Sensitivity-driven adaptive construction of reduced-space surrogates

Manav Vohra[1], Alen Alexanderian[2], Cosmin Safta[3], Sankaran Mahadevan[1]

[1]Department of Civil and Environmental Engineering
Vanderbilt University
Nashville, TN 37235

[2]Department of Mathematics
North Carolina State University
Raleigh, NC 27695

[3]Sandia National Laboratories
Livermore, CA 94550

Corresponding Author: Sankaran Mahadevan
Department of Civil and Environmental Engineering
Vanderbilt University
272 Jacobs Hall, VU Mailbox: PMB 351831
Nashville, TN 37235

Phone: (615) 322-3040
Fax: (615) 343-3773
Email: sankaran.mahadevan@vanderbilt.edu

Submitted to: *Journal of Scientific Computing*
July 2018

Abstract

Surrogate modeling has become a critical component of scientific computing in situations involving expensive model evaluations. However, training a surrogate model can be remarkably challenging and even computationally prohibitive in the case of intensive simulations and large-dimensional systems. We develop a systematic approach for surrogate model construction in reduced input parameter spaces. A sparse set of model evaluations in the original input space is used to approximate derivative based global sensitivity measures (DGSMs) for individual uncertain inputs of the model. An iterative screening procedure is developed that exploits DGSM estimates in order to identify the *unimportant* inputs. The screening procedure forms an integral part of an overall framework for adaptive construction of a surrogate in the reduced space. The framework is tested for computational efficiency through an initial implementation in simple test cases such as the classic Borehole function, and a semilinear elliptic PDE with a random source function. The framework is then deployed for a realistic application from chemical kinetics, where we study the ignition delay in an H_2/O_2 reaction mechanism with 19 and 33 uncertain rate-controlling parameters. It is observed that significant computational gains can be attained by constructing accurate low-dimensional surrogates using the proposed framework.

Keywords: Global sensitivity analysis, polynomial chaos, parameter screening, surrogate modeling

1 Introduction

Computing statistical properties of complex systems containing uncertain parameters is challenging due to the need for a large number of expensive model evaluations. Surrogate models make such computations possible by replacing expensive model runs with cheap evaluations of an approximate model. Commonly used approaches for surrogate modeling include polynomial chaos expansions (PCEs) [1–3], multivariate adaptive regression splines (MARS) [4], Gaussian process (GP) and Kriging models [5,6], and neural networks [7,8]. A pertinent challenge associated with surrogate construction involves the process of *training* that typically requires a large number of model runs. In situations involving large-dimensional systems and computationally intensive simulations, the training of a surrogate becomes challenging and even prohibitive in some cases. A possible approach for mitigating this challenge involves computation of global sensitivity measures such as Sobol' indices [9] to identify the *important parameters*—i.e., parameters that predominantly contribute towards the variability of the model output.

Sensitivity-based dimension reduction can help reduce the computational effort associated with surrogate construction, significantly. However, numerically estimating the Sobol' indices, which involve multi-dimensional integrals, can be a demanding task in itself. In fact, because of the large number of samples required, computation of Sobol' indices is often done using surrogate models, as shown in a variety of applications including in ocean modeling [10, 11], geosciences [12–14], and chemical kinetics [15–17] to name a few. We are thus, faced with a "chicken-and-egg" problem: on the one hand, computing global sensitivity measures enables dimension reduction, which in turn enables efficient surrogate model construction. On the other hand, computing sensitivity measures is expensive and is often made possible by using surrogate models. Development of an approach for overcoming this chicken-and-egg problem is a major goal of this article.

In this article, we propose a practical and efficient approach that focuses on optimal use of computational resources for surrogate construction in a reduced input space. Specifically, we reduce the dimensionality of the input space by using derivative-based global sensitivity analysis [18–22], which enables a tractable approach for global sensitivity analysis [22].

The links between derivative-based global sensitivity measures (DGSMs) and total Sobol' indices [18, 21, 22] provide a strong basis for their use in identifying important uncertainty contributors. In addition to the construction of an efficient surrogate in the reduced space, dimension reduction highlights key features of the input-output relationship encapsulated by the model, and allows for an efficient approach to calibration of the important model parameters.

We present a strategy for identifying and screening uncertain model inputs that are significantly less important than the rest, thereby reducing the dimensionality of the problem and enabling the construction of a reduced-space surrogate (RSS). Our approach combines DGSMs and surrogate modeling in an iterative manner. To make optimum use of computational resources, batches of model evaluations are performed iteratively, and convergence of our DGSM-based screening metric is tested successively. Moreover, a series of verification steps incorporated in our method enable monitoring the accuracy of the parameter screening and the resulting surrogate model. Our approach is agnostic to the choice of methodology for constructing the surrogate. However, in the present work, we rely on sparse polynomial chaos expansions (PCEs) to demonstrate the suitability of the proposed strategy.

The contributions of this article are as follows: (i) We establish a robust and practical framework for dimension reduction and surrogate modeling using derivative-based global sensitivity measures. Our approach is general in that it is applicable to a wide range of applications. (ii) We present comprehensive numerical results demonstrating the viability of our strategy using motivating applications: the classic borehole function, and a semi-linear elliptic PDE. (iii) We deploy our strategy in an application problem from chemical kinetics with 19 uncertain rate-controlling parameters. The problem is studied in multiple regimes. It is shown that the 19-parameter problem can be efficiently reduced to a 3- or 4-dimensional problem. The strategy is further implemented to a higher dimensional problem involving 33 uncertain rate-controlling parameters. In this case, it is shown that only 2 out of 33 parameters predominantly contribute to the uncertainty in the quantity of interest (QoI). Moreover, computational gains using the proposed strategy are shown to increase significantly with dimensionality.

The remainder of this article is structured as follows. In section 2, we provide a brief

introduction to DGSMs as well as the polynomial chaos methodology used in the present work. In section 3, we present our proposed approach, where we also provide a detailed numerical algorithm and a flow diagram to aid practitioners in implementing the presented framework. Section 4 is devoted to numerical examples examining various aspects of our approach. This is followed by implementation of our framework in a H_2/O_2 chemical kinetics problem, in section 5. Finally, concluding remarks are provided in section 6.

2 Background

In this section, we introduce the notations used in the rest of the article, and present the requisite background material on derivative-based global sensitivity measures and surrogate modeling using polynomial chaos expansions.

2.1 Derivative-based global sensitivity analysis

Let G be a mathematical model that is a function of N_p uncertain inputs, $\theta_1, \theta_2, \ldots, \theta_{N_\mathrm{p}}$. The goal of sensitivity analysis is measuring the influence of each component of the input vector $\boldsymbol{\theta} = \begin{bmatrix}\theta_1 & \theta_2 & \ldots & \theta_{N_\mathrm{p}}\end{bmatrix}^T$ on the model output. In the present work, we consider the case where the inputs are statistically independent.

Derivative-based global sensitivity analysis is performed by computing derivative based global sensitivity measures (DGSMs) [18] for each uncertain parameter in the model. Specifically, we consider the following DGSMs,

$$\mu_i = \mathbb{E}\left[\left(\frac{\partial G(\boldsymbol{\theta})}{\partial \theta_i}\right)^2\right], \quad i = 1, \ldots, N_\mathrm{p}. \tag{1}$$

Here $\mathbb{E}$ denotes expectation over the uncertain parameters. Notice that this formulation assumes that the function G is differentiable with respect to θ_i, $i = 1, \ldots, N_\mathrm{p}$ and the partial derivatives, $\left(\frac{\partial G(\boldsymbol{\theta})}{\partial \theta_i}\right)$ are square integrable.

If an analytic expression for G is available, the derivative in the above expression can be computed directly. In real-world applications, however, G is often defined in terms of a solution of a mathematical model. In the present work, we consider a generic computational

model and only assume that the model output depends differentiably on the parameter $\boldsymbol{\theta}$. A simple approach to computing the gradient is to use finite-differences:

$$\frac{\partial G(\boldsymbol{\theta})}{\partial \theta_i} \approx \frac{G(\theta_1, \ldots, \theta_{i-1}, \theta_i + \Delta\theta_i, \theta_{i+1}, \ldots, \theta_d) - G(\boldsymbol{\theta})}{\Delta\theta_i}, \quad i = 1, \ldots, N_{\mathrm{p}}. \tag{2}$$

Then, (1) can be evaluated by Monte Carlo (MC) sampling in the uncertain parameter space. The total number of model realizations or function evaluations needed to compute μ_i for a function G of N_{p} random inputs and using N samples is therefore, $N \times (N_{\mathrm{p}} + 1)$. It is noted in previous studies [21, 22], and also observed in the numerical experiments in the present work, that a modest MC sample size is often sufficient for computing (1) with reasonable accuracy to identify the unimportant inputs. Moreover, the computational efficiency for estimating μ_i can be enhanced by using techniques such as automatic differentiation [23] or adjoint-based gradient computation [24–27].

Consider the total Sobol' sensitivity index [28],

$$\mathcal{T}(\theta_i) = 1 - \frac{\mathbb{V}[\mathbb{E}(G|\boldsymbol{\theta}_{\sim i})]}{\mathbb{V}(G)}, \tag{3}$$

where $\boldsymbol{\theta}_{\sim i}$ is the random vector $\boldsymbol{\theta}$ with ith component removed, and $\mathbb{V}$ denotes the variance. The total Sobol' index quantifies the total contribution of θ_i to variance of the model G. Components of $\boldsymbol{\theta}$ with small total Sobol' index can be considered inessential and can be fixed at nominal values. However, computing the total Sobol' index is a computationally expensive task for expensive-to-evaluate models with large number of uncertain parameters. Fortunately, for parameters with continuous distributions, an upper bound on $\mathcal{T}_i$ can be expressed in terms of μ_i as follows:

$$\mathcal{T}(\theta_i) \leq \frac{\mathcal{C}_i \mu_i}{\mathbb{V}(G)}, \tag{4}$$

where $\mathcal{C}_i$ is the corresponding "Poincaré constant" and $\mathbb{V}(G)$ is the total variance of the model output [20]. The upper bound in the above inequality is proportional to the product of $\mathcal{C}_i$ and μ_i. For the purpose of parameter screening as discussed later in Section 3, we consider a normalized product, $\widehat{\mathcal{C}_i \mu_i}$ to ensure that it lies between 0 and 1:

$$\widehat{\mathcal{C}_i \mu_i} = \frac{\mathcal{C}_i \mu_i}{\sum_i \mathcal{C}_i \mu_i}. \tag{5}$$

The Poincaré constant, $\mathcal{C}_i$ is specific to the probability distribution of θ_i. For $\theta_i \sim \mathcal{U}[a,b]$, $\mathcal{C}_i = (b-a)^2/\pi^2$, and for $\theta_i \sim \mathcal{N}(\mu,\sigma^2)$, we have $\mathcal{C}_i = \sigma^2$ (see [20] for more details). Here $\mathcal{N}(\mu,\sigma^2)$ denotes a normal distribution with mean μ and variance σ^2, and $\mathcal{U}[a,b]$ denotes a uniform distribution on the interval $[a,b]$.

2.2 Polynomial chaos expansion

We consider models with N_p random inputs, $\theta_1,\ldots,\theta_{N_\mathrm{p}}$ that are modeled as statistically independent random variables. The variables θ_i will take in physically meaningful ranges; it is common to parameterize input uncertainties with canonical random variables $\xi_1,\ldots,\xi_{N_\mathrm{p}}$, which can be then shifted and scaled to obtain the corresponding $\theta_i's$. Typical choices for distribution of ξ_i include standard normal and uniform distribution on the interval $[-1,1]$. Let

$$\boldsymbol{f}(\boldsymbol{x}) = \prod_{i=1}^{N_\mathrm{p}} f_i(x_i), \quad \boldsymbol{x} \in \mathbb{R}^{N_\mathrm{p}}$$

where f_i are probability density functions of ξ_i, $i = 1,\ldots,N_\mathrm{p}$.

Consider a square integrable random variable $G : \mathbb{R}^{N_\mathrm{p}} \to \mathbb{R}$; i.e., $\int_{\mathcal{D}} G(\boldsymbol{\xi})^2\, \boldsymbol{f}(\boldsymbol{\xi}) d\boldsymbol{\xi} < \infty$, where $\mathcal{D}$ is the support of the distribution law of the random vector $\boldsymbol{\xi}$. The PCE of G is a mean-square convergent series expansion [1–3] of the form:

$$G(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} c_k \Psi_k(\boldsymbol{\xi}), \tag{6}$$

where Ψ_k's form a multivariate orthogonal polynomial basis—orthogonal with respect to the joint probability distribution of $\boldsymbol{\xi}$. In practice, a truncated expansion is used. Moreover, in applications, G is a mathematical model of interest that takes a parameter vector $\boldsymbol{\theta}$ (with components in physically meaningful ranges) as input. Therefore, we write the truncated PC representation of a model G as follows:

$$G(\boldsymbol{\theta}) \approx G^{\mathrm{PC}}(\boldsymbol{\theta}) := \sum_{k=0}^{N_{\mathrm{PC}}} c_k \Psi_k(\boldsymbol{\xi}(\boldsymbol{\theta})), \tag{7}$$

where $\boldsymbol{\xi}(\boldsymbol{\theta})$ is found by a simple linear transformation.

Computational strategies available for estimating the PC coefficients (c_k's) typically involve techniques based on projection or regression. Projection-based methods consider the

orthogonal projection of G on the PC basis $\{\Psi_k\}_{k=0}^{N_{\text{PC}}}$ and compute the resulting expansion coefficients via quadrature [3]. Regression-based methods such as least angle regression (LAR) [29], and least absolute shrinkage and selection operator (LASSO) [30] aim to construct a sparse PCE [31] by solving a penalized least-squares problem. Specifically in the case of LAR, a penalty term comprising the ℓ_1-norm of the PC coefficients is used:

$$\hat{\boldsymbol{c}} = \text{argmin } \mathbb{E}_{\boldsymbol{\theta}} \left[\left(\sum_{k=0}^{N_{\text{PC}}} c_k \Psi_k(\boldsymbol{\xi}(\boldsymbol{\theta})) - G(\boldsymbol{\theta}) \right)^2 \right] + \lambda \left\| \boldsymbol{c} \right\|_1 , \tag{8}$$

where $\|\boldsymbol{c}\|_1 = \sum_{k=0}^{N_{\text{PC}}} |c_k|$. The penalty term forces the minimization towards sparse coefficient vectors resulting in sparse PC representations. In this work, we construct sparse PCEs with LAR using UQLab [32], a general purpose uncertainty quantification software developed at ETH Zurich.

3 Methodology

In this section, we outline the underlying framework for *adaptively* constructing a reduced-space surrogate (RSS) using sensitivity analysis. The proposed methodology is described as adaptive since the RSS is constructed only in situations where it is expected to yield computational dividend as discussed further below. The term reduced-space implies that the surrogate is constructed in a reduced parameter space that sufficiently captures the uncertainty in the model output. We begin by outlining an algorithm for parameter screening to assess the importance of individual parameters for potential dimension reduction and construction of an RSS. The overall adaptive framework that incorporates parameter screening as an integral step is thereafter presented. Finally, we present metrics used for assessing the convergence and accuracy of the RSS followed by a brief discussion on salient features of the proposed framework.

Parameter screening. Dimension reduction and subsequent construction of the RSS is enabled by identifying key contributors to the uncertainty in the QoI. For this purpose, we estimate the upper-bound $\widehat{\mathcal{C}_i \mu_i}$, given in (4), on total Sobol' index ($\mathcal{T}(\theta_i)$) for each param-

eter θ_i; the *screening metrics*, $\{\widehat{\mathcal{C}_i\mu_i}\}_{i=1}^{N_\mathrm{p}}$, are used to identify parameters that are relatively unimportant.

An initial set of n_1 samples is used to obtain a rough estimate of the metric. Based on the associated metric value, an initial rank ($\mathcal{R}_i^{old}$) is assigned to each parameter. At each iteration, a new set of samples is generated based on the joint probability distribution of $\boldsymbol{\theta}$ and corresponding model output at each sample point is computed. The new set of gradient evaluations combined with prior evaluations is used to update parameter ranks. Additionally, deviation in the derivative-based sensitivity measure between successive iterations normalized by the measure in the previous iteration is recorded for each parameter. The iterative process is continued until parameter ranks between successive iterations are observed to be consistent as well as the maximum deviation among all parameters ($\Delta\mu_s$) is below a certain tolerance (τ). The amount of computational effort associated with the screening process is limited by the choice of maximum number of iterations, $s_{\max}$.

Key inputs to the screening procedure are as follows: (1) a limiting value τ of the maximum relative change in the sensitivity measure between successive iterations; (2) a limiting ratio τ_{screen} of the sensitivity metric relative to its maximum value; (3) a real number $\beta \in (0,1)$ to guide the number of new samples $\lceil \beta n_1 \rceil$ at each iteration ($\lceil \beta n_1 \rceil$ is the smallest integer greater than or equal to βn_1); (4) a set of samples $\{\boldsymbol{\theta}_k\}_{k=1}^{n_1}$ for the initial screening step in the algorithm and the corresponding gradient evaluations $\{\boldsymbol{g}^k\}_{k=1}^{n_1}$, where $\boldsymbol{g}^k = \nabla_{\boldsymbol{\theta}} G(\boldsymbol{\theta}_k)$. The outputs are the set of active indices $\mathcal{I}_{\text{active}}$ corresponding to the *important parameters*, the total number of available model evaluations N_{total}, and the enriched set of gradient evaluations $\{\boldsymbol{g}^k\}_{k=1}^{N_{\text{total}}}$. A general methodology for parameter screening is provided below in Algorithm 1.

Algorithm 1 Parameter screening with DGSMs: A generalized approach.

Input: $\tau > 0$, $\tau_{\text{screen}} > 0$, $s_{\min} \geq 1$, $s_{\max} \geq 1$, $\beta > 0$, $\{\boldsymbol{\theta}_k\}_{k=1}^{n_1}$, $\{\boldsymbol{g}^k\}_{k=1}^{N_{\text{total}}}$.

Output: $\mathcal{I}_{\text{active}}$, $\{\boldsymbol{g}^k\}_{k=1}^{N_{\text{total}}}$, N_{total}.

1: **procedure** SCREENING

2: Compute $\boldsymbol{g}^k = \nabla_{\boldsymbol{\theta}} G(\boldsymbol{\theta}_k)$, $k = N_{\text{total}} + 1, \ldots, N_{\text{total}} + n_1$.

3: $N_{\text{total}} = N_{\text{total}} + n_1$

Compute $\mu_{1,i} = \frac{1}{N_{\text{total}}} \sum_{k=1}^{N_{\text{total}}} (g_i^k)^2$

Compute $\nu_i = \widehat{\mathcal{C}_i \mu_{1,i}}$, for each θ_i, $i = 1, \ldots, N_{\text{p}}$.

Determine initial ranks: let $\mathcal{R}^{old} = \{\nu_{i_1}, \nu_{i_2}, \ldots, \nu_{i_{N_{\text{p}}}}\}$ such that

$$\nu_{i_1} \geq \nu_{i_2} \geq \cdots \geq \nu_{i_{N_{\text{p}}}}.$$

Set $s = 1$ and done = false.

while done == false **AND** $s \leq s_{\max}$ **do**

$s = s + 1$.

Draw $n_s = \lceil \beta n_1 \rceil$ new samples $\boldsymbol{\theta}_k$, $k = N_{\text{total}} + 1, \ldots, N_{\text{total}} + n_s$

$N_{\text{total}} = N_{\text{total}} + n_s$.

Compute $\boldsymbol{g}^k = \nabla_{\boldsymbol{\theta}} G(\boldsymbol{\theta}_k)$, $k = n_{s-1} + 1, \ldots, n_{s-1} + n_s$.

Compute $\{\mu_{s,i}\}_{i=1}^{N_{\text{p}}}$ using the augmented sample $\{\boldsymbol{g}_k\}_{k=1}^{N_{\text{total}}}$.

Compute $\nu_i = \widehat{\mathcal{C}_i \mu_{s,i}}$, $i = 1, \ldots N_{\text{p}}$.

Determine new ranks $\mathcal{R}^{new}$ based on $\{\nu_i\}_{i=1}^{N_{\text{p}}}$.

Compute $\Delta\mu_s = \max\limits_{1 \leq i \leq N_{\text{p}}} \left(\dfrac{|\mu_{s,i} - \mu_{s-1,i}|}{\mu_{s-1,i}} \right)$.

if $\mathcal{R}^{new} = \mathcal{R}^{old}$ **AND** $\Delta\mu_s \leq \tau$ **AND** $s \geq s_{\min}$ **then**

done = true

else

Set $\mathcal{R}^{old} = \mathcal{R}^{new}$

end if

end while

$\mathcal{I}_{\text{active}} = \{i \in \{1, \ldots, N_{\text{p}}\} : \dfrac{\nu_i}{\|\boldsymbol{\nu}\|_\infty} > \tau_{\text{screen}}\}$.

end procedure

Adaptive surrogate model construction. We begin by allocating computational resources for constructing a cross-validation test suite to be used for assessing the accuracy of the resulting surrogate. Naturally, the resources allocated for this purpose depend upon the application as well as total amount of available resources. The set of required inputs for parameter screening are initialized, and model evaluations at n_1 random samples in the full-space are computed. These evaluations are used to construct a surrogate in the full-space

(FSS) using regression-based techniques. If the surrogate is found to be sufficiently accurate for the given application, the process is terminated. However, it is likely that a full-space surrogate constructed using a small number of model evaluations would not provide a faithful representation of the input-output relationship.

The available set of model evaluations are utilized and further enhanced during parameter screening as discussed earlier. At the end of screening, the set of active indices, $\mathcal{I}_{\text{active}}$, is used to evaluate α, referred to as the degree of dimension reduction:

$$\alpha = \frac{|\mathcal{I}_{\text{active}}|}{N_p}, \tag{9}$$

where $|\mathcal{I}_{\text{active}}|$ denotes the cardinality of $\mathcal{I}_{\text{active}}$. Scope for dimension-reduction increases as α decreases. Hence, if α is considered to be small and computational gains are expected owing to dimension reduction, the RSS is constructed and verified for accuracy using a combination of model evaluations used for screening and those associated with the cross-validation test suite. On the other hand, if α is close to 1, the set of inputs required for screening are updated as needed, and a new set of n_1 samples and corresponding model evaluations are generated. The FSS is reconstructed using the enriched set of evaluations and the aforementioned analysis is repeated as illustrated in the flow-diagram in Figure 1 that shows the overall parameter screening and surrogate model construction method.

Assessment of the surrogate. To assess accuracy of the resulting surrogate, one could estimate the leave-one-out cross validation error as follows:

$$\epsilon_{\text{LOO}} = \frac{\sum_{i=1}^{N_l} \left(G(\boldsymbol{\theta}_i) - G^{\text{PC}\setminus\text{i}}(\boldsymbol{\xi}(\boldsymbol{\theta}_i))\right)^2}{\sum_{i=1}^{N_l} \left(G(\boldsymbol{\theta}_i) - \widetilde{\mu}\right)^2}, \tag{10}$$

where N_l is the number of training points, $\widetilde{\mu} = \frac{1}{N_l}\sum_{i=1}^{N_l} G(\boldsymbol{\theta}_i)$ is the sample mean of the model response, and $G^{\text{PC}\setminus\text{i}}$ is the PC surrogate constructed using all but the i^{th} model realization. From (10), it appears that N_l PCEs are needed to evaluate ϵ_{LOO}. However, in practice a modified formulation for ϵ_{LOO} [33], independent of $G^{\text{PC}\setminus\text{i}}$ is used; for an easy reference, see [32, Eq. (1.27)]. Predictive accuracy of the surrogate is assessed by evaluating the relative

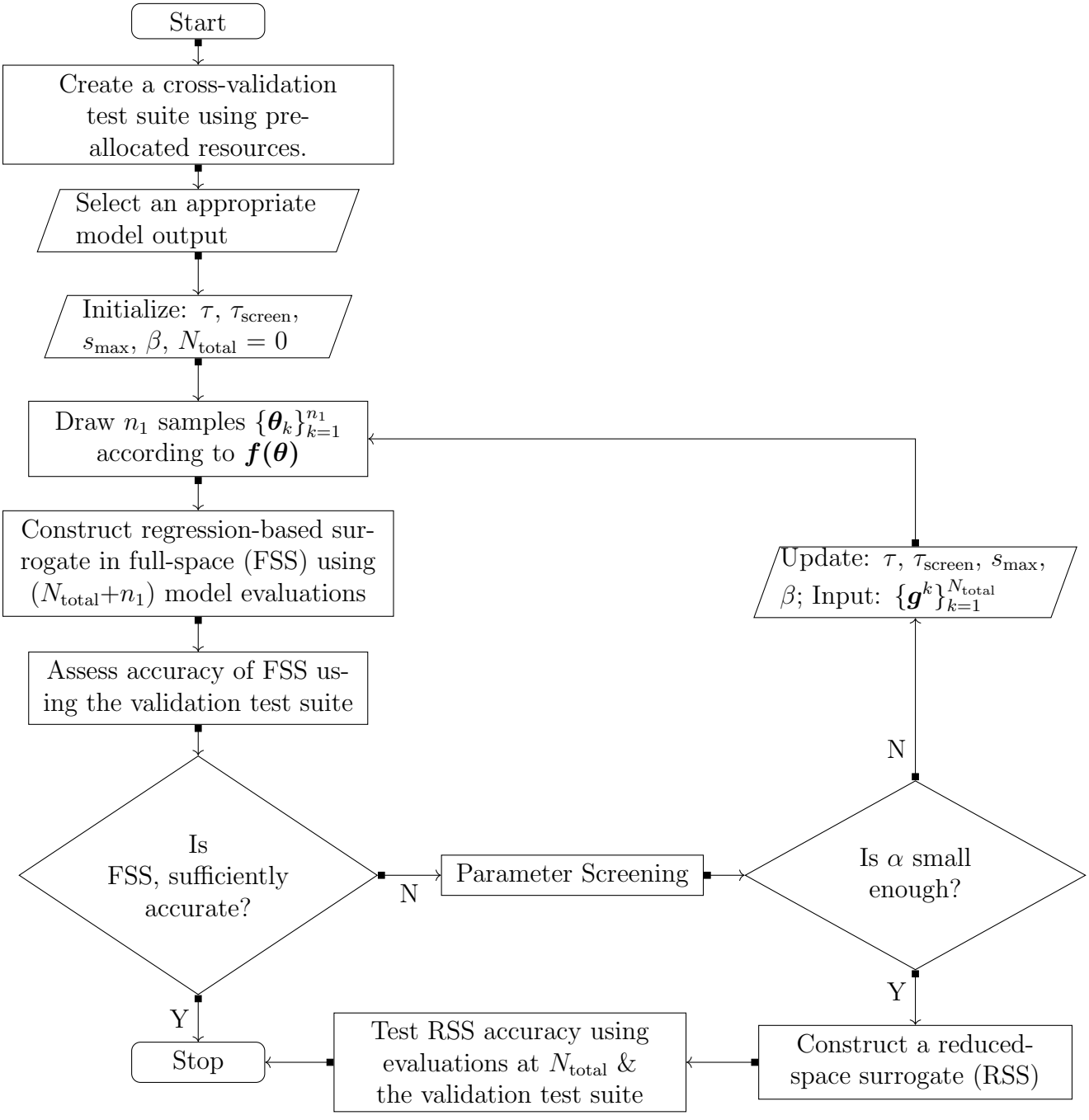

Figure 1: Flow-diagram outlining the adaptive strategy for constructing reduced-space surrogates.

ℓ_2-norm of the difference in predictions between the model and the surrogate ($\epsilon_{\text{L-2}}$), as follows:

$$\epsilon_{\text{L-2}} = \frac{\left[\sum_{i=1}^{N_v} \left(G(\boldsymbol{\theta}_i) - G^{\text{PC}}(\boldsymbol{\xi}(\boldsymbol{\theta}_i))\right)^2\right]^{\frac{1}{2}}}{\left[\sum_{i=1}^{N_v} \left(G(\boldsymbol{\theta}_i)\right)^2\right]^{\frac{1}{2}}}. \tag{11}$$

Here N_v is the number of sampling points in the full parameter space at which model evaluations are available; this, in the case of an RSS, is given by the augmented set of model evaluations used for validation and screening. Accuracy of the surrogate could be further investigated by comparing probability density functions (PDFs) of the model output based on model evaluations in the full parameter space and the RSS predictions corresponding to a large number of samples (say, 10^6 for a high-dimensional input space). However, in realistic problems involving complex, compute-intensive simulations, constructing the PDF based on model evaluations would be infeasible. A practical alternative would be to compare a (normalized) histogram based on sparse model evaluations with the surrogate-based PDF in order to gain some insight into the statistical quality of the surrogate.

Discussion on the proposed methodology. The amount of computational effort associated with the presented methodology can be mainly attributed to two steps: I. Parameter Screening, and II. Constructing a converged RSS. Computational gains are realized in situations where constructing the surrogate in the full parameter space is more expensive than the combined cost associated with these steps. Determining the optimal allocation of computational resources for these steps, however, is not possible a priori. Hence, in the proposed framework, we exploit the set of model evaluations used in parameter screening to simultaneously construct the FSS while keeping a track of its accuracy using the cross-validation test suite. This would help address situations where significant dimension reduction is not possible, and hence, constructing the RSS might result in a computational disadvantage. We suggest using a small number of samples in the initial screening step (say, $n_1 = 5$) and a relatively large τ (say, $\mathcal{O}(10^{-1})$) as a starting point with possible reduction in τ during subsequent screenings. Pseudo-random sampling approaches such as Latin hypercube sampling (LHS) and quasi Monte Carlo (QMC) could be used to generate samples in the input space.

Careful assessment and decision-making is required on whether or not to proceed with

the construction of the RSS at the end of each screening step. The user should account for factors such as the possible degree of dimension reduction, accuracy of the concurrent FSS, and availability of computational resources. Moreover, the stopping criterion for the screening procedure i.e. the choice of $\Delta\mu_s$ should be based on the application.

The applicability of the proposed framework depends upon the choice of the model output. Since the screening metric involves computation of partial derivatives in the full parameter space, the output must exhibit differentiable dependence on each parameter. It is therefore likely that for a given application involving multiple outputs, the RSS can only be constructed for a selected few, using the approach presented above. Hence, it is important to assess the nature of the input-output relationship for a given model prior to implementing the present framework.

Additionally, in some cases, the partial derivative of the output with respect to each uncertain input is not available analytically. In these cases, one could use finite difference (FD) to approximate the gradient as illustrated in 2. However, since FD requires model evaluations at neighboring points, the underlying computational cost is expected to increase by a factor, N_p+1, with N_p being the number of inputs. A possible, more efficient alternative to FD, which might be suitable in some cases, involves the use of adjoints for gradient computation [34]. In the adjoint approach, each gradient evaluation requires a solution of the state equation (forward solve) and that of the corresponding adjoint equation; see e.g., [24–26]. The adjoint method, however, requires the availability of an adjoint solver. Another alternative for efficient gradient computation is the use of automatic differentiation [23].

Using the framework proposed in this section, we aim to construct a reliable surrogate in the most efficient manner within the constraints of the computational budget. However, it might be possible that for a given application, the RSS is not found to be sufficiently accurate. In such a scenario, we suggest enriching the set of important inputs by incorporating the least unimportant model input as determined after a series of screening steps, and re-constructing the RSS. This process could be repeated depending upon the availability of resources and the desired accuracy of the surrogate.

4 Motivating Examples

In section 3, we presented a framework for constructing an RSS (if deemed advantageous) by identifying unimportant parameters based on estimates of the screening metric, $\widehat{\mathcal{C}_i\mu_i}$, for individual parameters. In this section, we motivate the proposed methodology by applying it to two test problems, namely, the borehole function, and a semilinear elliptic PDE. Model evaluations in these test problems are inexpensive. Therefore, we are able to compare the relative importance of model parameters based on the screening metric (computed by sampling the model) with those obtained from converged estimates of $\mathcal{T}(\theta_i)$ (computed using the surrogate constructed in the full parameter space (FSS)). Additionally, to illustrate the computational gains, we compare convergence trends as a function of training runs for the RSS and the FSS using ϵ_{LOO} in Eq. 10. Furthermore, as discussed earlier in section 3, we compare PDFs of the model output, obtained using the RSS, the FSS, as well as true model evaluations, for the purpose of verification.

4.1 Borehole function

The borehole function [32, 35] is a benchmark reference problem in sensitivity analysis. It models the discharge of water ($\mathcal{Q}$) through a borehole in terms of geometrical and physical inputs:

$$\mathcal{Q} = \frac{2\pi T_u (H_u - H_l)}{\ln(r/r_w)\left[1 + \dfrac{2LT_u}{\ln(r/r_w) r_w^2 K_w} + \dfrac{T_u}{T_l}\right]}. \tag{12}$$

The radius of influence, r is fixed at 3698.30 m whereas all other parameters in the right hand side of (12) are considered as uncertain. Hence, $\mathcal{Q} = \mathcal{Q}(\boldsymbol{\theta})$ with

$$\boldsymbol{\theta} = \begin{bmatrix} r_w & L & T_u & H_u & T_l & H_l & K_w \end{bmatrix}^T,$$

being the vector of uncertain parameters. Table 1 provides distributions of the uncertain input parameters.

Table 1: Description and distributions of uncertain inputs in the borehole function given by (12).

Parameter	Distribution
Borehole radius, r_w (m)	$\mathcal{N}(0.1,0.016)$
Borehole length, L (m)	$\mathcal{U}[1120,1680]$
Transmissivity of upper aquifer, T_u (m^2/yr)	$\mathcal{U}[63070,115600]$
Potentiometric head of upper aquifer, H_u (m)	$\mathcal{U}[990,1110]$
Transmissivity of lower aquifer, T_l (m^2/yr)	$\mathcal{U}[63.1,116]$
Potentiometric head of lower aquifer, H_l (m)	$\mathcal{U}[700,820]$
Borehole hydraulic conductivity, K_w (m/yr)	$\mathcal{U}[9855,12045]$

Cheap function evaluations of the discharge $\mathcal{Q}(\boldsymbol{\theta})$ enables construction of the FSS with minimal effort. FSS predictions at a large set of MC samples in the input space are used to obtain converged estimates of $\mathcal{T}(\theta_i)$. Shown in Figure 2(a) are estimates of these indices corresponding to the uncertain parameters in the borehole function using 10^6 MC samples in the input parameter space. These estimates are used to verify fidelity of parameter screening based on the methodology presented in Section 3.

In Figure 2(b), we plot estimates of the screening parameter $\widehat{\mathcal{C}_i\mu_i}$ for a wide range of the number of samples used for approximating μ_i using (1). Additionally, to illustrate a comparison of the estimates of $\widehat{\mathcal{C}_i\mu_i}$ obtained at the end of the initial screening step, i.e. $N = 5$, with subsequent steps, we plot them for individual parameters at $N =$'5, 10, 15, and 20 samples in Figure 2(c). Estimates for $\widehat{\mathcal{C}_i\mu_i}$ are found to be in excellent agreement with $\mathcal{T}(\theta_i)$ even when small number of samples (5–10) are used. Consequently, the relative importance of uncertain parameters in the borehole function is found to be consistent with predictions based on the Sobol' index. In the considered intervals for the uncertain parameters, it is clear that the discharge is insensitive to T_u and T_l. Moreover, the sensitivity to K_w is also small. We exploit these findings to reduce the dimensionality of the problem: we can discount the variabilities in T_u, T_l, and K_w by fixing them at their respective nominal values.

Our goal as discussed earlier is to gain computational advantage by constructing surrogates in a reduced input parameter space. To this end, we use LAR to construct PCEs in 5D and 4D spaces by fixing $\{T_u, T_l\}$ in the former and additionally fixing K_w in the latter at their respective mean values. In Figure 3(a), we compare convergence of PCEs constructed in the full space (7D) with those constructed in the two reduced spaces (4D and 5D) using ϵ_{LOO} (Eq. 10).

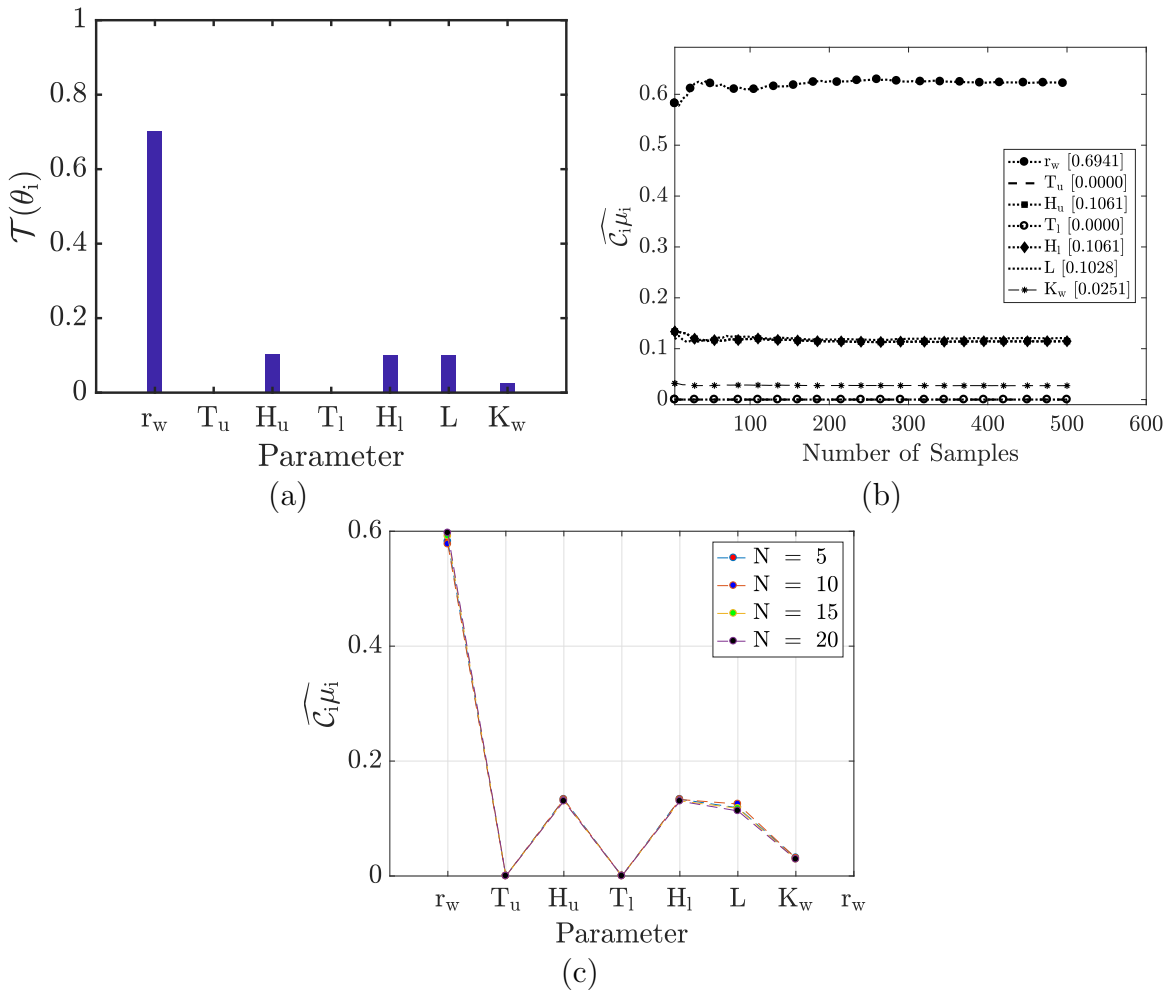

Figure 2: (a) Sobol' total sensitivity index, $\mathcal{T}(\theta_i)$ for uncertain parameters in the borehole discharge function in (12); (b) Estimates of the screening metric ($\widehat{\mathcal{C}_i\mu_i}$), plotted against number of samples. Also included in the legend are estimates of $\mathcal{T}(\theta_i)$ in each case in the legend; and (c) Estimates of $\widehat{\mathcal{C}_i\mu_i}$ for individual parameters using $N = 5$, 10, 15, and 20 samples in the full parameter space.

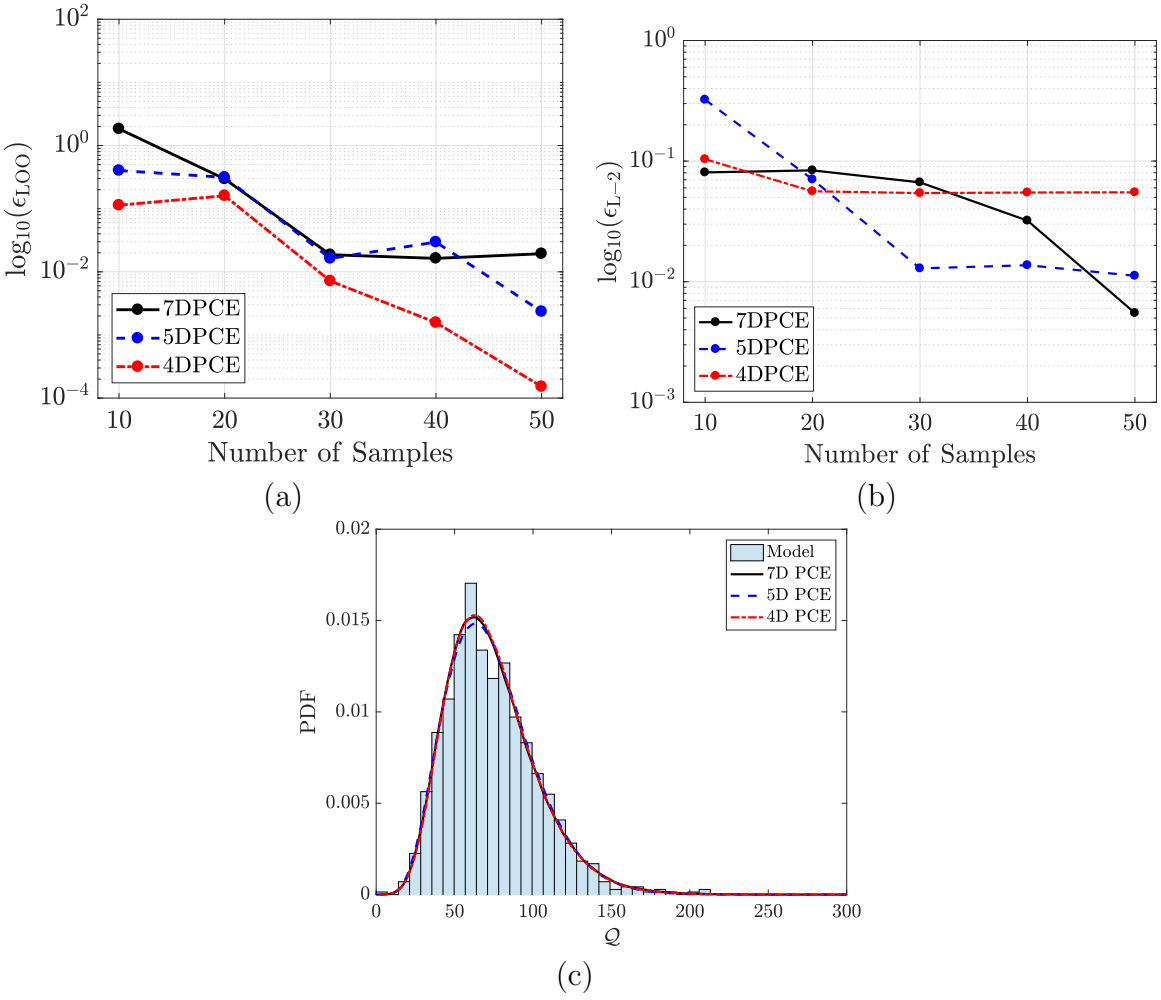

Figure 3: (a) A comparison of order of the leave-one-out-error (ϵ_{LOO}) as a function of number of regression samples used for constructing the PCE in 4, 5, and 7 dimensions. The degree of the PCE constructed using 50 training points was found to be 3, 2,and 2 in the 4D, 5D, and 7D case respectively; (b) Semilog plot of the relative L-2 error norm ($\epsilon_{\mathrm{L\text{-}2}}$) for the PCEs constructed in 4D, 5D, and 7D; and (c) A comparison of PDFs of the discharge, $\mathcal{Q}$, generated using 10^6 samples from the marginal distributions of the uncertain parameters in each case.

As expected, it is observed that the PCE constructed in the 4D space converges at a much faster rate. For instance, if a PCE with $\mathcal{O}(10^{-4})$ accuracy is sought, we need function evaluations at only about 50 sample points in the 4D parameter space whereas the number of samples needed in the full 7D space seems much higher. Latin hypercube sampling (LHS) was used in each case. It must be pointed out that the error in Figure 3(a) is not expected to decrease monotonically with the increase in sample size owing to the penalty term in the regularized optimization problem in Eq. 8. However, the plot of $\epsilon_{\text{L-2}}$ as a function of sample size for the three surrogate models in Figure 3(b) clearly demonstrates the trade-off between computational effort associated with surrogate construction and its predictive accuracy. For a small sample size of 20, the relative error in the case of 4D PCE is found to be the smallest. However, as the sample size increases, the higher order PCE becomes more accurate. This is expected since the higher order PCE inherently contains more information pertaining to the input-output relationship. Note that the same set of 50 samples in the validation test suite were used in each case. Hence, when limited computational resources are available, the proposed methodology could yield substantial gains due to fast convergence as well as higher accuracy of the surrogate in the reduced space. The gains are expected to increase significantly as the dimensionality of the input space increases.

Specifically, the estimates for $\epsilon_{\text{L-2}}$ were found to be 0.0551 and 0.0112 for the 4D and 5D PCE's, respectively. In other words, the 4D PCE is accurate within 5.52% and the 5D PCE is accurate within 1.12% of predictions based on the borehole function. Note that ϵ_{LOO} however, is lower in the case of 4D PCE (Figure 3). Generally, the required level of accuracy is problem dependent. The present framework allows for moving towards higher fidelity reduced-space surrogates based on the ranking of the parameter sensitivities.

Figure 3(c) illustrates a comparison of the PDFs of the discharge, $\mathcal{Q}$ obtained by propagating 10^6 random samples through the 7D PCE in the original input parameter domain as well as the reduced-space PCEs constructed in 4 and 5 dimensions. A normalized histogram plot using 1000 model evaluations in the validation test suite is also included. It is evident from this plot that the PDFs agree quite favorably with each other as well as the original model-based histogram with respect to the modal estimate as well as the uncertainty associated with $\mathcal{Q}$. Consequently, it can be said that the reduced-space PCE is verified in a

probabilistic sense. In other words, the mode as well as the uncertainty in the output is reliably captured and predicted by the reduced-space PCE.

4.2 Semilinear elliptic PDE with random source term

We consider the following semilinear elliptic PDE:

$$\begin{aligned} -\kappa\Delta u + cu^3 &= q \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{13}$$

Here $\Omega = (0,1) \times (0,1)$, u is the state variable, and κ and c are coefficients of the diffusion term and the nonlinear term in the above equation, respectively. We consider uncertainties in κ, c, and the source term. The right hand side function q is defined by

$$q(x,y) = \sum_{i=1}^{N=8} \alpha_i \sin\left(\frac{i\pi x}{8}\right) \cos\left(\frac{i\pi y}{8}\right), \tag{14}$$

where α_i, $i = 1, \ldots, 8$ are random coefficients. Hence, $u = u(\boldsymbol{\theta})$, where

$$\boldsymbol{\theta} = \begin{bmatrix} \kappa & c & \alpha_1 & \alpha_2 & \cdots & \alpha_8 \end{bmatrix}^T$$

is the vector of uncertain parameters. Distributions of the uncertain input parameters are tabulated in Figure 4 (left). The solution of (13) for a fixed set of values of the uncertain parameters is also illustrated.

θ_i	**Distribution**
κ	$\mathcal{U}$[0.05,0.1]
c	$\mathcal{U}$[1.0,2.0]
α_i	$\mathcal{U}$[0.0,4.0]

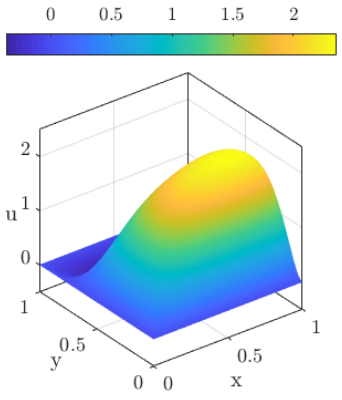

Figure 4: Left: Table providing distributions of the individual uncertain parameters in (13). Right: Solution of the 2D semilinear elliptic PDE (13) using $\kappa = 0.075$, $c = 1.5$, and $\alpha_i = 4.0$

We aim to construct a reduced-space surrogate for the following QoI:

$$\mathcal{F}(\boldsymbol{\theta}) = \frac{1}{|D|} \int_D u(\boldsymbol{x}; \boldsymbol{\theta})\, d\boldsymbol{x}, \tag{15}$$

where D is the region $[2/5, 3/5] \times [2/5, 3/5] \subset \Omega$, and $|D|$ denotes the area of D. While this model is considerably more complex than the borehole example, it can still be solved efficiently. The equation was discretized using finite differences, and Newton's method was used to solve the resulting nonlinear system on a 100×100 2D cartesian grid. We computed converged estimates of the Sobol' total-effect index $\mathcal{T}(\theta_i)$, reported in Figure 5 (left) using FSS predictions at 10^6 MC samples in the input space. The FSS was constructed using model predictions at 500 training points in the 10-dimensional input space. Corresponding value of ϵ_{LOO} was found to be 9.729×10^{-4}. Sensitivity predictions based on the screening metric, $\widehat{\mathcal{C}_i \mu_i}$,

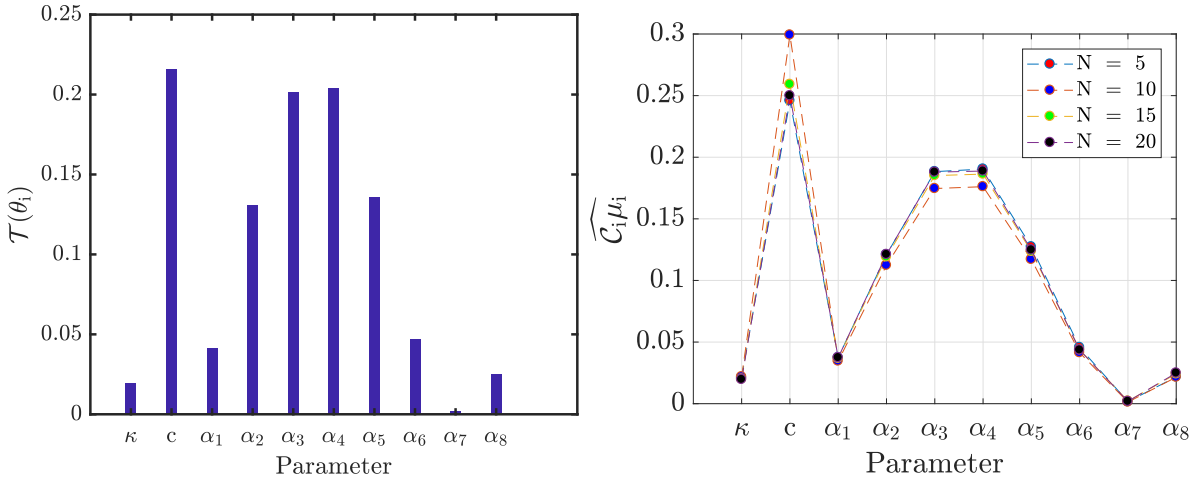

Figure 5: Left: Sobol' total sensitivity index, $\mathcal{T}(\theta_i)$ for uncertain parameters in the semi-linear elliptic PDE (13). Right: Estimates of the screening metric ($\widehat{\mathcal{C}_i \mu_i}$) for each uncertain parameter, obtained using $N = 5$, 10, 15, and 20 samples in the full parameter space.

plotted in Figure 5 (right), are found to be in close agreement with $\mathcal{T}(\theta_i)$, even for the case when $N = 5$. As N is increased from 5 to 20, estimates of the screening metric are observed to converge. Based on the trends observed in Figure 5, it can be said that the uncertainty in the QoI in (15) is largely dependent on c, α_2, α_3, α_4, and α_5. These observations underscore the potential for computational gains by constructing an RSS in the 5D parameter space. We illustrate the comparison of convergence characteristics of the PCEs constructed in the full parameter space (10D) and the reduced space (5D) in Figure 6(a). As expected, the RSS

converges considerably faster. Using model evaluations at 90 sample points, ϵ_{LOO} is found to be two orders of magnitude smaller than that in the case of full-surrogate ($\mathcal{O}(10^{-4})$ versus $\mathcal{O}(10^{-2})$). Consequently, the computational effort for constructing the RSS in the present test problem is expected to be much smaller. The comparison of predictive accuracy of the

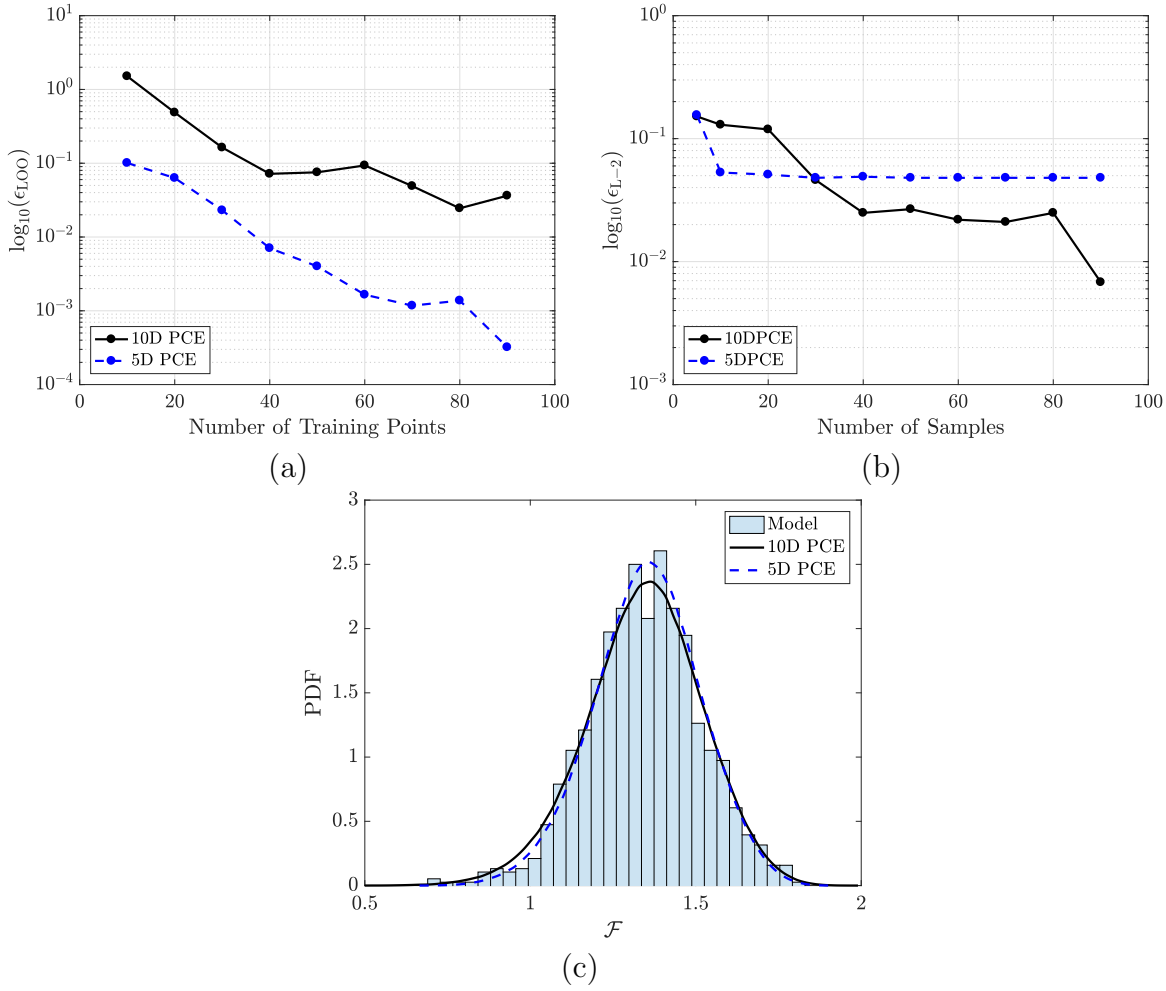

Figure 6: (a) Logarithm of ϵ_{LOO} is plotted against sample size for PCEs constructed in 10 and 5 dimensions to compare their convergence characteristics. The degree of the PCE constructed using 90 training points was found to be 2 and 3 in the 10D and 5D cases respectively; (b) Semilog plot of the relative L-2 error norm ($\epsilon_{\text{L-2}}$) for the PCEs constructed in 5D and 7D

; and (c) PDF of the QoI, $\mathcal{F}(\boldsymbol{\theta})$ in (13) is plotted using the full-surrogate and the reduced-space PC surrogate.

PCEs using $\epsilon_{\text{L-2}}$ is illustrated in Figure 6(b). It is observed that for a small sample size of 20, the surrogate constructed in the reduced space performs better in terms of predictability.

However, as the sample size increases, the FSS is observed to be slightly more accurate although the order of the accuracy is found to be consistent in both cases until the sample size is less than 80. When the sample size increases to 90, the 10D PCE is observed to be more accurate by an order of magnitude. However, the corresponding value of ϵ_{LOO} is $\mathcal{O}(10^{-1})$. Hence, the 10D PCE does not seem to have fully converged and requires additional model evaluations. Therefore, considering fast convergence of the surrogate in the 5D space and predictive accuracy comparable to the 10D PCE, the proposed methodology significantly reduces computational effort in this case.

Using model evaluations at 1000 independent MC samples in the 10D parameter space, the RSS was found to be accurate within 5%. In order to bolster confidence in the RSS, we compare PDFs of the QoI as well as a normalized histogram plot based on sparse model evaluations in the validation test-suite, in Figure 6(c). While the two PDFs are in favorable agreement, the modal estimate and the spread in the QoI based on the histogram is also captured by them. Hence, the RSS could be used with a reasonable degree of confidence to quantify the uncertainty in $\mathcal{F}(\boldsymbol{\theta})$ thereby leading to a computational advantage in this case.

5 Application: H_2/O_2 Reaction Kinetics

The proposed framework in section 3 is implemented to the H_2/O_2 reaction mechanism provided in [36]. The H_2/O_2 reaction is gaining a lot of attention as a potential source of clean energy for applications such as transportation [37] and fuel cell applications [38, 39]. We begin by providing the necessary background information for setting up the problem in 5.1. Results and discussion based on the implementation of the proposed framework are provided in 5.2. Finally, in 5.3, we provide a comparative analysis of the computational cost involved in determining relative parameter importance using DGSMs and the total-effect Sobol' index as the dimensionality increases from 19 to 33 for the present application.

5.1 Problem Setup

The mechanism comprises of 19 reactions including chain reactions, dissociation/recombination reactions, and formation and consumption of intermediate species as provided below in Ta-

ble 2. The reaction rate for the i^{th} reaction as a function of temperature is given as follows:

Reaction #	Reaction		
$\mathcal{R}_1$	$H + O_2$	$\rightleftharpoons$	$O + OH$
$\mathcal{R}_2$	$O + H_2$	$\rightleftharpoons$	$H + OH$
$\mathcal{R}_3$	$H_2 + OH$	$\rightleftharpoons$	$H_2O + H$
$\mathcal{R}_4$	$OH + OH$	$\rightleftharpoons$	$O + H_2O$
$\mathcal{R}_5$	$H_2 + M$	$\rightleftharpoons$	$H + H + M$
$\mathcal{R}_6$	$O + O + M$	$\rightleftharpoons$	$O_2 + M$
$\mathcal{R}_7$	$O + H + M$	$\rightleftharpoons$	$OH + M$
$\mathcal{R}_8$	$H + OH + M$	$\rightleftharpoons$	$H_2O + M$
$\mathcal{R}_9$	$H + O_2 + M$	$\rightleftharpoons$	$HO_2 + M$
$\mathcal{R}_{10}$	$HO_2 + H$	$\rightleftharpoons$	$H_2 + O_2$
$\mathcal{R}_{11}$	$HO_2 + H$	$\rightleftharpoons$	$OH + OH$
$\mathcal{R}_{12}$	$HO_2 + O$	$\rightleftharpoons$	$O_2 + OH$
$\mathcal{R}_{13}$	$HO_2 + OH$	$\rightleftharpoons$	$H_2O + O_2$
$\mathcal{R}_{14}$	$HO_2 + HO_2$	$\rightleftharpoons$	$H_2O_2 + O_2$
$\mathcal{R}_{15}$	$H_2O_2 + M$	$\rightleftharpoons$	$OH + OH + M$
$\mathcal{R}_{16}$	$H_2O_2 + H$	$\rightleftharpoons$	$H_2O + OH$
$\mathcal{R}_{17}$	$H_2O_2 + H$	$\rightleftharpoons$	$HO_2 + H_2$
$\mathcal{R}_{18}$	$H_2O_2 + O$	$\rightleftharpoons$	$OH + HO_2$
$\mathcal{R}_{19}$	$H_2O_2 + OH$	$\rightleftharpoons$	$HO_2 + H_2O$

Table 2: Reaction mechanism for H_2/O_2 from [36]

.

$$k_i(T) = A_i T^{n_i} \exp(-E_{a,i}/RT), \tag{16}$$

where A_i, n_i, and $E_{a,i}$ denote the pre-exponent, the index of T, and and the activation energy corresponding to the i^{th} reaction; and R is the universal gas constant. The TChem [40] software package is used to model homogeneous ignition at constant pressure for a range of initial conditions for the fuel-oxidizer mixture. During the simulation, the fuel-oxidizer mixture goes through a radical build-up phase followed by a sharp increase in temperature as heat is released during the thermal runaway. We focus on quantifying the uncertainty in the *ignition delay* due to uncertainty associated with the pre-exponent, A_i, for each reaction.

The ignition delay is defined as the inflection point on the temperature profile during the thermal runaway. The total number of uncertain parameters in the present case is 19. The A_i's are considered to be uniformly distributed in the interval: $[0.9A_i^*, 1.1A_i^*]$; A_i^* being the nominal estimate corresponding to the i^{th} reaction. The set of nominal values used in the computations, for parameters in (16) are provided in [36].

While the dimensionality of the problem is moderate, constructing a surrogate in the 19-dimensional parameter space could still be expensive. Hence, we explore the possibility of constructing a reduced-space surrogate (RSS) using the framework presented in section 3. In the present study, we focus on two scenarios: fuel(H_2)-rich, and fuel(H_2)-lean. Consider the global reaction:

$$2H_2 + O_2 \rightarrow 2H_2O \tag{17}$$

The equivalence ratio ϕ is defined as follows:

$$\phi = \frac{(M_{H_2}/M_{O_2})_{\text{obs}}}{(M_{H_2}/M_{O_2})_{\text{st}}} \tag{18}$$

The numerator in the right-hand-side represents the observed (obs) fuel-oxygen mass ratio at a given condition and the denominator represents the stoichiometric (st) ratio of the same quantity. Hence, $\phi = 1$ at stoichiometric conditions. The equivalence ratio can be altered by changing the amount of O_2 in the mixture. In the case of a lean mixture, (17) can be written as follows:

$$2H_2 + \alpha O_2 \rightarrow 2H_2O + (\alpha - 1)O_2 \quad (\alpha > 1) \tag{19}$$

Similarly, for the case when the mixture if fuel rich, (17) is modified as follows:

$$2H_2 + \alpha O_2 \rightarrow 2\alpha H_2O + 2(1 - \alpha)H_2 \quad (\alpha < 1) \tag{20}$$

Eqs. (19) and (20) can be generalized as follows:

$$2H_2 + \alpha O_2 \rightarrow 2\min(1, \alpha)H_2O + \max(\alpha - 1, 0)O_2 + \max(0, 2 - 2\alpha)H_2 \tag{21}$$

From the above set of chemical equations, the relationship between ϕ and α can be easily obtained as $\phi = \frac{1}{\alpha}$. Since $\phi > 1$ corresponds to a rich mixture, and $\phi < 1$ corresponds to a lean mixture, we consider $\phi = 2.0$ and 0.5 to investigate the two scenarios respectively.

5.2 DGSM-guided surrogate construction

We apply the parameter screening algorithm with the following parameters: τ_{screen}, $s_{\min}$, $s_{\max}$, β are fixed at 0.2, 3, 10, and 1.0 respectively for both cases. Additionally, the value of τ is considered to be 1.0×10^{-17} and 5.0×10^{-17} in the rich and lean case respectively. Such a small value of τ for this application is a consequence of the nature of convergence exhibited by the sensitivity measures. Moreover, the screening procedure is carried out for atleast $s_{\min}$ number of iterations in order to bolster our confidence in the estimates.

Following the steps outlined in the flow-diagram in Figure 1, model evaluations are initially generated at $n_1 = 5$ samples. The evaluations are used to construct a regression-based surrogate in the full-space. As expected, the surrogate is found to be highly inaccurate. Moreover, unlike the test problems in section 4, we do not estimate the Sobol' total-effect sensitivity indices in the interest of following the overall framework closely. Hence, we proceed to the screening step to estimate the screening metric for the uncertain pre-exponents, A_i's. Results are plotted below in Figure 7 (top row) for both cases. Furthermore, we illustrate the decay in the value of $\Delta\mu_s$ with iterations in Figure 7 (bottom row).

The screening metric estimates in the above plots are observed to converge with only a few samples (5–10). Moreover, out of the 19 uncertain pre-exponents, only A_1, A_9, A_{15}, and A_{17} seem to be important in the fuel-rich case, whereas, only A_1, A_9, and A_{15} seem important in the fuel-lean case, based on the value of τ_{screen}. These observations are indicative of the potential for significant reduction in the dimensionality of this problem. A reduced-space surrogate constructed using the proposed framework could thus lead to large computational gains. The decay of $\Delta\mu_s$ with iterations is expected and builds our confidence in the screening procedure in both cases.

A reduced-space surrogate (RSS) was constructed in 4D for the fuel-rich case, and in 3D for the fuel-lean case. Figure 8 (left) illustrates a comparison of convergence characteristics for the PCEs constructed in the full-space and the reduced-space for the fuel-rich case. Note that the plot is generated using the implementation of least angle regression (LAR) for sparse PCEs in UQLab. The leave-one-out cross validation error is observed to drop initially and plateau with the increase in training points for the 19-dimensional PCE. However, in the case

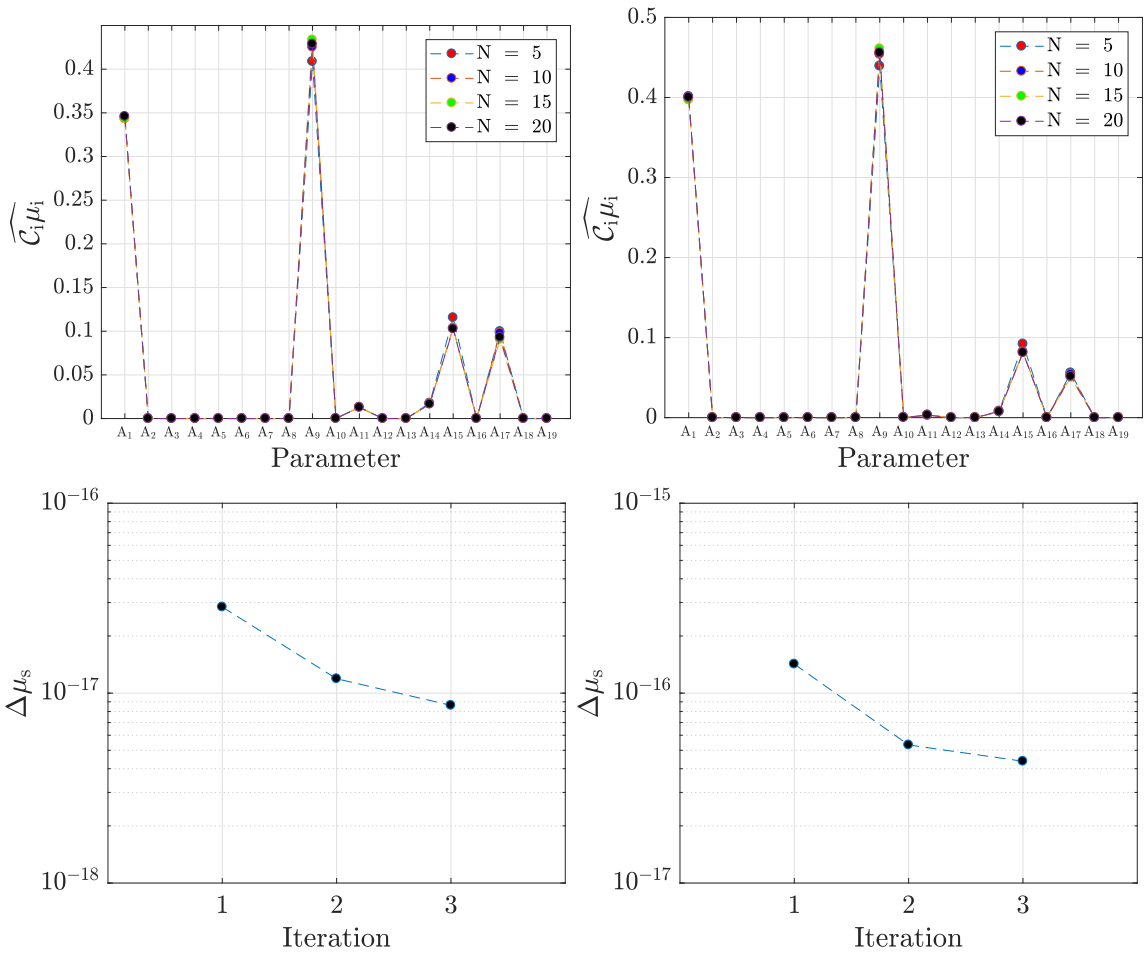

Figure 7: Top: Estimates of $\widehat{\mathcal{C}_i\mu_i}$ for A_i's in the case of fuel-rich mixture (left) and fuel-lean mixture (right). Bottom: The value of $\Delta\mu_s$ during three iterations within the screening step are plotted for the case of fuel-rich mixture (left) and fuel-lean mixture (right).

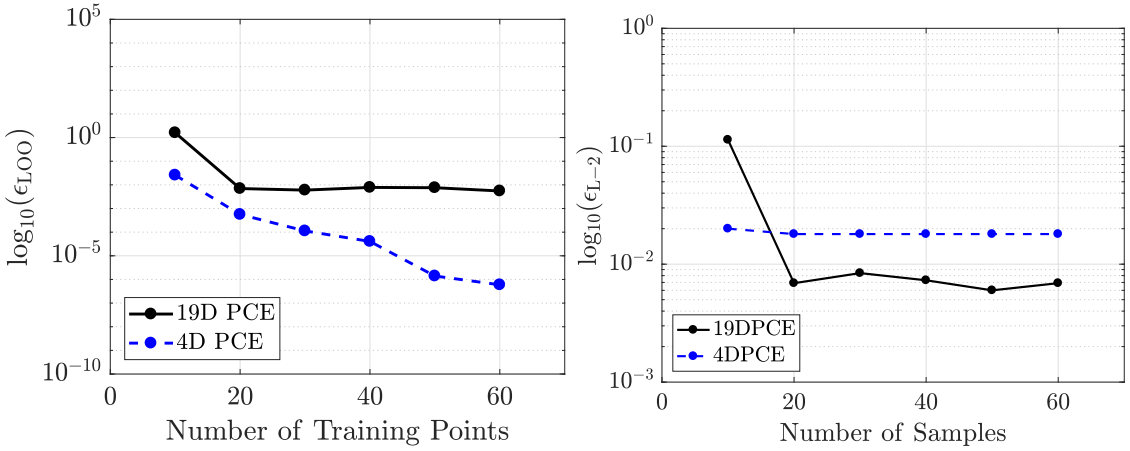

Figure 8: Left: A semi-log plot of ϵ_{LOO} as a function of number of model evaluations in the full-space (19D) and the reduced-space (4D) for the fuel (H_2)-rich case i.e. $\phi = 2.0$. The degree of the PCE constructed using 60 training points was found to be 1 and 3 in the 19D and the 4D cases respectively. Right: Semilog plot of the relative L-2 error norm ($\epsilon_{\text{L-2}}$) for the PCEs constructed in 4D and 19D.

of 4-dimensional PCE, the error exhibits a monotonic behavior and is found to be smaller than $\mathcal{O}(10^{-5})$ at 60 training points. Clearly, the RSS shows a much faster rate of convergence. Additionally, the convergence of the relative error norm, $\epsilon_{\text{L-2}}$ in Figure 8 (right) indicates that for the case of 10 samples, the 4D PCE is observed to be more accurate. However, the 19D PCE is observed to be more accurate by an order of magnitude as the sample size increases. Specifically, the 4D PCE was found to be accurate within 1.8% in the fuel-rich case, and within 3.1% in the fuel-lean case. Therefore, the 4D PCE is still observed to be reasonably accurate in this case. Moreover, its faster convergence as mentioned earlier should help reduce computational effort associated with surrogate model construction. Similar trends (not included) were observed in the fuel-lean case.

Model evaluations at 1000 samples in the test suite are further used to plot a normalized histogram of the ignition time in Figure 9. To verify the accuracy of the RSS in a probabilistic-sense, we compare the histogram plot with a PDF of ignition time using surrogate predictions at 10^6 samples in the reduced space in both cases. Clearly, the RSS captures the spread as well as the modal estimate of the ignition delay in both scenarios. Hence, the proposed framework has enabled significant dimension reduction and construction of an accurate RSS for multiple scenarios pertaining to the H_2/O_2 reaction mechanism.

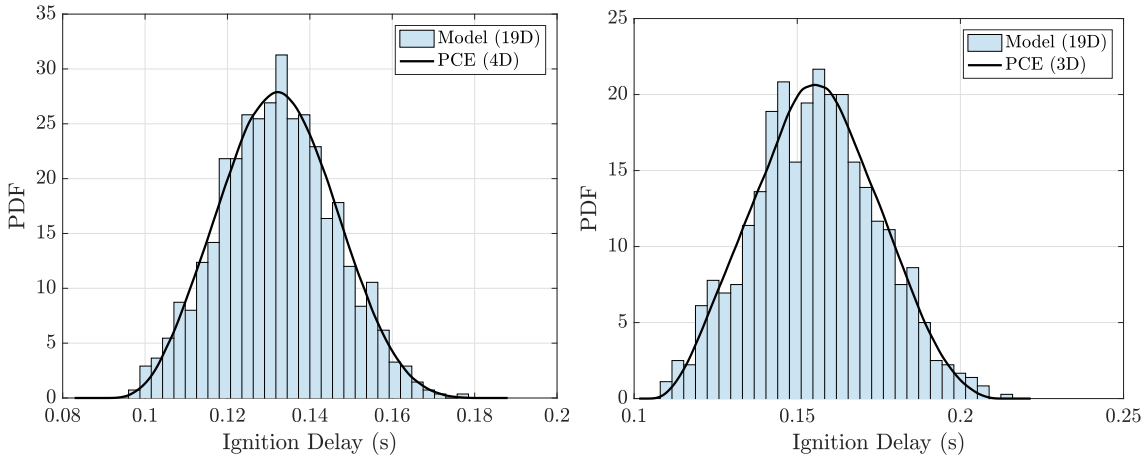

Figure 9: A normalized histogram based on model evaluations at 1000 samples is plotted along with a PDF of ignition delay for the fuel-rich case (left) and the fuel-lean case (right).

5.3 Comparative cost analysis: higher-dimensional setting

In this section, we perform a comparative analysis of the computational cost associated with obtaining converged estimates of parametric sensitivities using DGSM, and the total-effect Sobol' indices ($\mathcal{T}(\theta_i)$). The former is computed using the parameter screening algorithm (Algorithm 1), and the latter is estimated using a sparse-basis PCE. Note that estimating $\mathcal{T}(\theta_i)$ using a PCE has been shown to be more efficient compared to sampling techniques [41, 42]. Moreover, a sparse-basis PCE requires fewer training points which enhances computational savings [33, 43]. To investigate the impact of the dimensionality, the analysis is performed with 19 uncertain parameters, i.e. the pre-exponents (A_i's) as discussed earlier in 5.1 and 5.2, and 33 uncertain parameters wherein the activation energies ($E_{a,i}$'s) are considered to be uncertain in addition to the A_i's. The $E_{a,i}$'s are also considered to be uniformly distributed in the interval: $[0.9E^*_{a,i}, 1.1E^*_{a,i}]$, where $E^*_{a,i}$ is the nominal value for the i^{th} reaction as provided in [36]. Note that only those $E_{a,i}$'s with a non-zero nominal value are considered as uncertain, and therefore the total number of uncertain parameters in the high-dimensional case is 33 as opposed to 38.

The 33-dimensional sparse basis PCE was assessed for its convergence characteristics and its predictive accuracy using ϵ_{LOO} (see (10)) and $\epsilon_{\text{L-2}}$ (see (11)) respectively as shown in Figure 10. The sparse-basis PCEs constructed in the 19 and the 33 dimensional parameter

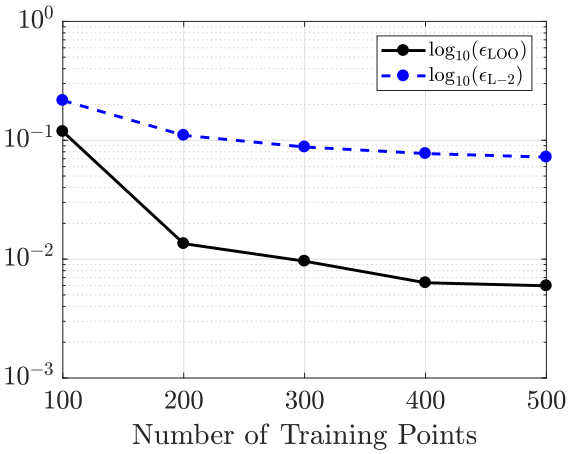

Figure 10: A semi-log plot of ϵ_{LOO} and $\epsilon_{\text{L-2}}$ as a function of the number of training points in the 33D parameter space for the fuel (H_2)-rich case i.e. $\phi = 2.0$. The degree of the PCE constructed using 500 training points was found to be 3.

space were considered to have converged once ϵ_{LOO} was found to be smaller than 6.0×10^{-3}. As observed in Figures 8 and 10, the number of training points required in the 19D and the 33D cases are found to be 20 and 500 respectively. As expected, the predictive accuracy is observed to increase with increase in the number of training points in both cases. The estimate of $\epsilon_{\text{L-2}}$ was found to be 7.22×10^{-2} using the converged 33D PCE and an independent set of model evaluations at 1000 samples.

Before comparing the computational cost pertaining to the two approaches (DGSM-based and Sobol'-based) and the impact of dimensionality, we verify that the parametric sensitivities are consistent in both cases. Sensitivity estimates for the 33 uncertain parameters obtained using the DGSM-based strategy, and by evaluating $\mathcal{T}(\theta_i)$ using the sparse-basis PCE are plotted and placed adjacent to each other for comparison in Figure 11. The plots clearly indicate that the relative importance of the uncertain parameters is consistent in both cases. Specifically, both approaches reveal that the uncertainty in the ignition delay is predominantly due to the uncertainty in $E_{a,1}$ and $E_{a,10}$ with a minor contribution from $E_{a,12}$, whereas, contributions from other parameters is either zero or negligible. These results are clearly indicative of the potential for dimension reduction in this case using the proposed sensitivity-driven approach. A reasonably accurate RSS in a 2D parameter space could potentially capture the uncertainty in the ignition delay due to the uncertainty in the

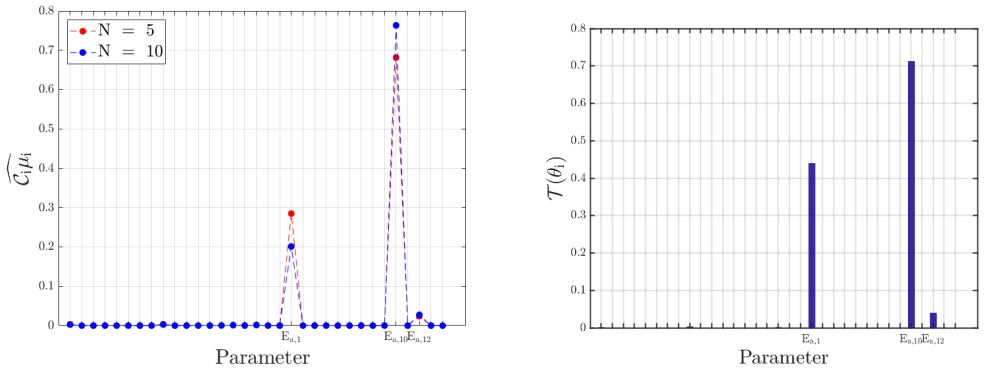

Figure 11: Left: Estimates of the screening metric ($\widehat{\mathcal{C}_i\mu_i}$), obtained using $N = 5$, 10 samples in the full parameter space; and Right: Total-effect Sobol' indices, $\mathcal{T}(\theta_i)$ for the 33 uncertain rate-controlling parameters (A_i's and $E_{a,i}$'s).

33 rate-controlling parameters.

The computational cost is estimated in terms of the number of function evaluations or model runs, denoted by $\mathcal{M}$ in each case. In the case of DGSM-based approach presented in this work, $\mathcal{M} = N(N_p + 1)$ (N: number of samples, N_p: number of parameters) as discussed earlier in 2.1. In the case of PCE-based computation of $\mathcal{T}(\theta_i)$, $\mathcal{M}$ is the sum total of the number of training points used for constructing the sparse-basis PCE and the number of evaluations used for its verification. In Table 3, we provide a comparison of $\mathcal{M}$ for the two approaches in the case of 19 and 33 uncertain parameters for the H_2/O_2 reaction kinetics application. For the 19D case, estimating the converged estimates of the screening metric

Table 3: A comparison of computational cost for the DGSM-based and Sobol'-based parametric sensitivity analysis in the case of 19 and 33 uncertain rate-controlling parameters.

	# of Function Evaluations/Model Runs ($\mathcal{M}$)	
	19D	**33D**
$\widehat{\mathcal{C}_i\mu_i}$ **(DGSM-based)**	$5(19+1) = 100 + N_{v_1}$	$5(33+1) = 170 + N_{v_1}$
$\mathcal{T}(\theta_i)$ **(PCE-based)**	$20 + N_{v_2}$	$500 + N_{v_2}$

($\widehat{\mathcal{C}_i\mu_i}$) required 5 samples. However, since finite difference was used in this work for estimating the gradient of the model output, a total of 100 model runs were required. Additionally,

the screening procedure is continued for one iteration using $N = 10$ samples to ensure the convergence of $\widehat{\mathcal{C}_i\mu_i}$. Since these additional runs are essentially used for verification, we denote them as N_{v_1}. On the other hand, constructing the sparse-basis PCE requires only 20 samples in the 19D parameter space with uncertain pre-exponents. Additional model runs (N_{v_2}) typically ranging from $\mathcal{O}(10^2)$ - $\mathcal{O}(10^3)$ are needed to build a cross validation test suite to verify the accuracy of the PCE. The DGSM-based approach could thus yield computational gains especially when efficient gradient computation techniques are used. The comparison is significantly more favorable for the proposed DGSM-based framework in the higher dimensional case involving 33 uncertain parameters. Once again, converged estimates of the screening metric ($\widehat{\mathcal{C}_i\mu_i}$) are obtained using only 5 samples in the 33D parameter space. For verification, the screening procedure is continued for one iteration by evaluating the model output at 10 samples in 33 dimensions. Therefore, a total of 340 model runs are needed in this case including for gradient computation using finite difference as well as verification. Whereas, as discussed earlier, constructing the sparse PCE itself requires model runs at 500 training points to sufficiently converge. Additional model runs ($\mathcal{O}(10^2)$ - $\mathcal{O}(10^3)$) are required to further verify the accuracy of the resulting PCE. Note that the analysis was performed for the fuel-rich case.

Based on our findings, it appears that the proposed DGSM-based approach can offer a significant computational advantage especially in higher dimensions due to multiple factors: (1) Computational effort required for estimating $\mathcal{T}(\theta_i)$ is observed to increase substantially with dimensionality even when using sparse-basis PCEs. Whereas, the DGSM-based estimates converge with only a few samples, $\mathcal{O}(10^1)$ even for a relatively higher-dimensional case. Moreover, it must be noted that the computational gains associated with the proposed DGSM-based approach could be enhanced significantly by employing efficient gradient computation techniques such as those involving adjoints and automatic differentiation as mentioned earlier in 2.1; (2) The number of model runs needed to verify the convergence of $\widehat{\mathcal{C}_i\mu_i}$ can be controlled using the iterative strategy in this work and is expected to be much smaller than the number of runs needed for verifying the accuracy of $\mathcal{T}(\theta_i)$ especially in high-dimensional settings. Specifically, for the 33D case, model runs at 5 additional samples (corresponding to the first iteration) were used to verify the convergence

of $\widehat{\mathcal{C}_i\mu_i}$ as opposed to 1000 model evaluations used to verify the accuracy of the PCE. It can thus be understood that for higher dimensional problems involving hundreds of parameters, conventional approaches (even sparse-basis PCEs) can quickly become prohibitive and the proposed DGSM-based approach could help reduce the computational effort by several orders of magnitude.

6 Summary and Conclusion

In this work, we have presented an efficient and practical approach for constructing a reduced-space surrogate for scientific and engineering applications. Dimension reduction is accomplished by identifying uncertain parameters that contribute relatively less towards the uncertainty in the QoI. These parameters deemed as *unimportant* are determined using a screening metric (5) involving derivative-based sensitivity measures. Initially, the metric is estimated using model evaluations at a small set of samples in the parameter domain. These estimates are refined by subsequent enrichment of the sample set during the screening procedure presented in Algorithm 1. The outcome of parameter screening is assessed for the scope of dimension reduction. In a favorable scenario, a reduced-space surrogate (RSS) is constructed. The RSS is tested for accuracy in a least-squares sense as well as a probabilistic sense using a cross-validation test suite. In the proposed framework, a surrogate in the full-space (FSS) is constructed in tandem with parameter screening using the available set of model evaluations. Both, RSS and FSS are constructed using regression-based sparse PCEs. Note however that the FSS is constructed in an independent manner to ensure that the computational effort associated with the proposed methodology does not overshoot the effort required to construct the FSS directly. Therefore, it does not impact the accuracy of the RSS which is only constructed in situations where computational gains are expected.

Parameter screening methodology was implemented to low-to-moderate dimensional test problems and an accurate RSS was constructed to demonstrate potential for computational gains in each case. Furthermore, the overall framework was implemented to a relatively higher dimensional application involving kinetics of the H_2/O_2 reaction mechanism. Signif-

icant dimension reduction (19 dimensions to 3 or 4 dimensions) was accomplished for two different scenarios involving a fuel-rich and a fuel-lean mixture. In both cases, the resulting RSS was able to capture the input-output relationship as well as the uncertainty in the QoI with reasonable accuracy. Moreover, it was shown that the parameter screening procedure (Algorithm 1) was able to determine the relative importance of the parameters using only 5 samples in an even higher dimensional case involving 33 uncertain rate-controlling parameters. It was observed that the parameter dimension could be reduced from 33 to 2 in this case. Additional highlights of the proposed framework are as follows:

1. Although PCEs were used in this work, the proposed framework is agnostic to the choice of the surrogate model construction method.

2. Substantial computational gains are expected in situations involving compute-intensive simulations even if the scope for dimension reduction is small. However, computational gains using the proposed framework are expected to increase significantly with the increase in dimensionality of the parameter space.

3. Significant gains can be realized in situations where multiple surrogates need to be constructed as illustrated in the kinetics application. Other possible scenarios may include inverse problems involving parameter estimation in a Bayesian setting.

4. Dimension reduction based on the proposed methodology could help reduce the effort required for model calibration wherein only the important parameters are calibrated.

Based on the results presented for the test problems and the kinetics application, the proposed framework seems quite promising in its potential for identifying the unimportant model inputs. In fact, in the numerical tests and the chemical kinetics application presented in this work, reasonable estimates of the screening metric are obtained during the initial screening step with a few samples. This is indicative of a small degree of variability in the gradient of the model output with respect to individual parameters, and therefore, a low variance of the Monte Carlo estimator. These observations could be exploited to construct efficient model surrogates in a reduced input space.

While the proposed approach has attractive features, it is important to remain cognizant about the limitations of the framework as well. For instance, the QoI is required to be differentiable with respect to each parameter in the considered domain. This condition once satisfied, enhances the accuracy of the PCE-based surrogates as well. Moreover, in the presence of severe non-linearity in the model output leading to significant variability in its gradient in the considered input domain, the number of samples and hence, the computational effort associated with the estimation of screening metrics would expectedly increase. The extent of increase in the effort would of course depend upon the application. Also, we note that the proposed methodology aims to reduce the computational effort pertaining to the total-effect Sobol' index, which is a variance-based measure. For severely nonlinear models, whose QoIs might exhibit heavy-tailed or multimodal distributions, the variance might not provide an adequate representation of uncertainty. Possible remedies for such cases include the so-called moment independent measures; see e.g., [44, 45].

Additionally, the proposed framework does not account for the existence of possible correlations between the uncertain inputs of the model. However, while the assumption of independent inputs is not always justified, in many cases, correlations between inputs are not well understood a priori, and assuming mutual independence could be reasonable at least in initial screening using DGSMs. On the other hand, if approximate correlations are known, we recommend using a Gaussian process or Kriging-based surrogate since it provides a means for incorporating the correlation between inputs. Implementation to applications involving strongly correlated parameters could enhance the applicability of the proposed framework. We consider that to be a potential direction for future studies related to this work.

Acknowledgment

M. Vohra and S. Mahadevan gratefully acknowledge funding support from the National Science Foundation (Grant No. 1404823, CDSE Program). C. Safta was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, as part of the Computational Chemical Sciences Program. The research of A. Alexanderian was partially supported

by the National Science Foundation through the grant DMS-1745654. M. Vohra would also like to sincerely thank Dr. Xun Huan at Sandia National Labs for his guidance pertaining to the usage of TChem for the chemical kinetics application in this work. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department Of Energy or the United States Government.

References

[1] D. Xiu and G.E. Karniadakis. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2):619–644, 2002.

[2] R.G. Ghanem and P.D. Spanos. *Stochastic finite elements: a spectral approach*. Courier Corporation, 2003.

[3] O. Le Maître and O.M. Knio. *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer Science & Business Media, 2010.

[4] J.H. Friedman. Fast MARS. Technical Report 110, Laboratory for Computational Statistics, Department of Statistics, Stanford University, 1993.

[5] C.E. Rasmussen. Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer, 2004.

[6] M.L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.

[7] K. Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural networks*, 2(3):183–192, 1989.

[8] D.F. Specht. Probabilistic neural networks. *Neural networks*, 3(1):109–118, 1990.

[9] I.M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Math. Mod. Comp. Exp.*, 1:407–414, 1993.

[10] A. Alexanderian, J. Winokur, I. Sraj, A. Srinivasan, M. Iskandarani, W.C. Thacker, and O.M. Knio. Global sensitivity analysis in an ocean general circulation model: a sparse spectral projection approach. *Computational Geosciences*, 16(3):757–778, 2012.

[11] G. Li, M. Iskandarani, M. Le Hénaff, J. Winokur, O.P. Le Maître, and O.M. Knio. Quantifying initial and wind forcing uncertainties in the gulf of mexico. *Computational Geosciences*, 20(5):1133–1153, 2016.

[12] A. Namhata, S. Oladyshkin, R.M. Dilmore, L. Zhang, and D.V. Nakles. Probabilistic assessment of above zone pressure predictions at a geologic carbon storage site. *Scientific reports*, 6:39536, 2016.

[13] G. Deman, K. Konakli, B. Sudret, J. Kerrou, P. Perrochet, and H. Benabderrahmane. Using sparse polynomial chaos expansions for the global sensitivity analysis of groundwater lifetime expectancy in a multi-layered hydrogeological model. *Reliability Engineering & System Safety*, 147:156–169, 2016.

[14] B. Saad, A. Alexanderian, S. Prudhomme, and O.M. Knio. Probabilistic modeling and global sensitivity analysis for *co_2* storage in geological formations: a spectral approach. *Applied Mathematical Modelling*, 53:584–601, 2018.

[15] A. Degasperi and S. Gilmore. Sensitivity analysis of stochastic models of bistable biochemical reactions. In *Formal Methods for Computational Systems Biology*, pages 1–20. Springer, 2008.

[16] M Navarro J., O.P. Le Maître, and O.M. Knio. Global sensitivity analysis in stochastic simulators of uncertain reaction networks. *The Journal of Chemical Physics*, 145(24):244106, 2016.

[17] M. Vohra, J. Winokur, K.R. Overdeep, P. Marcello, T.P. Weihs, and O.M. Knio. Development of a reduced model of formation reactions in zr-al nanolaminates. *Journal of Applied Physics*, 116(23):233501, 2014.

[18] I.M. Sobol' and S. Kucherenko. Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79(10):3009–3017, 2009.

[19] I.M. Sobol and S. Kucherenko. Derivative based global sensitivity measures. *Procedia-Social and Behavioral Sciences*, 2(6):7745–7746, 2010.

[20] M. Lamboni, B. Iooss, A.L. Popelin, and F. Gamboa. Derivative-based global sensitivity measures: general links with sobol' indices and numerical tests. *Mathematics and Computers in Simulation*, 87:45–54, 2013.

[21] S. Kucherenko, M. Rodriguez-Fernandez, C. Pantelides, and N. Shah. Monte carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering & System Safety*, 94(7):1135–1148, 2009.

[22] S. Kucherenko and B. Iooss. *Derivative-based global sensitivity measures*. Springer, 2017.

[23] A. Kiparissides, S.S. Kucherenko, A. Mantalaris, and E.N. Pistikopoulos. Global sensitivity analysis challenges in biological systems modeling. *Industrial & Engineering Chemistry Research*, 48(15):7168–7180, 2009.

[24] A. Jameson. Aerodynamic design via control theory. *Journal of scientific computing*, 3(3):233–260, 1988.

[25] M.D. Gunzburger. *Perspectives in flow control and optimization*, volume 5. Siam, 2003.

[26] A. Borzì and V. Schulz. *Computational optimization of systems governed by partial differential equations*, volume 8. SIAM, 2011.

[27] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas. Mean-variance risk-averse optimal control of systems governed by PDEs with random parameter fields using quadratic approximations. *SIAM Journal on Uncertainty Quantification*, 5:1166–1192, 2017.

[28] I.M. Sobol'. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1):271–280, 2001.

[29] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, et al. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

[30] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[31] G. Blatman and B. Sudret. Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. *Comptes Rendus Mécanique*, 336(6):518–523, 2008.

[32] S. Marelli and B. Sudret. Uqlab: A framework for uncertainty quantification in matlab. In *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, pages 2554–2563. 2014.

[33] G. Blatman. *Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis*. PhD thesis, Clermont-Ferrand 2, 2009.

[34] A. Griewank and A. Walther. *Evaluating derivatives: principles and techniques of algorithmic differentiation*, volume 105. Siam, 2008.

[35] M.D. Morris, T.J. Mitchell, and D. Ylvisaker. Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. *Technometrics*, 35(3):243–255, 1993.

[36] R.A. Yetter, F.L. Dryer, and H. Rabitz. A comprehensive reaction mechanism for carbon monoxide/hydrogen/oxygen kinetics. *Combustion Science and Technology*, 79(1-3):97–128, 1991.

[37] L.M. Das. Hydrogen-oxygen reaction mechanism and its implication to hydrogen engine combustion. *International Journal of Hydrogen Energy*, 21(8):703–715, 1996.

[38] B. Loges, A. Boddien, H. Junge, and M. Beller. Controlled generation of hydrogen from formic acid amine adducts at room temperature and application in H_2/O_2 fuel cells. *Angew. Chem. Int. Ed.*, 47:3962–3965, 2008.

[39] S. Cosnier, A.J. Gross, A. Le Goff, and M. Holzinger. Recent advances on enzymatic glucose/oxygen and hydrogen/oxygen biofuel cells: Achievements and limitations. *J. Power Sources*, 325:252–263, 2016.

[40] C. Safta, H.N. Najm, and O.M. Knio. Tchem-a software toolkit for the analysis of complex kinetic models. *Sandia Report, SAND2011-3282*, 2011.

[41] T. Crestaux, O.P. Le Maitre, and J.-M. Martinez. Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94(7):1161 – 1172, 2009. Special Issue on Sensitivity Analysis.

[42] G. Blatman and B. Sudret. Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Reliability Engineering & System Safety*, 95(11):1216–1229, 2010.

[43] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964 – 979, 2008.

[44] E. Borgonovo and B. Iooss. *Moment-independent and reliability-based importance measures*. Springer, 2017.

[45] B. Iooss and P. Lemaître. A review on global sensitivity analysis methods. In *Uncertainty management in simulation-optimization of complex systems*, pages 101–122. Springer, 2015.