



Combining multiple surrogate models to accelerate failure probability estimation with expensive high-fidelity models



Benjamin Peherstorfer*, Boris Kramer, Karen Willcox

Department of Aeronautics & Astronautics, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:

Received 7 July 2016

Received in revised form 11 January 2017

Accepted 4 April 2017

Available online 6 April 2017

Keywords:

Multifidelity

Model reduction

Uncertainty quantification

Failure probability estimation

Surrogate modeling

Rare event simulation

ABSTRACT

In failure probability estimation, importance sampling constructs a biasing distribution that targets the failure event such that a small number of model evaluations is sufficient to achieve a Monte Carlo estimate of the failure probability with an acceptable accuracy; however, the construction of the biasing distribution often requires a large number of model evaluations, which can become computationally expensive. We present a mixed multifidelity importance sampling (MMFIS) approach that leverages computationally cheap but erroneous surrogate models for the construction of the biasing distribution and that uses the original high-fidelity model to guarantee unbiased estimates of the failure probability. The key property of our MMFIS estimator is that it can leverage multiple surrogate models for the construction of the biasing distribution, instead of a single surrogate model alone. We show that our MMFIS estimator has a mean-squared error that is up to a constant lower than the mean-squared errors of the corresponding estimators that uses any of the given surrogate models alone—even in settings where no information about the approximation qualities of the surrogate models is available. In particular, our MMFIS approach avoids the problem of selecting the surrogate model that leads to the estimator with the lowest mean-squared error, which is challenging if the approximation quality of the surrogate models is unknown. We demonstrate our MMFIS approach on numerical examples, where we achieve orders of magnitude speedups compared to using the high-fidelity model only.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

System inputs are often modeled as random variables to account for uncertainties in the inputs due to measurement errors, noise, or small perturbations in the manufacturing processes. A high-fidelity model of the system of interest maps inputs onto model outputs that approximate the system outputs with the accuracy required by the problem at hand. With an input random variable, the model output becomes a random variable as well. An important task in the context of reliability analysis is the estimation of probabilities of failure events. Basic Monte Carlo estimation draws realizations from the input random variable, evaluates the high-fidelity model at the realizations, and derives a failure probability estimate from the corresponding outputs. Because failure probabilities are typically small, basic Monte Carlo estimation typically requires a large number of high-fidelity model evaluations to derive estimates of an acceptable accuracy. Therefore, basic Monte Carlo estimation can become computationally intractable if the high-fidelity model is expensive to solve.

* Corresponding author.

E-mail address: pehersto@mit.edu (B. Peherstorfer).

One class of methods that accelerate Monte Carlo estimation derives problem-dependent sampling strategies such that fewer samples are necessary to obtain an acceptable estimate of the failure probability than in basic Monte Carlo estimation. One such method is importance sampling. The first step of importance sampling is constructing a problem-dependent biasing distribution in which the failure event has a larger probability than in the original nominal distribution. In the second step, the biasing distribution is used to derive an estimate of the failure probability, where the change in distribution is compensated via re-weighting the samples of the biasing distribution. The key challenge of importance sampling is constructing a suitable biasing distribution. A general overview of importance sampling is given in [1–3]. Adaptive importance sampling techniques iterate between the biasing distribution construction and the estimation step to iteratively construct and adapt the biasing distribution as the computation proceeds [4–8]. The cross-entropy method [9,10] can be used to iteratively build the biasing distribution. Another popular approach to accelerate Monte Carlo estimation of failure probabilities is the subset method [11–13]. The subset method starts with an event that includes the failure event and that has a large probability compared to the failure probability. The subset method then successively refines the event until the actual failure event of interest is recovered. During the refinement, a problem-dependent sampling strategy is constructed with which the failure event can be estimated with few samples. A multilevel extension to the subset method is introduced in [14].

Another class of methods that accelerate Monte Carlo estimation uses surrogate models so that expensive high-fidelity model evaluations can be replaced with cheap surrogate evaluations. Examples of surrogate models include projection-based reduced models [15,16], data-fit interpolation and regression models [17], machine-learning-based models such as support vector machines [18], and other simplified models [19]. However, simply replacing the high-fidelity model with a surrogate model in a basic Monte Carlo estimator can lead to significantly biased estimates of the failure probability [20,21]. Multifidelity methods combine surrogate model and high-fidelity model outputs to control, or even avoid, the bias that is introduced by a surrogate model. In [22,23], the surrogate model is successively adapted during the Monte Carlo estimation so that the surrogate model becomes an accurate approximation of the high-fidelity model at the failure event. In [21,24], an approach is presented that switches between a surrogate model and the high-fidelity model depending on the error of the surrogate model. The obtained estimators are unbiased if tolerance parameters are chosen well. The work [25] uses *a posteriori* error estimators of a surrogate model to decide whether the surrogate model or the high-fidelity model should be used.

The two-fidelity importance sampling approach introduced in [26] uses the surrogate model for the construction of the biasing distribution and the high-fidelity model to derive the estimate of the failure probability. The key property of the two-fidelity approach [26] is that the two-fidelity estimator is an unbiased estimator of the failure probability, independent of the availability of error guarantees on the surrogate model. A surrogate model that is a poor approximation of the high-fidelity model can lead to a two-fidelity estimator that requires more evaluations of the high-fidelity model than a basic Monte Carlo estimator with the same mean-squared error (MSE); however, the unbiasedness of the two-fidelity estimator is guaranteed. If the surrogate model provides a reasonable approximation of the high-fidelity model, then the two-fidelity approach can lead to significant speedups compared to basic Monte Carlo estimation, see the numerical examples in [26].

The two-fidelity approach presented in [26], as well as the methods [22,23,21,24,25] discussed above, can leverage only a single surrogate model. If multiple surrogate models are given, one of the surrogate models has to be selected. Error bounds for the surrogate model outputs are available only in limited problem settings, which means that it is challenging in many situations to make an informed decision which of the surrogate models to use in a two-fidelity approach. Even if error bounds for the surrogate model outputs are available, it often remains challenging to propagate the bounds for the outputs onto bounds for the MSE of two-fidelity estimators that use the surrogate model. We present here a mixed multifidelity importance sampling (MMFIS) method that builds on the mixed importance sampling methodology of [27] and the two-fidelity approach of [26] to derive a biasing distribution from multiple surrogate models, instead of using a single surrogate model alone. We show that the MSE of our MMFIS estimator is up to a constant as low as the MSE of the best two-fidelity estimator (i.e., the one that uses the surrogate model that leads to the lowest MSE). Our MMFIS estimator avoids the problem of selecting one of the given surrogate models and is still guaranteed to achieve an MSE that is up to a constant as low as the MSE of any of the two-fidelity estimators derived from the given surrogate models. The constant is equal to the number of surrogate models and therefore is typically small. Furthermore, our MMFIS estimator is an unbiased estimator of the failure probability.

The outline of the presentation is as follows. Section 2 gives a problem definition and briefly discusses basic Monte Carlo estimation, importance sampling, and the two-fidelity importance sampling method of [26]. Section 3 introduces our MMFIS method that combines outputs of multiple surrogate models for the construction of a biasing distribution. Numerical results in Section 4 demonstrate our MMFIS method on a structural reliability example and a shock propagation problem. The conclusions in Section 5 close the presentation.

2. Failure probability estimation with importance sampling

This section gives the problem formulation and briefly discusses Monte Carlo approaches to estimate failure probabilities. Section 2.1 defines the problem, and Section 2.2 introduces basic Monte Carlo estimation and importance sampling. Section 2.3 introduces two-fidelity importance sampling, which fuses outputs of a surrogate model with outputs of the high-fidelity model to accelerate the estimation of failure probabilities compared to basic Monte Carlo estimation and importance sampling.

2.1. Problem definition

Consider a system with $d \in \mathbb{N}$ inputs and $d' \in \mathbb{N}$ outputs. Let $\mathcal{D} \subseteq \mathbb{R}^d$ be the input domain and $\mathcal{Y} \subseteq \mathbb{R}^{d'}$ the output domain. The system is modeled by a high-fidelity model $f : \mathcal{D} \rightarrow \mathcal{Y}$, which maps inputs $\mathbf{z} \in \mathcal{D}$ onto outputs $f(\mathbf{z}) \in \mathcal{Y}$. The high-fidelity model output $f(\mathbf{z})$ approximates the actual system output in a suitable metric with the accuracy required by the current problem at hand. The costs of evaluating f are $0 < w \in \mathbb{R}$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with sample space Ω , σ -algebra \mathcal{F} , and probability measure \mathbb{P} on (Ω, \mathcal{F}) . Let $Z : \Omega \rightarrow \mathcal{D}$ be a random variable with the probability density function p . Let further $g : \mathcal{Y} \rightarrow \mathbb{R}$ be a limit state function that defines the set

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

The failure probability $P \in [0, 1]$ is

$$P = \mathbb{P}[Z^{-1}(\mathcal{G})].$$

Define the indicator function $\mathbb{I}_{\mathcal{G}} : \mathcal{D} \rightarrow \{0, 1\}$ as

$$\mathbb{I}_{\mathcal{G}}(\mathbf{z}) = \begin{cases} 1, & \mathbf{z} \in \mathcal{G}, \\ 0, & \text{else.} \end{cases}$$

The failure probability P equals the expected value $\mathbb{E}_p[\mathbb{I}_{\mathcal{G}}]$ of the indicator factor $\mathbb{I}_{\mathcal{G}}$ with respect to the density p , i.e.,

$$P = \mathbb{E}_p[\mathbb{I}_{\mathcal{G}}] = \int_{\mathcal{D}} \mathbb{I}_{\mathcal{G}}(\mathbf{z})p(\mathbf{z})d\mathbf{z}.$$

The goal of the following methodology is to estimate the failure probability P in the case where P is small, i.e., $P \ll 1$.

2.2. Failure probability estimation with Monte Carlo

This section discusses basic Monte Carlo estimation and importance sampling.

2.2.1. Basic Monte Carlo estimation

Let Z_1, \dots, Z_m be $m \in \mathbb{N}$ independent random variables that are distributed as the input random variable Z . The basic Monte Carlo estimator of the failure probability P is

$$P_m^{\text{MC}} = \frac{1}{m} \sum_{i=1}^m \mathbb{I}_{\mathcal{G}}(Z_i).$$

The costs of evaluating the basic Monte Carlo estimator P_m^{MC} are

$$c(P_m^{\text{MC}}) = mw,$$

because each of the m evaluations of the indicator function $\mathbb{I}_{\mathcal{G}}$ entails an evaluation of the model f and an evaluation of the limit state function g . The costs of one evaluation of f are w . The costs of evaluating the limit state function g are typically negligible compared to the costs of f , and therefore we ignore the costs of g in the costs $c(P_m^{\text{MC}})$ of the basic Monte Carlo estimator.

The relative MSE of the basic Monte Carlo estimator P_m^{MC} is

$$e_r(P_m^{\text{MC}}) = \frac{\mathbb{V}_p[P_m^{\text{MC}}]}{(\mathbb{E}_p[P_m^{\text{MC}}])^2} = \frac{1}{m} \frac{\mathbb{V}_p[\mathbb{I}_{\mathcal{G}}]}{(\mathbb{E}_p[\mathbb{I}_{\mathcal{G}}])^2},$$

if the variance $\mathbb{V}_p[\mathbb{I}_{\mathcal{G}}]$ is finite with respect to the density p . Because the basic Monte Carlo estimator P_m^{MC} is an unbiased estimator of P , the relative MSE of P_m^{MC} is

$$e_r(P_m^{\text{MC}}) = \frac{P(1-P)}{P^2m} = \frac{1-P}{Pm}. \tag{1}$$

The relative MSE $e_r(P_m^{\text{MC}})$ shows that the basic Monte Carlo estimator requires many evaluations m of the high-fidelity model f to accurately estimate a small failure probability P . For example, if $P = 10^{-5}$, then $m > 10^7$ evaluations of f are required to obtain a basic Monte Carlo estimate of P with relative MSE $e_r(P_m^{\text{MC}}) < 10^{-2}$.

2.2.2. Importance sampling

Importance sampling is a variance reduction approach that samples from a problem-dependent biasing distribution and compensates the change in the distribution with a re-weighting [1–3]. Importance sampling consists of two steps in the setting that we consider here. First, the biasing density q is constructed. Second, an estimate of the failure probability is derived.

In the first step, importance sampling constructs a biasing density q with support $\text{supp}(p) \subseteq \text{supp}(q)$, where p is the density of the random variable Z and the support is defined as

$$\text{supp}(p) = \{\mathbf{z} \in \mathcal{D} \mid p(\mathbf{z}) > 0\}. \quad (2)$$

The random variable Z and the density p are the nominal random variable and the nominal density, respectively. The random variable Z' with biasing density q is the biasing random variable.

In the second step, the biasing density q and (2) are used to write the failure probability P as

$$P = \mathbb{E}_p[\mathbb{I}_G] = \int_{\mathcal{D}} \mathbb{I}_G(\mathbf{z}) p(\mathbf{z}) d\mathbf{z} = \int_{\mathcal{D}} \mathbb{I}_G(\mathbf{z}) \frac{p(\mathbf{z})}{q(\mathbf{z})} q(\mathbf{z}) d\mathbf{z} = \mathbb{E}_q \left[\mathbb{I}_G \frac{p}{q} \right], \quad (3)$$

where the ratio p/q serves as a weighting factor to compensate that the expected value is taken with respect to the biasing density q . Using (3) leads to the importance sampling estimator P_m^{IS} of P

$$P_m^{\text{IS}} = \frac{1}{m} \sum_{i=1}^m \mathbb{I}_G(Z'_i) \frac{p(Z'_i)}{q(Z'_i)}, \quad (4)$$

with the independent and identically distributed (i.i.d.) random variables Z'_1, \dots, Z'_m that are distributed as the biasing random variable Z' . Note that P_m^{IS} is an unbiased estimator of P , if $\text{supp}(p) \subseteq \text{supp}(q)$.

The costs $c(P_m^{\text{IS}})$ of the importance sampling estimator are split into the costs of constructing the biasing density and the costs of evaluating the model f at m realizations of the biasing random variable Z' . The relative MSE of the importance sampling estimator P_m^{IS} is

$$e_r(P_m^{\text{IS}}) = \frac{1}{m} \frac{\mathbb{V}_q \left[\mathbb{I}_G \frac{p}{q} \right]}{\left(\mathbb{E}_q [P_m^{\text{IS}}] \right)^2},$$

if the variance $\mathbb{V}_q[\mathbb{I}_G \frac{p}{q}]$ is finite. If the biasing density q is constructed such that $\mathbb{V}_q[\mathbb{I}_G \frac{p}{q}] < \mathbb{V}_p[\mathbb{I}_G]$, then the importance sampling estimator P_m^{IS} achieves a smaller relative MSE than the basic Monte Carlo estimator P_m^{MC} with the same number of realizations m .

2.3. Two-fidelity importance sampling

Let $f^{(1)} : \mathcal{D} \rightarrow \mathcal{Y}$ be a surrogate model of the high-fidelity model f . Evaluating the surrogate model $f^{(1)}$ has costs $0 < w_1 \in \mathbb{R}$. The costs w_1 are typically significantly lower than the costs w of the high-fidelity model. No assumptions on the error of the surrogate output with respect to the high-fidelity output are made here. Furthermore, no assumptions on the type of the surrogate model are made either. The following methodology is applicable for heterogeneous types of surrogate models, reaching from projection-based reduced models to data-fit models to simplified-physics models, see the discussion in Section 1.

In [26], a multifidelity importance sampling approach is introduced that uses a surrogate model to construct the biasing density q , and the high-fidelity model to derive an estimate of the failure probability P . In the following, we refer to the approach presented in [26] as two-fidelity importance sampling because two models are used—a surrogate and the high-fidelity model. Step one of the two-fidelity importance sampling method is constructing a biasing density. Let

$$\mathbf{z}_1, \dots, \mathbf{z}_n \quad (5)$$

be $n \in \mathbb{N}$ realizations of the random variable Z . The two-fidelity approach of [26] evaluates the surrogate model $f^{(1)}$ at the realizations (5) to obtain the surrogate model outputs $f^{(1)}(\mathbf{z}_1), \dots, f^{(1)}(\mathbf{z}_n)$. The set

$$\mathcal{Z} = \left\{ \mathbf{z}_i \mid g(f^{(1)}(\mathbf{z}_i)) < 0, i = 1, \dots, n \right\} \quad (6)$$

contains the realizations for which the surrogate model leads to a limit state function value below zero. Note that typically $|\mathcal{Z}| \ll n$, because only a few of the realizations (5) correspond to the failure described by the limit state function g if $P \ll 1$. A normal distribution is then fitted to the realizations in \mathcal{Z} . In [26], the expectation–maximization algorithm is used, but any other density estimation approach [28–30] can be used. The result of the first step is the biasing density $q_n^{(1)}$, derived from n evaluations of the surrogate model $f^{(1)}$, and the corresponding biasing random variable $Z^{(1)}$.

The second step of the two-fidelity importance sampling approach is estimating the failure probability P by using the biasing random variable $Z^{(1)}$. Let $Z_1^{(1)}, \dots, Z_m^{(1)}$ be m i.i.d. random variables that are distributed as the biasing random variable $Z^{(1)}$. The two-fidelity importance sampling estimator of P is then given by

$$P_{n,m}^{(1)} = \frac{1}{m} \sum_{i=1}^m \mathbb{I}_{\mathcal{G}}(Z_i^{(1)}) \frac{p(Z_i^{(1)})}{q_n^{(1)}(Z_i^{(1)})}. \tag{7}$$

The high-fidelity model f is evaluated in (7) by evaluating the indicator function $\mathbb{I}_{\mathcal{G}}$. Therefore, the estimator $P_{n,m}^{(1)}$ is a two-fidelity estimator because it combines outputs of the surrogate model $f^{(1)}$ (construction of biasing density) with outputs of the high-fidelity model f . Note that the superscript (1) in the notation $P_{n,m}^{(1)}$ indicates that the random variables $Z_1^{(1)}, \dots, Z_m^{(1)}$ in (7) are distributed as the biasing random variable $Z^{(1)}$, which is derived from surrogate model $f^{(1)}$.

The costs $c(P_{n,m}^{(1)})$ of the estimator $P_{n,m}^{(1)}$ are split into the costs of generating the biasing density $q_n^{(1)}$, and the costs for deriving the estimate of P . Constructing the biasing density $q_n^{(1)}$ requires n evaluations of the surrogate model. The high-fidelity model is evaluated at m realizations of the biasing random variable. Therefore, the costs of $c(P_{n,m}^{(1)})$ are

$$c(P_{n,m}^{(1)}) = nw_1 + mw.$$

Note that the costs of fitting the biasing density to the realizations in \mathcal{Z} are ignored here, because these costs are typically negligible compared to the costs of model evaluations, see [26].

3. Multifidelity importance sampling with multiple surrogate models

We now consider the situation where we have given $k > 1$ surrogate models $f^{(1)}, \dots, f^{(k)} : \mathcal{D} \rightarrow \mathcal{Y}$ with costs $0 < w_1, \dots, w_k \in \mathbb{R}$, and where any further information about these surrogate models is unavailable. In particular, no notion of their approximation quality with respect to the high-fidelity model is available. Therefore, it is unclear which of these surrogate models leads to the biasing density that reduces the variance most in the two-fidelity importance sampling approach.

We propose here a mixed multifidelity importance sampling (MMFIS) estimator that combines all k surrogate models and the high-fidelity model, and therefore our approach avoids the problem of selecting the particular surrogate model that most reduces the variance. Our MMFIS estimator builds on the mixed importance sampling approach introduced in [27] and the two-fidelity importance sampling method discussed in Section 2.3. By combining all surrogate models in the biasing construction step in a judicious way following [27], our MMFIS estimator guarantees that the relative MSE is up to the factor k as low as the variance of the two-fidelity importance sampling estimator with the (single) surrogate model that leads to the lowest variance of all k given surrogate models; but without selecting that particular surrogate model that reduces the variance most.

Our MMFIS approach is split into the biasing density construction and the estimation step. Section 3.1 gives details on constructing the biasing density from k surrogate models. Section 3.2 presents the MMFIS estimator. The computational procedure is summarized in Section 3.3. A cost and error analysis is given in Section 3.4.

3.1. Construction of biasing density

Our MMFIS approach constructs a biasing density $q_n^{(i)}$, and the corresponding biasing random variable $Z^{(i)}$, from each surrogate model $f^{(i)}$ for $i = 1, \dots, k$. For $i = 1, \dots, k$, MMFIS draws n realizations $\mathbf{z}_1, \dots, \mathbf{z}_n$ from Z and evaluates the surrogate model $f^{(i)}$ at these realizations to derive the set $\mathcal{Z}^{(i)}$

$$\mathcal{Z}^{(i)} = \{\mathbf{z}_j \mid g(f^{(i)}(\mathbf{z}_j)) < 0, j = 1, \dots, n\},$$

cf. the set \mathcal{Z} defined in (6). A separate set of realizations is drawn for each set $\mathcal{Z}^{(1)}, \dots, \mathcal{Z}^{(k)}$. From each set $\mathcal{Z}^{(i)}, i = 1, \dots, k$, a biasing density $q_n^{(i)}$ is constructed as described in Section 2.3. Note that if $\mathcal{Z}^{(i)} = \emptyset$ (i.e., the set is empty), then we set $q_n^{(i)} = p$ to the nominal density, for $i = 1, \dots, k$, see [26] for details. The number of realizations n can have a significant influence on the quality of the biasing densities with respect to variance reduction. In particular, in high-dimensional settings, constructing a biasing density $q_n^{(i)}$ from a set $\mathcal{Z}^{(i)}$ with only few elements is challenging. In [26], an adaptive selection of the number of realizations n is proposed, where the number of realizations with negative limit state function are used to decide if an increase in the number n is necessary.

The biasing densities $q_n^{(1)}, \dots, q_n^{(k)}$ are combined into the mixture density q_n as

$$q_n(\mathbf{z}) = \frac{1}{k} \sum_{i=1}^k q_n^{(i)}(\mathbf{z}). \tag{8}$$

In the sense of the sum in the definition of the mixture density q_n in (8), the mixture density q_n combines information from all surrogate models $f^{(1)}, \dots, f^{(k)}$. Note that q_n is in general different than directly constructing a density from the union $\mathcal{Z}^{(1)} \cup \dots \cup \mathcal{Z}^{(k)}$.

3.2. Mixed multifidelity importance sampling estimator

We follow [27] and use the mixture density q_n and the densities $q_n^{(1)}, \dots, q_n^{(k)}$ to derive an estimator of P as

$$P_{n,m}^{\text{MMFIS}} = \frac{1}{m} \sum_{i=1}^m \frac{\mathbb{I}_{\mathcal{G}}(Z'_i) p(Z'_i) - \sum_{j=1}^k \beta_j q_n^{(j)}(Z'_i)}{q_n(Z'_i)} + \sum_{j=1}^k \beta_j. \tag{9}$$

The estimator $P_{n,m}^{\text{MMFIS}}$ combines importance sampling and control variates [31,32]. The mixture density q_n is the biasing density and the random variables Z'_1, \dots, Z'_m are i.i.d. as the random variable Z' with the density q_n . The densities $q_n^{(1)}, \dots, q_n^{(k)}$ serve as control variates with the control variate coefficients $\beta = [\beta_1, \dots, \beta_k]^T \in \mathbb{R}^k$. The estimator $P_{n,m}^{\text{MMFIS}}$ is an unbiased estimator of P , which follows from [27, Theorem 2] and [26, Theorem 1]. We refer to [27] for further details on combining importance sampling and control variates in estimators of the form (9).

3.2.1. Determining the control variate coefficients

The estimator $P_{n,m}^{\text{MMFIS}}$ depends on the control variate coefficients β . We define the optimal control variate coefficients β^* to solve

$$\beta^* = \arg \min_{\beta \in \mathbb{R}^k} \mathbb{V}_{q_n} [P_{n,m}^{\text{MMFIS}}],$$

which means that β^* minimizes the variance of the $P_{n,m}^{\text{MMFIS}}$ estimator. The optimal coefficients β^* are usually unavailable. We therefore derive approximate control variate coefficients $\hat{\beta}$ following [27]. Let $\mathbf{f} \in \mathbb{R}^m$ be a vector defined as

$$\mathbf{f} = \begin{bmatrix} \frac{\mathbb{I}_{\mathcal{G}}(\mathbf{z}_1) p(\mathbf{z}_1)}{q_n(\mathbf{z}_1)} \\ \vdots \\ \frac{\mathbb{I}_{\mathcal{G}}(\mathbf{z}_m) p(\mathbf{z}_m)}{q_n(\mathbf{z}_m)} \end{bmatrix} \in \mathbb{R}^m,$$

and let $\mathbf{Q} \in \mathbb{R}^{m \times k}$ be the matrix

$$\mathbf{Q} = \begin{bmatrix} \frac{q_n^{(1)}(\mathbf{z}_1)}{q_n(\mathbf{z}_1)} & \cdots & \frac{q_n^{(k)}(\mathbf{z}_1)}{q_n(\mathbf{z}_1)} \\ \vdots & \ddots & \vdots \\ \frac{q_n^{(1)}(\mathbf{z}_m)}{q_n(\mathbf{z}_m)} & \cdots & \frac{q_n^{(k)}(\mathbf{z}_m)}{q_n(\mathbf{z}_m)} \end{bmatrix} \in \mathbb{R}^{m \times k}.$$

With the vector \mathbf{f} and the matrix \mathbf{Q} , we rewrite the variance $\mathbb{V}_{q_n} [P_{n,m}^{\text{MMFIS}}]$ as

$$\mathbb{V}_{q_n} [P_{n,m}^{\text{MMFIS}}] = \mathbb{E}_{q_n} \left[\left(\frac{\mathbf{1}^T (\mathbf{f} - \mathbf{Q} \beta)}{m} + \sum_{j=1}^k \beta_j - P \right)^2 \right].$$

Since we want to find coefficients β that lead to a small variance, we minimize the first term

$$\frac{\mathbf{1}^T (\mathbf{f} - \mathbf{Q} \beta)}{m}$$

in the variance $\mathbb{V}_{q_n} [P_{n,m}^{\text{MMFIS}}]$. Consider therefore the least-squares problem

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^k} \|\mathbf{f} - \mathbf{Q} \beta\|_2^2, \tag{10}$$

with the solution $\hat{\beta} \in \mathbb{R}^k$ and Euclidean norm $\|\cdot\|_2$. The columns of \mathbf{Q} sum to the vector $\mathbf{1}$ with all components set to 1. This means that the residual $\|\mathbf{f} - \mathbf{Q} \hat{\beta}\|_2^2$ of the least-squares problem (10) is zero at the solution $\hat{\beta}$.

Algorithm 1 Mixed multifidelity importance sampling.

- 1: **procedure** MMFIS($Z, f, f^{(1)}, \dots, f^{(k)}, g, n, m$)
- 2: **for** $i = 1, \dots, k$ **do**
- 3: Draw realizations $\mathbf{z}_1, \dots, \mathbf{z}_n$ from Z
- 4: Set $\mathcal{Z}^{(i)} = \{\mathbf{z}_j \mid g(f^{(i)}(\mathbf{z}_j)) < 0, j = 1, \dots, n\}$
- 5: Fit distribution $q_n^{(i)}$ to realizations in $\mathcal{Z}^{(i)}$ (see Section 2.3 and Section 3.1)
- 6: **end for**
- 7: Define mixture density

$$q_n(\mathbf{z}) = \frac{1}{k} \sum_{i=1}^k q_n^{(i)}(\mathbf{z})$$

- 8: Draw realizations $\mathbf{z}'_1, \dots, \mathbf{z}'_m$ of Z' with density q_n
 - 9: Assemble vector $\mathbf{f} = [\mathbb{I}_G(\mathbf{z}'_1)p(\mathbf{z}'_1)/q_n(\mathbf{z}'_1), \dots, \mathbb{I}_G(\mathbf{z}'_m)p(\mathbf{z}'_m)/q_n(\mathbf{z}'_m)]^T \in \mathbb{R}^m$
 - 10: Assemble matrix $\mathbf{Q} \in \mathbb{R}^{m \times k}$ with $Q_{ij} = q_n^{(j)}(\mathbf{z}'_i)/q_n(\mathbf{z}'_i), i = 1, \dots, m, j = 1, \dots, k$
 - 11: Compute $\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^k} \|\mathbf{f} - \mathbf{Q}\boldsymbol{\beta}\|_2^2$
 - 12: Set $\hat{p}_{n,m}^{\text{MMFIS}} = \sum_{i=1}^k \hat{\beta}_i$
 - 13: **return** $\hat{p}_{n,m}^{\text{MMFIS}}$
 - 14: **end procedure**
-

3.2.2. The MMFIS estimator

Consider now the MMFIS estimator

$$\hat{p}_{n,m}^{\text{MMFIS}} = \frac{1}{m} \sum_{i=1}^m \frac{\mathbb{I}_G(Z'_i)p(Z'_i) - \sum_{j=1}^k \hat{\beta}_j q_n^{(j)}(Z'_i)}{q_n(Z'_i)} + \sum_{j=1}^k \hat{\beta}_j, \quad (11)$$

which is the estimator $p_{n,m}^{\text{MMFIS}}$ defined in (9) with coefficients $\hat{\boldsymbol{\beta}}$. The MMFIS estimator $\hat{p}_{n,m}^{\text{MMFIS}}$ simplifies to

$$\hat{p}_{n,m}^{\text{MMFIS}} = \sum_{j=1}^k \hat{\beta}_j, \quad (12)$$

because the residual of (10) is zero. The effect of using the approximate coefficients $\hat{\boldsymbol{\beta}}$, instead of the optimal coefficients $\boldsymbol{\beta}^*$, on the relative MSE of the estimator $\hat{p}_{n,m}^{\text{MMFIS}}$ is asymptotically negligible under certain conditions. We refer to [27] for details.

3.3. Computational procedure

The computational procedure of deriving an estimate of P with the MMFIS approach is summarized in Algorithm 1. The inputs are the nominal random variable Z , the high-fidelity model f , the surrogate models $f^{(1)}, \dots, f^{(k)}$, the limit state function g , the number of realizations n for constructing the biasing density, and the number of realizations m for deriving the MMFIS estimate.

In the loop of Algorithm 1 (lines 2–6), a biasing density $q_n^{(i)}$ is constructed from each surrogate model $f^{(i)}, i = 1, \dots, k$. For each surrogate model, realizations are drawn from the nominal random variable Z , and then the realizations for which the surrogate model predicts a failure are collected in the set $\mathcal{Z}^{(i)}$. A distribution is then fitted to the realizations in $\mathcal{Z}^{(i)}$ with, e.g., the expectation–maximization algorithm, as described in Section 2.3 and Section 3.1.

Line 7 of Algorithm 1 defines the mixture density q_n , which serves as biasing density. Realizations $\mathbf{z}'_1, \dots, \mathbf{z}'_m$ are drawn from the biasing random variable Z' with density q_n , from which the vector \mathbf{f} and the matrix \mathbf{Q} are assembled. The coefficients $\hat{\boldsymbol{\beta}} \in \mathbb{R}^k$ are obtained by solving the least-squares problem (10). The MMFIS estimate is given as the sum of the coefficients $\hat{\boldsymbol{\beta}}$. The MMFIS estimate is returned.

3.4. Cost and error analysis of MMFIS estimator

We first analyze the costs of deriving an MMFIS estimate with Algorithm 1. We consider here only the costs of model evaluations. Typically, evaluating the limit state function g and sampling the random variables have negligible costs compared to the costs of evaluating the surrogate and the high-fidelity models. The matrix \mathbf{Q} and the vector \mathbf{f} in the least-squares problem (10) can be derived from the high-fidelity model outputs and the density outputs required in the MMFIS estimator at no additional costs. In particular, no additional model evaluations are necessary to obtain the coefficients $\hat{\boldsymbol{\beta}}$ with the least-squares problem (10). Additionally, we ignore the costs of fitting a distribution to data points.

The costs of the MMFIS estimation can be split into the costs of constructing the biasing density and the costs of deriving the estimate of P . Each surrogate model is evaluated at n inputs, and the high-fidelity model is evaluated at m inputs. The costs of the MMFIS estimator therefore are

$$c(\hat{P}_{n,m}^{\text{MMFIS}}) = n \sum_{i=1}^k w_i + mw,$$

where the $n \sum_{i=1}^k w_i$ are the costs for constructing the biasing densities and mw are the costs for deriving the estimate using m high-fidelity model evaluations. Note that the costs of the two-fidelity estimator $P_{n,m}^{(i)}$, which uses a single surrogate model $f^{(i)}$ only, are $c(P_{n,m}^{(i)}) = nw_i + mw$.

We compare the relative MSE $e_r(\hat{P}_{n,m}^{\text{MMFIS}})$ of the MMFIS estimator to the relative MSE $e_r(P_{n,m}^{(i)})$ of the two-fidelity estimator that uses a single surrogate model $f^{(i)}$ only. Consider first the estimator $P_{n,m}^{\text{MMFIS}}$ with the optimal control variate coefficients β^* . It is shown in [27] that

$$\frac{1}{k} \mathbb{V}_{q_n}[P_{n,m}^{\text{MMFIS}}] \leq \min_{1 \leq i \leq k} \mathbb{V}_{q_n^{(i)}}[P_{n,m}^{(i)}] \quad (13)$$

holds under mild assumptions. The result (13) shows that the variance of the estimator $P_{n,m}^{\text{MMFIS}}$ with the optimal control variate coefficients β^* is up to the factor k as low as the variance of any of the multifidelity estimators $P_{n,m}^{(i)}$ that use a single surrogate model $f^{(i)}$ only, for $i = 1, \dots, k$. Since the $P_{n,m}^{\text{MMFIS}}$ estimator and the estimators $P_{n,m}^{(1)}, \dots, P_{n,m}^{(k)}$ are unbiased with respect to P , the relative MSEs satisfy

$$\frac{1}{k} e_r(P_{n,m}^{\text{MMFIS}}) \leq \min_{1 \leq i \leq k} e_r(P_{n,m}^{(i)}). \quad (14)$$

The bound (14) on the relative MSE holds only for the estimator $P_{n,m}^{\text{MMFIS}}$ with the optimal control variate coefficients β^* . Our numerical results will show, however, that the approximate coefficients $\hat{\beta}$ lead to similar behavior.

4. Numerical experiments

This section demonstrates the MMFIS estimator on two numerical examples. Section 4.1 considers a cantilever beam. A force is applied to the beam and a failure occurs if the displacement of the beam exceeds a certain threshold. Section 4.2 considers a hyperbolic problem, where a shock travels through the spatial domain. Closure terms are necessary to derive surrogate models that capture the shock front. Closure terms are typically derived in an ad-hoc way and with domain expertise, and therefore it is often unclear which of the closure terms leads to an accurate surrogate model. Our MMFIS estimator combines surrogate models obtained from various closure terms to derive an estimate of the probability that the shock front moved beyond a threshold after a certain amount of time.

4.1. Displacement of cantilever beam

Consider a model that describes a cantilever beam in a three-dimensional spatial domain [33], see Fig. 1a. At time $t = 0$, a force is applied at the tip of the beam, which leads to a response in the displacement of the beam in direction x_3 .

4.1.1. Beam example: problem setup

The high-fidelity model is the finite element model described in [33]. We use the implicit Euler method to discretize the time domain. We simulate the beam until time $T = 1.5$ with a time step size $\delta t = 10^{-4}$. This leads to 1.5×10^4 time steps, where in each time step a linear system with 240 unknowns is solved. Let $\mathcal{D} = \mathbb{R}$ be the input domain. The high-fidelity model $f : \mathcal{D} \rightarrow \mathcal{Y}$ maps an input $z \in \mathcal{D}$ of the beam onto the output. The input $z \in \mathcal{D}$ defines the length $|z|$ of the beam. To define the output, let $u_3(t)$ be the trajectory of the displacement of the tip of the beam in direction x_3 over the time $t \in [0, T]$. Let further $u_3^{(1)}, u_3^{(2)}, \dots \in \mathbb{R}$ be the local maxima of the trajectory $u_3(t)$ ordered with respect to the time t , and let \bar{u}_3 be the mean of the trajectory $u_3(t)$. The output of f is defined as $y = u_3^{(5)} - \bar{u}_3$, see Fig. 1b.

The input random variable Z is normally distributed with mean 1.1 and standard deviation 0.05. The task is to estimate the probability P that the output is larger than 1.65×10^{-4} . The corresponding limit state function is $g_{\text{beam}}(y) = 1.65 \times 10^{-4} - y$. We use an importance sampling estimate $P_{n,m}^{\text{IS}} = 5.0066 \times 10^{-4}$ as reference probability. The reference probability is computed from $m = 10^5$ realizations. The biasing density is constructed from $n = 10^5$ high-fidelity model evaluations by fitting a normal distribution to the model outputs with negative limit state function values, cf. Section 2.3. The number of realizations n for constructing the biasing density can significantly influence the quality of the estimate of the failure probability, see the discussion in Section 3.1. To ensure that our reference estimate $P_{n,m}^{\text{IS}}$ is accurate, we set the number of realizations m for the estimate high. The estimated variance of the reference importance sampling estimator corresponding to $P_{n,m}^{\text{IS}}$ is less than 10^{-12} .

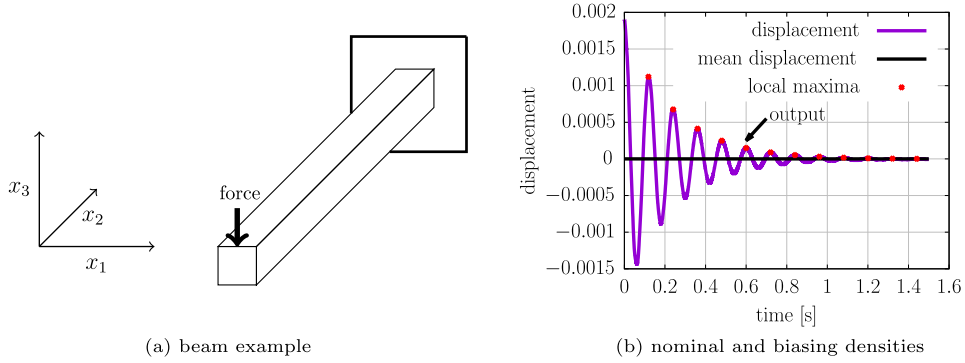


Fig. 1. Beam example: A force is applied to a cantilever beam as shown in (a). The plot in (b) visualizes the output of the beam problem. The output is the difference between the mean displacement in direction x_3 and the fifth local maximum of the displacement trajectory.

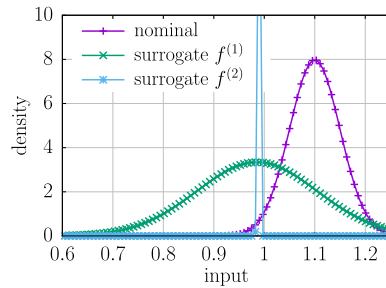


Fig. 2. The plot shows that surrogate model $f^{(1)}$ leads to a biasing density with mean near 0.95, whereas the mean of the nominal density is 1.1. The biasing density corresponding to the surrogate model $f^{(2)}$ has a high peak near 1 and is low for large parts of the input domain, which reflects that $f^{(2)}$ is a poor approximation of the high-fidelity model.

4.1.2. Beam example: surrogate models

We construct two surrogate models $f^{(1)}, f^{(2)} : \mathcal{D} \rightarrow \mathcal{Y}$. Surrogate model $f^{(1)}$ is constructed with proper orthogonal decomposition (POD) [16]. Let therefore $z_1, \dots, z_{100} \in \mathcal{D}$ be 100 realizations of the input random variable Z . For each realization, the high-fidelity model is simulated, and the corresponding states are stored at every 150-th time step. This leads to $100 \times 100 = 10,000$ states, from which we derive the POD basis. We keep the first 15 POD basis vectors. The surrogate model operators are obtained via Galerkin projection of the high-fidelity model operators onto the space spanned by the POD basis. Surrogate model $f^{(2)}$ is derived with balanced truncation for the high-fidelity model at $z = 1$. The surrogate model $f^{(2)}$ uses 15 basis vectors. Note that we construct surrogate model $f^{(2)}$ for the input $z = 1$ only on purpose, so that we expect that the surrogate model $f^{(2)}$ is accurate near $z = 1$ only and fails to provide an accurate approximation of the high-fidelity model at other inputs.

4.1.3. Beam example: results

Fig. 2 shows the nominal density of the random variable Z and the biasing densities $q_n^{(1)}$ and $q_n^{(2)}$ derived from the surrogate model $f^{(1)}$ and $f^{(2)}$ from $n = 10^5$ realizations. The biasing density $q_n^{(2)}$ has a sharp peak near input $z = 1$, which is the input around the surrogate model was created. The density $q_n^{(1)}$ corresponding to the surrogate model $f^{(1)}$ is significantly smoother than the density $q_n^{(2)}$.

We now compare the relative MSE of four estimators:

- The importance sampling estimator $P_{n,m}^{IS}$, where the biasing density is constructed by evaluating the high-fidelity model at n realizations of Z and fitting a normal distribution to the model outputs with negative limit state function values.
- The two-fidelity estimators $P_{n,m}^{(1)}$ and $P_{n,m}^{(2)}$ that are derived as in Section 2.3 and that use a single surrogate model for the construction of the biasing density.
- The MMFIS estimator $\hat{P}_{n,m}^{MMFIS}$ that uses both surrogate models $f^{(1)}$ and $f^{(2)}$ as described in Section 3 and Algorithm 1.

The four estimators use the high-fidelity model to derive the estimate of the failure probability P . They are unbiased estimators of the failure probability P , and they use $n = 10^5$ realizations to construct the biasing densities. The importance sampling estimator $P_{n,m}^{IS}$ serves as our benchmark estimator in terms of the relative MSE. The two-fidelity estimators $P_{n,m}^{(1)}$ and $P_{n,m}^{(2)}$ serve as benchmarks in terms of runtime. This means that we want our MMFIS estimator to have a relative MSE

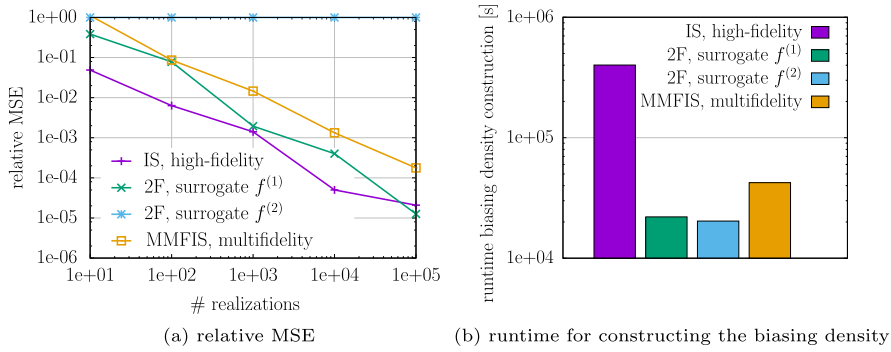


Fig. 3. Beam example: To avoid the selection of a surrogate model for variance reduction in the two-fidelity approach, our MMFIS approach combines all surrogates. In this example, our MMFIS approach shows a relative MSE that is similar to the relative MSE of the two-fidelity estimator with the lowest variance (“2F, surrogate $f^{(1)}$ ”). The runtime for constructing the biasing density with our MMFIS estimator is larger than with the two-fidelity estimators; however, our runtime is still significantly lower than the runtime of using the high-fidelity model only.

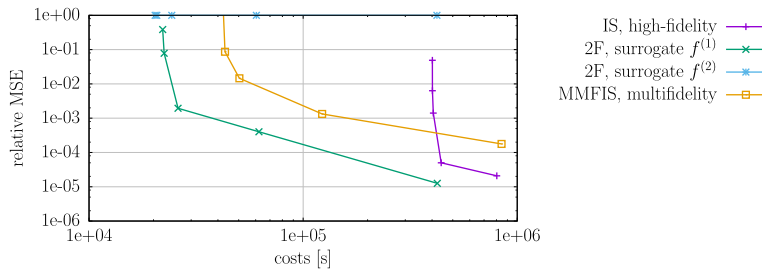


Fig. 4. Beam example: In case the construction of the biasing density dominates the costs of the failure probability estimation, our MMFIS approach (“MMFIS, multifidelity”) achieves an error/cost ratio that lies between the two-fidelity estimator that uses surrogate model $f^{(1)}$ (“2F, surrogate $f^{(1)}$ ”), which requires selecting the surrogate model that leads to the lowest variance of the two surrogate models, and the estimator that uses the high-fidelity model (“IS, high-fidelity”).

similar to the importance sampling estimator $P_{n,m}^{IS}$, with costs that are similar to the costs of the two-fidelity estimators $P_{n,m}^{(1)}$ and $P_{n,m}^{(2)}$.

Fig. 3a shows the relative MSEs of the four estimators. The relative MSE of the two-fidelity estimator $P_{n,m}^{(1)}$ that uses surrogate model $f^{(1)}$ shows a similar behavior to the relative MSE of the importance sampling estimator $P_{n,m}^{IS}$ that uses the high-fidelity model for constructing the biasing density. This confirms that surrogate model $f^{(1)}$ provides a reasonably good approximation of the high-fidelity model for creating a biasing density. In contrast, the relative MSE of the two-fidelity estimator $P_{n,m}^{(2)}$, which uses surrogate model $f^{(2)}$, stays constant near 1 even for a large number of realizations $m = 10^5$. Surrogate model $f^{(2)}$ therefore seems to be a poor approximation of the high-fidelity model, leading to a biasing density that cannot reduce the variance. This experiment shows that the selection of a surrogate model can significantly influence the variance reduction that is achieved with two-fidelity importance sampling. Our MMFIS estimator avoids the problem of selecting a single surrogate model and uses both surrogate models to derive a biasing density. The relative MSE of our MMFIS estimator is similar to the relative MSE of the importance sampling estimator, even though both surrogate models are used.

Consider now the costs of the estimators. The costs of the estimators differ only in the costs of constructing the biasing densities because all four estimators use the high-fidelity model to derive the estimate. The costs of constructing the biasing densities are reported in Fig. 3b. Our MMFIS estimator has significantly lower costs than the importance sampling estimator, and only slightly higher costs than the two-fidelity estimators. This is confirmed in Fig. 4, which shows the relative MSE and cost ratios of the estimators. The costs are the total costs, including the costs of constructing the biasing densities and the costs of the high-fidelity model evaluations for deriving the estimate. The MMFIS estimator $P_{n,m}^{MMFIS}$ has a lower MSE/cost ratio than the importance sampling estimator $P_{n,m}^{IS}$, except for large numbers of realizations m where the costs of constructing the biasing density become negligible.

4.1.4. Beam example: effect of the control variates

The control variate coefficients $\hat{\beta}$ for $n = 10^5$ and $m = 10^4$ are reported in Table 1. The reported numbers are the averages of the coefficients over five runs. First, notice that the control variate coefficients vary by more than one order of magnitude in the beam example (first row of Table 1). The coefficient $\hat{\beta}_1$ corresponding to surrogate model $f^{(1)}$ is larger than the coefficient $\hat{\beta}_2$ of surrogate $f^{(2)}$, which reflects that surrogate $f^{(1)}$ provides a more accurate approximation of the

Table 1

The table reports the control variate coefficients $\hat{\beta}$ obtained with the optimization problem (10) for $n = 10^5$ and $m = 10^4$ realizations.

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
beam	5.1836×10^{-4}	-1.3524×10^{-5}	–
shock propagation	-2.4844×10^{-8}	-1.4552×10^{-4}	1.3282×10^{-3}

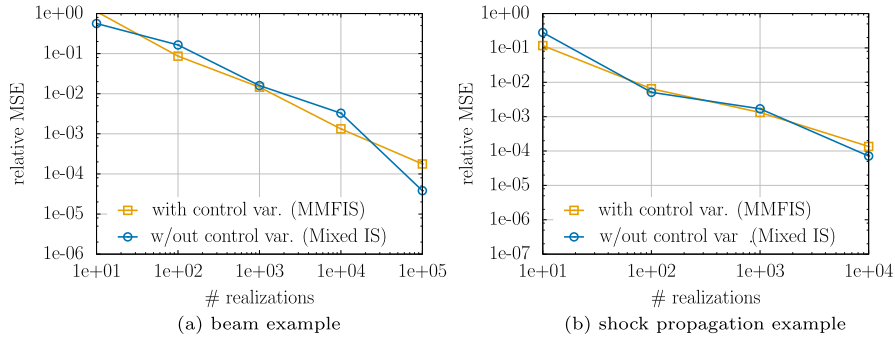


Fig. 5. The plots show that the control variates in our MMFIS estimator have a small effect on the relative MSE in the beam and the shock propagation example; however, the analysis discussed in Section 3.4 shows that our MMFIS estimator, which uses control variates, guarantees bounds on the relative MSE that can be significantly lower than the bounds on the relative MSE of mixture importance sampling estimators, which ignore control variates [36].

high-fidelity model than $f^{(2)}$. Since the MMFIS estimator (12) is the sum of the control variate coefficients, the value of β_1 dominates the value of the MMFIS estimate, and therefore the biasing density derived from $f^{(1)}$ has a larger effect on the MMFIS estimator than the biasing density derived from $f^{(2)}$.

Let us now consider the mixture importance sampling estimator without control variates, which is defined as

$$\hat{P}_{n,m}^{\text{MixIS}} = \frac{1}{m} \sum_{i=1}^m \frac{\mathbb{I}_{\mathcal{G}}(Z'_i) p(Z'_i)}{q_n(Z'_i)}, \tag{15}$$

where the random variables Z'_1, \dots, Z'_m are i.i.d. with the random variable with density q_n . Note that $\hat{P}_{n,m}^{\text{MixIS}}$ uses the same mixture density q_n as our MMFIS estimator $\hat{P}_{n,m}^{\text{MMFIS}}$ but ignores the control variates used in our MMFIS estimator. We refer to, e.g., [34,27] for details on such mixture importance sampling estimators without control variates. The estimator $\hat{P}_{n,m}^{\text{MixIS}}$ equals the multiple importance sampling estimator if the weights are selected with the balance heuristic, see [35] and [36, p. 9]. Fig. 5a compares our MMFIS estimator to the estimator $\hat{P}_{n,m}^{\text{MixIS}}$. In this example, the estimator without control variates $\hat{P}_{n,m}^{\text{MixIS}}$ achieves a similar relative MSE as our MMFIS estimator; however, our MMFIS estimator is guaranteed to achieve a relative MSE that is up to a constant equal or less than the relative MSE of a two-fidelity estimator that uses any of the surrogate models alone, see the discussion in Section 3.4. Mixture importance sampling estimators without control variates achieve weaker bounds on the relative MSE in general, see, e.g., [34,36].

4.1.5. Beam example with two parameters

We extend the beam example introduced in Section 4.1.1 to have the force applied to the tip of the beam as another input. Thus, the input is a vector $\mathbf{z} = [z_1, z_2]^T$, where z_1 is the length input as defined in Section 4.1.1 and z_2 is the force. The corresponding random vector $Z = [Z_1, Z_2]^T$ is normally distributed with mean $[1.1, 1.1]^T$ and covariance matrix

$$\begin{bmatrix} 0.0025 & 0 \\ 0 & 0.0025 \end{bmatrix}.$$

We estimate the probability that the output is larger than 1.86×10^{-4} and obtain a reference probability of $P_{n,m}^{\text{IS}} = 9.3317 \times 10^{-4}$ from $m = 10^4$ high-fidelity model evaluations. We build surrogate models as described in Section 4.1.2, except that we keep 30, instead of 15, basis vectors. The surrogate model $f^{(2)}$ is constructed for the input $\mathbf{z} = [1, 1]^T$. Fig. 6 reports the relative MSE and the runtime of the importance sampling estimator, the two-fidelity estimators, and our MMFIS estimator. Our MMFIS estimator achieves similar results as in the beam example with a single input by providing an estimate that is about as accurate as the importance sampling estimator that uses the high-fidelity model alone with significantly lower runtime costs.

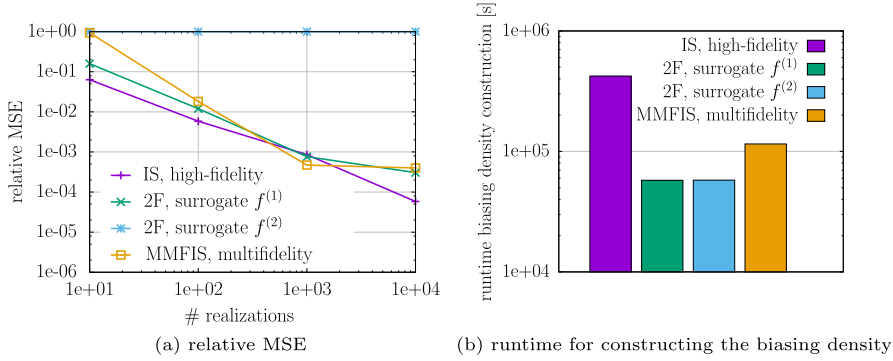


Fig. 6. Beam example with two inputs: Our MMFIS approach provides estimates with similar accuracy and costs as in the beam example with a single input.

4.2. Shock propagation

Consider the Burgers' equation

$$\frac{\partial u(x, t; \nu, z)}{\partial t} + u(x, t; \nu, z) \frac{\partial u(x, t; \nu, z)}{\partial x} = \nu \frac{\partial^2 u(x, t; \nu, z)}{\partial x^2}, \quad (16)$$

where $x \in [0, 1]$ is the spatial variable, $t \in (0, \infty)$ is time, $\nu \in [10^{-3}, 10^{-4}]$ is the viscosity parameter, $z \in \mathcal{D} = \mathbb{R}$ is the input, and $u : [0, 1] \times [0, \infty) \times [10^{-3}, 10^{-4}] \times \mathcal{D} \rightarrow \mathbb{R}$ is the solution function. The initial condition is

$$u(x, 0; \nu, z) = \begin{cases} 0, & \text{if } x - z \geq 0.5, \\ 1, & \text{else.} \end{cases} \quad (17)$$

The input $z \in \mathcal{D}$ defines the initial position of the shock in the spatial domain $[0, 1]$ through the initial condition (17), and the output is the position of the shock at time $t = 1$.

4.2.1. Shock propagation: high-fidelity and surrogate models

We discretize the Burgers' equation (16) in the spatial domain with the finite element method with piecewise linear finite elements. We use an equidistant grid of 512 grid points x_1, \dots, x_{512} in the spatial domain $[0, 1]$. The time domain is discretized with the implicit Euler method at t_1, \dots, t_{1000} time steps with time step size $\delta t = 10^{-3}$. The high-fidelity model $f : \mathcal{D} \rightarrow \mathcal{Y}$ maps the input onto the shock position at time $t = 1$, with the viscosity parameter set to $\nu = 10^{-4}$. The shock position is determined by searching for the maximal value of the state at the grid points x_1, \dots, x_{512} .

Define a trajectory as a matrix $\mathbf{U}(\nu, z) \in \mathbb{R}^{512 \times 1000}$ with

$$\mathbf{U}(\nu, z) = \begin{bmatrix} u(x_1, t_1; \nu, z) & \dots & u(x_1, t_{1000}; \nu, z) \\ \vdots & \ddots & \vdots \\ u(x_{512}, t_1; \nu, z) & \dots & u(x_{512}, t_{1000}; \nu, z) \end{bmatrix} \in \mathbb{R}^{512 \times 1000},$$

which depends on the parameter ν and the input z . We construct a global POD basis from the matrix

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}(10^{-3}, -0.2)^T \\ \mathbf{U}(10^{-4}, -0.2)^T \\ \mathbf{U}(10^{-3}, -0.125)^T \\ \mathbf{U}(10^{-4}, -0.125)^T \\ \mathbf{U}(10^{-3}, -0.05)^T \\ \mathbf{U}(10^{-4}, -0.05)^T \end{bmatrix}.$$

We keep 30 POD basis vectors and create the POD reduced model operators by (Galerkin) projection of the high-fidelity model operators onto the space spanned by the first 30 POD basis vectors.

For low viscosities, the convection term dominates the flow behavior. It has been observed that surrogate models based on low-dimensional POD spaces can lose long-term stability. Therefore, such models have to be stabilized, e.g., through additive closure terms [37–42], which are motivated by Large Eddy Simulation. This is an active area of research, and finding a proper closure model is still an open problem. In many situations, closure terms are derived in an ad-hoc way and with domain expertise. We consider three closure terms discussed in [37]. Surrogate model $f^{(1)}$ uses the closure term that is called “POD-ROM-MK” in [37, Section 3.3.1] and which is based on the spectral vanishing viscosity concept of [43].

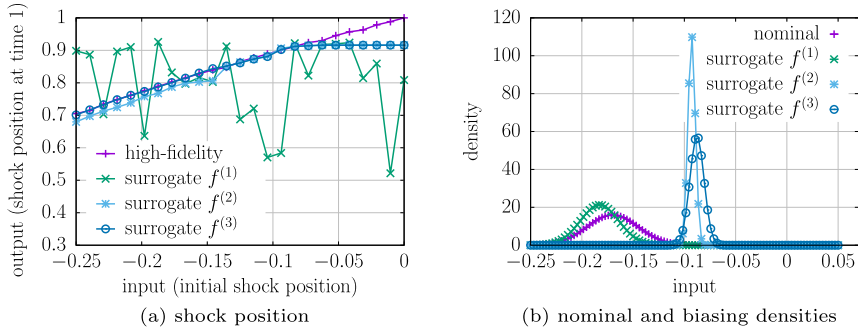


Fig. 7. Burgers' equation: The plot in (a) shows the shock positions at final time $t = 1$ as predicted by the three surrogate models and compares the positions to the position obtained with the high-fidelity model. Surrogate model $f^{(1)}$ is a poor approximation of the high-fidelity model, which is reflected in the corresponding biasing densities as shown in (b).

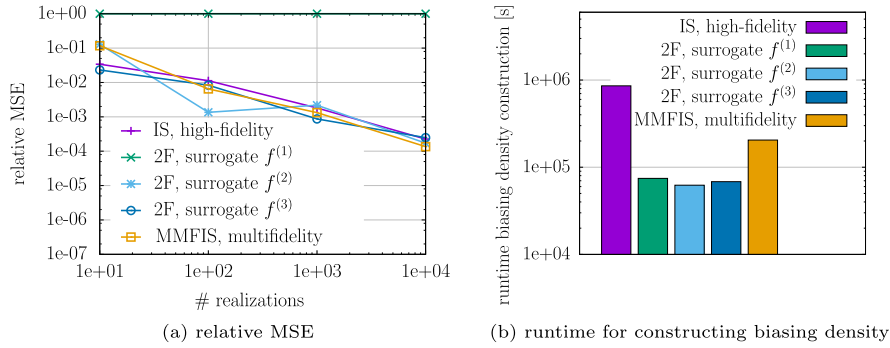


Fig. 8. Burgers' equation: The relative MSE of our MMFIS estimator is similar to the relative MSEs of the two-fidelity estimators that either use surrogate model $f^{(2)}$ or $f^{(3)}$. Additionally, the runtime for constructing the biasing density with our MMFIS approach is significantly lower than the runtime of the importance sampling estimator that uses the high-fidelity model.

The closure term “POD-ROM-H” [37, Section 3.3.1] of surrogate model $f^{(2)}$ uses the eddy viscosity concept to compensate for the lost energy due to the POD representation [44]. Surrogate model $f^{(3)}$ employs the closure term “POD-ROM-CL” [37, Section 3.3.1], which is another closure term based on the spectral vanishing viscosity concept. We refer to [37] for details on these closure terms and for further references.

4.2.2. Shock propagation: failure probability estimation

We consider the input random variable Z that is normally distributed with mean -0.17 and standard deviation 0.025 . The input random variable Z determines the initial position of the shock. We are interested in the limit state function $g_{\text{Burgers}}(y) = -y + 0.9$, which means we estimate the probability that the shock position is larger than 0.9 at time $t = 1$. We set $\nu = 10^{-4}$. The importance sampling estimator $P_{n,m}^{\text{IS}}$ gives the reference probability 1.18×10^{-3} , with $m = 10^6$ realizations for the estimation, and $n = 10^5$ samples for constructing the biasing density with the high-fidelity model.

4.2.3. Shock propagation: results

Fig. 7a shows the output of the high-fidelity model and the three surrogate models for inputs in $[-0.25, 0]$. The surrogate model $f^{(1)}$ leads to a poor approximation of the high-fidelity model. The surrogate models $f^{(2)}$ and $f^{(3)}$ predict a similar shock position as the high-fidelity model for inputs in the range $[-0.25, -0.08]$. Consider now the corresponding biasing densities $q_n^{(1)}, q_n^{(2)}, q_n^{(3)}$ plotted in Fig. 7b. The biasing densities reflect the result of Fig. 7a. The surrogate models $f^{(2)}$ and $f^{(3)}$ lead to biasing densities with mean near -0.1 . Fig. 7a shows that inputs near -0.1 lead to a shock position near 0.9 , and therefore the biasing densities $q_n^{(2)}$ and $q_n^{(3)}$ seem to be reasonable. In contrast, the biasing density $q_n^{(1)}$ derived from the surrogate model $f^{(1)}$ corresponds to a similar mean as the nominal random variable Z , which means that the corresponding two-fidelity estimator has a similar variance as the basic Monte Carlo estimator.

Fig. 8 shows the relative MSE of the two-fidelity estimators $P_{n,m}^{(1)}, P_{n,m}^{(2)}, P_{n,m}^{(3)}$ that use a single surrogate model only, the relative MSE of the importance sampling estimator $P_{n,m}^{\text{IS}}$, and the relative MSE of our MMFIS estimator $\hat{P}_{n,m}^{\text{MMFIS}}$. The relative MSE of the two-fidelity estimator $P_{n,m}^{(1)}$ reflects that the surrogate model $f^{(1)}$ is a poor approximation of the high-fidelity model. In contrast, our MMFIS estimator achieves a similar relative MSE to the importance sampling estimator that uses the high-fidelity model for constructing the biasing density. Table 1 reports the control variate coefficients $\hat{\beta}$, which reflect that $f^{(2)}$ and $f^{(3)}$ provide a sufficiently accurate approximation of f for constructing a biasing density. A comparison of

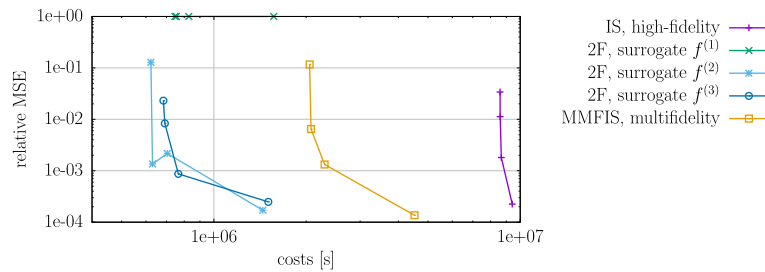


Fig. 9. Burgers' equation: The plot reports the relative MSE and costs ratios of the estimators that use a single model only and our MMFIS estimator. Our MMFIS estimator has a better error/cost ratio than using the high-fidelity model only, and a significantly better ratio than the two-fidelity estimator that uses surrogate model $f^{(1)}$.

our MMFIS estimator with the importance sampling estimator that uses the mixture density (15) is shown in Fig. 5b. Our MMFIS estimator and the importance sampling estimator achieve comparable relative MSEs. Note the discussion on the variance bounds of the two estimators in Section 4.1.4.

Fig. 8b compares the costs of constructing the biasing densities of the five estimators. The MMFIS estimator is more expensive than the two-fidelity estimators, because all surrogate models are evaluated, but the MMFIS estimator is significantly cheaper than the importance sampling estimator $P_{n,m}^{IS}$ that uses the high-fidelity model only. In that sense, our MMFIS is cheaper than using the high-fidelity model only and it guarantees to achieve a variance reduction up to a constant as selecting the two-fidelity estimator with the lowest variance. Fig. 9 compares the five estimators with respect to the ratio of relative MSE and costs. The results confirm that the MMFIS estimator achieves a relative MSE/costs ratio between the best two-fidelity estimator and the importance sampling estimator that uses the high-fidelity model only. However, the MMFIS estimator is robust, since its error/costs ratio does not depend upon *a priori* selection of the best surrogate model.

5. Conclusions

We presented a multifidelity approach to importance sampling that leverages multiple surrogate models for speeding up the construction of a biasing distribution and that uses the high-fidelity model to derive an unbiased estimate of the failure probability. The key property of our multifidelity approach is that it can leverage all given surrogate models and so avoids the problem of selecting a single surrogate model. In that sense, our multifidelity approach goes beyond the usual two-fidelity approaches that combine a single surrogate model with the high-fidelity model. Following [27], we proved that our multifidelity estimator has an MSE that is up to a constant as low as using only the surrogate model that leads to the lowest MSE of all given surrogate models. Our multifidelity approach therefore is more robust than a two-fidelity approach that relies on an *a priori* selection of a surrogate model. In particular, the multifidelity approach is a good option in situations where multiple surrogate models are available, but their relative quality is unknown. Our numerical results provided evidence of the theoretical analysis.

Acknowledgements

This work was supported by the DARPA EQUiPS Program, Award UTA15-001067, Program Manager F. Fahroo, and by the AFOSR MURI on multi-information sources of multi-physics systems, Award Number FA9550-15-1-0038, Program Manager J.-L. Cambier. Several examples were computed on the computer clusters of the Munich Centre of Advanced Computing.

References

- [1] C. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, 2004.
- [2] R. Srinivasan, Importance Sampling, Springer, 2002.
- [3] J. Buckle, Introduction to Rare Event Simulation, Springer, 2003.
- [4] J.-M. Marin, A. Mira, C. Robert, Adaptive multiple importance sampling, Scand. J. Stat. 39 (4) (2012) 798–812.
- [5] R. Douc, A. Guillin, J.-M. Marin, C. Robert, Minimum variance importance sampling via population Monte Carlo, ESAIM, Probab. Stat. 11 (2007) 427–447.
- [6] R. Douc, A. Guillin, J.-M. Marin, C. Robert, Convergence of adaptive mixtures of importance sampling schemes, Ann. Stat. 35 (1) (2007) 420–448.
- [7] O. Cappé, R. Douc, A. Guillin, J.-M. Marin, C. Robert, Adaptive importance sampling in general mixture classes, Stat. Comput. 18 (4) (2008) 447–459.
- [8] O. Cappé, A. Guillin, J.M. Marin, C. Robert, Population Monte Carlo, J. Comput. Graph. Stat. 13 (4) (2004) 907–929.
- [9] P.-T.d. Boer, D. Kroese, S. Mannor, R. Rubinstein, A tutorial on the cross-entropy method, Ann. Oper. Res. 134 (1) (2005) 19–67.
- [10] R. Rubinstein, Optimization of computer simulation models with rare events, Eur. J. Oper. Res. 99 (1) (1997) 89–112.
- [11] S.-K. Au, J.L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probab. Eng. Mech. 16 (4) (2001) 263–277.
- [12] K.M. Zuev, J.L. Beck, S.-K. Au, L.S. Katafygiotis, Bayesian post-processor and other enhancements of subset simulation for estimating failure probabilities in high dimensions, Comput. Struct. 92–93 (2012) 283–296.
- [13] V. Papadopoulos, D.G. Giovanis, N.D. Lagaros, M. Papadrakakis, Accelerated subset simulation with neural networks for reliability analysis, Comput. Methods Appl. Mech. Eng. 223–224 (2012) 70–80.
- [14] E. Ullmann, I. Papaioannou, Multilevel estimation of rare events, SIAM/ASA J. Uncertain. Quantificat. 3 (1) (2015) 922–953.
- [15] G. Rozza, D. Huynh, A. Patera, Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations, Arch. Comput. Methods Eng. 15 (3) (2007) 1–47.

- [16] P. Benner, S. Gugercin, K. Willcox, A survey of projection-based model reduction methods for parametric dynamical systems, *SIAM Rev.* 57 (4) (2015) 483–531.
- [17] A. Forrester, A. Sóbester, A. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*, Wiley, 2008.
- [18] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (3) (1995) 273–297.
- [19] A.J. Majda, B. Gershgorin, Quantifying uncertainty in climate change science through empirical information theory, *Proc. Natl. Acad. Sci. USA* 107 (34) (2010) 14958–14963.
- [20] M. Valdebenito, H. Pradlwarter, G. Schuëller, The role of the design point for calculating failure probabilities in view of dimensionality and structural nonlinearities, *Struct. Saf.* 32 (2) (2010) 101–111.
- [21] J. Li, D. Xiu, Evaluation of failure probability via surrogate models, *J. Comput. Phys.* 229 (23) (2010) 8966–8980.
- [22] A. Basudhar, S. Missoum, An improved adaptive sampling scheme for the construction of explicit boundaries, *Struct. Multidiscip. Optim.* 42 (4) (2010) 517–529.
- [23] B. Peherstorfer, K. Willcox, Online adaptive model reduction for nonlinear systems via low-rank updates, *SIAM J. Sci. Comput.* 37 (4) (2015) A2123–A2150.
- [24] J. Li, J. Li, D. Xiu, An efficient surrogate-based method for computing rare failure probability, *J. Comput. Phys.* 230 (24) (2011) 8683–8697.
- [25] P. Chen, A. Quarteroni, Accurate and efficient evaluation of failure probability for partial differential equations with random input data, *Comput. Methods Appl. Mech. Eng.* 267 (2013) 233–260.
- [26] B. Peherstorfer, T. Cui, Y. Marzouk, K. Willcox, Multifidelity importance sampling, *Comput. Methods Appl. Mech. Eng.* 300 (2016) 490–509.
- [27] A. Owen, Y. Zhou, Safe and effective importance sampling, *J. Am. Stat. Assoc.* 95 (449) (2000) 135–143.
- [28] P. Mills, Efficient statistical classification of satellite measurements, *Int. J. Remote Sens.* 32 (21) (2011) 6109–6132.
- [29] B. Peherstorfer, D. Pflüger, H.-J. Bungartz, Density estimation with adaptive sparse grids for large data sets, in: *Proceedings of the 2014 SIAM International Conference on Data Mining*, SIAM, 2014, pp. 443–451, ch. 50.
- [30] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 1986.
- [31] B. Nelson, On control variate estimators, *Comput. Oper. Res.* 14 (3) (1987) 219–225.
- [32] G. Fishman, *Monte Carlo*, Springer, 1996.
- [33] H. Panzer, J. Hubele, R. Eid, B. Lohmann, *Generating a Parametric Finite Element Model of a 3D Cantilever Timoshenko Beam Using Matlab*, Tech. rep., Chair of Automatic Control, Technische Universität München, 2009.
- [34] T. Hesterberg, Weighted average importance sampling and defensive mixture distributions, *Technometrics* 37 (2) (1995) 185–194.
- [35] E. Veach, L.J. Guibas, Optimally combining sampling techniques for Monte Carlo rendering, in: *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '95*, ACM, New York, NY, USA, 1995, pp. 419–428.
- [36] H.Y. He, A.B. Owen, Optimal mixture weights in multiple importance sampling, arXiv e-prints, arXiv:1411.3954.
- [37] O. San, T. Iliescu, Proper orthogonal decomposition closure models for fluid flows: Burgers equation, *Int. J. Numer. Anal. Model., Ser. B* 5 (3) (2014) 217–237.
- [38] Z. Wang, I. Akhtar, J. Borggaard, T. Iliescu, Two-level discretizations of nonlinear closure models for proper orthogonal decomposition, *J. Comput. Phys.* 230 (1) (2011) 126–146.
- [39] Z. Wang, I. Akhtar, J. Borggaard, T. Iliescu, Proper orthogonal decomposition closure models for turbulent flows: a numerical comparison, *Comput. Methods Appl. Mech. Eng.* 237–240 (2012) 10–26.
- [40] S. Sirisup, G.E. Karniadakis, A spectral viscosity method for correcting the long-term behavior of POD models, *J. Comput. Phys.* 194 (1) (2004) 92–116.
- [41] O. San, J. Borggaard, Basis selection and closure for POD models of convection dominated Boussinesq flows, in: *21st International Symposium on Mathematical Theory of Networks and Systems*, 2014, pp. 132–139, Groningen, The Netherlands.
- [42] M. Benosman, B. Kramer, P.T. Boufounos, P. Grover, Learning-based reduced order model stabilization for partial differential equations: application to the coupled Burgers' equation, in: *2016 American Control Conference (ACC)*, Boston, MA, 2016, pp. 1673–1678, <http://dx.doi.org/10.1109/ACC.2016.7525157>.
- [43] E. Tadmor, Convergence of spectral methods for nonlinear conservation laws, *SIAM J. Numer. Anal.* 26 (1) (1989) 30–44.
- [44] N. Aubry, P. Holmes, J.L. Lumley, E. Stone, The dynamics of coherent structures in the wall region of a turbulent boundary layer, *J. Fluid Mech.* 192 (1988) 115–173.