

Multifidelity conditional value-at-risk estimation by dimensionally decomposed generalized polynomial chaos-Kriging

Dongjin Lee^{a,*}, Boris Kramer^{a,*}

^a*Department of Mechanical and Aerospace Engineering, University of California San Diego, CA, United States*

March 14, 2023

Abstract

We propose novel methods for Conditional Value-at-Risk (CVaR) estimation for nonlinear systems under high-dimensional dependent random inputs. We develop a novel DD-GPCE-Kriging surrogate that merges dimensionally decomposed generalized polynomial chaos expansion and Kriging to accurately approximate nonlinear and nonsmooth random outputs. We use DD-GPCE-Kriging (1) for Monte Carlo simulation (MCS) and (2) within multifidelity importance sampling (MFIS). The MCS-based method samples from DD-GPCE-Kriging, which is efficient and accurate for high-dimensional dependent random inputs, yet introduces bias. Thus, we propose an MFIS-based method where DD-GPCE-Kriging determines the biasing density, from which we draw a few high-fidelity samples to provide an unbiased CVaR estimate. To accelerate the biasing density construction, we compute DD-GPCE-Kriging using a cheap-to-evaluate low-fidelity model. Numerical results for mathematical functions show that the MFIS-based method is more accurate than the MCS-based method when the output is nonsmooth. The scalability of the proposed methods and their applicability to complex engineering problems are demonstrated on a two-dimensional composite laminate with 28 (partly dependent) random inputs and a three-dimensional composite T-joint with 20 (partly dependent) random inputs. In the former, the proposed MFIS-based method achieves 104x speedup compared to standard MCS using the high-fidelity model, while accurately estimating CVaR with 1.15% error.

Keywords: Risk measures, Conditional Value-at-Risk, Dimensionally decomposed generalized polynomial chaos expansion, Kriging, Importance sampling, Multifidelity importance sampling

1. Introduction

Risk measures are indispensable to designing safe and reliable systems in the presence of uncertainties that arise from many sources, such as manufacturing processes, material properties, or operating environments. In risk-averse design optimization [4, 44, 32], risk measures are used to avoid or handle a *risky* output via objective and/or constraint functions. Here, we focus on two representative risk measures (originally employed in finance [35] but nowadays used in engineering fields): the Value-at-Risk (VaR) and the Conditional Value-at-Risk (CVaR). While the VaR aggregates losses (related to costs or target violation in any performance function) with a quantile of the distribution for the losses, the CVaR reflects the mean or average value of losses exceeding VaR. The CVaR has superior quantitative or qualitative benefits over VaR, which has been demonstrated in portfolio management [34, 35, 23] and engineering design [48, 36]. For example, CVaR (as a coherent risk measure) measures the average tail risk and preserves subadditivity and convexity; it thus facilitates optimization [34, 15, 32, 4]. Nevertheless, optimization may require smoothed approximations of CVaR [15, 16]. These benefits make CVaR appealing compared to other possibilities of measuring risks, such as reliability or failure

*Corresponding author

Email addresses: dongjin-lee@ucsd.edu (Dongjin Lee), bmramer@ucsd.edu (Boris Kramer)

probability, which is widely used in engineering [33, 4, 44]. Recently, CVaR has been incorporated into design of civil [33], energy [40], naval [36], and aerospace [22, 36, 5] applications.

CVaR and related risk measures are mostly estimated via sampling, e.g., Monte Carlo Simulation (MCS), especially when the quantity of interest does not follow a parametric distribution. Such sampling methods generally require a large number of samples to capture the tail risk of the quantity of interest. When each sample requires an expensive high-fidelity model run, e.g., a finite element analysis (FEA), standard MCS becomes computationally intensive, if not prohibitive. Numerous surrogate methods (e.g., reduced-order modeling (ROM) [12, 49, 11], polynomial chaos expansion (PCE) [2], Kriging or Gaussian process modeling [6], support vector machine [9], and neural networks [38]) have been used to break that computational bottleneck. Recently, [14] presented a surrogate model strategy to conservatively estimate risk measures from given data. Many of these methods, especially the expansion or decomposition methods, are restricted to the assumption that the input random variables are independent. In reality, there is often correlation or dependence among some, or all, of the random inputs. Whether emanating from loads, material properties, or manufacturing variables, neglecting these correlations or dependencies may produce inefficient or risky designs [28, 18].

Some PCE methods, e.g., the generalized PCE (GPCE) [31], dimensionally decomposed GPCE (DD-GPCE) [21] or other PCE variants [27, 13], can handle dependent input random variables directly without a potentially detrimental measure transformation between dependent and independent random variables. A recent GPCE-based surrogate [18] generates a multivariate orthonormal polynomial basis consistent with any non-product-type probability measure of inputs numerically instead of an analytical expression by a Rodrigues-type formula in [31]. DD-GPCE has been demonstrated to mitigate the curse of dimensionality to some extent by reshuffling and pruning the GPCE basis in a dimensionwise manner [21]. In [17], we used DD-GPCE to estimate CVaR for a nonlinear output under high-dimensional and dependent random inputs and showed that DD-GPCE can be efficiently computed via a bifidelity method. However, DD-GPCE or other PCE-related methods generally have limitations when handling highly nonlinear or nonsmooth outputs due to the smoothness of their polynomial bases. The PCE-Kriging method from [37] produces a more accurate and efficient approximation of highly nonlinear outputs compared to PCE. In PCE-Kriging, the PCE part approximates the global behavior of the output, while the Kriging part stochastically interpolates the local output behavior. However, computing the PCE component suffers from the curse of dimensionality, and constructing the PCE’s basis functions is based on the assumption of independent inputs.

Variance reduction methods (e.g., importance sampling [39, 29] or control variates [3]) exploit a problem-dependent sampling strategy to reduce the required number of expensive high-fidelity model evaluations. The efficiency of an importance-sampling method relies on the choice of importance-sampling density (or biasing density), which is often selected from a parameterized family of distributions. Importance-sampling methods can drastically reduce the number of samples required to estimate risk [29] or reliability [41]. To speed up the importance sampling process, the multifidelity importance sampling (MFIS) method from [29] suggests using a surrogate model for learning the biasing density. In the context of MFIS for risk measure estimation, the authors in [12] used ROMs to construct a biasing density that samples in the *risk region* of the parameter space, where input samples produce quantities of interest that are deemed “risky”. The risk region is referred to as the ϵ -risk region when it includes ROM error bounds. The work [12] shows that MFIS can estimate CVaR efficiently and without bias, even when the ROM has significant error compared to the high-fidelity model. However, that work does not consider high-dimensional or dependent random inputs, as the ROM construction becomes computationally intensive as the number of inputs increases.

This paper presents novel computational methods for CVaR estimation for nonlinear systems under high-dimensional and dependent inputs. First, we propose a novel *DD-GPCE-Kriging* surrogate model, a fusion of DD-GPCE [18] and Kriging to approximate a highly nonlinear and nonsmooth random output. Second, we leverage this new surrogate model in sampling-based CVaR estimation via (1) Monte Carlo sampling and (2) multifidelity importance sampling. The proposed MCS-based method replaces the high-fidelity model (expensive to evaluate) with DD-GPCE-Kriging (cheap to

evaluate) and is shown to be accurate in the presence of high-dimensional and dependent random inputs. However, the proposed method may produce a biased CVaR estimate as it relies on the approximation quality of the DD-GPCE-Kriging. Thus, we also suggest leveraging MFIS in connection with DD-GPCE-Kriging to determine the biasing density efficiently. The high-fidelity model is used for the importance-sampling-based CVaR estimate. For the proposed MFIS-based method, there are two major contributions. First, we suggest using the confidence interval (CI) or the probabilistic error bound of DD-GPCE-Kriging to define the ϵ -risk region for the biasing density. Second, we compute DD-GPCE-Kriging by using computationally cheap low-fidelity output samples to further speed up the MFIS process. We investigate how much fidelity (and/or correlation with the high-fidelity model) is needed to construct a good biasing density. We demonstrate the performance of the two proposed CVaR estimation methods on lightweight composite structures with high stiffness and preferable heat resistance. It is essential to measure risk for these composites (used in high-performance air- and spacecraft) to produce risk-averse designs that can perform missions under harsh operating environments, e.g., high temperature and/or pressure. This work differs from our prior work [17] as follows. First, we introduce the novel DD-GPCE-Kriging surrogate to improve the accuracy of the DD-GPCE for nonlinear outputs under dependent random inputs. Second, we suggest the MFIS-based method in connection with DD-GPCE-Kriging, which produces unbiased estimates and achieves higher efficiency than the MCS-based method in computing CVaR estimates.

The paper is organized as follows. Section 2 covers the theoretical background, including the definition of the input and output random variables and introduces CVaR. We also summarize the DD-GPCE and necessary material on (multifidelity) importance sampling. Section 3 presents the new DD-GPCE-Kriging surrogate model, and uses it for MCS-based CVaR estimation. Section 4 proposes an MFIS-based method leveraging DD-GPCE-Kriging for efficient CVaR computation. Numerical results are reported in Section 5. In Section 6, we offer conclusions and an outlook for future work.

2. Theoretical background

This section presents the preliminaries for the CVaR problem setup in Section 2.1, defines input and output random variables in Sections 2.2 and 2.3, respectively, and then gives the CVaR definition in Section 2.4. We briefly summarize DD-GPCE in Section 2.5 and review importance sampling in Section 2.6 and multifidelity importance sampling in Section 2.7.

2.1. Preliminaries

Let \mathbb{N} , \mathbb{N}_0 , \mathbb{R} , and \mathbb{R}_0^+ be the sets of positive integers, non-negative integers, real numbers, and non-negative real numbers, respectively. For a positive integer $N \in \mathbb{N}$, denote by $\mathbb{A}^N \subseteq \mathbb{R}^N$ a subspace of \mathbb{R}^N .

2.2. Input random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an abstract probability space, with a sample space Ω , a σ -algebra \mathcal{F} on Ω , and a probability measure $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$. Consider an N -dimensional random vector $\mathbf{X} := (X_1, \dots, X_N)^\top : \Omega \rightarrow \mathbb{A}^N$, that models the uncertainties in a stochastic problem. We refer to \mathbf{X} as the random input vector or the input random variables. Denote by $F_{\mathbf{X}}(\mathbf{x}) := \mathbb{P}[\cap_{i=1}^N \{X_i \leq x_i\}]$ the joint distribution function of \mathbf{X} , admitting the joint probability density function $f_{\mathbf{X}}(\mathbf{x}) := \partial^N F_{\mathbf{X}}(\mathbf{x}) / \partial x_1 \cdots \partial x_N$. For $(\Omega, \mathcal{F}, \mathbb{P})$, the image probability space is $(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}}(\mathbf{x})d\mathbf{x})$, where \mathbb{A}^N is the image of Ω under the mapping $\mathbf{X} : \Omega \rightarrow \mathbb{A}^N$ and $\mathcal{B}^N := \mathcal{B}(\mathbb{A}^N)$ is the Borel σ -algebra on $\mathbb{A}^N \subset \mathbb{R}^N$.

2.3. Output random variables

Given an input random vector \mathbf{X} with a known probability density function $f_{\mathbf{X}}(\mathbf{x})$ on $\mathbb{A}^N \subseteq \mathbb{R}^N$, denote by $y(\mathbf{X})$ a real-valued, square-integrable transformation on (Ω, \mathcal{F}) . Here, $y : \mathbb{A}^N \rightarrow \mathbb{R}$ describes an output that an application engineer deems relevant for risk assessment. In this work, we assume that y belongs to the weighted L^2 space $\left\{ y : \mathbb{A}^N \rightarrow \mathbb{R} : \int_{\mathbb{A}^N} |y(\mathbf{x})|^2 f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} < \infty \right\}$, which is the Hilbert

space $\left\{ Y = y(\mathbf{X}) : \Omega \rightarrow \mathbb{R} : \int_{\Omega} |y(\mathbf{X}(\omega))|^2 d\mathbb{P}(\omega) < \infty \right\}$ for the abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If there is more than one output variable, then each component is associated with a measurement function $y_i, i = 1, 2, \dots$. The generalization for a multivariate output random vector is straightforward.

2.4. Conditional Value-at-Risk (CVaR)

Given a random input vector \mathbf{X} , consider an output function $y(\mathbf{X}) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Let $\text{CVaR}_{\beta}[y(\mathbf{X})]$ and $\text{VaR}_{\beta}[y(\mathbf{X})]$ be the Conditional Value-at-Risk (CVaR) and the Value-at-Risk (VaR), respectively, of $y(\mathbf{X})$ at a given risk level $\beta \in (0, 1)$. The $\text{VaR}_{\beta}[y(\mathbf{X})]$ is the β -quantile of $y(\mathbf{X})$, i.e.,

$$\text{VaR}_{\beta}[y(\mathbf{X})] = \arg \min_{t \in \mathbb{R}} \{\mathbb{P}[y(\mathbf{X}) \leq t] \geq \beta\}. \quad (1)$$

Here $\mathbb{P}[y(\mathbf{X}) \leq t] = \int_{\mathbb{A}^N} \mathbb{I}_{\{y(\mathbf{x}) \leq t\}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$, where the indicator function $\mathbb{I}_{\{y(\mathbf{x}) \leq t\}}(\mathbf{x})$ is *one*, if $y(\mathbf{x}) \leq t$, and *zero*, otherwise. The $\text{CVaR}_{\beta}[y(\mathbf{X})]$ is the mean of the outputs $y(\mathbf{X})$ exceeding $\text{VaR}_{\beta}[y(\mathbf{X})]$. At level $\beta \in (0, 1)$, the CVaR_{β} can be determined as follows:

$$\text{CVaR}_{\beta}[y(\mathbf{X})] = \text{VaR}_{\beta}[y(\mathbf{X})] + \frac{1}{1-\beta} \mathbb{E}[(y(\mathbf{X}) - \text{VaR}_{\beta}[y(\mathbf{X})])_+], \quad (2)$$

where \mathbb{E} is the expectation operator with respect to $f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$, see [34, 35]. If the cumulative distribution function $\mathbb{P}[Y \leq t]$ is continuous at $t = \text{VaR}_{\beta}[y(\mathbf{X})]$, equation (2) is equivalent to

$$\text{CVaR}_{\beta}[y(\mathbf{X})] = \frac{1}{1-\beta} \mathbb{E}[y(\mathbf{X}) \cdot \mathbb{I}_{\mathcal{G}_{\beta}[y(\mathbf{X})]}(\mathbf{X})], \quad (3)$$

where the *risk region* for CVaR_{β} is defined as

$$\mathcal{G}_{\beta}[y(\mathbf{X})] := \{\mathbf{x} : y(\mathbf{x}) \geq \text{VaR}_{\beta}[y(\mathbf{X})]\} \subset \mathbb{A}^N. \quad (4)$$

With the definition of CVaR at hand, we can formalize the problem that we focus on in this work.

Problem 1. Consider a high-dimensional and dependent random input vector $\mathbf{X} \in \mathbb{A}^N$. Given a highly nonlinear and/or nonsmooth output $y : \mathbb{A}^N \rightarrow \mathbb{R}$, requiring a computationally expensive model evaluation, our goal is to compute an unbiased estimate of $\text{CVaR}_{\beta}[y(\mathbf{X})]$ accurately and efficiently.

2.5. Dimensionally decomposed generalized polynomial chaos expansion

For risk or reliability problems under high-dimensional inputs (e.g., $N \geq 20$), *regular* GPCE [18] or GPCE-based design methods [19, 20] require a relatively large number of basis functions, thus suffering from the curse of dimensionality. Appendix A provides details of *regular* GPCE. This section briefly summarizes the more efficient DD-GPCE approximation [21], which reorders and truncates the regular GPCE to select a subset of the basis functions of the regular GPCE. The truncation is based on the degree of interaction among input variables, thereby tackling the curse of dimensionality to some extent. The chosen multivariate orthonormal polynomial basis functions are consistent with an arbitrary, non-product-type probability measure $f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$ of \mathbf{X} . The bases can be determined by a three-step algorithm using a whitening transformation of the monomial basis, see Appendix B.

Given an input random vector \mathbf{X} with known probability density function $f_{\mathbf{X}}(\mathbf{x})$, consider an output $y(\mathbf{X})$ as defined in Section 2.3. To define the S -variate, m th-order DD-GPCE basis (see Appendix B for details), we require $0 \leq S \leq N$ and $S \leq m < \infty$ and define

$$\Psi_{S,m}(\mathbf{x}) := (\Psi_1(\mathbf{x}), \dots, \Psi_{L_{N,S,m}}(\mathbf{x}))^{\top} \quad (5)$$

an $L_{N,S,m}$ -dimensional column vector of orthonormal polynomials that are consistent with the probability measure $f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$. Here, the dimension of $\Psi_{S,m}(\mathbf{x})$ is $L_{N,S,m} := 1 + \sum_{s=1}^S \binom{N}{s} \binom{m}{s}$. The

S -variate, m th-order DD-GPCE approximation is then

$$y_{S,m}(\mathbf{X}) = \mathbf{c}^\top \Psi_{S,m}(\mathbf{X}) \simeq y(\mathbf{X}) \quad (6)$$

with the column vector $\mathbf{c} = (c_1, \dots, c_{L_{N,S,m}})^\top$, whose elements are the expansion coefficients

$$c_i := \int_{\mathbb{A}^N} y(\mathbf{x}) \Psi_i(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad i = 1, \dots, L_{N,S,m}. \quad (7)$$

Based on (5), a DD-GPCE approximation retains the degree of interaction among input variables less than or equal to S and preserves polynomial orders less than or equal to m .

2.6. Importance sampling

Importance sampling (IS) is a variance reduction method to improve the efficiency of MCS, and it follows two steps: (1) construct a biasing density, which is biased towards a specific region in input space (e.g., the risk region) and (2) sample from the biasing density and evaluate the high-fidelity model to estimate statistics of the output. The resulting IS estimator re-weights the drawn samples to compensate for the increased occurrence of biased samples in the sampling estimate [12]. To illustrate this for CVaR estimation, consider a random input vector \mathbf{X} with a probability density function $f_{\mathbf{X}}(\mathbf{x})$ on $\mathbb{A}^N \subseteq \mathbb{R}^N$, consider the output $y(\mathbf{X})$. Let $f_{\mathbf{Z}}(\mathbf{z})$ be a biasing density that defines a random vector $\mathbf{Z} \in \bar{\mathbb{A}}^N$ such that $\mathbb{A}^N \subseteq \bar{\mathbb{A}}^N$. Consider the risk region (4) corresponding to $\text{CVaR}_\beta[y(\mathbf{X})]$, that is a small region in input space associated with the output exceeding $\text{VaR}_\beta[y(\mathbf{X})]$. From (3), we have

$$\text{CVaR}_\beta[y(\mathbf{X})] = \frac{1}{1-\beta} \mathbb{E}[\mathbb{I}_{\mathcal{G}_\beta[y(\mathbf{Z})]}(\mathbf{Z}) y(\mathbf{Z}) w(\mathbf{Z})], \quad (8)$$

where $w(\mathbf{Z}) := \frac{f_{\mathbf{X}}(\mathbf{Z})}{f_{\mathbf{Z}}(\mathbf{Z})}$ is a weight function and the expectation operator \mathbb{E} is with respect to the biasing density $f_{\mathbf{Z}}(\mathbf{z})$ of the random input vector \mathbf{Z} . In IS, it is sufficient that the biasing density $f_{\mathbf{Z}}(\mathbf{z}) > 0$ for $\mathbf{z} \in \mathcal{G}_\beta[y(\mathbf{X})] \subset \bar{\mathbb{A}}^N$. The expectation in (8) can be approximated by MCS with samples $\mathbf{z}_1, \dots, \mathbf{z}_M$, $M \in \mathbb{N}$, drawn from the biasing density $f_{\mathbf{Z}}(\mathbf{z})$. Therefore, the IS estimator of $\text{CVaR}_\beta[y(\mathbf{X})]$ is

$$\widehat{\text{CVaR}}_\beta^{\text{IS}}[y(\mathbf{X})] = \frac{1}{M} \sum_{l=1}^M \mathbb{I}_{\mathcal{G}_\beta[y(\mathbf{X})]}(\mathbf{z}^{(l)}) w(\mathbf{z}^{(l)}). \quad (9)$$

Compared to MCS, one typically needs much fewer samples M for IS to obtain converged CVaR. In IS, a biasing density $f_{\mathbf{Z}}(\mathbf{z})$ is designed so that it minimizes the variance of the IS estimator (9).

2.7. Multifidelity importance sampling

Multifidelity importance sampling (MFIS) uses a surrogate model to design the biasing density while using the high-fidelity model to evaluate the IS estimator. In [12], the authors introduced a ROM-based MFIS method tailored to CVaR estimation. Given an error bound relating the output y and its ROM approximation \tilde{y} , i.e. $|y(\mathbf{x}) - \tilde{y}(\mathbf{x})| \leq \epsilon(\mathbf{x})$ for $\mathbf{x} \in \mathbb{A}^N$, the ϵ -risk region is defined by

$$\mathcal{G}_\beta^\epsilon[\tilde{y}(\mathbf{X})] := \{\mathbf{x} : \tilde{y}(\mathbf{x}) + \epsilon(\mathbf{x}) \geq \text{VaR}_\beta[\tilde{y}(\mathbf{X}) - \epsilon(\mathbf{X})]\}. \quad (10)$$

By using an asymptotic result of the variance of $\widehat{\text{CVaR}}_\beta^{\text{IS}}[y(\mathbf{X})]$, [12, Section 4], showed that the optimal biasing density and the weight function in the context of CVaR estimation are

$$f_{\mathbf{Z}}(\mathbf{z}) := \frac{\mathbb{I}_{\mathcal{G}_\beta^\epsilon[\tilde{y}(\mathbf{X})]}(\mathbf{z}) f_{\mathbf{X}}(\mathbf{z})}{\mathbb{P}[\mathcal{G}_\beta^\epsilon[\tilde{y}(\mathbf{X})]]} \quad \text{and} \quad w(\mathbf{z}) := \frac{\mathbb{P}[\mathcal{G}_\beta^\epsilon[\tilde{y}(\mathbf{X})]]}{\mathbb{I}_{\mathcal{G}_\beta^\epsilon[\tilde{y}(\mathbf{X})]}(\mathbf{z})}. \quad (11)$$

The MFIS estimator for CVaR then samples from $f_{\mathbf{Z}}(\mathbf{z})$ and reweighs the samples according to (11).

3. Dimensionally decomposed GPCE-Kriging for CVaR estimation

This section proposes a new DD-GPCE-Kriging method to efficiently and accurately estimate CVaR. Compared to the DD-GPCE introduced in Section 2.4, the DD-GPCE-Kriging leverages both DD-GPCE and Kriging. This allows for a more accurate approximation than DD-GPCE for highly nonlinear and nonsmooth outputs under dependent and high-dimensional random inputs. A combination of standard PCE and Kriging was first proposed in [37] in the context of reliability analysis. We propose DD-GPCE-Kriging in Section 3.1 and give details on the calculation of the DD-GPCE-Kriging predictor in Section 3.2. In Section 3.3, we present an algorithm to estimate CVaR via DD-GPCE-Kriging.

3.1. DD-GPCE-Kriging model

For $y(\mathbf{X}) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, we combine the S -variate, m th-order DD-GPCE approximation (6) with a Kriging model. The combination (referred to as DD-GPCE-Kriging) results in a Gaussian random variable of $y(\mathbf{x})$, i.e.,

$$\bar{Y}_{S,m}(\mathbf{x}) = \mathbf{c}^\top \Psi_{S,m}(\mathbf{x}) + \sigma^2 Z(\mathbf{x}; \boldsymbol{\theta}), \quad (12)$$

where the first and the second terms in the right-hand side of (12) indicate the S -variate, m th-order DD-GPCE and Kriging, respectively. In (12), $\mathbf{c}^\top \Psi_{S,m}(\mathbf{x})$ follows the definition in (6), σ^2 is the Gaussian process variance, and $Z(\mathbf{x}; \boldsymbol{\theta})$ is a zero-mean, unit-variance stationary Gaussian process, which is fully determined by the autocorrelation function between two distinct input realizations \mathbf{x} and \mathbf{x}' , i.e.,

$$R(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}) = R(|\mathbf{x} - \mathbf{x}'|; \boldsymbol{\theta}).$$

Here, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)^\top$ is the N -dimensional column vector of hyper-parameters to be computed. In this work, we chose either the Gaussian autocorrelation function

$$R(|\mathbf{x} - \mathbf{x}'|; \boldsymbol{\theta}) = \exp \left[- \sum_{i=1}^N \left(\frac{x_i - x'_i}{\theta_i} \right)^2 \right] \quad (13)$$

or the exponential autocorrelation function

$$R(|\mathbf{x} - \mathbf{x}'|; \boldsymbol{\theta}) = \exp \left[- \sum_{i=1}^N \frac{|x_i - x'_i|}{\theta_i} \right]. \quad (14)$$

We choose the autocorrelation function that most accurately predicts the quantity of interest for the particular application. There are also other choices (e.g., the Matern autocorrelation function), see [8, 37]. In (13) or (14), an N -dimensional vector $\boldsymbol{\theta}$ of hyper-parameters can be determined by a maximum likelihood estimate [24] or a leave-one-out cross-validation estimate [1]. In this work, we select the leave-one-out cross-validation estimate method to evaluate all numerical examples since it provides more stable solutions than the maximum likelihood estimate method when the autocorrelation function family for outputs is unknown (see [1, 37]). In Appendix C, we give details on the leave-one-out cross-validation estimate.

Remark 1. We can view DD-GPCE-Kriging as an extension of PCE-Kriging [37]. In contrast to the conventional PCE-Kriging, the DD-GPCE-Kriging method can handle arbitrary, dependent random inputs directly without relying on a detrimental transformation (e.g., Rosenblatt or Nataf transformation) that may slow the convergence rate of the estimate to the true output. Moreover, the DD-GPCE part itself alleviates the curse of dimensionality to some extent by effectively truncating the basis functions and expansion coefficients in a dimensionwise manner.

3.2. Calculation of DD-GPCE-Kriging predictor

Kriging is a stochastic interpolation method that provides the best predictor and an estimate of its accuracy. Given the known distribution of a random input \mathbf{X} and an output model $y : \mathbb{A}^N \rightarrow \mathbb{R}$, consider an input-output data set $\{\mathbf{x}^{(l)}, y(\mathbf{x}^{(l)})\}_{l=1}^{L'}$ of size $L' \in \mathbb{N}$ constructed by evaluating the quantity of interest y at each input data \mathbf{x} , which are drawn as part of a sampling method (e.g., standard MCS, quasi MCS, or Latin hypercube sampling). We use the input-output data set to determine the S th-variate, m th-order DD-GPCE-Kriging at a new realization \mathbf{x} . Then, as *the best predictor* for $y(\mathbf{x})$, we have the mean $\bar{y}_{S,m}(\mathbf{x})$ of the Gaussian random variable $\bar{Y}_{S,m}$ in (12), i.e.,

$$\bar{y}_{S,m}(\mathbf{x}) = \hat{\mathbf{c}}^\top \Psi_{S,m}(\mathbf{x}) + \mathbf{r}(\mathbf{x})^\top \mathbf{R}^{-1}(\mathbf{b} - \mathbf{A}\hat{\mathbf{c}}), \quad (15)$$

and the variance of $\bar{Y}_{S,m}$ is

$$\begin{aligned} \bar{\sigma}_{S,m}^2(\mathbf{x}) = & \hat{\sigma}^2 \left\{ 1 - \mathbf{r}(\mathbf{x})^\top \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}) + [\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}) - \Psi_{S,m}(\mathbf{x})]^\top (\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A})^{-1} \right. \\ & \left. \times [\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}) - \Psi_{S,m}(\mathbf{x})] \right\}, \end{aligned} \quad (16)$$

where

$$\begin{aligned} \mathbf{A} &:= \begin{bmatrix} \Psi_1(\mathbf{x}^{(1)}) & \cdots & \Psi_{L_{N,S,m}}(\mathbf{x}^{(1)}) \\ \vdots & \ddots & \vdots \\ \Psi_1(\mathbf{x}^{(L')}) & \cdots & \Psi_{L_{N,S,m}}(\mathbf{x}^{(L')}) \end{bmatrix}, \\ \mathbf{R} &:= \begin{bmatrix} R(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}; \boldsymbol{\theta}) & \cdots & R(\mathbf{x}^{(1)}, \mathbf{x}^{(L')}; \boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ R(\mathbf{x}^{(L')}, \mathbf{x}^{(1)}; \boldsymbol{\theta}) & \cdots & R(\mathbf{x}^{(L')}, \mathbf{x}^{(L')}; \boldsymbol{\theta}) \end{bmatrix}, \end{aligned} \quad (17)$$

and $\mathbf{r}(\mathbf{x}) = (R(\mathbf{x}, \mathbf{x}^{(1)}; \boldsymbol{\theta}), \dots, R(\mathbf{x}, \mathbf{x}^{(L')}; \boldsymbol{\theta}))^\top$ and $\Psi_{S,m}(\mathbf{x}) = (\Psi_1(\mathbf{x}), \dots, \Psi_{L_{N,S,m}}(\mathbf{x}))^\top$. In (15), $\mathbf{b} = (y(\mathbf{x}^{(1)}), \dots, y(\mathbf{x}^{(L')}))^\top$ is an L' -dimensional column vector of outputs evaluated at each input $\mathbf{x}^{(l)}$, $l = 1, \dots, L'$, and $\hat{\mathbf{c}}$ is an $L_{N,S,m}$ -dimensional solution vector to the DD-GPCE's coefficients \mathbf{c} in (12), determined by

$$\hat{\mathbf{c}} = (\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{b}. \quad (18)$$

In (16), the variance term is

$$\hat{\sigma}^2 = \frac{1}{L'} (\mathbf{b} - \mathbf{A}\hat{\mathbf{c}})^\top \mathbf{R}^{-1} (\mathbf{b} - \mathbf{A}\hat{\mathbf{c}}). \quad (19)$$

In DD-GPCE-Kriging, the DD-GPCE model expands the output $y(\mathbf{x})$ into a set of polynomials to fit the mean behavior of the output. The Kriging model interpolates local variations with the weighted neighboring samples. As a result, DD-GPCE-Kriging leverages both DD-GPCE and Kriging, thus effectively predicting a highly nonlinear, nonsmooth output.

3.3. Sampling-based CVaR estimation by DD-GPCE-Kriging

The estimation of VaR (1) and CVaR (2) for nontrivial examples requires a sampling method (e.g., standard MCS, quasi MCS, or Latin hypercube sampling). In this work, we follow [12, Alg. 2.1] and use standard MCS. We use the predictor $\bar{y}_{S,m}$ of an S -variate, m th-order DD-GPCE-Kriging (see Section 3.2) as a surrogate model to replace an expensive high-fidelity function $y(\mathbf{X})$. Therefore, we can afford a large number L of output samples to obtain statistically converged CVaR estimates when using the DD-GPCE-Kriging surrogate. The sampling-based estimation method evaluates $\bar{y}_{S,m}(\mathbf{x})$, and we denote the resulting estimates by $\widehat{\text{VaR}}_\beta[\bar{y}_{S,m}(\mathbf{X})]$ and $\widehat{\text{CVaR}}_\beta[\bar{y}_{S,m}(\mathbf{X})]$.

Algorithm 1 summarizes the steps of the proposed CVaR estimation process. Estimating VaR is

straightforward (see Steps 3-4). From the definition of $\text{CVaR}_\beta[y(\mathbf{X})]$ in (2), we obtain the estimate

$$\widehat{\text{CVaR}}_\beta[\bar{y}_{S,m}(\mathbf{X})] = \widehat{\text{VaR}}_\beta[\bar{y}_{S,m}(\mathbf{X})] + \frac{1}{1-\beta} \frac{1}{L} \sum_{l=1}^L \left(\bar{y}_{S,m}(\mathbf{x}^{(l)}) - \widehat{\text{VaR}}_\beta[\bar{y}_{S,m}(\mathbf{X})] \right)_+. \quad (20)$$

We use (20) in the final step of Algorithm 1.

Algorithm 1 Sampling-based estimation of CVaR_β by the DD-GPCE-Kriging predictor $\bar{y}_{S,m}$.

Input: Input samples $\mathbf{x}^{(l)} = (x_1^{(l)}, \dots, x_N^{(l)})^\top$, $l = 1, \dots, L$, $L \gg 1$, via MCS, quasi MCS, or Latin hypercube sampling with corresponding probabilities $p^{(l)} = \int_{\mathbf{X}} \delta_{\mathbf{x}^{(l)}} d\mathbf{x}^{(l)}$; Set sample number $L' < L$, used for the computation of DD-GPCE-Kriging; Set a risk level $\beta \in (0, 1)$.

Output: Estimates $\widehat{\text{VaR}}_\beta[\bar{y}_{S,m}(\mathbf{X})]$ and $\widehat{\text{CVaR}}_\beta[\bar{y}_{S,m}(\mathbf{X})]$.

- 1: **procedure** CALCULATE DD-GPCE-KRIGING($\bar{y}_{S,m}(\mathbf{x})$)
- 2: Create measure-consistent orthonormal basis vector $\Psi_{S,m}(\mathbf{x})$. We refer to Appendix B.
- 3: Determine hyper-parameters θ by either of a maximum likelihood estimate or a leave-one-out cross-validation estimate, see Section 3.1 and Appendix C.
- 4: Calculate DD-GPCE's coefficients \mathbf{c} (18).
- 5: **end procedure**
- 6: **procedure** CREATE OUTPUT SAMPLES($\{\bar{y}_{S,m}(\mathbf{x}^{(l)})\}_{l=1}^L$)
- 7: Create L output samples of the predictor $\bar{y}_{S,m}$ (15) of DD-GPCE-Kriging for y .
- 8: **end procedure**
- 9: Sort values of $\bar{y}_{S,m}$ in descending order and relabel the samples so that

$$\bar{y}_{S,m}(\mathbf{x}^{(1)}) > \bar{y}_{S,m}(\mathbf{x}^{(2)}) > \dots > \bar{y}_{S,m}(\mathbf{x}^{(L)}),$$

and reorder the probabilities accordingly (so that $p^{(l)}$ corresponding to $\mathbf{x}^{(l)}$).

- 10: Compute the index $\tilde{k}_\beta \in \mathbb{N}$ such that

$$\sum_{l=1}^{\tilde{k}_\beta-1} p^{(l)} \leq 1-\beta < \sum_{l=1}^{\tilde{k}_\beta} p^{(l)}.$$

- 11: Set $\text{VaR}_\beta[y(\mathbf{X})] \approx \widehat{\text{VaR}}_\beta[\bar{y}_{S,m}(\mathbf{X})] = \bar{y}_{S,m}(\mathbf{x}^{(\tilde{k}_\beta)})$.
 - 12: Set $\text{CVaR}_\beta[y(\mathbf{X})] \approx \widehat{\text{CVaR}}_\beta[\bar{y}_{S,m}(\mathbf{X})]$ (20).
-

To satisfy the accuracy of the DD-GPCE-Kriging's predictor, we need to determine the best-fit parameters (e.g., $\hat{\mathbf{c}}$, $\hat{\sigma}$, or θ in (18) and (19)). We select the number of output samples to be at least greater than the number $L_{N,S,m}$ of DD-GPCE's coefficients or basis functions. Consequently, when faced with high-dimensional inputs (e.g., $N \geq 20$), DD-GPCE-Kriging may require thousands of output samples. In such cases, computing a DD-GPCE-Kriging surrogate may require prohibitive computational cost when each sample is determined by an expensive high-fidelity model evaluation. Moreover, when not using enough samples in the computation of DD-GPCE-Kriging, a significant bias in the CVaR estimate may be present. In the following section, we propose to use the DD-GPCE-Kriging surrogate as a part of an MFIS framework. This further reduces the computational expense, while obtaining unbiased CVaR estimates via importance sampling.

4. Multifidelity importance sampling for CVaR estimation

We present a novel multifidelity importance sampling strategy that leverages the DD-GPCE-Kriging surrogates (tailored to CVaR) to obtain an unbiased CVaR estimate for a highly nonlinear

and nonsmooth random output. In contrast to existing methods [29, 12, 17], the proposed MFIS-based method works for high-dimensional and dependent random inputs (see Problem 1). In the proposed MFIS-based method, we use the DD-GPCE-Kriging surrogate to determine the biasing density, from which we then draw a few ($M \ll L$) high-fidelity output samples to estimate VaR_β and CVaR_β . We use the confidence interval (CI) of the predictor $\bar{y}_{S,m}$ to estimate the risk region for an optimal biasing density (11) which is determined via the CI-based ϵ -risk region. We define the CI-based ϵ -risk region in Section 4.1, present the importance sampling in Section 4.2, and introduce the complete algorithm to obtain unbiased CVaR_β estimates in Section 4.3.

4.1. Confidence interval based ϵ -risk region

Algorithm 2 CI-based ϵ -risk region by the DD-GPCE-Kriging predictor $\bar{y}_{S,m}$.

Input: Samples $\mathbf{x}^{(l)} = (x_1^{(l)}, \dots, x_N^{(l)})^\top$, $l = 1, \dots, L$, via MCS, quasi MCS, or Latin hypercube sampling with corresponding probabilities $p^{(l)} = f_{\mathbf{X}}(\mathbf{x}^{(l)})d\mathbf{x}^{(l)}$; Generate the input-output data set $\{\mathbf{x}^{(l)}, y(\mathbf{x}^{(l)})\}_{l=1}^{\bar{L}}$, $\bar{L} < L$; Set risk level $\beta \in (0, 1)$.

Output: CI-based ϵ -risk region $\hat{\mathcal{G}}_\beta^\alpha$.

- 1: Calculate the S -variate, m th-order DD-GPCE-Kriging $\bar{Y}_{S,m}$ through Steps 1–5 of Algorithm 1.
- 2: Evaluate the mean value $\bar{y}_{S,m}(\mathbf{x}^{(l)})$ (15) and the standard deviation $\bar{\sigma}_{S,m}(\mathbf{x}^{(l)})$ (16) of $\bar{Y}_{S,m}$ at samples $\{\mathbf{x}^{(l)}\}_{l=1}^{\bar{L}}$ and determine the corresponding the lower or upper limits of CI, $\epsilon_{S,m}^\alpha(\mathbf{x}^{(l)})$ (22).
- 3: Sort values of $(\bar{y}_{S,m}(\mathbf{x}^{(l)}) - \epsilon_{S,m}^\alpha(\mathbf{x}^{(l)}))$ in descending order and relabel those, i.e.,

$$\bar{y}_{S,m}(\mathbf{x}^{(1)}) - \epsilon_{S,m}^\alpha(\mathbf{x}^{(1)}) > \dots > \bar{y}_{S,m}(\mathbf{x}^{(L)}) - \epsilon_{S,m}^\alpha(\mathbf{x}^{(L)}).$$

- 4: Compute an index \bar{k}_β such that

$$\sum_{l=1}^{\bar{k}_\beta-1} p^{(l)} \leq 1 - \beta < \sum_{l=1}^{\bar{k}_\beta} p^{(l)}.$$

- 5: Set

$$\widehat{\text{VaR}}_\beta[\bar{y}_{S,m}(\mathbf{X}) - \epsilon_{S,m}^\alpha(\mathbf{X})] = \bar{y}_{S,m}(\mathbf{x}^{(\bar{k}_\beta)}).$$

- 6: Determine the $(1 - \alpha)100\%$ CI-based ϵ -risk region by a discrete set:

$$\hat{\mathcal{G}}_\beta^\alpha := \{\mathbf{x}^{(j)} \in \mathbb{A}^N : \bar{y}_{S,m}(\mathbf{x}^{(j)}) + \epsilon_{S,m}^\alpha(\mathbf{x}^{(j)}) \geq \widehat{\text{VaR}}_\beta[\bar{y}_{S,m}(\mathbf{X}) - \epsilon_{S,m}^\alpha(\mathbf{X})], j = 1, \dots, \bar{L} < L\}.$$

The random variable $\bar{Y}_{S,m}$ of DD-GPCE-Kriging is assumed to follow the Gaussian distribution with the mean $\bar{y}_{S,m}(\mathbf{x})$ in (15) and variance $\bar{\sigma}_{S,m}^2(\mathbf{x})$ in (16). Given $\alpha \in [0, 1]$, the upper and lower limits of the $(1 - \alpha)100\%$ confidence interval of the DD-GPCE-Kriging predictor $\bar{y}_{S,m}$ at input \mathbf{x} satisfy

$$-Q_{1-\alpha/2}\bar{\sigma}_{S,m}(\mathbf{x}) \leq y(\mathbf{x}) - \bar{y}_{S,m}(\mathbf{x}) \leq Q_{1-\alpha/2}\bar{\sigma}_{S,m}(\mathbf{x}). \quad (21)$$

Here, $Q_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -level quantile of the standard normal distribution and $\bar{\sigma}_{S,m}^2(\mathbf{x})$ is the predicted variance at \mathbf{x} by DD-GPCE-Kriging. By expressing $Q_{1-\alpha/2}\bar{\sigma}_{S,m}(\mathbf{x})$ as $\epsilon_{S,m}^\alpha(\mathbf{x})$, equation (21) becomes

$$-\epsilon_{S,m}^\alpha(\mathbf{x}) \leq y(\mathbf{x}) - \bar{y}_{S,m}(\mathbf{x}) \leq \epsilon_{S,m}^\alpha(\mathbf{x}). \quad (22)$$

In reference to Section 2.6, by replacing ϵ in (10) with $\epsilon_{S,m}^\alpha$ in (22), the ϵ -risk region corresponding to $\text{CVaR}_\beta[y(\mathbf{X})]$ becomes the CI-based ϵ -risk region

$$\mathcal{G}_\beta^\alpha := \{\mathbf{x} \in \mathbb{A}^N : \bar{y}_{S,m}(\mathbf{x}) + \epsilon_{S,m}^\alpha(\mathbf{x}) \geq \text{VaR}_\beta[\bar{y}_{S,m}(\mathbf{X}) - \epsilon_{S,m}^\alpha(\mathbf{X})]\}. \quad (23)$$

The procedure for computing the CI-based ϵ -risk region (23) is presented in Algorithm 2.

Remark 2. *In contrast to the previous work [12], the CI (as a probabilistic bound) does not guarantee that the true risk region is included in the CI-based ϵ -risk region at all times. We note that the ϵ -risk region is only used to define the biasing density. Thus, since the MFIS estimator uses high-fidelity output samples, it is always unbiased, even if the CI-based ϵ -risk region is inaccurate. The quality of the ϵ -risk region only impacts the variance of the CVaR estimate.*

4.2. Importance sampling for CVaR estimation

For importance sampling, we can determine an optimal biasing density by minimizing the asymptotic variance of the CVaR_β estimator (e.g., see Section 4.2 of [12]). We use the CI-based ϵ -risk region to construct the optimal biasing density as

$$f_{\mathbf{z}}(\mathbf{z}) := \frac{\mathbb{I}_{\mathcal{G}_\beta^\alpha[\bar{y}_{S,m}(\mathbf{x})]}(\mathbf{z}) f_{\mathbf{X}}(\mathbf{z})}{\mathbb{P}[\mathcal{G}_\beta^\alpha[\bar{y}_{S,m}(\mathbf{X})]]}, \quad (24)$$

where $\mathbb{I}_{\mathcal{G}_\beta^\alpha[\bar{y}_{S,m}(\mathbf{x})]}(\mathbf{z})$ is the indicator function associated with the CI-based ϵ -risk region. Inserting $f_{\mathbf{z}}(\mathbf{z})$ from equation (24) into the weight function (11) leads to

$$w(\mathbf{z}) := \frac{\mathbb{P}[\mathcal{G}_\beta^\alpha[\bar{y}_{S,m}(\mathbf{X})]]}{\mathbb{I}_{\mathcal{G}_\beta^\alpha[\bar{y}_{S,m}(\mathbf{x})]}(\mathbf{z})}. \quad (25)$$

We draw a few high-fidelity output samples $\{y(\mathbf{z}^{(l)})\}_{l=1}^M$, where $M \ll L$, from (24) and use (25) and (9) to obtain the importance sampling estimate of CVaR_β , namely $\widehat{\text{CVaR}}_\beta^{\text{IS}}[y(\mathbf{X})]$ (9).

4.3. Proposed multifidelity importance sampling for CVaR estimation

Algorithm 3 outlines the procedure of the proposed MFIS-based method to obtain $\widehat{\text{CVaR}}_\beta^{\text{IS}}[y(\mathbf{X})]$, including Algorithms 2 and 3. In Step 1 of Algorithm 3, we suggest computing the DD-GPCE-Kriging surrogate with two options:

- (a) HF (High-fidelity) that uses the high-fidelity model evaluations,
- (b) LF (Low-fidelity) that uses the low-fidelity model evaluations.

We include here the low-fidelity model to speed up the computation of DD-GPCE-Kriging. We investigate how much fidelity/correlation of a surrogate model and the high-fidelity model is really needed to obtain good biasing densities. Oftentimes, the only option one may have is to use a coarse-grid approximation in order to generate enough data to train a surrogate model. We emphasize again that the proposed MFIS-based method with LF can provide an unbiased CVaR estimate even when the low-fidelity model has a significant error compared to the high-fidelity model. Absolute error, as it turns out, is not the sole driving factor, as *correlation* with the high-fidelity output is a key driver, see also [30]. We demonstrate this with a variety of (artificially created) low-fidelity models on a mathematical example in Section 5.

In Algorithm 3, the cost to determine the CVaR estimate is dominated by constructing two sets of input-output data of size L' and M . For the first input-output data set, its size L' is determined as a multiple of the number of DD-GPCE's coefficients or basis functions $L_{N,S,m}$, satisfying the necessary condition ($L' \leq L_{N,S,m}$). In the proposed algorithm of MFIS, the computational cost to determine the first input-output data set can be reduced by replacing a high-fidelity output model with its low-fidelity version. For the second input-output data set, its sample number M is usually chosen to be less than L' , and in this work, the chosen number is $L'/2$, but one can select a different number via a convergence test for the CVaR estimator in other problems.

Let $c_T \in \mathbb{R}_0^+$ be the total computational budget and $c_H \in \mathbb{R}_0^+$ and $c_L \in \mathbb{R}_0^+$ be the costs of the high-fidelity and low-fidelity model evaluations, respectively. Then, the total cost c_T of Algorithm 3 is, in the HF option

$$c_T = (L' + M)c_H$$

and, in the LF option

$$c_T = L'c_L + Mc_H.$$

Given a total computational budget c_T for LF, we can determine the cost of a low-fidelity model evaluation such that

$$c_L \leq \frac{c_T - Mc_H}{L'}.$$

Generally, $M < L'$ and $c_L < c_H$, therefore, the proposed MFIS-based method with LF has higher efficiency than its HF version. More optimal use of the computational budget can be achieved, see [47].

Algorithm 3 Proposed MFIS-based method for CVaR_β estimation.

Input: Samples $\mathbf{x}^{(l)} = (x_1^{(l)}, \dots, x_N^{(l)})^\top$, $l = 1, \dots, L$, $L \gg 1$, via MCS, quasi MCS, or Latin hypercube sampling with corresponding probabilities $p^{(l)} = f_{\mathbf{X}}(\mathbf{x}^{(l)})d\mathbf{x}^{(l)}$; Set a risk level $\beta \in (0, 1)$.

Output: $\widehat{\text{CVaR}}_\beta^{\text{IS}}[\bar{y}_{S,m}(\mathbf{X})]$.

- 1: Calculate the S -variate, m th-order DD-GPCE-Kriging $\bar{Y}_{S,m}$ through Steps 1–5 of Algorithm 1. For Steps 3 and 4 of Algorithm 1,
 - (a) HF (High-fidelity): Use high-fidelity outputs of size L' (e.g., fine mesh model for FEA).
 - (b) LF (Low-fidelity): Use low-fidelity outputs of size L' (e.g., coarse mesh model for FEA).
 - 2: Determine the $(1 - \alpha)100\%$ CI-based ϵ -risk region by a discrete set through Steps 2–6 of Algorithm 2.
 - 3: Select input samples $\mathbf{x}^{(l)}$, $l = 1, \dots, M$, within the CI-based ϵ -risk region. Then, construct high-fidelity output samples of y at $\mathbf{x}^{(l)}$.
 - 4: Assign values of the weight function (25), i.e., $w(\mathbf{x}^{(1)}), \dots, w(\mathbf{x}^{(M)}) \equiv \mathbb{P}[\widehat{\mathcal{G}}_\beta^\alpha[y(\mathbf{X})]] \approx |\widehat{\mathcal{G}}_\beta^\alpha[y(\mathbf{X})]|/L$, where $|\cdot|$ denotes the cardinality of a set.
 - 5: Obtain $\text{CVaR}_\beta[y(\mathbf{X})] \approx \widehat{\text{CVaR}}_\beta^{\text{IS}}[\bar{y}_{S,m}(\mathbf{X})]$ from (9) with $y(\mathbf{x}^{(l)})$, $p^{(l)} = w(\mathbf{x}^{(l)})/M$, $l = 1, \dots, M$.
-

Figure 1 presents a flow chart for the proposed MFIS-based method. It starts with drawing the samples for the random input vector \mathbf{X} . Given the computational budget c_T , we choose to evaluate either the high-fidelity or low-fidelity model to calculate DD-GPCE-Kriging according to $c_T \leq (L' + M)c_H$. We calculate the mean $\bar{y}_{S,m}(\mathbf{x}^{(l)})$ and the standard deviation $\bar{\sigma}_{S,m}(\mathbf{x}^{(l)})$ of $\bar{Y}_{S,m}(\mathbf{x}^{(l)})$ for $l = 1, \dots, L$. From these, we estimate $\widehat{\text{VaR}}_\beta[\bar{y}_{S,m}(\mathbf{X}) - \epsilon_{S,m}^\alpha(\mathbf{X})]$ and then determine the CI-risk region \tilde{G}_β^α . We conduct importance sampling by selecting M input samples from \tilde{G}_β^α . The importance sampling process produces the CVaR estimate $\widehat{\text{CVaR}}_\beta^{\text{IS}}[y(\mathbf{X})]$ with only M high-fidelity outputs.

The next section demonstrates the computational efficiency of the proposed MCS and MFIS-based methods via practical engineering examples.

5. Numerical results

Four numerical examples are presented to evaluate the proposed DD-GPCE-Kriging-based MCS and MFIS methods for CVaR estimation. Sections 5.1 and 5.2 describe the details of the numerical setup and error measures, respectively, used for Examples 1–4. In Sections 5.3 and 5.4, our instructional examples consider two closed-form mathematical functions for which we construct a suite of surrogate models analytically. In Sections 5.5 and 5.6, we demonstrate the scalability of the proposed

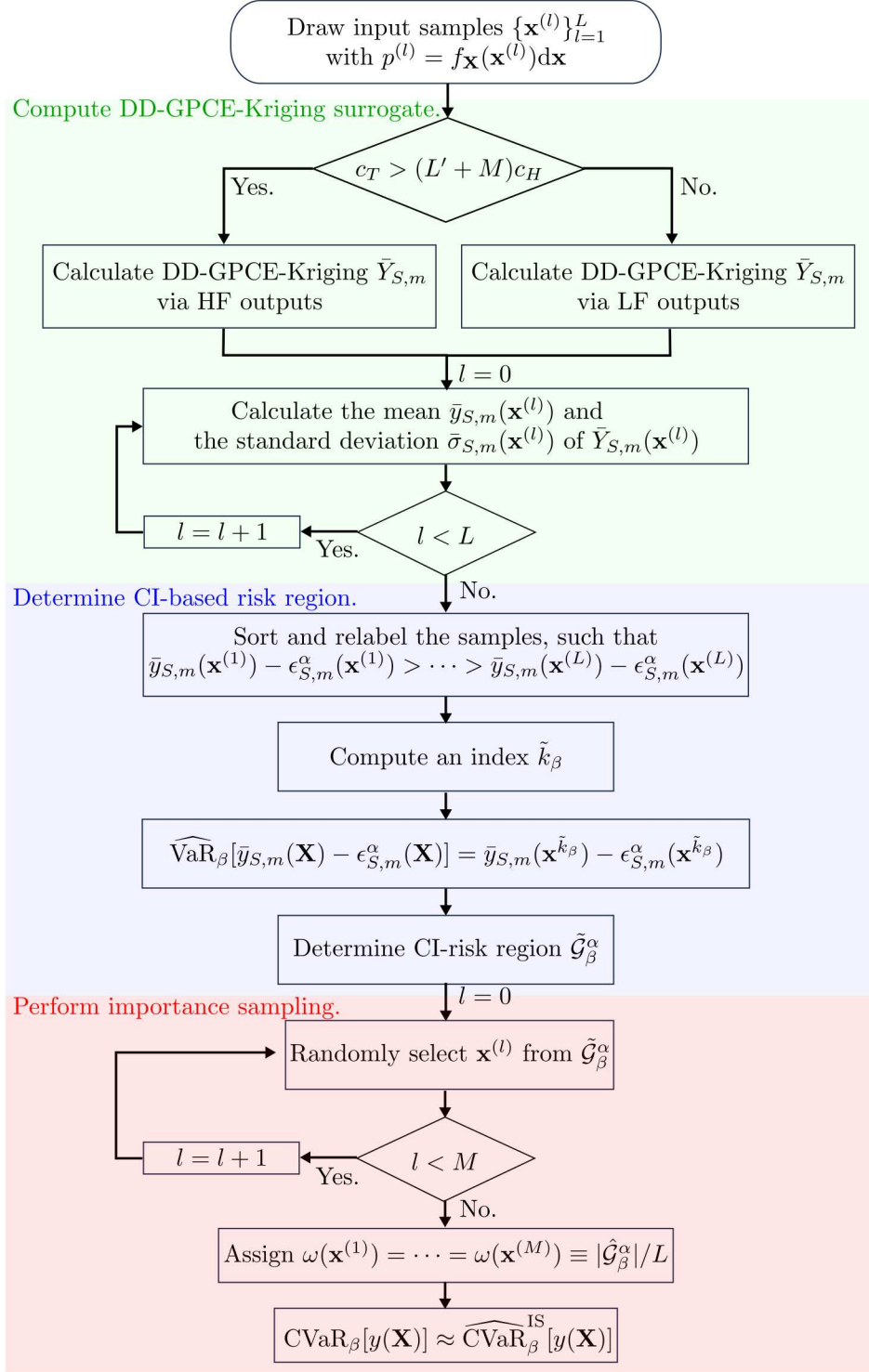


Figure 1: Flow chart for the proposed MFIS-based method to estimate Conditional Value-at-Risk.

methods and their applicability to complex engineering problems via a 2D glass/vinylester composite plate example and a 3D composite T-joint example, respectively.

5.1. Numerical setup

In Examples 1–3, the sample number for MCS is $L = 10,000$ and in Example 4, the sample number is $L = 6,000$ due to the higher computational cost associated with that model. In all four examples, the chosen sample numbers led to converged CVaR estimates. We deem estimates converged if the difference between the previous and current CVaR value differs by less than 0.1%.

In all four examples, the monomial moment matrix $\mathbf{G}_{S,m}$ (see (41) in Appendix B) used for creating $\Psi_{S,m}(\mathbf{x})$ in DD-GPCE-Kriging (12) is determined by quasi MCS with 5×10^6 samples together with the Sobol sequence. To determine the hyper-parameters θ via a leave-one-out cross-validation (presented in Appendix C), in all examples, we employed the trust-region reflective optimization algorithm, which is known to perform well for nonlinear optimization problems [7]. We obtained all numerical results via MATLAB [25] on an Intel Core i7-10850H 2.70 GHz processor with 64 GB of RAM.

5.2. Error measures

To measure the deviation of a CVaR estimate relative to its standard MCS estimate (which we consider the benchmark), we use the mean relative difference (MRD). The MRD is an average error measure from $K \in \mathbb{N}$ independent estimates, i.e.,

$$\text{MRD} = \frac{\frac{1}{K} \sum_{k=1}^K \left| \mathcal{Y}_k - \widehat{\text{CVaR}}_{\beta}[y(\mathbf{X})] \right|}{\left| \widehat{\text{CVaR}}_{\beta}[y(\mathbf{X})] \right|}, \quad (26)$$

where, $\mathcal{Y}_k = \widehat{\text{CVaR}}_{\beta}^{(k)}[\bar{y}_{S,m}(\mathbf{X})]$ or $\widehat{\text{CVaR}}_{\beta}^{\text{IS}(k)}[y(\mathbf{X})]$ is the estimate at the k th independent trial run for DD-GPCE-Kriging-based MCS or MFIS, respectively. We also provide the normalized root-mean-square deviation (N-RMSD) as

$$\text{N-RMSD} = \sqrt{\frac{\frac{1}{K} \sum_{k=1}^K \left(\mathcal{Y}_k - \widehat{\text{CVaR}}_{\beta}[y(\mathbf{X})] \right)^2}{\left(\widehat{\text{CVaR}}_{\beta}[y(\mathbf{X})] \right)^2}}, \quad (27)$$

which measures the standard deviation of a CVaR estimator relative to the benchmark. The N-RMSD has a normalized value to the benchmark.

Additionally, in Examples 1, 3, and 4, we determine the Pearson correlation coefficient (PCC) between the high-fidelity model output y_H and low-fidelity model output y_L from $P \gg 1$ samples, i.e.,

$$\rho_{H,L} = \frac{\frac{1}{P} \sum_{l=1}^P \left(y_H(\mathbf{x}^{(l)}) - \frac{1}{P} \sum_{l=1}^P y_H(\mathbf{x}^{(l)}) \right) \left(y_L(\mathbf{x}^{(l)}) - \frac{1}{P} \sum_{l=1}^P y_L(\mathbf{x}^{(l)}) \right)}{\sqrt{\frac{1}{P} \sum_{l=1}^P \left(y_H(\mathbf{x}^{(l)}) - \frac{1}{P} \sum_{l=1}^P y_H(\mathbf{x}^{(l)}) \right)^2 \frac{1}{P} \sum_{l=1}^P \left(y_L(\mathbf{x}^{(l)}) - \frac{1}{P} \sum_{l=1}^P y_L(\mathbf{x}^{(l)}) \right)^2}}, \quad (28)$$

where $\rho_{H,L} \in [-1, 1]$. In Examples 1, 3, and 4, the selected values of P in (28) are 10,000, 3,000, and 3,000, respectively. The correlation coefficient $\rho_{H,L}$ guides us in selecting a low-fidelity model y_L for computing DD-GPCE-Kriging in the proposed MFIS-based method with LF, similar to [30]. We

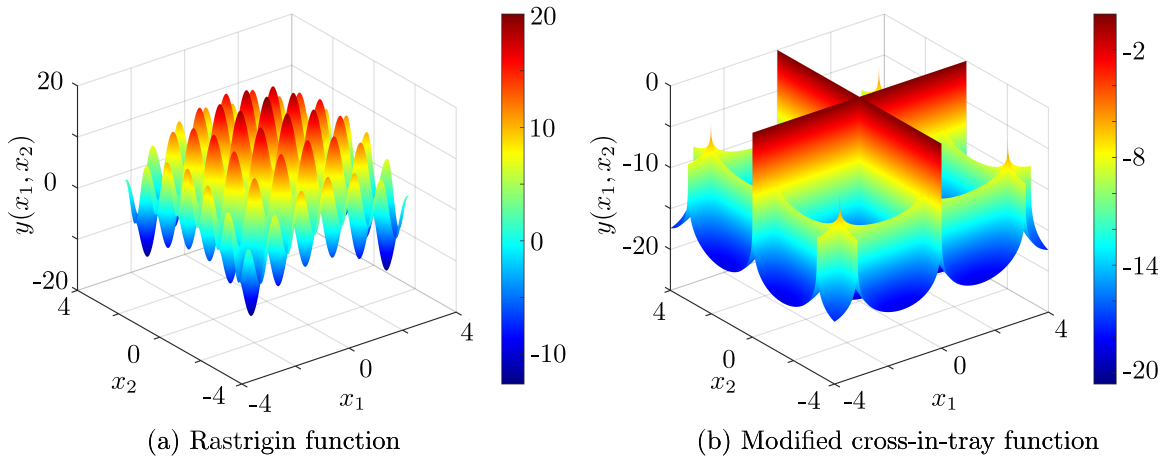


Figure 2: Graphs of two two-dimensional functions evaluated for CVaR_β in Examples 1 and 2. The Rastrigin function in Figure 2a exhibits high nonlinearity yet is smooth, while the modified cross-in-tray function in Figure 2b exhibits high nonlinearity and nonsmoothness. Each color map indicates the values of $y(x_1, x_2)$ on the domain $[-4, 4] \times [-4, 4]$.

ignore a low-fidelity model when its correlation coefficient with the high-fidelity model is close to *zero* (i.e., $|\rho_{H,L}| \ll 1$). In Example 1, we provide four distinct low-fidelity models $y_{L,i}$ and we compute the corresponding ρ_{H,L_i} by replacing y_L with y_{L_i} in (28) for $i = 1, 2, 3, 4$. The following example illustrates how retaining qualitative features of the high-fidelity model in the low-fidelity model can impact the accuracy of the estimator $\widehat{\text{CVaR}}_\beta^{\text{IS}}$.

5.3. Example 1: Rastrigin function and its surrogates

This first example illustrates the DD-GPCE-Kriging-based MCS and MFIS methods for a highly nonlinear analytical mathematical function for which we construct simple surrogates.

5.3.1. Problem definition

For a bivariate Gaussian random vector $\mathbf{X} = (X_1, X_2)^\top$ with mean vector $(0, 0)^\top$ and standard deviations $\sqrt{\text{var}[X_i]} = 2$, $i = 1, 2$, consider the Rastrigin function

$$y(X_1, X_2) = 10 - \sum_{i=1}^2 (X_i^2 - 5 \cos(2\pi X_i)). \quad (29)$$

We consider two distinct cases of \mathbf{X} , where the correlation coefficient between X_1 and X_2 is 0.9 in the first case, and zero in the second case. Since the random input variables follow a Gaussian distribution, the *zero* correlation further means that they are independent.

Figure 2a shows the Rastrigin function on the domain $[-4, 4] \times [-4, 4]$. We see that the Rastrigin function has smooth, continuous, and highly oscillatory behavior. The function is notoriously difficult to approximate with a polynomial basis as it requires high-degree polynomial bases.

5.3.2. High-fidelity and low-fidelity models

This example seeks to illustrate that correlation between the low-fidelity and high-fidelity model is instrumental to construct good biasing densities. We consider the Rastrigin function as the high-fidelity quantity of interest (‘truth’). We propose the following four low-fidelity models:

(1) Low-fidelity model #1 (Output offset by a constant 90):

$$y_{L_1}(\mathbf{X}) = 100 - \sum_{i=1}^2 (X_i^2 - 5 \cos(2\pi X_i)) = y(\mathbf{X}) + 90, \quad (30)$$

(2) Low-fidelity model #2 (Magnification of the output by a factor of 10):

$$y_{L_2}(\mathbf{X}) = 100 - \sum_{i=1}^2 (10X_i^2 - 50 \cos(2\pi X_i)) = 10y(\mathbf{X}), \quad (31)$$

(3) Low-fidelity model #3 (Phases shifted by $\pi/2$):

$$y_{L_3}(\mathbf{X}) = 10 - \sum_{i=1}^2 (X_i^2 - 5 \cos(2\pi X_i + \pi/2)), \quad (32)$$

(4) Low-fidelity model #4 (Frequencies scaled by $1/2$):

$$y_{L_4}(\mathbf{X}) = 10 - \sum_{i=1}^2 (X_i^2 - 5 \cos(\pi X_i)). \quad (33)$$

The correlation coefficients ρ_{H,L_i} from (28) for $i = 1, 2, 3, 4$ between the high-fidelity model output y_H and the low-fidelity model output y_{L_i} are 1, 1, 0.72, and 0.72, respectively. These indicate that the low-fidelity models 1 and 2 have a stronger relationship with the high-fidelity model than the low-fidelity models 3 and 4. A perfect correlation of *one* implies that the low-fidelity models 1 and 2 capture the trends of the high-fidelity model output completely, even when the low-fidelity models 1 and 2 have significant bias. We use the above four surrogates for MFIS to study how correlation and accuracy of the low-fidelity models relative to the high-fidelity model affect the estimator.

5.3.3. Results

Table 1 presents the sampling-based CVaR estimates for the Rastrigin function (29) when $\beta = 0.99$. We show MCS estimates that use the DD-GPCE-Kriging, DD-GPCE, and PCE-Kriging as well as the high-fidelity model for comparison. We also show the DD-GPCE-Kriging-based MFIS estimator. The DD-GPCE-Kriging, DD-GPCE, and PCE-Kriging surrogates (whether using first, second or third-order approximation) are calculated using the same number (300) of high-fidelity output samples. We consider two separate cases where the two input variables have correlation coefficients 0.9 and 0, respectively. We provide benchmark estimates obtained by standard MCS with 10,000 high-fidelity output samples.

When the inputs X_1 and X_2 are highly correlated with correlation coefficient 0.9, the univariate ($S = 1$) DD-GPCE-Kriging methods of first ($m = 1$) through third ($m = 3$) order approximations provide MCS-based CVaR estimates that are close to the standard MCS. Their MRD over $K = 50$ trials are around 0.20%. In contrast, the MCS-based CVaR estimates by univariate ($S = 1$) DD-GPCE or PCE-Kriging methods of the same order ($m = 1 - 3$) approximations are less accurate compared with the standard MCS solution. The accuracy of CVaR estimates by the DD-GPCE or PCE-Kriging method improves from 71.14% to 45.92% or from 3.93% to 3.90%, respectively, in MRD as the order of approximation increases from $m = 1$ to $m = 3$. For the highly nonlinear Rastrigin function, the MCS-based CVaR estimate that uses the DD-GPCE-Kriging surrogate is therefore significantly more accurate than the estimate by MCS that samples from the DD-GPCE and PCE-Kriging surrogates.

When the inputs X_1 and X_2 have zero correlation, the accuracy of the proposed DD-GPCE-Kriging-based MCS method is still better than the DD-GPCE method. However, the PCE-Kriging

Table 1: CVaR $_{\beta}$ estimates ($\beta = 0.99$) for the Rastrigin function (29) via the DD-GPCE-Kriging, DD-GPCE, and PCE-Kriging-based MCS methods, the DD-GPCE-Kriging-based MFIS, and standard MCS.

Methods	CVaR $_{\beta}$ estimate ^(a)	MRD ^(a) (%)	N-RMSD ^(a) (%)	Model evaluations		
				HF ^(b)	LF I ^(c)	LF II ^(d)
Correlation coefficient between X_1 and $X_2 = 0.9$						
DD-GPCE-Kriging-based MCS						
$S = 1, m = 1$	18.7363	0.2046	0.2667	300 ^(e)	-	-
$S = 1, m = 2$	18.7366	0.2042	0.2640	300 ^(e)	-	-
$S = 1, m = 3$	18.7366	0.2043	0.2641	300 ^(e)	-	-
DD-GPCE-based MCS						
$S = 1, m = 1$	5.4175	71.1383	71.8302	300 ^(e)	-	-
$S = 1, m = 2$	10.1197	46.0874	46.1142	300 ^(e)	-	-
$S = 1, m = 3$	10.1499	45.9263	45.9534	300 ^(e)	-	-
PCE-Kriging-based MCS						
$m = 1$	18.0600	3.9322	4.3663	300 ^(e)	-	-
$m = 2$	18.0648	3.9104	4.3472	300 ^(e)	-	-
$m = 3$	18.0676	3.8971	4.3317	300 ^(e)	-	-
DD-GPCE-Kriging-based MFIS						
HF	18.8607	0.8361	1.0674	300 ^(f)	-	10,000 ^(g)
LF						
LF model #1 (30)	18.8217	0.8119	1.0545	150 ^(h)	150 ⁽ⁱ⁾	10,000 ^(g)
LF model #2 (31)	18.8624	0.8538	1.0720	150 ^(h)	150 ⁽ⁱ⁾	10,000 ^(g)
LF model #3 (32)	16.0848	14.3078	14.5028	150 ^(h)	150 ⁽ⁱ⁾	10,000 ^(g)
LF model #4 (33)	18.2318	2.8697	3.0861	150 ^(h)	150 ⁽ⁱ⁾	10,000 ^(g)
Standard MCS (Benchmark)	18.7705	-	-	10,000	-	-
Correlation coefficient between X_1 and $X_2 = 0$						
DD-GPCE-Kriging-based MCS ($S = 1, m = 3$)						
	18.3036	0.4438	0.6065	300 ^(e)	-	-
DD-GPCE-based MCS ($S = 1, m = 3$)						
	10.0410	45.3839	45.4445	300 ^(e)	-	-
PCE-Kriging-based MCS ($m = 3$)						
	18.3007	0.4623	0.6245	300 ^(e)	-	-
Standard MCS (Benchmark)	18.3848	-	-	10,000	-	-

- a. The estimates are averaged over $K = 50$ trials.
- b. The high-fidelity output is obtained by the Rastrigin function (29).
- c. The low-fidelity output is obtained by either of (30)–(33).
- d. The low-fidelity output is obtained by DD-GPCE-Kriging.
- e. The high-fidelity output samples are used to calculate DD-GPCE-Kriging, DD-GPCE, or PCE-Kriging.
- f. Two different 50% high-fidelity output samples are used to estimate the CI-based ϵ -risk region and CVaR $_{\beta}$, respectively.
- g. The low-fidelity output samples are used to estimate the CI-based ϵ risk region.
- h. The high-fidelity output samples are used to estimate the CVaR $_{\beta}$.
- i. The low-fidelity output samples are used to estimate DD-GPCE-Kriging.

method performs similar to DD-GPCE-Kriging, showing only a 0.02% difference between their MRD values. The N-RMSD values of DD-GPCE-Kriging and PCE-Kriging methods are both around 0.6%, indicating almost the same standard deviation of both estimators relative to the benchmark. This is because both DD-GPCE-Kriging and PCE-Kriging use the same orthonormal basis functions that are measure-consistent with independent random inputs. However, for dependent inputs, only DD-GPCE-Kriging can use such basis functions without any transformation from dependent to independent random inputs. Therefore, the DD-GPCE-Kriging outperforms PCE-Kriging in prediction accuracy when the random input variables are dependent. We also conclude that the additional variance term introduced in DD-GPCE-Kriging can better capture the local oscillations in the Rastrigin function compared to the more global approximation that is DD-GPCE.

We next present results for the proposed MFIS-based method. Recall, that option ‘‘HF’’ uses the high-fidelity model to learn the DD-GPCE-Kriging surrogate, and option ‘‘LF’’ uses the low-fidelity model to learn the DD-GPCE-Kriging surrogate. In both cases, those surrogates are only used to learn the biasing density for importance sampling. The univariate ($S = 1$), third-order ($m = 3$) DD-GPCE-Kriging was chosen to determine the CI-based ϵ -risk region for all MFIS cases tabulated in the second through sixth rows from the bottom of Table 1. The proposed MFIS-based CVaR estimate is 18.42 with HF, using 300 high-fidelity output samples. That estimate is close to the benchmark estimate of 18.38 by standard MCS. However, its MRD value of 0.80% is slightly higher than the 0.36–0.37% computed for the DD-GPCE-Kriging methods that also use 300 high-fidelity output samples. To speed up the process of learning the biasing density for MFIS, we consider the LF option. Specifically, in this example, we are interested in the accuracy of the MFIS-based CVaR estimates by LF when each of the four distinct surrogate models (30)–(33) is used. The LF option with low-fidelity models #1 and #2 (ρ_{H,L_1} and $\rho_{H,L_2} = 1$) yield a similar accuracy to the HF option. Contrarily, the LF option with low-fidelity models #3 and #4, for which ρ_{H,L_3} and $\rho_{H,L_4} = 0.72$, provides less accurate CVaR estimates. The results indicate that for the proposed MFIS-based method, it is acceptable for the low-fidelity models (y_{L_i} , $i=1-4$) to have a bias (which low-fidelity models #1 and #2 certainty do) as long as they have a strong correlation with the high-fidelity output y_H .

5.4. Example 2: Modified cross-in-tray function

This second example illustrates the DD-GPCE-Kriging-based MCS and MFIS methods for a non-smooth analytical mathematical function.

5.4.1. Problem definition

Define a bivariate Gaussian random vector $\mathbf{X} = (X_1, X_2)^\top$ with mean vector $(0, 0)^\top$ and standard deviations $\sqrt{\text{var}[X_i]} = 2$, $i = 1, 2$. The correlation coefficient between X_1 and X_2 is 0.9. For the input vector \mathbf{X} , consider the modified cross-in-tray function:

$$y(X_1, X_2) = -0.001 \left(\left| \sin(X_1) \sin(X_2) \exp \left(\left| 100 - \sqrt{\frac{X_1^2 + X_2^2}{\pi}} \right| \right) \right| + 1 \right)^{0.1}. \quad (34)$$

Figure 2b shows the modified cross-in-tray function on the domain $[-4, 4] \times [-4, 4]$. The function has two ridges along the $x_1 = 0$ and $x_2 = 0$ axes, respectively. It sharply drops on each side of the ridges, forming sharp corners and exhibiting oscillatory behavior. The sharp cross is not differentiable. The function is notoriously difficult to be approximate with polynomial bases due to its nonsmoothness.

5.4.2. Results

Table 2 presents the sampling-based CVaR estimates of the modified cross-in-tray function (34) when $\beta = 0.99$. We show MCS estimates that use the DD-GPCE-Kriging, DD-GPCE, and PCE-Kriging; we provide a benchmark estimate obtained by standard MCS of the high-fidelity model with 10,000 samples for comparison. We also show the DD-GPCE-Kriging-based MFIS estimator.

Table 2: CVaR $_{\beta}$ estimates ($\beta = 0.99$) of the modified cross-in-tray function (34) via the DD-GPCE-Kriging, DD-GPCE, and PCE-Kriging-based MCS methods, the DD-GPCE-Kriging-based MFIS method, and standard MCS.

Methods	CVaR $_{\beta}$ estimate ^(a)	MRD ^(a) (%)	N-RMSD ^(a) (%)	Model evaluations	
				HF ^(b)	LF ^(c)
DD-GPCE-Kriging-based MCS					
$S = 1, m = 3$	-10.4649	6.1486	6.8785	400 ^(d)	-
$S = 1, m = 4$	-10.4040	5.6964	6.4619	400 ^(d)	-
$S = 1, m = 5$	-10.3035	4.9523	5.7967	400 ^(d)	-
DD-GPCE-based MCS					
$S = 1, m = 3$	-13.4384	35.9874	36.7051	400 ^(d)	-
$S = 1, m = 4$	-14.2902	44.6070	45.5094	400 ^(d)	-
$S = 1, m = 5$	-13.9834	41.5772	43.5414	400 ^(d)	-
PCE-Kriging-based MCS					
$m = 3$	-11.1909	13.2435	14.7770	400 ^(d)	-
$m = 4$	-10.8560	12.6933	14.8892	400 ^(d)	-
$m = 5$	-10.1499	12.4506	19.6315	400 ^(d)	-
DD-GPCE-Kriging-based MFIS (HF)					
$m = 3$	-10.2050	3.3645	3.9706	400 ^(e)	10,000 ^(f)
$m = 4$	-10.1884	3.1556	3.6153	400 ^(e)	10,000 ^(f)
$m = 5$	-10.2141	3.3589	3.7319	400 ^(e)	10,000 ^(f)
Standard MCS (Benchmark)	-9.8821	-	-	10,000	-

a. The estimates are averaged over $K = 50$ trials.

b. The high-fidelity output is obtained by the modified cross-in-tray function (34).

c. The low-fidelity output is obtained by the DD-GPCE-Kriging.

d. The high-fidelity output samples are used to calculate DD-GPCE-Kriging, DD-GPCE, or PCE-Kriging.

e. Two different 50% high-fidelity output samples are used to estimate the CI-based ϵ -risk region and CVaR $_{\beta}$, respectively.

f. The low-fidelity output samples are used to estimate the CI-based ϵ -risk region.

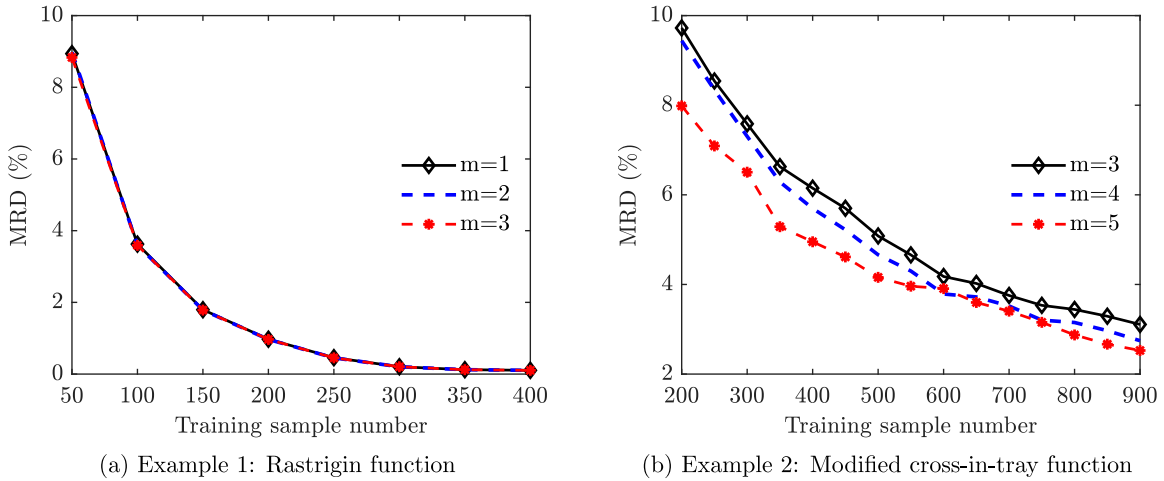


Figure 3: MRD (%) versus training data size when estimating CVaR by Monte Carlo sampling of the DD-GPCE-Kriging surrogates. Various degrees (m) of the polynomial approximation are illustrated.

We evaluate the accuracy of the DD-GPCE-Kriging-based MCS and MFIS methods when the quantity of interest includes nonsmooth features, as shown in Figure 2b. While the DD-GPCE-Kriging-based MCS improves the accuracy of CVaR estimation as the order increases from $m = 3$ to $m = 5$, the MFIS-based CVaR estimates with the HF option are the most accurate with 3.16–3.36% MRD and are relatively insensitive to the order (m) increase of the DD-GPCE-Kriging. This is because for the CVaR estimate, MFIS uses high-fidelity output samples drawn from the biasing density. Here, we only use a small number of 200 high-fidelity output samples for CVaR estimation to demonstrate its efficiency. While the DD-GPCE-Kriging surrogate is somewhat limited in predicting the nonsmooth output, the surrogate is only used to learn the biasing density. The standard deviation of the MFIS estimators relative to the benchmark is also the smallest compared to the MCS-based methods, in the range of 3.6–3.97% N-RMSD. In sum, for the nonsmooth modified cross-in-tray function, the proposed MFIS-based method is more accurate in providing CVaR estimates than DD-GPCE-Kriging-based MCS.

5.4.3. Convergence properties of the DD-GPCE-Kriging method

We examine the convergence of the DD-GPCE-Kriging-based Monte Carlo estimator with respect to the sample size that is used for training the DD-GPCE-Kriging surrogate. We illustrate those on Examples 1 and 2. Figures 3(a) and (b) show MRD versus training sample size when the CVaR estimates are obtained by sampling directly from the DD-GPCE-Kriging models. The MRD is computed relative to the MCS benchmark in Example 1 and 2. In Example 1, the MRD decreases from 8.8% to 0.1% as the training sample number increases from a small number of 50 to a modest 400. The behavior is almost identical for the three DD-GPCE-Kriging surrogates with different order approximations ($m = 1 - 3$). The figure illustrates that the DD-GPCE-Kriging method is—as expected in such high-dimensional problems—less accurate when a small number of samples are used to train the surrogate. When considering the results for Example 2 in Figure 3(b), we observe a similar monotone decrease in MRD as the sample number increases. This time, however, there is also a noticeable improvement from the $m = 3$ to $m = 4$ and $m = 5$ DD-GPCE-Kriging approximations. In both cases of heavily nonlinear and/or nonsmooth functions, and for all m , the DD-GPCE-Kriging method becomes more accurate as the training sample number increases.

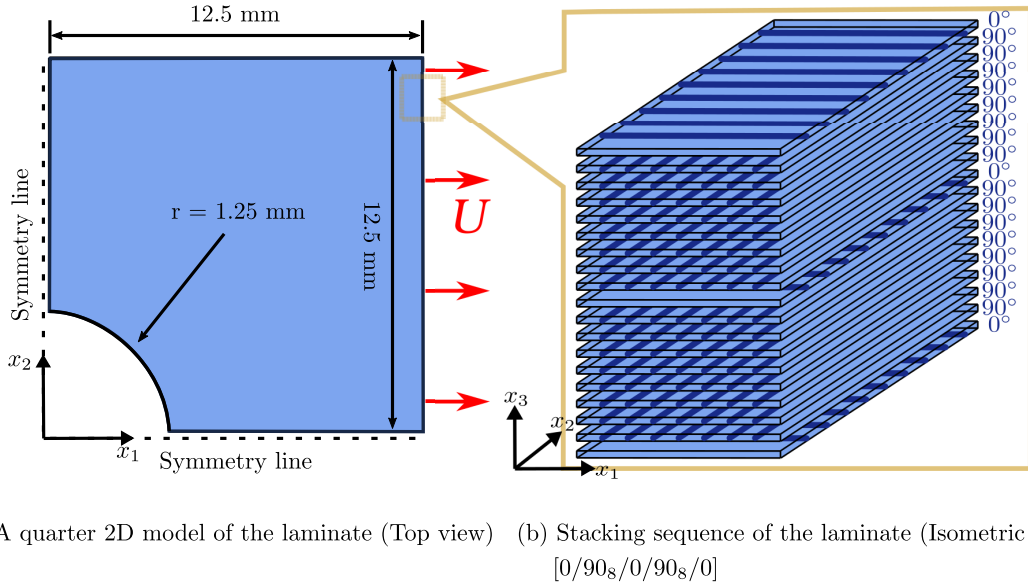


Figure 4: Geometry, loading, and boundary conditions of a glass/vinylester composite plate: We consider one quarter of the full plate domain, which represents the original square plate with domain $\mathcal{D} = (25 \text{ mm} \times 25 \text{ mm})$ including a circular hole of radius $r = 1.25 \text{ mm}$ in the center due to the symmetry conditions in the x_1 and x_2 directions of the center lines.

5.5. Example 3: A 2D glass/vinylester composite plate

This example studies a bolted or riveted composite laminate, which is widely used in aircraft structures. We are interested in measuring the risk of failure that occurs by stress concentration near rivet holes. The quantity of interest is obtained by solving a nonlinear quasi-static two-dimensional FEA. The model has a high number ($N = 28$) of random inputs, some of which are modeled as dependent random variables. This provides a challenging case for risk assessment.

5.5.1. Problem description

Figure 4a presents a quarter model of a two-dimensional square plate (spatial domain $\mathcal{D} = 25 \text{ mm} \times 25 \text{ mm}$) with a circular hole of radius $r = 1.25 \text{ mm}$ in the center. Due to symmetry conditions with respect to x_1 and x_2 , we can consider the quarter model in the domain $\bar{\mathcal{D}} = (12.5 \text{ mm} \times 12.5 \text{ mm})$ to save computational cost. Figure 4b shows the arrangement of the quarter model made of the glass/vinylester laminate (Fiberite/HyE 9082Af) with 19 stacking sequences [0/90_s/0/90_s/0], where ‘0’ indicates a ply having fibers in x_1 direction and 90_s denotes that eight plies have fibers in x_2 direction. The plate is subjected to a uniaxial tensile displacement loading U (mm/s) that acts on the entire right side edge. This example can also be found in [26, 17]. The local spatial coordinates are x_1 , x_2 , and x_3 , and we define E_i , ν_{ij} , and G_{ij} to be Young’s modulus, Poisson’s ratio, and shear modulus of the plies in the corresponding local spatial coordinates for $i, j = 1, 2, 3$. Let S_{ti} , S_{ci} , and S_{sij} be the tensile, compressive, and shear strengths of the plies. The measured material properties and ply thicknesses of the laminate vary so we model them as $N = 28$ random variables, as presented in Table 3. The 19 random variables that model the ply thickness are modeled as correlated via a multivariate lognormal distribution with a correlation coefficient of 0.5. The remaining 9 random variables are modeled as independent and distributed uniformly.

5.5.2. Quantity of interest

The high-fidelity model is a finite element model with a fine mesh, and the low-fidelity model is a finite element model with a coarse mesh, see Figure 5a. For both models, the element types were chosen as S4R, a 4-node, quadrilateral, stress/displacement shell element with reduced integration,

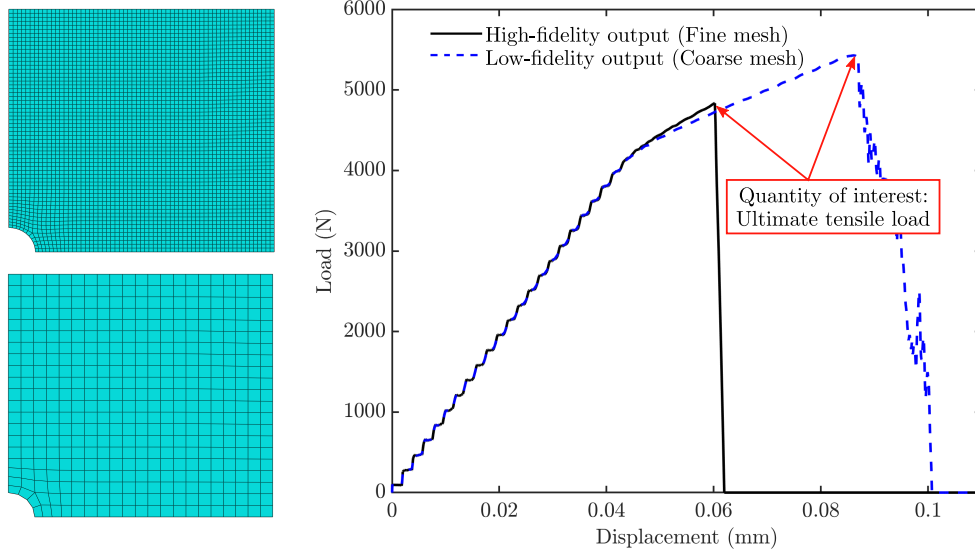
Table 3: Properties of the random inputs in Fiberite/HyE 9082Af

Random variable	Property	Mean	Coefficient of variation ($\text{COV}_i^{(a)}$, %)	Lower boundary	Upper boundary	Probability distribution
X_1	E_1 (MPa)	44,700	11.55	35760	53640	Uniform
X_2	E_2 (MPa)	12,700	11.55	10,160	15,240	Uniform
X_3	ν_{12}	0.297	11.55	0.238	0.356	Uniform
X_4	G_{12} (MPa)	5,800	11.55	4,640	6,960	Uniform
X_5	S_{t1} (MPa)	1,020	11.55	816	1,224	Uniform
X_6	S_{t2} (MPa)	40	11.55	32	48	Uniform
X_7	S_{c1} (MPa)	620	11.55	496	744	Uniform
X_8	S_{c2} (MPa)	140	11.55	126	168	Uniform
X_9	$S_{s12}^{(b)}$ (MPa)	60	11.55	48	72	Uniform
$X_{10}-X_{28}^{(c)}$	Plies 1-19 thicknesses (mm)	0.144	6	0	∞	Multivariate Lognormal

a. $\text{COV}_i = 100 \times \sqrt{\text{var}[X_i]/\mathbb{E}[X_i]}$, $i = 1, \dots, 28$.

b. $S_{s12} = S_{s23}$.

c. Correlation coefficients among $X_{10}-X_{28}$ are 0.5.



(a) Fine mesh (Top) and coarse mesh (Bottom)

(b) Load-displacement curve

Figure 5: Comparison of meshes and the load-displacement curve between high-fidelity and low-fidelity models: Figure (a) presents the fine mesh with 3,887 elements (Top) and the coarse mesh with 441 elements (Bottom). The fine and coarse mesh models generate high-fidelity and low-fidelity output data. In plot (b), high-fidelity and low-fidelity outputs in the load-displacement curve have different maximum values (ultimate tensile load).

Table 4: CVaR $_{\beta}$ estimates ($\beta = 0.99$) of the ultimate tensile load of the glass/vinylester composite plate via the DD-GPCE-Kriging, DD-GPCE, and PCE-Kriging-based MCS methods, DD-GPCE-Kriging-based MFIS, and standard MCS.

Methods	Ultimate tensile load (N)			Model evaluations			CPU time (hours) ^(e)
	CVaR $_{\beta}$ estimate ^(a)	MRD (%) ^(a)	N-RMSD ^(a) (%)	HF ^(b)	LF I ^(c)	LF II ^(d)	
DD-GPCE-Kriging-based MCS							
$S = 1, m = 1$	5558.0651	0.4383	0.4985	250 ^(f)	-	-	21.5
$S = 1, m = 2$	5556.7570	0.4861	0.6189	250 ^(f)	-	-	21.5
$S = 1, m = 3$	5556.7570	0.4861	0.5815	250 ^(f)	-	-	21.5
DD-GPCE-based MCS							
$S = 1, m = 1$	5529.3800	0.9381	0.9570	250 ^(f)	-	-	21.5
$S = 1, m = 2$	5533.2096	0.8743	0.9164	250 ^(f)	-	-	21.5
$S = 1, m = 3$	5542.4489	0.7334	0.8220	250 ^(f)	-	-	21.5
DD-GPCE-Kriging-based MFIS							
HF	5577.0842	0.5441	0.6784	250 ^(g)	-	10,000 ^(h)	21.5
LF	5577.1421	1.1459	1.6112	50 ⁽ⁱ⁾	200 ^(j)	10,000 ^(h)	8.2
Standard MCS (Benchmark)	5582.1484	-	-	10,000	-	-	859.8

- a. The estimates are averaged over $K = 30$ trials.
- b. The high-fidelity output is obtained by the fine mesh model in Figure 5a (Top).
- c. The low-fidelity output is obtained by the coarse mesh model in Figure 5a (Bottom).
- d. The low-fidelity output is obtained by the DD-GPCE-Kriging.
- e. The total CPU time is (the number of FEA) \times (FEA runtime) in each trial.
- f. The high-fidelity output samples are used to calculate the DD-GPCE-Kriging or DD-GPCE.
- g. The 250 samples are split as follows: 200 samples are used to compute DD-GPCE-Kriging and 50 samples are used to estimate CVaR $_{\beta}$.
- h. The low-fidelity output samples are used to estimate the CI-based ϵ -risk region.
- i. The high-fidelity output samples are used to estimate the CVaR $_{\beta}$.
- j. The low-fidelity output samples are used to estimate the DD-GPCE-Kriging.

built-in ABAQUS/Explicit, version 6.14-2. The total number of degrees of freedom for the high-fidelity and low-fidelity models are 24,084 and 2,910, respectively.

Figure 5b shows the load versus displacement curves of both the high-fidelity and low-fidelity models. Given the increasing applied displacement U in the range of 0–0.04 mm at the right edges of both models, the resultant tensile loads increase almost linearly in each case. After a displacement of about 0.04mm, the rate of change in both load curves decreases. The load values peak right before fracture. We then see a failure/drop right after the maximum tensile load is reached. We consider the peak load values, referred to as an ultimate tensile load, as the output of interest. The damage emanates from the hole and spreads outwards to the plate until it is completely fractured. To describe the damage, we use the Hashin damage model [10] built-in ABAQUS/Explicit, version 6.14-2, see [17] for a detailed explanation.

The correlation coefficient between the high-fidelity model $\rho_{H,L}$ and the low-fidelity model y_L is $\rho_{H,L} = 0.80$, see (28) for its computation. This indicates that the low-fidelity and high-fidelity model are strongly correlated. Thus, we use the low-fidelity model to compute the DD-GPCE-Kriging surrogate, which is then used to learn the biasing density for MFIS.

5.5.3. Results

Table 4 presents the CVaR estimates obtained from the DD-GPCE-Kriging and DD-GPCE-based MCS methods and DD-GPCE-Kriging-based MFIS methods with two options (Option “LF” trains the DD-GPCE-Kriging with the low-fidelity model and Option “HF” with the high-fidelity model). For comparison, we also provide a benchmark estimate obtained by the standard MCS with 10,000 samples in the last row of Table 4. The CVaR estimates are accurate with 0.43% through 1.15% in MRD over $K = 30$ trials. The CVaR estimates by the univariate ($S = 1$) DD-GPCE-Kriging methods of the first through third-order ($m = 1 - 3$) approximations are more precise (about two times lower MRD) than those by DD-GPCE when the same number (300) of high-fidelity output samples are used.

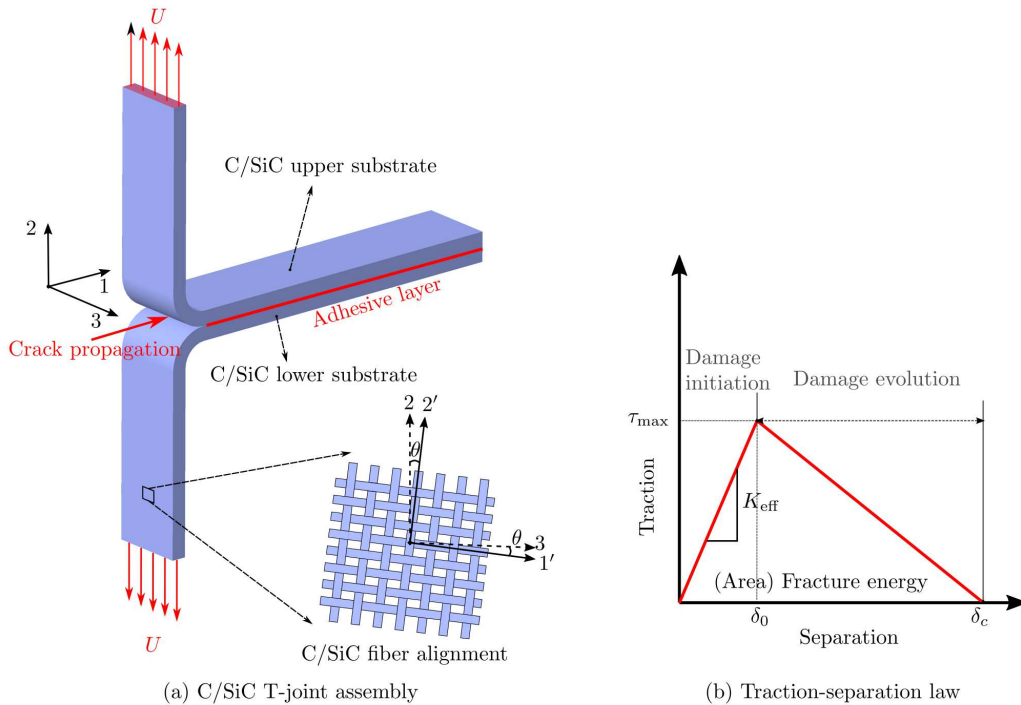


Figure 6: Geometry, loading, and boundary conditions of the 3D C/SiC composite T-joint assembly: (a) The two substrates consisting of interlaced C/SiC fibers are jointed by an adhesive layer; (b) The traction-separation law determines the damage of the adhesive layer under the peel loading U , leading to the deposition of the C/SiC T-joint. In Figure 6b, τ_{\max} , K_{eff} , and the fracture energies are all considered to be uncertain, see Table 5.

The MFIS with two options (HF and LF) exhibits good accuracy (0.54% and 1.15%, respectively in MRD) of the CVaR estimation. The MFIS with LF has an N-RMSD of 1.6%, which is relatively large compared to those (0.5%–0.6%) of MCS-based methods or the MFIS method with HF. The LF option requires a CPU time of 8.2 hours, which is an almost 62% reduction compared to those (21.5 hours) required by the HF option. Such computational gains are made feasible by using the cheaper low-fidelity surrogate for constructing the DD-GPCE-Kriging model.

5.6. Example 4: A 3D composite T-joint

Our final example considers a three-dimensional carbon fiber-reinforced silicon carbide (C/SiC) T-joint that requires a nonlinear quasi-static FEA to obtain the quantity of interest. This model has $N = 20$ random inputs, some of which are dependent. The complexity of the 3D FEA and the high-dimensional inputs make this a challenging case for risk analysis.

5.6.1. Problem description

Figure 6a shows a three-dimensional T-joint of interlaced C/SiC composite materials. It also illustrates the interlaced fiber alignments in the upper and lower substrates. Since the composite consists of high-strength carbon fibers and a high modulus, oxidation-resistant matrix, it exhibits high specific strength, low coefficient of thermal expansion, moderate thermal conductivity, and is lightweight. This makes this composite material appealing for super and hyper-sonic transport propulsion systems (e.g., space- or aircraft). The T-joint assembly allows the assembled components to avoid stress concentration occurring in the region near rivet holes and to achieve a lightweight assembled structure, see [46].

The T-joint structure is subjected to a uni-axial tensile displacement loading U that acts on both the top and bottom sides of the edges (known as a peel test). The peel test evaluates the performance of

Table 5: Properties of the random inputs for the C/SiC material and the adhesive layer.

Random variable	Property	Mean	Coefficient of variation ($\text{COV}_i^{(a)}$, %)	Lower boundary	Upper boundary	Probability distribution
Material properties of C/SiC composite						
X_1	E_{11} (MPa)	92,140	5	0	∞	Lognormal ^(b)
X_2	E_{22} (MPa)	92,140	5	0	∞	Lognormal ^(b)
X_3	E_{33} (MPa)	40,000	5	0	∞	Lognormal ^(b)
X_4	ν_{12}	0.01	5	0	∞	Lognormal ^(b)
X_5	ν_{13}	0.01	5	0	∞	Lognormal ^(b)
X_6	ν_{23}	0.01	5	0	∞	Lognormal ^(b)
X_7	G_{12} (MPa)	23,180	5	0	∞	Lognormal ^(b)
X_8	G_{13} (MPa)	23,180	5	0	∞	Lognormal ^(b)
X_9	G_{23} (MPa)	23,180	5	0	∞	Lognormal ^(b)
Cohesive parameters for the T-joint deposition						
X_{10}	K_{11} (MPa/mm)	679,800	10	611,820	747,780	Uniform
X_{11}	K_{22} (MPa/mm)	283,240	10	254,916	311,564	Uniform
X_{12}	K_{33} (MPa/mm)	283,240	10	254,916	311,564	Uniform
X_{13}	$\tau_{11,\max}$ (MPa)	9.5	10	8.55	10.45	Uniform
X_{14}	$\tau_{22,\max}$ (MPa)	8.8	10	7.92	9.68	Uniform
X_{15}	$\tau_{33,\max}$ (MPa)	8.8	10	7.92	9.68	Uniform
X_{16}	Normal fracture energy(mJ)	0.15	10	0.135	0.165	Uniform
X_{17}	1st shear fracture energy(mJ)	0.15	10	0.135	0.165	Uniform
X_{18}	2nd shear fracture energy(mJ)	0.15	10	0.135	0.165	Uniform
X_{19}	Fiber alignment in upper substrate(degree)	0	-(c)	-10	10	Uniform
X_{20}	Fiber alignment in lower substrate(degree)	0	-(c)	-10	10	Uniform

a. $\text{COV}_i = 100 \times \sqrt{\text{var}[X_i]/\mathbb{E}[X_i]}$, $i = 1, \dots, 20$.

b. Correlation coefficients among X_1-X_9 are 0.5.

c. COV_i for $i = 19, 20$ does not exist in its standard definition due to the *zero* mean.

interfacial adhesion in the T-joint assembly. We describe interfacial adhesion through cohesive surfaces with the traction-separation law illustrated in Figure 6b. In the graph of Figure 6b, the traction and separation represent the nominal stress and the crack opening displacement, respectively, which occur at the crack front line of the adhesive layer. The traction-separation model assumes initially linear elastic behavior of the adhesive layer. After the nominal traction stress approaches τ_{\max} , the layer damage initiates at the separation δ_0 . Then, the damage propagates until complete separation occurs at δ_c . Given τ_{\max} and K_{eff} , we determine the separation or crack opening displacement by the fracture energy (i.e., the integral of the traction over the separation in Figure 6b).

The measured material properties of C/SiC and adhesive parameters vary. Thus, we model them as $N = 20$ random variables X_i , $i = 1, \dots, 20$, following either multivariate lognormal or uniform distributions, see Table 5. For the material properties of the C/SiC composite, E_{ij} , ν_{ij} , and G_{ij} indicate elastic modulus, poisson ratio, and shear modulus, respectively, in the local j -direction on the surface normal to the local i -direction, $i, j = 1, 2, 3$. For the cohesive parameters associated with the T-joint deposition (detachment between two substrates), K_{ii} and $\tau_{ii,\max}$ indicate the stiffness of the adhesive layer in the local i -direction, $i = 1, 2, 3$, and the components of the maximum nominal traction τ_{\max} , see Figure 6b. The mean values of X_i , $i = 1, \dots, 20$, were obtained from [42].

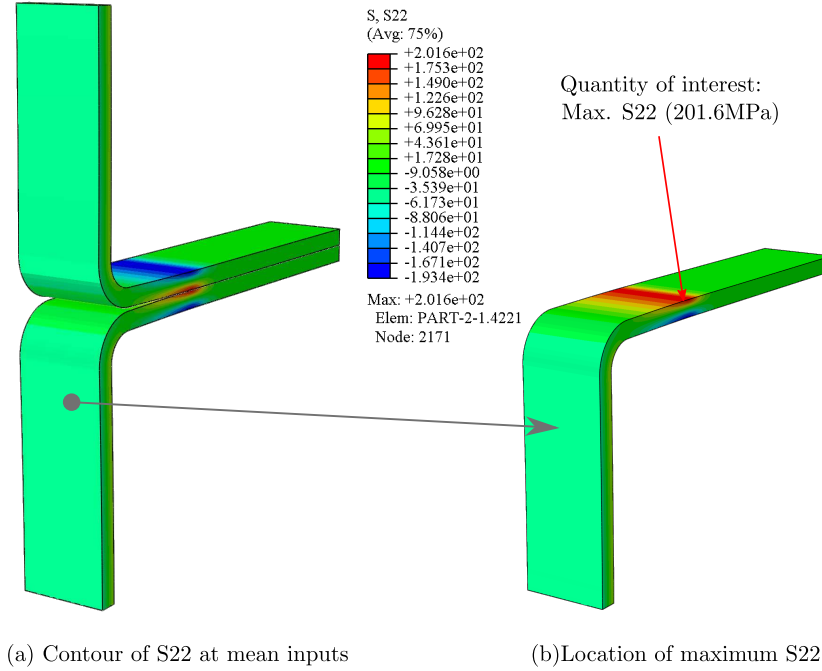


Figure 7: Deterministic FEA results for the C/SiC composite T-joint. Figures a and b show the stress in the 2-direction (S22) when the peeling displacement U is applied to the T-joint and when setting the inputs to their mean values.

5.6.2. Quantity of interest

The peel test determines the peeling behavior of the C/SiC joint (i.e., the crack initiation, propagation, and arrest on the interface layer between two C/SiC substrates). Figures 7a and 7b show the contour plots of the resultant stress (S22) in the 2-direction when the peeling displacement load U is applied to the T-joint. Figure 7b also shows the maximum value of S22. The location of the maximum value in the S22 contour can vary from sample to sample due to the randomness of the inputs \mathbf{X} . We consider the maximum value of S22 as the output of interest for CVaR estimation since the maximum S22 can be critical for the damage initiation of the C/SiC composite materials during the peeling behavior.

5.6.3. High-fidelity and low-fidelity model

The high-fidelity output is computed from a fine-meshed finite element model with 48,000 elements, and the low-fidelity model is based on a coarse mesh with only 3,000 elements. Those meshes are shown in Figures 8a and b, respectively. The chosen element type is an eight-node brick element with reduced integration (C3D8R), built in ABAQUS/Explicit-version 6.14-2. Thus, the degrees of freedom (DOF) of the fine and coarse mesh models are 180,648 and 15,048, respectively. We determined the mesh size of the high-fidelity model by a convergence study with respect to the output of interest. We selected the low-fidelity model (which is hence not converged) such that it stays below a given computational budget.

The correlation coefficient between the high-fidelity model y_H and the low-fidelity model y_L is $\rho_{H,L} = 0.55$, see (28) for its computation. This indicates that the low-fidelity and high-fidelity model are still fairly correlated. Thus, we use the low-fidelity model to compute the DD-GPCE-Kriging surrogate, which is then used to learn the biasing density for MFIS.

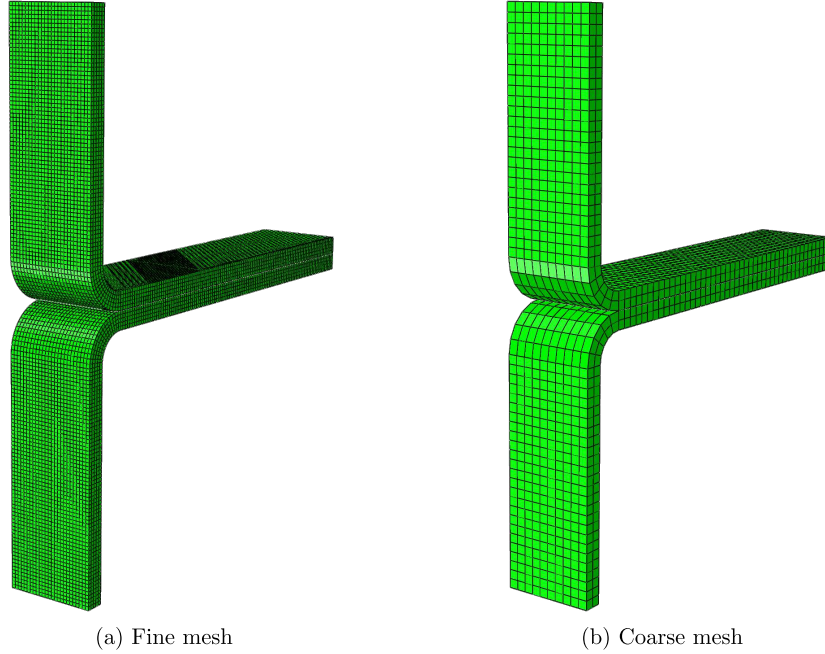


Figure 8: The finite element mesh for the C/SiC T-joint. The fine mesh in (a) with 48,000 elements is used to generate high-fidelity output data, while the coarse mesh in (b) with 3,000 elements is used to generate low-fidelity output data.

Table 6: CVaR_β estimates ($\beta = 0.99$) of the maximum value of S22 in the C/SiC T-joint via the DD-GPCE-Kriging and DD-GPCE-based MCS methods, DD-GPCE-Kriging-based MFIS, and standard MCS.

Methods	Maximum value of S22 (MPa)			Model evaluations			CPU time (hours) ^(e)
	CVaR_β estimates ^(a)	MRD ^(a) (%)	N-RMSD ^(a) (%)	HF ^(b)	LF I ^(c)	LF II ^(d)	
DD-GPCE-Kriging-based MCS							
$S = 1, m = 1$	242.6210	2.6970	2.8467	400 ^(f)	-	-	159.80
$S = 1, m = 2$	245.9401	2.2590	2.9568	400 ^(f)	-	-	159.80
$S = 1, m = 3$	249.2416	2.2392	3.0945	400 ^(f)	-	-	159.80
DD-GPCE-based MCS							
$S = 1, m = 1$	191.1512	22.5908	23.8108	400 ^(f)	-	-	159.80
$S = 1, m = 2$	208.8489	15.4653	15.7966	400 ^(f)	-	-	159.80
$S = 1, m = 3$	227.6264	8.4541	9.0896	400 ^(f)	-	-	159.80
DD-GPCE-Kriging-based MFIS							
HF	246.3827	0.6644	0.7156	400 ^(g)	-	6,000 ^(h)	159.80
LF	246.9583	0.9792	1.2357	150 ⁽ⁱ⁾	250 ^(j)	6,000 ^(h)	67.69
Standard MCS (Benchmark)	247.0036	-	-	6,000	-	-	1597.97

- a. The estimates are averaged over $K = 30$ independent trials.
- b. The high-fidelity output is obtained by the fine mesh model in Figure 8a.
- c. The low-fidelity output is obtained by the coarse mesh model in Figure 8b.
- d. The low-fidelity output is obtained by DD-GPCE-Kriging.
- e. The total CPU time is (the number of FEA) \times (FEA run-time) in each trial.
- f. The high-fidelity output samples are used to calculate DD-GPCE-Kriging or DD-GPCE.
- g. Two different 50% high-fidelity output samples are used to estimate the CI-based ϵ -risk region and CVaR_β , respectively.
- h. The low-fidelity output samples are used to estimate the CI-based ϵ -risk region.
- i. The high-fidelity output samples are used to estimate the CVaR_β .
- j. The low-fidelity output samples are used to calculate DD-GPCE-Kriging.

5.6.4. Results

Table 6 summarizes the CVaR estimates, when $\beta = 0.99$, by the DD-GPCE-Kriging and DD-GPCE-based MCS methods and DD-GPCE-Kriging-based MFIS with two options (Option “LF” trains the DD-GPCE-Kriging with the low-fidelity model and Option “HF” with the high-fidelity model). Compared to the benchmark CVaR estimate by the standard MCS of 6,000 high-fidelity output samples, the univariate ($S = 1$) DD-GPCE-Kriging methods of the first through third-order ($m = 1 - 3$) approximations have a MRD of 2.2–2.7%. While the MRD values of both DD-GPCE-Kriging and DD-GPCE decrease as the number of m increases, DD-GPCE-Kriging-based estimates are more accurate than those of the DD-GPCE method. It takes 159.80 CPU hours to obtain CVaR estimates by the DD-GPCE or the DD-GPCE-Kriging method, which is only 10% of the 1,597.97 CPU hours needed for the standard MCS estimate.

The proposed MFIS-based method with HF (when the high-fidelity model is used for computing DD-GPCE-Kriging) yields the most accurate CVaR estimate, showing only 0.66% in MRD. Moreover, the MFIS method with HF has an N-RMSD of 0.7%, which is the lowest among all methods in Table 6. It requires the same CPU time (159.80 hours) as the DD-GPCE-Kriging or DD-GPCE methods. When the proposed MFIS-based method is integrated with LF, as expected, the required CPU time decreases to 67.69 hours, which is only 42% of the 159.80 hours by the HF option. This is because the DD-GPCE-Kriging is calculated from cheap low-fidelity model evaluations in the LF option. The DD-GPCE-Kriging-based MFIS with the LF option yields a CVaR estimate with 0.98% in MRD, which is more accurate when compared with the MCS strategy that uses DD-GPCE-Kriging to sample from. This example illustrates that the proposed MFIS-based method is not only accurate but also computationally efficient when the LF option is used.

5.7. Guidelines for choosing a CVaR estimation strategy

We summarize our findings that are supported by the numerical results in this paper. The intent is to provide guidelines for the use of surrogates in risk estimation. We categorize “efficiency” into low, medium, and high based on the computational cost that it takes to get accurate estimates. The standard MCS has low efficiency due to the prohibitive cost of evaluating the high-fidelity model.

Using the DD-GPCE-Kriging surrogate for Monte-Carlo simulation.

The method is appropriate for treating moderately to highly nonlinear and smooth output functions. However, it only has medium efficiency. The accuracy of CVaR estimates highly relies on the precision of the DD-GPCE-Kriging model. Thus, the MCS method is limited in providing accurate estimates for a nonsmooth output function.

Using the DD-GPCE-Kriging surrogate for multifidelity importance sampling.

- (a) Using high-fidelity evaluations to learn the surrogate (‘HF’). The method is appropriate for treating highly nonlinear and nonsmooth output functions, and its efficiency is medium. The MFIS method provides unbiased CVaR estimates obtained by high-fidelity output samples.
- (b) Using low-fidelity evaluations to learn the surrogate (‘LF’). The method is appropriate for treating highly nonlinear and nonsmooth output functions, and its efficiency is high. The LF option determines DD-GPCE-Kriging using low-fidelity output samples to further speed up the process of the MFIS.

Lastly, we note that both proposed methods outperformed the MCS-based CVaR estimation when the DD-GPCE model is used for sampling. Consequently, the proposed DD-GPCE-Kriging model is a superior surrogate for uncertainty quantification, and, specifically, risk estimation. The DD-GPCE-Kriging-based MFIS method is especially suited to highly nonlinear and nonsmooth output responses. Moreover, the proposed MFIS-based method can achieve higher efficiency when the DD-GPCE-Kriging surrogate is computed with the low-fidelity surrogate.

6. Conclusions

We presented novel computational methods for Conditional Value-at-Risk (CVaR) estimation for nonlinear systems under high-dimensional and dependent inputs. First, we proposed a new surrogate model, *DD-GPCE-Kriging*, as a fusion of DD-GPCE and Kriging to approximate a highly nonlinear and nonsmooth random output when the inputs are high-dimensional and dependent. We integrated DD-GPCE-Kriging with two sampling-based CVaR estimation methods: standard Monte Carlo Sampling (MCS) and Multifidelity Importance Sampling (MFIS). The proposed MCS-based method samples from the computationally efficient DD-GPCE-Kriging surrogate and is shown to be accurate in the presence of high-dimensional and dependent random inputs. However, the proposed method may produce a biased CVaR estimate as it relies on the approximation quality of the DD-GPCE-Kriging. Therefore, we leveraged the MFIS method and used the newly-constructed DD-GPCE-Kriging surrogate to determine the biasing density for importance sampling. The importance-sampling-based CVaR estimate uses the high-fidelity model only. To further speed up the MFIS process, we computed a second DD-GPCE-Kriging surrogate from lower-fidelity model simulations, which in all numerical results, still provided a good biasing density.

The numerical results for two mathematical functions with bivariate random inputs demonstrated that the proposed DD-GPCE-Kriging method is better suited to handle a highly nonlinear or nonsmooth output than the DD-GPCE method. The example also shows that the proposed MFIS-based method provides more accurate CVaR estimates than the MCS-based method when the quantity of interest is nonsmooth.

We evaluated the proposed CVaR estimation methods in aerospace and defense engineering applications focusing on composite structures. The scalability of the proposed methods and their applicability to complex engineering problems were demonstrated by solving the two-dimensional glass/vinylester laminate (lightweight and high stiffness) composite problem, including 28 (partly dependent) random inputs and the three-dimensional C/SiC (lightweight and heat resistant) composite T-joint problem, including 20 (partly dependent) random inputs. In the glass/vinylester composite laminate problem, the proposed MFIS-based method achieves a speedup of 104x compared to standard MCS using the high-fidelity model, while producing a CVaR estimate with 1.15% error.

Acknowledgment

Funding This material is based on research sponsored by the Air Force Research Lab (AFRL) and Defense Advanced Research Projects Agency (DARPA) under agreement number FA8650-21-2-7126. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

Disclaimer The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the AFRL and DARPA or the U.S. Government.

References

- [1] F. Bachoc. Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification. *Computational Statistics & Data Analysis*, 66:55–69, 2013.
- [2] F. Bernal, E. Gobet, and J. Printems. Volatility uncertainty quantification in a stochastic control problem applied to energy. *Methodology and Computing in Applied Probability*, 22(1):135–159, 2020.
- [3] R. Billinton and A. Jonnavithula. Composite system adequacy assessment using sequential monte carlo simulation with variance reduction techniques. *IEE Proceedings-Generation, Transmission and Distribution*, 144(1):1–6, 1997.

- [4] A. Chaudhuri, B. Kramer, M. Norton, J. O. Royset, and K. Willcox. Certifiable risk-based engineering design optimization. *AIAA Journal*, 60(2):551–565, 2022.
- [5] A. Chaudhuri, B. Peherstorfer, and K. Willcox. Multifidelity cross-entropy estimation of conditional value-at-risk for risk-averse design optimization. In *AIAA Scitech 2020 Forum*, page 2129, 2020.
- [6] X. Chen, B. L. Nelson, and K.-K. Kim. Stochastic Kriging for conditional value-at-risk and its sensitivities. In *Proceedings of the 2012 Winter Simulation Conference (WSC)*, pages 1–12. IEEE, 2012.
- [7] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*. Society for Industrial and Applied Mathematics, 2009.
- [8] N. Durrande, D. Ginsbourger, and O. Roustant. Additive covariance kernels for high-dimensional Gaussian process modeling. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 21, pages 481–499, 2012.
- [9] J.-y. Gotoh and S. Uryasev. Support vector machines based on convex risk functions and general norms. *Annals of Operations Research*, 249(1):301–328, 2017.
- [10] Z. Hashin. Failure criteria for unidirectional fiber composites. *Journal of Applied Mechanics*, 47(2):329–334, 1980.
- [11] M. Heinkenschloss, B. Kramer, and T. Takhtaganov. Adaptive reduced-order model construction for conditional value-at-risk estimation. *SIAM/ASA Journal on Uncertainty Quantification*, 8(2):668–692, 2020.
- [12] M. Heinkenschloss, B. Kramer, T. Takhtaganov, and K. Willcox. Conditional-value-at-risk estimation via reduced-order models. *SIAM/ASA Journal on Uncertainty Quantification*, 6(4):1395–1423, 2018.
- [13] J. D. Jakeman, F. Franzelin, A. Narayan, M. Eldred, and D. Pflüger. Polynomial chaos expansions for dependent random variables. *Computer Methods in Applied Mechanics and Engineering*, 351:643–666, 2019.
- [14] J. D. Jakeman, D. P. Kouri, and J. G. Huerta. Surrogate modeling for efficiently, accurately and conservatively estimating measures of risk. *Reliability Engineering & System Safety*, 221:108280, 2022.
- [15] D. P. Kouri and T. M. Surowiec. Risk-averse PDE-constrained optimization using the conditional value-at-risk. *SIAM Journal on Optimization*, 26(1):365–396, 2016.
- [16] D. P. Kouri and T. M. Surowiec. Epi-regularization of risk measures. *Mathematics of Operations Research*, 45(2):774–795, 2020.
- [17] D. Lee and B. Kramer. Bi-fidelity conditional-value-at-risk estimation by dimensionally decomposed generalized polynomial chaos expansion. *Structural and Multidisciplinary Optimization*, 66(2):33, 2023.
- [18] D. Lee and S. Rahman. Practical uncertainty quantification analysis involving statistically dependent random variables. *Applied Mathematical Modelling*, 84:324–356, 2020.
- [19] D. Lee and S. Rahman. Robust design optimization under dependent random variables by a generalized polynomial chaos expansion. *Structural and Multidisciplinary Optimization*, 63(5):2425–2457, 2021.

- [20] D. Lee and S. Rahman. Reliability-based design optimization under dependent random variables by a generalized polynomial chaos expansion. *Structural and Multidisciplinary Optimization*, 65(1), 2022.
- [21] D. Lee and S. Rahman. High-dimensional stochastic design optimization under dependent random variables by a dimensionally decomposed generalized polynomial chaos expansion. *International Journal for Uncertainty Quantification*, 13(4), 2023.
- [22] A. Mancuso, M. Compare, A. Salo, and E. Zio. Optimal prognostics and health management-driven inspection and maintenance strategies for industrial systems. *Reliability Engineering & System Safety*, 210:107536, 2021.
- [23] R. Mansini, W. Ogryczak, and M. G. Speranza. Conditional value at risk and related linear programming models for portfolio optimization. *Annals of Operations Research*, 152:227–256, 2007.
- [24] A. Marrel, B. Iooss, F. Van Dorpe, and E. Volkova. An efficient methodology for modeling complex computer codes with Gaussian processes. *Computational Statistics & Data Analysis*, 52(10):4731–4744, 2008.
- [25] MATLAB. *version 9.10.0.1613233 (R2021b)*. The MathWorks Inc., Natick, Massachusetts, 2021.
- [26] M. Moure, S. Sanchez-Saez, E. Barbero, and E. Barbero. Analysis of damage localization in composite laminates using a discrete damage model. *Composites Part B: Engineering*, 66:224–232, 2014.
- [27] M. Navarro, J. A. S. Witteveen, and J. G. Blom. Polynomial chaos expansion for general multivariate distributions with correlated variables. *arXiv preprint arXiv:1406.5483*, 2014.
- [28] Y. Noh, K. K. Choi, and L. Du. Reliability-based design optimization of problems with correlated input variables using a Gaussian copula. *Structural and Multidisciplinary Optimization*, 38:1–16, 2009.
- [29] B. Peherstorfer, T. Cui, Y. Marzouk, and K. Willcox. Multifidelity importance sampling. *Computer Methods in Applied Mechanics and Engineering*, 300:490–509, 2016.
- [30] B. Peherstorfer, K. Willcox, and M. Gunzburger. Optimal model management for multifidelity Monte Carlo estimation. *SIAM Journal on Scientific Computing*, 38(5):A3163–A3194, 2016.
- [31] S. Rahman. A polynomial chaos expansion in dependent random variables. *Journal of Mathematical Analysis and Applications*, 464(1):749–775, 2018.
- [32] R. Rocchetta and L. G. Crespo. A scenario optimization approach to reliability-based and risk-based design: Soft-constrained modulation of failure probability bounds. *Reliability Engineering & System Safety*, 216:107900, 2021.
- [33] R. T. Rockafellar and J. O. Royset. Engineering decisions under risk averseness. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 1(2):04015003, 2015.
- [34] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at risk. *Journal of Risk*, 3:21–41, 2000.
- [35] R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.
- [36] J. O. Royset, L. Bonfiglio, G. Vernengo, and S. Brizzolara. Risk-adaptive set-based design and applications to shaping a hydrofoil. *Journal of Mechanical Design*, 139(10):101403, 2017.

- [37] R. Schobi, B. Sudret, and J. Wiart. Polynomial-chaos-based Kriging. *International Journal for Uncertainty Quantification*, 5(2), 2015.
- [38] T. Soma and Y. Yoshida. Statistical learning with conditional value at risk. *arXiv preprint arXiv:2002.05826*, 2020.
- [39] R. Srinivasan. *Importance sampling: Applications in communications and detection*. Springer Science & Business Media, 2002.
- [40] O. Stover, P. Karve, and S. Mahadevan. Reliability and risk metrics to assess operational adequacy and flexibility of power grids. *Reliability Engineering & System Safety*, 231:109018, 2023.
- [41] A. Tabandeh, G. Jia, and P. Gardoni. A review and assessment of importance sampling methods for reliability analysis. *Structural Safety*, 97:102216, 2022.
- [42] Y. Wang, X. Wang, Z. Guan, J. Xu, and X. Guo. Experimental and numerical investigation on c/sic composite z-pinned/bonded hybrid single-lap joints. *Materials*, 14(5):1130, 2021.
- [43] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, 1938.
- [44] J. Wu and P. Wang. Risk-averse optimization for resilience enhancement of complex engineering systems under uncertainties. *Reliability Engineering & System Safety*, 215:107836, 2021.
- [45] D. Xiu and G. E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal of Scientific Computing*, 24:619–644, 2002.
- [46] S. Yan, X. Zeng, and A. Long. Experimental assessment of the mechanical behaviour of 3D woven composite T-joints. *Composites Part B: Engineering*, 154:108–113, 2018.
- [47] H. Yang, Y. Fujii, K. Wang, and A. A. Gorodetsky. Control variate polynomial chaos: Optimal fusion of sampling and surrogates for multifidelity uncertainty quantification. *arXiv preprint arXiv:2201.10745*, 2022.
- [48] H. Yang and M. Gunzburger. Algorithms and analyses for stochastic optimization for turbofan noise reduction using parallel reduced-order modeling. *Computer Methods in Applied Mechanics and Engineering*, 319:217–239, 2017.
- [49] Z. Zou, D. Kouri, and W. Aquino. An adaptive local reduced basis method for solving PDEs with uncertain inputs and evaluating risk. *Computer Methods in Applied Mechanics and Engineering*, 345:302–322, 2019.

Appendix A Generalized polynomial chaos expansion

When $\mathbf{X} = (X_1, \dots, X_N)^\top$ consists of statistically dependent random variables, the resultant probability measure is in general not of a product type, meaning that the joint distribution of \mathbf{X} cannot be obtained strictly from its marginal distributions. Consequently, measure-consistent multivariate orthonormal polynomials in $\mathbf{x} = (x_1, \dots, x_N)^\top$ cannot be built from an N -dimensional tensor product of measure-consistent univariate orthonormal polynomials. In this case, a three-step algorithm based on a whitening transformation of the monomial basis can be used to determine multivariate orthonormal polynomials consistent with an arbitrary, non-product-type probability measure $f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$ of \mathbf{X} , see [18].

Let $\mathbf{j} := (j_1, \dots, j_N) \in \mathbb{N}_0^N$ be an N -dimensional multi-index. For a realization $\mathbf{x} = (x_1, \dots, x_N)^\top \in \mathbb{A}^N \subseteq \mathbb{R}^N$ of \mathbf{X} , a monomial in the real variables x_1, \dots, x_N is the product $\mathbf{x}^{\mathbf{j}} = x_1^{j_1} \dots x_N^{j_N}$ with a total degree $|\mathbf{j}| = j_1 + \dots + j_N$. Consider for each $m \in \mathbb{N}_0$ the elements of the multi-index set

$$\mathcal{J}_m := \{\mathbf{j} \in \mathbb{N}_0^N : |\mathbf{j}| \leq m\},$$

which is arranged as $\mathbf{j}^{(1)}, \dots, \mathbf{j}^{(L_{N,m})}$, $\mathbf{j}^{(1)} = \mathbf{0}$, according to a monomial order of choice. The set \mathcal{J}_m has cardinality $L_{N,m}$ obtained as

$$L_{N,m} := |\mathcal{J}_m| = \sum_{l=0}^m \binom{N+l-1}{l} = \binom{N+m}{m}. \quad (35)$$

Let us denote by

$$\mathbf{\Psi}_m(\mathbf{x}) = (\Psi_1(\mathbf{x}), \dots, \Psi_{L_{N,m}}(\mathbf{x}))^\top \quad (36)$$

an $L_{N,m}$ -dimensional vector of multivariate orthonormal polynomials that is consistent with the probability measure $f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$ of \mathbf{X} . Consequently, any output random variable $y(\mathbf{X}) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ can be approximated by the m th-order GPCE¹

$$y_m(\mathbf{X}) = \sum_{i=1}^{L_{N,m}} c_i \Psi_i(\mathbf{X}) \quad (37)$$

of $y(\mathbf{X})$, comprising $L_{N,m}$ basis functions with expansion coefficients

$$c_i := \int_{\mathbb{A}^N} y(\mathbf{x}) \Psi_i(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad i = 1, \dots, L_{N,m}. \quad (38)$$

Here, the orthonormal polynomials $\Psi_i(\mathbf{X})$, $i = 1, \dots, L_{N,m}$, are determined numerically via three steps, which we implement by replacing $\mathcal{J}_{N,S,m}$ with \mathcal{J}_m , see the following Appendix B. We refer to [18] for more details. In this work, we refer to the GPCE in (37) as regular GPCE to distinguish it from the DD-GPCE.

Appendix B Three steps for measure-consistent orthonormal polynomials for DD-GPCE

For $N \in \mathbb{N}$, denote by $\{1, \dots, N\}$ an index set and $\mathcal{U} \subseteq \{1, \dots, N\}$ a subset (including the empty set \emptyset) with cardinality $0 \leq |\mathcal{U}| \leq N$. The complementary subset of \mathcal{U} is denoted by $\mathcal{U}^c := \{1, \dots, N\} \setminus \mathcal{U}$. For each $m \in \mathbb{N}_0$ and $0 \leq S \leq N$, we define the reduced multi-index set

$$\mathcal{J}_{S,m} := \left\{ \mathbf{j} = (\mathbf{j}_{\mathcal{U}}, \mathbf{0}_{\mathcal{U}^c}) \in \mathbb{N}_0^N : \mathbf{j}_{\mathcal{U}} \in \mathbb{N}^{|\mathcal{U}|}, |\mathcal{U}| \leq |\mathbf{j}_{\mathcal{U}}| \leq m, 0 \leq |\mathcal{U}| \leq S \right\}, \quad (39)$$

which is arranged as $\mathbf{j}^{(1)}, \dots, \mathbf{j}^{(L_{N,S,m})}$, $\mathbf{j}^{(1)} = \mathbf{0}$, according to a monomial order of choice and where $|\mathbf{j}_{\mathcal{U}}| := j_{i_1} + \dots + j_{i_{|\mathcal{U}|}}$. Here, $(\mathbf{j}_{\mathcal{U}}, \mathbf{0}_{\mathcal{U}^c})$ denotes an N -dimensional multi-index whose i th component is j_i if $i \in \mathcal{U}$ and 0 if $i \notin \mathcal{U}$. The set $\mathcal{J}_{S,m}$ represents a subset of \mathcal{J}_m determined from the chosen S , where only at most S -variate basis functions are preserved, that are relevant for the S th-variate DD-GPCE approximation. As a result, we have that

$$L_{N,S,m} \leq L_{N,m},$$

i.e., the DD-GPCE never has more terms than the regular GPCE; in most cases, it will have significantly less terms. For $\mathbf{x} = (x_1, \dots, x_N)^\top \in \mathbb{A}^N \subseteq \mathbb{R}^N$ we then define the basis vector for the DD-GPCE as

$$\mathbf{\Psi}_{S,m}(\mathbf{x}) := (\Psi_i(\mathbf{x}), \dots, \Psi_{L_{N,S,m}}(\mathbf{x}))^\top,$$

¹The GPCE in (37) should not be confused with that of [45]. The GPCE, presented here, is meant for an arbitrary dependent probability distribution of random input. In contrast, the existing PCE, whether classical [43] or generalized [45], still requires independent random inputs.

which is an $L_{N,S,m}$ -dimensional vector of multivariate orthonormal polynomials that is consistent with the probability measure $f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$ of \mathbf{x} . The orthonormal polynomials are determined by the following three steps.

Step 1. Given $0 \leq S \leq N$ and $S \leq m < \infty$, create an $L_{N,S,m}$ -dimensional column vector

$$\mathbf{M}_{S,m}(\mathbf{x}) = (\mathbf{x}^{\mathbf{j}^{(1)}}, \dots, \mathbf{x}^{\mathbf{j}^{(L_{N,S,m})}})^{\top} \quad (40)$$

of monomials whose elements are the monomials $\mathbf{x}^{\mathbf{j}}$ for $\mathbf{j} \in \mathcal{J}_{S,m}$ arranged in the aforementioned order. For $\mathcal{U} \subseteq \{1, \dots, N\}$, let $\mathbf{x}_{\mathcal{U}} := (x_{i_1}, \dots, x_{i_{|\mathcal{U}|}})^{\top}$, $1 \leq i_1 < \dots < i_{|\mathcal{U}|} \leq N$, be a subvector of \mathbf{x} . The complementary subvector is defined by $\mathbf{x}_{\mathcal{U}^c} := \mathbf{x}_{\{1, \dots, N\} \setminus \mathcal{U}}$. Then, for $\mathbf{j} \in \mathcal{J}_{S,m}$,

$$\mathbf{x}^{\mathbf{j}} = \mathbf{x}_{\mathcal{U}}^{\mathbf{j}_{\mathcal{U}}} \mathbf{0}_{\mathcal{U}^c}^{\mathbf{j}_{\mathcal{U}^c}} = \mathbf{x}_{\mathcal{U}}^{\mathbf{j}_{\mathcal{U}}}.$$

Hence, $\mathbf{M}_{S,m}(\mathbf{x})$ is the monomial vector in $\mathbf{x}_{\mathcal{U}} = (x_{i_1}, \dots, x_{i_{|\mathcal{U}|}})^{\top}$ of degree $0 \leq |\mathcal{U}| \leq S$ and $|\mathcal{U}| \leq |\mathbf{j}_{\mathcal{U}}| \leq m$.

Step 2. Construct an $L_{N,S,m} \times L_{N,S,m}$ monomial moment matrix of $\mathbf{M}_{S,m}(\mathbf{X})$, defined as

$$\begin{aligned} \mathbf{G}_{S,m} &:= \mathbb{E}[\mathbf{M}_{S,m}(\mathbf{X})\mathbf{M}_{S,m}^{\top}(\mathbf{X})] \\ &:= \int_{\mathbb{A}^N} \mathbf{M}_{S,m}(\mathbf{x})\mathbf{M}_{S,m}^{\top}(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}. \end{aligned} \quad (41)$$

For an arbitrary probability density $f_{\mathbf{X}}(\mathbf{x})$, the matrix $\mathbf{G}_{S,m}$ cannot be determined exactly, yet it can be accurately estimated with numerical integration and/or sampling methods [18].

Step 3. Select the $L_{N,S,m} \times L_{N,S,m}$ whitening matrix $\mathbf{W}_{S,m}$ from the Cholesky decomposition of the symmetric positive-definite monomial moment matrix $\mathbf{G}_{S,m}$ [31], leading to

$$\mathbf{G}_{S,m} = \mathbf{W}_{S,m}^{-1} \mathbf{W}_{S,m}^{-\top}. \quad (42)$$

The whitening transformation is then used to generate multivariate orthonormal polynomials as follows:

$$\mathbf{\Psi}_{S,m}(\mathbf{x}) = \mathbf{W}_{S,m} \mathbf{M}_{S,m}(\mathbf{x}). \quad (43)$$

Appendix C Leave-one-out cross validation estimate

From the known distribution of random inputs \mathbf{X} and an output function $y : \mathbb{A}^N \rightarrow \mathbb{R}$, consider an input-output data set $\{\mathbf{x}^{(l)}, y(\mathbf{x}^{(l)})\}_{l=1}^{L'}$ of number $L' \in \mathbb{N}$. Given an autocorrelation function $R(|\mathbf{x} - \mathbf{x}'|; \boldsymbol{\theta})$ between two input realizations \mathbf{x} and \mathbf{x}' , the hyper-parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)^{\top}$ can be determined by the leave-one-out cross-validation estimate method [1] which solves

$$\arg \min_{\boldsymbol{\theta} \in \mathbb{R}^N} [\mathbf{b}^{\top} \mathbf{R}^{-1} \text{diag}(\mathbf{R}^{-1})^{-2} \mathbf{R}^{-1} \mathbf{b}], \quad (44)$$

where

$$\mathbf{R} := \begin{bmatrix} R(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}; \boldsymbol{\theta}) & \cdots & R(\mathbf{x}^{(1)}, \mathbf{x}^{(L')}; \boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ R(\mathbf{x}^{(L')}, \mathbf{x}^{(1)}; \boldsymbol{\theta}) & \cdots & R(\mathbf{x}^{(L')}, \mathbf{x}^{(L')}; \boldsymbol{\theta}) \end{bmatrix} \quad (45)$$

and $\mathbf{b} = (y(\mathbf{x}^{(1)}), \dots, y(\mathbf{x}^{(L')}))^\top$ is an L' -dimensional column vector of output evaluated at each input $\mathbf{x}^{(l)}$, $l = 1, \dots, L'$.

The optimal value of $\boldsymbol{\theta}$ from (44) is used to determine the mean $\bar{y}_{S,m}(\mathbf{x})$ and variance $\bar{\sigma}_{S,m}^2(\mathbf{x})$ of the S -variate, m th-order DD-GPCE-Kriging, as explained in Section 3.2.