arXiv:2611.05087v1 [cs.CE] 3 Nov 2026

Enabling risk and reliability analysis in next-generation computational tools via multifidelity methods

Anirban Chaudhuri and Karen Willcox

University of Texas at Austin, TX 78712, USA

Abstract

Risk and reliability analysis are fundamental to the design, assessment, and operation of safety-critical complex systems; however, their integration into next-generation computational tools is constrained by the cost of uncertainty quantification with expensive computational models. The computational challenge is particularly prohibitive for tail-sensitive quantities, such as probabilities of failure, conditional values-at-risk, and buffered probabilities of failure, where accurate estimation can require a large number of model evaluations. Multifidelity methods provide a framework for addressing this computational bottleneck by combining information from models of varying accuracy and cost while retaining the predictive accuracy of the high-fidelity model. This review summarizes the state of the art in methods that enable scalable risk and reliability analysis using multifidelity control variates, multifidelity importance sampling, and multifidelity active learning.

1 Introduction

Risk assessment and reliability analysis are critical aspects of engineering and scientific workflows, playing particularly important roles in system analysis, assessment, certification, system design, experimental design, and decision under uncertainty. The emergence of powerful next-generation computational tools such as digital twins [1, 2, 3] makes these workflows increasingly computational, yet there is a gap in capabilities to achieve quantitative risk and reliability assessments at scale [4]. For example, AI-enabled and simulation-based design of aerospace systems and energy infrastructure requires statistical analysis to estimate tail risks and rare-event probabilities using computational models that are increasingly expensive to evaluate [5, 6, 7]. Similarly, AI-enabled digital twins that support medical decision-making must go beyond point estimates; they must provide the decision-maker with characterization and quantification of uncertainties [8, 9]. Over the past two decades a rich set of uncertainty quantification methods have emerged to address these needs (see e.g., [10]), but it remains a challenge to achieve computationally tractable uncertainty quantification, especially when the digital twin embeds expensive high-fidelity simulations that are needed to characterize low-probability, high-consequence events. This computational challenge is being addressed by the recent advancements in multifidelity uncertainty quantification methods.

Multifidelity uncertainty quantification methods provide a structured way to assess risk and reliability at acceptable computational cost without compromising accuracy. Multifidelity methods

leverage multiple computational models of varying degrees of fidelity and computational cost, including detailed high-fidelity models, reduced-order and surrogate models, reduced-resolution models, data-driven models, and simplified-physics models. This review examines multifidelity methods that accelerate risk and reliability analysis while maintaining rigorous error control, and discusses their integration into next-generation computational tools such as digital twins.

As one concrete example, Figure 1 illustrates the risk-aware digital twin paradigm for precision oncology, in which patient-specific data calibrate a hierarchy of tumor growth models, multifidelity uncertainty quantification is used to estimate distributions of clinically meaningful quantities of interest (QoIs), and risk or reliability measures inform treatment selection. As described in [11], incorporating uncertainty quantification into medical digital twins is essential to account for uncertainty in the digital twin models by continuous integration of incoming patient data, to forecast the parameter uncertainty to estimate risk associated with the therapeutic outcomes, and to support risk-aware clinical decision-making under uncertainty. Multifidelity methods play a key role in making uncertainty-aware, patient-specific decision support computationally tractable. Digital twin frameworks across medicine and engineering share a common need to integrate models and data under uncertainty, making multifidelity methods especially attractive for efficient prediction, risk assessment, and operational decision-making.

Section 2 introduces common risk and reliability measures—probability of failure, conditional value-at-risk (CVaR), and buffered probability of failure—and discusses their important role in engineering and scientific workflows. Section 3 then presents the multifidelity setting and discusses three specific classes of multifidelity methods for improving estimator efficiency for these risk and reliability measures: multifidelity control variates, multifidelity importance sampling, and multifidelity active learning with probabilistic surrogates. Finally, Section 4 concludes with an outlook and discussion of how these methods can enable next-generation computational tools with scalable reliability and risk assessment.

2 Risk and reliability measures

Risk and reliability measures quantify how a system responds to uncertain inputs. Let $\boldsymbol{\xi} \in \Xi$ denote a vector of uncertain inputs with probability density $\pi(\boldsymbol{\xi})$. A computational model maps these uncertain inputs to an output QoI as

$$y = f(\boldsymbol{\xi}), \tag{1}$$

where $f : \Xi \to \mathbb{R}$ represents the computational simulation mapping from input space Ξ to a scalar QoI $y \in \mathbb{R}$, such as a structural response, temperature, stress, energy output, or treatment response. The uncertainty in $\boldsymbol{\xi}$ induces uncertainty in the QoI, and so in (1) y is a random variable.

The mean QoI response is given by $\mu = \mathbb{E}_\pi[y] = \int_\Xi y\pi(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}$ and the variance by $\sigma^2 = \mathbb{V}_\pi[y] = \int_\Xi (y-\mu)^2\,\pi(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}$. Box 1 shows four commonly used risk and reliability measures that are computed by estimating different statistics of the random variable y. Risk and reliability analysis focus on the tails of the distribution, and are typically used when the tails correspond to undesirable or unsafe system behavior. Let y_{th} denote a prescribed failure threshold, with failure defined by $y > y_{\text{th}}$. Then the probability of failure (PoF) is

$$p_f = \mathbb{P}_\pi\left(y > y_{\text{th}}\right) = \int_\Xi \mathbb{I}\left[y > y_{\text{th}}\right]\pi(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} = \mathbb{E}_\pi\left[\mathbb{I}\left[y > y_{\text{th}}\right]\right], \tag{2}$$

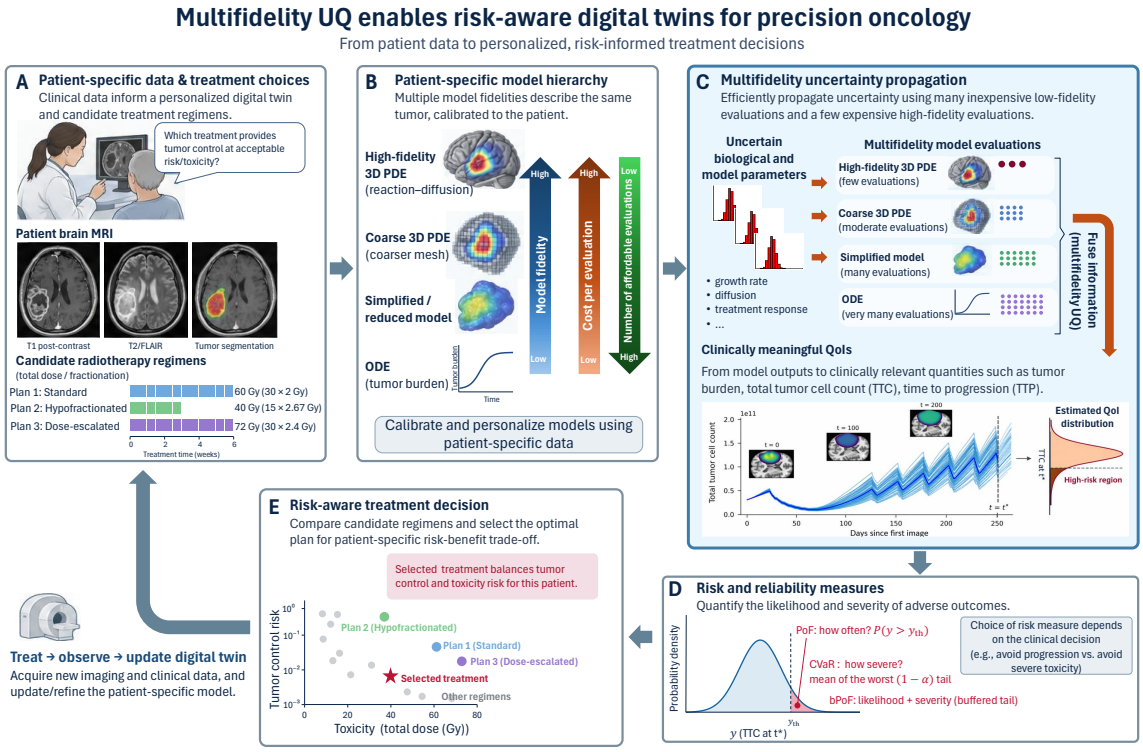


Figure 1: Risk-aware digital twin framework for precision oncology using multifidelity methods. (A) Patient-specific data and candidate treatment regimens define the clinical decision problem. (B) These data are used to personalize a hierarchy of tumor growth models with varying fidelities and computational costs. (C) Multifidelity methods efficiently propagate uncertainty across these models to estimate predictive distributions of clinically meaningful QoIs, (D) These distributions are summarized using risk and reliability measures. (E) Candidate treatments are compared on the basis of tumor control risk and toxicity trade-offs. The digital twin is iteratively updated as new patient data become available.

where $\mathbb{I}[\cdot]$ is the indicator function. Equivalently, if F_y denotes the cumulative distribution function of y, then $p_f = 1 - F_y(y_{\text{th}})$. The PoF is a popular measure of reliability because it directly quantifies the likelihood of violating a prescribed performance threshold. However, the indicator function in Eq. (2) does not distinguish between outcomes that barely violate the threshold and outcomes that substantially exceed it. This distinction between the likelihood and severity of exceedances motivates tail-sensitive risk measures.

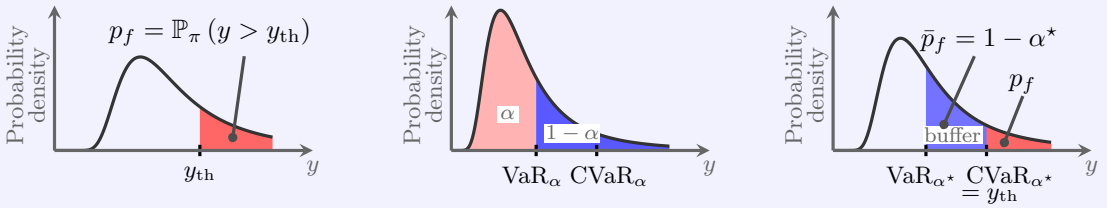
Tail-sensitive risk measures account for the magnitude of unsafe outcomes. For a given distribution of y, the value-at-risk (VaR) is defined as the α-quantile,

$$\text{VaR}_\alpha(y) = \inf\{y \in \mathbb{R} : F_y(y) \geq \alpha\} = F_y^{-1}(\alpha), \qquad 0 < \alpha < 1. \tag{3}$$

The α-quantile identifies the threshold below which a fraction α of the outcomes lie, with the remaining fraction $1 - \alpha$ constituting the upper tail of the distribution, as illustrated in Box 1.

Box 1 — Key concepts in risk and reliability analysis

Term	Meaning
Quantity of interest (QoI)	A model output used to characterize system behaviour.
Probability of failure (PoF)	The probability that the QoI exceeds a prescribed failure threshold, estimating *how often the threshold is exceeded.*
Value-at-risk (VaR)	The α-quantile of a QoI distribution. It identifies a value below which a fraction α of outcomes lie, indicating *where the tail begins.*
Conditional value-at-risk (CVaR)	A tail-risk measure that quantifies the average severity of outcomes in the upper tail beyond a specified quantile. It estimates *how severe the outcomes are, on average, in the tail.* It is also known as the superquantile or expected shortfall.
Buffered probability of failure (bPoF)	A reliability measure based on the superquantile that accounts for both the likelihood and severity of threshold exceedances. It provides a conservative alternative to the conventional probability of failure and estimates *what probability level corresponds to a tail average reaching the failure threshold.*



(a) PoF **(b) CVaR** **(c) buffered PoF**

The conditional value-at-risk (CVaR) characterizes the average severity of outcomes in the upper $(1-\alpha)$ tail [12, 13, 14, 15]. CVaR can be expressed in terms of VaR as

$$\text{CVaR}_\alpha(y) = \frac{1}{1-\alpha}\int_\alpha^1 \text{VaR}_\gamma(y)\,\mathrm{d}\gamma. \tag{4}$$

For a continuous distribution of y, this corresponds to the expected value of y in the upper $(1-\alpha)$ tail, given by $\mathbb{E}_\pi[y \mid y \geq \text{VaR}_\alpha(y)]$, and thus estimates the average severity of outcomes in this tail, as illustrated in Box 1. CVaR can be equivalently written as

$$\text{CVaR}_\alpha(y) = \text{VaR}_\alpha(y) + \frac{1}{1-\alpha}\mathbb{E}_\pi\left[(y - \text{VaR}_\alpha(y))_+\right] = \inf_{\gamma\in\mathbb{R}}\left\{\gamma + \frac{1}{1-\alpha}\mathbb{E}_\pi\left[(y-\gamma)_+\right]\right\}, \tag{5}$$

where $(\cdot)_+ \equiv \max(\cdot, 0)$. The latter optimization-based representation in (5) is particularly useful computationally because it expresses CVaR in terms of an expectation and a scalar optimization problem.

A related tail-sensitive measure for reliability analysis is the buffered probability of failure (bPoF) [16] given by

$$\bar{p}_f = 1-\alpha^\star, \qquad \text{where } \text{CVaR}_{\alpha^\star}(y) = y_{\text{th}} \tag{6}$$

when such an $\alpha^\star$ exists. The value $\alpha^\star$ is the quantile level at which the average response in the upper $1-\alpha^\star$ tail, measured by $\mathrm{CVaR}_{\alpha^\star}(y)$, reaches the prescribed failure threshold y_{th}, as illustrated in Box 1. Thus, bPoF is the tail probability corresponding to the CVaR level that reaches the failure threshold. Unlike PoF, bPoF therefore accounts for both the likelihood and severity of threshold exceedances. Under typical conditions, the bPoF is at least as large as the conventional PoF, providing a more conservative characterization of reliability [16, 17, 18, 19, 20].

For engineering and scientific problems where the input-to-QoI map takes a general nonlinear form (1), PoF, bPoF and CVaR are estimated using Monte Carlo simulation, which generates many realizations of the uncertain inputs, evaluates the computational model for each realization to obtain samples of the QoI using (1), and then estimates from the QoI samples the desired statistics of interest, such as the risk and reliability measures. For example, given N independent QoI samples $\{y_i\}_{i=1}^N$, the Monte Carlo estimator of the mean is $\mathbb{E}_\pi[y] \approx \frac{1}{N}\sum_{i=1}^N y_i$. The cost of estimating these risk and reliability measures thus depends on the cost of a single model evaluation (one evaluation per Monte Carlo sample) multiplied by the number of evaluations required. This computational burden is typically severe for tail quantities and small probabilities, since Monte Carlo methods require sample sizes of thousands or more in order to accurately characterize the tails of the distribution. Multifidelity uncertainty quantification methods described next address this computational challenge by using cheaper models to reduce the number of expensive high-fidelity evaluations required.

3 Multifidelity risk and reliability analysis

3.1 Introducing the multifidelity setting

The QoI in Eq. (1) is generally obtained from a high-fidelity computational model whose results are trusted by the engineer or scientist but whose evaluation can require substantial computational resources. Estimation of risk and reliability measures via Monte Carlo simulation requires that this high-fidelity model be evaluated thousands of times, and this typically goes well beyond the available budget. A multifidelity method seeks to use the available budget optimally by leveraging cheaper, approximate models. Different multifidelity uncertainty quantification methods use different theoretical bases for optimally distributing evaluations across multiple models and for ensuring that the resulting estimated risk and reliability measures remain accurate.

We denote the highest fidelity model by $f^{(0)}(\boldsymbol{\xi})$, where the superscript (0) denotes the high-fidelity model. We consider a collection of L lower-fidelity models, $\{f^{(1)}, \ldots, f^{(L)}\}$. The low-fidelity models provide cheaper, but generally less accurate, approximations of the same QoI. For a model of fidelity level ℓ, we have $y^{(\ell)} = f^{(\ell)}(\boldsymbol{\xi}), \ell = 0, \ldots, L$, with corresponding computational cost c_ℓ. We order the models so that $c_0 > c_1 \geq \cdots \geq c_L$. The models do not need to form a nested sequence or possess a strict hierarchy in accuracy except knowing the highest-fidelity model. In the context of the oncology digital twin illustrated in Figure 1, the high-fidelity model can be a fully-resolved three-dimensional parametric partial differential equation model of tumor growth. Lower-fidelity alternatives can include models with coarser spatial discretizations, lumped-parameter ordinary differential equation models, or projection-based reduced-order models, which provide varying trade-offs between computational cost and predictive accuracy while describing the same underlying tumor growth process.

The central objective of multifidelity risk and reliability analysis is to reduce the number of expensive, high-fidelity evaluations required for risk and reliability estimation by supplementing with

cheaper, lower-fidelity model solves. Multifidelity methods have been shown to provide orders of magnitude reductions in overall computational cost without compromising accuracy. An important consideration in multifidelity methods is the statistical relationship between the outputs of models at different fidelity levels. In particular, lower-fidelity models are useful when their outputs provide information about the high-fidelity response. This relationship can be quantified through correlation, which measures how strongly the outputs of two models vary together as the uncertain inputs change. For example, if a lower-fidelity model consistently predicts larger responses for the same uncertain input realizations that produce larger responses in the high-fidelity model, the two models exhibit positive correlation. Note that strong correlation does not require the models to produce identical predictions or even have small prediction errors; a lower-fidelity model can have systematic bias while still capturing the correct trends of the high-fidelity response. The multifidelity strategies differ in how lower-fidelity information is incorporated. Information from lower-fidelity models can: (i) be used as correlated control variates (multifidelity control variates), (ii) concentrate computational effort in rare-event regions using lower-fidelity models to identify important regions (multifidelity importance sampling), or (iii) identify where and at what fidelity additional evaluations are most informative to reduce uncertainty in approximations for high-fidelity predictions (multifidelity active learning). Multifidelity methods can also exploit temporal or workflow-level information reuse, treating previously generated data, samples, or surrogate models as additional information sources. Box 2 shows the key concepts for multifidelity methods described in this review.

3.2 Multifidelity control variates

Control variates reduce the variance of Monte Carlo estimators by exploiting the correlation between the QoI and an auxiliary quantity whose expectation is known or can be estimated [21, 22]. In multifidelity control variates approaches, the auxiliary quantities are obtained from evaluating lower-fidelity models, enabling information from multiple models of varying fidelity to be combined. Because lower-fidelity models are cheaper to evaluate, many more samples can be obtained from them than from the high-fidelity model. In order to provide computational benefit, the lower-fidelity models require strong correlation (i.e., captures similar variations in the QoI across different realizations of the uncertain inputs) and sufficiently lower cost with respect to the high-fidelity model. Stronger correlation between models allows the more frequently evaluated lower-fidelity models to provide information about the high-fidelity response and reduce the variance of the estimated statistics. The multilevel Monte Carlo (MLMC) method uses a control variates approach in which estimators are constructed from a telescoping sum of differences between successive model fidelities [23, 24, 25, 26, 27, 28]. The multifidelity Monte Carlo (MFMC) method combines the control variates principle with optimal resource allocation across a collection of multifidelity models [29, 30, 31, 32, 33]. The optimal allocation balances the cost of each model against its correlation with the high-fidelity model via an explicit sample allocation strategy that minimizes estimator variance for a given computational budget [30]. The more general approximate control variate (ACV) estimators [34] and multifidelity best linear unbiased estimator (MLBLUE) [35, 36, 37] relax some of the sampling and nesting assumptions of MFMC and MLMC.

To introduce the setup and notation, first consider the multifidelity control variate estimator

Box 2 — Key concepts in multifidelity methods

Term	Meaning
High-fidelity model	A computational model that provides a high-accuracy prediction of the QoI, typically at relatively high computational cost.
Low-fidelity model	A computationally cheaper model that provides an approximation of the QoI. Although generally less accurate than a high-fidelity model, it can provide useful information.
Correlation	The correlation measures how strongly the model outputs from different fidelity levels vary together as the uncertain inputs change.
Multifidelity methods	Computational methods that combine information from models with different accuracy and computational costs to estimate a high-fidelity QoI statistic more efficiently than could be done by using the high-fidelity model alone.
Budget	The amount of resources available to conduct a computational task, such as estimation of a risk or reliability measure. Typically budget is practically wall clock time or processing time, but is often expressed in terms of the number of executions of the high-fidelity model.

Multifidelity uncertainty quantification techniques

Control variates
Exploit correlation across fidelities

Importance sampling
Use cheap models to identify tail regions

Active learning
Improve surrogate by querying where uncertainty matters

for the expected value, $\mu = \mathbb{E}_\pi[y^{(0)}]$:

$$\mathbb{E}_\pi\left[y^{(0)}\right] \approx \frac{1}{N_0}\sum_{i=1}^{N_0} y_i^{(0)} + \sum_{\ell=1}^{L}\beta_\ell\left[\frac{1}{N_\ell}\sum_{i=1}^{N_\ell} y_i^{(\ell)} - \frac{1}{N_{\ell-1}}\sum_{i=1}^{N_{\ell-1}} y_i^{(\ell)}\right], \tag{7}$$

where $y_i^{(\ell)}$ denotes the ith QoI sample evaluated using model ℓ and N_ℓ is the total number of samples of model ℓ. We have $N_L \geq N_{L-1} \geq \cdots \geq N_0 > 0$, that is, cheaper models are sampled more often than more expensive models. The first term on the right-hand side of (7) is the standard Monte Carlo estimate of the mean using N_0 samples of the high-fidelity model $y^{(0)}$. The second term on the right-hand side of (7) provides a correction to this mean estimate using the lower-fidelity models $\ell = 1, \ldots, L$, weighted by the control variate coefficients β_ℓ, $\ell = 1, \ldots, L$. Notice that the correction terms associated with model ℓ is zero in expectation, so that the estimator (7) remains unbiased even when the low-fidelity models themselves have error. The advantage of introducing the control variate terms is that they reduce the variance of the estimator, so that for a given computational budget (7) provides a better estimate of $\mathbb{E}_\pi[y^{(0)}]$ than would be obtained by sampling only $y^{(0)}$.

To illustrate the variance reduction achieved with multifidelity control variates, consider the MFMC method. In the MFMC setup, the samples are nested and reused; that is, the term $\sum_{i=1}^{N_{\ell-1}} y_i^{(\ell)}$ is computed with samples that are a subset of those used to compute the term $\sum_{i=1}^{N_\ell} y_i^{(\ell)}$.

For a fixed computational budget, the sample allocation can be optimized by jointly accounting for the correlation between each lower-fidelity model and the high-fidelity model, and the corresponding model evaluation cost. After solving the optimal MFMC allocation problem, the ratio of mean squared errors (MSEs) of MFMC ($\mathrm{MSE}_{\mathrm{MFMC}}$) compared to standard single-fidelity Monte Carlo using only the high-fidelity model ($\mathrm{MSE}_{\mathrm{MC}}$) shows the cost–correlation trade-off required for variance reduction as given by [30]:

$$\frac{\mathrm{MSE}_{\mathrm{MFMC}}}{\mathrm{MSE}_{\mathrm{MC}}} = \left[\sum_{\ell=0}^{L} \sqrt{\frac{c_\ell}{c_0}\left(\rho_{0,\ell}^2 - \rho_{0,\ell+1}^2\right)}\right]^2, \tag{8}$$

where $\rho_{0,\ell}$ is the correlation between the high-fidelity QoI $y^{(0)}$ and the QoI $y^{(\ell)}$ for $\ell = 0, \ldots, L$. We note that $\rho_{0,0} = 1$ and define $\rho_{0,L+1} = 0$. Equation (8) illustrates the trade-off between model cost and correlation in determining the computational benefit of MFMC. Stronger correlation between the high- and low-fidelity models generally enables greater variance reduction, whereas the relative computational cost determines how many additional lower-fidelity evaluations can be afforded. The differences between successive squared correlation coefficients also account for the additional information provided by each fidelity level relative to the preceding models. Thus, a lower-fidelity model is most beneficial when it provides substantial additional correlation with the high-fidelity response at a sufficiently low computational cost.

The multifidelity control variate approach applies directly to the risk and reliability measures introduced in Section 2 whenever the target statistic can be expressed as an expectation. The PoF, written as the expectation of the high-fidelity failure indicator as in Eq. (2), can be approximated using multifidelity control variates as

$$\mathbb{E}_\pi\left[\mathbb{I}\left[y^{(0)} > y_{\mathrm{th}}\right]\right] \approx \frac{1}{N_0}\sum_{i=1}^{N_0}\mathbb{I}\left[y_i^{(0)} > y_{\mathrm{th}}\right] + \sum_{\ell=1}^{L}\beta_\ell\left[\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\mathbb{I}\left[y_i^{(\ell)} > y_{\mathrm{th}}\right] - \frac{1}{N_{\ell-1}}\sum_{i=1}^{N_{\ell-1}}\mathbb{I}\left[y_i^{(\ell)} > y_{\mathrm{th}}\right]\right]. \tag{9}$$

Along these lines, MLMC has been extended to PoF estimation by applying control variates to the indicator function in Eq. (2) [38, 39]. For CVaR estimation, multifidelity control variates can be applied through the expectation in Eq. (5). For a fixed γ, the quantity $\mathbb{E}_\pi\left[\left(y^{(0)} - \gamma\right)_+\right]$ can be estimated using lower-fidelity control variates as

$$\mathbb{E}_\pi\left[\left(y^{(0)} - \gamma\right)_+\right] \approx \frac{1}{N_0}\sum_{i=1}^{N_0}\left(y_i^{(0)} - \gamma\right)_+ + \sum_{\ell=1}^{L}\beta_\ell\left[\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left(y_i^{(\ell)} - \gamma\right)_+ - \frac{1}{N_{\ell-1}}\sum_{i=1}^{N_{\ell-1}}\left(y_i^{(\ell)} - \gamma\right)_+\right], \tag{10}$$

after which the scalar minimization over γ can be performed as in Eq. (5). Note that multifidelity control variates are effective when lower-fidelity models are correlated with respect to the particular statistic being estimated and not just the raw model output QoI. For PoF estimation, this requires correlation between $\mathbb{I}\left[y^{(\ell)} > y_{\mathrm{th}}\right]$, which depends on how consistently the models classify the same input realization as safe or failed. For CVaR estimation, the correlation is required between $\left(y^{(\ell)} - \gamma\right)_+$, which measures the exceedance beyond a given threshold γ.

An alternative to directly estimating individual risk measures is to estimate the cumulative density function (CDF) of the high-fidelity QoI, $F_{y^{(0)}}(t) = \mathbb{P}(y^{(0)} \leq t)$, over a range of response thresholds t. Once the CDF has been estimated, several risk and reliability measures, such as

PoF and quantile-based measures can be obtained directly from the estimated CDF (see Eqs. (2) and (3)). Multilevel approaches for CDF estimation include estimators based on smoothing the discontinuous indicator function and adaptive selection of the smoothing parameter [40, 41]. CVaR estimation through MLMC-based CDF estimation is shown in [42, 43]. Recent work has also shown the ACV version of estimating $F_{y^{(0)}}$ [44].

To summarize, multifidelity control variates reduce computational cost by exploiting the correlation between the high- and low-fidelity estimators of the target statistic of the QoI while factoring in the relative computational costs of the models. The multifidelity control variate strategy differs from multifidelity importance sampling described in the next section, which instead uses lower-fidelity information to change where high-fidelity samples are generated. Multifidelity control variates are therefore attractive when lower-fidelity models provide accurate global information about the high-fidelity response, whereas importance sampling is advantageous when the estimator is dominated by localized rare-event regions.

3.3 Multifidelity importance sampling

Importance sampling is particularly useful for risk and reliability analysis when the regions that contribute most to the estimator are rarely sampled under the nominal input density $\pi(\boldsymbol{\xi})$. For example, when estimating a small PoF, most samples from $\pi(\boldsymbol{\xi})$ may correspond to the safe regions, providing limited information about the failure probability. The main idea in importance sampling is to shift the sampling distribution toward the important regions, such as the failure regions, so that these regions are sampled more frequently. Instead of drawing samples directly from π, importance sampling draws samples from a biasing density $q(\boldsymbol{\xi})$ that places more probability mass in important regions and corrects for the change of measure using the likelihood ratio π/q [21, 22]. Again to introduce the setup and notation, we start with rewriting the expectation of the QoI under a biasing density q as

$$\mathbb{E}_\pi\left[y^{(0)}\right] = \mathbb{E}_q\left[y^{(0)}\frac{\pi(\boldsymbol{\xi})}{q(\boldsymbol{\xi})}\right] \approx \frac{1}{N}\sum_{i=1}^{N} y_i^{(0)}\frac{\pi(\boldsymbol{\xi}_i)}{q(\boldsymbol{\xi}_i)}, \qquad \boldsymbol{\xi} \sim q, \tag{11}$$

such that $\text{supp}(\pi) \subseteq \text{supp}(q)$. The effectiveness of importance sampling therefore depends on constructing a biasing density that concentrates samples in regions important to the target quantity without excluding regions with non-zero probability under π. Multifidelity importance sampling (MFIS) approaches use lower-fidelity models to construct or guide this biasing density [45]. For MFIS approaches, the lower-fidelity models need to provide useful information about where important events occur with regard to the high-fidelity simulations.

The lower-fidelity models can be evaluated at lower computational cost and can therefore be used to explore Ξ more extensively than the high-fidelity model. The role of the lower-fidelity model in MFIS is to identify regions of the uncertain input space where high-fidelity evaluations are likely to be more informative. For reliability analysis, the failure region estimated using model ℓ can be written as $\mathcal{F}^{(\ell)} = \left\{\boldsymbol{\xi} \in \Xi : y^{(\ell)}(\boldsymbol{\xi}) > y_{\text{th}}\right\}, \ell = 0, \ldots, L$. The target reliability measure is determined by the high-fidelity failure region $\mathcal{F}^{(0)}$, but the lower-fidelity regions $\mathcal{F}^{(1)}, \ldots, \mathcal{F}^{(L)}$ can provide inexpensive approximations of where important samples are likely to occur, and thus provide information to construct a multifidelity biasing density q_{MF}. We write $q_{\text{MF}}(\boldsymbol{\xi}) = \mathcal{Q}\left(\boldsymbol{\xi}; y^{(1)}, \ldots, y^{(L)}\right)$, where $\mathcal{Q}$ denotes the procedure used to construct the sampling distribution from the available lower-fidelity information to approximate the theoretically optimal biasing density $q^* = \frac{\mathbb{I}\left[y^{(0)} > y_{\text{th}}\right]\pi(\boldsymbol{\xi})}{p_f}$. The theoretical optimal biasing density concentrates all sampling effort in the high-fidelity failure

region to eliminate non-informative samples; however, one cannot construct it in practice since it requires the knowledge of the PoF in advance. MFIS addresses this challenge by constructing q_{MF} to approximate q^* using samples from the hierarchy of lower-fidelity models to capture the failure region and then fitting a Gaussian mixture model to the sampled failure points. The MFIS estimator for the PoF defined in Eq. (2) is given by

$$\widehat{p}_{f,\mathrm{MFIS}} = \mathbb{E}_{q_{\mathrm{MF}}}\left[\mathbb{I}\left[y^{(0)} > y_{\mathrm{th}}\right]\frac{\pi(\boldsymbol{\xi})}{q_{\mathrm{MF}}(\boldsymbol{\xi})}\right] \approx \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\left[y_i^{(0)} > y_{\mathrm{th}}\right]\frac{\pi(\boldsymbol{\xi}_i)}{q_{\mathrm{MF}}(\boldsymbol{\xi}_i)}, \qquad \boldsymbol{\xi} \sim q_{\mathrm{MF}}. \tag{12}$$

The high-fidelity model therefore remains responsible for determining whether a sampled input constitutes a failure. Lower-fidelity models instead determine where high-fidelity models are preferentially evaluated. Provided $\mathrm{supp}(\pi) \subseteq \mathrm{supp}(q_{\mathrm{MF}})$ holds, errors in the lower-fidelity models affect the efficiency of the estimator through the quality of q_{MF}, but the PoF estimator $\widehat{p}_{f,\mathrm{MFIS}}$ remains unbiased. Related strategies are also used in delayed acceptance Markov chain Monte Carlo, where cheaper approximations are first used to screen proposed samples, and the expensive high-fidelity model is evaluated only for proposals that pass the initial screening [46, 47].

The same change-of-measure principle can be applied to tail-risk measures. In particular, the expectation appearing in the CVaR formulation in Eq. (5) can be evaluated using importance sampling, with lower-fidelity models used to concentrate high-fidelity evaluations in regions contributing to the upper tail. There are two approaches to estimating CVaR through MFIS. The first is to rewrite CVaR estimation for a continuous distribution of $y^{(0)}$ as $\mathrm{CVaR}_\alpha(y^{(0)}) = \mathbb{E}_\pi[y^{(0)} \mid y^{(0)} \geq \mathrm{VaR}_\alpha(y^{(0)})] = \frac{1}{1-\alpha}\mathbb{E}_\pi\left[y^{(0)}\,\mathbb{I}[y^{(0)} \geq \mathrm{VaR}_\alpha(y^{(0)})]\right]$ and then use MFIS to obtain [48, 49]

$$\widehat{\mathrm{CVaR}}_{\alpha,\mathrm{MFIS}}\left(y^{(0)}\right) = \frac{1}{1-\alpha}\mathbb{E}_{q_{\mathrm{MF}}}\left[y^{(0)}\,\mathbb{I}[y^{(0)} \geq \mathrm{VaR}_\alpha(y^{(0)})]\frac{\pi(\boldsymbol{\xi})}{q_{\mathrm{MF}}(\boldsymbol{\xi})}\right], \qquad \boldsymbol{\xi} \sim q_{\mathrm{MF}}. \tag{13}$$

In Eq. (13), the $\mathrm{VaR}_\alpha(y^{(0)})]$ is typically estimated jointly with the importance sampling procedure. A second approach is to use MFIS for the expectation appearing in the CVaR formulation in Eq. (5), with lower-fidelity models used to concentrate high-fidelity evaluations in regions contributing to the upper tail. For a fixed value of γ,

$$\mathbb{E}_\pi\left[(y^{(0)} - \gamma)_+\right] = \mathbb{E}_{q_{\mathrm{MF}}}\left[\left(y^{(0)} - \gamma\right)_+\frac{\pi(\boldsymbol{\xi})}{q_{\mathrm{MF}}(\boldsymbol{\xi})}\right], \qquad \boldsymbol{\xi} \sim q_{\mathrm{MF}}. \tag{14}$$

Consequently, the CVaR can be estimated by solving the same scalar minimization problem as in Eq. (5), but evaluating its expectation under the multifidelity biasing density. Note that Eq. (14) does not require the distribution of $y^{(0)}$ to be continuous.

Several approaches exploit the idea of creating multifidelity biasing densities. Mixed multifidelity importance sampling extends MFIS to multiple lower-fidelity models by using control variates to combine importance sampling estimators from several fidelity-specific q_{MF} distributions [50]. Another approach fuses multiple MFIS estimators into a weighted ensemble estimator with the weights selected to minimize the variance of the fused estimator [51]. Other approaches combine importance sampling with multilevel and multifidelity control variates, using importance sampling to concentrate evaluations in important regions while exploiting correlated lower-fidelity information to further reduce estimator variance [52, 53, 54]. In the context of adaptive importance sampling strategies, multifidelity cross-entropy (MFCE) uses lower-fidelity models to construct and progressively refine a biasing density before introducing more expensive high-fidelity evaluations [55]. This

provides a systematic way of adapting the sampling distribution to rare-event regions without requiring the high-fidelity model to explore the entire input space. Using Eq. (13), MFCE approaches have also been developed for CVaR estimation and risk-averse optimization [56]. Related multilevel approaches apply the adaptive biasing density construction idea to sequential importance sampling and subset simulation, using a hierarchy of increasingly accurate model discretizations to progressively concentrate samples toward rare-event regions. Multilevel subset simulation constructs a sequence of intermediate failure domains across model resolutions [57], while multilevel sequential importance sampling combines sequential importance sampling with adaptive model refinement to estimate the high-fidelity PoF more efficiently [58]. Similarly, one can apply MFIS and MFCE approaches for bPoF estimation using the importance sampling estimators described in [59, 60].

To summarize, MFIS approaches reduce computational cost through a mechanism distinct from multifidelity control variates. Whereas control variates use correlated lower-fidelity quantities to reduce the variance of an estimator, MFIS uses lower-fidelity information to change where high-fidelity evaluations are performed. MFIS is particularly useful when the target risk or reliability measure is dominated by a small region of the uncertainty space, such as a rare-event or extreme-tail region.

3.4 Multifidelity active learning

Active learning is an adaptive approach to refine an approximate surrogate model by sequentially selecting what information to add next for maximum improvement in the surrogate predictions. Rather than evaluating the computational model at a predetermined set of input locations to fit the surrogate model, active learning uses the accumulated information from previous evaluations to determine most valuable additional evaluations. Multifidelity active learning provides a third strategy for reducing the computational cost of risk and reliability analysis by adaptively selecting (i) where to evaluate the computational model and (ii) at what fidelity evaluations are most informative. The objective is to reduce uncertainty in a target risk or reliability estimate at minimum cost. For multifidelity active learning, lower-fidelity models are useful when their evaluations provide information that reduces goal-oriented prediction uncertainty in the high-fidelity response or in a quantity derived from it, such as a failure boundary.

An effective multifidelity probabilistic surrogate model that can quantify uncertainty in predictions as well as provide a natural framework for combining information from multiple fidelity models is central to multifidelity active learning. Gaussian processes (GPs) are particularly well suited to this setting because they provide a probabilistic prediction of the model response at each input location, characterized by a normal distribution of that leads to both a prediction mean and a prediction variance [61]. The predictive mean provides the estimated response, while the prediction variance quantifies the uncertainty associated with that response that can be used to identify regions of input space where additional model evaluations are likely to improve the surrogate. The GP framework also naturally allows information from multiple computational models to be combined into a multifidelity GP. Several multifidelity GPs formulations have been developed that include co-Kriging, autoregressive, nonlinear autoregressive, and more flexible covariance-based formulations [62, 63, 64, 65, 66, 67]. Related multi-output and multi-task GP formulations provide alternative ways to model correlations across fidelity levels [68, 69]. In a multifidelity GP, observations from inexpensive lower-fidelity models can reduce predictive uncertainty in the high-fidelity response, provided that the models exhibit a sufficiently informative statistical relationship. For example, if a lower-fidelity model captures the general variation of the high-fidelity response, obser-

vations from that model can improve predictions at input locations where high-fidelity evaluations are unavailable. The multifidelity GP learns this relationship from the available model evaluations while accounting for discrepancies between fidelity levels. Beyond GPs, other probabilistic surrogate models can provide improved scalability for high-dimensional inputs and large datasets and greater flexibility for modeling complex dynamical systems. Some such options for probabilistic surrogate models include probabilistic versions of neural networks, such as Bayesian neural networks, deep ensembles, and dropout-based approaches [70, 71, 71, 72], as well as physics-aware Bayesian reduced-order models [73, 74, 75, 76]. Recent work has begun to extend Bayesian neural networks to multifidelity surrogate modeling [77], but substantial opportunities remain for developing scalable, physics-informed multifidelity probabilistic surrogates.

The role of multifidelity probabilistic surrogate models in active learning is to quantify how observations at one fidelity reduce uncertainty in the high-fidelity response. The predictive uncertainty of the high-fidelity response is used to construct an acquisition function that balances the information gained from an additional evaluation against the computational cost of gaining that information. The construction of the acquisition function depends on the statistics of interest. For estimating the mean, an acquisition function that reduces global error, such as maximizing prediction variance with a cost penalty, can be suitable; however, for reliability analysis reducing global prediction error is not necessarily the appropriate objective and instead one should identify the failure boundary accurately [78]. This distinction motivates goal-oriented active learning.

In multifidelity active learning, the next evaluation is selected jointly in terms of its input location $\boldsymbol{\xi}$ and fidelity level ℓ. Let $a(\ell, \boldsymbol{\xi}; \mathcal{D}, c_\ell)$ denote an acquisition function representing the expected utility of evaluating fidelity ℓ at input $\boldsymbol{\xi}$, given the current multifidelity training data $\mathcal{D} = \bigcup_{\ell=0}^{L} \mathcal{D}_\ell$ and the fidelity-specific computational cost c_ℓ, where $\mathcal{D}_\ell = \{(\xi_i, y_i^{(\ell)})\}_{i=1}^{N_\ell}, \ell = 0, \ldots, L$ denotes the N_ℓ training samples obtained to date from model ℓ. A general cost-aware acquisition strategy is

$$(\ell^\star, \boldsymbol{\xi}^\star) = \underset{\ell \in \{0,\ldots,L\}, \boldsymbol{\xi} \in \Xi}{\arg\max} \; a\left(\ell, \boldsymbol{\xi}; \mathcal{D}, c_\ell\right). \tag{15}$$

The optimization statement (15) shows the two decisions inherent in multifidelity active learning: where to evaluate ($\boldsymbol{\xi}^\star$) and which fidelity to use ($\ell^\star$). A lower-fidelity evaluation is preferred when it is expected to substantially reduce the uncertainty in the target relevant to the target risk or reliability measure at a relatively small computational cost, whereas a high-fidelity evaluation becomes preferable when lower-fidelity information is insufficient.

For reliability analysis, the PoF depends primarily on the location of the failure boundary of the high-fidelity model output, $\Gamma^{(0)} = \left\{\boldsymbol{\xi} \in \Xi : y^{(0)}(\boldsymbol{\xi}) = y_{\text{th}}\right\}$. Consequently, the most valuable model evaluations are those that reduce the uncertainty about whether points near the failure boundary belong to the safe or the failure domain. The goal-oriented acquisition functions developed for the purpose of accurately identifying the failure boundary include probability of feasibility, expected feasibility function (EFF) [79], U-learning [80], and integrated measures such as the targeted integrated mean squared error [81]. Although these criteria differ in their definitions of utility of information, they concentrate evaluations in regions where uncertainty in the failure classification or failure boundary is high [78].

Multifidelity active learning can exploit lower-fidelity models to identify regions likely to contain the failure boundary and then determine whether a lower- or higher-fidelity evaluation is most informative. This is particularly useful when the failure boundary is localized in a small portion of Ξ; a low-fidelity model may accurately classify most of the input domain as clearly safe or clearly failed, leaving only a narrow region near $y^{(0)} = y_{\text{th}}$ that requires high-fidelity evaluations. The method

can therefore focus computational effort on the failure boundary region rather than constructing a globally accurate surrogate. Several multifidelity reliability methods can be viewed within this framework. CLoVER [82, 83] uses an entropy-based acquisition function to guide the selection of additional evaluations. mfEGRA [84] uses a two-stage process with EFF for location selection and a weighted information-gain-based criterion for fidelity selection. The two-stage process can be extended to any single-fidelity acquisition function to create a multifidelity analog [85]. CAMERA combines a similar cost-aware strategy that can utilize existing single-fidelity acquisition functions with importance sampling for PoF estimation [86]. Ref. [87] expands U-learning acquisition to a multifidelity setting and combines it with subset simulation. These approaches illustrate different strategies for location and fidelity-selection.

Beyond PoF, a similar goal-oriented formulation applies to other risk measures. For a generic risk or reliability quantity $\mathcal{G}[y^{(0)}]$, an ideal acquisition function would select the next model evaluation according to its expected reduction in uncertainty in $\mathcal{G}$,

$$(\ell^\star, \boldsymbol{\xi}^\star) = \underset{\ell \in \{0,\dots,L\}, \boldsymbol{\xi} \in \Xi}{\arg\max} \frac{1}{c_\ell} \mathbb{E}\left[\mathcal{U}\left(\mathcal{G}[y^{(0)}] \mid \mathcal{D}\right) - \mathcal{U}\left(\mathcal{G}[y^{(0)}] \mid \mathcal{D} \cup (\boldsymbol{\xi}, y^{(\ell)})\right)\right], \tag{16}$$

where $\mathcal{U}$ denotes an uncertainty measure associated with the target risk or reliability quantity and the expectation is taken over the possible outcome of the new model evaluation. The computational cost of each fidelity is incorporated into the acquisition function as a normalization factor, so that an inexpensive evaluation is favored when it is expected to provide substantial information about the target quantity, while a more expensive evaluation is selected when lower-fidelity evaluations provide insufficient information. For quantities such as CVaR, the relevant region may extend beyond the failure boundary to the portion of the input space contributing to the upper tail of the response. Goal-oriented acquisition functions for such quantities would therefore target uncertainty in the corresponding tail behavior or directly in the risk estimate as presented in the idealized acquisition function in Eq. (16). Developing efficient multifidelity acquisition functions specifically tailored to such risk measures remains an important direction for future work.

To summarize, multifidelity active learning reduces the cost of risk and reliability analysis by adaptively allocating model evaluations to the locations and fidelity levels that provide the greatest information about the target quantity. Within this framework, multifidelity probabilistic surrogate models are a key ingredient because they combine information from multiple sources while quantifying predictive uncertainty. Multifidelity active learning is complementary to multifidelity control variates and multifidelity importance sampling: rather than modifying an estimator or its sampling distribution, active learning reduces the number of high-fidelity evaluations required to construct a sufficiently accurate estimate of the target risk or reliability measure.

4 Outlook and future perspectives

Multifidelity optimization under uncertainty. An important step is to integrate multifidelity risk and reliability analysis into decision-making through optimization under uncertainty (OUU) formulations, where uncertainty quantification is performed repeatedly as the design evolves through the optimization iterations. For example, OUU can be used for risk-informed treatment decisions in oncology digital twins as illustrated in Figure 1 (panel E). Let $\boldsymbol{x} \in \mathcal{X}$ denote the vector of design variables within the design space $\mathcal{X}$ that represent the decisions to be made, such as shape definition, process variables, and control actions. In the OUU setting, computational model ℓ now

maps both the uncertain inputs $\boldsymbol{\xi}$ and the design variables $\boldsymbol{x}$ to a QoI $y^{(\ell)}$, which we write as $y^{(\ell)} = f^{(\ell)}(\boldsymbol{x}, \boldsymbol{\xi})$, where $f^{(\ell)} : \mathcal{X} \times \Xi \to \mathbb{R}$ denotes the model response for design $\boldsymbol{x}$ and uncertain input $\boldsymbol{\xi}$ as estimated by model ℓ. The goal is to solve the high-fidelity OUU problem, which can be generically written as

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x} \in \mathcal{X}} \mathcal{J}^{(0)}(\boldsymbol{x}) \quad \text{subject to} \quad \mathcal{R}^{(0)}(\boldsymbol{x}) \leq 0, \tag{17}$$

where $\mathcal{J}^{(0)}$ and $\mathcal{R}^{(0)}$ are respectively the objective function and constraint function evaluated with the high-fidelity model. Both the objective function and constraint function might include risk or reliability measures such as PoF, bPoF, or CVaR [12, 15, 20, 88]. The computational approaches to estimate these measures are the same as discussed in Section 3 (i.e., using sampling to integrate over the uncertain input $\boldsymbol{\xi}$) but the OUU setting requires repeated estimation of the risk or reliability measure for each value of the design variables $\boldsymbol{x}$ visited during the optimization iterations. Multifidelity methods can substantially reduce the computational cost of the repeated uncertainty quantification required at each optimization iteration, particularly for operational or time-sensitive decision-making. Multifidelity control variate approaches have been used for robust optimization with objectives involving linear combinations of mean and variance [29, 89, 90, 91], as well as for CVaR-based optimization [92]. The systematic extension of these multifidelity methods to bPoF-based optimization remains relatively unexplored. More broadly, despite emerging work on multifidelity OUU, fundamental questions remain regarding how to optimally allocate computational effort jointly across the design space, uncertain input space, and model fidelities. The multidimensional resource allocation and problem formulation for multifidelity OUU spans decisions over $\{\mathcal{R}, \boldsymbol{x}, \boldsymbol{\xi}, \ell\}$ to answer: (i) which risk or reliability measure to use in the OUU problem formulation, (ii) where in the design space to evaluate, (iii) how many uncertain input samples are required, (iv) where in the uncertain input space should those samples be concentrated, given the selected risk or reliability measure, and (v) which fidelity to use. An additional challenge arises when model fidelity is input-dependent, such that the correlation between high- and low-fidelity models varies across the design space and changes as the optimizer explores different regions. Finally, beyond multifidelity estimation of risk and reliability themselves, scalable high-dimensional OUU also requires efficient multifidelity estimation of their gradients with respect to the design variables.

Information reuse as a multifidelity information source. An alternative way to construct multifidelity methods is to treat information generated earlier in a computational workflow as an additional source of correlated information. Information reuse exploits previously generated model evaluations, surrogate models, or sampling distributions when estimating a risk or reliability measure at a new design point or iteration. This perspective can be implemented through several complementary variance reduction and active learning mechanisms. In a control variate setting, model evaluations obtained at a previous design point can be incorporated as correlated control variates for a new risk or reliability estimate, provided that the responses remain sufficiently correlated [93, 94, 95]. For importance sampling, previously constructed biasing densities and simulation data can be reused at nearby design points rather than restarting the sampling procedure, thus avoiding repeated identification of important regions [96, 97, 98]. Information reuse can also be combined with multifidelity active learning, where surrogates fit during earlier design iterations are retained and used as additional information sources for subsequent risk and reliability estimation in an optimization workflow [99]. Probabilistic re-analysis is a related line of research that can

also be interpreted as an information reuse approach [100, 101, 102]. These illustrate a broader opportunity to view information accumulated throughout an evolving computational workflow as a dynamic multifidelity resource, potentially enabling control variates, importance sampling, and active learning to share and reuse information across repeated risk and reliability analyses, specifically in optimization under uncertainty frameworks.

Downstream applications and integration into digital twins. Multifidelity risk and reliability analysis can support design, certification, and operational decision-making in safety-critical applications. In aerospace engineering, for example, high-fidelity computational fluid dynamics and structural mechanics models are used to assess the performance under uncertain operating conditions. Repeated reliability and tail-risk evaluations during design optimization and certification can be computationally prohibitive. Multifidelity methods address this challenge by combining detailed simulations with reduced-order models, coarser discretizations, and data-driven surrogates. Similar opportunities arise in energy infrastructure, where digital twins of wind turbines and power systems could use multifidelity methods to continuously assess system reliability as operating conditions and sensor measurements evolve. In medicine, multifidelity methods can advance patient-specific digital twins for precision medicine. As illustrated in Figure 1, a hierarchy of mechanistic and data-driven models can predict patient-specific treatment outcomes while accounting for uncertainty in model parameters, measurements, and treatment response. Risk and reliability measures can quantify the likelihood of inadequate tumor control or adverse treatment outcomes and support comparisons among candidate treatment strategies. Beyond individual patients, similar approaches could enable population-level risk assessment by characterizing variability in treatment outcomes across patient populations and identifying vulnerable patient subgroups.

Realizing these opportunities requires development of application-specific model hierarchies that preserve information relevant to target risk measures, which needs domain expertise. Future research should therefore develop application-aware multifidelity frameworks that integrate physical knowledge, observational data, and computational models. Such multifidelity frameworks can make risk and reliability assessment a routine component of next-generation digital twins rather than a computationally prohibitive analysis performed only during isolated design or assessment stages.

Acknowledgments

The authors acknowledge support from AFOSR grant FA9550-24-1-0327 under the Multidisciplinary University Research Initiative (MURI), NSF FDT-Biotech award 2436499, DARPA TRS award HR00112590114, and Department of Energy MMMIC grant DE-SC002317.

References

[1] Tuegel, E. J., Ingraffea, A. R., Eason, T. G., and Spottswood, S. M., "Reengineering Aircraft Structural Life Prediction Using a Digital Twin," *International Journal of Aerospace Engineering*, Vol. 2011, 2011, pp. 154798.

[2] Glaessgen, E. and Stargel, D., "The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles," *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, No. AIAA 2012-1818, 2012, pp. 1–14.

[3] Thelen, A., Zhang, X., Fink, O., Lu, Y., Ghosh, S., Youn, B. D., Todd, M. D., Mahadevan, S., Hu, C., and Hu, Z., "A Comprehensive Review of Digital Twin – Part 1: Modeling and Twinning Enabling Technologies," *Structural and Multidisciplinary Optimization*, Vol. 65, No. 12, 2022, pp. 354.

[4] National Academies of Sciences, Engineering, and Medicine, *Foundational Research Gaps and Future Directions for Digital Twins*, The National Academies Press, Washington, DC, 2024.

[5] Karve, P. M., Guo, Y., Kapusuzoglu, B., Mahadevan, S., and Haile, M. A., "Digital twin approach for damage-tolerant mission planning under uncertainty," *Engineering Fracture Mechanics*, Vol. 225, 2020, pp. 106766.

[6] Kapteyn, M. G., Pretorius, J. V. R., and Willcox, K. E., "A Probabilistic Graphical Model Foundation for Enabling Predictive Digital Twins at Scale," *Nature Computational Science*, Vol. 1, No. 5, 2021, pp. 337–347.

[7] Ferrari, A. and Willcox, K., "Digital Twins in Mechanical and Aerospace Engineering," *Nature Computational Science*, Vol. 4, No. 3, 2024, pp. 178–183.

[8] Sel, K., Hawkins-Daarud, A., Chaudhuri, A., Osman, D., Bahai, A., Paydarfar, D., Willcox, K., Chung, C., and Jafari, R., "Survey and perspective on verification, validation, and uncertainty quantification of digital twins for precision medicine," *npj Digital Medicine*, Vol. 8, No. 1, 2025, pp. 40.

[9] Laubenbacher, R., Mehrad, B., Shmulevich, I., and Trayanova, N., "Digital twins in medicine," *Nature Computational Science*, Vol. 4, No. 3, 2024, pp. 184–191.

[10] Smith, R. C., *Uncertainty Quantification: Theory, Implementation, and Applications, Second Edition*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2024.

[11] Chaudhuri, A., Pash, G., Hormuth, D. A., Lorenzo, G., Kapteyn, M., Wu, C., Lima, E. A. B. F., Yankeelov, T. E., and Willcox, K., "Predictive digital twin for optimizing patient-specific radiotherapy regimens under uncertainty in high-grade gliomas," *Frontiers in Artificial Intelligence*, Vol. 6, Oct. 2023, pp. 1222612.

[12] Rockafellar, R. T. and Uryasev, S., "Optimization of conditional value-at-risk," *Journal of Risk*, Vol. 2, 2000, pp. 21–42.

[13] Rockafellar, R. T. and Uryasev, S., "Conditional value-at-risk for general loss distributions," *Journal of Banking & Finance*, Vol. 26, No. 7, 2002, pp. 1443–1471.

[14] Rockafellar, R. T. and Royset, J. O., "Superquantiles and their applications to risk, random variables, and regression," *Theory Driven by Influential Applications*, Informs, 2013, pp. 151–167.

[15] Rockafellar, R. T. and Royset, J. O., "Engineering decisions under risk averseness," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, Vol. 1, No. 2, 2015, pp. 04015003.

[16] Rockafellar, R. T. and Royset, J. O., "On buffered failure probability in design and optimization of structures," *Reliability Engineering & System Safety*, Vol. 95, No. 5, 2010, pp. 499–510.

[17] Mafusalov, A. and Uryasev, S., "Buffered probability of exceedance: mathematical properties and optimization," *SIAM Journal on Optimization*, Vol. 28, No. 2, 2018, pp. 1077–1103.

[18] Norton, M., Khokhlov, V., and Uryasev, S., "Calculating CVaR and bPOE for common probability distributions with application to portfolio optimization and density estimation," *Annals of Operations Research*, 2019, pp. 1–35.

[19] Basova, H. G., Rockafellar, R. T., and Royset, J. O., "A computational study of the buffered failure probability in reliability-based design optimization," *Proceedings of International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP), Zurich, Switzerland*, Citeseer, 2011.

[20] Chaudhuri, A., Kramer, B., Norton, M., Royset, J. O., and Willcox, K., "Certifiable risk-based engineering design optimization," *AIAA Journal*, Vol. 60, No. 2, 2022, pp. 551–565.

[21] Bratley, P., Fox, B. L., and Schrage, L. E., *A Guide to Simulation*, Springer New York, New York, NY, 1987.

[22] Owen, A. B., *Monte Carlo theory, methods and examples*, https://artowen.su.domains/mc/, 2013.

[23] Heinrich, S., "Multilevel Monte Carlo methods," *International Conference on Large-Scale Scientific Computing*, Springer, 2001, pp. 58–67.

[24] Giles, M. B., "Multilevel Monte Carlo Path Simulation," *Operations Research*, Vol. 56, No. 3, 2008, pp. 607–617.

[25] Cliffe, K. A., Giles, M. B., Scheichl, R., and Teckentrup, A. L., "Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients," *Computing and Visualization in Science*, Vol. 14, No. 1, 2011, pp. 3.

[26] Barth, A., Schwab, C., and Zollinger, N., "Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients," *Numerische Mathematik*, Vol. 119, No. 1, 2011, pp. 123–161.

[27] Giles, M. B., "Multilevel Monte Carlo methods," *Acta Numerica*, Vol. 24, 2015, pp. 259–328.

[28] Giles, M. B. and Haji-Ali, A.-L., "Multilevel nested simulation for efficient risk estimation," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 7, No. 2, 2019, pp. 497–525.

[29] Ng, L. W. T. and Willcox, K. E., "Multifidelity Approaches for Optimization Under Uncertainty," *International Journal for Numerical Methods in Engineering*, Vol. 100, No. 10, 2014, pp. 746–772.

[30] Peherstorfer, B., Willcox, K., and Gunzburger, M., "Optimal Model Management for Multifidelity Monte Carlo Estimation," *SIAM Journal on Scientific Computing*, Vol. 38, No. 5, 2016, pp. A3163–A3194.

[31] Peherstorfer, B., "Multifidelity Monte Carlo estimation with adaptive low-fidelity models," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 7, No. 2, 2019, pp. 579–603.

[32] Farcaș, I.-G., Peherstorfer, B., Neckel, T., Jenko, F., and Bungartz, H.-J., "Context-aware learning of hierarchies of low-fidelity models for multi-fidelity uncertainty quantification," *Computer Methods in Applied Mechanics and Engineering*, Vol. 406, 2023, pp. 115908.

[33] Gruber, A., Gunzburger, M., Ju, L., and Wang, Z., "A multifidelity Monte Carlo method for realistic computational budgets," *Journal of Scientific Computing*, Vol. 94, No. 1, 2023, pp. 2.

[34] Gorodetsky, A. A., Geraci, G., Eldred, M. S., and Jakeman, J. D., "A Generalized Approximate Control Variate Framework for Multifidelity Uncertainty Quantification," *Journal of Computational Physics*, Vol. 408, 2020, pp. 109257.

[35] Schaden, D. and Ullmann, E., "On Multilevel Best Linear Unbiased Estimators," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 8, No. 2, 2020, pp. 801–838.

[36] Schaden, D. and Ullmann, E., "Asymptotic analysis of multilevel best linear unbiased estimators," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 9, No. 3, 2021, pp. 953–978.

[37] Croci, M., Willcox, K. E., and Wright, S. J., "Multi-output multilevel best linear unbiased estimators via semidefinite programming," *Computer Methods in Applied Mechanics and Engineering*, Vol. 413, 2023, pp. 116130.

[38] Fagerlund, F., Hellman, F., Målqvist, A., and Niemi, A., "Multilevel Monte Carlo methods for computing failure probability of porous media flow systems," *Advances in Water Resources*, Vol. 94, 2016, pp. 498–509.

[39] Elfverson, D., Hellman, F., and Målqvist, A., "A multilevel Monte Carlo method for computing failure probabilities," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 4, No. 1, 2016, pp. 312–330.

[40] Giles, M. B., Nagapetyan, T., and Ritter, K., "Multilevel Monte Carlo approximation of distribution functions and densities," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 3, No. 1, 2015, pp. 267–295.

[41] Lu, D., Zhang, G., Webster, C., and Barbier, C., "An improved multilevel Monte Carlo method for estimating probability distribution functions in stochastic oil reservoir simulations," *Water Resources Research*, Vol. 52, No. 12, 2016, pp. 9642–9660.

[42] Krumscheid, S. and Nobile, F., "Multilevel Monte Carlo approximation of functions," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 6, No. 3, 2018, pp. 1256–1293.

[43] Ayoul-Guilmard, Q., Ganesh, S., Krumscheid, S., and Nobile, F., "Quantifying uncertain system outputs via the Multi-level Monte Carlo Method- Distribution and robustness measures," *International Journal for Uncertainty Quantification*, Vol. 13, No. 5, 2023.

[44] Han, R., Kramer, B., Lee, D., Narayan, A., and Xu, Y., "An approximate control variates approach to multifidelity distribution estimation," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 12, No. 4, 2024, pp. 1349–1388.

[45] Peherstorfer, B., Cui, T., Marzouk, Y., and Willcox, K., "Multifidelity importance sampling," *Computer Methods in Applied Mechanics and Engineering*, Vol. 300, 2016, pp. 490–509.

[46] Christen, J. A. and Fox, C., "Markov chain Monte Carlo using an approximation," *Journal of Computational and Graphical Statistics*, Vol. 14, No. 4, 2005, pp. 795–810.

[47] Lykkegaard, M. B., Dodwell, T. J., Fox, C., Mingas, G., and Scheichl, R., "Multilevel Delayed Acceptance MCMC," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 11, No. 1, 2023, pp. 1–30.

[48] Heinkenschloss, M., Kramer, B., Takhtaganov, T., and Willcox, K., "Conditional-value-at-risk estimation via reduced-order models," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 6, No. 4, 2018, pp. 1395–1423.

[49] Lee, D. and Kramer, B., "Multifidelity conditional value-at-risk estimation by dimensionally decomposed generalized polynomial chaos-Kriging," *Reliability Engineering & System Safety*, Vol. 235, 2023, pp. 109208.

[50] Peherstorfer, B., Kramer, B., and Willcox, K., "Combining multiple surrogate models to accelerate failure probability estimation with expensive high-fidelity models," *Journal of Computational Physics*, Vol. 341, 2017, pp. 61–75.

[51] Kramer, B., Marques, A. N., Peherstorfer, B., Villa, U., and Willcox, K., "Multifidelity probability estimation via fusion of estimators," *Journal of Computational Physics*, Vol. 392, 2019, pp. 385–402.

[52] Kebaier, A. and Lelong, J., "Coupling importance sampling and multilevel Monte Carlo using sample average approximation," *Methodology and Computing in Applied Probability*, Vol. 20, No. 2, 2018, pp. 611–641.

[53] Pham, T. and Gorodetsky, A. A., "Ensemble approximate control variate estimators: Applications to multifidelity importance sampling," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 10, No. 3, 2022, pp. 1250–1292.

[54] Chakroborty, P., Dhulipala, S. L., and Shields, M. D., "Covariance-Free Bifidelity Control Variates Importance Sampling for Rare Event Reliability Analysis," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 13, No. 3, 2025, pp. 1406–1451.

[55] Peherstorfer, B., Kramer, B., and Willcox, K., "Multifidelity preconditioning of the cross-entropy method for rare event simulation and failure probability estimation," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 6, No. 2, 2018, pp. 737–761.

[56] Chaudhuri, A., Peherstorfer, B., and Willcox, K., "Multifidelity Cross-Entropy Estimation of Conditional Value-at-Risk for Risk-Averse Design Optimization," *AIAA SciTech 2020 Forum*, AIAA, 2020.

[57] Ullmann, E. and Papaioannou, I., "Multilevel estimation of rare events," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 3, No. 1, 2015, pp. 922–953.

[58] Wagner, F., Latz, J., Papaioannou, I., and Ullmann, E., "Multilevel sequential importance sampling for rare event estimation," *SIAM Journal on Scientific Computing*, Vol. 42, No. 4, 2020, pp. A2062–A2087.

[59] Harajli, M. M., Rockafellar, R. T., and Royset, J. O., "Importance sampling in the evaluation and optimization of buffered failure probability," 2015.

[60] Budhiraja, A., Lu, S., Yu, Y., and Tran-Dinh, Q., "Minimization of a class of rare event probabilities and buffered probabilities of exceedance," *Annals of Operations Research*, Vol. 302, No. 1, 2021, pp. 49–83.

[61] Rasmussen, C. E. and Williams, C. K. I., *Gaussian Processes for Machine Learning*, The MIT Press, 11 2005.

[62] Kennedy, M. C. and O'Hagan, A., "Predicting the output from a complex computer code when fast approximations are available," *Biometrika*, Vol. 87, No. 1, 2000, pp. 1–13.

[63] Forrester, A. I., Sóbester, A., and Keane, A. J., "Multi-fidelity optimization via surrogate modelling," *Proceedings: Mathematical, Physical and Engineering Sciences*, 2007, pp. 3251–3269.

[64] Le Gratiet, L. and Garnier, J., "Recursive co-kriging model for design of computer experiments with multiple levels of fidelity," *International Journal for Uncertainty Quantification*, Vol. 4, No. 5, 2014, pp. 365–386.

[65] Perdikaris, P., Venturi, D., Royset, J. O., and Karniadakis, G. E., "Multi-fidelity modelling via recursive co-kriging and Gaussian–Markov random fields," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 471, No. 2179, 07 2015, pp. 20150018.

[66] Perdikaris, P., Raissi, M., Damianou, A., Lawrence, N. D., and Karniadakis, G. E., "Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 473, No. 2198, 2017, pp. 20160751.

[67] Poloczek, M., Wang, J., and Frazier, P. I., "Multi-information source optimization," *Advances in Neural Information Processing Systems*, Vol. 30, 2017.

[68] Teh, Y. W., Seeger, M., and Jordan, M. I., "Semiparametric latent factor models," *International Workshop on Artificial Intelligence and Statistics*, PMLR, 2005, pp. 333–340.

[69] Bonilla, E. V., Chai, K., and Williams, C., "Multi-task Gaussian process prediction," *Advances in neural information processing systems*, Vol. 20, 2007.

[70] Blundell, C., Cornebise, J., Kavukcuoglu, K., and Wierstra, D., "Weight uncertainty in neural network," *International conference on machine learning*, PMLR, 2015, pp. 1613–1622.

[71] Gal, Y. and Ghahramani, Z., "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," *International Conference on Machine Learning*, PMLR, 2016, pp. 1050–1059.

[72] Lakshminarayanan, B., Pritzel, A., and Blundell, C., "Simple and scalable predictive uncertainty estimation using deep ensembles," *Advances in Neural Information Processing Systems*, Vol. 30, 2017.

[73] Guo, M., McQuarrie, S. A., and Willcox, K. E., "Bayesian operator inference for data-driven reduced-order modeling," *Computer Methods in Applied Mechanics and Engineering*, Vol. 402, 2022, pp. 115336.

[74] McQuarrie, S. A., Chaudhuri, A., Willcox, K. E., and Guo, M., "Bayesian learning with Gaussian processes for low-dimensional representations of time-dependent nonlinear systems," *Physica D: Nonlinear Phenomena*, Vol. 475, 2025, pp. 134572.

[75] Cheng, K. and Zimmermann, R., "Adaptive data-driven probabilistic reduced-order models for parameterized dynamical systems," *SIAM Journal on Scientific Computing*, Vol. 47, No. 2, 2025, pp. C426–C450.

[76] McQuarrie, S. A., Guo, M., and Chaudhuri, A., "Active learning for data-driven reduced models of parametric differential systems with Bayesian operator inference," *Journal of Scientific Computing*, Vol. 109, No. 2, 2026, pp. 42.

[77] Kerleguer, B., Cannamela, C., and Garnier, J., "A Bayesian neural network approach to multifidelity surrogate modeling," *International Journal for Uncertainty Quantification*, Vol. 14, No. 1, 2024.

[78] Moustapha, M., Marelli, S., and Sudret, B., "Active learning for structural reliability: Survey, general framework and benchmark," *Structural Safety*, Vol. 96, 2022, pp. 102174.

[79] Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., and McFarland, J. M., "Efficient global reliability analysis for nonlinear implicit performance functions," *AIAA Journal*, Vol. 46, No. 10, 2008, pp. 2459–2468.

[80] Echard, B., Gayton, N., and Lemaire, M., "AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation," *Structural Safety*, Vol. 33, No. 2, 2011, pp. 145–154.

[81] Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R. T., and Kim, N.-H., "Adaptive Designs of Experiments for Accurate Approximation of a Target Region," *Journal of Mechanical Design*, Vol. 132, No. 7, 06 2010, pp. 071008.

[82] Marques, A. N., Lam, R. R., and Willcox, K. E., "Contour Location via Entropy Reduction Leveraging Multiple Information Sources," *Advances in Neural Information Processing Systems*, Vol. 31, 2018, pp. 5217–5227.

[83] Marques, A. N., Opgenoord, M. M., Lam, R. R., Chaudhuri, A., and Willcox, K. E., "Multifidelity method for locating aeroelastic flutter boundaries," *AIAA Journal*, Vol. 58, No. 4, 2020, pp. 1772–1784.

[84] Chaudhuri, A., Marques, A. N., and Willcox, K. E., "mfEGRA: Multifidelity Efficient Global Reliability Analysis through Active Learning for Failure Boundary Location," *Structural and Multidisciplinary Optimization*, Vol. 64, No. 2, 2021, pp. 797–811.

[85] Chaudhuri, A. and Willcox, K., "Generalized multifidelity active learning for Gaussian-process-based reliability analysis," *Dynamic Data Driven Applications Systems*, Springer Nature Switzerland, 2022, pp. 17–26.

[86] Renganathan, S. A., Rao, V., and Navon, I. M., "CAMERA: A method for cost-aware, adaptive, multifidelity, efficient reliability analysis," *Journal of Computational Physics*, Vol. 472, 2023, pp. 111698.

[87] Dhulipala, S. L., Shields, M. D., Spencer, B. W., Bolisetti, C., Slaughter, A. E., Labouré, V. M., and Chakroborty, P., "Active learning with multifidelity modeling for efficient rare event simulation," *Journal of Computational Physics*, Vol. 468, 2022, pp. 111506.

[88] Rockafellar, R. T. and Uryasev, S., "The fundamental risk quadrangle in risk management, optimization and statistical estimation," *Surveys in Operations Research and Management Science*, Vol. 18, No. 1-2, 2013, pp. 33–53.

[89] Chaudhuri, A., Jasa, J., Martins, J. R., and Willcox, K. E., "Multifidelity optimization under uncertainty for a tailless aircraft," *2018 AIAA Non-Deterministic Approaches Conference*, 2018.

[90] Van Barel, A. and Vandewalle, S., "Robust Optimization of PDEs with Random Coefficients Using a Multilevel Monte Carlo Method," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 7, No. 1, 2019, pp. 174–202.

[91] Hyun, J., Chaudhuri, A., Willcox, K. E., and Kim, H. A., "Robust topology optimization of continuum structures under material uncertainties using multifidelity Monte Carlo," *Computer Methods in Applied Mechanics and Engineering*, Vol. 451, 2026, pp. 118714.

[92] Ganesh, S. and Nobile, F., "Gradient-based optimisation of the conditional-value-at-risk using the multi-level Monte Carlo method," *Journal of Computational Physics*, Vol. 495, 2023, pp. 112523.

[93] Ng, L. W. T. and Willcox, K. E., "Monte Carlo Information-Reuse Approach to Aircraft Conceptual Design Optimization Under Uncertainty," *Journal of Aircraft*, Vol. 53, No. 2, 2016, pp. 427–438.

[94] Cook, L. W., Jarrett, J. P., and Willcox, K. E., "Generalized information reuse for optimization under uncertainty with non-sample average estimators," *International Journal for Numerical Methods in Engineering*, Vol. 115, No. 12, 2018, pp. 1457–1476.

[95] Law, F., Cerfon, A., Peherstorfer, B., and Wechsung, F., "Meta variance reduction for Monte Carlo estimation of energetic particle confinement during stellarator optimization," *Journal of Computational Physics*, Vol. 495, 2023, pp. 112524.

[96] Beaurepaire, P., Jensen, H., Schuëller, G., and Valdebenito, M., "Reliability-based optimization using bridge importance sampling," *Probabilistic Engineering Mechanics*, Vol. 34, 2013, pp. 48–57.

[97] Chaudhuri, A., Kramer, B., and Willcox, K., "Information Reuse for Importance Sampling in Reliability-Based Design Optimization," *Reliability Engineering & System Safety*, Vol. 201, 2020, pp. 106853.

[98] Ehre, M., Papaioannou, I., Willcox, K. E., and Straub, D., "Conditional reliability analysis in high dimensions based on controlled mixture importance sampling and information reuse," *Computer Methods in Applied Mechanics and Engineering*, Vol. 381, 2021, pp. 113826.

[99] Chaudhuri, A., Marques, A. N., Lam, R., and Willcox, K. E., "Reusing Information for Multifidelity Active Learning in Reliability-Based Design Optimization," *AIAA SciTech 2019 Forum*, 2019.

[100] Kuczera, R. C., Mourelatos, Z. P., Latcha, M., and Nikolaidis, E., "An RBDO Method for Multiple Failure Region Problems using Probabilistic Reanalysis and Approximate Metamodels," *SAE International Journal of Materials and Manufacturing*, Vol. 2, No. 1, 2009, pp. 108–120.

[101] Kuczera, R., Mourelatos, Z., and Nikolaidis, E., "A re-analysis methodology for system RBDO using a trust region approach with local metamodels," *SAE International Journal of Materials and Manufacturing*, Vol. 3, No. 1, 2010, pp. 356–371.

[102] Chaudhuri, A., Waycaster, G., Price, N., Matsumura, T., and Haftka, R. T., "NASA uncertainty quantification challenge: an optimization-based methodology and validation," *Journal of Aerospace Information Systems*, Vol. 12, No. 1, 2015, pp. 10–34.