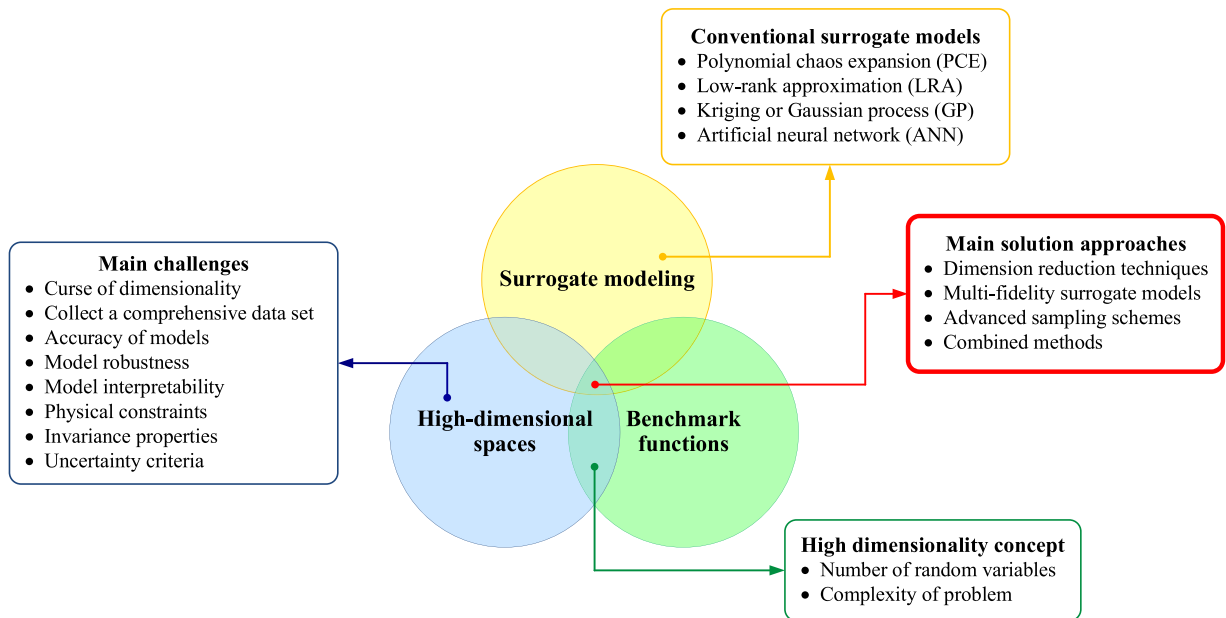


Fig. 2. - Overview of the comprehensive study

high-dimensionality from various perspectives.

When addressing problems involving a large number of variables, these variables can be classified into two main categories based on their characteristics: unstructured and structured [6]. Unstructured inputs lack inherent order and are typically identified by model parameters. The dimensionality of these inputs is generally within the range of 10^{0-2} in practical applications. An example of unstructured input is constant concentrated loads applied to a numerical model. In contrast, structured inputs are associated with an inherent order or a strong correlation between a set of meaningful physical coordinates. The dimensionality of this type of input typically ranges from 10^{2-6} in most practical problems. Examples of structured inputs include time-dependent excitations applied to dynamic systems [86] and data-driven processes such as time series or spatial data [6]. The varied ability of different surrogate modeling methods to organize these two types of input variables motivates this classification. Therefore, selecting a suitable surrogate model for high-dimensional spaces may be influenced by the type and number of input variables. The upcoming sections will introduce conventional surrogate modeling methods developed for these two categories.

Following this preface, high dimension generally refers to problems with numerous independent random input variables [6,29]. Alternatively, high-dimensionality can indicate a feature or covariate space with a dimensionality greater than or equal to the number of samples (observations) [108,109], consistent with its definition in field of statistics [110]. Some engineering literature also considers problems with 10 or more input variables as high-dimensional [45], while those with fewer than 10 are classified as low-dimensional [111]. When assessing the capabilities of surrogate models for high-dimensional versus low-dimensional problems, a common boundary is often set at 20 input variables for high-dimensionality and 10 for low-dimensionality [111].

It is noteworthy that while some studies interpret high dimensionality as having dimensions (d) much greater than 1 ($d \gg 1$) [1], as mentioned previously, the substantial dimensionality of such problems significantly affects their computational complexity. This complexity arises from the intricate relationships required to accurately describe system behaviors in these problems, which are difficult to express using fundamental physical or mathematical principles. Consequently, the precise definition of high dimensionality is highly problem-dependent. In other words, the boundary for classifying a problem as high-dimensional is not fixed and can vary depending on the context. As a result, a clear and consistent definition of high dimensions is lacking in the engineering literature. What is crucial is the researcher's understanding of the complexity involved in solving a given problem. Examining datasets used in the literature often leads to an interpretation of high dimensionality. For instance, some problems may be considered high-dimensional with as few as 3 to 5 input variables due to practical reasons, while the complexity of solving others may reach high-dimensional levels for dimensions greater than or equal to 10 [112]. In brief, without providing a precise mathematical criterion for high dimensions, this study focuses on the practical challenges in dealing with complex systems and adopts an engineering pragmatic perspective to define dimensionality.

3. Surrogate modeling methods

Surrogate models are mathematical constructs that mimic the real behavior of systems using a limited input-output dataset [2]. They are instrumental in reducing the computational burden of primary simulation models and serve as a crucial tool for quantifying response uncertainties [6]. Surrogate models can be represented as either black-box or gray-box functions that map input to output [113]. By accurately approximating these mappings through simplified mathematical models [114,115], surrogates enhance efficiency across a wide range of applications.

As shown in Table 1, researchers have employed a variety of surrogate models to tackle diverse engineering applications, thereby enhancing either the model's flexibility or the interpretability of results in high-dimensional problems. This paper focuses on improving the performance of surrogate modeling in high-dimensional systems. Consequently, this section only provides a brief overview of the fundamentals of surrogate modeling methods frequently applied in advanced forms to quantify uncertainties in high-dimensional spaces. This section aims to foster a realistic understanding of the challenges associated with developing surrogate models—particularly the curse of dimensionality—while also introducing some commonly used surrogates as examples for scenarios involving numerous variables.

3.1. Polynomial chaos expansions (PCE)

Polynomial Chaos Expansion (PCE) represents a type of surrogate modeling that is widely employed for quantifying uncertainties in engineering problems due to its flexibility and interpretability [6,116]. PCE provides a spectral decomposition of the model response into a series of orthogonal polynomials with respect to the random input variables [45]. This expansion allows for the discovery of the correlation between different input parameters and their effect on the model response [117]. A literature survey on high-dimensional systems demonstrates that developed PCEs have been effectively utilized for sensitivity analysis and the propagation of uncertainty in structured inputs [6,46]. However, PCEs are rarely used in reliability analysis because they have low accuracy in estimating the extreme values of the model's response distribution [45,55].

Equation (1) presents the polynomial chaos expansion $M(X)$ for the d -dimensional input vector X , which includes random variables with independent components defined by the joint probability density function (PDF) f_X , and the output Y [6,118,119].

$$Y = M(X) = \sum_{\alpha \in \mathbb{N}^d} \theta_\alpha \Psi_\alpha(X) \quad (1)$$

where $\Psi_\alpha(X)$ represents multivariate orthogonal polynomials with regard to f_X , $\alpha \in \mathbb{N}^d$ is the multi-index identifying the components of the multivariate polynomials Ψ_α in each of the input variables X_i , and $\theta_\alpha \in \mathbb{R}$ are PCE coefficients.

Equation (2) determines the polynomial $\Psi_\alpha(X)$, which is created using the tensor product of the univariate orthogonal polynomials $\phi_k^{(i)}(x_i)$ concerning the marginal probability density function (PDF) of the corresponding variable [14].

$$\Psi_\alpha(X) \equiv \prod_{i=1}^M \phi_k^{(i)}(x_i) \quad (2)$$

Equation (1) can be truncated after a limited number of terms for computational purposes. In the case, the approximate polynomial expansion can be obtained using Equation (3) [14].

$$Y = M(X) \approx \sum_{\alpha \in \mathcal{A}} \theta_\alpha \Psi_\alpha(X) \equiv \theta^T \Psi(X) \quad (3)$$

$\mathcal{A} \subset \mathbb{N}^d$ in Equation (3) is a set of selected multivariate polynomials up to order p . $\mathcal{A}$ can also be interpreted as a set of a limited number of cardinality indices $\text{Card } \mathcal{A} \equiv P = \binom{d+p}{p}$ [14].

Different intrusive [120,121] and non-intrusive [122-129] approaches have been suggested to derive PCE coefficients θ_α [6] in the literature. Non-intrusive methods investigate PCE coefficients on a set of N samples, known as the experimental design [14]. Therefore, many large-scale, complex systems and/or scarce datasets prefer to utilize these methods over intrusive ones [6,29].

As an example, one of the non-intrusive approaches, known as least-squares minimization (or regression), is introduced. In this method, the coefficients θ for the model output set $\mathcal{Y} = \{y^{(1)}, \dots, y^{(N)}\}$, with respect to the input set in the experimental design space $\mathcal{X} = \{x^{(1)}, \dots, x^{(N)}\}$, are determined by minimizing the expected value of the least square residual using Equation (4) [6].

$$\theta = \underset{\theta \in \mathcal{J}_{\theta, \mathcal{A}}}{\operatorname{argmin}} \mathbb{E} \left[(\theta^T \Psi(\mathcal{X}) - \mathcal{Y})^2 \right] \quad (4)$$

where the model matrix Ψ contains the values of all polynomial basis functions assessed in the experimental design samples, with the components $\Psi_{ij} = \Psi_{\alpha_j}(x^{(i)})$.

To ensure the accuracy of the regression in determining the hyperparameters of Equation (4), the number of samples N is usually chosen to be greater than the cardinality P , often around 2 to 3 times larger [14,130]. However, a prominent obstacle arises as the number of terms in the PCE representation of Equation (3) grows exponentially with the increase in input dimensions. This growth makes the required sample evaluations for constructing an effective surrogate model computationally impractical, resulting in the curse of dimensionality [14].

Moreover, there is difficulty in estimating accurate out-of-sample predictions [5] since the model's ability to generalize beyond the training data may be restricted by the high-dimensional input space. A significant lack of information makes it more challenging to quantify uncertainties from the available data. It is also crucial to take into account the joint probability density function (PDF) f_X or fully infer it from the data, as this can impact the convergence of the PCE model. Further, it is imperative to precisely generalize the PCE model to encompass the underlying relationships over the entire dataset. Performing inference on marginal distributions, which can be computationally intensive for high-dimensional spaces, or ensuring the PCE model's robustness against potential noise in the inputs, is necessary [23]. These challenges are associated with effectively applying PCE models, especially when dealing with a large number of input variables that are common in data-driven processes.

3.2. Low-rank tensor approximations (LRA)

In low-rank tensor approximations (LRA), a polynomial basis tensor product is used to represent the response as a sum of many rank-1 functions. The reduction in the number of unknown coefficients resulting from this representation has made LRA an effective surrogate model for uncertainty propagation, especially in problems with numerous random inputs [33-35,79,131-134]. To the authors' knowledge, despite the mathematical development of these surrogates for dynamical systems with structured inputs [135], the literature reports on their application to real-world engineering problems with unstructured inputs are limited [29,33-35,79,131-134,136].

The output Y of the low-rank tensor approximation $M(X)$ can be canonically decomposed as a sum of rank-1 functions [33-35,79,131-134] in equation (5).

$$Y = M(X) = \sum_{l=1}^R b_l \left(\prod_{i=1}^d v_l^{(i)}(X_i) \right) \quad (5)$$

Here, $v_l^{(i)}(X_i)$ is a univariate function of the i -th input variable in the l -th rank-1 component, and b_l are normalizing constants. R denotes the decomposition order, and d represents the dimension of the input variable X .

Equation (6) is obtained by decomposing each univariate function $v_l^{(i)}(X_i)$ into orthogonal basis polynomials with respect to the probability density function (PDF) f_X [33-35,79,131-134].

$$Y = M(X) = \sum_{l=1}^R b_l \left(\prod_{i=1}^d \left(\sum_{k=0}^{p_i} z_{k,i}^{(i)} P_k^{(i)}(X_i) \right) \right) \quad (6)$$

where $P_k^{(i)}$ is the k -th degree of univariate polynomial in the i -th input variable, p_i is the maximum polynomial order of $P_k^{(i)}$, and $z_{k,i}^{(i)}$ is the coefficient of $P_k^{(i)}$ in the l -th rank-1 term.

LRA can be constructed using greedy approaches. These approaches gradually increase the approximation rank while sequentially updating the polynomial coefficients along each dimension. The researches present several algorithms for the non-intrusive computation of the LRA coefficients [35].

This context introduces a conventional algorithm that includes a series of correction steps and an update step for assigning coefficients [137]. The algorithm calculates Y_r , which is the r -th order approximation of the model $Y = M(X)$, using Equations (7) and (8) [33-35,79,131-134].

$$Y_r = \sum_{l=1}^r b_l w_l \quad (7)$$

$$w_l = \prod_{i=1}^d \left(\sum_{k=0}^{p_i} z_{k,l}^{(i)} P_k^{(i)}(X_i) \right) \quad (8)$$

The rank-1 tensor w_r in the r -th correction step is created according to Equation (9) [33-35,79,131-134].

$$w_r = \underset{w \in \mathcal{W}}{\operatorname{argmin}} \| R_{r-1} - w \|^2_\varepsilon \quad (9)$$

where $\mathcal{W}$ is the space of rank-1 tensors and R_{r-1} represents the residual after the $(r-1)$ -th step. The first adjustment step takes $R_0 = Y$ into account.

Using a sequential minimization algorithm, the optimization problem in Equation (9) is replaced by a series of smaller ones, each containing coefficients in only a single dimension, as described by Equation (10) [33-35,79,131-134].

$$z_r^{(i)} = \underset{\zeta \in \mathbb{R}^{p_i}}{\operatorname{argmin}} \left\| R_{r-1} - \left(\prod_{i \neq j} \sum_{k=0}^{p_i} z_{k,r}^{(i)} P_k^{(i)} \right) \left(\sum_{k=0}^{p_j} \zeta_k P_k^{(j)} \right) \right\|_\varepsilon^2 \quad (10)$$

Notably, Several iterations over dimensions $\{1, \dots, d\}$ may be required for a correction step, and a termination criterion has been proposed for this purpose [137,138]. This criterion choice can affect the model's accuracy and warrant further investigation.

After a correction step has run out, the coefficients $b = \{b_1, \dots, b_r\}$ is determined and updated sequentially according to Equation (11) [33-35,79,131-134].

$$b = \underset{\beta \in \mathbb{R}^r}{\operatorname{argmin}} \left\| M - \sum_{l=1}^r \beta_l w_l \right\|_\varepsilon^2 \quad (11)$$

Since the system output predictions in LRA are performed using a smaller number of rank-1 tensors, the number of coefficients required to construct these models in Equation (11) rises linearly with the number of input variables [29,33]. As a result, LRA has been proposed as a more suitable surrogate model compared to PCE for use in high-dimensional systems [79]. However, this model requires further development when combined with other methods to optimize its usage in uncertainty quantification of various engineering problems, particularly those with structured inputs.

3.3. Kriging

Kriging models, also called Gaussian process (GP) regression, are one of the most widely used surrogate modeling methods in engineering problems, particularly in simulation-based design, spatial data analysis, and uncertainty quantification [139]. Kriging is a stochastic interpolation technique that considers the model response to be a manifestation of a GP [140]. This approach, particularly in addressing problems involving unknown input-output functions, has rendered it a flexible instrument for a wide range of applications. A review study states that Kriging has been employed in advanced formulations to propagate uncertainty in structured inputs [6].

Equation (12) specifies the Kriging model $M(X)$, which has a d -dimensional input vector X and a model output Y [141].

$$Y = M(X) \approx \beta^T f(X) + \sigma^2 Z(X, \omega) \quad (12)$$

$\beta^T f(X)$ represents the mean value of the GP, commonly referred to as the trend. σ^2 denotes the variance of the GP. $Z(X, \omega)$ is a GP with a zero mean and unit variance, which is characterized by a correlation function $R(|X^{(i)} - X^{(j)}|; \theta)$ and hyperparameters θ .

In application practices, the Kriging model hyperparameters θ are usually estimated on a set of N samples consisting of inputs $\mathcal{X} = \{x^{(1)}, \dots, x^{(N)}\}$ and corresponding model outputs $\mathcal{Y} = \{y^{(1)}, \dots, y^{(N)}\}$, which is called the experimental design space [6,142]. Maximum likelihood estimation, cross-validation, and Bayesian inference are methods used in the literature to determine the Kriging parameters [6,143,144].

For instance, in the maximum likelihood approach, the parameters of the Kriging model are chosen to maximize the probability of observing the output data. Because the model output is a Gaussian vector, by maximizing the likelihood of the multivariate normal distribution, the parameters β and σ^2 are obtained from Equations (13) and (14), which are known as the generalized least-squares estimates [143].

$$\beta = \beta(\theta) = (F^T R^{-1} F)^{-1} F^T R^{-1} \mathcal{Y} \quad (13)$$

$$\sigma^2 = \sigma^2(\theta) = \frac{1}{N} (\mathcal{Y} - F\beta)^T R^{-1} (\mathcal{Y} - F\beta) \quad (14)$$

The correlation matrix of the experimental design samples is denoted as $R_{ij} = R(|x^{(i)} - x^{(j)}|; \theta)$, while the information matrix containing the values of all regression functions evaluated on the input set X is represented as $F_{ij} = f_j(x^{(i)})$ [145].

The values of the hyperparameters θ are obtained by solving the optimization problem described in Equation (15) [6].

$$\theta = \underset{\theta \in \mathcal{D}_\theta}{\operatorname{argmin}} (-\log \mathcal{L}(\mathcal{Y}|\beta, \sigma^2, \theta)) \quad (15)$$

The precision of regression analysis generally relies on a large sample size. As previously mentioned, in high-dimensional spaces, acquiring a large number of samples might pose considerable difficulty due to computational resource and time restrictions [16]. Additionally, if the estimation of the inherent model hyperparameters requires a large number of samples, the size of the Kriging covariance matrix in Equations (13) and (14) increases noticeably. In such cases, the inversion of the covariance matrix into high dimensions for optimization and hyperparameter determination in Equation (15) becomes computationally expensive [64,66]. The number of tuning parameters in the correlation function also grows with the input dimensions [7,20]. Consequently, the Kriging model may lose its efficiency in uncertainty quantification approaches and become susceptible to the curse of dimensionality.

3.4. Artificial neural network (ANN)

Surrogate models based on artificial neural networks (ANNs) have received the attention of researchers as powerful tools for handling high-dimensional systems, particularly in the field of uncertainty quantification [11,26,53,72,83,146]. These models exhibit high interpretability and, by leveraging the capabilities of deep neural networks (DNNs), offer efficient substitutions for traditional surrogates, which are often challenged by the curse of dimensionality, especially in structured inputs [53]. However, unlike common approaches, ANN-based surrogate models do not inherently provide a probabilistic characterization of predictions, which is essential for models applied to calibration and statistical model updating [52].

The standard architecture of ANNs is characterized by a structure of interconnected, dense layers of neurons. These layers establish nonlinear relationships between the input and output by combining the weighted sums of the inputs and applying specialized functions known as activation functions [147,148].

This work describes a surrogate model leveraging an ANN architecture. The ANN comprises an input layer, followed by K hidden layers $L^{(k)} : \mathbb{R}^{q_{k-1}} \rightarrow \mathbb{R}^{q_k}, k = 1, \dots, K$ and an output layer $O : \mathbb{R}^{q_K} \rightarrow \mathbb{R}^m$, which is employed to approximate the mapping between the input $X \in \mathbb{R}^d$ and output $Y \in \mathbb{R}^m$ of the problem [147]. It is worth noting that by adjusting the number of hidden layers and the number of neurons in each layer, ANN models with the desired complexity can be obtained. The number of neurons in the input and output layers corresponds to the dimensions of the input and output variables, respectively [53]. Fig. 3 illustrates a standard ANN with three hidden layers.

In Equation (16), each hidden layer k is composed of a weight matrix $W^{(k)} \in \mathbb{R}^{q_{k-1} \times q_k}$ and a bias vector $b^{(k)} \in \mathbb{R}^{q_k}$.

$$W^{(k)} = \begin{bmatrix} w_0^{(k)} \\ w_1^{(k)} \\ \vdots \\ w_{q_k}^{(k)} \end{bmatrix} = \begin{bmatrix} w_{0,0}^{(k)} & w_{0,1}^{(k)} & \dots & w_{0,q_k-1}^{(k)} \\ w_1^{(k)} & \ddots & & w_{1,q_k-1}^{(k)} \\ \vdots & & \ddots & \vdots \\ w_{q_k,0}^{(k)} & w_{q_k,1}^{(k)} & \dots & w_{q_k,q_k-1}^{(k)} \end{bmatrix}, \quad b^{(k)} = \begin{bmatrix} b_0^{(k)} \\ b_1^{(k)} \\ \vdots \\ b_{q_k}^{(k)} \end{bmatrix} \quad (16)$$

The nonlinear relationship in ANN-based models is only expressed through a nonlinear activation function $\sigma(\cdot)$, according to Equation (17).

$$z^{(k-1)} \in \mathbb{R}^{q_{k-1}} \mapsto L^{(k)}(z^{(k-1)}) = \sigma(W^{(k)}z^{(k-1)} + b^{(k)}) = \begin{bmatrix} \sigma(w_0^{(k)}, z^{(k-1)} + b_0^{(k)}) \\ \sigma(w_1^{(k)}, z^{(k-1)} + b_1^{(k)}) \\ \vdots \\ \sigma(w_{q_k}^{(k)}, z^{(k-1)} + b_{q_k}^{(k)}) \end{bmatrix} \in \mathbb{R}^{q_k} \quad (17)$$

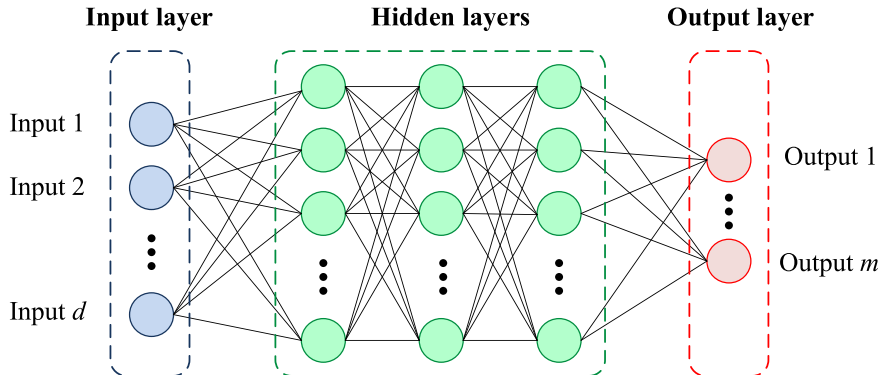


Fig. 3. - Standard artificial neural network

where $w_i^{(k)}$ and $b_i^{(k)}$ for $i = 0, \dots, q_k$ are the i -th row of the weight matrix $W^{(k)}$ and the i -th element of the bias vector $b^{(k)}$, respectively. The weights and biases $\theta = \{W^{(k)}, b^{(k)}\}$ are known as the network parameters, which fully characterize the ANN architecture. $z^{(k-1)}$ is an intermediate input variable, whose role is shown in Equation (18).

$$z^{(0)} = X; z^{(k)} = L^{(k)}(z^{(k-1)}), k = 1, \dots, K \quad (18)$$

As shown in Equation (19), the output layer neurons of ANN are determined by a weight matrix $W^{(O)} \in \mathbb{R}^{q_k \times m}$, without any bias terms or activation functions.

$$z^{(k)} \in \mathbb{R}^{q_k} \mapsto y = O(z^{(k)}) = W^{(O)} z^{(k)} \in \mathbb{R}^m \quad (19)$$

Equation (20) combines the layers to establish the input-output relationship. This structure is powerful for a diverse array of engineering problems due to its sufficient complexity [147].

$$L^{(K)} \circ \dots \circ L^{(1)}(X) = W^{(O)} \left[\sigma \left(W^{(k)} \left[\sigma \left(W^{(k-1)} [\dots] + b^{(k-1)} \right) \right] + b^{(k)} \right) \right] Y = O \quad (20)$$

The network parameters θ are computed by solving the optimization problem in Equation (21) using a set of N input-output samples [53]. It is worth underscoring that the solution to this optimization problem is obtained through iterative algorithms.

$$\theta = \underset{\theta}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(\theta; y_i) \quad (21)$$

The cost (loss) function is denoted as $\mathcal{L}(\theta; y_i)$.

The accuracy of ANN-based surrogate models is highly dependent on the quantity and quality of the available data [16-20]. In complex high-dimensional systems, access to comprehensive input-output datasets is often limited due to the computational demand of data. Additionally, the sparsity and presence of noise in the datasets present obstacles to the accuracy and robustness of these models [7]. Recently, research efforts have been geared towards overcoming the challenges associated with ANN-based surrogates in real-world engineering problems. This includes the development of fast optimization techniques for parameter estimation, the construction of interpretable surrogate models, and, specifically, the incorporation of physical constraints into physics-informed neural networks (PINNs) to enable gray-box rather than black-box modeling approaches [90,97,103,146,149,150].

4. Developed surrogate modeling techniques for high-dimensional spaces

The high computational cost associated with model evaluations, coupled with other challenges in uncertainty quantification, underscores the necessity of developing surrogate models for applications with high complexity and dimensions. This section reviews the main approaches from the past years, including dimensionality reduction, multi-fidelity surrogate models, and advanced sampling-based surrogate modeling, aimed at overcoming the curse of dimensionality. Alongside outlining these solutions, this section also discusses proposed techniques for enhancing model performance in addressing other challenges specific to high-dimensional spaces.

Fig. 4 illustrates the contributions of various techniques to develop surrogate models for quantifying uncertainties in high-dimensional spaces, based on the 90 studies reviewed in this paper over the past years, as elaborated in the subsequent subsections. It highlights that research has predominantly focused on three main solutions compared to other approaches [29,33-35,48,68].

Figs. 5 (a) and 5 (b) display the bar and cumulative charts for the 90 studies reviewed in this paper, covering the period from 2013 to 2024, as detailed in the following subsections. These studies are categorized by the different techniques employed in constructing surrogate models to manage a large number of variables. Fig. 5 (a) clearly shows an annual increase in the use of dimensionality reduction, multi-fidelity surrogate models, and advanced sampling-based surrogate modeling. This trend is not unexpected, as the

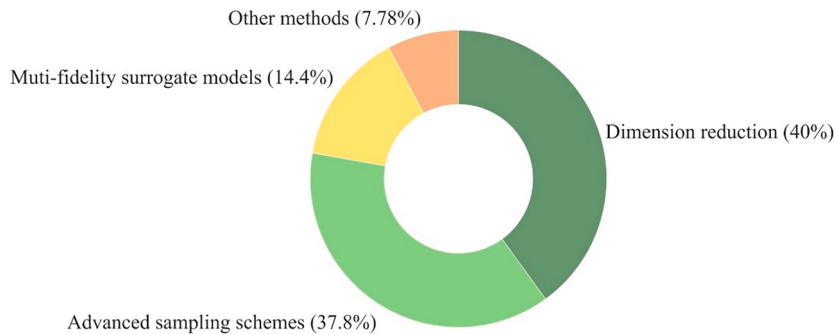


Fig. 4. - Distribution of the reviewed papers from 2013 to 2024 based on techniques used to construct surrogate models for uncertainty quantification in high-dimensional spaces

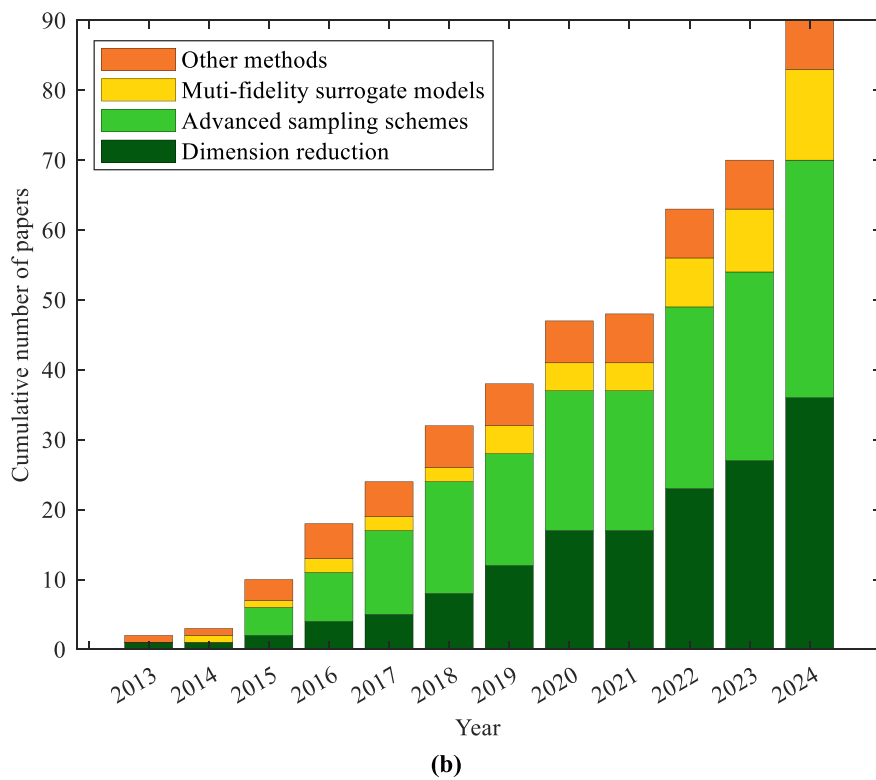
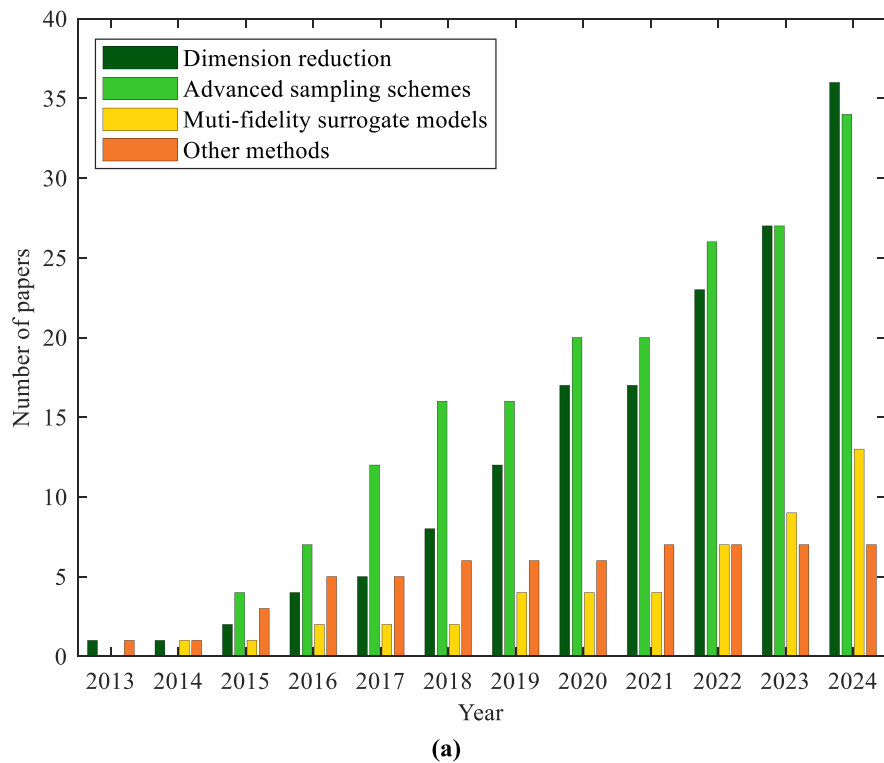


Fig. 5. - Cumulative number of the reviewed papers per year from 2013 to 2024, categorized by techniques used to construct surrogate models for uncertainty quantification in high-dimensional spaces; (a) bar, (b) stacked charts

advancement of machine learning algorithms in surrogate modeling for high-dimensional systems is a crucial factor in these approaches. In contrast, other methods have exhibited virtually no growth in recent years. A review of these methods indicates that researchers are actively seeking to innovate frameworks and develop more generalizable approaches to tackle various engineering problems. Furthermore, the upward trend depicted in Fig. 5 (b) suggests that this field of uncertainty quantification has garnered increased attention in recent years. Note that this review was written in mid-2024 and does not include articles published in the latter half of the year. The growing trend of research in this area is expected to continue throughout 2024.

4.1. Dimension reduction

Dimension reduction employs a subset of input variables to map the uncertainty of inputs to outputs, often at the expense of some accuracy and information. This approach simplifies the process of extracting intrinsic patterns and uncovering the underlying data structure, thereby mimicking full-scale simulations at a reduced computational cost [109]. Various dimension reduction techniques have garnered attention due to their remarkable abilities to improve model efficiency, enhance data visualization, and enable better model generalization with fewer computational resources. The literature often utilizes dimension reduction independently for unstructured inputs, while it is commonly combined with other solutions for structured inputs to address the curse of dimensionality. As illustrated in Fig. 4, a significant portion of research has adopted this practical strategy.

Among dimension reduction techniques, Principal Component Analysis (PCA) creates new independent features by combining properties of the original dataset. PCA has emerged as a prominent tool for feature extraction in various engineering applications, leading to reduced problem complexities, simplified calculations, and fundamental information extraction from data [109]. As a part of its application, Hombal and Mahadevan [10] addressed the challenge of the curse of dimensionality and the difficulty of identifying space-filling designs for sampling due to the exponential increase in data space volume. They proposed a method for developing a non-parametric representation of nonlinear crack growth under multi-axial loading with variable amplitudes. This method enables robust surrogate modeling for non-planar crack growth with complex shapes using a relatively small number of samples. Their approach employs three-dimensional spline curves with a fixed number of nodes to represent the crack fronts. After feature extraction in a lower-dimensional space using PCA, they constructed a GP model to describe the crack growth process. The literature demonstrates that the PCA technique has been employed for both input and output variables. For instance, to tackle the computational challenges of high-dimensional outputs, Li et al. [57,58] combined surrogate modeling with a dimension reduction technique known as Dimension Reduction and Surrogate-based Sensitivity Analysis (DRE-SSA). This method uses PCA to reduce the dimensionality of high-dimensional outputs and Kriging modeling to compute covariance matrices in the reduced output space. Sensitivity indices for the high-dimensional outputs are then calculated based on the covariance matrices. The validity of this approach has been demonstrated using a complex, two-dimensional analytical function. The efficiency and accuracy of DRE-SSA were further assessed through a sensitivity analysis of the peak water level in the coastal areas of San Francisco Bay.

The method proposed by Liang and Mahadevan [37] intends to reduce the computational burden of high-dimensional problems by using PCA on Bayesian Networks (BNs). BNs are elegant tools for extracting information from complex systems with a large number of variables. By determining the probability distribution of variables in BNs, this method provides a more effective generalization. It is particularly suitable for probabilistic applications that involve tracking random and knowledge-based uncertainties [146]. This approach, designed for uncertainty propagation in stochastic multidisciplinary analysis involving numerous coupling variables, comprises three components: BNs, copula-based sampling, and PCA. It constructs the BN by inferring the joint distribution of the coupling variables through interdisciplinary compatibility conditions, requiring only a few iterations. The copula-based sampling utilizes joint and conditional distributions. Given the dependence of this approach on first-order approximation, its accuracy is significantly affected by the nonlinearity of the problem. The efficiency and accuracy of the proposed method have been evaluated on a mathematical problem and the aeroelastic analysis of an aircraft wing model. Furthermore, Bassamzadeh and Ghanem [54] leveraged the BN tool to estimate the probability of oil production rates in locations with limited data from oil fields. They employed the BN to select the most relevant variables for constructing a model that relates unknown variables within a high-dimensional complex system. The performance of this BN structure was compared to that of local linear embedding (LLE), a tool for reducing the dimension of input variables to discover nonlinear structures. The results of the BN model, based on a real dataset from the Gulf of Mexico, were also compared with the outputs of other models, including neural network-based and co-Kriging models.

The use of nonlinear dimension reduction techniques has continued in Lataniotis's study [6] to quantify uncertainty in high-dimensional data-driven systems. This study replaces the sequential approach with a supervised dimension reduction and surrogate modeling (DRSM) method. The DRSM method optimizes the dimension reduction parameters to generate reduced variables that are suitable for a more accurate surrogate model. This is accomplished by updating the compression parameters in the outer loop and the surrogate model in the inner loop of a nested optimization problem. Several nonlinear dimension reduction techniques, such as kernel PCA applied to Kriging and PCE models, were applied within the proposed DRSM framework. The effectiveness of the DRSM method is investigated by determining the probability density function (PDF) of the response for the Sobol function, an electrical resistance network, a two-dimensional heat diffusion problem, and structural health monitoring data of a wind turbine.

It is well known that PCA focuses solely on the input space to identify patterns, handling maximum variance in this space as the basis for determining the main features. The partial least squares (PLS) method, on the other hand, maps both the input and output variables to a new feature space based on maximum covariance, decreasing the number of dimensions to improve the performance of the surrogate model [109]. The literature has explored PLS in combination with various surrogate models. For example, Ehre et al. [63] utilized latent-variable-based PCE and PLS to calculate variance-based sensitivity indices for high-dimensional problems. By re-transforming the surrogate from the latent variable space back to the original input variable space, they derived analytical

expressions dependent on the model coefficients for the desired sensitivity. Therefore, the variance-based sensitivity indices are computed by constructing the surrogate model without requiring further sampling. The accuracy of the proposed method has been assessed on a 23-rod truss and a two-dimensional steel plate. Meanwhile, Zhou et al. [64] introduced a method that embeds dimension reduction within the data-driven PCE approach to mitigate the curse of dimensionality in high-dimensional problems. The proposed surrogate model uses a completely non-intrusive technique called sparse PLS to identify the most predictive projection directions. In constructing the polynomial basis of the suggested model, the components are not assumed to be independent in the reduced space. As a result, this method can be applied to complex systems with implicit information about the underlying distribution. The efficacy of this approach has been validated on problems with different complexities, including a cantilever tube structure, a heat diffusion problem, and an isotropic plate. In another study, Liu et al. [74] combined the Kriging model and PLS for surrogate modeling. This approach identifies the principal components of the input and output variables using PLS and constructs the relationship between each pair of these components using the Kriging model. Additionally, the error-based weights updating method has been applied to improve the integration of the Kriging model and PLS. This method has been demonstrated to be accurate and effective for a time-variant system, cantilever tube, heat conduction, and oscillator models.

Consistent with these investigations, researchers have also explored manifold-based techniques for handling data spaces with complex structures and nonlinear relationships. These methods represent high-dimensional space on a manifold with two or more dimensions [109]. These approaches offer several key advantages, including effective dimensionality reduction without significant information loss, the ability to visualize data by revealing relationships between variables in the reduced space, and compatibility with various data configurations, such as noisy data. One application of these techniques is the method proposed by Thimmisetty et al. [52], which aims to reduce uncertainty in high-dimensional predictions. This method applies GP regression to the data using the intrinsic manifold metric, and then reinterprets the scattered data around the nonlinear manifold. A set of standard well log data has confirmed the method's superiority in describing geological properties. Further application of the manifold concept is found in Soize and Ghanem's study [46], which introduced a robust and efficient method for high-dimensional problems and large datasets. To generate samples for PCE modeling, their method offers an analytical characterization of the dataset on the manifold. This method creates a PCE from a random vector along with an associated generator, sampling the vector from the target probability distribution of the data concentrated near the sample manifold. Two datasets and a petrophysical database based on experimental measurements have shown the effectiveness of this approach. In the meantime, classical dimension reduction techniques have not escaped the attention of researchers. For example, Chen and Quarteroni [15] adopted a weighted reduced basis method to address the curse of dimensionality and reduce the computational burden. To quantify uncertainties in high-dimensional spaces characterized by sparsity and reducibility, their method develops an adaptive greedy algorithm coupled with a generalized sparse grid. This approach leverages hierarchical surpluses to automatically identify the importance and interactions of different dimensions. It also allows the method to exit the stagnation region before achieving the desired accuracy by examining specific stopping criteria. Initially, the proposed algorithm's performance was verified in two dimensions. Subsequently, two numerical experiments were conducted to assess its efficiency and accuracy.

In other studies, the active subspace (AS) method has been applied to reduce problem complexity by identifying the most influential subspace of lower dimensions that encapsulates the essential characteristics of the main space. The AS method, with fewer search iterations, facilitates more efficient exploration in high-dimensional spaces, significantly reducing computational costs. Identifying the AS can reveal potential dependencies and symmetries among the variables, providing insights into the problem's underlying structure. Generally, the AS method requires calculating the gradient of the target function at different points. Tripathy et al. [40] succeeded in eliminating the gradient requirement. To overcome the curse of dimensionality, these researchers developed a probabilistic version of the AS that is both gradient-free and robust to observational noise. This method involves developing a GP model with dimension reduction, where the orthogonal projection matrix of the AS serves as one of the hyperparameters of the GP covariance function. This is achieved through a two-stage maximum likelihood optimization. The automatic selection of AS dimensions is guided by the Bayesian information criterion. This approach has been employed for various problems, including artificial samples, an elliptic stochastic partial differential equation with random conductivity, and solitary wave propagation in one-dimensional granular crystals. Subsequently, Vohra et al. [7] established a combined method of PCA and AS, termed PCAS, for reliability analyses and global sensitivity analyses in problems with high-dimensional inputs and outputs. This approach reduces the dimensions of both input and output spaces. Initially, the features of the output field are extracted through PCA. Then, given a map from the inputs to each of these features, the relationship between them is captured in a low-dimensional input subspace using the AS method. The accuracy of the PCAS method has been demonstrated on a finite element model for simulating the residual stress distribution in a manufactured component. Other researchers have utilized the concept of combining active subspaces with different methods. For instance, Wang et al. [98] created a novel framework called AS-NIPC, which combines the active subspace (AS) approach with the non-intrusive polynomial chaos (NIPC) method. The goal was to tackle the curse of dimensionality in quantifying the uncertainties of high-dimensional problems. The AS-NIPC method involves creating orthogonal PCE basis functions using the active subspace variables and developing an effective quadrature rule in the original input space to estimate the PCE coefficients. The proposed framework is further enhanced by accelerating model evaluations on tensor grids using computational graph transformations to construct the quadrature rule with a desirable tensor structure. This approach benefits from the sparsity of the computational graph, leading to improved model efficiency. The effectiveness of this methodology compared to other methods has been shown by computing the relative errors of output prediction for a nonlinear piston simulation model and an air-taxi trajectory optimization scenario. The development of the Support Vector Machine (SVM) method into a regression version known as Support Vector Regression (SVR) has prepared this model to apply in high-dimensional problems. Although these models are sensitive to missing data and require tuning kernel functions and hyperparameters, predictions generated by these surrogates using a small number of training data can lead to satisfactory results [3]. To

illustrate, Zhou et al. [93] aimed to quantify uncertainties by developing a dimension-reduced SVR modeling approach. This method achieves improved dimension reduction performance by utilizing a sparse active subspace. The sparse active subspace identifies a set of sparse projection vectors, each representing a group of input variables that affect the target quantities. Integrating the dimension reduction step into the kernel structure of the SVR model within a unified framework leads to improved results in high-dimensional spaces. The kernel trick is used to identify nonlinear manifolds, thereby overcoming the limitation of linearity. The suggested method has been employed to predict the probability density function (PDF) of the response on the Sobel function, a steady-state diffusion problem, and a lid-driven cavity problem, demonstrating its effectiveness. Simultaneously, Mufti et al. [88] designed a reduced-order modeling (ROM) framework that is multi-fidelity, parametric, and non-intrusive to handle multiple evaluations in high-dimensional design spaces. This method uses machine learning techniques to align manifolds and perform proper orthogonal decomposition (POD) to reduce the output dimension. Additionally, it employs active subspaces based on a multi-fidelity model to decrease the input dimension. These two techniques are then combined to construct a multi-fidelity reduced-order model in the input space with a large number of dimensions. The suggested method has been validated for two aerodynamic analysis problems. Compared to the single-fidelity PCAS method, evaluating different fidelity levels, sample sizes, and training data ratios indicates that this method enhances prediction accuracy using fewer computational resources. Furthermore, the findings imply that this approach has better performance in managing the high-dimensional input space compared to the reduced manifold-aligned ROM. In a different study, Bharti and Ghosh [89] investigated the capabilities of the POD technique in reducing input dimensions in time-dependent data sets. They suggested a non-intrusive ROM for quantifying uncertainty in dynamical systems subjected to random excitation. Instead of discretizing the large random input, which leads to high-dimensional interpolation or regression in the ROM, they use a neural network-based surrogate model to compute the regression on the random excitation. The reduction of the random excitation by POD is carried out sequentially, where the POD captures the most important modes of variation to reduce the system complexity and enable more efficient modeling. The accuracy and efficiency of this model have been assessed for estimating the probability density function (PDF), cumulative distribution function (CDF), and failure probability in a mistuned-bladed disk problem, as well as analyzing the soil-structure interaction in a beam on the Winkler foundation.

Combining bootstrapping with dimension reduction techniques leverages the strengths of both methods to enhance the performance of clustering algorithms, especially when dealing with complex datasets. Bootstrap-based clustering can effectively mitigate the curse of dimensionality and improve model prediction by concentrating on the most informative features. From this perspective, Sheikholeslami et al. [59] proposed a method for grouping input factors based on bootstrap clustering. This method groups inputs with similar sensitivities, which can effectively reduce the input space and facilitate global sensitivity analysis in high-dimensional models. It employs either the elbow method or the minimum robustness method to determine the optimal number of groups. Furthermore, a criterion is provided to assess the robustness, stability, and convergence of the analysis. The results of a global sensitivity analysis with active grouping have been demonstrated on the Sobol function and a hydrological model. In another study, Catalanotti [102] utilized a combination of bootstrap and Bayesian analysis to deal with the computational challenges of high-dimensional simulations of composite material damage mechanisms, which exhibited a notable level of complexity. The bootstrap process generates the input vector by sampling from the pool of experimental data, feeds it into the computational model, and records the predictions in each iteration. Bayesian analysis is then applied to aggregate the predictions. This researcher estimates the response distribution hyperparameters and stores the values from each iteration for subsequent analysis. The analysis is terminated by monitoring the convergence of the hyperparameters in successive iterations. Additional simulations are deemed unnecessary due to negligible improvement or no change in the hyperparameter values or prediction accuracy. This approach to estimating the hyperparameters of the response distribution in a 40-dimensional problem, including open-hole specimens under both tension and compression, has led to a significant reduction in the number of required simulations while maintaining accuracy.

Further research has evaluated the performance of various linear and nonlinear dimension reduction techniques. Guo et al. [16,20] investigated dimension reduction for both input and output variables. They proposed a combined method that applies techniques such as singular value decomposition (SVD), random projection, randomized singular value decomposition (rSVD), and diffusion maps to the output variables. This method utilizes rSVD to combine the accuracy of classical SVD with the speed of random projection. Diffusion maps provide a low-dimensional nonlinear embedding by rearranging data in a high-dimensional space based on the underlying geometry of the parameters [109]. To reduce the input dimension, the researchers employed variance-based global sensitivity analysis, leveraging its capabilities to identify the most important variables in unstructured datasets [109,151]. Subsequently, they applied GP modeling and tree-based regression techniques in the reduced-dimensional space. The effectiveness of the proposed method was evaluated through the analysis of an aircraft fuselage panel and a gas turbine blade.

To highlight the importance of preserving the invariance properties of surrogate models, Wycoff et al. [70] developed a framework incorporating a preprocessing stage for rotating and rescaling the inputs for high-dimensional surrogate modeling. This approach achieves GP modeling by executing fewer iterations of sparse local models with short-range correlations. The input space rotation, using AS-based algorithms and automatic relevance determination, is conducted so that sensitivity analysis demonstrates the importance of all directions. By learning the globally significant directions, each local model can focus on its prediction region without knowing the overall trend. The input rotation also enables enhanced neighborhood selection and improves the value of distance in prediction. The suggested method has been validated using two sets of observational datasets, benchmark test functions, and an automotive industry simulation problem.

With the growing popularity of DNNs in recent years, the use of adaptable autoencoder tools as unsupervised nonlinear dimensionality reduction methods has become widespread [109]. Autoencoders consist of two main components: an encoder and a decoder. The encoder compresses the input variables into a lower-dimensional representation, often referred to as the latent or encoded space. The decoder then reconstructs the input from this compressed representation. This process leads to the discovery of the underlying data

structures. The versatility and adaptability of autoencoder tools have made them a valuable asset for deep learning (DL) in addressing high-dimensional problems. As an illustration, Zhu et al. [26] designed convolutional encoder-decoder neural network (NN) physics-constrained models for quantifying uncertainties in high-dimensional systems. In this approach, the partial differential equation (PDE) model incorporates the governing equations of the physical model into the loss or likelihood functions. The physics-constrained DL model is trained without the need for labeled data. It predicts the response based on the convolutional encoder-decoder NN model and a conditional generative model. This method has been used to perform uncertainty quantification for the steady-state flow in random heterogeneous media. In another study, Tripathy and Bilonis [53] tackled high-dimensional problems and introduced a method based on parameterizing the structure of DNNs to train surrogate models. The DNN in this methodology consists of an encoder and a single-layer multilayer perceptron network. Parameterizing the DNN enables the interpretation of recovering a nonlinear, low-dimensional AS. This method has been implemented to propagate uncertainty in a stochastic elliptic partial differential equation.

The use of a low-dimensional latent space to map the distribution of input data to a lower-dimensional space and then map between the latent space distribution and the output to construct NN-based surrogate models has also been investigated in other studies. In this regard, Shustin et al. [72] employed a two-stage method to train surrogate models in high-input systems under uncertainty. This method first applies a NN-based Bayesian unsupervised learning approach to map the distribution of input data to a lower-dimensional latent normal distribution. Then, a PCE model is trained to map between the latent and output distributions. The performance of the proposed method are presented on three machine-learning datasets and two PDEs. Low-dimensional latent space has also been a focus of the latest advancements in the field. For example, Shi et al. [96] utilized a deep convolutional dimension reduction network (ConvDR) to construct surrogate models and quantify structural uncertainties in high-dimensional problems. This approach addresses the inherent high dimensionality by transforming the spatial data into a low-dimensional latent space using a neural network. The knowledge-based uncertainties in the latent space are formulated using a loss regularization. Evolutionary algorithms are used to train this network, and a linear regression model is employed as a surrogate for response prediction. The proposed model has been validated and assessed by determining the root mean square error (RMSE). Additionally, the accuracy of the approach for quantifying uncertainty has been examined by comparing probability density function (PDF) curves and statistical moments with other methods. The results of ConvDR have been presented to address high-dimensional challenges on the Messerschmitt-Bölkow-Blohm beam and a system of compliant force inverters. Furthermore, Song et al. [95] introduced the improved sufficient dimension reduction-based Kriging modeling (KISDR) method to deal with the curse of dimensionality in high-dimensional uncertainty propagation problems. This approach begins with an improved enough dimension reduction technique to estimate the projection matrix. This estimation relies on the concept of martingale difference divergence, which allows for the representation of high-dimensional inputs in a low-dimensional latent space while preserving sufficient information for response prediction. The Ladle estimator then combines the eigenvalues and eigenvectors of this matrix to enhance the accuracy in determining the dimensions of the latent space. The new Kriging correlation function is created by integrating the reduced-dimensional information and the correlation structure, leading to a significant decrease in the number of model hyperparameters. An innovative local optimization strategy is used to tune the hyperparameters of the new Kriging model. The performance of the KISDR method in calculating the probability density function (PDF) of the response has been investigated on the Dixon-Price, Griewank, Ackley, Rosenbrock, and Morris functions, as well as on a cantilever beam, a truss bridge, and a hybrid joint structure. Continuing the motivation towards dimensionality reduction methods, Kontolati et al. [71] applied various dimensionality reduction techniques to overcome the curse of dimensionality while constructing surrogate models for quantifying uncertainty in complex PDEs of physics-based systems. After applying thirteen linear and nonlinear dimensionality reduction techniques, PCEs were created on the compressed inputs. This approach, known as manifold PCE, explores the latent space manifold for each of the dimensionality reduction techniques. The strengths and limits of the proposed approach have been compared on black-box systems, including stochastic Poisson's equation, the stochastic steady-state heat equation, and the time-dependent stochastic Brusselator model. In the most recent research, Kim et al. [97] extracted a dimension-reduced stochastic surrogate model (DR-SM) to tackle the curse of dimensionality while quantifying uncertainties and handling computationally intensive physics-based models. This method first performs dimension reduction in the input and output space of the computational model, thereby eliminating the necessity for reconstruction mapping. The low-dimensional features are then linked to the output by establishing a conditional distribution using a heterogeneous GP. Subsequently, a stochastic surrogate model is constructed to predict the output. The key idea of the DR-SM approach is to create a transition kernel that alternates between the dimension reduction algorithm and a conditional distribution of the feature space. The proposed method has been implemented using three dimension reduction techniques: PCA, kernel PCA, and autoencoder. It estimates the probability density function (PDF) and complementary cumulative distribution function (CCDF) of the response of a uniaxial elastic-linear bar under tension load and a 5-story nonlinear hysteretic shear building under stochastic excitation.

To highlight the role of optimal mid-dimension selection as an idea in the dimension reduction procedure, Bigoni et al. [77] proposed a method for nonlinear dimensionality reduction in the adaptive polynomial function space to address the curse of dimensionality in small-data regimes (when only a few evaluations are available). The objective of this approach is to identify a nonlinear feature map through an intermediate space that has a significantly lower dimension compared to the output. This approach relies on model gradient evaluations and is carried out in a two-stage procedure. In the first step, the feature map is constructed from the input dimension to the mid-dimension, and in the second step, the profile function is built from the mid-dimension to the output dimension. Numerical examples such as isotropic functions, the Borehole function, the deep combination of functions, and a PDE-based model for determining the smallest resonance frequency in a two-dimensional bridge have been used to study the influence of the mid-dimension on the performance of the proposed algorithm.

A survey of the literature indicates that convolutional autoencoders have been successfully employed for dimensionality reduction

in image processing, as well as enhancing the accuracy of geological model predictions, particularly in cases where traditional methods may encounter difficulties. As Gavallas et al. [152] demonstrated in their study on the random fields of mechanical properties of composite materials, employing convolutional neural networks rather than the expensive and highly repetitive finite element homogenization procedure enables nearly instant predictions of random property fields derived from microstructure images. Specifically, Sun et al. [83] utilized a dense convolutional network based on a deep autoregressive neural network (ARnet) to develop a surrogate model for quantifying the uncertainty of parameters in land surface hydrological models. The implementation of an encoder-decoder architecture in the proposed deep convolutional network has proven highly effective in handling high-dimensional images and extracting spatial features. Furthermore, this model uses an autoregressive technique to manage the time-dependent input-output relationship. The soil moisture simulation over a river in China has illustrated significant enhancement in the performance and accuracy of spatial pattern prediction compared to the long-short-term memory (LSTM) model.

Despite advances in the development of DNN-based surrogate models, efforts to increase their accuracy using dimensionality reduction techniques in combination with other modeling methods have continued. For example, Liu and Jiang [85] designed the deep kernel polynomial chaos expansion (DKPCE) model to alleviate the curse of dimensionality in high-dimensional uncertainty propagation problems. The DKPCE model comprises multiple neural network layers, including the deep kernel and a PCE layer. By restricting the number of neurons in the layer connected to the PCE layer, a balance can be established between the precision and quality of the input dimension reduction to the PCE layer. The orthogonal polynomial bases of this model are computed during the forward propagation step, while their coefficients are determined during the backpropagation step. The effectiveness of the DKPCE model has been investigated in comparison to the performance of PCE methods coupled with PCA, DNN, autoencoder-based PCE, and Monte Carlo sampling on benchmark functions of dimensions 150, 200, and 300, as well as a plane heat conduction problem.

Sliced inverse regression is an effective dimensionality reduction technique that replaces high-dimensional covariate variables with a few linear combinations [153]. Due to its advantages, such as flexibility, efficiency, and versatility in various data environments, this method has been a potent tool for dimensionality reduction in constructing surrogate models. For example, Zhou and Lu [66] recommended the concurrent use of Kriging modeling and dimensionality reduction to simulate the behavior of systems with high-dimensional input spaces. This approach relies on sliced inverse regression to construct a novel projection vector, thereby diminishing the input dimensions. Then, a Kriging correlation function is created in the reduced subspace using the tensor product of multiple correlation functions for each direction. The proposed method has been validated on Rackwitz and Morris functions. Furthermore, the finite element simulation of an aero-engine compressor disc has confirmed the enhanced accuracy and robustness of the surrogate model. Considering the significant impact of the correlation of uncertain variables on uncertainty quantification, researchers have explored this issue in their studies [154]. In this context, Cheng and Zimmermann [84] devised the sliced gradient-enhanced Kriging (SGE-Kriging) method for global sensitivity analysis in high-dimensional problems. This method decreases the size of the correlation matrix and the number of hyperparameters in the model. Assuming the independence of distant samples, a simplified likelihood function is derived from Bayes' theorem. By dividing the training sample set into multiple batches, several small correlation matrices are employed to capture the correlation of the sample set. Finally, a surrogate model is developed by learning the relationship between the model hyperparameters and the derivative-based global sensitivity indices in a low-dimensional space. The accuracy and robustness of the model have been justified by several benchmark functions and a high-dimensional aerodynamic simulation problem, with significantly lower training costs compared to the gradient-enhanced Kriging model.

In recent studies, Schär et al. [86] introduced the manifold nonlinear autoregressive model with exogenous inputs (mNARX) for predicting the response of complex dynamical systems under time-varying excitations. The fundamental concept of this approach is to decompose the problem into a series of smaller and less complex subproblems compared to the original problem. This method sequentially constructs a chain of nonlinear autoregressive models with exogenous inputs (NARX). These models, which are trained on the exogenous inputs manifold, incorporate the predictions of the models in the initial chain together with their features. As the modeling chain is extended and the number of features increases, the capability to represent more complex behaviors is provided. The improved performance of the mNARX in accurately and efficiently approximating the response of a mass-spring system subjected to random excitation has been successfully established, compared to the autoregressive model. In addition, a case study has been conducted on an onshore wind turbine simulator to evaluate the feasibility of the method for high-dimensional engineering systems.

4.2. Multi-fidelity surrogate models

Surrogate models can be generated at different levels of quality to describe the system of interest in numerous engineering problems. These models differ in terms of evaluation cost and level of fidelity. Computationally expensive high-fidelity (HF) models can accurately capture system behavior, whereas low-fidelity (LF) models have lower evaluation costs and reduced prediction accuracy compared to HF models [155]. However, the optimal construction of surrogate models within a limited computational budget requires versatile tools capable of making a wide range of predictions. These tools can ultimately synthesize new boundaries based on the analysis results [42]. The multi-fidelity approach can be applied to construct surrogate models by combining the outputs of LF and HF models [155]. The literature review reveals that using LF models for surrogate modeling can make it faster, while HF models provide accurate evaluations and (or) convergence of model outputs [155]. Therefore, leveraging the strengths of multiple levels of fidelity models can be effective in enhancing the accuracy and efficiency of surrogate models for quantifying uncertainties in complex, high-dimensional systems. This approach can be especially effective when multiple data sources are available. Fig. 4 displays the participation of multi-fidelity models in the progress of surrogate modeling research, as observed in studies over the past years.

Among the multi-fidelity management techniques, information fusion, which merges data from several computational sources, has received significant attention from researchers. For instance, Perdikaris et al. [42] introduced a method that combines multi-fidelity

information and predictive inference of dynamic system response surfaces. This method is designed to handle high-dimensional input spaces in the presence of massive datasets. It scales to high dimensions via data-driven dimensionality reduction techniques based on hierarchical functional decompositions and encodes custom autocorrelation structures in GP priors using a graph theoretic approach. Machine learning algorithms along with stochastic autoregressive schemes then facilitate the fusion of multi-fidelity information. The combined approach has been demonstrated to be effective on the Borehole function, an elliptic problem subjected to random forcing, and the Sobol function. Prior research has shown that hierarchical surrogate models have been well-received. The fusion technique aims to establish a hierarchy of models where LF models are used as trends to construct higher fidelity surrogates [140]. Abdallah et al. [143] investigated the application of hierarchical Kriging as a multi-fidelity surrogate model. This approach combines the outputs from multiple simulators with different complexities to improve prediction and reduce model uncertainty compared to conventional Kriging. The selection of the optimal surrogate models in the approach is determined by evaluating all possible combinations of Kriging parameter candidates. The success of the proposed method has been indicated for predicting bending moments at the root of a large wind turbine, where model uncertainty and heterogeneous noisy outputs are present. With the approach of combining multi-fidelity data, Conti et al. [80] developed a balance between model accuracy and efficiency. It particularly works in time-parameter-dependent problems with limited HF data. Their approach embeds LSTM neural networks to construct surrogate multi-fidelity models, utilizing LF data for faster computations and limited HF data for improved approximation. The LSTM neural networks, as a technique of recurrent NNs, serve as the main building blocks in the construction of this surrogate multi-fidelity model. The LSTM in the proposed model, in addition to data compatibility, also identifies the main pattern of the model's gradual temporal evolution. This model is considered suitable for data-driven and non-intrusive problems due to its accurate predictions and generalization capabilities. The suggested approach has been applied to various problems, including the propagation of an electrical signal in excitable cells, the estimation of fluid flow around a cylindrical obstacle, and the Lotka-Volterra system representing the nonlinear interaction between prey and predator. In addition, Shang et al. [82] devised a method based on the multi-fidelity Kriging (co-Kriging) model to determine the Sobol index in the global sensitivity analysis of high-dimensional input variables. The co-Kriging model is first constructed for variance decomposition-based sensitivity analysis. Then, by fusing the information from HF and LF models, the accuracy of the Sobol index prediction is increased. The model formulates the sensitivity indices through several one-dimensional integrals, thereby reducing the computational burden. The performance of the method has been validated by comparing the accuracy of estimates on the Ishigami function, the Borehole model, the stochastic PDE problem, and the two-dimensional airfoil problem against multi-fidelity PCE, Kriging, and Latin Hypercube Sampling (LHS) methods. In a separate study, Le Gratiet and Garnier [156] implemented a recursive formulation for the multi-fidelity co-Kriging model. The core principle of this method is to construct the co-Kriging model through a series of independent Kriging models. This structure facilitates the execution of cross-validation operations at a higher speed. Employing the recursive formulation leads to a significant reduction in the complexity of the co-Kriging model compared to the original approach. The proposed method also has the potential to decrease the size and number of covariance matrix conditions required for constructing the surrogate model. The uncertainty quantification results of the recursive co-Kriging model predictions are compared to the conventional Kriging approach for emulating a complex hydrodynamic simulator.

Researchers in this field have highlighted the extensive applications of multi-fidelity modeling across a diverse range of prominent problems. For instance, Yang and Perdikaris [56] utilized a probabilistic DL method to facilitate uncertainty propagation in high-dimensional dynamical systems and maintain the robustness of surrogate model predictions in complex, noisy processes. This method enables surrogate training on stochastic samples from information sources of varying fidelity. The effectiveness of this approach has been illustrated through the regression of data containing random noise, multi-fidelity modeling of random processes, and the non-linear time-dependent Burgers equation.

Another multi-fidelity management technique focuses on adapting LF model improvements using HF model information during surrogate modeling [155]. This fundamental adaptation technique performs the correction of LF model outputs through either additive or multiplicative updates. In additive updating, the correction is determined by calculating the difference between high- and LF outputs. Conversely, multiplicative updating applies a correction based on the ratio between high- and LF outputs [155,157]. In this way, Yildiz et al. [79] provided the multi-fidelity LRA method as a substitute for HF analyses in problems affected by the curse of dimensionality. Improved model accuracy is achieved by using a limited number of HF simulations, with only a linear increase in the number of coefficients required to construct the LRA model. This results in the efficiency of the approach in quantifying the uncertainty of high-dimensional problems. The study utilized both additive and multiplicative corrections to implement multi-fidelity surrogate modeling. The effectiveness of the developed method was assessed on benchmarks, including four-dimensional Parkers, Borehole, and Sobol functions. Furthermore, to evaluate this method for real engineering problems in comparison with techniques such as single- and multi-fidelity PCE and Monte Carlo, the probability density function (PDF) results for a supersonic aircraft design model were presented. In another study, Li and Montomoli [92] introduced multi-fidelity deep neural networks (MF-DNNs) approach for uncertainty quantification in high-dimensional problems. This method addresses the curse of dimensionality by designing a DNN that utilizes multi-fidelity data, thereby overcoming the limitation of access to computational or experimental resources. The goal of this approach is to establish a multi-fidelity model using a combination of additive and multiplicative correction coefficients. The efficiency of MF-DNNs has first been demonstrated by approximating various benchmark functions. The accuracy of the method for aleatory uncertainty propagation is then evaluated by predicting the probability density function (PDF) and statistical moments of the target quantities on the benchmark functions. Additionally, the MF-DNNs approach has also been employed to model the physical flow inside an aircraft propulsion system in the presence of potential uncertainties arising from experimental measurement errors. Recent studies have explored the use of physics-based surrogate models to correct low-fidelity models. Xian and Wang [90] presented a coupled physics-data-driven surrogate modeling method to estimate the probability of rare events in high-dimensional structural and mechanical systems. This approach builds the error corrections of prediction in the low-fidelity model as a function of the physics-based

surrogate model output, therefore mitigating the curse of dimensionality in quantifying input uncertainties. The key components of this approach are parametric optimization for physics-based surrogates, active learning for efficient surrogate training, and GP regression for error correction. In this method, an importance sampling formula has been proposed to approximate the correction coefficient in estimating the probability of rare events using a surrogate model. The proposed method has been examined to construct physics-based surrogate models in static and dynamic system analysis problems, representing input uncertainties as a random field and a stochastic process, respectively. In another application of the adaptation technique, multi-fidelity adaptation through reduced models was performed to more effectively adapt to HF data. With this insight, Navaneeth and Souvik [75] used sparse AS algorithms to identify high-dimensional inputs on a low-dimensional manifold. This approach integrates the AS technique with surrogate modeling. The LF surrogate model, specifically the sparse polynomial chaos expansion (SPCE) in this case, assists in computing the gradient information of the response variable. Moreover, the AS helps train the surrogate model. The relationship between the low-dimensional manifold inputs and the output is mapped through the HF surrogate model known as a hybrid polynomial correlated function expansion. The accuracy and efficiency of the recommended framework for solving reliability problems have been proven on the Sobol function, a composite beam, and a 25-element spatial truss. Zaroni et al. [99] also investigated the idea of discovering a set of active inputs in conjunction with multi-fidelity models. They focused on two multi-fidelity Monte Carlo estimators that rely on linear and nonlinear dimension reduction techniques based on active subspaces and autoencoders, respectively, to propagate uncertainty using computationally expensive models. The active subspaces and autoencoders are utilized not only for the purpose of reducing dimensionality but also as tools to enhance the sampling process in constructing the multi-fidelity Monte Carlo estimator. The discovery of a separate set of active inputs for high-fidelity and low-fidelity models determines the important directions. By transforming the subspaces produced by these directions into a common distribution, a shared space is formed that can be used for all levels of model fidelity. This method relies on the concept that the important directions of the high-fidelity and low-fidelity models are well aligned. Accordingly, these two approaches can lead to improved sampling in the common correlation space between the high-fidelity and low-fidelity models, thereby creating a multi-fidelity estimator with lower variance. A simple example and two high-dimensional functions are used to highlight the advantages of the approach over the single-fidelity Monte Carlo method.

Filtering is another technique for multi-fidelity management. This technique explores the random space using the LF model and only invokes the HF model if the filter or criterion indicates the inaccuracy of the LF model for the candidate data [155]. In an example of combining adaptation and filtering techniques, Dhulipala et al. [78] suggested a framework for active learning with multi-fidelity modeling, emphasizing the effective estimation of rare events in high-dimensional spaces. This method begins by establishing an initial LF model. To increase the accuracy and correctness of the LF model predictions, a GP correction term derived from previous HF simulation calls is added. Active learning functions are then used to dynamically guide the decision to call the HF model, enhancing the robustness of the model in estimating the probability of smaller failures. The proposed framework has been applied to various benchmarks, such as the four-branch, Rastrigin, and Borehole functions, as well as two finite element model case studies, to accurately calculate the failure probability while requiring only a few HF model calls. Gong and Pan [100] have also employed the criterion approach. They developed a multi-fidelity sequential Bayesian experimental design framework to quantify statistical quantities, focusing on rare events. This method consists of two main components: a cheap multi-fidelity GP surrogate model trained on multi-fidelity samples and an acquisition function that measures the expected gain at each computational budget. Maximizing the acquisition function allows the method to adaptively select the best next sample based on the fidelity level and location in the sampling space. Introducing an analytical formula to compute the acquisition function and its derivative has enabled the implementation of this algorithm through gradient-based optimization for high-dimensional problems. The effectiveness of the method has been demonstrated in comparison to single-fidelity and bi-fidelity with hierarchical fidelity approaches for the Forrester function, a two-dimensional stochastic oscillator, and a hydrological borehole model. Additionally, the advantage of this approach has been demonstrated in approximating the probability density function (PDF) of rare ship roll motion in irregular ocean waves.

4.3. Surrogate models based on advanced sampling techniques

Developing surrogate models typically requires a large number of input-output sample data points. Collecting this data for complex, high-dimensional engineering problems is often computationally expensive and scarce. As a result, numerous researchers have focused on efficient information management to construct surrogate models using fewer data points. In this context, experimental design emphasizes sample selection to achieve surrogate models with sufficient accuracy. Two primary approaches to experimental design are nonadaptive and adaptive methods, based on the classification of the literature [18,43,158]. Traditional nonadaptive methods blindly fill the input space without considering the physical relationship between input and output variables, often leading to an excessive waste of computational resources during sampling. On the other hand, adaptive methods incorporate the problem's physics during new sampling [18,158]. When using sequential adaptive sampling to construct surrogate models, the goal is to effectively explore unknown regions or areas of particular interest by repeatedly updating and refining the intermediate models. This is achieved by leveraging information obtained from previous data to intelligently add new samples until convergence to the optimal solution. Therefore, utilizing adaptive methods can enhance the efficiency, accuracy, and flexibility of the surrogate model in dealing with high-dimensional spaces with fewer samples. This approach also enables dynamic sampling to create efficient surrogate models, particularly for data with rare populations, uneven distributions, or difficult access. Recent literature reviews reveal ongoing efforts to not only address the growing trend toward advanced adaptive sampling approaches but also to improve traditional methods. As shown in Fig. 4, the development of surrogate models based on advanced sampling methods, including both adaptive and nonadaptive (in developed form) has occupied a substantial portion of the studies reviewed in the recent decade.

Researchers have been motivated by the development of sequential importance sampling methods, traditional approaches

employed when direct sampling from the target distribution is challenging or impossible. Papaioannou et al. [38] conducted a study utilizing a Metropolis-Hastings conditional sampling algorithm. This algorithm includes a parameter that adapts to the optimal sample selection probability. Moreover, this approach extends the sequential importance sampling method to handle high dimensions in estimating failure probability. The method obtains samples from each distribution by sequentially resampling weighted samples from the previous distribution and moving the resulting samples from the Markov Chain Monte Carlo method. The success of the proposed method has been assessed on two high-dimensional problems: a noisy limit-state function and a high-dimensional linear limit-state function. In another study, Xu and Wang [61,62] proposed an efficient probability estimation technique for rare events with narrow bounds in high-dimensional problems with small sample sizes. The method employs sequential importance sampling, which sequentially identifies and updates the optimal sampling distribution. To estimate the narrow bounds, the technique leverages the concept of confidence intervals, which rely on the asymptotic property of the proposed estimator. The suggested method has been evaluated for accuracy and efficiency in handling a linear function of random variables. Subsequently, the idea of extending importance sampling methods to enhance efficiency in high-dimensional spaces was further explored in other investigations. Uribe et al. [67] suggested that the cross-entropy method for importance sampling can be an efficient tool to estimate failure probability in rare events with high dimensionality. This approach focuses on identifying the intrinsic, low-dimensional structure of rare events. The method searches for a connection between rare event simulation and Bayesian inverse problems, allowing for the adaptation of dimensionality reduction from Bayesian inference. It enables the approximation of the optimal low-dimensional biased distribution in the cross-entropy method. This computational framework has been applied to linear and nonlinear reliability problems, encompassing linear and quadratic limit-state functions as well as a two-dimensional steel plate under plane stress.

Improving the efficiency and accuracy of classical methods, which depend on the number and location of experimental samples around the design points, has also been studied in reliability analysis. For instance, Hadidi et al. [47] implemented the exponential response surface method along with sample location selection to decrease the computational cost of high-dimensional problems while increasing the accuracy of failure probability estimation. Compared to the quadratic response surface, applying the exponential surrogate model results in a significant reduction in the number of experimental samples required to determine the model coefficients. Moreover, precise sample location selection based on the vector length of the random variable and the angle between two vectors improves the model's accuracy. The efficiency and accuracy of the proposed method have been illustrated on three non-structural two-dimensional examples, as well as a 23-rod truss and a two-dimensional frame structure. Concurrently with these investigations, efforts have been made to develop sampling selection methods for determining the location of a finite number of samples. The goal is to produce a regression result as close as possible to the results obtained from larger sets of candidate samples. For example, Shin and Xiu [39] designed a nonadaptive quasi-optimal sample set to address the challenge of collecting a large number of dense candidate samples while ensuring the accuracy of the regression result. The quasi-optimal set is chosen by maximizing a quantity that measures the mutual column orthogonality and the determinant of the model matrix. This method selects the sampling locations before collecting the samples, thus obtaining a near-optimal regression result for any number of samples, making it a cost-effective approach. The efficiency of this method has been confirmed through a greedy algorithm in comparison with the generalized polynomial chaos Galerkin method on Genz benchmark functions and a stochastic diffusion equation. Simultaneously with the growing development of optimal sampling techniques, other researchers have focused on advancing classical criteria for selecting sample points. For example, Hadigol and Doostan [50] established a hybrid sampling method called Alphabetic-Correlation-Optimal for high-dimensional problems. This approach utilized alphabetic optimality criteria to optimize experimental design within the coherence-optimal framework. Compared to other sampling methods, it can more accurately find the PCE coefficients through least-squares approximation. Numerical examples, such as manufactured functions, the nonlinear Duffing oscillator, and the useful life of batteries, have been utilized to compare the performance of different sampling methods. In another study, Rushdi et al. [43] employed a Voronoi piecewise surrogate (VPS) without inducing the sampling location. This approach overcomes the curse of dimensionality associated with standard domain decomposition and enhances local modeling accuracy in high-dimensional spaces. The main idea behind VPS is to decompose the high-dimensional space using the implicit Voronoi tool around the sample points as seeds. This tool assigns the samples to cells and identifies the neighboring cells through local hyperplane sampling by simply searching for the nearest seed. The VPS method then constructs a global surrogate model by establishing a neighborhood network between the cells and constructing a low-order local piece using the neighborhood information of each cell. This approach has been employed on several benchmark functions and the integral of a high-dimensional function.

Constructing adaptive surrogate models typically involves three steps: (1) starting from an initial experimental design in accordance with the response; (2) calibrating an initial surrogate model; and (3) evaluating the constructed model through an iterative process. In the iterative process, the initial surrogate model is updated by creating intermediate models after selecting the next sample points to incorporate into the experimental design. The quantity of interest is then estimated from the current surrogate model, and the model's performance is evaluated for each data point until convergence is achieved. An example of this approach is provided by Schöbi et al. [41], who improved the accuracy of surrogate models while reducing the number of runs required for complex computer models. This was achieved by combining the polynomial chaos Kriging (PC-Kriging) method with an experimental design enrichment algorithm, also known as adaptive experimental design. The formulation of the PC-Kriging method equalizes the small probability of failure in rare events and extreme quantiles. This approach leverages the computational benefits of distributed computing by emphasizing the selection of the most suitable additional samples in the experimental design and adding multiple samples in each iteration of the algorithm. By equipping the algorithm with a new stopping criterion, the monitoring of convergence is enhanced compared to existing criteria, thereby reducing computational resources. The results of the proposed method are presented in the four-branch and Borehole functions, as well as the two-dimensional truss structure problems.

The distribution of sample points over the domain to obtain maximum information about the system's behavior is one of the

challenges in constructing adaptive surrogate models. Therefore, adaptive modeling requires a trade-off between filling the design space and data density distribution. Filling the design space without considering the outputs allows for the acquisition of knowledge from the entire design space. On the other hand, data density distribution through sample output analysis identifies regions of interest, such as those near optimal points or with discontinuities, to capture complex system behavior or discover regions with large errors in intermediate models. Accordingly, Sadoughi et al. [51] suggested a sequential sampling method called Sequential Exploration-Exploitation with Dynamic Trade-off (SEEDT) to study the reliability of complex systems with high dimensions. This method sequentially places an efficient set of sample points to construct the Kriging model. By dynamically introducing a trade-off coefficient, a beneficial balance between exploration and exploitation is established. The effectiveness and accuracy of SEEDT have been evaluated using four mathematical functions and a piezoelectric energy harvester, demonstrating its advantages over existing sampling methods.

The literature generally focuses on two approaches for developing sequential sampling methods and addressing surrogate model error to evaluate the performance of intermediate models. The first approach involves sequential selection using enrichment criteria. The second approach focuses on adaptive selection of the optimal experimental design (or learning function). Fajraoui et al. [44] used adaptive sequential sampling to deal with high-dimensional challenges and improve PCE modeling accuracy. Their method sequentially creates the optimal experimental design space, focusing on the sparsity characteristic of the underlying model to determine the expansion coefficients in the least-squares minimization. The efficiency, accuracy, and stable behavior of this method have been proven on Ishigami and Sobol functions, a two-dimensional truss, and a two-dimensional diffusion problem compared to several advanced methods. Applying learning functions to select the most suitable samples from a set of candidate points for stepwise refinement of surrogate models has also been pursued in reliability analysis studies. For instance, Xiao et al. [49] recommended a sequential adaptive sampling method to increase surrogate model accuracy while reducing the number of training samples. This approach chooses an NN-based surrogate model, which is well-suited for high-dimensional spaces and can capture multiple failure modes accurately. This method entails the development of three learning functions for each iteration, designed to select samples further away from current points and closer to the limit-state functions. The new sampling process incorporates knowledge gained from each iteration. The accuracy and efficiency of this proposed method have been estimated on a system with four performance functions, a nonlinear oscillator, and a cantilever tube. In another study, Kapusuzoglu et al. [18] applied adaptive sampling and output dimensionality reduction for surrogate modeling to quantify the uncertainties of multi-physics dynamic systems, which have high-dimensional spatio-temporal outputs. This method first projects the high-dimensional output onto a lower-dimensional space using the rSVD dimensionality reduction technique. Then, it creates a surrogate model in the low-dimensional space based on several ensemble learning methods. Next, the model accuracy is estimated using GP, and an adaptive sampling technique based on the learning function is employed to identify new training points. This iterative process continues until the stopping criteria, i.e. achieving acceptable model accuracy or reaching computational resource limitations (such as the maximum number of training points), are met. The effectiveness of the proposed method in minimizing the required model evaluations has been investigated on seven benchmark functions with varying complexities, as well as a turbine blade thermo-mechanical analysis problem for a gas turbine engine blade. The literature review indicates that the adaptive sampling approach leverages available information to enhance the power of surrogate models. In a study, Xu and Wang [76] developed an adaptive surrogate modeling method to address high dimensions and missing values in multi-dimensional or multi-fidelity datasets. This method combines the Bayesian GP latent variable model with the adaptive sampling approach to effectively utilize partially observed information and improve training sample selection. A mathematical example and a problem with multiple outputs and multi-dimensional inputs demonstrate the performance of this combined method in the well-organized use of all available observed data. A quick overview of the most recent research in this field demonstrates the sequential selection of new samples that combines a learning function with other surrogate modeling techniques for high-dimensional systems. As an illustration of the research, Ji et al. [81] used an adaptive dimensionality reduction technique and the Kriging model to alleviate the curse of dimensionality in the reliability analysis of systems with high-dimensional outputs. This method performs Kriging modeling in the reduced space, following PCA on the time-dependent primary output. A learning function and a stopping criterion are defined to iteratively select the optimal training samples. The proposed learning function focuses on ensuring prediction accuracy over the entire time range, and the failure probability is directly affected by the stopping criterion. This single-loop adaptive process simultaneously selects training samples, performs dimensionality reduction, and constructs the Kriging model. The efficiency and accuracy of the method have been confirmed on a time-dependent mathematical model, a roof truss structure, a beam under stochastic loads, and a stone arch bridge under hurricane loads.

Active learning algorithms have been widely employed in the literature on adaptive models. These algorithms can focus on the most informative regions of the data and intelligently select the next samples by evaluating the current surrogate model status. Furthermore, these algorithms iteratively refine the intermediate model to achieve optimal exploration of the design space. Characteristics such as the ability to interactively query the oracle (or other information sources) to determine the desired outputs at new sample points and assess the model accuracy on-the-fly have also made active learning a suitable tool for optimal adaptive experimental design of complex systems. As an example, Hong et al. [73] established an active learning function called the potential risk function to update surrogate models in reliability analysis. A trade-off coefficient maintains the balance between exploration and exploitation in the learning function. Defining a convergence criterion ensures the prevention of surrogate models prone to premature convergence. The suggested method has been validated to demonstrate the robustness and performance of Kriging and Radial Basis Function (RBF) models on various cases, including the four-branch function, a 23-bar truss structure, a cantilever tube, and a journal bearing of a gear pump. It is noteworthy that recent findings indicate that employing a hybrid surrogate model of the radial basis function neural network (RBFNN) can significantly improve the efficiency of uncertainty propagation analysis [3]. In recent studies, Kim et al. [91] also developed an adaptive metamodeling approach for reliability analysis in high-dimensional problems and rare event simulation.

This approach combines three key components: active subspace, heterogeneous GP, and active learning. The active subspace algorithm identifies the prominent features of high-dimensional models, effectively reducing the dimensionality of the problem. For the reduced feature space, the heterogeneous GP is constructed as a surrogate in the lower dimensions. The heterogeneous GP enables the estimation of the prediction errors caused by the surrogate modeling in the reduced space. The active learning algorithm adaptively guides the surrogate model's training toward the target region. A nonlinear mathematical function, a space truss structure, and a steel lattice transmission tower have investigated the efficiency and accuracy of the heterogeneous GP-based adaptive subspace method. Simultaneously, Nguyen et al. [104] developed a robust active learning framework using uncertainty quantification and flexible metamodels (AUQ-meta) for reliability analysis in nonlinear and high-dimensional problems. This hybrid method consists of three main components: a model-based prediction using advanced machine learning models, a quantile regression module for uncertainty assessment, and an adaptive reliability analysis process. The use of advanced machine learning models such as LGBM and XGBoost, along with quantile regression, significantly improves the performance of structural reliability analysis. Selecting potential candidate samples using active learning for model training in this method leads to better performance and accelerates the reliability analysis. The effectiveness of the AUQ-meta method in terms of accuracy and efficiency of computations has been verified in three benchmark reliability analysis functions, a hyperstatic truss, a 1-story frame structure with a viscous damper under ground motion, and a post-tensioned reinforced concrete beam.

A general overview of the studies indicates that the bootstrap method has been utilized to measure the local error of surrogate models. This tool subsequently identifies the most suitable candidate points through active learning algorithms to add to the experimental design space. For instance, Marelli and Sudret [45] proposed an adaptive method called Active Polynomial-Chaos-Bootstrap Monte Carlo Simulation (APCB-MCS) to address problems with high dimensions and a small initial experimental design space. In this method, an active learning algorithm dynamically improves the experimental design. In this greedy algorithm, the bootstrap method is applied to estimate the local error of the expansion parameters. The most desirable candidate points are then selected to add to the experimental design. The performance of the APCB-MCS method has been evaluated on the four-branch function and the slope stability problem. In a separate investigation [55], these researchers developed a hybrid approach combining bootstrap sampling and SPCE to estimate the local error of surrogate models. This approach iteratively updates a small initial experimental design using an active learning algorithm. This greedy algorithm relies on selecting the most informative set of samples to enrich the experimental design space of complex systems. The success of the suggested approach for determining the failure probability in the four-branch function as well as two high-dimensional engineering applications, including a two-dimensional truss and a frame structure, has been demonstrated.

Given the substantial reduction in computational costs facilitated by surrogate modeling, researchers have embraced the idea of constructing adaptive models in combination with other proposed solutions to tackle the challenge of high dimensionality in the uncertainty quantification of complex systems. For example, Zhou and Peng [8] presented a modified method known as Active Learning-Based Sampling Combined with DL and GP regression (AL-DLGP) to mitigate the curse of dimensionality in reliability problems. This method integrates a DL architecture, specifically an autoencoder, with the GP regression model. The parameters of these two components are obtained using a joint optimization scheme. Adaptive sampling in this hybrid model is performed based on active learning principles. The DLGP model is then combined with the probability density evolution method to analyze reliability. This method for handling high-dimensional reliability problems has been demonstrated to achieve a balance between accuracy and efficiency in the static and dynamic analysis of a steel frame structure. In another study, Yin [11] conducted a comparison of three techniques for quantifying uncertainty in input spaces with a high number of dimensions. The first two methods deal with numerical inputs, while the third combines numerical and image data. The first method carries out reliability analysis after reducing the dimensions by fixing the less significant variables in the first most probable point step. In the second method, GP modeling is performed following dimension reduction using generalized sliced inverse regression. An active learning algorithm identifies the failure boundary and iteratively adds optimal training samples. The third method first converts the numerical data to images and then combines them with the available image data. Convolutional neural networks and GP regression combine to predict model uncertainty. The improved accuracy of these methods has been evaluated on various benchmark problems, including the parabolic function, a cantilever beam, a 52-bar geodesic truss, a 25-story shear frame under stochastic seismic excitation, the MNIST dataset, and heat transfer problems. Furthermore, Guo et al. [87] developed an adaptive surrogate modeling approach for iterative computations in high-dimensional problems, addressing both input and output spaces. This method utilized the SVD dimension reduction technique to identify the output space features. Next, it employs the active subspace methodology for input dimension reduction, establishing a mapping from the input space to each feature in a low-dimensional subspace. This approach balances exploration and exploitation to select limited samples and improves modeling through an adaptive learning technique for both input and output. Specifically, the active learning function is used for new sampling in each iteration until the stopping criterion is satisfied and the desired accuracy level is reached. The proposed method has been implemented on three benchmark functions, as well as simulating a component under a residual stress field. He et al. [94] also pursued the approach of using adaptive learning in conjunction with other methods. These researchers suggested an adaptive data-driven subspace polynomial-dimensional decomposition method for uncertainty quantification in high dimensions. This method involves reconstructing the probability density function (PDF) of the input variables based on a data-driven zero-entropy criterion. Subsequently, data-driven orthogonal polynomial bases are proposed utilizing the generated PDFs. An approximate active subspace method is used to improve the efficiency of determining the desired subspace. An adaptive learning algorithm is also employed to further enhance the data-driven polynomial-dimensional decomposition model based on sparse Bayesian learning theory. The proposed method has been validated using three synthetic functions, as well as a cantilever tube structure, a 72-bar space truss, and a box structure, and compared to four dimension reduction techniques. Simultaneously, Wan et al. [105] employed the SS-MASVM method, which combines the multi-class adaptive support vector machine (MASVM) model and subset simulation (SS), to estimate the

failure probability of complex systems with large dimensions. The core of this method includes learning functions, K-means clustering integration, and the SS process. SS enables more efficient and accurate estimation by dividing the problem into multiple failure regions. The MASVM model provides more reliable and accurate predictions by establishing correlations between different failure modes. Furthermore, it allows the surrogate model to adapt to different failure regions, thereby improving computational efficiency and accuracy while requiring fewer samples. Various scenarios have confirmed the effectiveness of the SS-MASVM method, including the multi-sharp corner example, nonlinear exponential and Rackwitz functions, a classic oscillator problem, an arc-stiffened plate, and a carriage structure.

A skim of recent research indicates the use of ensemble learning capabilities to merge the predictions of multiple surrogate models to achieve better effectiveness in adaptive models [60]. This approach can simultaneously leverage the best features of each model. For example, Cheng and Lu [60] studied a competitive multiple surrogate model ensemble as a substitute for a single surrogate model in the reliability analysis of complex systems. Their method introduces an active learning framework for reliability analysis based on the ensemble learning of multiple surrogate models. The final prediction is made using the weighted average of several surrogate models, with the local error estimated through the variance of the multiple surrogate models. Consequently, the active learning algorithm is iteratively applied by adding new sample points to the initial experimental design in regions with high error and near the limit state to enhance prediction accuracy. The proposed method has been validated for Kriging, PCE, and SVR in estimating the failure probability of the four-branch series system, a 2-bar supporting structure, a 25-bar truss structure, and a heat conduction problem. In other research, the technique of using an adaptive ensemble of surrogate models in combination with adaptive sampling has been employed. In this technique, researchers also utilize hybrid modeling approaches to combine the advantages of various models in order to overcome the challenge of insufficient accuracy in surrogate models [159]. For instance, Zhou et al. [21] proposed an Adaptive Ensemble of Surrogate Models based on Hybrid Measures (AESMHM) to improve prediction accuracy in regions farther from the highest sample density. The AESMHM method calculates the hybrid weight of the surrogate models, such as Kriging, Support Vector Machines (SVM), and RBF, by modifying the overall weight matrix at representative sample points with the local weight matrix. The impact matrix adjusts the ratio between the global and local weights. The learning function simultaneously selects the points adaptively and updates the surrogates until the convergence criterion is met. Three numerical examples of solving high-dimensional and nonlinear reliability problems demonstrate the improved modeling efficiency and accuracy across all sample points. The suggested method also presents effective solutions to real-world reliability issues in an engineering application and a composite material structure.

A review of the literature reveals that SPCE is a widely used tool for quantifying uncertainty in complex or high-dimensional systems. This approach retains only a small number of the significant basis functions in the polynomial chaos approximation [128]. It reduces computational cost by introducing sparsity in the expansion coefficients. The primary concept behind SPCE is to express random variables using a sparse set of basis functions rather than expanding them in terms of all possible monomials. These basis functions are selected through various techniques, such as adaptive algorithms that iteratively update and upgrade the basis set while considering the importance of approximating the objective function. Acknowledging the effective portion of adaptive sampling methods in the development of SPCEs, this approach is introduced as part of adaptive surrogate models in this review. It is evident that the adaptive algorithms employed in constructing these models not only strike a balance between accuracy and efficiency but also enable a more efficient space search and provide insights into the system's behavior. The compatibility and flexibility of the method make models appropriate for a wide range of applications, especially when dealing with high-dimensional spaces. The investigation of current algorithms for creating SPCEs, as well as the metrics introduced in research by Lüthen et al. [111], have provided guidance on selecting the most appropriate approach for constructing these models in practical applications. Therefore, this review focuses on showcasing instances of research conducted on high-dimensional systems. As part of the studies, Kersaudy et al. [31] reduced computation time in numerical dosimetry and investigated the influence of uncertain input parameters on the specific absorption rate. They presented the LARS-Kriging-PC model, which implements least-angle regression (LARS) to maintain the most influential polynomials in the sparse representation of PCEs and then uses these polynomials as regression functions in a universal Kriging model. The proposed method has been applied to three benchmark functions and a full-scale exposure problem. Finally, a global sensitivity analysis was executed on the fetal exposure to electromagnetic fields model, and the Sobol sensitivity indices were computed using Monte Carlo simulation. In another study, Kubicek et al. [30] presented a non-intrusive high-dimensional model representation (Cut-HDMR) methodology for sensitivity analysis. Their method decomposes the random space into sub-domains, each of which undergoes separate interpolation using a multi-dimensional Lagrange interpolation technique. This sampling method, based on the sparse grid idea, overcomes the curse of dimensionality by reducing the number of samples in high-dimensional spaces. Subsequently, the sensitivity of a variable or a combination of variables is determined using Monte Carlo simulation for each interpolated sub-domain, utilizing Sobol sensitivity indices. The recommended interpolation method has been applied to the well-known Borehole problem. Furthermore, Deman et al. [36] utilized the SPCE method to compute the Sobol indices of a multi-layer hydrological model. This approach allowed them to calculate the mean lifetime expectancy of water molecules leaving the boundary of a highly confining layer within a numerical model. The model used the sparsity feature and powerful regression solvers to reduce the number of computational evaluations required for high-dimensional problems [111]. It is highly effective in accurately representing the mean lifetime expectancy in the hydrological model. Other researchers have pursued the necessity of addressing the curse of dimensionality in the development of PCEs using concepts such as simultaneous sparsity. For example, Jakeman et al. [32] provided a basis selection method to accurately solve high-dimensional problems where the determination of high-order PCE bases is impossible. This method can apply ℓ_1 -minimization to the adaptive determination of large coefficients in PCEs. This adaptive modeling produces anisotropic basis sets containing more important terms and a limited number of unimportant ones, improving mutual coherence and enhancing ℓ_1 -minimization's performance. To assess the key features and accuracy of this method, the corner peak function, random oscillator,

diffusion equation, and resistor network have been employed. Additional studies have explored the ℓ_1 -minimization method, which has been the focus of other research. For instance, Savin and Hantrais-Gervois [65] developed a non-intrusive approach to identify SPCE surrogates for handling high-dimensional variable spaces and mitigating the curse of dimensionality. This methodology is based on an ℓ_1 -minimization algorithm for polynomial reconstruction in a regression framework. It offers the advantage of requiring a smaller sample size compared to extensive sampling. The proposed method also provides the range of expansion coefficients and their orders on a polynomial basis. This technique has been implemented to determine surrogates for PDFs and sensitivity indices in the numerical simulation of fluid flow on an aircraft wing. The study by Loukrezis et al. [101] also pursued the adaptive selection of polynomial bases utilizing alternative approaches. They employed the multivariate sensitivity-adaptive polynomial chaos expansion (MVSA PCE) method to tackle the curse of dimensionality in the rapid growth of polynomial bases, which arises due to the large number of input dimensions and high polynomial degrees. This method uses multivariate sensitivity analysis metrics to select the adaptive bases. These metrics can be estimated by post-processing the computed PCE and providing a common anisotropic polynomial basis for response estimation. Using a common polynomial basis enables vectorized ordinary least squares (OLS) solvers to lead to a significant reduction in computation time. The performance and accuracy of the MVSA PCE method to estimate uncertainties in calculating the deflection of a simply supported beam under a uniform load, simulating the response of an induction motor connected to the power grid and mechanical load, and modeling the power flow in the European high-voltage transmission network have been assessed.

Recent studies have investigated the effectiveness of using PINNs in combination with uniform sampling and coverage of the input space. In this field, Baisthakur and Fitzgerald [103] aimed to overcome the evaluation of a large set of simulations, especially in reliability analysis and fatigue estimation of wind turbines, by utilizing PINN surrogate models. PINNs can leverage prior knowledge of the system to guide the optimization process, leading to solutions that are more physically meaningful, reliable, and feasible. Since PINNs can incorporate physic-based constraints into the model training process, the approach converges faster than purely data-driven NNs. This method uses a weighting factor to achieve an optimal balance between data fitting and physic-based constraints during optimization. Furthermore, employing the Sobol sequence enhances the sampling rate and ensures uniform coverage of the input space. Applying PINNs to data produced from dynamic numerical simulations of a 15MW wind turbine replaced the time-consuming root-finding process in blade element momentum, resulting in a more efficient and accurate response evaluation compared to conventional data-driven approaches.

5. High-dimensional mathematical benchmark functions

Validation is fundamental in surrogate modeling. To evaluate the effectiveness of surrogate modeling techniques in uncertainty quantification, it is necessary to test these algorithms across various scenarios. Evaluating the efficiency of surrogate models is not solely reliant on high-dimensional benchmark functions. Instead, each benchmark problem, whether low or high-dimensional, offers

Table 2

- High-dimensional mathematical benchmark functions employed to evaluate the performance of surrogate modeling techniques for uncertainty quantification of engineering problems

Function	Input Dimensions	Input Variables Distribution	Reference	Year
Sobol function	20	Uniform	Kersaudy et al. [31]	2015
			Konakli and Sudret [33]	2016
			Moustapha et al. [114]	2018
			Lataniotis [6]	2019
	30, 60	Uniform	Peng et al. [69]	2021
Morris' function	20	Uniform	Arseniyev et al. [160]	2014
			Gu and Wu [161]	2014
			Blatman and Sudret [162]	2015
Specific function	20	Exponential	Sundar and Shields [163]	2016
			Zhou et al. [21]	2022
Ellipsoid function	20	Uniform	Zhao et al. [164]	2020
Griewank function	20, 40, 60, 80, 100	Uniform	Xu and Wang [68]	2021
Rackwitz function	≤ 25	Normal	Sadoughi et al. [51]	2017
	100, 250, 2000	Lognormal	Zhou and Lu [66]	2020
	40, 100, 250	Lognormal	Yin [11]	2022
	40	Lognormal	Wan et al. [105]	2024
Rosenbrock function	30	Uniform	Cheng and Zimmermann [84]	2022
Dixon-Price function	30, 50	Uniform	Zhao et al. [164]	2020
			Cheng and Zimmermann [84]	2022
d -dimensional paraboloid	100	Standard normal	Depina et al. [165]	2016
			Nguyen et al. [104]	2024
High-dimensional nonlinear function	100	Uniform	Lüthen et al. [111]	2021
			UQLab example [166]	2022
Parabolic function	100	Standard normal	Yin [11]	2022
Linear function	100, 1000	Standard normal	Papaioannou et al. [38]	2016
			Xu and Wang [61,62]	2020
Corner-peak function	100, 1000, 10000	Uniform	Jakeman et al. [32]	2015

valuable insights into efficiency metrics of surrogate models. Therefore, it is important to select appropriate benchmark problems to assess the performance of surrogate modeling techniques, irrespective of dimensionality.

The comprehensive literature review (as shown in Table 1), covering practical problems with a wide range of input variables and not strictly confined to high dimensionality, supports this point. As a result, selecting suitable benchmark problems for specific research purposes is at the researcher's discretion and is beyond the scope of this article. This article focuses solely on a selection of mathematical benchmark functions commonly used in the literature for surrogate modeling to propagate uncertainties in high-dimensional engineering problems, irrespective of the specific surrogate modeling techniques.

As the number of input variables or problem dimensions rises, the search space required for sampling in constructing surrogate models expands exponentially. As indicated in Table 2, the threshold for defining high-dimensional problems is set at 20 random input variables. This boundary is not selected based on the complexity of the problem but provides a clear framework for assessing the performance of surrogate modeling in propagating input uncertainties.

Furthermore, Table 2 highlights that the literature has employed a variety of high-dimensional benchmark functions. However, there is no uniform framework available to validate the proposed methods. The selection of appropriate benchmark functions depends on the problem type, as multiple criteria must be evaluated [167]. Consequently, researchers may choose to utilize other benchmark functions to validate the effectiveness of their techniques [168]. The benchmark functions discussed in this section aim solely to demonstrate the capabilities of a particular set of high-dimensional problems for quantifying response uncertainty. The details provided in the upcoming subsections enable researchers to consider several factors: (1) the utilization of each function in various approaches to uncertainty quantification as highlighted in the reviewed literature, (2) the significance of the function structure for its specific application, and (3) the formulation of the functions for different numbers of input random variables when employing benchmarks to validate their developed methods.

5.1. Sobol function

The literature commonly employs the *Sobol function*, also known as the *g-function*, as a benchmark to evaluate the performance of developed surrogate modeling techniques in high-dimensional problems for quantifying uncertainty. Equation (22) presents the Sobol function [31,33,169].

$$g_1 = \prod_{i=1}^d \frac{|4x_i - 2| + c_i}{1 + c_i} \quad (22)$$

where $x_i = \{x_1, \dots, x_d\}$ are independent random input variables with uniform distributions between 0 and 1. The dimension d can be tuned to handle a varying number of input variables. The coefficients $c_i = \{c_1, \dots, c_d\}$ are non-negative constants. The coefficients c_i can be adapted to control the relative significance of a specific dimension or feature of the input variables [114].

Some studies have set the coefficients c_i for $d = 20$ using the values of Equation (23) [31,33].

$$c_i = \{1, 2, 5, 10, 20, 50, 100, 500, 500, \dots, 500\} \quad (23)$$

It is abundantly clear that the effect of each input variable x_i on the function output has an inverse relationship with the coefficients c_i . It means that a small (usually large) value of c_i leads to a significant (or negligible) contribution of x_i to the variance of the function output [6]. In general, the dimension index i is important when $c_i \rightarrow 0$ and its importance becomes negligible for $c_i \geq 100$ [114].

The literature recommends utilizing coefficients c_i from the values of Equation (24) to evaluate surrogate models for high-dimensional real-world applications where the less important input variables are probable. In Equation (24), the level of importance across dimensions alternates [114].

$$c_i = \begin{cases} 3(i-1) & \text{if } i \text{ is odd,} \\ 3i & \text{if } i \text{ is even.} \end{cases} \quad (24)$$

The Sobol function is usually employed as a benchmark in global sensitivity analysis due to its complexity (highly nonlinear and non-monotonic relationship) and the availability of analytical sensitivity indices [31,33,169]. Moreover, an overview of studies reveals that structural reliability analysis has utilized the Sobol function with dimensions $d = 30, 60$ [69].

5.2. Morris' function

The *Morris function* is applied to assess the effectiveness of surrogate models in high-dimensional spaces while conducting global sensitivity analysis [162-164]. This function is identified in Equation (25) [170].

$$g_2 = \beta_0 + \sum_{i=1}^{20} \beta_0 w_i + \sum_{i<j}^{20} \beta_{ij} w_i w_j + \sum_{i<j<l}^{20} \beta_{ijl} w_i w_j w_l + \sum_{i<j<l<s}^{20} \beta_{ijls} w_i w_j w_l w_s \quad (25)$$

where w_i is obtained from Equation (26).

$$w_i = \begin{cases} 2 \left(1.1 \frac{x_i}{x_i + 0.1} - 0.5 \right) & \text{if } i = 3, 5, 7 \\ 2(x_i - 0.5) & \text{otherwise} \end{cases} \quad (26)$$

where $x_i = \{x_1, \dots, x_{20}\}$ are random input variables with uniform distributions between 0 and 1. The coefficients β_i can control the importance of the inputs variables. Equation (27) specifies the coefficients of the function [170].

$$\begin{cases} \beta_i = 20 & \text{for } i = 1, \dots, 10 \\ \beta_{ij} = -15 & \text{for } i = 1, \dots, 6 \\ \beta_{ijl} = -10 & \text{for } i = 1, \dots, 5 \\ \beta_{ijls} = 5 & \text{for } i = 1, \dots, 4 \end{cases} \quad (27)$$

Other coefficients $\beta_0 = 0$, $\beta_i = (-1)^i$, and $\beta_{ij} = (-1)^{i+j}$ are considered.

5.3. Specific function

A *specific function* has been employed as a benchmark to investigate the high-dimensionality effects on the performance of surrogate models in reliability analysis [21,165]. Equation (28) displays the global formula of this function [171,172].

$$g_3 = \pm \sum_{i=1}^{20} x_i \mp c \leq 0 \quad (28)$$

where $x_i = \{x_1, \dots, x_{20}\}$ are independent random input variables that follow exponential distributions with a mean value of 1. The function in the u -space becomes highly nonlinear when a probability transformation is applied, as demonstrated in equation (29).

$$g_3 = \mp \sum_{i=1}^{20} \ln[\Phi(-U_i)] \pm c \leq 0 \quad (29)$$

In equation (29), U_i are independent variables with normal distributions. The value of c is $-c = \{16.175, 11.077, 8.951, 7.435, 6.277, 5.343\}$ for the convex surface (sum of random variables) and $+c = \{25.900, 31.856, 36.720, 41.050, 45.067, 48.932\}$ for the concave surface (difference of random variables) [171,172]. In reliability analysis, the negative and positive signs in equation (29) correspond to the lower and upper tails of the failure probability distribution for this function, respectively [172].

5.4. Ellipsoid function

The *ellipsoid function* has been utilized to evaluate the effectiveness of surrogate modeling techniques in global sensitivity analysis of high-dimensional problems [166]. Equation (30) describes the ellipsoid function [166].

$$g_4 = \pm \sum_{i=1}^{20} i x_i^2 \quad (30)$$

where the input random variables $x_i = \{x_1, \dots, x_{20}\}$ have uniform distributions between -5 and 5. As seen in equation (30), the global sensitivity of x_j is greater than x_i for $1 \leq i \leq j \leq 20$ [166]. Accordingly, the ellipsoid function is an appropriate benchmark to utilize in global sensitivity analysis.

5.5. Griewank function

The *Griewank function* has been employed in the literature as a high-dimensional benchmark for testing surrogate models aimed at complex systems design, reliability-based design, and uncertainty analyses [68]. Finding the solution to this function requires a challenging endeavor due to its large number of local minima, with the number of local minima increasing exponentially with the dimension d [173]. Equation (31) defines the Griewank function [68,173].

$$g_5 = \sum_{i=1}^d \frac{x_i^2}{4000} - \prod_{i=1}^d \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1 \quad (31)$$

The random input variables $x_i = \{x_1, \dots, x_d\}$ follow uniform distributions between -600 and 600 [68,173]. The dimension d of the Griewank function can be adjusted. For instance, the dimensions $d = \{20, 40, 60, 80, 100\}$ are considered in one of the studies while evaluating surrogate models for high-dimensional inputs [68]. It is important to note that the references also suggest more advanced formulations of the Griewank function [173], which can be applied as needed.

5.6. Rackwitz function

The *Rackwitz function* has been proposed for studying the combined performance of surrogate modeling and sampling techniques in reliability analysis of high-dimensional complex engineering systems with a wide range of reliability levels [51]. Equation (32) expresses the dependency of this function on the input variables [11,51,66,174].

$$g_6 = \left(d + c\sigma\sqrt{d} \right) - \sum_{i=1}^d x_i \quad (32)$$

where $x_i = \{x_1, \dots, x_d\}$ are the independent random input variables. The dimension d in this function is changeable.

A study on surrogate modeling for reliability analysis presumes that the random variables are normal where ions with a mean of 1 and a standard deviation of 0.1. The study selects dimensions $d = 2 - 25$. A keynote is that the term $d + c\sigma\sqrt{d}$ maintains reliability at the same level for all chosen values of dimension [51].

In two further studies, the random variables adhere to lognormal distributions with a mean of 1 and a standard deviation of 0.2. In these investigations, the parameter $c = 3$ and the dimensions $d = \{100, 250, 1000\}$ as well as $d = \{40, 100, 250\}$ are assigned to evaluate the effects of the input dimension on the effectiveness of the surrogate modeling in high-dimensional spaces [11,66]. In another study, the random variables of this function have a lognormal distribution with a mean of 1 and a standard deviation of 0.22. This research selects the value of $c = 3$ and the dimension $d = 40$ [105].

5.7. Rosenbrock function

The developed form of the *Rosenbrock function* also called the full Rosenbrock function, has been leveraged to study the performance of surrogate models in the global sensitivity analysis of high-dimensional problems [84]. This function is defined by Equation (33) [175].

$$g_7 = \sum_{i=1}^{d-1} (x_i - 1)^4 + \sum_{i=2}^d \sqrt{i} (x_i - x_{i-1}^2)^2 \quad (33)$$

where $x_i = \{x_1, \dots, x_d\}$ are input random variables with uniform distributions between -1 and 1. The dimension d is adjustable and can be changed based on the researcher's demands [175]. The studies review indicates that some researchers consider $d = 30$ while evaluating surrogate models for engineering problems [84].

In the Rosenbrock function, the variance of each input variable x_i increases significantly with the rise in dimension d , as each new random variable is directly dependent on the square of the previous variable [176]. Therefore, this function is considered an appropriate benchmark for sensitivity analysis. This characteristic also struggles with challenges in Markov Chain-Monte Carlo sampling [176].

5.8. Dixon-Price function

The *Dixon-Price function* has been utilized to assess the success of surrogate modeling for global sensitivity analysis in high-dimensional spaces [84,166]. Equation (34) presents the Dixon-Price function [84,166].

$$g_8 = (x_1 - 1)^2 + \sum_{i=2}^d i(2x_i^2 - x_{i-1})^2 \quad (34)$$

The input random variables $x_i = \{x_1, \dots, x_d\}$ are uniformly distributed between -10 and 10. The dimension d is mutable, which has been assigned the value $d = \{30, 50\}$ in studies investigating the effectiveness of surrogate models for high-dimensional industrial applications [84,166].

Due to the dependence of each new random variable on the previous one in the Dixon-Price function, the variance of each input variable x_i increases as the dimension d grows. Consequently, this function can be valuable for studying the influences of changes in the input variables on the output predictions. Additionally, the literature has employed this function as a benchmark for determining sensitivity indices in global sensitivity analysis [177].

5.9. d -dimensional paraboloid

The d -dimensional paraboloid has been deployed to examine the combined utility of surrogate modeling and sampling techniques in the reliability analysis of high-dimensional practical problems [104,165]. Equation (35) represents the d -dimensional paraboloid [104,165].

$$g_9 = a \sum_{i=2}^d x_i^2 - x_1 - b \quad (35)$$

The input random variables $x_i = \{x_1, \dots, x_d\}$ have independent standard normal distributions with a mean of 0 and a standard

deviation of 1. The dimension d can be adapted according to the researcher's demands. The constant values $a = 0.1$ and $b = 4.5$ for the paraboloid with dimension $d = 100$ are defined [104,165].

5.10. High-dimensional nonlinear function

The performance of surrogate models can be estimated using a *high-dimensional nonlinear function* [111,168]. This function, specifically designed for sensitivity analysis, is defined according to Equation (36) [111].

$$g_{10} = 3 - \frac{5}{d} \sum_{i=1}^d ix_i + \frac{1}{d} \sum_{i=1}^d ix_i^3 + \frac{1}{3d} \sum_{i=1}^d i \ln(x_i^2 + x_i^4) + x_1x_2^2 + x_2x_4 - x_3x_5 + x_{51} + x_{50}x_{54}^2 \quad (36)$$

where $x_i = \{x_1, \dots, x_d\}$ are input random variables with uniform distributions. The dimension d of this function is adjustable, but it is greater than or equal to 55. The input variable distribution for $d = 100$ follows Equation (37) [168].

$$\begin{aligned} x_i &\sim u(1, 2) \quad i = 1, \dots, 100, \quad i \neq 20 \\ x_{20} &\sim u(1, 3) \end{aligned} \quad (37)$$

Since all input variables are uniformly distributed, the function's sensitivity pattern generally has a nonlinear increasing trend as the dimensionality increases. In this case, distinct peaks occur for the variables x_2 , x_{20} , x_{51} , and x_{54} . Additionally, four interaction-related two-term peaks between x_1x_2 , x_2x_4 , x_3x_5 , and $x_{50}x_{54}$ are predicted. It expects two pairs of interaction terms (i.e. x_1x_2 , $x_{50}x_{54}$ and x_2x_4 , x_3x_5) to have identical sensitivity indices [168].

5.11. Parabolic function

Researchers leverage a *parabolic function* to assess the performance of surrogate modeling techniques in reliability analysis because of the large number of input variables [11]. Equation (38) represents this function [11].

$$g_{11} = 20 - 3 \sum_{i=1}^5 x_i(1 + 0.1x_i) - \sum_{i=6}^{100} c_i x_i \quad (38)$$

The input random variables $x_i = \{x_1, \dots, x_{100}\}$ have independent standard normal distributions with a mean of 0 and a standard deviation of 1. The coefficients of the linear term for $i = 6, 7, \dots, 100$ in Equation (38) are $c_i = 0.08$ [11].

5.12. Linear function

A *linear function* has been used in the literature to examine the effectiveness of surrogate modeling and sampling techniques in reliability analysis [38,61] and particularly to estimate the probability of rare events in high-dimensional problems [62]. Equation (39) shows this function [38,61,62].

$$g_{12} = -\frac{1}{\sqrt{d}} \sum_{i=1}^d x_i + \beta \quad (39)$$

This function is a linear combination of independent variables $x_i = \{x_1, \dots, x_d\}$ with standard normal distributions. The dimension d is adjustable and can be changed. Studies indicate that dimensions $d = \{100, 1000\}$ have been considered to show high dimensionality.

The failure probability for this specific type of limit-state function depends on the value of β . Reliability analyses researches apply the value of $\beta = 3.5$ to correspond to the failure probability of $\Phi(-\beta) = 2.33 \times 10^{-4}$ [38,61,62]. This function is characterized by the independence of the failure probability on dimension d [61].

5.13. Corner-peak function

The *corner-peak function* has been used to evaluate surrogate modeling performance for quantifying uncertainties in high-dimensional spaces. Equation (40) introduces this function [32].

$$g_{13} = \left(1 + \sum_{i=1}^d c_i x_i\right)^{-(d+1)} \quad (40)$$

where $x_i = \{x_1, \dots, x_d\}$ are input random variables with uniform distributions between 0 and 1. Notably, the dimension d is changeable.

By adjusting the coefficients c_i , it is possible to control the significance and compressibility of the space dimension. In a study, the performance of the suggested surrogate modeling technique for various dimensionalities $d = \{100, 1000, 10000\}$ is evaluated by selecting the specific coefficient values c_i from Equation (41) [32].

$$c_i^{(1)} = \frac{i - \frac{1}{2}}{d}, \quad c_i^{(2)} = \frac{1}{i^2}, \quad c_i^{(3)} = \exp\left(\frac{i \log(10^{-8})}{d}\right) \quad (41)$$

6. Conclusions

Uncertainty quantification is an essential process for supporting crucial technical decision-making. Additionally, surrogate models can effectively alleviate the computational demands of quantifying response uncertainties in real-world problems. However, as the number of input variables and the complexity of problems increase, constructing these surrogates becomes challenging. This complexity can make the propagation of uncertainty from inputs to outputs difficult or even infeasible in practice. By exploring the literature, these challenges can be categorized into the curse of dimensionality, the need for comprehensive datasets, model accuracy, robustness and generalization, interpretability, adherence to physical constraints, preservation of invariant properties, and the establishment of uncertainty criteria.

Given the expanding efforts in recent years to develop efficient and accurate surrogate modeling methods, a comprehensive overview of pioneering solutions is crucial. Such an overview can help address upcoming challenges in solving various practical problems and revitalize research efforts in quantifying uncertainties in high-dimensional systems.

The following stages summarize the selection of surrogate model development methods for quantifying uncertainties in high-dimensional input problems, based on a general overview of the algorithms utilized in the literature.

- **Selection of the initial surrogate model:** In the first step, the selection of the initial surrogate model is based on the characteristics of the problem, including the size of the datasets, the number of input-output variables, and the complexity of the problem. At this stage, the capabilities of various methods for constructing suitable surrogate models, in relation to the number of datasets and the nature of the variables (structured or unstructured), must be taken into consideration.
- **Adequate understanding of upcoming challenges:** In addition to the challenges posed by high-dimensional spaces, this step necessitates identifying the limitations of the selected surrogate modeling method as well as the inherent challenges of the data. The properties of the data can present challenges, such as ensuring the accuracy and stability of the model when dealing with time-dependent excitations in dynamic systems and data-driven processes.
- **Integration of algorithms for development of surrogate models:** Finally, the techniques for developing the surrogate model can be selected based on the challenges identified in the previous step. A literature review can play a significant role in choosing optimal tools and solution strategies. It is crucial to emphasize how the chosen techniques for developing solution algorithms are integrated to overcome the challenges associated with surrogate modeling in accordance with the characteristics of the problem.

Focusing on core ideas from the literature, this extensive study is addressed the curse of dimensionality and identify areas requiring further research. The discussed solutions encompass dimensionality reduction, multi-fidelity surrogate models, and advanced sampling-based surrogate models. In addition to these general solutions, the study introduces strategies to tackle other challenges in the development of surrogate models. It emphasizes the importance of incorporating machine learning techniques to handle the increasing complexity of engineering problem-solving. This review represents a pivotal step toward future developments in this open research area. In the following, some research gaps and areas for future directions are outlined.

High dimensionality poses challenges when utilizing optimization algorithms as a main component for determining parameters of surrogate models. It is expected that further efforts in both theoretical and empirical fields to address these challenges could lead to promising results. In this context, future research could benefit greatly from the development of active subspace approaches that enhance the efficiency of exploring high-dimensional spaces for uncertainty quantification problems. Additionally, developing methods for generating synthetic data to address sparse datasets is a promising area for investigating uncertainty analysis in high-dimensional input problems. Furthermore, recent progress in physics-informed surrogate models such as PINNs, which incorporate prior physical knowledge into machine learning models, offers new opportunities to tackle complex uncertainty quantification problems that once appeared computationally infeasible. This research area could particularly influence the prediction of missing values or the simulation of additional data points. Despite the scarcity of studies in this specific area, the extensive potential of these tools, particularly in managing large and variable datasets, along with the creative application of machine learning capabilities and insights from physical knowledge, could open new avenues in various approaches to modeling uncertainties.

At the end of this review paper, the improved performance of the proposed approaches in benchmark problems and engineering applications are evaluated. Furthermore, it demonstrates that a specific set of high-dimensional problems can validate surrogate models for quantifying uncertainties by providing several benchmark functions with a diverse range of dimensions.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Zeynab Azarhoosh: Writing – original draft, Methodology, Investigation, Conceptualization. **Majid Ilchi Ghazaan:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

On behalf of all authors, the corresponding author states that there is no competing financial interests or personal relationships that could have appeared to influence the study reported in this paper.

References

- [1] R. Ghanem, D. Higdon, H. Owhadi, Handbook of uncertainty quantification, Springer, 2017.
- [2] C. Wang, X. Qiang, M. Xu, T. Wu, Recent advances in surrogate modeling methods for uncertainty quantification and propagation, *Symmetry* 14 (2022) 1219.
- [3] C. Wang, L. Hong, X. Qiang, M. Xu, Novel numerical method for uncertainty analysis of coupled vibro-acoustic problem considering thermal stress, *Computer Methods in Applied Mechanics and Engineering* 420 (2024) 116727.
- [4] H.N. Najm, R.G. Ghanem, M.S. Eldred, Design Optimization under Uncertainty in Large Scale Computational Models, Sandia National Lab.(SNL-CA), Livermore, CA United States, 2018. Sandia National Laboratories.
- [5] K. Kontolati, D. Loukrezis, K.R. dos Santos, D.G. Giovanis, M.D. Shields, Manifold learning-based polynomial chaos expansions for high-dimensional surrogate models, *International Journal for Uncertainty Quantification* 12 (2022).
- [6] C. Lataniotis, Data-driven uncertainty quantification for high-dimensional engineering problems, ETH Zurich, 2019.
- [7] M. Vohra, P. Nath, S. Mahadevan, Y.-T.T. Lee, Fast surrogate modeling using dimensionality reduction in model inputs and field output: Application to additive manufacturing, *Reliability engineering & system safety* 201 (2020) 106986.
- [8] T. Zhou, Y. Peng, Efficient reliability analysis based on deep learning-enhanced surrogate modelling and probability density evolution method, *Mechanical Systems and Signal Processing* 162 (2022) 108064.
- [9] J.N. Fuhg, A. Fau, U. Nackenhorst, State-of-the-art and comparative review of adaptive sampling methods for kriging, *Archives of Computational Methods in Engineering* 28 (2021) 2689–2747.
- [10] V. Hombal, S. Mahadevan, Surrogate modeling of 3D crack growth, *International journal of fatigue* 47 (2013) 90–99.
- [11] J. Yin, Efficient Uncertainty quantification with high dimensionality, Purdue University Graduate School, 2022.
- [12] F.Y. Kuo, I.H. Sloan, Lifting the curse of dimensionality, *Notices of the AMS* 52 (2005) 1320–1328.
- [13] J.B. Nagel, J. Rieckermann, B. Sudret, Uncertainty quantification in urban drainage simulation: fast surrogates for sensitivity analysis and model calibration, *arXiv preprint arXiv:1709.03283*, (2017) 1–37.
- [14] N. Fajraoui, S. Marelli, B. Sudret, Sequential design of experiment for sparse polynomial chaos expansions, *SIAM/ASA Journal on Uncertainty Quantification* 5 (2017) 1061–1085.
- [15] P. Chen, A. Quarteroni, A new algorithm for high-dimensional uncertainty quantification based on dimension-adaptive sparse grid approximation and reduced basis methods, *Journal of Computational Physics* 298 (2015) 176–193.
- [16] B. Kapsuzoglu, Y. Guo, S. Mahadevan, S. Matsumoto, M. Yoshitomo, S. Taba, D. Watanabe, Dimension reduction for efficient surrogate modeling in high-dimensional applications, in: *AIAA SCITECH 2022 Forum*, 2022, p. 1440.
- [17] Y. Guo, S. Mahadevan, S. Matsumoto, S. Taba, D. Watanabe, Surrogate modeling with high-dimensional input and output, in: *AIAA Scitech 2021 forum*, 2021, p. 0182.
- [18] B. Kapsuzoglu, S. Mahadevan, S. Matsumoto, Y. Miyagi, D. Watanabe, Adaptive surrogate modeling for high-dimensional spatio-temporal output, *Structural and Multidisciplinary Optimization* 65 (2022) 300.
- [19] B. Kapsuzoglu, S. Mahadevan, Information fusion and machine learning for sensitivity analysis using physics knowledge and experimental data, *Reliability Engineering & System Safety* 214 (2021) 107712.
- [20] Y. Guo, S. Mahadevan, S. Matsumoto, S. Taba, D. Watanabe, Investigation of Surrogate Modeling Options with High-Dimensional Input and Output, *AIAA Journal* 61 (2023) 1334–1348.
- [21] C. Zhou, H. Zhang, Q. Chang, X. Song, C. Li, An adaptive ensemble of surrogate models based on hybrid measure for reliability analysis, *Structural and Multidisciplinary Optimization* 65 (2022) 16.
- [22] A. Pires, M. Moustapha, S. Marelli, B. Sudret, Surrogate-based reliability analysis for noisy models, in: *14th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP14)*, Trinity College Dublin, 2023.
- [23] E. Torre, S. Marelli, P. Embrechts, B. Sudret, Data-driven polynomial chaos expansion for machine learning regression, *Journal of Computational Physics* 388 (2019) 601–623.
- [24] M.R. Karim, M. Shajalal, A. Graß, T. Döhmen, S.A. Chala, A. Boden, C. Beecks, S. Decker, Interpreting black-box machine learning models for high dimensional datasets, in: *2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, 2023, pp. 1–10.
- [25] J. Chan-Lau, M.J.A. Chan-Lau, R. Hu, M. Ivanyina, R. Qu, C. Zhong, Surrogate data models: interpreting large-scale machine learning crisis prediction models, *International Monetary Fund*, 2023.
- [26] Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, P. Perdikaris, Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data, *Journal of Computational Physics* 394 (2019) 56–81.
- [27] H. Zheng, Z. Huang, G. Lin, A physics-constrained neural network for multiphase flows, *Physics of Fluids* 34 (2022).
- [28] C. Qu, Q. Yu, B.L. Van Hoozen Jr, J.M. Bowman, R.A. Vargas-Hernández, Assessing Gaussian process regression and permutationally invariant polynomial approaches to represent high-dimensional potential energy surfaces, *Journal of Chemical Theory and Computation* 14 (2018) 3381–3396.
- [29] A. Doostan, A. Validi, G. Iaccarino, Non-intrusive low-rank separated approximation of high-dimensional stochastic models, *Computer Methods in Applied Mechanics and Engineering* 263 (2013) 42–55.
- [30] M. Kubicek, E. Minisci, M. Cisternino, High dimensional sensitivity analysis using surrogate modeling and high dimensional model representation, *International Journal for Uncertainty Quantification* 5 (2015).
- [31] P. Kersaudy, B. Sudret, N. Varsier, O. Picon, J. Wiart, A new surrogate modeling technique combining Kriging and polynomial chaos expansions—Application to uncertainty analysis in computational dosimetry, *Journal of Computational Physics* 286 (2015) 103–117.
- [32] J.D. Jakeman, M.S. Eldred, K. Sargsyan, Enhancing ℓ_1 -minimization estimates of polynomial chaos expansions using basis selection, *Journal of Computational Physics* 289 (2015) 18–34.
- [33] K. Konakli, B. Sudret, Global sensitivity analysis using low-rank tensor approximations, *Reliability Engineering & System Safety* 156 (2016) 64–83.
- [34] K. Konakli, B. Sudret, Uncertainty quantification in high-dimensional spaces with low-rank tensor approximations, in: *1st International Conference on Uncertainty Quantification in Computational Sciences and Engineering*, 2015.
- [35] K. Konakli, B. Sudret, Reliability analysis of high-dimensional models using low-rank tensor approximations, *Probabilistic Engineering Mechanics* 46 (2016) 18–36.
- [36] G. Deman, K. Konakli, B. Sudret, J. Kerrou, P. Perrochet, H. Benabderrahmane, Using sparse polynomial chaos expansions for the global sensitivity analysis of groundwater lifetime expectancy in a multi-layered hydrogeological model, *Reliability Engineering & System Safety* 147 (2016) 156–169.
- [37] C. Liang, S. Mahadevan, Stochastic multidisciplinary analysis with high-dimensional coupling, *AIAA Journal* 54 (2016) 1209–1219.
- [38] I. Papaioannou, C. Papadimitriou, D. Straub, Sequential importance sampling for structural reliability analysis, *Structural safety* 62 (2016) 66–75.
- [39] Y. Shin, D. Xiu, Nonadaptive quasi-optimal points selection for least squares linear regression, *SIAM Journal on Scientific Computing* 38 (2016) A385–A411.
- [40] R. Tripathy, I. Bilonis, M. Gonzalez, Gaussian processes with built-in dimensionality reduction: Applications to high-dimensional uncertainty propagation, *Journal of Computational Physics* 321 (2016) 191–223.

- [41] R. Schöbi, B. Sudret, S. Marelli, Rare event estimation using polynomial-chaos kriging, *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 3 (2017) D4016002.
- [42] P. Perdikaris, D. Venturi, G.E. Karniadakis, Multifidelity information fusion algorithms for high-dimensional systems and massive data sets, *SIAM Journal on Scientific Computing* 38 (2016) B521–B538.
- [43] A. Rushdi, L.P. Swiler, E.T. Phipps, M. D'Elia, M.S. Ebeida, VPS: Voronoi piecewise surrogate models for high-dimensional data fitting, *International Journal for Uncertainty Quantification* 7 (2017).
- [44] N. Fajraoui, S. Marelli, B. Sudret, On optimal experimental designs for sparse polynomial chaos expansions, *arXiv preprint arXiv:1703.05312*, (2017).
- [45] S. Marelli, B. Sudret, Adaptive Designs and sparse polynomial chaos expansions for structural reliability analysis, in: *Proceedings of the 12th International Conference on Structural Safety and Reliability (ICOSSAR 2017)*, TU Verlag, 2017, pp. 227–236.
- [46] C. Soize, R. Ghanem, Polynomial chaos representation of databases on manifolds, *Journal of Computational Physics* 335 (2017) 201–221.
- [47] A. Hadidi, B.F. Azar, A. Rafiee, Efficient response surface method for high-dimensional structural reliability analysis, *Structural Safety* 68 (2017) 15–27.
- [48] C.V. Mai, Polynomial chaos expansions for uncertain dynamical systems, Applications in earthquake engineering, *IBK Bericht* 502 (2018).
- [49] N.-C. Xiao, M.J. Zuo, C. Zhou, A new adaptive sequential sampling method to construct surrogate models for efficient reliability analysis, *Reliability Engineering & System Safety* 169 (2018) 330–338.
- [50] M. Hadigol, A. Doostan, Least squares polynomial chaos expansion: A review of sampling strategies, *Computer Methods in Applied Mechanics and Engineering* 332 (2018) 382–407.
- [51] M.K. Sadoughi, C. Hu, C.A. MacKenzie, A.T. Eshghi, S. Lee, Sequential exploration-exploitation with dynamic trade-off for efficient reliability analysis of complex engineered systems, *Structural and Multidisciplinary Optimization* 57 (2018) 235–250.
- [52] C.A. Thimmisetty, R.G. Ghanem, J.A. White, X. Chen, High-dimensional intrinsic interpolation using Gaussian process regression and diffusion maps, *Mathematical Geosciences* 50 (2018) 77–96.
- [53] R.K. Tripathy, I. Bilonis, UQ Deep, Learning deep neural network surrogate models for high dimensional uncertainty quantification, *Journal of computational physics* 375 (2018) 565–588.
- [54] N. Bassamzadeh, R. Ghanem, Probabilistic data-driven prediction of wellbore signatures in high-dimensional data using Bayesian networks, *SPE Journal* 23 (2018) 1090–1104.
- [55] S. Marelli, B. Sudret, An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis, *Structural Safety* 75 (2018) 67–74.
- [56] Y. Yang, P. Perdikaris, Conditional deep surrogate models for stochastic, high-dimensional, and multi-fidelity systems, *Computational Mechanics* 64 (2019) 417–434.
- [57] M. Li, G. Jia, R.Q. Wang, Surrogate modeling for sensitivity analysis of models with high-dimensional outputs, in: *13 th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP13*, Seoul, South Korea, 2019, pp. 26–30.
- [58] M. Li, R.-Q. Wang, G. Jia, Efficient dimension reduction and surrogate-based sensitivity analysis for expensive models with high-dimensional outputs, *Reliability Engineering & System Safety* 195 (2020) 106725.
- [59] R. Sheikholeslami, S. Razavi, H.V. Gupta, W. Becker, A. Haghnegahdar, Global sensitivity analysis for high-dimensional problems: How to objectively group factors and measure robustness and convergence while reducing computational cost, *Environmental modelling & software* 111 (2019) 282–299.
- [60] K. Cheng, Z. Lu, Structural reliability analysis based on ensemble learning of surrogate models, *Structural Safety* 83 (2020) 101905.
- [61] Y. Xu, P. Wang, Sequential sampling based reliability analysis for high dimensional rare events with confidence intervals, in: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, 2020. V11BT11A039.
- [62] Y. Xu, P. Wang, Rare event estimation of high dimensional problems with confidence intervals, in: *IIE Annual Conference. Proceedings*, Institute of Industrial and Systems Engineers (IISE), 2020, pp. 31A–36A.
- [63] M. Ehre, I. Papaioannou, D. Straub, Global sensitivity analysis in high dimensions with PLS-PCE, *Reliability Engineering & System Safety* 198 (2020) 106861.
- [64] Y. Zhou, Z. Lu, J. Hu, Y. Hu, Surrogate modeling of high-dimensional problems via data-driven polynomial chaos expansions and sparse partial least square, *Computer Methods in Applied Mechanics and Engineering* 364 (2020) 112906.
- [65] E. Savin, J.-L. Hantrais-Gervois, Sparse polynomial surrogates for non-intrusive, high-dimensional uncertainty quantification of aeroelastic computations, *Probabilistic Engineering Mechanics* 59 (2020) 103027.
- [66] Y. Zhou, Z. Lu, An enhanced Kriging surrogate modeling technique for high-dimensional problems, *Mechanical Systems and Signal Processing* 140 (2020) 106687.
- [67] F. Uribe, I. Papaioannou, Y.M. Marzouk, D. Straub, Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation, *SIAM/ASA Journal on Uncertainty Quantification* 9 (2021) 818–847.
- [68] Y. Xu, P. Wang, An enhanced squared exponential kernel with Manhattan similarity measure for high dimensional Gaussian process models, in: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, 2021. V03BT03A025.
- [69] Y. Peng, T. Zhou, J. Li, Surrogate modeling immersed probability density evolution method for structural reliability analysis in high dimensions, *Mechanical Systems and Signal Processing* 152 (2021) 107366.
- [70] N. Wycoff, M. Binois, R.B. Gramacy, Sensitivity Prewarping for Local Surrogate Modeling, *Technometrics* 64 (2022) 535–547.
- [71] K. Kontolati, D. Loukrezis, D.G. Giovanis, L. Vandanapu, M.D. Shields, A survey of unsupervised learning methods for high-dimensional uncertainty quantification in black-box-type problems, *Journal of Computational Physics* 464 (2022) 111313.
- [72] P.F. Shustin, S. Ubaru, V. Kalantzis, L. Hoesch, H. Avron, PCENet: High dimensional surrogate modeling for learning uncertainty, *arXiv preprint arXiv:2202.05063*, (2022).
- [73] L. Hong, H. Li, J. Fu, A novel surrogate-model based active learning method for structural reliability analysis, *Computer Methods in Applied Mechanics and Engineering* 394 (2022) 114835.
- [74] Y. Liu, L. Li, S. Zhao, A global surrogate model for high-dimensional structural systems based on partial least squares and Kriging, *Mechanical Systems and Signal Processing* 164 (2022) 108246.
- [75] N. Navaneeth, S. Chakraborty, Surrogate assisted active subspace and active subspace assisted surrogate—A new paradigm for high dimensional structural reliability analysis, *Computer Methods in Applied Mechanics and Engineering* 389 (2022) 114374.
- [76] Y. Xu, P. Wang, Adaptive surrogate models for uncertainty quantification with partially observed information, in: *AIAA SCITECH 2022 Forum*, 2022, p. 1439.
- [77] D. Bigoni, Y. Marzouk, C. Prieur, O. Zahm, Nonlinear dimension reduction for surrogate modeling using gradient information, *Information and Inference: A Journal of the IMA* 11 (2022) 1597–1639.
- [78] S.L. Dhulipala, M.D. Shields, B.W. Spencer, C. Bolisetti, A.E. Slaughter, V.M. Labouré, P. Chakraborty, Active learning with multifidelity modeling for efficient rare event simulation, *Journal of Computational Physics* 468 (2022) 111506.
- [79] S. Yildiz, H. Pehlivan Solak, M. Nikbay, Multi-Fidelity Low-Rank Approximations for Uncertainty Quantification of a Supersonic Aircraft Design, *Algorithms* 15 (2022) 250.
- [80] P. Conti, M. Guo, A. Manzoni, J.S. Hesthaven, Multi-fidelity surrogate modeling using long short-term memory networks, *Computer methods in applied mechanics and engineering* 404 (2023) 115811.
- [81] Y. Ji, H. Liu, N.-C. Xiao, H. Zhan, An efficient method for time-dependent reliability problems with high-dimensional outputs based on adaptive dimension reduction strategy and surrogate model, *Engineering Structures* 276 (2023) 115393.
- [82] X. Shang, L. Su, H. Fang, B. Zeng, Z. Zhang, An efficient multi-fidelity Kriging surrogate model-based method for global sensitivity analysis, *Reliability Engineering & System Safety* 229 (2023) 108858.
- [83] R. Sun, B. Pan, Q. Duan, A surrogate modeling method for distributed land surface hydrological models based on deep learning, *Journal of Hydrology* 624 (2023) 129944.

- [84] K. Cheng, R. Zimmermann, Sliced Gradient-Enhanced Kriging for High-Dimensional Function Approximation, *SIAM Journal on Scientific Computing* 45 (2023) A2858–A2885.
- [85] J. Liu, C. Jiang, Surrogate modeling for high dimensional uncertainty propagation via deep kernel polynomial chaos expansion, *Applied Mathematical Modelling* (2023).
- [86] S. Schär, S. Marelli, B. Sudret, Emulating the dynamics of complex systems using autoregressive models on manifolds (mNARX), *Mechanical Systems and Signal Processing* 208 (2024) 110956.
- [87] Y. Guo, P. Nath, S. Mahadevan, P. Witherell, Active learning for adaptive surrogate model improvement in high-dimensional problems, *Structural and Multidisciplinary Optimization* 67 (2024) 122.
- [88] B. Mufti, C. Perron, D.N. Mavris, A Multi-Fidelity Methodology for Reduced Order Models with High-Dimensional Inputs, *arXiv preprint arXiv:2402.17061*, (2024).
- [89] C. Bharti, D. Ghosh, A novel non-intrusive ROM for randomly excited linear dynamical systems with high stochastic dimension using ANN, *Probabilistic Engineering Mechanics* 75 (2024) 103570.
- [90] J. Xian, Z. Wang, A physics and data co-driven surrogate modeling method for high-dimensional rare event simulation, *Journal of Computational Physics* 510 (2024) 113069.
- [91] J. Kim, Z. Wang, J. Song, Adaptive active subspace-based metamodeling for high-dimensional reliability analysis, *Structural Safety* 106 (2024) 102404.
- [92] Z. Li, F. Montomoli, Aleatory uncertainty quantification based on multi-fidelity deep neural networks, *Reliability Engineering & System Safety* 245 (2024) 109975.
- [93] Y. Zhou, X. Gong, X. Zhang, An active-subspace-enhanced support vector regression model for high-dimensional uncertainty quantification, (2024).
- [94] W. He, G. Li, Y. Zeng, Y. Wang, C. Zhong, An adaptive data-driven subspace polynomial dimensional decomposition for high-dimensional uncertainty quantification based on maximum entropy method and sparse Bayesian learning, *Structural Safety* 108 (2024) 102450.
- [95] Z. Song, Z. Liu, H. Zhang, P. Zhu, An improved sufficient dimension reduction-based Kriging modeling method for high-dimensional evaluation-expensive problems, *Computer Methods in Applied Mechanics and Engineering* 418 (2024) 116544.
- [96] L. Shi, Z. Kai, Z. Wang, Convolutional Dimension-Reduction with Knowledge Reasoning for Reliability Approximations of Structures under High-Dimensional Spatial Uncertainties, *Journal of Mechanical Design* (2024) 1–11.
- [97] J. Kim, S.-r. Yi, Z. Wang, Dimensionality reduction can be used as a surrogate model for high-dimensional forward uncertainty quantification, *arXiv preprint arXiv:2402.04582*, (2024).
- [98] B. Wang, N.C. Orndorff, M. Sperry, J.T. Hwang, Extension of graph-accelerated non-intrusive polynomial chaos to high-dimensional uncertainty quantification through the active subspace method, *arXiv preprint arXiv:2405.05556*, (2024).
- [99] A. Zanon, G. Geraci, M. Salvador, K. Menon, A.L. Marsden, D.E. Schiavazzi, Linear and nonlinear dimension reduction strategies for multifidelity uncertainty propagation of nonparametric distributions, in: *AIAA SCITECH 2024 Forum*, 2024, p. 0389.
- [100] X. Gong, Y. Pan, Multifidelity Bayesian Experimental Design to Quantify Rare-Event Statistics, *SIAM/ASA Journal on Uncertainty Quantification* 12 (2024) 101–127.
- [101] D. Loukrezis, E. Diehl, H. De Gerssem, Multivariate sensitivity-adaptive polynomial chaos expansion for high-dimensional surrogate modeling and uncertainty quantification, *arXiv preprint arXiv:2310.09871v3*, (2024).
- [102] G. Catalanotti, Navigating the unknown: Tackling high-dimensional challenges in composite damage modeling with bootstrapping and Bayesian uncertainty quantification, *Composites Science and Technology* 248 (2024) 110462.
- [103] S. Baisthakur, B. Fitzgerald, Physics-Informed Neural Network surrogate model for bypassing Blade Element Momentum theory in wind turbine aerodynamic load estimation, *Renewable Energy* 224 (2024) 120122.
- [104] T.-T. Nguyen, V.-H. Dang, D.-M. Hoang, X.-D. Pham, T.-H. Nguyen, V.-T. Dinh, Robust active learning framework for structural reliability analysis using uncertainty quantification and flexible meta-model, in: *Structures*, Elsevier, 2024 106465.
- [105] H.-P. Wan, J.-R. Gan, Y.-K. Zhu, Z. Meng, SS-MASVM: An advanced technique for assessing failure probability of high-dimensional complex systems using the multi-class adaptive support vector machine, *Computer Methods in Applied Mechanics and Engineering* 418 (2024) 116568.
- [106] M. Moustapha, S. Marelli, B. Sudret, A generalized framework for active learning reliability analysis in UQLab, in: *2nd international workshop on the Computational Challenges in the Reliability Assessment of Engineering Structures*, 2020, p. 2020.
- [107] A. Hirvoas, Development of a data assimilation method for the calibration and continuous update of wind turbines digital twins, *Université Grenoble Alpes [2020-....]*, 2021.
- [108] N.N. Narisetty, Bayesian model selection for high-dimensional data, in: *Handbook of statistics*, Elsevier, 2020, pp. 207–248.
- [109] C.K.J. Hou, K. Behdian, Dimensionality reduction in surrogate modeling: A review of combined methods, *Data Science and Engineering* 7 (2022) 402–427.
- [110] T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Springer, 2017.
- [111] N. Lüthen, S. Marelli, B. Sudret, Sparse polynomial chaos expansions: Literature survey and benchmark, *SIAM/ASA Journal on Uncertainty Quantification* 9 (2021) 593–649.
- [112] H. Rabitz, Ö.F. Aliş, General foundations of high-dimensional model representations, *Journal of Mathematical Chemistry* 25 (1999) 197–233.
- [113] X. Qiang, C. Wang, H. Fan, Hybrid interval model for uncertainty analysis of imprecise or conflicting information, *Applied Mathematical Modelling* 129 (2024) 837–856.
- [114] M. Moustapha, J.-M. Bourinet, B. Guillaume, B. Sudret, Comparative study of Kriging and support vector regression for structural engineering applications, *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 4 (2018) 04018005.
- [115] T. Chatterjee, K.K. Bera, A. Banerjee, Machine learning enabled quantification of stochastic active metadamping in acoustic metamaterials, *Journal of Sound and Vibration* 567 (2023) 117938.
- [116] Z. Liu, D. Lesselier, B. Sudret, J. Wiart, Surrogate modeling based on resampled polynomial chaos expansions, *Reliability Engineering & System Safety* 202 (2020) 107008.
- [117] N. Fajraoui, M. Fahs, A. Younes, B. Sudret, Analyzing natural convection in porous enclosure with polynomial chaos expansions: Effect of thermal dispersion, anisotropic permeability and heterogeneity, *International Journal of Heat and Mass Transfer* 115 (2017) 205–224.
- [118] C. Soize, R. Ghanem, Physical systems with random uncertainties: chaos representations with arbitrary probability measure, *SIAM Journal on Scientific Computing* 26 (2004) 395–410.
- [119] R.G. Ghanem, P.D. Spanos, *Stochastic finite elements: a spectral approach*, Courier Corporation, 2003.
- [120] M.K. Deb, I.M. Babuška, J.T. Oden, Solution of stochastic partial differential equations using Galerkin finite element techniques, *Computer Methods in Applied Mechanics and Engineering* 190 (2001) 6359–6372.
- [121] O. Le Maître, O. Knio, H. Najm, R. Ghanem, Uncertainty propagation using Wiener–Haar expansions, *Journal of computational Physics* 197 (2004) 28–57.
- [122] A. Keese, H.G. Matthies, Hierarchical parallelisation for the solution of stochastic finite element equations, *Computers & Structures* 83 (2005) 1033–1047.
- [123] D.M. Ghiocel, R.G. Ghanem, Stochastic finite-element analysis of seismic soil–structure interaction, *Journal of Engineering Mechanics* 128 (2002) 66–77.
- [124] O.P. Le Maître, M.T. Reagan, H.N. Najm, R.G. Ghanem, O.M. Knio, A stochastic projection method for fluid flow: II. Random process, *Journal of computational Physics* 181 (2002) 9–44.
- [125] D. Xiu, G.E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, *SIAM journal on scientific computing* 24 (2002) 619–644.
- [126] D. Xiu, J.S. Hesthaven, High-order collocation methods for differential equations with random inputs, *SIAM Journal on Scientific Computing* 27 (2005) 1118–1139.
- [127] M. Berveiller, B. Sudret, M. Lemaire, Stochastic finite element: a non intrusive approach by regression, *European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique* 15 (2006) 81–92.
- [128] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, *Journal of computational Physics* 230 (2011) 2345–2367.

- [129] A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, R. Tempone, Discrete least squares polynomial approximation with random evaluations— application to parametric and stochastic elliptic PDEs, *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique* 49 (2015) 815–837.
- [130] S. Hosder, R.W. Walters, M. Balch, Efficient sampling for non-intrusive polynomial chaos applications with multiple uncertain input variables, (2007).
- [131] K. Konakli, B. Sudret, Polynomial meta-models with canonical low-rank approximations: Numerical insights and comparison to sparse polynomial chaos expansions, *Journal of Computational Physics* 321 (2016) 1144–1169.
- [132] K. Konakli, B. Sudret, Low-rank tensor approximations for reliability analysis, in: 12th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP12), 2015.
- [133] K. Konakli, C. Mylonas, S. Marelli, B. Sudret, Report UQLab-V1, 2019, pp. 1–108.
- [134] K. Konakli, B. Sudret, Addressing high dimensionality in reliability analysis using low-rank tensor approximations, in: European Safety and Reliability Conference ESREL, 2015, p. 2015.
- [135] E. Kieri, B. Vandereycken, Projection methods for dynamical low-rank approximation of high-dimensional problems, *Computational Methods in Applied Mathematics* 19 (2019) 73–92.
- [136] K. Konakli, B. Sudret, Variance-based sensitivity indices from low-rank tensor approximations with polynomial basis, in: 2nd International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2017), National Technical University of Athens (NTUA), 2017.
- [137] M. Chevreuil, R. Lebrun, A. Nouy, P. Rai, A least-squares method for sparse low rank approximation of multivariate functions, *SIAM/ASA Journal on Uncertainty Quantification* 3 (2015) 897–921.
- [138] K. Konakli, B. Sudret, Low-rank tensor approximations versus polynomial chaos expansions for meta-modeling in high-dimensional spaces, *arXiv preprint arXiv:1511.07492*, (2015).
- [139] Y. Xu, P. Wang, A comparison of numerical optimizers in developing high dimensional surrogate models, in: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, 2019. V02BT03A037.
- [140] C. Lataniotis, S. Marelli, B. Sudret, The Gaussian process modelling module in UQLab, *arXiv preprint arXiv:1709.09382*, (2017).
- [141] C. Lataniotis, S. Marelli, B. Sudret, Report UQLab-V0, 2015, pp. 9–105.
- [142] G. Abbiati, R. Schöbi, B. Sudret, B. Stojadinovic, Structural reliability analysis using deterministic hybrid simulations and adaptive kriging metamodeling, in: Proceedings of the 16th World Conference on Earthquake Engineering (16WCEE), 2017, pp. 9–13.
- [143] I. Abdallah, C. Lataniotis, B. Sudret, Parametric hierarchical kriging for multi-fidelity aero-servo-elastic simulators—Application to extreme loads on wind turbines, *Probabilistic Engineering Mechanics* 55 (2019) 67–77.
- [144] B. Sudret, K. Konakli, C.V. Mai, S. Marelli, J. Nagel, R. Schöbi, Recent developments in surrogate modelling for uncertainty quantification, in: 3rd International Conference on Vulnerability and Risk Analysis and Management (ICVRAM 2018), ETH Zurich, Risk, Safety and Uncertainty Quantification, 2018.
- [145] G. Abbiati, I. Abdallah, S. Marelli, B. Sudret, B. Stojadinovic, Hierarchical Kriging Surrogate of the Seismic Response of a Steel Piping Network Based on Multi-Fidelity Hybrid and Computational Simulators, in: Proceedings of the 7th International Conference on Advances in Experimental Structural Engineering (7AESE), 2017, pp. 6–8.
- [146] A. Dadras Eslamlou, S. Huang, Artificial-neural-network-based surrogate models for structural health monitoring of civil structures: a literature review, *Buildings* 12 (2022) 2067.
- [147] R.G. Cooper, Augmented Neural Network Surrogate Models for Polynomial Chaos Expansions and Reduced Order Modeling, Virginia Tech, 2021.
- [148] J. Mohammed J. Zaki, Wagner Meira, Data Mining and Machine Learning: Fundamental Concepts and Algorithms, Second Edition ed., Cambridge University Press, 2020.
- [149] K. Giannoukou, S. Marelli, B. Sudret, A comprehensive framework for multi-fidelity surrogate modeling with noisy data: a gray-box perspective, *arXiv preprint arXiv:2401.06447*, (2024).
- [150] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational physics* 378 (2019) 686–707.
- [151] P. Zhang, A novel feature selection method based on global sensitivity analysis with application in machine learning-based prediction model, *Applied Soft Computing* 85 (2019) 105859.
- [152] P. Gavallas, G. Stefanou, D. Savvas, C. Mattrand, J.-M. Bourinet, CNN-based prediction of microstructure-derived random property fields of composite materials, *Computer Methods in Applied Mechanics and Engineering* 430 (2024) 117207.
- [153] Z. Cai, R. Li, L. Zhu, Online sufficient dimension reduction through sliced inverse regression, *Journal of Machine Learning Research* 21 (2020) 1–25.
- [154] C. Wang, Z. Song, H. Fan, Novel evidence theory-based reliability analysis of functionally graded plate considering thermal stress behavior, *Aerospace Science and Technology* 146 (2024) 108936.
- [155] B. Beherstorfer, K. Willcox, M. Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization, *Siam Review* 60 (2018) 550–591.
- [156] L. Le Gratiet, J. Garnier, Recursive co-kriging model for design of computer experiments with multiple levels of fidelity, *International Journal for Uncertainty Quantification* 4 (2014).
- [157] M. Berchier, Multi-fidelity surrogate modelling with polynomial chaos expansions, ETH Zurich Msc thesis 400 (2016).
- [158] Y. Guo, P. Nath, S. Mahadevan, Adaptive learning for surrogate models in active subspace for high dimensional problems, *Structural and Multidisciplinary Optimization* 67 (2024) 10–26.
- [159] H. Fan, C. Wang, S. Li, Novel method for reliability optimization design based on rough set theory and hybrid surrogate model, *Computer Methods in Applied Mechanics and Engineering* 429 (2024) 117170.
- [160] I. Arsenyev, F. Duddeck, A. Fischerswöring-Bunk, Global Sensitivity Analysis for Multidisciplinary Studies of Vane Clusters. Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management, 2014, pp. 1554–1563.
- [161] L. Gu, C.J. Wu, A Unified Framework for Uncertainty and Sensitivity Analysis of Computational Models with Many Input Parameters, in: The sixth international conference on advances in system simulation, Citeseer, 2014, pp. 276–280.
- [162] G. Blatman, B. Sudret, Efficient computation of global sensitivity indices using sparse polynomial chaos expansions, *Reliability Engineering & System Safety* 95 (2010) 1216–1229.
- [163] V. Sundar, M.D. Shields, Surrogate-enhanced stochastic search algorithms to identify implicitly defined functions for reliability analysis, *Structural Safety* 62 (2016) 1–11.
- [164] L. Zhao, P. Wang, B. Song, X. Wang, H. Dong, An efficient kriging modeling method for high-dimensional design problems based on maximal information coefficient, *Structural and Multidisciplinary Optimization* 61 (2020) 39–57.
- [165] I. Depina, T.M.H. Le, G. Fenton, G. Eiksund, Reliability analysis with metamodel line sampling, *Structural Safety* 60 (2016) 1–15.
- [166] UQLab, version 2.0, <https://www.uqlab.com/sensitivity-high-dimension>.
- [167] J. Unger, D. Roos, Investigation and benchmarks of algorithms for reliability analysis, *Proceedings Weimarer Optimierungs-und Stochastiktage* 1 (2004).
- [168] K. Hussain, M.N.M. Salleh, S. Cheng, R. Naseem, Common benchmark functions for metaheuristic evaluation: A review, *JOIV: International Journal on Informatics Visualization* 1 (2017) 218–223.
- [169] A. Marrel, B. Iooss, F. Van Dorpe, E. Volkova, An efficient methodology for modeling complex computer codes with Gaussian processes, *Computational Statistics & Data Analysis* 52 (2008) 4731–4744.
- [170] M.D. Morris, Factorial sampling plans for preliminary computational experiments, *Technometrics* 33 (1991) 161–174.
- [171] S. Engelund, R. Rackwitz, A benchmark study on importance sampling techniques in structural reliability, *Structural safety* 12 (1993) 255–276.
- [172] M. Fujita, R. Rackwitz, Updating first-and second-order reliability estimates by importance sampling, *Doboku Gakkai Ronbunshu* 1988 (1988) 53–59.
- [173] M. Locatelli, A note on the Griewank test function, *Journal of global optimization* 25 (2003) 169–174.
- [174] R. Rackwitz, Reliability analysis—a review and some perspectives, *Structural safety* 23 (2001) 365–395.

- [175] Z.-H. Han, Y. Zhang, C.-X. Song, K.-S. Zhang, Weighted gradient-enhanced kriging for high-dimensional surrogate modeling and design optimization, *Aiaa Journal* 55 (2017) 4330–4346.
- [176] F. Pagani, M. Wiegand, S. Nadarajah, An n-dimensional Rosenbrock distribution for MCMC testing, arXiv preprint arXiv:1903.09556, (2019).
- [177] A. Spagnol, R.L. Riche, S.b.D. Veiga, Global sensitivity analysis for optimization with variable selection, *SIAM/ASA Journal on uncertainty quantification* 7 (2019) 417–443.