Multifidelity probability estimation via fusion of estimators

ACDL TR-2017-3

Boris Kramer [*] Alexandre Noll Marques[*]
Benjamin Peherstorfer[†] Umberto Villa[‡] Karen Willcox[*]

October 30, 2017

This paper develops a multifidelity method that enables estimation of failure probabilities for expensive-to-evaluate models via a new combination of techniques, drawing from information fusion and importance sampling. We use low-fidelity models to derive biasing densities for importance sampling and then fuse the importance sampling estimators such that the fused multifidelity estimator is unbiased and has mean-squared error lower than or equal to that of any of the importance sampling estimators alone. The presented general fusion method combines multiple probability estimators with the goal of further variance reduction. By fusing all available estimators, the method circumvents the challenging problem of selecting the best biasing density and using only that density for sampling. A rigorous analysis shows that the fused estimator is optimal in the sense that it has minimal variance amongst all possible combinations of the estimators. The asymptotic behavior of the proposed method is demonstrated on a convection-diffusion-reaction PDE model for which $n = 10^5$ samples can be afforded. To illustrate the proposed method at scale, we consider a model of a free plane jet and quantify how uncertainties at the flow inlet propagate to a quantity of interest related to turbulent mixing. The computed fused estimator has similar root-mean-squared error to that of an importance sampling estimator using a density computed from the high-fidelity model. However, it reduces the CPU time to compute the biasing density from 2.5 months to three weeks.

[*]Department of Aeronautics and Astronautics, Massachusetts Institute of Technology (bokramer@mit.edu, noll@mit.edu, kwillcox@mit.edu).

[†]Department of Mechanical Engineering and Wisconsin Institute for Discovery, University of Wisconsin-Madison (peherstorfer@wisc.edu).

[‡]Institute for Computational Engineering & Sciences, The University of Texas at Austin (uvilla@ices.utexas.edu).

Keywords: Multifidelity modeling, uncertainty quantification, information fusion, importance sampling, reduced-order modeling, failure probability estimation, PDEs, turbulent jet

1 Introduction

This paper considers estimation of failure probabilities for large-scale applications with expensive-to-evaluate models. Such uncertainty propagation is needed in reliable engineering design, where quantitative information about the probability of a systems failure to meet specified requirements is often required. However, small-probability estimation is computationally challenging with expensive-to-evaluate models. On the one hand, standard Monte-Carlo approaches require a large number of samples for low-variance estimates. On the other hand, the numerical solution of large-scale computational models—such as those arising from discretization of partial differential equations (PDEs)—incurs prohibitive computational costs when repeated simulations for different input parameters are required. Therefore, to make computations feasible, efficient methods aim to reduce the number of samples at which the expensive model is evaluated, by exploiting good sampling strategies and/or lower fidelity models.

An efficient variance reduction technique is importance sampling [Owen, 2013], which allows for order-of-magnitude reductions in the number of samples needed to reliably estimate a small probability. However, importance sampling shifts the problem towards finding a good biasing distribution which in turn requires insight into the system. Surrogate models can provide such information at much lower computational cost. Multifidelity approaches that use surrogates for failure probability estimation via sampling have seen great interest recently [Li et al., 2011, Chen and Quarteroni, 2013, Li and Xiu, 2014, Elfverson et al., 2016, Ullmann and Papaioannou, 2015, Peherstorfer et al., 2016, Fagerlund et al., 2016], but require that the user selects a good biasing density. To circumvent the selection of densities, multifidelity methods that use a general suite of surrogate models to generate importance sampling distributions were proposed in [Peherstorfer et al., 2017a, Peherstorfer et al., 2017b]. Mixture sampling can increase the variance of the estimator compared to sampling from the best biasing density by a factor that is equal to the number of biasing densities [Owen and Zhou, 2000, Peherstorfer et al., 2017a], but it avoids the selection of a single biasing density. Nevertheless, this framework requires all knowledge about the small probability event to be available in the form of biasing densities, and is therefore only applicable for importance sampling estimators.

This work considers a more general variance reduction method, namely information fusion, that allows for the combination of probability estimators. Those could be derived from experimental results, expert elicitations, analytical models, etc. Experimentalists and computational scientists alike have successfully used information fusion to include all available information about a probability in a single estimator, see [Clemen and Winkler, 1999, O'Hagan et al., 2006, Marín-Martínez and Sánchez-Meca, 2010]. Moreover, a weighted multifidelity method that combines information from multiple

models and sources has been successfully applied in data assimilation [Narayan et al., 2012].

We build on the information fusion framework in the context of sampling-based failure probability estimators, where in addition to the variance reduction from importance sampling, we obtain further variance reduction through information fusion. In particular, we propose a new combination of techniques to enable small probability estimation for large-scale, computationally expensive models that draws from prior work in information fusion, importance sampling, and multifidelity modeling. In addition to the computationally expensive model, more information about the system is available in form of, e.g., analytical models, expert elicitation, and surrogate modeling. The proposed multifidelity framework uses those surrogates to compute multiple unbiased failure probability estimators. We then combine them *optimally* into a new unbiased estimator that has minimal variance amongst all possible linear combinations of those estimators. The method therefore avoids the selection of the best biasing density to be used for sampling.

In Section 2 we formally introduce the problem of computing failure probabilities. Section 3 details the proposed combination of techniques of information fusion, importance sampling and multifidelity modeling. We then present in Section 4 a relatively cheap convection-diffusion-reaction test case, where we illustrate the asymptotic behavior of our approach. Section 5 discusses a turbulent jet model and demonstrates the computational efficiency of our proposed methods for this computationally expensive model. We close with conclusions in Section 6.

2 Small probability events

We are interested in computing events with small probabilities, e.g., failure events, where the system fails to meet critical constraints. Let Ω be a sample space which, together with a sigma algebra and probability measure, defines a probability space. Define a d-dimensional random variable $Z : \Omega \mapsto \mathcal{D} \subseteq \mathbb{R}^d$ with probability density p, and let $\mathbf{z}$ be a realization of Z. Let $f : \mathcal{D} \subseteq \mathbb{R}^d \mapsto \mathbb{R}^{d'}$ be an expensive-to-evaluate model of high fidelity with corresponding d'-dimensional quantity of interest $f(\mathbf{z}) \in \mathbb{R}^{d'}$. Let $g : \mathbb{R}^{d'} \mapsto \mathbb{R}$ denote a limit state function that defines failure of the system. If $g(f(\mathbf{z})) < 0$, then $\mathbf{z} \in \mathcal{D}$ is a configuration where the system fails. This defines a failure set

$$\mathcal{G} := \{\mathbf{z} \in \mathcal{D} \mid g(f(\mathbf{z})) < 0\}.$$

Define the indicator function $I_{\mathcal{G}} : \mathcal{D} \mapsto \{0, 1\}$ via

$$I_{\mathcal{G}}(\mathbf{z}) = \begin{cases} 1\,, & \mathbf{z} \in \mathcal{G}\,, \\ 0\,, & \text{otherwise}\,. \end{cases}$$

The failure probability is then evaluated by the average of the indicator function over the parameter domain, i.e.,

$$P = \mathbb{E}_p[I_{\mathcal{G}}[Z]] = \int_{\mathcal{D}} I_{\mathcal{G}}(\mathbf{z}) p(\mathbf{z}) d\mathbf{z} = \int I_{\{g(f(\mathbf{z}))<0\}}(\mathbf{z}) p(\mathbf{z}) d\mathbf{z}. \tag{1}$$

The standard Monte Carlo estimator of the failure probability integral (1) uses n realizations $\mathbf{z}_1, \ldots, \mathbf{z}_n$ of the random variable Z and estimates

$$P_n = \frac{1}{n} \sum_{i=1}^{n} I_{\mathcal{G}}(\mathbf{z}_i). \tag{2}$$

In the special case of small probabilities, standard Monte Carlo may be unfeasible due to the large number of samples needed to obtain good estimators. Since failure probabilities are generally small, most realizations $\mathbf{z}_i$ will be outside the failure domain $\mathcal{G}$, and conversely, only a small fraction of the n samples lies in the failure region. The coefficient of variation (also called relative root-mean-squared error) of the estimator P_n is given by

$$e^{\mathrm{CV}}(P_n) = \sqrt{\frac{\mathbb{V}[P_n]}{(\mathbb{E}[P_n])^2}} = \sqrt{\frac{P(1-P)}{nP^2}} = \sqrt{\frac{1-P}{nP}}. \tag{3}$$

Thus, to obtain estimators with a small coefficient of variation, a large number of samples is necessary. For instance, if the small probability is $P = 10^{-4}$ and if we want $e^{\mathrm{CV}} = 10^{-1}$ we would need $n = \mathcal{O}(10^6)$ samples via standard Monte Carlo approaches. This challenge is amplified by the presence of an expensive-to-evaluate model, such as the model of a free plane jet in Section 5.

3 Computational Methods and Analysis

We present a new combination of techniques, drawing from information fusion, importance sampling and multifidelity modeling that allows us to estimate small-probabilities as introduced above. Section 3.1 presents a method to optimally fuse probability estimators, which applies to a broad class of estimators that are derived, e.g., from analytical models, expert elicitation [O'Hagan et al., 2006] as well as computational models. In this work, we obtain the estimators via multifidelity importance sampling, as detailed in Section 3.2, where we also give the complete algorithm. In Section 3.3, we present measures of convergence of the estimators and introduce sampling-based estimates of the statistics.

3.1 Fusion of probability estimators

We formulate the fusion of probability estimators in Section 3.1.1 and derive expressions for the mean and variance of the fused estimator in Section 3.1.2. In Section 3.1.3, we derive the optimal weights for the fused estimator. Section 3.1.4 then discusses the special case of uncorrelated estimators.

3.1.1 Problem formulation

Consider the situation where k unbiased estimators, $P_1, \ldots, P_k$, of the probability P are available, that is

$$\mathbb{E}[P_i] = P, \ i = 1, \ldots, k.$$

These estimators have corresponding variances $0 < \sigma_i^2 < \infty,\ i = 1, \dots k$. We next specify the computational problem that we solve in this paper.

Problem 1. *Find an unbiased probability estimator for P by fusion of the available estimators $P_1, \dots, P_k$ of the form*

$$P_{\boldsymbol{\alpha}} = \sum_{i=1}^{k} \alpha_i P_i, \tag{4}$$

such that it attains minimal variance amongst all estimators of the form (4)*. That is, find the optimal weights $\alpha_i \in \mathbb{R},\ i = 1, \dots, k$ such that*

$$\min_{\boldsymbol{\alpha}} \ \mathbb{V}[P_{\boldsymbol{\alpha}}] \quad \text{s.t.} \quad \mathbb{E}[P_{\boldsymbol{\alpha}}] = P. \tag{5}$$

3.1.2 Mean and variance of fused estimator

We start with the observation that if the weights α_i sum to one, then the fused estimator $P_{\boldsymbol{\alpha}}$ is unbiased:

$$\sum_{i=1}^{k} \alpha_i = 1 \quad \Leftrightarrow \quad \mathbb{E}[P_{\boldsymbol{\alpha}}] = \sum_{i=1}^{k} \alpha_i \mathbb{E}[P_i] = P \sum_{i=1}^{k} \alpha_i = P.$$

To compute the variance of the fused estimator $P_{\boldsymbol{\alpha}}$ we have to consider covariances between the individual estimators. Define the Pearson product-moment correlation coefficient as

$$\rho_{i,j} = \frac{\mathbb{C}\text{ov}(P_i, P_j)}{\sigma_i \sigma_j}, \tag{6}$$

where $\mathbb{C}\text{ov}(P_i, P_j) = \mathbb{E}[(P_i - \mathbb{E}[P_i])(P_j - \mathbb{E}[P_j])] = \mathbb{E}[P_i P_j] - P^2$. We also define the symmetric, positive semi-definite *covariance matrix* $\boldsymbol{\Sigma}_{ij} = \mathbb{C}\text{ov}(P_i, P_j)$ as:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{1,2} & \dots & \dots & \sigma_1\sigma_k\rho_{1,k} \\ \sigma_2\sigma_1\rho_{2,1} & \sigma_2^2 & \sigma_2\sigma_3\rho_{2,3} & \dots & \sigma_2\sigma_k\rho_{2,k} \\ \vdots & \ddots & & & \\ \vdots & & & \sigma_{k-1}^2 & \sigma_{k-1}\sigma_k\rho_{k-1,k} \\ \sigma_k\sigma_1\rho_{k,1} & \sigma_k\sigma_2\rho_{k,2} & \dots & \sigma_k\sigma_{k-1}\rho_{k,k-1} & \sigma_k^2 \end{bmatrix}. \tag{7}$$

It is worth noticing that if the estimators $P_1, \dots, P_k$ are independent, then Σ is diagonal. With this notation, we obtain an explicit expression for the variance of the fused estimator.

Proposition 1. *Let the estimators $P_1, \dots, P_k$ and the covariance matrix Σ be given. Then the variance of the fused estimator $P_{\boldsymbol{\alpha}}$ from* (4) *is*

$$\mathbb{V}[P_{\boldsymbol{\alpha}}] = \boldsymbol{\alpha}^T \boldsymbol{\Sigma} \boldsymbol{\alpha} = \sum_{i=1}^{k} \alpha_i^2 \sigma_i^2 + 2 \sum_{i=1}^{k} \sum_{j>i}^{k} \alpha_i \alpha_j \sigma_i \sigma_j \rho_{i,j}. \tag{8}$$

Proof.

$$\mathbb{V}[P_{\boldsymbol{\alpha}}] = \mathbb{V}\left[\sum_{i=1}^{k} \alpha_i P_i\right] = \sum_{i=1}^{k} \alpha_i^2 \mathbb{V}[P_i] + 2\mathbb{C}\text{ov}\left(\sum_{i=1}^{k} \alpha_i P_i, \sum_{j=1}^{k} \alpha_j P_j\right)$$
$$= \sum_{i=1}^{k} \alpha_i^2 \sigma_i^2 + 2\sum_{i=1}^{k}\sum_{j>i}^{k} \alpha_i \alpha_j \sigma_i \sigma_j \rho_{i,j}.$$

□

In the following section, we provide an explicit formula to find the optimal weights $\boldsymbol{\alpha}$ for the general case of (possibly)-correlated estimators $P_1, \ldots, P_k$; while in Section 3.1.4 we discuss the case of independent estimators, such as those constructed in Section 3.2.

3.1.3 Optimizing the weights for minimum-variance estimate

The goal of the optimization problem (5) is to find the optimal $\boldsymbol{\alpha}$ such that the variance in (8) is minimized and $P_{\boldsymbol{\alpha}}$ remains unbiased. In this section, we show that such weights exist, are unique, and present a closed-form solution, provided that the covariance matrix Σ is invertible. This is summarized in the following result.

Proposition 2. *Let $\boldsymbol{P} = [P_1, \ldots, P_k]^T$ be the vector of probability estimators and assume that Σ is not singular. Define $\mathbf{1}_k = [1, \ldots, 1]^T$ as a column-vector of length k. The optimization problem* (5) *has the unique solution*

$$\boldsymbol{\alpha} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}_k}{\mathbf{1}_k^T \boldsymbol{\Sigma}^{-1}\ \mathbf{1}_k}.$$

That is, the minimal variance unbiased estimator $P_{\boldsymbol{\alpha}}$ is such that

$$P_{\boldsymbol{\alpha}} = \frac{\mathbf{1}_k^T \boldsymbol{\Sigma}^{-1}\ \boldsymbol{P}}{\mathbf{1}_k^T \boldsymbol{\Sigma}^{-1}\ \mathbf{1}_k}, \qquad \mathbb{V}[P_{\boldsymbol{\alpha}}] = \frac{1}{\mathbf{1}_k^T \boldsymbol{\Sigma}^{-1}\ \mathbf{1}_k}.$$

Proof. We have seen above that $\sum_{i=1}^{k} \alpha_i = 1$ if and only if $\mathbb{E}[P_{\boldsymbol{\alpha}}] = P$. Define the cost function $J(\boldsymbol{\alpha}) := \mathbb{V}[P_{\boldsymbol{\alpha}}] = \boldsymbol{\alpha}^T \boldsymbol{\Sigma} \boldsymbol{\alpha}$ by using Proposition 1. Therefore, the optimization problem (5) can be written as the quadratic program

$$\min_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^T \boldsymbol{\Sigma} \boldsymbol{\alpha}, \quad \text{s.t. } \boldsymbol{\alpha}^T \mathbf{1} = 1. \tag{9}$$

Letting $\mathcal{L}(\boldsymbol{\alpha}, \lambda) := \boldsymbol{\alpha}^T \boldsymbol{\Sigma} \boldsymbol{\alpha} + \lambda(\boldsymbol{\alpha}^T \mathbf{1} - 1)$ denote the Lagrangian cost function associated to (9), the optimality conditions are $\nabla_{\boldsymbol{\alpha}} \mathcal{L}(\boldsymbol{\alpha}, \lambda) = \mathbf{0}$ and $\frac{d\mathcal{L}}{d\lambda}(\boldsymbol{\alpha}, \lambda) = 0$. This optimality system is written as

$$\begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{1}_k \\ \mathbf{1}_k^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{0}_k \\ 1 \end{bmatrix}. \tag{10}$$

For invertible Σ, the unique weights to this quadratic program are then obtained by

$$\boldsymbol{\alpha} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}_k^T \boldsymbol{\Sigma}^{-1}\, \mathbf{1}_k}, \tag{11}$$

and the expression for the variance follows by inserting these weights into (9). The estimator is obtained by inserting the weights into (4). □

The weights can be expressed explicitly in terms of the components of the covariance matrix as

$$\alpha_i = \frac{1}{\sigma_i^2}\left[\frac{1}{\sum_{l=1}^k \frac{1}{\sigma_l^2}}\left(1 + \sum_{l=1}^k \frac{1}{\sigma_l^2}\sum_{j>l}^k \alpha_j \sigma_l \sigma_j \rho_{l,j}\right) - \sum_{j>i}^k \alpha_j \sigma_i \sigma_j \rho_{i,j}\right]. \tag{12}$$

Note, that the weights are inversely proportional to the variance of the individual estimators and the weight α_i depends on the covariance between the estimators P_i and P_j. Also, note that if P_i are correlated some weights may be negative, while for a diagonal Σ all weights α_i are *strictly* positive. In the next section, we have a closer look at the uncorrelated case.

3.1.4 The special case of uncorrelated estimators

In the situation where all estimators are uncorrelated, we recover the classical result of the *inverse variance-weighted mean* [Meier, 1953]. As a corollary from Proposition 2 we get the following result.

Corollary 1. *Consider the setting from Proposition 2, and let* $\Sigma = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_k^2)$ *be diagonal. Then the unique solution to the optimization problem* (5) *is given by*

$$\alpha_i = \frac{1}{\sigma_i^2 \sum_{i=1}^k \frac{1}{\sigma_i^2}}, \qquad \mathbb{V}[P_{\boldsymbol{\alpha}}] = \frac{1}{\sum_{i=1}^k \frac{1}{\sigma_i^2}}. \tag{13}$$

A few observations about this special case are in order:

1. The optimal coefficients α_i of the combined estimator $P_{\boldsymbol{\alpha}}$ are inversely proportional to the asymptotic variance σ_i of the corresponding estimator P_i. To reduce the variance via a weighted combination of estimators, smaller weights are assigned to estimators with larger variance.

2. If one variance is small compared to all other ones, say $\sigma_1^2 \ll \sigma_i^2,\ i = 2, \ldots, k$, then $\sum_{i=1}^k \frac{1}{\sigma_i^2} \approx \frac{1}{\sigma_1^2}$ so that $\mathbb{V}[P_{\boldsymbol{\alpha}}] \approx \sigma_1^2$. The estimators with large variance can not reduce the variance of the fused estimator much more.

3. If all estimators have equal variance, $\sigma_1^2 = \ldots = \sigma_k^2$, then $\sum_{i=1}^k \frac{1}{\sigma_i^2} = \frac{k}{\sigma_1}$ so that $\mathbb{V}[P_{\boldsymbol{\alpha}}] = \frac{\sigma_1^2}{k}$. Hence, combining the estimators reduces the variance by a factor of k.

4. Since $0 < \alpha_i < 1, \forall i$, it follows from both equations in (13) that

$$\mathbb{V}[P_{\boldsymbol{\alpha}}] = \frac{1}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}} = \sigma_i^2 \alpha_i < \sigma_i^2, \forall i \quad \Rightarrow \quad \mathbb{V}[P_{\boldsymbol{\alpha}}] < \min_{i=1,\dots,k} \sigma_i^2. \tag{14}$$

Consequently, we are guaranteed to reduce the variance in $P_{\boldsymbol{\alpha}}$ by combining all estimators in the optimal way described above.

3.2 Fusion of multifidelity importance sampling estimators

We now use the general fusion framework to obtain a failure probability estimate, where the individual importance-sampling estimators are derived with the help of surrogate models. We briefly review importance sampling in Section 3.2.1. We then introduce surrogate models and briefly review the multifidelity importance sampling framework from [Peherstorfer et al., 2016] in Section 3.2.2. We then present the complete algorithm for our fusion-based framework in Section 3.2.3.

3.2.1 Importance sampling

Importance sampling achieves variance reduction by using realizations of a random variable $Z' : \Omega \mapsto \mathcal{D}$ with probability density q. This random variable Z' is chosen such that its probability density function q has higher mass (compared to the nominal density p) in the region of the event of interest. For a general introduction to importance sampling, see [Owen, 2013, Sec.9]. Define the support $\text{supp}(p) = \{\mathbf{z} \in \mathcal{D} \mid p(\mathbf{z}) > 0\}$, and let $\text{supp}(p) \subseteq \text{supp}(q)$. Then

$$P = \int_{\mathcal{D}} I_{\mathcal{G}}(\mathbf{z}) p(\mathbf{z}) d\mathbf{z} = \int_{\mathcal{D}} I_{\mathcal{G}}(\mathbf{z}) \frac{p(\mathbf{z})}{q(\mathbf{z})} q(\mathbf{z}) d\mathbf{z} \tag{15}$$

is well defined, where $p(\mathbf{z})/q(\mathbf{z})$ is the *likelihood ratio*—in the context of importance sampling also called *importance weight*. The importance-sampling estimate of the failure probability P then draws n realizations $\mathbf{z}'_1, \dots, \mathbf{z}'_n$ of the random variable Z' with density q and evaluates

$$P_n^{\text{IS}} = \frac{1}{n} \sum_{i=1}^{n} I_{\mathcal{G}}(\mathbf{z}'_i) \frac{p(\mathbf{z}'_i)}{q(\mathbf{z}'_i)}. \tag{16}$$

The variance of the importance sampling estimator is

$$\mathbb{V}[P_n^{\text{IS}}] = \frac{\sigma_q^2}{n}, \tag{17}$$

where

$$\sigma_q^2 = \int_{\mathcal{D}} \left(\frac{I_{\mathcal{G}}(\mathbf{z}') p(\mathbf{z}')}{q(\mathbf{z}')} - P \right)^2 q(\mathbf{z}') d\mathbf{z}'. \tag{18}$$

If $\text{supp}(p) \subseteq \text{supp}(q)$, and by using (15), one can show that the importance sampling estimator P_n^{IS} is an unbiased estimator of the failure probability, i.e.,

$$\mathbb{E}_q[P_n^{\text{IS}}] = \mathbb{E}_p[I_{\mathcal{G}}(Z)] = P.$$

The importance sampling estimator P_n^{IS} has mean P and variance σ_q^2/n, and by the central limit theorem converges in distribution to the normal random variable $\mathcal{N}(P, \sigma_q^2/n)$.

Constructing a good biasing density that leads to small σ_q^2 is challenging [Owen, 2013]. We next introduce low-fidelity surrogate models, which are then used to construct biasing densities.

3.2.2 Multifidelity Importance Sampling (MFIS)

Recall that by $f : \mathcal{D} \mapsto \mathbb{R}^{d'}$ we denote an expensive-to-evaluate model of high fidelity with corresponding quantity of interest $f(\mathbf{z}) \in \mathbb{R}^{d'}$. Let k surrogates

$$f^{(i)} : \mathcal{D} \mapsto \mathbb{R}^{d'},\ i = 1, \ldots, k$$

of lower fidelities be available, which are faster to evaluate than the high-fidelity model $f(\cdot)$. We do not assume any information about the accuracy of the $f^{(i)}(\cdot)$ with respect to the high-fidelity model $f(\cdot)$. Section 4.2 and 5.3 detail the specific surrogate models used for the respective applications.

We use the MFIS method from [Peherstorfer et al., 2016] to obtain k estimators of the failure probability P. First, MFIS evaluates the surrogate models $f^{(i)}$ at m_i samples to obtain a surrogate-model specific failure set $\mathcal{G}^{(i)}$. Second, MFIS computes a biasing density q_i by fitting a distribution in form of a Gaussian mixture model to the parameters in the failure set. We use the expectation-minimization algorithm to find the best-fit Gaussian mixture model. We refer to [Peherstorfer et al., 2016] for details of the MFIS method. The above steps produce k biasing densities $q_1, \ldots, q_k$.

Let $\mathbf{z}_{i,j},\ j = 1, \ldots, n_i$ be independent samples from the density q_i. The importance sampling estimate of the failure probability P with n_i samples is

$$P_{n_i}^{\text{IS}} = \frac{1}{n_i} \sum_{j=1}^{n_i} I_{\mathcal{G}}(\mathbf{z}_{i,j}) \frac{p(\mathbf{z}_{i,j})}{q_i(\mathbf{z}_{i,j})}, \qquad i = 1, \ldots, k. \tag{19}$$

The variance of the importance sampling estimator is given by (17) with $n = n_i$ and $\sigma_q = \sigma_{q_i}$, with $\sigma_{q_i}^2$ being the asymptotic variance from (18) with $q = q_i$. The importance sampling estimate (19) requires evaluating the high-fidelity model at n_i (biased) samples. While not required, we use $n_i = n/k,\ i = 1, \ldots, k$ to distribute the computational load evenly. This results in k estimators $P_{n_i}^{\text{IS}}$ for $i = 1 \ldots, k$. The fusion framework then assigns the largest weight to the estimator with the lowest variance.

3.2.3 Fused multifidelity importance sampling—Complete Algorithm

We solve Problem 1 in the context of sampling-based failure probability estimators so that $P_i = P_{n_i}^{\text{IS}}$. Our proposed method optimally fuses the k MFIS estimators from (19), such that

$$P_{\boldsymbol{\alpha}} = \sum_{i=1}^{k} \alpha_i P_{n_i}^{\text{IS}}, \tag{20}$$

with the optimal weights chosen as in Proposition 2 and $\sum_{i=1}^{k} n_i = n$. Algorithm 1 describes the computational procedure; we denote sampling-based estimates as $\hat{P}_{n_i}^{\mathrm{IS}}$, which are realizations of the estimator $P_{n_i}^{\mathrm{IS}}$.

Algorithm 1 Computing failure probability estimate $\hat{P}_{\boldsymbol{\alpha}}$ via fused importance sampling

---

Input: Nominal distribution p, biasing distributions $\{q_i\}_{i=1}^{k}$, # of evaluations $\{n_i\}_{i=1}^{k}$, limit state function $g(\cdot)$.

Output: Failure probability estimate $\hat{P}_{\boldsymbol{\alpha}}$ and variance estimate $\mathbb{V}[\hat{P}_{\boldsymbol{\alpha}}]$

1: **for** $j = 1 : k$ **do** {Loop over all surrogates}

2: Draw $\mathbf{z}_{j,1}, \ldots, \mathbf{z}_{j,n_j}$ independent realizations from Z_j with density q_j and compute

$$\hat{P}_{n_j}^{\mathrm{IS}} = \frac{1}{n_j} \sum_{i=1}^{n_j} I_{\mathcal{G}}(\mathbf{z}_{j,i}) \frac{p(\mathbf{z}_{j,i})}{q_j(\mathbf{z}_{j,i})} \tag{21}$$

3: Compute the sample variances

$$\hat{\sigma}_{q_j}^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} \left(I_{\mathcal{G}}(\mathbf{z}_{j,i}) \frac{p(\mathbf{z}_{j,i})}{q_j(\mathbf{z}_{j,i})} - \hat{P}_{n_j}^{\mathrm{IS}} \right)^2 \tag{22}$$

4: **end for**

5: Define the vector $\boldsymbol{P} = [\hat{P}_{n_1}^{\mathrm{IS}}, \ldots, \hat{P}_{n_k}^{\mathrm{IS}}]^T$

6: Let $\hat{\boldsymbol{\Sigma}} = \mathrm{diag}(\hat{\sigma}_{q_1}^2/n_1, \ldots, \hat{\sigma}_{q_k}^2/n_k)$

7: Compute the fused estimate as in (9):

$$\hat{P}_{\boldsymbol{\alpha}} = \frac{\mathbf{1}_k^T \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{P}}{\mathbf{1}_k^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}_k}, \qquad \mathbb{V}[\hat{P}_{\boldsymbol{\alpha}}] = \frac{1}{\mathbf{1}_k^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}_k} \tag{23}$$

The estimator $P_{\boldsymbol{\alpha}}$ fuses the k estimators $P_{n_i}^{\mathrm{IS}}$ with biasing distribution q_i, with estimator i computed from n_i samples. Thus $P_{\boldsymbol{\alpha}}$ uses $n = \sum_{i=1}^{k} n_i$ samples. We now discuss how $P_{\boldsymbol{\alpha}}$ compares to a single importance sampling estimator with n samples. Consider the estimator $P_{j'}^{\mathrm{IS}}$ that uses n samples drawn from a single biasing density $q_{j'}$ for $j' \in \{1, \ldots, k\}$. This estimator requires selection of the best biasing density, a formidable task. The next results compares $P_{\boldsymbol{\alpha}}$ and $P_{j'}^{\mathrm{IS}}$, and gives a criterion for which the former has lower variance than the latter.

Proposition 3. *Let k estimators $P_{n_i}^{IS}$ with $n_1 = n_2 = \ldots = n_k$ samples be given. Let $j' \in \{1, \ldots, k\}$, and $q_{j'}$ be a biasing density that is used to derive an IS estimator $P_{j'}^{IS}$ with $n = kn_1$ samples. If*

$$k - 1 < \sum_{i \neq j'} \frac{\sigma_{q'_j}^2}{\sigma_{q_i}^2}, \tag{24}$$

then the variance of the fused estimator $P_{\boldsymbol{\alpha}}$ in (20) with n samples is smaller than the variance of the estimator with biasing density $q_{j'}$ with n samples, i.e.,

$$\mathbb{V}[P_{\boldsymbol{\alpha}}] < \mathbb{V}[P_{j'}^{IS}]. \tag{25}$$

Proof. Set $n_i = n/k,\ i = 1, \dots, k$, so that all estimators use the same number of samples. According to equation (13),

$$\mathbb{V}[P_{\boldsymbol{\alpha}}] = \frac{1}{\sum_{i=1}^{k} \frac{n_i}{\sigma_{q_i}^2}} = \frac{k}{n \sum_{i=1}^{k} \frac{1}{\sigma_{q_i}^2}} \tag{26}$$

as well as $\mathbb{V}[P_{j'}^{\mathrm{IS}}] = \frac{\sigma_{q_{j'}}^2}{n}$, so that

$$\mathbb{V}[P_{\boldsymbol{\alpha}}] = \frac{k}{n \sum_{i=1}^{k} \frac{1}{\sigma_{q_i}^2}} < \frac{\sigma_{q_{j'}}^2}{n} = \mathbb{V}[P_{j'}^{\mathrm{IS}}] \tag{27}$$

$$\Leftrightarrow \quad k < \sigma_{q_{j'}}^2 \sum_{i=1}^{k} \frac{1}{\sigma_{q_i}^2} \tag{28}$$

$$\Leftrightarrow \quad k - 1 < \sum_{i \neq j'}^{k} \frac{\sigma_{q_{j'}}^2}{\sigma_{q_i}^2}. \tag{29}$$

□

3.3 Error measures and practical computation

The failure probability estimate $\hat{P}_{n_i}^{\mathrm{IS}}$ is computed as in (21) and the sample variance $\hat{\sigma}_{q_i}^2$ as in (22). The root-mean-squared-error (RMSE) of the estimate $\hat{P}_{n_i}$ is

$$e^{\mathrm{RMSE}}(\hat{P}_{n_i}) = \sqrt{\frac{\hat{\sigma}_{q_i}^2}{n_i}}, \tag{30}$$

and the relative mean-squared-error, or coefficient of variation is computed as

$$e^{\mathrm{CV}}(\hat{P}_{n_i}) = \sqrt{\frac{\hat{\sigma}_{q_i}^2}{n_i (\hat{P}_{n_i}^{\mathrm{IS}})^2}}. \tag{31}$$

4 Test case: Convection-diffusion-reaction

We first consider a PDE model whose solution can be numerically evaluated with moderate computational cost. Here, we demonstrate the asymptotic behavior of our method because we can afford to sample the high-fidelity model $n = 10^5$ times, which will be too costly for the model in Section 5. The test problem is the convection-diffusion-reaction PDE introduced in Section 4.1. Its discretizations and reduced-order models are described in Section 4.2. Numerical results are presented in Section 4.3.

4.1 Convection-diffusion-reaction PDE model

We consider a simplified model of a premixed combustion flame at constant and uniform pressure, and follow the notation and setup in [Buffoni and Willcox, 2010, Sec.3]. The model includes a one-step reaction of the species

$$2H_2 + O_2 \ \rightarrow 2H_2O$$

in the presence of an additional non-reactive species, nitrogen. The physical combustor domain is $18mm$ in length (x-direction), and $9mm$ in height (y-direction), as shown in Figure 1.

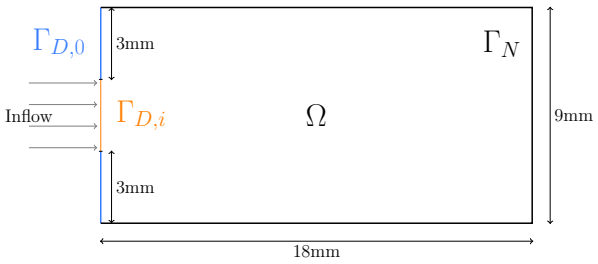


Figure 1: Set-up of combustor, with details of the boundary conditions in Table 1.

The velocity field U is set to be constant in the positive x direction, and divergence free. The molecular diffusivity κ is modeled as constant, equal and uniform for all species. The partial differential equation model is given by

$$0 = \kappa \Delta s - U \nabla s + \mathcal{F}(s, \mathbf{z}) \quad \in \Omega \tag{32}$$

where the state is comprised of the components $s = [T, Y_{H_2}, Y_{O_2}, Y_{H_2O}]$, with the Y_i being the mass fractions of the species (fuel, oxidizer, product), and T denoting the temperature. Referring to Figure 1, we have that $\Gamma_D = \Gamma_{D,i} \cup \Gamma_{D,0}$ is the Dirichlet part of the boundary and Γ_N combines the top, bottom and right boundary, where Neumann conditions are prescribed. In sum, $\partial\Omega = \Gamma_D \cup \Gamma_N$; the boundary conditions are imposed as given in Table 1. The nonlinear reaction term $\mathcal{F}(s, \mathbf{z}) = [\mathcal{F}_T, \mathcal{F}_{H_2}, \mathcal{F}_{O_2}, \mathcal{F}_{H_2O}](s, \mathbf{z})$ is of Arrhenius type [Cuenot and Poinsot, 1996], and modeled as

$$\mathcal{F}_i(s, \mathbf{z}) = -\nu_i \left(\frac{W_i}{\rho}\right) \left(\frac{\rho Y_F}{W_F}\right)^{\nu_F} \left(\frac{\rho Y_O}{W_O}\right)^{\nu_O} A \exp\left(-\frac{E}{RT}\right), \quad i = H_2,\ O_2,\ H_2O \tag{33}$$

$$\mathcal{F}_T(s, \mathbf{z}) = Q\ \mathcal{F}_{H_2O}(s, \mathbf{z}). \tag{34}$$

The parameters of the model are defined in Table 2. The uncertain parameters are the pre-exponential factor A and the activation energy E of the Arrhenius model. The domain for these parameters is denoted as $\mathcal{D}$. In particular, we have that

$$\mathbf{z} = [A, E] \in \mathcal{D} = [5.5 \times 10^{11},\ 1.5 \times 10^{13}] \times [1.5 \times 10^{3},\ 9.5 \times 10^{3}].$$

Table 1: Boundary conditions for the combustion model from [Buffoni and Willcox, 2010].

Boundary	Temperature	Species
$\Gamma_{D,i}$	$T = 950\text{K}$	$Y_{H_2} = 0.0282,\ Y_{O_2} = 0.2259,\ Y_{H_2O} = 0$
$\Gamma_{D,0}$	$T = 300\text{K}$	$Y_{H_2} = 0,\ Y_{O_2} = 0,\ Y_{H_2O} = 0$
Γ_N	$\nabla T \cdot \boldsymbol{n} = 0$	$\nabla Y_i \cdot \boldsymbol{n} = 0$

Table 2: Parameters for the combustion model from [Buffoni and Willcox, 2010].

quantity	physical meaning	assumptions	value
κ	molecular diffusivity	const., equal, uniform $\forall i$	$2.0\frac{\text{cm}^2}{\text{s}}$
U	velocity	const.	$50\frac{\text{cm}}{\text{s}}$
W_{H_2}	molecular weight	const.	$2.016\frac{\text{g}}{\text{mol}}$
W_{O_2}	molecular weight	const.	$31.9\frac{\text{g}}{\text{mol}}$
W_{H_2O}	molecular weight	const.	$18\frac{\text{g}}{\text{mol}}$
ρ	density of mixture	const.	$1.39 \times 10^{-3}\frac{\text{g}}{\text{cm}^3}$
R	univ. gas constant	const.	$8.314472\frac{\text{J}}{\text{mol K}}$
Q	heat of reaction	const.	9800K
ν_{H_2}	stochiometric coefficient	const.	2
ν_{O_2}	stochiometric coefficient	const.	1
ν_{H_2O}	stochiometric coefficient	const.	2

4.2 Discretization and reduced-order models

The model is discretized using a Finite Difference approximation in two spatial dimensions, with 72 nodes in x direction, and 36 nodes in y direction, leading to $10,804$ unknowns in the model. The nonlinear system is solved with Newton's method. Let $\mathbf{T}(\mathbf{z})$ be the vector with components corresponding to the approximations of the temperature $T(x, y; \mathbf{z})$ at the grid points. The high-fidelity model (HFM) is $f : \mathcal{D} \mapsto \mathbb{R}$ and the quantity of interest is the maximum temperature over all grid points:

$$f(\mathbf{z}) = \max\ \mathbf{T}(\mathbf{z}).$$

Reduced-order models provide a powerful framework to obtain surrogates for expensive-to-evaluate models. In the case of nonlinear systems, reduced-order models can be obtained via reduced-basis methods [Jan S Hesthaven et al., 2016], dynamic mode decomposition [Kutz et al., 2016], proper orthogonal decomposition [Berkooz et al., 1993], and many others; for a survey, see [Benner et al., 2017]. Here, we compute reduced-order models $f^{(i)}$ for our multifidelity approach via Proper Orthogonal Decomposition and the Discrete Empirical Interpolation Method (DEIM) for an efficient evaluation of the nonlinear term. The training snapshots are generated from solutions to the high-fidelity model on a parameter grid of 50×50 equally spaced values $\mathbf{z} \in \mathcal{D}$. The three surrogate models are built from $2, 10, 15$ POD basis vectors, and accordingly $2, 5, 10$ DEIM interpolation

points. The corresponding models are denoted as ROM1, ROM2, ROM3, respectively. We denote by $\mathbf{T}_r^{(i)}(\mathbf{z})$ the approximation to the temperature $T(x, y; \mathbf{z})$ via the ith ROM. The surrogate models $f^{(i)}$ are the mappings $f^{(i)} : \mathcal{D} \mapsto \mathbb{R}$ with corresponding quantity of interest denoted as

$$f^{(i)}(\mathbf{z}) = \max \ \mathbf{T}_r^{(i)}(\mathbf{z}), \quad i = 1, \ldots, k.$$

We refer the reader to [Buffoni and Willcox, 2010] for more details on the discretization and ROM construction for this convection-diffusion-reaction model.

4.3 Results for multifidelity fusion of failure probabilities

We define a failure of the system when the maximum temperature in the combustor exceeds 2430K, so that the limit state function is

$$g(f(\mathbf{z})) = 2430 - f(\mathbf{z}), \tag{35}$$

and likewise for the reduced-order models $g(f^{(i)}(\mathbf{z})) = 2430 - f^{(i)}(\mathbf{z})$.

To compute the biasing densities, we draw $\hat{m} = 20,000$ samples from the uniform distribution on $\mathcal{D}$, compute surrogate-based solutions, and evaluate the limit state function for those solutions. If the limit state function indicates failure of the system for a solution obtained from the ith surrogate model, the corresponding parameter is added to $\mathcal{G}^{(i)}$, the failure set computed from the ith surrogate model. Next, we compute the biasing densities q_1, q_2, q_3 via MFIS (see Section 3.2.2) as Gaussian mixture distributions with a single component.

For reference purposes, a biasing density is constructed by performing the same steps as above using the HFM with $\hat{m} = 20,000$ samples. Based on this density, we compute an importance sampling estimate of the failure probability with $n = 10^5$ samples, resulting in $\hat{P}_{10^5}^{\mathrm{IS}} = 2.3707 \times 10^{-4}$.

In Figure 2 we show the quantity of interest, i.e., the maximum temperature. The plots are obtained by generating $m = 10^5$ samples from the nominal distribution (left) and the respective biasing distributions (right), and evaluating the HFM at those samples. Figure 2, left, shows that the typical range of the quantity of interest is between approximately 1200K and 2440K. However, only the events where the quantity of interest is above 2430K are relevant for the failure probability computation. By using the biasing distributions in Figure 2, right, a large portion of the outputs leads to a failure of the system. This indicates that the biasing distributions are successful in generating samples at the failure region of the high-fidelity model.

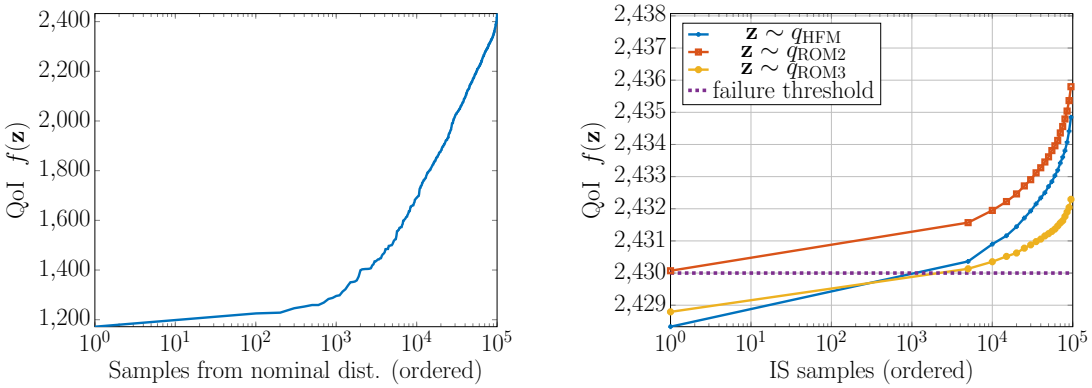


Figure 2: Quantity of interest $f(\mathbf{z})$ in [K] of HFM ordered by magnitude versus # of samples $\mathbf{z}$, for $m = 10^5$ samples. Left: Samples are from the nominal (uniform) distribution. Right: The parameter samples are drawn from different biasing distributions (biased towards failure above 2430K). This demonstrates that the biasing distributions are good since the outputs are largely above the failure threshold. Here, ROM1 did not have any parameters in the failure domain, and hence defaulted to being the nominal distribution and is therefore not plotted.

We compare the computational cost of computing the biasing distributions from the various ROMs and the HFM in Figure 3. Computing a biasing density using the high-fidelity model with $m = 10^5$ samples costs approximately 11.4 hours in CPU time. By evaluating the low-fidelity models ROM2 and ROM3 at $m = 10^5$ samples to construct the biasing density, we reduce the computational time by three orders of magnitude, i.e. to approximately one minute in CPU time. Note, that ROM1 is the reduced-order model that is cheapest to execute per model evaluation, but it is also the least accurate. In our case, ROM1 did not produce any samples in the failure region, even after $m = 5 \times 10^5$ samples. This higher number of samples led to increased computational cost to compute the biasing distribution in Figure 3. It is not unexpected that ROM1 is so inaccurate, since only two POD modes are not enough to resolve the important character of this problem. ROM1 is included to demonstrate how the fusion approach can be effective even in the presence of highly inaccurate surrogate models.

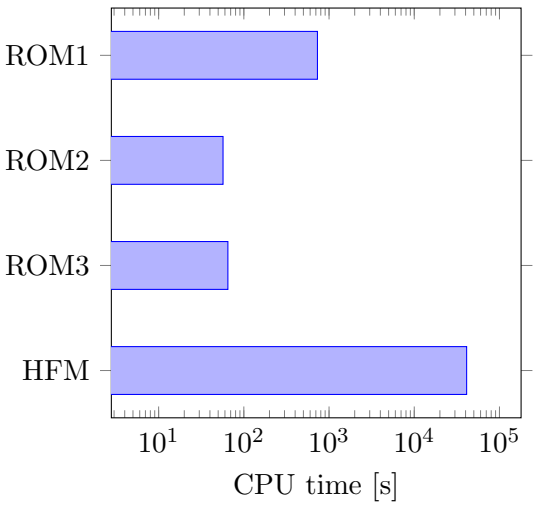


Figure 3: CPU time to generate the biasing densities from three reduced-order models described in Section 4.2. HFM, ROM2 and ROM3 were evaluated at $m = 10^5$ and produced enough samples to fit a good biasing distribution. ROM1 was evaluated $m = 5 \times 10^5$ times, and still did not produce any samples in the failure region.

To assess the quality of the fused estimator $P_{\boldsymbol{\alpha}}$, we consider the error measures introduced in Section 3.3. In Figure 4, left, we show the root mean-squared error of the importance sampling estimators $\hat{P}_{n_i}^{\mathrm{IS}}$ as well as the combined estimator $\hat{P}_{\boldsymbol{\alpha}}$. Figure 4, right, shows the coefficient of variation defined in (31) for the estimators. The fused estimator is competitive in RMSE and coefficient of variation with the estimator using the high-fidelity biasing density, but comes at a much cheaper computational cost.

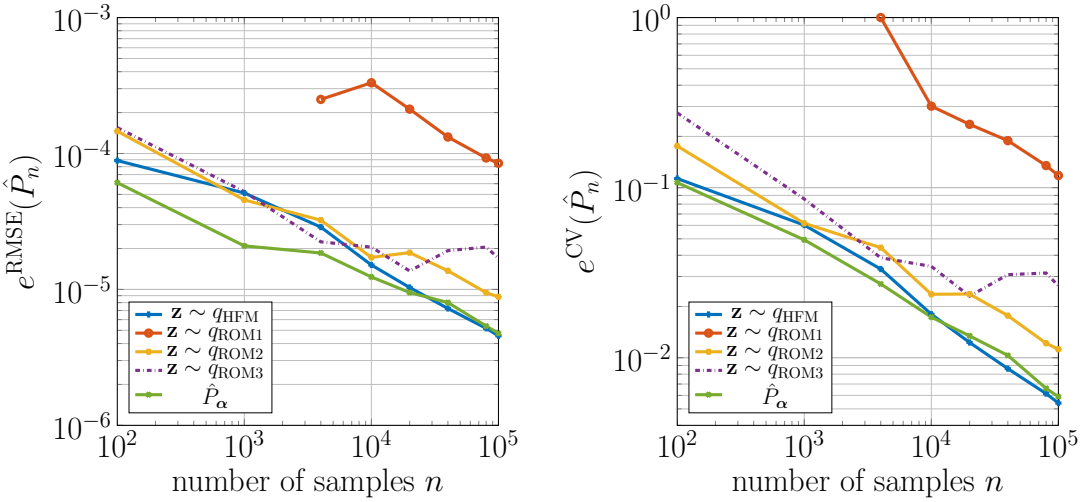


Figure 4: Left: Root mean-squared error from (30); Right: Coefficient of variation as defined in (31) for the convection-diffusion-reaction simulation.

5 Failure probability estimation related to a free plane jet

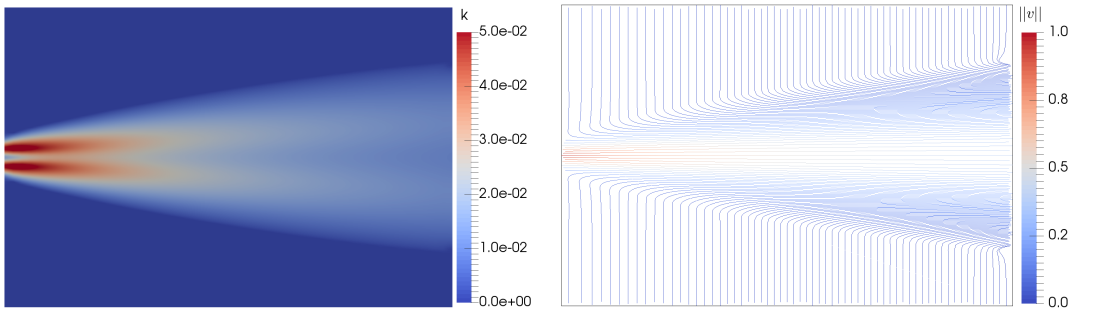
We apply the proposed fusion of estimators to quantify the influence of uncertain parameters at the inlet of a free plane jet on a measure of turbulent mixing. This is a challenging problem, since it involves an expensive-to-evaluate model for which the naive computation of low probabilities requires thousands of hours of computation. We reduce this number significantly with our multifidelity importance sampling framework via a fusion of estimators.

The remainder of this section is organized as follows. Section 5.1 introduces the free plane jet, followed by details of the model and its governing equations in Section 5.2. In Section 5.3 we discuss the low-fidelity surrogate models we use in our fusion of estimators. Finally, we present the results for multifidelity fusion of small probability estimators in Section 5.4.

5.1 Large-scale application: Free plane jet

Free turbulent jets are prototypical flows believed to represent the dynamics in many engineering applications, such as combustion and propulsion. As such, free jet flows are the subject of several experimental [Gutmark and Wygnanski, 1976, Gutmark et al., 1978, Krothapalli et al., 1981] and numerical investigations [Zhou et al., 1999, Ribault et al., 1999, Stanley et al., 2002, Klein et al., 2003, Klein et al., 2015] and constitute an important benchmark for turbulent flows. These investigations show that the evolution of the jet is very dependent on the conditions at the inlet.

Here we focus our attention on a free plane jet flow at Reynolds number ranging from 5,000 to 15,000. Figure 5 shows a flow field typical of the cases considered here. Our



(a) Contours of turbulent kinetic energy.

(b) Streamlines colored by the intensity of the velocity.

Figure 5: Flow field of a two-dimensional plane jet at Reynolds number 10,000, computed with a RANS model.

expensive-to-evaluate model is based on the two-dimensional incompressible Reynolds-averaged Navier-Stokes (RANS) equations. Although this model is incapable of resolving all relevant turbulent features of the flow, it still represents a challenging large-scale application for the computation of small probabilities.

In particular, we investigate the influence of uncertain parameters at the inlet of the jet on the amount of turbulent mixing produced by the jet, with the caveat that this is the mixing predicted by our RANS model. We quantify the mixing using a relatively simple metric: the width of the jet. We consider two uncertain parameters: the intensity of the velocity and the mixing length at the inlet boundary. The first is a physical parameter that reflects uncertainty in the setup of the flow (e.g., operating conditions of an engine). Since we keep other physical parameters constant, by varying the velocity intensity we effectively change the Reynolds number of the flow. The second parameter is part of the $k - \epsilon$ turbulence model used to describe the dynamics of the flow, as detailed in Section 5.2. Thus, this parameter reflects an uncertainty due to model inadequacy.

5.2 Modeling and governing equations

We consider a free plane jet in conditions similar to the ones reported in [Klein et al., 2003, Klein et al., 2015]. Namely, the flow exits a rectangular nozzle into quiescent surroundings with a prescribed top-hat velocity profile and turbulence intensity. The nozzle has width D, and is infinite along the span-wise direction. The main difference between the free plane jet we considered here and the one described in [Klein et al., 2003, Klein et al., 2015] is the Reynolds number at the exit nozzle. Here, the Reynolds number varies between 5,000 and 15,000.

Our simulation model computes the flow in a rectangular domain Ω located at a

distance $5D$ downstream from the exit of the jet nozzle, as illustrated in Figure 6. By doing so, modeling the conditions at the exit plane of the jet nozzle is avoided. Instead, direct numerical simulation data are used to define inlet conditions at the surface Γ_{in}. The dynamics are modeled with the incompressible Reynolds-averaged Navier-Stokes

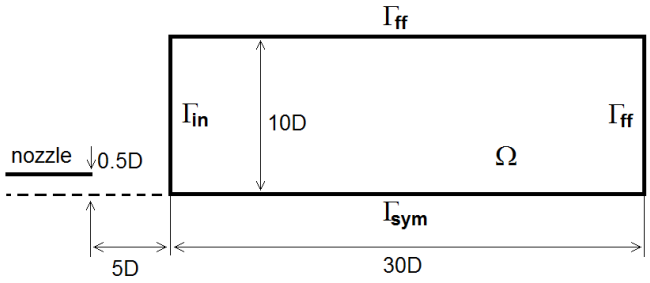


Figure 6: Illustration of the free plane jet setup. The diameter of the nozzle is denoted by D. The simulation domain Ω is composed of a $30D \times 10D$ box situated at a distance $5D$ downstream to the nozzle exit.

equations, complemented by the $k-\epsilon$ turbulence model [Launder and Spalding, 1974]:

$$(\boldsymbol{v}\cdot\nabla)\boldsymbol{v} + \frac{1}{\rho}\nabla p - \nabla\cdot((\nu+\nu_t)\bar{\bar{S}}(\boldsymbol{v})) = 0, \tag{36}$$

$$\nabla\cdot\boldsymbol{v} = 0, \tag{37}$$

$$\boldsymbol{v}\cdot\nabla k - 2\nu_t(\bar{\bar{S}}(\boldsymbol{v}):\bar{\bar{S}}(\boldsymbol{v})) + \epsilon - \nabla\cdot\left(\left(\nu+\frac{\nu_t}{\sigma_k}\right)\nabla k\right) = 0, \tag{38}$$

$$\boldsymbol{v}\cdot\nabla\epsilon - 2C_{1\epsilon}\frac{\epsilon\nu_t}{k}(\bar{\bar{S}}(\boldsymbol{v}):\bar{\bar{S}}(\boldsymbol{v})) + C_{2\epsilon}\frac{\epsilon^2}{k} - \nabla\cdot\left(\left(\nu+\frac{\nu_t}{\sigma_\epsilon}\right)\nabla\epsilon\right) = 0, \tag{39}$$

where $\boldsymbol{v} = [v_x, v_y]$ denotes the velocity vector, p denotes pressure, $\rho = 1.0$ kg/m^3 is the density, $\nu = 0.0001$ m^2/s is the kinematic viscosity, and $\bar{\bar{S}}$ is the strain rate tensor given by

$$\bar{\bar{S}}(\boldsymbol{v}) = \frac{1}{2}(\nabla\boldsymbol{v} + (\nabla\boldsymbol{v})^T).$$

In the $k-\epsilon$ turbulence model, k denotes the turbulent kinetic energy, ϵ denotes the turbulent dissipation, ν_t denotes the turbulent kinematic viscosity,

$$\nu_t = C_\mu\frac{k^2}{\epsilon},$$

and the constants[1] of the model are

$$C_\mu = 0.09 \qquad \sigma_k = 1.00, \qquad \sigma_\epsilon = 1.30, \qquad C_{1\epsilon} = 1.44, \qquad \text{and} \qquad C_{2\epsilon} = 1.92.$$

[1]We use σ_k and σ_ϵ here as model constants, which is typical notation in this community. These are only used in this section, and throughout the paper σ's are variances.

At the inlet surface Γ_{in} Dirichlet boundary conditions are imposed. Data obtained by the direct numerical simulation described in [Klein et al., 2015] are used to determine reference inlet profiles for velocity, $\boldsymbol{v}_{\text{ref}}$, and for turbulent kinetic energy, k_{ref}. Furthermore, inlet conditions are allowed to vary by defining a velocity intensity (U) scale, which is applied to the reference profiles. Finally, turbulent dissipation at the inlet is estimated by assuming a mixing length model. Thus, the boundary conditions at the inlet surface are given by

$$\boldsymbol{v}|_{\Gamma_{\text{in}}} = U\boldsymbol{v}_{\text{ref}}, \qquad k|_{\Gamma_{\text{in}}} = U^2 k_{\text{ref}}, \qquad \epsilon|_{\Gamma_{\text{in}}} = C_\mu \frac{k^{3/2}}{\ell_m},$$

where ℓ_m denotes the mixing length parameter.

At the symmetry axis surface, Γ_{sym}, no-flux boundary conditions are imposed through a combination of Dirichlet and Neumann conditions of the form

$$v_y|_{\Gamma_{\text{sym}}} = 0, \qquad \left.\frac{\partial v_x}{\partial n}\right|_{\Gamma_{\text{sym}}} = 0, \qquad \left.\frac{\partial k}{\partial n}\right|_{\Gamma_{\text{sym}}} = 0, \qquad \left.\frac{\partial \epsilon}{\partial n}\right|_{\Gamma_{\text{sym}}} = 0.$$

Finally, at the surface Γ_{ff} "far-field" conditions that allow the entrainment of air around the jet are imposed through weak Dirichlet conditions, as detailed in [Villa and Marques, 2017].

The complete model includes additional features that make it more amenable to numerical discretization. The most delicate issue in the solution of the RANS model is the possible loss of positivity of the turbulence variables. To avoid this issue, we introduce an appropriately mollified (and thus smoothly differentiable) max function to ensure positivity of k and ε. In addition, if inflow is detected at any point on the far-field boundary, the boundary condition is switched from Neumann to Dirichlet by means of a suitably mollified indicator of the inflow region. Finally, we stabilize the discrete equations using a strongly consistent stabilization technique (Galerkin Least Squares, GLS, stabilization) to address the convection-dominated nature of the RANS equations. The complete formulation is shown in [Villa and Marques, 2017].

The model equations described above are solved numerically using a finite element discretization. The discretization is implemented in FEniCS [Alnæs et al., 2015] by specifying the weak form of the residual, including the GLS stabilization and mollified versions of the positivity constraints on k and ϵ and the switching boundary condition on the outflow boundary. To solve the nonlinear system of equations that arise from the finite element discretization, we employ a damped Newton method. The bilinear form of the state Jacobian operator is computed using FEniCS's symbolic differentiation capabilities. Finally, we use pseudo-time continuation to guarantee global convergence of the Newton method to a physically stable solution (if such solution exists) [Kelley and Keyes, 1998]. The finite element solver is detailed in [Villa and Marques, 2017].

In this investigation, the uncertain parameters are considered to be the velocity intensity (U) and the mixing length (ℓ_m) at the inlet surface:

$$\mathbf{z} = [U, \ell_m].$$

The parameter domain is $\mathbf{z} \in \mathcal{D} = [0.5, 1.5] \times [0.05, 0.15]$, and the nominal distribution of parameters is assumed to be uniform in $\mathcal{D}$. Figure 7 illustrates a typical solution

behavior for this turbulent jet by plotting contours of the turbulent kinetic energy for selected samples in $\mathcal{D}$.

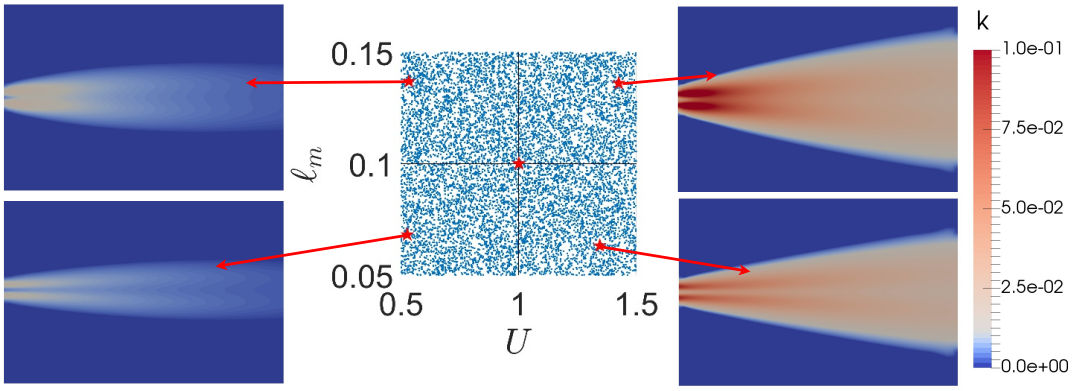


Figure 7: Samples of the flow solution computed in different regions of the input parameter space $\mathcal{D}$. The plots show the turbulent kinetic energy.

The quantity of interest is the width of jet measured at $x = 27.5D$:

$$w(\boldsymbol{v}; \mathbf{z}) = \frac{1}{v_{x_0} D} \int_0^{10D} v_x(x = 27.5D, y; \mathbf{z})\, dy, \tag{40}$$

where $v_{x_0} = v_x(x = 27.5D, y = 0; \mathbf{z})$.

5.3 Simplified-physics surrogate models

We consider four surrogate models to represent the dynamics of the free plane jet flow. The models are based on two distinct computational grids (fine and coarse), and on two representations of turbulent effects. The fine computational grid contains 10,000 elements and 5,151 nodes, while the coarse grid contains 2,500 elements and 1,326 nodes. Furthermore, the models are based either on the complete $k - \epsilon$ turbulence model described in the previous section, or on a prescribed turbulent viscosity field. In the latter case, the turbulent viscosity field is estimated by a bilinear interpolation based on 81 conditions that span the input parameter space $\mathcal{D}$ uniformly. At each of these 81 conditions, the turbulent viscosity field is computed with the $k - \epsilon$ turbulence model and the fine computational grid. The following four low-fidelity models are increasingly complex in terms of either modeled physics or grid resolution:

- **LFM1–CI**: Coarse, interpolated; combines the interpolated turbulence viscosity field with the coarse computational grid (3,978 degrees of freedom); average computational time 13.9s
- **LFM2–FI**: Fine, interpolated; combines the interpolated turbulence viscosity field with the fine computational grid (15,453 degrees of freedom); average computational time 89.1s

- **LFM3–CKE**: Coarse $k-\epsilon$; combines the $k-\epsilon$ turbulence model with the coarse computational grid (6,630 degrees of freedom); average computational time 105.3s
- **HFM**: High-fidelity model; combines the $k-\epsilon$ turbulence model with the fine computational grid (25,755 degrees of freedom); average computational time 669.3s

Note that the models based on an interpolated turbulent viscosity field run approximately 7.5 times faster than the corresponding models based on the $k-\epsilon$ turbulence model. This speedup results from eliminating (38)–(39) from the governing equations, which leads to a reduction in the total number of degrees of freedom (elimination of variables k and ϵ) and simplifications in the numerical discretization.

Let $\boldsymbol{v}_i$, $i =$ HFM, LFM1, LFM2, LFM3, denote the velocity field computed with the models above. The high-fidelity model is the mapping from the inputs to the quantity of interest (jet width from (40)) for a velocity field computed with the most complex representation of the flow dynamics, $\boldsymbol{v}_{\text{HFM}}$:

$$f : \mathcal{D} \mapsto \mathbb{R}, \qquad f(\mathbf{z}) = w(\boldsymbol{v}_{\text{HFM}}; \mathbf{z}).$$

The surrogate models are defined in a similar fashion as

$$f^{(i)} : \mathcal{D} \mapsto \mathbb{R}, \qquad f^{(i)}(\mathbf{z}) = w(\boldsymbol{v}_i; \mathbf{z}), \quad i = \text{LFM1, LFM2, LFM3.}$$

5.4 Results for multifidelity fusion of small probability estimators

We define a design failure when the jet width is below 2.10. Hence, the limit state function is given by

$$g(f(\mathbf{z})) = f(\mathbf{z}) - 2.10. \tag{41}$$

We compute the biasing distributions q_i, $i = 1, 2, 3$ from the three low-fidelity surrogate models via MFIS (see Section 3.2.2). In each case, we draw $\hat{m} = 10,000$ parameter samples from the uniform distribution on $\mathcal{D}$ and evaluate the limit state function applied to the resulting quantity of interest. If the limit state function indicates failure of the system for a solution obtained from the ith surrogate model, the corresponding parameter is added to $\mathcal{G}^{(i)}$, the failure set computed from the ith surrogate model. We then fit a multivariate Gaussian to the samples in $\mathcal{G}^{(i)}$, resulting in the biasing densities q_1, q_2, q_3. As reference, we repeat the same process with the high-fidelity model, resulting in the biasing distribution q_{HFM}.

The CPU run times in hours on an Intel Core i7-2600 to compute the biasing densities via this approach are listed in Table 3. The significant time difference between the high-fidelity model and the lower-fidelity models can have important implications for engineering practice. For illustration, LFM1 would need 1.5 days in CPU time for the exploration, whereas LFM2 and LFM3 would need seven and twelve days of CPU time, respectively. These are considerable savings compared to using the high-fidelity model to get a good biasing distribution, which would require 2.5 months of CPU time in this case.

Table 3: CPU time in hours on an Intel Core i7-2600 for 10,000 model evaluations, and the number of samples in the failure domain. The samples are drawn from the uniform distribution on $\mathcal{D}$.

	LFM1	LFM2	LFM3	HFM
# of samples drawn	10,000	10,000	10,000	10,000
# of samples in failure domain $\mathcal{G}^{(i)}$	36	26	9	45
time needed	38.7[h]	247.6[h]	292.6[h]	1859.1[h]

Next, we investigate the quality of the biasing distributions. For reference, Figure 8, left, shows the result of 10^3 uniform sample evaluations with the four computational models. Note, that hardly any samples are below the failure threshold. In contrast, the quantity of interest computed from samples of the four biasing distributions is shown in Figure 8, right. Here, the situation reversed, and all models have more than 90% of the samples in the failure domain, contributing in turn to the evaluation of the failure probability from (19). Comparing the y-axis scaling of both figures, note that even the samples above the failure threshold in Figure 8, right, are relatively close to the threshold level. This leads to the conclusion that the biasing distributions are indeed biased towards failure of the samples, and therefore the multifidelity strategy provides a viable way of saving computational time to inform a biasing distribution.

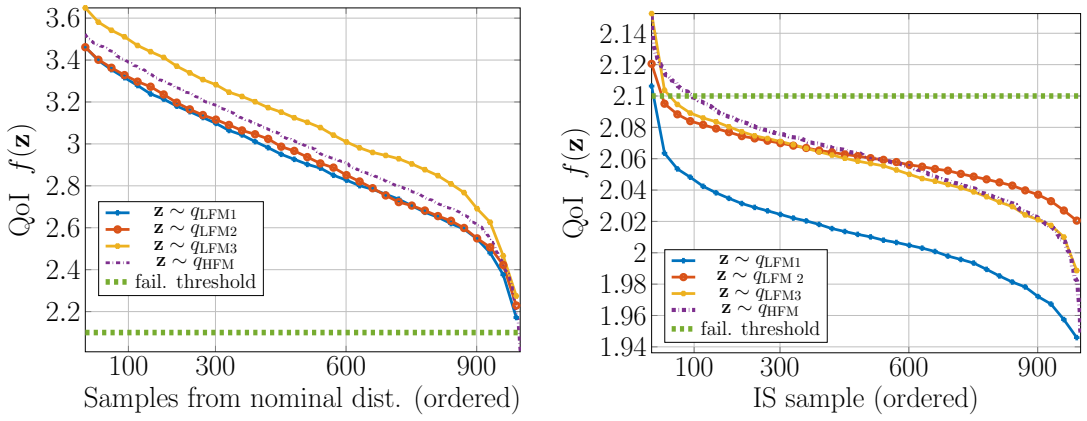


Figure 8: Quantity of interest, the width of the jet at $x = 27.5D$, for $n = 10^3$ samples. Left: The input parameters are drawn from the nominal distribution. Right: The input parameters are drawn from different biasing distributions (biased towards failure below 2.10). Note that the majority of outputs falls below the failure threshold, thus indicating the effectiveness of the biasing distributions.

The reference failure probability is computed via importance sampling with $n = 10,000$ samples drawn from the HFM biasing distribution and is $\hat{P}^{\text{IS}}_{10,000,q_{\text{HFM}}} = 3.5763 \times 10^{-3}$.

Table 4: Weights of the fused estimator $\hat{P}_{\boldsymbol{\alpha}}$ with n samples.

	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1000$
α_1	0.8210	0.8075	0.8202	0.8740	0.8662
α_2	0.1016	0.1044	0.0734	0.0637	0.0673
α_3	0.0773	0.0881	0.1064	0.0623	0.0664

We then compute the estimators $P_{n_i}^{\text{IS}}, i = 1, 2, 3$ with $n = 10^3$ samples from the biasing densities q_i from the three surrogate models. We compare these estimators with an estimator that uses $n = 10^3$ samples from the HFM biasing density q_{HFM}. We obtain the fused multifidelity estimator $P_{\boldsymbol{\alpha}}$ as described above in Algorithm 1 with $n_i = \lfloor n/3 \rfloor, i = 1, 2, 3$, samples by fusing the three surrogate-model-based importance sampling estimators. The fused estimator thus uses a total of $n = 10^3$ samples. The estimators and the error measures below are averaged over three independent runs. The root-mean-square-error from (30) and coefficient of variation (31) are shown in Figure 9. Table 4 shows the three weights for the fused estimator $\hat{P}_{\boldsymbol{\alpha}}$ as given in Proposition 2. As seen in that proposition, the estimates with the lowest variance (which is proportional to the RSME in Figure 9) get assigned the largest weights.

The fused estimator has the lowest root-mean-squared-error among the five computed estimators. While there is no guarantee that this is the case, we are always guaranteed to improve over the worst estimator, see Proposition 3. With respect to the coefficient of variation, the high-fidelity model gives the best improvement among all the models. The fused estimator is only slightly worse in coefficient of variation than the estimator with the high-fidelity biasing density, but is much cheaper.

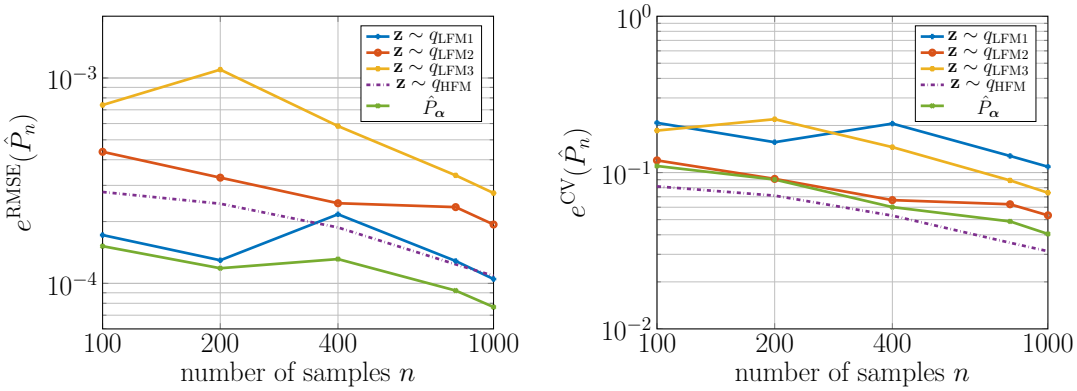


Figure 9: Left: Root-mean-squared error from (30); Right: Coefficient of variation as defined in (31) for the free plane jet application.

6 Conclusions

We enabled the estimation of small probabilities for expensive-to-evaluate models via a new combination of techniques from importance sampling, multifidelity modeling and information fusion. The effectiveness of the proposed approach is demonstrated on a convection-diffusion-reaction PDE, where asymptotic numerical results could be obtained. The strength of the proposed framework is then shown on the target application of the turbulent jet, a challenging problem for small-probability computation due to its high computational cost. The proposed framework was illustrated for the special case of importance-sampling based estimators, but in fact applies to a much broader class of estimators, as long as the estimators are unbiased. By fusing different estimators, we avoid the complicated biasing density selection problem. We also showed that this strategy always outperforms sampling from the worst biasing density. The numerical results suggest that the fused estimator is often comparable to an estimator that samples from the best biasing density only.

Acknowledgements

The authors thank Prof. M. Klein for sharing the DNS data in [Klein et al., 2015] with us. This work was supported by the Defense Advanced Research Projects Agency [EQUiPS program, award W911NF-15-2-0121, Program Manager F. Fahroo]; the Air Force [Center of Excellence on Multi-Fidelity Modeling of Rocket Combustor Dynamics, award FA9550-17-1-0195]; and the US Department of Energy, Office of Advanced Scientific Computing Research (ASCR) [Applied Mathematics Program, awards DE-FG02-08ER2585 and DE-SC0009297, as part of the DiaMonD Multifaceted Mathematics Integrated Capability Center].

References

[Alnæs et al., 2015] Alnæs, M. S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M. E., and Wells, G. N. (2015). The FEniCS project version 1.5. *Archive of Numerical Software*, 3(100).

[Benner et al., 2017] Benner, P., Ohlberger, M., Cohen, A., and Willcox, K. (2017). *Model Reduction and Approximation: Theory and Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

[Berkooz et al., 1993] Berkooz, G., Holmes, P., and Lumley, J. L. (1993). The proper orthogonal decomposition in the analysis of turbulent flows. *Annual review of fluid mechanics*, 25(1):539–575.

[Buffoni and Willcox, 2010] Buffoni, M. and Willcox, K. (2010). Projection-based model reduction for reacting flows. In *40th Fluid Dynamics Conference and Exhibit*, page 5008.

[Chen and Quarteroni, 2013] Chen, P. and Quarteroni, A. (2013). Accurate and efficient evaluation of failure probability for partial different equations with random input data. *Computer Methods in Applied Mechanics and Engineering*, 267:233–260.

[Clemen and Winkler, 1999] Clemen, R. T. and Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk analysis*, 19(2):187–203.

[Cuenot and Poinsot, 1996] Cuenot, B. and Poinsot, T. (1996). Asymptotic and numerical study of diffusion flames with variable Lewis number and finite rate chemistry. *Combustion and Flame*, 104(1):111–137.

[Elfverson et al., 2016] Elfverson, D., Hellman, F., and Mlqvist, A. (2016). A multilevel Monte Carlo method for computing failure probabilities. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):312–330.

[Fagerlund et al., 2016] Fagerlund, F., Hellman, F., Mlqvist, A., and Niemi, A. (2016). Multilevel Monte Carlo methods for computing failure probability of porous media flow systems. *Advances in Water Resources*, 94:498 – 509.

[Gutmark et al., 1978] Gutmark, E., Wolfshtein, M., and Wygnanski, I. (1978). The plane turbulent impinging jet. *Journal of Fluid Mechanics*, 88(4):737756.

[Gutmark and Wygnanski, 1976] Gutmark, E. and Wygnanski, I. (1976). The planar turbulent jet. *Journal of Fluid Mechanics*, 73(3):465495.

[Jan S Hesthaven et al., 2016] Jan S Hesthaven, J. S., Rozza, G., and Stamm, B. (2016). *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. Springer.

[Kelley and Keyes, 1998] Kelley, C. T. and Keyes, D. E. (1998). Convergence analysis of pseudo-transient continuation. *SIAM Journal on Numerical Analysis*, 35(2):508–523.

[Klein et al., 2003] Klein, M., Sadiki, A., and Janicka, J. (2003). Investigation of the influence of the Reynolds number on a plane jet using direct numerical simulation. *International Journal of Heat and Fluid Flow*, 24(6):785 – 794.

[Klein et al., 2015] Klein, M., Wolff, C., and Tangermann., E. (2015). A-priori analysis of a les model for scalar flux based on interscale energy transfer. In *9th International Symposium on Turbulence and Shear Flow*.

[Krothapalli et al., 1981] Krothapalli, A., Baganoff, D., and Karamcheti, K. (1981). On the mixing of a rectangular jet. *Journal of Fluid Mechanics*, 107:201220.

[Kutz et al., 2016] Kutz, J., Brunton, S., Brunton, B., and Proctor, J. (2016). *Dynamic Mode Decomposition*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

[Launder and Spalding, 1974] Launder, B. E. and Spalding, D. B. (1974). The numerical computation of turbulent flows. *Computer Methods in Applied Mechanics and Engineering*, 3(2):269 – 289.

[Li et al., 2011] Li, J., Li, J., and Xiu, D. (2011). An efficient surrogate-based method for computing rare failure probability. *Journal of Computational Physics*, 230(24):8683–8697.

[Li and Xiu, 2014] Li, J. and Xiu, D. (2014). Surrogate based method for evaluation of failure probability under multiple constraints. *SIAM Journal on Scientific Computing*, 36(2):A828–A845.

[Marín-Martínez and Sánchez-Meca, 2010] Marín-Martínez, F. and Sánchez-Meca, J. (2010). Weighting by inverse variance or by sample size in random-effects meta-analysis. *Educational and Psychological Measurement*, 70(1):56–73.

[Meier, 1953] Meier, P. (1953). Variance of a weighted mean. *Biometrics*, 9(1):59–73.

[Narayan et al., 2012] Narayan, A., Marzouk, Y., and Xiu, D. (2012). Sequential data assimilation with multiple models. *Journal of Computational Physics*, 231(19):6401 – 6418.

[O'Hagan et al., 2006] O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). Multiple experts. In *Uncertain Judgements: Eliciting Experts' Probabilities*, pages 179–192. John Wiley and Sons, Ltd.

[Owen and Zhou, 2000] Owen, A. and Zhou, Y. (2000). Safe and effective importance sampling. *Journal of the American Statistical Association*, 95(449):135–143.

[Owen, 2013] Owen, A. B. (2013). *Monte Carlo theory, methods and examples.*

[Peherstorfer et al., 2016] Peherstorfer, B., Cui, T., Marzouk, Y., and Willcox, K. (2016). Multifidelity importance sampling. *Computer Methods in Applied Mechanics and Engineering*, 300:490 – 509.

[Peherstorfer et al., 2017a] Peherstorfer, B., Kramer, B., and Willcox, K. (2017a). Combining multiple surrogate models to accelerate failure probability estimation with expensive high-fidelity models. *Journal of Computational Physics*, 341:61–75.

[Peherstorfer et al., 2017b] Peherstorfer, B., Kramer, B., and Willcox, K. (2017b). Multifidelity preconditioning of the cross-entropy method for rare event simulation and failure probability estimation. Technical report, University of Wisconsin-Madison.

[Ribault et al., 1999] Ribault, C. L., Sarkar, S., and Stanley, S. A. (1999). Large eddy simulation of a plane jet. *Physics of Fluids*, 11(10):3069–3083.

[Stanley et al., 2002] Stanley, S. A., Sarkar, S., and Mellado, J. P. (2002). A study of the flow-field evolution and mixing in a planar turbulent jet using direct numerical simulation. *Journal of Fluid Mechanics*, 450:377407.

[Ullmann and Papaioannou, 2015] Ullmann, E. and Papaioannou, I. (2015). Multilevel estimation of rare events. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):922–953.

[Villa and Marques, 2017] Villa, U. and Marques, A. (2017). A finite element solver of free plane jets. Technical Report.

[Zhou et al., 1999] Zhou, X., Sun, Z., Durst, F., and Brenner, G. (1999). Numerical simulation of turbulent jet flow and combustion. *Computers & Mathematics with Applications*, 38(9):179 – 191.