HAL
open science

Computational Challenges in Sampling and Representation of Uncertain Reaction Kinetics in Large Dimensions

Saja M Almohammadi, Olivier Le Maitre, Omar M Knio

To cite this version:

Saja M Almohammadi, Olivier Le Maitre, Omar M Knio. Computational Challenges in Sampling and Representation of Uncertain Reaction Kinetics in Large Dimensions. International Journal for Uncertainty Quantification, 2021, 12 (1), pp.1-24. 10.1615/Int.J.UncertaintyQuantification.2021035691 . hal-03282657

HAL Id: hal-03282657

https://hal.science/hal-03282657

Submitted on 18 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Computational Challenges in Sampling and Representation of Uncertain Reaction Kinetics in Large Dimensions

Saja M. Almohammadi[a], Olivier P. Le Maître[b], Omar M. Knio[a,*]

[a] *King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia*
[b] *Centre de Mathématiques Appliquées, CNRS, Inria, École Polytechnique, 91128 Palaiseau, France*

Abstract

This work focuses on constructing functional representations of quantities of interest (QoIs) of an uncertain system in high dimension. Attention is focused on the ignition delay time of an iso-octane air mixture, using a detailed chemical mechanism with 3,811 elementary reactions. Uncertainty in all reaction rates is directly accounted for using associated uncertainty factors, assuming independent log-uniform priors. A Latin hypercube sample (LHS) of the ignition delay times was first generated, and the resulting database was then exploited to assess the possibility of constructing polynomial chaos (PC) representations in terms of the canonical random variables parametrizing the uncertain rates. We explored two avenues, namely sparse regression (SR) using LASSO, and a coordinate transform (CT) approach. Preconditioned variants of both approaches were also considered, namely using the logarithm of the ignition delay time as QoI.

Both approaches resulted in representations of the ignition delay with similar representation errors. However, the CT approach was able to reproduce better the empirical distribution of the underlying LHS ensemble, and also preserved the positivity of the ignition delay time. When preconditioned representations were considered, however, similar performances were obtained using CT and SR representations. The results also revealed that both the CT and SR representations yield consistent global sensitivity estimates. The results were finally used to test a reduced dimension representation, and to outline potential extensions of the work.

Keywords: Polynomial chaos, regularized regression, basis adaptation, preconditioned representation, reduction, global sensitivity analysis

1. Introduction

The analysis of chemical ignition phenomena plays a key role in the design, characterization, and performance assessment of engines and fuels. Different experimental setups have been specifically designed to measure fuel oxidation properties, including shock tubes (e.g. [1, 2, 3, 4]) and rapid compression machines (e.g. [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]). In most of these setups, experimental observables include the pressure trace, as well as the time histories of temperature and key radicals.

A fundamental approach to modeling fuel oxidation properties in the above systems relies on homogeneous oxidation models under adiabatic, constant-volume conditions. To adequately describe the reacting mixture's behavior, comprehensive chemical mechanisms have been developed which, for heavier fuels, may involve hundreds to thousands of reacting species, and thousands to tens of thousands of elementary reactions. One of the fundamental challenges in developing and applying such detailed models concerns quantifying uncertainties in the rate parameters of elementary reactions and the sensitivity of observables on these parameters.

*Corresponding author
Email addresses: saja.almohammadi@kaust.edu.sa (Saja M. Almohammadi), olivier.le-maitre@polytechnique.edu (Olivier P. Le Maître), omar.knio@kaust.edu.sa (Omar M. Knio)

Polynomial chaos (PC) methods [18, 19, 20] have emerged as a popular approach for addressing uncertainties in chemical systems and address problems involving uncertainty propagation and representation, sensitivity analysis, model calibration and experimental design (e.g., [21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32]). Despite these and other successes, the application of PC methods was restricted, for far, to simple problems involving a parametrization of the uncertain rate parameters with stochastic germs having low or moderate dimension. In contrast, the direct application to uncertainties in the rate parameters of detailed hydrocarbon mechanisms, involving tens of uncertain reactions, remains exceptionally challenging without *a priori* reduction of the germ, for instance, exploiting pathway analysis or rate rule correlations (e.g., [29, 33]).

A notable exception concerns the recent work of Malpica *et al.* [34], which addressed the problem of model reduction, under uncertainty, of a detailed n-butane oxidation mechanism with 1,111 uncertain reactions. The study in [34] relied on a direct Monte-Carlo (MC) sampling of the 1111-dimensional uncertainty germ used to parametrize the rates of elementary reactions. Their results, in particular, highlighted the possibility of constructing a family of reduced models based on selected thresholds on the importance of **individual rates** and marginal probabilities.

Inspired by the above developments, the present work also addresses an oxidation problem involving a high-dimensional germ. As further discussed below, we consider a detailed iso-octane mechanism, involving 874 species and 3811 uncertain reactions, and treat the corresponding pre-exponential rate parameters as independent random variables. However, rather than targeting model reduction, we focus on constructing suitable polynomial representations of a quantity of interest (QoI) in terms of the high-dimensional germ parametrizing the uncertain reaction rates and exploiting the resulting representations for global sensitivity analysis [35, 36, 37, 38, 39]. In other words, we investigate means of *a posteriori* reduction of the dimensionality of the random germ. To tackle the forward propagation and representation problems, we explore methods based on a random sampling of the uncertain rates, and that exploit sparsity in the representation of the QoI. Such methods are necessary because spectral projection and sparse grid techniques (e.g., [21, 22, 23, 40, 41, 30, 33]) do not provide the efficiency needed to tackle the curse of dimensionality in the problem at hand. We specifically focus on two approaches respectively falling in the broad classes of regularized regression (e.g., [42, 43, 44, 45, 46, 47, 48, 49, 50, 51]) and subspace (e.g., [52, 53, 54, 55, 56, 57, 58]) methods. We further restrict ourselves to LASSO and CT [52, 53] approaches as representative members of the first and second classes, respectively.

As further discussed in section 2, we perform a single experiment of iso-octane combustion in air and focus on the ignition delay time as our sole QoI. The parametrization of the reaction rates uses independent, canonical random variables, each uniformly distributed over $[-1, 1]$, with associated uncertainty factors. Observations of the QoI are generated from deterministic simulations, using values of the canonical random variables generated using Latin Hypercube Sampling (LHS). The resulting database is then used to assess LASSO and CT approaches, whose constructions are briefly outlined in section 3. We present the results of the analysis in section 4, which includes results obtained for direct representation of the QoI, as well as preconditioned representations [59, 60] based on approximating the logarithm of the ignition delay time. We exploit the simulations to estimate global sensitivity indices and explore a potential reduction in the random input vector's dimensionality. Major conclusions are given in section 5.

2. Setup

2.1. Detailed reaction mechanism

As mentioned in the introduction, we consider propagating uncertainty in the reaction rate parameters of a detailed iso-octane mechanism. Attention is specifically focused on a single experiment, consisting in the adiabatic, constant-volume, oxidation of stoichiometric iso-octane (C_8H_{18}) air mixture, at initial temperature $T_0 = 850$ K and pressure $p_0 = 20$ bar. Also, we focus on a single quantity of interest (QoI), namely the ignition delay time. The evolution of the reaction is described in terms of the detailed mechanism of Mehl *et al.* [61, 62, 63]; it involves $N_s = 874$ species and $N_r = 3811$ elementary reactions. As further discussed below, the rates of the individual elementary reactions are assumed to be uncertain and individually parameterized by independent and identically-distributed (iid) canonical random variables; this immediately leads to a forward propagation and representation problem in high-dimensional parameter space.

2.2. Uncertainty model

As is commonly done in recent UQ studies [21, 22, 64, 65, 33], the uncertainty in elementary reaction rates is modeled in terms of uncertainty factors [66, 67, 68]. Specifically, an elementary reaction rate k is expressed in terms of the modified Arrhenius expression,

$$k(T) = A\left(\frac{T}{T_{\text{ref}}}\right)^n \exp\left(-\frac{E}{RT}\right), \tag{1}$$

where A, E, n and T_{ref} are the pre-exponential factor, activation energy, temperature exponent and reference temperature of the considered reaction, while R is the universal gas constant. Letting A_j^0 denote the nominal pre-exponential factor of reaction j, as given in [61], we express the variability in the corresponding reaction rate by defining uncertain pre-exponential factor, namely according to:

$$A_j(\xi_j) = A_j^0 \mathrm{UF_j}^{\xi_j}. \tag{2}$$

In (2), ξ_j is the canonical random variable associated to the j-th reaction, and $\mathrm{UF_j} \geq 1$ is the corresponding uncertainty factor. The ξ_j are uniformly distributed in $[-1, 1]$, *i.e.*, $\xi_j \sim \mathcal{U}[-1, 1]$, such that $A_j(\xi_j)$ follows a log-uniform distribution with bounds $A_j^{\min} = A_j^0/\mathrm{UF_j}$ and $A_j^{\max} A_j^0 \times \mathrm{UF_j}$. In the following, we denote $\boldsymbol{\xi} = (\xi_1 \cdots \xi_{N_r})$ the collection of canonical random variables. We shall refer to $\boldsymbol{\xi}$ as the uncertainty germ.

Assuming that the uncertainty factors are temperature independent, Eq. (1) shows that the rates $k_j(T)$ are also log-uniformly distributed with uncertainty factors $\mathrm{UF_j}$ and bounds depending on T, the activation energies and exponential exponents. For the present study, the uncertainty factors associated with individual elementary reactions are specified based on a rate rule approach [69, 70, 71, 72, 73, 74], which groups individual elementary reactions in individual reaction classes. In the present implementation of this approach, the uncertainty factor $\mathrm{UF_j}$ of a reaction is set equal to its reaction class's uncertain factor. Table 1 lists the reaction classes and their associated uncertainty factors for the iso-octane mechanism. Note that, as indicated in Table 2, a finer granularity is used for H-abstraction reactions. We emphasize that uncertain elementary reactions remain independent even though they share the same uncertainty factors. In other words, the uncertainty factors of elementary reactions, inherited from their respective reaction classes, are not presently used to imply any *a priori* correlation or dependence between the class member's reaction rates.

2.3. Sampled QoI

In order to assess the performance of various forward UQ approaches, we first sampled an ensemble of realizations of the uncertain rates, using a uniform Latin Hyper-Cube Sampling (LHS), in $[-1, 1]^{N_r}$, of the germ. Specifically, the LHS sample comprised $M = 38,110$ members. For each LHS realization of the germ, denoted $\boldsymbol{\xi}^{(i)}$, the ignition process was simulated using the TChem software [75] using the corresponding pre-exponential factors with $\xi_j = \xi_j^{(i)}$ in (2). The simulation results, which include the time history of the temperature, pressure, and species concentration, were then used to estimate the ignition delay time, τ_{ign}, defined as the time coordinate at which the temperature of the mixture first reaches $T_0 + 400$ K. In other words, $T(\tau_{\text{ign}}) = T_0 + 400$ K. The analyses described below thus exploit a database comprising the M sampled (N_r-dimensional) coordinates of the germ and the corresponding QoI values. The data set may be freely accessed from [76].

Figure 1 shows the distribution of τ_{ign} estimated from the LHS sample. The pdf resembles a log-normal distribution with a long tail toward the larger values. The dashed line in the figure represents the empirical mean, $\mu = 0.0203$ s. The empirical variance of the distribution is $\sigma^2 = 9.17393 \times 10^{-5}$ s^2. The coefficient of variation is large, $COV \equiv \sigma/\mu \simeq 47\%$. Note that the nominal value of the ignition delay, $\tau_0 = 0.0227$ s, is close to but appreciably larger than the mean.

Remark 1. *As can be appreciated from [76], the size of the data set is quite large, requiring approximately 1Gb of storage. In the present context, recording the coordinates of the sampled germ, i.e. the realization of the canonical random variables $\xi_j^{(i)}$, consumes most of the storage. The storage requirement for the*

Table 1: Reaction classes and their uncertainty factors (UF) for iso-octane mechanism.

25 Reaction Classes in iso-octane mechanism	UF
1. Unimolecular fuel decomposition	2
2. H atom abstraction from the fuel	Table 2
3. Alkyl radical decomposition	2
4. Alkyl radical + O2 to produce olefin + HO2 directly	4
5. Alkyl radical isomerization	3.5
6. Abstraction reactions from Olefin by OH, H, O, CH3	3
7. Addition of radical species O and OH to olefin	3
8. Alkenyl radical decomposition	2
9. Olefin decomposition	5
10. Addition of alkyl radicals to O2 [R + O2]	2
11. R + RO2 = RO + RO	2
12. Alkyl peroxy radical isomerization [ROO = QOOH]	4
13. RO2 + HO2 = RO2H +O2	3
14. RO2 + H2O2 = RO2H + HO2	3
15. RO2 + CH3O2 = RO + CH3O +O2	3
16. RO2 + RO2 = RO + RO + O2	3
17. RO2H = RO + OH	3
18. RO decomposition	3
19. QOOH = cyclic ether + OH (cyclic ether formation via cyclisation of diradical)	4
20. QOOH = olefin + HO2 (radical site β to OOH group)	3
21. QOOH = olefin + carbonyl radical (radical site γ to OOH group)	3
22. Addition of QOOH to O2	2
23. Isomerization of O2QOOH and formation of carbonylhydroperoxide and OH	4
24. Decomposition of carbonylhydroperoxide to form oxygenated radical species and OH	3
25. Cyclic ether reactions with OH and HO2	3

Table 2: Uncertainty factors (UF) for H-abstraction reactions.

H atom abstraction from the fuel by	UF
1. H, O, OH	2
2. O2, HO2	3
3. CH3, CH3O, CH3O2, O2CHO, C2H5, C2H3, C4H6OH, C8H17O2	2

QoIs $\tau_{\text{ign}}^{(i)}$ is, in contrast, relatively insignificant, in fact, N_r times less. This observation highlights the coordinate storage challenge in sampling and representing high-dimensional systems. The present study does not consider this challenge, and focuses instead on propagation and representation challenges.

3. Methodology

In this section, we outline our approach to the construction of PC representations [19, 20, 77, 78, 79, 30, 80] of the ignition delay time, $\tau_{\text{ign}}(\boldsymbol{\xi})$, using the database described in section 2. We briefly outline these PC expansions in section 3.1 below. In sections 3.2 and 3.3, we respectively describe our implementation of sparse regression (SR) and Coordinate Transformation (CT) approaches to the construction of the representations. We then outline a preconditioning of the quantity of interest, consisting in the representation of the logarithm of the ignition delay time, in section 3.4.

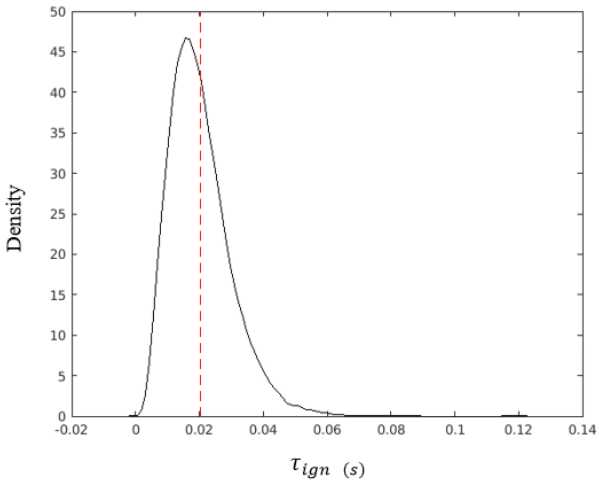

Figure 1: PDF of the ignition delay time (s). The distribution is estimated from the LHS data using a KDE method. The dashed line shows the mean value, estimated as the sample's empirical average.

3.1. PC representation

We will primarily focus on establishing suitable PC expansions (PCEs) [19, 20, 77, 80] of the ignition delay time, τ_{ign}, in terms of the canonical random variables parameterizing the uncertain reaction rates. Specifically, we seek generic representations of the form:

$$\tau_{\text{ign}}(\boldsymbol{\xi}) \approx \sum_{\alpha \in \mathbb{N}^d} c_\alpha \Psi_\alpha(\boldsymbol{\xi}). \tag{3}$$

In the previous expansion, $d = N_r$ is the dimension of the PC expansion, the Ψ_α are multi-dimensional Legendre polynomials in $\boldsymbol{\xi}$, and the c_α are the PC coefficients that need to be determined using the database of realizations outlined in the previous section. The multi-dimensional polynomials Ψ_α are constructed by taking the product of univariate orthogonal polynomials with degree defined by the d-dimensional multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$. Specifically, we have [20]:

$$\Psi_\alpha(\boldsymbol{\xi}) = \prod_{j=1}^{d} \psi_{\alpha_j}(\xi_j) \tag{4}$$

where ψ_α denotes the univariate Legendre polynomial of order α.

In practice, the PC expansion (3) is truncated by retaining the polynomials whose total order is p or less. The summation is then restricted to $\alpha \in \mathcal{A}_p$, where the set of multi-indices $\mathcal{A}_p$ is defined by

$$\mathcal{A}_p = \left\{ \alpha \in \mathbb{N}^d, \sum_{j=1}^{d} \alpha_j \leq p \right\}. \tag{5}$$

For a d-dimensional expansion, the dimension of the truncated basis, P, is given by the cardinality of $\mathcal{A}_p$, which is

$$P = |\mathcal{A}_p| = \frac{(p+d)!}{p!d!}. \tag{6}$$

Remark 2. *It is clear from* (6) *that for the present high-dimensional setting with* $d = 3,811$*, the size of the basis is quite large unless a first-order* ($p = 1$) *truncation is adopted . For instance, with* $p = 2$*, the size of the truncated basis will be* $P = 7,267,578$*. The dimension of the basis underscores one of the key*

challenges that we faced in the present study, as the number of realizations in our database ($M = 38,110$) will generally be significantly smaller than the size of the truncated basis. As mentioned in the introduction, we will attempt to address this challenge by relying on SR and CT techniques.

Remark 3. *Another challenge arising from the high-dimensionality of the problem concerns the multi-index storage requirements. Specifically, if full storage of the multi-indices is adopted, one would need to reserve sufficient memory to store a total of P integer-valued d-dimensional vectors. To illustrate this requirement, we note that for a second-order basis, storage of all the multi-indices in $\mathcal{A}_p$ would translate to over 100Gb RAM, which is a prohibitive even on a large computer platform.*

To overcome this memory challenge, we developed a sparse multi-index storage format, exploiting the fact that the multi-indices $\alpha \in \mathcal{A}_p$ have at most $p \ll d$ non-zero entries. The sparse storage consists in utilizing two descriptors for each multi-index $\alpha \in \mathcal{A}_p$: the number $t_0(\alpha) \leq p$ of non-zero degree ($\alpha_j \neq 0$) univariate polynomials the definition of Ψ_α, see (4), and the list of $t_0(\alpha)$ couples (j, α_j) of the dimensions j and associated degrees α_j which are non-zero. This format effectively reduces the multi-index memory requirement virtually from $\mathrm{O}(P \times d)$ to $\mathrm{O}(P \times p)$.

3.2. Sparse Polynomial Regression

Given the realizations database, *i.e.*, the pairs $(\boldsymbol{\xi}^{(i)}, \tau_{\text{ign}}^{(i)})$ for $i = 1, \ldots, M$, and a PC basis $\{\Psi_\alpha, \alpha \in \mathcal{A}_p\}$, one has to determine the coefficients c_α of the expansion. Different methodologies have been proposed for this purpose. We focus on regression-type methods, which aim to minimize the distance between the Constructed PC approximation and the realizations of the database. For instance, the ordinary least squares (OLS) regression problem consider the minimization of

$$\mathrm{LS}\left(\{c_\alpha, \alpha \in \mathcal{A}_p\}\right) = \sum_{i=1}^{M} \left| \tau_{\text{ign}}^{(i)} - \sum_{\alpha \in Aset_p} c_\alpha \Psi_\alpha(\boldsymbol{\xi}^{(i)}) \right|^2 . \tag{7}$$

Introducing $\boldsymbol{c}$ the vector of coefficients c_α, $\boldsymbol{\tau}$ the vector of ignition time realizations $\tau_{\text{ign}}^{(i)}$ and $[\boldsymbol{\Psi}] \in \mathbb{R}^{M \times P}$ the regression matrix with entries $[\boldsymbol{\Psi}]_{i,\alpha} \doteq \Psi_\alpha(\boldsymbol{\xi}^{(i)})$, the OLS problem can be recast in

$$\widehat{\boldsymbol{c}} = \underset{\boldsymbol{c} \in \mathbb{R}^P}{\arg\min} \left\| \boldsymbol{\tau} - [\boldsymbol{\Psi}]\boldsymbol{c} \right\|_2^2 . \tag{8}$$

Although simple in principle, the OLS regression presents in practice severe challenges related to the well-posedness of (8) and computational aspects. Starting with the well-posedness, we observe that the solution of (8) satisfies $[\boldsymbol{\Psi}]^T[\boldsymbol{\Psi}]\boldsymbol{c} = [\boldsymbol{\Psi}]^T\boldsymbol{\tau}$, a system that has a unique solution only of $[\boldsymbol{\Psi}]^T[\boldsymbol{\Psi}]$ is full rank. A necessary condition to ensure that the problem is not under-determined is then $M > P$. In our problem, for $p \geq 2$, we have $M \ll P$, such that only a first-order model ($p = 1$) seems accessible to OLS using the total order truncation strategy. Concerning the computational aspect, the assembly and the storage of the full regression matrix $[\boldsymbol{\Psi}]$ is clearly challenging and may be impractical for $p > 1$.

To overcome these issues, one can consider the regularization of the OLS problem to the adaptive construction of $\mathcal{A}$, particularly to reduce the sizes of the regression basis and regression problem.

3.2.1. Regularized regression

In the present work, we rely on the least absolute shrinkage and selection operator (LASSO), which mostly consists of an approximate regression inherently comprising regularization and promoting sparsity. With our notations, the LASSO considers the following minimization problem:

$$\widehat{\boldsymbol{c}} = \underset{\boldsymbol{c} \in \mathbb{R}^P}{\arg\min} \left\| \boldsymbol{\tau} - [\boldsymbol{\Psi}]\boldsymbol{c} \right\|_2^2 + \lambda \left\| \boldsymbol{c} \right\|_1 , \tag{9}$$

where $\lambda > 0$ is a regularization parameter. The solution $\widehat{\boldsymbol{c}}$ of problem (9) becomes sparser as λ increases, because of the L_1-penalty on the coefficients norm. The value of λ is determined as part of a validation procedure [48, 81, 47, 51, 42]. In our numerical implementations, we rely on the MATLAB function `lasso` with a 10-fold cross-validation procedure to determine λ minimizing the cross-validation error.

3.2.2. Basis selection by two-stage sparse regression

As mentioned previously, a critical issue with the polynomial regression in high-dimension is the computational complexity associated with the use of a large basis with dimension P growing exponentially fast with p. Adaptive strategies have been proposed to mitigate this limitation by sequentially enriching the basis [51]. Classically, the set of multi-indices $\mathcal{A}^k$ of the k-th stage of the adaptation is enriched to $\mathcal{A}^{k+1} \supset \mathcal{A}^k$, by considering new multi-indices which have predecessors $\alpha \in \mathcal{A}^k$ with coefficients $\widehat{c}_\alpha \neq 0$.

Two-stage LASSO procedures are employed in the present work to select the regression basis and construct the approximation. In this first stage, the sparse regression stage, we solve the LASSO problem (9) using the first order basis defined by $\mathcal{A}^{1\text{st}} \doteq \mathcal{A}_{p=1}$ (see (5)); therefore, at this stage, $P_1 = N_r + 1 = 3,812 \ll M$. Let $\tau_{\text{ign}}^{1\text{st}}(\boldsymbol{\xi})$ be the solution of the first stage problem, with coefficients vector $\boldsymbol{c}^{1\text{st}}$. We denote $\widehat{\mathcal{A}}^{1\text{st}} \subset \mathcal{A}^{1\text{st}}$ the subset of active multi-indices associated to non-zero coefficients $c_\alpha^{1\text{st}}$. From $\widehat{\mathcal{A}}^{1\text{st}}$ we derive the set of active dimension indices $\mathcal{J}^{1\text{st}} \subset \{1, \ldots, d\}$. Indeed, for each $\alpha \in \mathcal{A}_1 \setminus (0, \ldots, 0)$ one can associate a *unique* direction index $j \in \{1, \ldots, d\}$ such that $\alpha_j = 1$. In our test problem (see details in the next section), the cross-validation of $\tau_{\text{ign}}^{1\text{st}}(\boldsymbol{\xi})$ yielded a subset of 328 active dimensions: $|\mathcal{J}^{1\text{st}}| = 328 < d \ll 3,812$.

In the second stage, we extend the active basis $\widehat{\mathcal{A}}^{1\text{st}}$ of the first stage up to prescribed order $p > 1$, defining the second set of multi-indices $\mathcal{A}^{2\text{nd}}$ as

$$\mathcal{A}_p^{2\text{nd}} = \widehat{\mathcal{A}}^{1\text{st}} \bigcup \left\{ \alpha \in \mathbb{N}^d, |\alpha|_1 \leq p, \alpha_i = 0 \ \forall i \notin \mathcal{J}^{1\text{st}} \right\}. \tag{10}$$

In other words, directions not selected in the first stage are disregarded in the second stage.

We explore two approaches for the second stage. In the first approach, we seek a correction

$$\widehat{\delta\tau}_{\text{ign}}^{2\text{nd}}(\boldsymbol{\xi}) = \sum_{\alpha \in \mathcal{A}_p^{2\text{nd}}} \widehat{\delta c}_\alpha^{2\text{nd}} \Psi_\alpha(\boldsymbol{\xi}),$$

of the first stage model $\tau_{\text{ign}}^{1\text{st}}(\boldsymbol{\xi})$. The correction is determined solving the second stage LASSO problem

$$\widehat{\boldsymbol{\delta c}}^{2\text{nd}} = \underset{\boldsymbol{\delta c} \in \mathbb{R}^{\left|\mathcal{A}_p^{2\text{nd}}\right|}}{\arg\min} \left\| \boldsymbol{\delta\tau}^{2\text{nd}} - [\boldsymbol{\Psi}]\boldsymbol{\delta c} \right\|_2^2 + \lambda \left\| \boldsymbol{\delta c} \right\|_1, \tag{11}$$

where the entries of residual vector $\boldsymbol{\delta\tau}^{2\text{nd}}$ are defined by the differences

$$\boldsymbol{\delta\tau}_i^{2\text{nd}} = \tau_{\text{ign}}^{(i)} - \tau^{1\text{st}}(\boldsymbol{\xi}^{(i)}), \quad i = 1, \ldots, M.$$

The resulting surrogate is finally expressed as

$$\tau_{\text{ign}}(\boldsymbol{\xi}) \approx \sum_{\alpha \in \mathcal{A}^{1\text{st}}} c_\alpha^{1\text{st}} \Psi_\alpha(\boldsymbol{\xi}) + \sum_{\alpha \in \mathcal{A}_p^{2\text{nd}}} \widehat{\delta c}_\alpha^{2\text{nd}} \Psi_\alpha(\boldsymbol{\xi}). \tag{12}$$

For the second approach, we directly seek an approximation of $\tau_{\text{ign}}(\boldsymbol{\xi})$ in the basis defined by $\mathcal{A}_p^{2\text{nd}}$. This second approach yields to the second stage LASSO problem,

$$\widehat{\boldsymbol{c}}^{2\text{nd}} = \underset{\boldsymbol{c} \in \mathbb{R}^{\left|\mathcal{A}_p^{2\text{nd}}\right|}}{\arg\min} \left\| \boldsymbol{\tau} - [\boldsymbol{\Psi}]\boldsymbol{c} \right\|_2^2 + \lambda \left\| \boldsymbol{c} \right\|_1, \tag{13}$$

with resulting surrogate

$$\tau_{\text{ign}}(\boldsymbol{\xi}) \approx \sum_{\alpha \in \mathcal{A}^{2\text{nd}}} \widehat{c}_\alpha^{2\text{nd}} \Psi_\alpha(\boldsymbol{\xi}). \tag{14}$$

In the results section, application of (11,12) is referred to as the SR-C (Sparse Regression-Correction) method, whereas application of (13,14) is referred to as the SR-D (Sparse Regression-Direct)) method.

3.3. Coordinate transformation

The two-stage SR methods introduced above rely on an achievable approximation (first-order expansion) to select relevant directions along which a higher-order sparse expansion is sought during the second stage. The next method we investigate belongs to the class of active subspace methods designed to address the curse of dimensionality. The key idea of these methods is the derivation of a new set of coordinates, with much lower dimensionality, and constructing the approximation in these new coordinates. Classically, the new coordinates $\boldsymbol{\eta}$ are defined by linear combinations of original ones through $\boldsymbol{\eta} = [\mathbf{W}]^T \boldsymbol{\xi}$, with transformation matrix $[\mathbf{W}]$ such that $[\mathbf{W}]^T[\mathbf{W}] = I$, so the columns $\boldsymbol{W}_i$ of $[\mathbf{W}]$ define orthogonal directions. Therefore, rather than selecting a subset of important *dimensions* (indices) in the germ $\boldsymbol{\xi}$, these approaches select few important *directions* (combinations of dimensions) in the germ. The choice of $[\mathbf{W}]$ is critical to come up with a good approximation of the QoI in the transformed coordinates $\boldsymbol{\eta}$ and achieving a significant dimensionality reduction. A classical approach consists of using the dominant eigenvectors of the correlation matrix of the gradient of the QoI for the column $[\mathbf{W}]$. Unfortunately, we do not have access to the gradient of the ignition times, $\nabla_{\boldsymbol{\xi}} \tau_{\text{ign}}(\boldsymbol{\xi})$, and we need to rely on an alternative to estimate the principal directions of variability.

Our CT method is then adapted from the technique developed for Gaussian germs in [52, 53]. It is a sequential approach that yields a new direction at each step. The approach uses an initial approximation of the first direction $\boldsymbol{W}_1$ constructed on the first-order approximations,

$$\tau_{\text{ign}}(\boldsymbol{\xi}) \approx C_0 + \sum_{j=1}^{d} C_j \xi_j = C_0 + \boldsymbol{C}^T \boldsymbol{\xi}, \tag{15}$$

corresponding to the first stage of the sparse regression in $\mathcal{A}_{p=1}$. The gradient of the model in (15),

$$\nabla_{\boldsymbol{\xi}} \tau_{\text{ign}}(\boldsymbol{\xi}) \approx \boldsymbol{C} = (C_1 \cdots C_d)^T, \tag{16}$$

is independent of $\boldsymbol{\xi}$ and its correlation matrix has rank one. Setting the first direction to $\boldsymbol{W}_1 = \boldsymbol{C}$ the first CT model is written as the degree-p polynomial expansion, for $p > 1$,

$$\tau_{\text{ign}}(\boldsymbol{\xi}) \approx \tau_1^{\text{CT}}\ (\boldsymbol{\xi}) \equiv \sum_{l=0}^{p} \widetilde{c}_l \psi_l(\eta_1), \quad \eta_1 = \boldsymbol{W}_1^T \boldsymbol{\xi}. \tag{17}$$

The coefficients $\widetilde{\boldsymbol{c}}$ of this univariate expansion are computed using, for instance, on the LASSO algorithm.

Remark 4. *We observe that using Legendre polynomials for the expansion in* (17) *is arbitrary; any basis of polynomials with degree less or equal to p would be suitable. Indeed,* $\eta_1 = \boldsymbol{W}_1^T \boldsymbol{\xi}$ *is not uniformly distributed in* $[-1, 1]$*, in general, so the expansion of* $\tau_1^{\text{CT}}\ (\boldsymbol{\xi})$ *is not orthogonal in the Legendre basis. The consequence of losing this orthogonality is further discussed later in section Appendix A.2, when detailing the computation of the sensitivity indices in the CT method.*

Once the first CT approximation has been determined, a second direction $\boldsymbol{W}_2$ can be constructed from a first order approximation of the residual $\tau_{\text{ign}}(\boldsymbol{\xi}) - \tau_1^{\text{CT}}\ (\boldsymbol{\xi})$ and enforcing orthogonality between $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$. Then, a two dimensional polynomial regression in (η_1, η_2) can be applied to obtain the degree $p > 1$ model $\tau_2^{\text{CT}}\ (\boldsymbol{\xi})$. This process can be iterated, as long as the first-order approximation of the residual yields a direction distinct from the previous ones, leading to higher-dimensional representations.

The greedy CT method just sketched involves two essential stages: the determination of the direction by first-order residual fit, and the polynomial regression in the resulting transformed coordinates. For the latter, the LASSO method discussed in section 3.2.1 is suitable as the problem dimension of the regression space is small. For the first stage, one can also rely on a LASSO regression with the basis induced by $\mathcal{A}_1$. In the present work, we avoided the overhead of solving the associated LASSO problems (and the cross-validation) and relied on the MC procedure proposed in [52, 53]. This approach estimates the first-order model's coefficients by the empirical correlation between the residual and the components of the germ:

$$\boldsymbol{W}_{k+1} = \frac{\boldsymbol{C}}{\|\boldsymbol{C}\|}, \quad \boldsymbol{C}_j = \frac{1}{M} \sum_{i=1}^{M} (\tau_{\text{ign}}^{(i)} - \tau_k^{\text{CT}}\ (\boldsymbol{\xi}^{(i)})) \xi_j^{(i)}, \tag{18}$$

where $\tau_{k=0}^{\text{CT}}\,(\boldsymbol{\xi}^{(i)}) \equiv 0$.

3.4. Preconditioned representations

We also explored the application of SR and CT approaches to construct representations of preconditioned QoI. The key idea of spectral preconditioning is to apply an invertible transformation to the QoI so that the transformed version is more easily amenable to the PC expansion, with a tighter spectrum [59, 60]. In the present work, we expand the logarithm of the ignition delay time, i.e. we consider $h(\boldsymbol{\xi}) \equiv \ln\left(\tau_{\text{ign}}(\boldsymbol{\xi})/\tau_{\text{ref}}\right)$ as the new QoI, where τ_{ref} is the ignition delay time corresponding to the nominal rate constants (i.e., $\boldsymbol{\xi} = 0$).

The motivation behind exploring the present preconditioned representations is that the positivity of the representation of the ignition delay time is guaranteed by construction [82]. Note that in the applications below, we analyze representation errors for the ignition delay time itself, exponentiating the SR or CT expansion of $\ln(\tau_{\text{ign}}/\tau_{\text{ref}})$. However, for simplicity we conduct global sensitivity analysis and estimate the Sobol indices of the preconditioned ignition delay time, i.e. $h(\boldsymbol{\xi})$. Practical techniques to compute Sobol's indices from generic preconditioned polynomial expansions in high dimension remain to be investigated.

4. Results

In this section, we first assess the performance of surrogate representations and then exploit these surrogates to conduct GSA. In section 4.1, we focus on direct representations based on SR and CT methodologies. The exercise is repeated in 4.2 using representations based on using $\log(\tau_{\text{ign}})$ as QoI; as in section 4.1, the SR and CT methods are considered, and the resulting surrogates are labeled as pSR and pCT, respectively. Results of global sensitivity analysis based on both direct and preconditioned representations are presented in section 4.3. Finally, in section 4.4, we explore the possibility of constructing reduced coordinate transform surrogates in a suitably defined subspace.

4.1. Direct representations

As discussed in section 3.2, we use a sequential approach to construct SR representations. In the first stage, the LASSO methodology is applied using a polynomial basis of total order less or equal to 1; for the present case, the basis has 3812 members. The regression leads to a first-order surrogate (SR1) with spectrum shown in Figs. 2(a,b). As mentioned in section 3.2.1 and reflected in these plots, SR1 only has 328 non-zero coefficients, which define the set of active dimensions $\mathcal{J}^{\text{1st}}$.

In the second stage, we determine second-order ($p = 2$) regularized representations using the two approaches outlined in section 3.2.2. In other words, we consider extended bases corresponding to the multi-index set $\mathcal{A}_2^{\text{2nd}}$ defined according to (10) to determine the regularized (total-order 2) surrogates SR-C and SR-D. Note that the cardinality of the basis, $|\mathcal{A}_2^{\text{2nd}}| = 54,285$, is much smaller than that of the full total second-order basis, $|\mathcal{A}_2| = 7,263,766$. In Figures 2(c,d) we plots the SR-C correction coefficients $\widehat{\boldsymbol{\delta c}}^{\text{2nd}}$ defined in (11), and in Figures 2(e,f) we plot the coefficients of SR-D. The results indicate that, for the present application, the second stage of the spare regression approach resulted in "selection" of approximately as many basis elements as in the first stage. This fact can be appreciated from Figs. 2(d,f), which reveal that about 600 non-zero coefficients are involved in the SR-C correction and about 800 in the SR-D surrogate. Comparing the results in Figs. 2(b) and (d), one observes that the amplitude of the peak coefficients in the second-order correction (about 1.5×10^{-3}) are appreciably smaller than that of the leading first-order coefficients, specifically those lying above the leftmost shoulder in the curve of Fig. 2(b). A similar observation was made for the SR-D spectrum (not shown).

To assess the respective global quality SR and CT representations, we report the RMS errors at the LHS points in Table 3. Also reported are the RMS errors normalized by the empirical L_2 norm of τ_{ign}, and the maximum and minimum values of τ_{ign} obtained by the SR and CT representations at the LHS points. The results indicate that the two approaches have comparable RMS error, with CT leading to slightly lower values. For the present application, however, the RMS errors decrease slightly as the representations are refined.

To gain additional insight into the quality of the SR and CT representations, we present QQ plots and PDFs of τ_{ign}. Figure 3 shows the QQ plots, whereas the PDFs of the SR and CT models are depicted in Fig. 4. Note that the QQ plots use the LHS points, whereas the PDFs use independently drawn MC

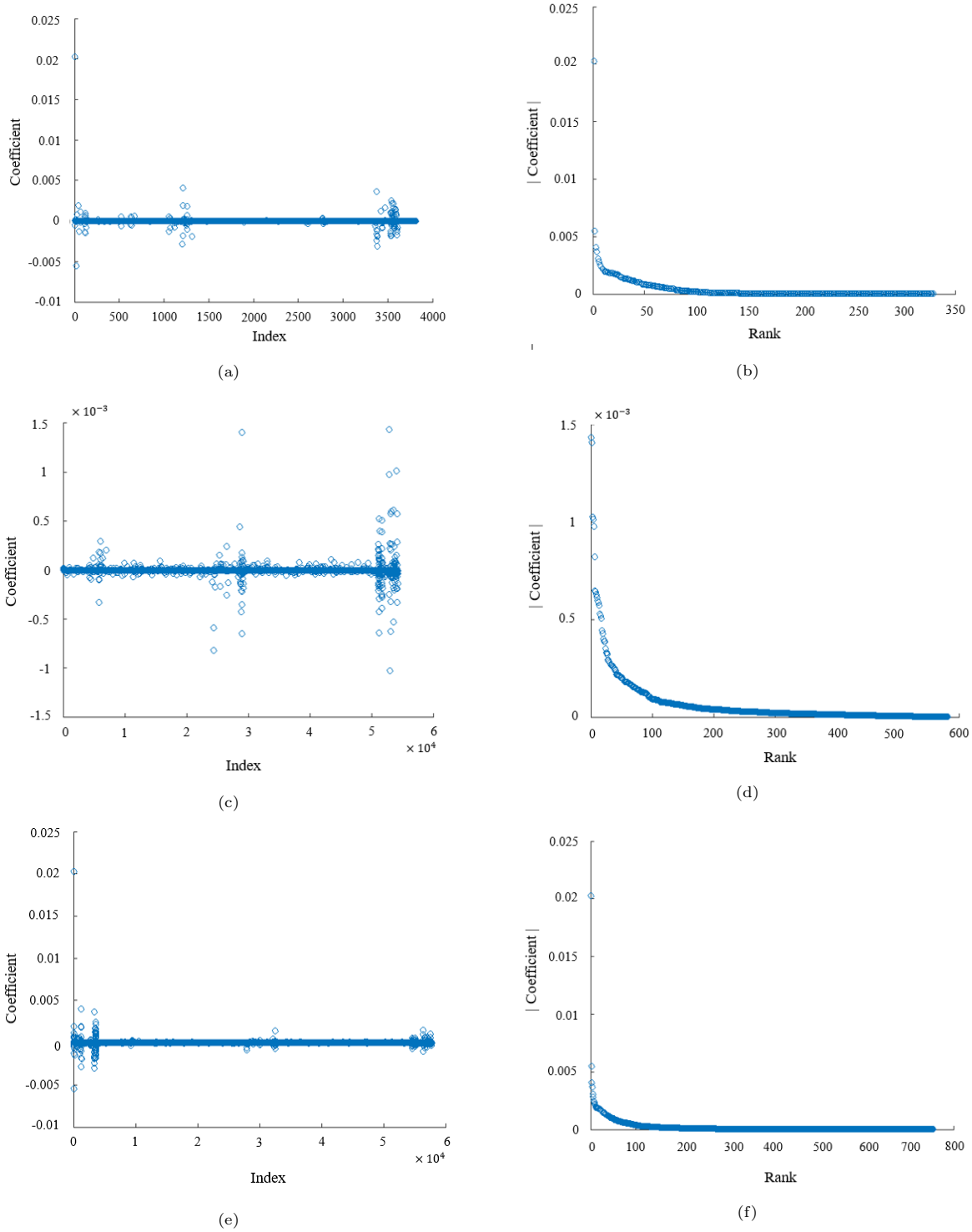

Figure 2: (a) Coefficient versus polynomial index of SR1. (b) Absolute value of non-zero coefficients in (a), sorted by decreasing magnitude. (c) Coefficient versus polynomial index for SR-C. (d) Absolute value of non-zero coefficients in (c), sorted by decreasing magnitude. (e) Coefficient versus polynomial index of SR-D. (f) Absolute value of the non-zero coefficients in (e) sorted by decreasing magnitude.

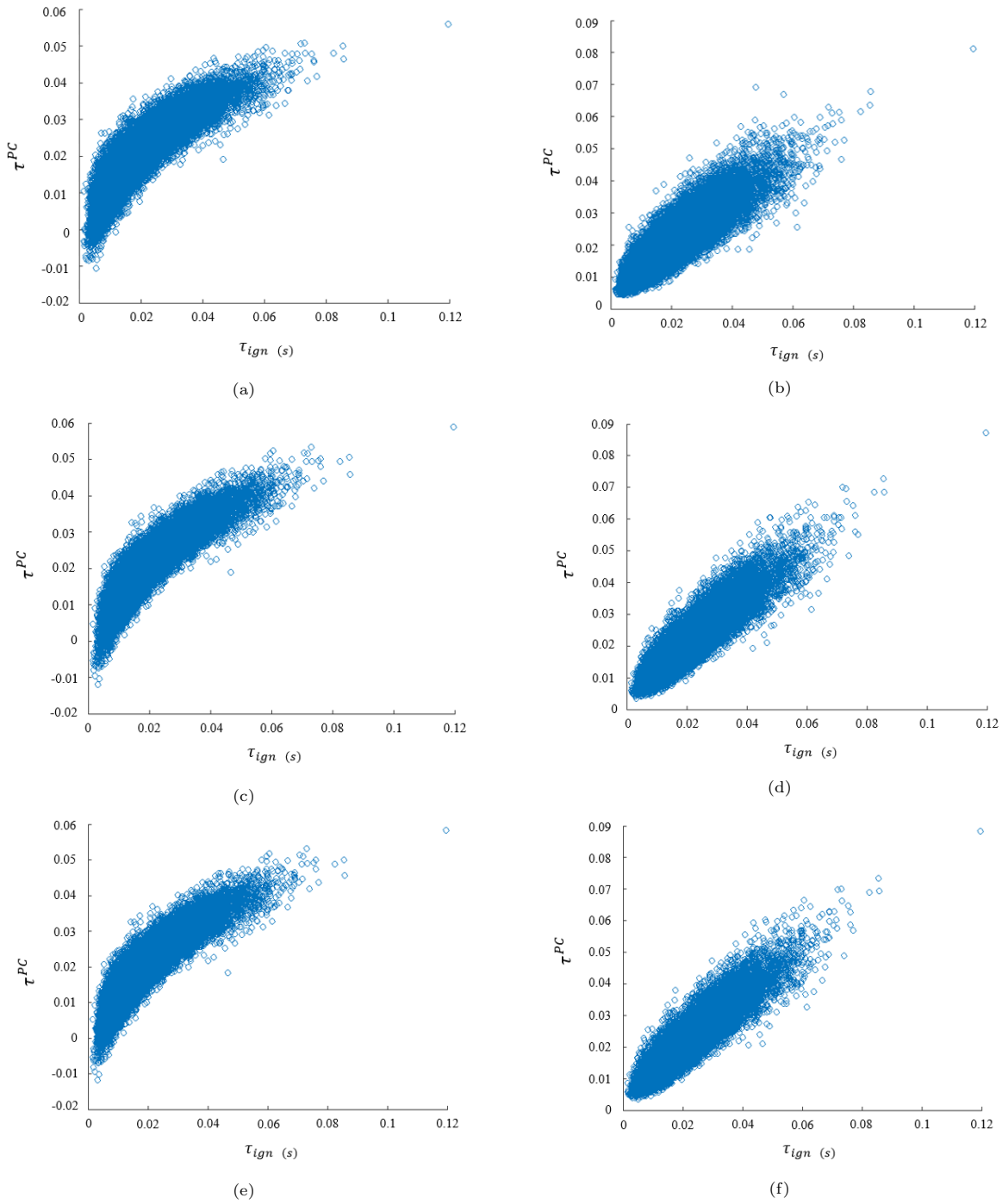

Figure 3: PC prediction versus computed values of τ at the LH sample points. Shown are results obtained using (a)SR1, (b)CT1, (c)SR-C, (d)CT2, (e)SR-D, and (f)CT3.

Table 3: RMS errors and normalized RMS errors for different methods SR1, SR-C, SR-D, CT1, CT2, and CT3. Also reported are the maximum and minimum values of τ_{ign} obtained from the different representations evaluated at the LH sample points.

Method	RMSE	Normalized RMSE	max	min
SR1	0.0045745	0.0386851	0.056008	−0.010671
SR-C	0.0041585	0.0351674	0.058907	−0.012045
SR-D	0.0041914	0.0354454	0.058275	−0.011722
CT1	0.0046484	0.0393099	0.081163	0.0043985
CT2	0.0039616	0.0335024	0.087131	0.0035425
CT3	0.0038830	0.0328370	0.088310	0.0035298

samples. The QQ plots indicate that whereas the SR and CT representations are in reasonable agreement with the simulated values of τ_{ign}, in both cases, appreciable deviations occur that are more pronounced at the extreme values of the underlying data. However, the QQ plots suggest that the CT representation is more satisfactory than that corresponding to SR. In particular, several drawbacks can be observed in the SR case: a substantial underprediction of the peak values of τ_{ign} and unphysical negative values in all the SR models; see also Table 3. In contrast, all three CT representations appear to respect the positivity of the ignition delay time and yield maxima and minima that are in better agreement with those of the LHS sample.

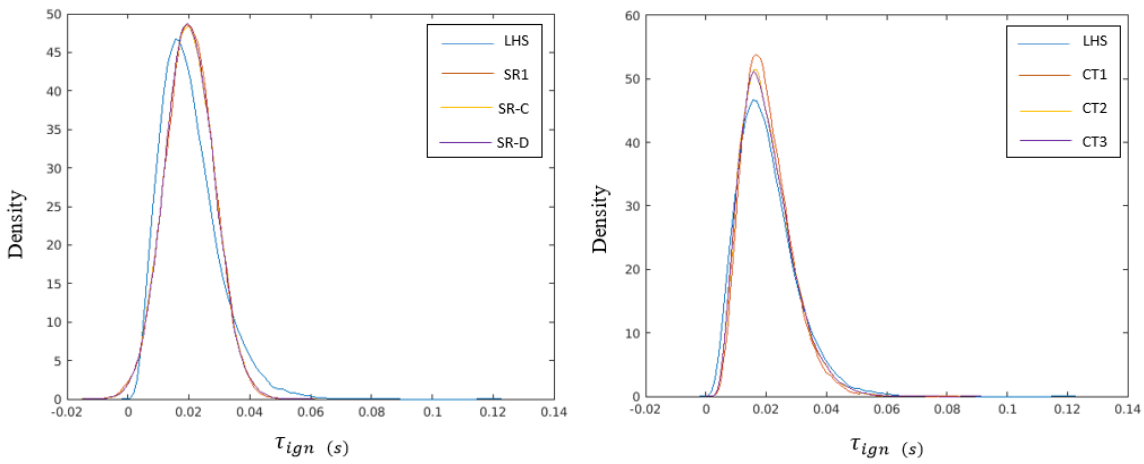

Figure 4: Left: PDFs of the ignition delay time obtained using SR1, SR-C, and SR-D; right: PDFs of the ignition delay time obtained using CT1, CT2, and CT3. In both cases, the empirical PDF generated using the LHS sample is shown for comparison.

The general trends observed in the QQ plots are also reflected in the PDFs plotted in Fig. 4. In particular, the plots show that both SR and CT reproduce reasonably well the empirical PDF, but in contrast to the SR representations, all three CT curves respect the positivity of τ_{ign}. CT also exhibits a better behavior than SR for the tails of the distribution. We also note that the three curves for SR are nearly identical, whereas the PDFs of CT tend closer to the empirical PDF as the representation is refined. Besides, one also observes that the PDFs obtained using SR have a consistent peak location, which differs from the empirical PDF. In contrast, all the CT representations' peaks align well with that of the empirical PDF.

In light of the observations above, we conclude that in the present high-dimensional application, the PC representation based on the CT methodology provides a much more robust approach for the direct representation of τ_{ign} compared to the SR approach.

4.2. Preconditioned representations

In this section, we explore the application of preconditioned representations, based on the construction of PC expansions of $\log(\tau_{\text{ign}})$. As in the previous section, both the SR and CT methodologies applied on the sampled values of $\log(\tau_{\text{ign}})$ as QoIs at the LHS coordinates.

The SR methodologies are applied similarly as in the direct representation, i.e., we first form first-order approximations of $\log(\tau_{\text{ign}})$, which we denote by P-SR1. As shown in Fig. 5(a), 307 coordinates are retained in P-SR1, a slightly smaller value than in the direct case (328 active coordinates retained in SR1). A comparison of the active coordinate sets in the direct and preconditioned representations reveals that their intersection includes 219 coordinates. Although the coordinates belonging to the two active sets of SR1 and P-SR1 represent a substantial fraction of each set, the preconditioning has a significant impact on the selected coordinates. Also note that the spectra of the non-zero elements in second-order preconditioned representations (Figs. 5c,e) are slightly broader than their direct representation counterparts in Figs. 2(d,f).

Figures 5b,c,e depict the QQ plots for the preconditioned representations, respectively P-SR1, P-SRC, and P-SRD. To generate these plots, we evaluate the PC representations of $\log(\tau_{\text{ign}})$ at the LHS points and exponentiate the result to estimate τ_{ign}. Note that because the preconditioned representation is by construction guaranteed to be positive, the QQ plots in Fig. 5 are, as expected, confined to the first quadrant. Besides, contrasting with the results in Fig. 2 reveals that the preconditioning achieves a better representation across the entire range of ignition times, in particular for the extremes. In contrast, however, the QQ plots obtained using the preconditioned CT representations (not shown) were found to be quite similar to those obtained using the direct approach. This similarity leads to the conclusion that, in the present case, preconditioning offers a significant enhancement of the representation using SR, but has an insignificant impact when using the CT approach.

To further examine the results, PDFs of τ_{ign} are generated using the preconditioned SR and CT representations and reported in Fig. 6, along with the empirical PDF. Consistent with the observations made from the QQ plots, one sees that preconditioning leads to a clear improvement in the PDFs obtained using sparse regression. In particular, the tails of the distribution are better reproduced, and the shift in the peaks observed for the direct SR approach is removed. In other words, with P-SR the most probable value of τ_{ign} appears to be well captured. With the CT approach, on the other hand, the PDFs are similar to those obtained using the direct representation approach.

Finally, we show in Table 4 the estimated values of the RMS errors estimated for the P-SR and P-CT representations. Also reported are the normalized values, as well as maximum and minimum values of τ_{ign} at the sampled coordinates. Consistent with earlier observations, the results indicate that preconditioning leads to a reduction of the estimated RMS errors for the SR representations, compared to those obtained in the direct approach. In particular, the smallest values are obtained for P-SRD, which are close to those obtained with CT3. On the other hand, a slight deterioration is observed for the preconditioned CT representations compared to the direct counterparts. One can conclude from the present experiences that preconditioning can be quite beneficial for direct representations, and should generally be explored in the associated representation context. Though preconditioning did not result in an improvement in the quality of the CT representations, due to the modest cost incurred in building the corresponding expansions, exploring the performance of preconditioned CT may still be desirable.

4.3. Global Sensitivity Analysis

GSA is applied in this section to briefly examine the results obtained using the various representations, namely to identify coordinates contributing significantly to the variability of the QoI. As outlined in Appendix A, when direct representations are applied, Sobol indices may be readily and efficiently computed using the coefficients involved in the representations. On the other hand, when preconditioned representations are used, the decomposition of the variance of τ_{ign} presents computational challenges due to the nonlinear transformation involved in the preconditioning. Rather than tackling these challenges, when using preconditioned representations we restrict our attention to estimating the Sobol indices of $\log(\tau_{\text{ign}})$ but still assess whether the simplified analysis yields consistent predictions concerning key reactions.

Figures 7 shows the first-order Sobol indices computed using (a) SR1 and (b) P-SR1, total sensitivity indices estimates using (c) SRD, (d) P-SRD, (e) CT1 and CT2, and (f) P-CT1 and P-CT2. Comparing the

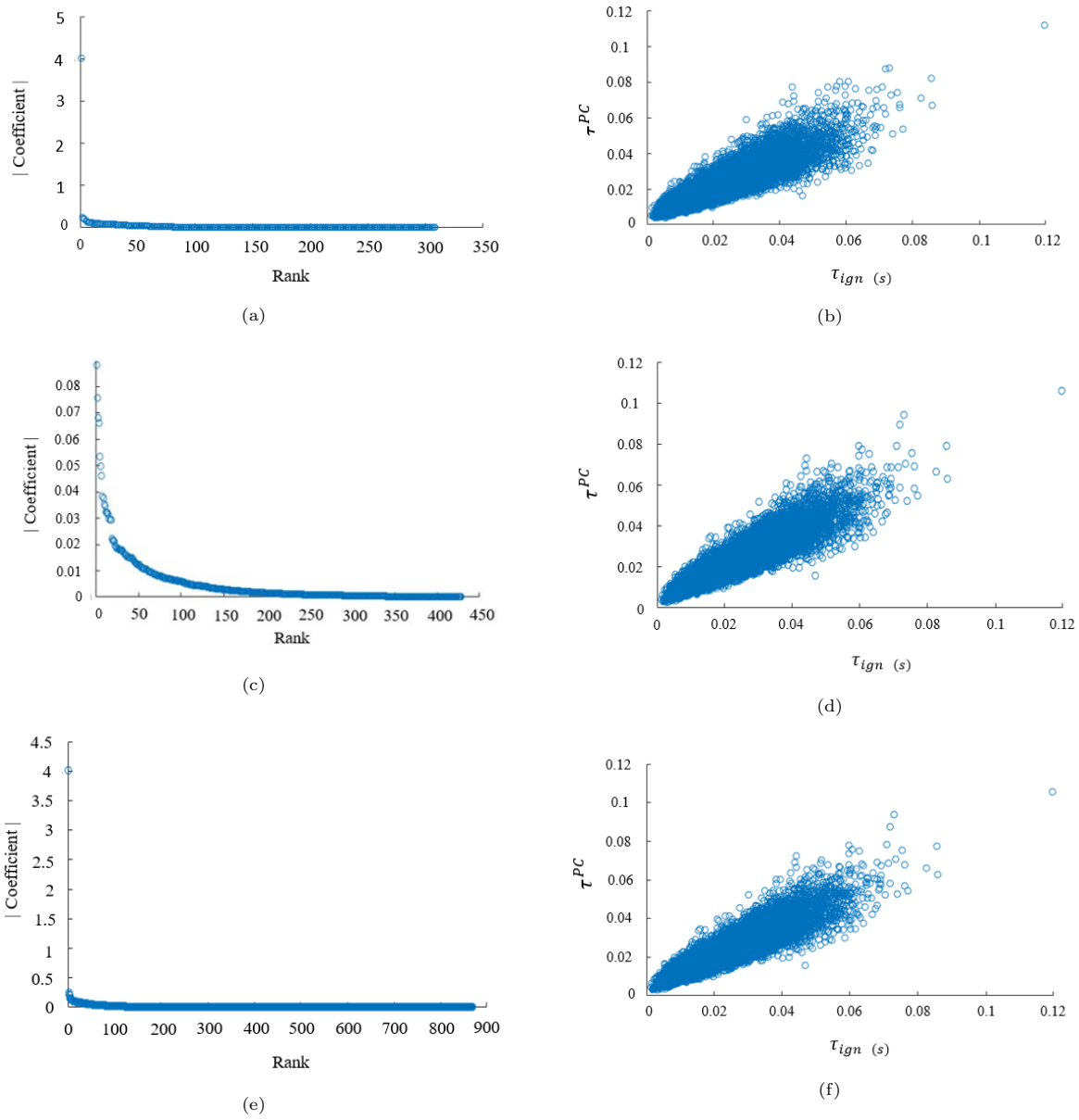

Figure 5: (a) Absolute value of non-zero coefficients in SR1 of $\log \tau$, sorted by decreasing magnitude. (b) Predicted values of τ using P-SR1 versus simulated values at the LH sample. (c) Absolute value of non-zero coefficients in SR-C for $\log \tau$, sorted by decreasing magnitude. (d) Predicted values of τ using P-SR-C versus simulated values at the LH sample. (e) Absolute value of non-zero coefficients in SR-D representation for $\log \tau$, sorted by decreasing magnitude. (f) Predicted values of τ using P-ST-D versus simulated values at the LH sample.

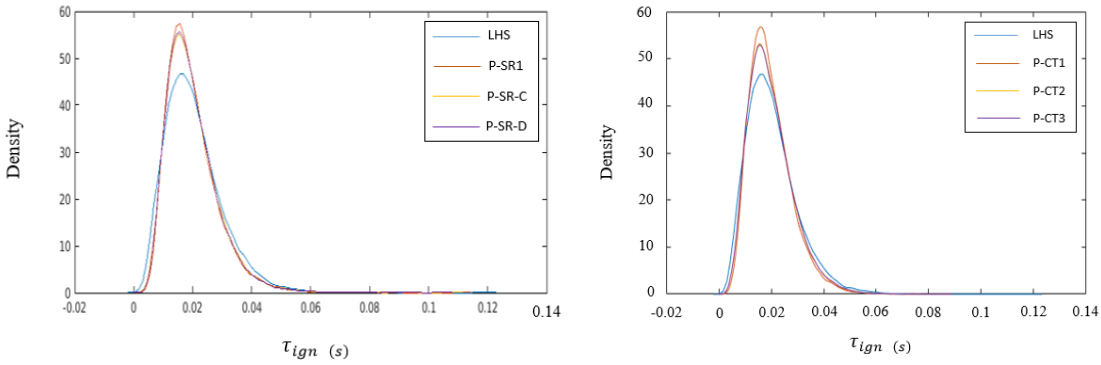

Figure 6: PDF of the ignition delay time. Curves are generated based on results obtained using (top) P-SR1, P-SR-C, and P-SR-D; and (bottom) P-CT1, P-CT2, and P-CT3. The distribution obtained from the LH sample is also shown for comparison.

Table 4: RMS errors and normalized RMS errors for different methods P-SR1, P-SRC, P-SRD, P-CT1, and P-CT2. Also reported are the maximum and minimum values of τ obtained from the different representations evaluated at the LH sample points.

Method	RMSE	Normalized RMSE	max	min
P-SR1	0.0043597	0.0368684	0.111737	0.003559
P-SRC	0.0038404	0.0324772	0.1061369	0.0030183
P-SRD	0.0038201	0.0323048	0.1053223	0.0030193
P-CT1	0.0048447	0.0409705	0.0813353	0.0034410
P-CT2	0.0041819	0.0353653	0.0848750	0.0024881

first and total order sensitivity indices obtained using the SR representations reveals that, in the present high-dimensional case, mixed interactions do not appear significant. Further, a substantial number of reactions contribute to the variance of the QoI. These observations are not surprising because, as noted earlier, higher-order corrections had moderate amplitudes yet were involved in the order of hundreds of individual coordinates. The results also illustrate that the various representations yield consistent predictions of the coordinates contributing significantly to the variance. Specifically, comparable estimates are obtained using expansions constructed using sparse regression or coordinate transform, and using direct and preconditioned representations.

Regardless of the approach used to construct the PC surrogate, the GSA analysis reveals nine reactions that dominate the variability of the ignition delay time. These reactions are listed in Table 5. The results are consistent with what one would anticipate in fuel oxidation applications, namely concerning the importance of abstraction reactions involving active radicals.

4.4. Reduced CT representation

In this section, we briefly explore the possibility of constructing CT representations in a reduced subspace. This idea is motivated by the observation that the inner products between the first eigenvector, W_1, used to construct CT1, and the directions of the individual components of $\boldsymbol{\xi}$, have a negligibly small value for a large number of indices; see Fig. 8. Letting $\mathcal{I}$ denote the set of indices, i, for which the i-th component of W_1 (direction cosine) is small, and $\mathcal{D} \equiv \{1, 2, \ldots, N_r\}$, this suggests constructing CT representations in the space spanned by the coordinates ξ_j, where $j \in \mathcal{A} \equiv \mathcal{D} \backslash \mathcal{I}$. In the experiments presented below, the subset $\mathcal{A}$ consisted of those indices whose direction cosines where located to the right of the shoulder seen in the curve depicted in Fig. 8. Specifically, in the tests presented below, 328 "active" coordinates ξ_j, $j \in \mathcal{A}$, were retained.

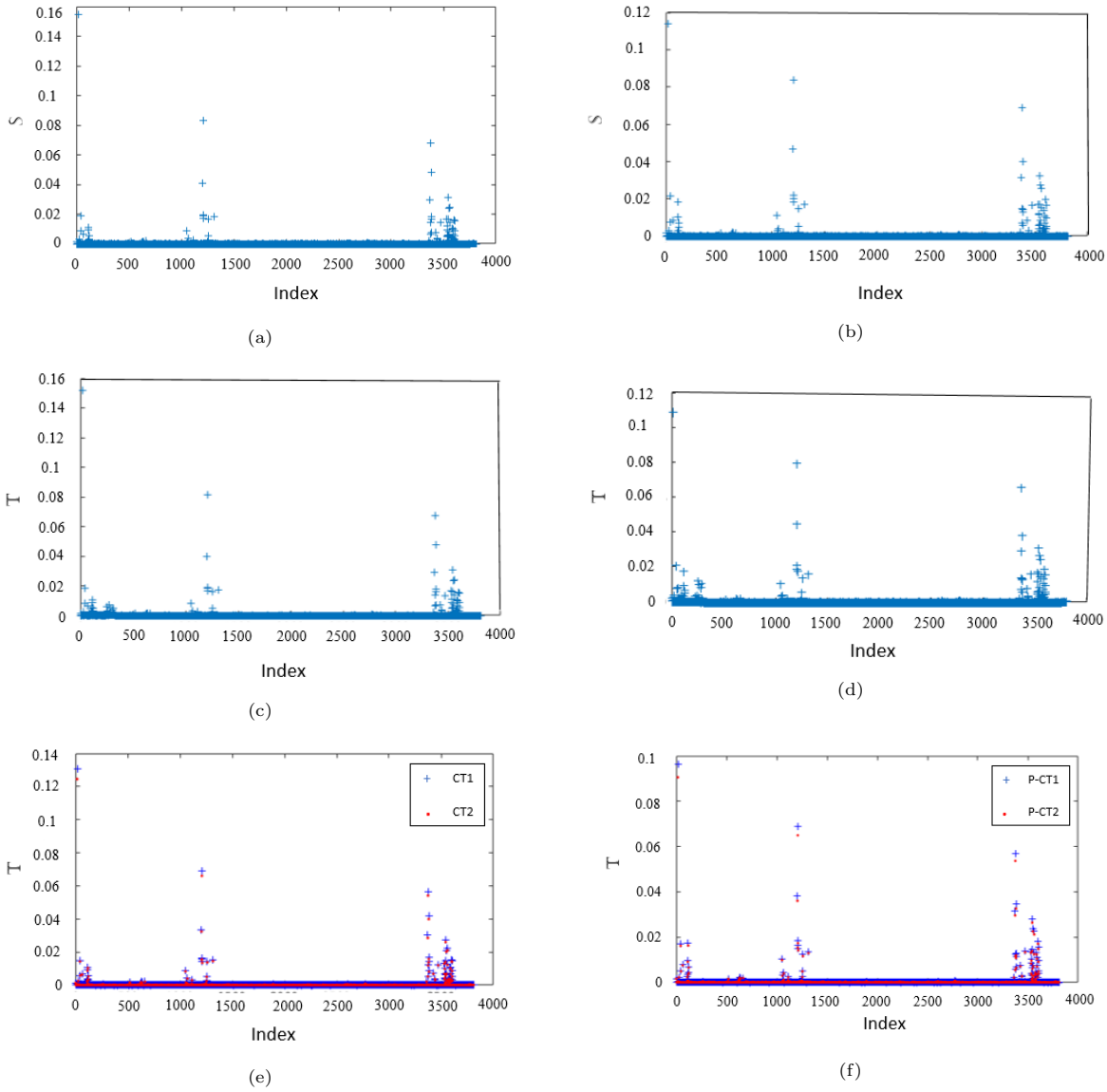

Figure 7: (a) First order sensitivity indices of τ_{ign} obtained using SR1. (b) First-order sensitivity indices of $\log(\tau_{\text{ign}})$ using P-SR1. (c) Total sensitivity indices of τ_{ign} obtained using SR-D. (d) Total sensitivity indices of $\log(\tau_{\text{ign}})$ obtained using P-SRD. (e) Total sensitivity indices of τ_{ign} obtained using CT1 and CT2. (f) Total sensitivity indices of $\log(\tau_{\text{ign}})$ obtained using P-CT1 and P-CT2.

We explored a simplified reduction approach based on constructing CT representations in the subspace spanned by the active coordinates. The methodology consisted in dropping the coordinates whose indices belong to $\mathcal{I}$, and accordingly restricting the eigenvectors involved in the CT methodology to the subspace spanned by the active coordinates. With this restriction, the CT algorithm was applied recursively to construct reduced 1D, 2D, and 3D representations, respectively labeled R-CT1, R-CT2, and R-CT3.

The performance of the CT representations in the reduced space was first examined by contrasting the reduced subspace models and corresponding full space counterparts using QQ plots. A sample of this exercise is shown in Fig. 9; it reports the QQ plot of the R-CT3 versus full space CT3 predictions. The results for R-CT1 and R-CT2 are similar to the case of R-CT3; therefore, we omit them. The QQ plot

Table 5: The most significant reactions from GSA with their reaction number.

Reaction number	Reaction
16	H2O2(+M) $\Longleftrightarrow$ OH+OH(+M)
1206	IC4H9O2 $\Longleftrightarrow$ IC4H8O2H-I
1210	TC4H9O2 $\Longleftrightarrow$ IC4H8+HO2
3373	IC8H18+OH $\Longleftrightarrow$ AC8H17+H2O
3375	IC8H18+OH $\Longleftrightarrow$ CC8H17+H2O
3382	IC8H18+HO2 $\Longleftrightarrow$ BC8H17+H2O2
3543	DC8H17O2 $\Longleftrightarrow$ DC8H16OOH-C
3550	AC8H16OOH-B $\Longleftrightarrow$ IC8ETERAB+OH
3563	BC8H16OOH-C $\Longleftrightarrow$ C8H16+HO2

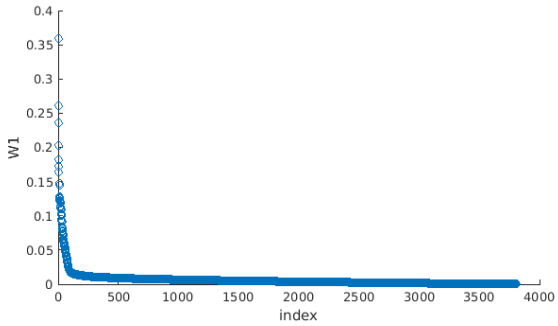

Figure 8: Absolute value of the coordinates of $\boldsymbol{W}_1$ sorted by decreasing magnitude.

indicates that the reduced R-CT3 is consistent with its full-space counterpart, with discrepancies smaller than those observed in the QQ plots contrasting the full space CT predictions and the underlying data.

A more informative assessment of the suitability of the reduced CT representations is gained by analyzing the corresponding PDFs of the ignition delay time. Results are depicted is in Fig. 10. The curves indicate that the PDFs obtained from the reduced CT representation are in excellent agreement with the empirical PDF estimated from the LHS sample. In particular, both the peak location and maxima are well captured as the dimensionality increases, with small underprediction of the tails' extent, reflecting the loss of variability in the disregarded dimensions. Further, the positivity of the representation is observed, as in the full-space, direct CT representation. Remarkably, one also observes that the PDFs obtained using the reduced CT representations exhibit a closer agreement with the simulation results than is obtained both direct and preconditioned SR or CT representations. This better agreement suggests that the reduction applied could mitigate the impact of using a moderate size sample, and consequently, of the statistical noise inherently present in estimating eigenvectors involved in the CT transforms. Of course, additional investigations would be needed to understand this phenomenon entirely, but these would be best conducted in a synthetic setting that affords the efficient generation of large samples.

5. Conclusions

This exploratory work focused on investigating the potential of constructing functional representations based on SR and CT techniques to represent the output of oxidation simulations in a high-dimensional context and exploiting these functional representations to perform global sensitivity analysis. Attention was focused on a detailed iso-octane mechanism comprising 3811 uncertain elementary reactions. With the ignition delay time used as sole QoI, a moderate size Latin Hypercube sample was generated and used to

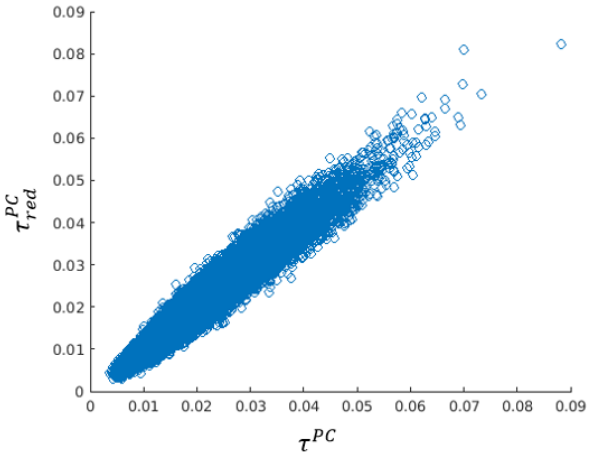

Figure 9: QQ plot of reduced dimensionality CT3 representation versus the original representation.

assess PC surrogates constructed using both direct and preconditioned representations of the QoI.

Assessment of the PC surrogates revealed that both SR and CT lead to representations having comparable mean square errors. When using direct representations, however, the SR representations exhibited drawbacks, including apparent bias in the location of the PDF's peak and prediction of negative values. The CT constructions avoided these drawbacks and are deemed more suitable. On the other hand, with the preconditioned QoI, both the SR and CT constructions yielded representations with comparable mean square errors and PDFs in good agreement with the empirical PDFs. Consequently, we recommend the preconditioning to enhance the quality of sparse regression algorithms in high dimension.

The direct and preconditioned representations were exploited to estimate global sensitivity indices of the QoI to the individual components of the germ (reactions). Computations indicated that the various functional representations of the QoI yielded consistent results. Finally, we briefly explored the potential of combining a subspace reduction methodology with the CT construction algorithms. Simulation results suggested that such methodology may be quite attractive in a high dimensional setting.

The present experiences in sampling and representation highlighted several computational challenges that are severe when considering high dimensional uncertainty germs. These include memory large or prohibitive memory requirements to store the coordinates of the sampled germ (LHS), the full storage of the multi-indices describing the structure of the elements constituting the approximation basis, and, finally, the assembly and storage of the full regression matrices. At the computational level, the inversion of the large full matrices associated with the regression problems and the computation of the Sobol indices in CT constructions called for specific solutions. Concerning the multi-index storage, we implemented a sparse multi-index methodology, which restricts the multi-index dimension to the order of the PC basis. For the present application, this effectively mitigated otherwise prohibitive memory requirements. The SR and CT methodologies were designed to involve regression matrices with manageable sizes, thanks to expansions limited to first-order or sought in reduced spaces for higher-order. For the estimation of Sobol indices from a CT approximation, it proved to be especially challenging because of the expansion in a non-orthogonal basis. We proposed an analytical transform to address this challenge and project the CT functionals onto an orthogonal basis and account, on-the-fly, for contributions to the Sobol indices associated with individual reactions.

The mitigation of the storage cost of large samples of a high-dimensional germ fell outside the scope of the present work and, consequently, is left for future work. In addition to the development of reduced storage approaches, it would also be worthwhile to explore the possibility of using alternative random variables to represent the a priori variability of the rate parameters. Specifically, whereas the parametrization of the pre-exponentials in the form $\boldsymbol{U}\boldsymbol{F}^{\boldsymbol{\xi}}$ is fundamentally attractive, a drawback of using independent random

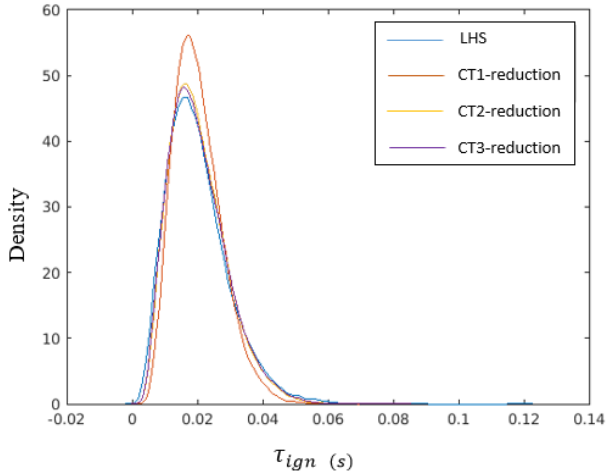

Figure 10: PDF of the ignition delay time reduced CT representations. Also shown for comparison is the PDF generated using LHS simulation results.

variables with uniform densities is that the sampled scaling vectors have norms significantly larger than 1 with high probability; see Fig. 11. As a result, unless an excessively large sample is used, none of the sampled realizations will lie close to the nominal value ($\boldsymbol{\xi} = 0$). Thus, it would appear worthwhile to consider germs with independent components having densities concentrated at the nominal value. Another avenue that we are currently exploring concerns the development of an efficient methodology to estimate derivative sensitivities in individual realizations and enable the determination of active subspaces for reduced dimensionality representations.

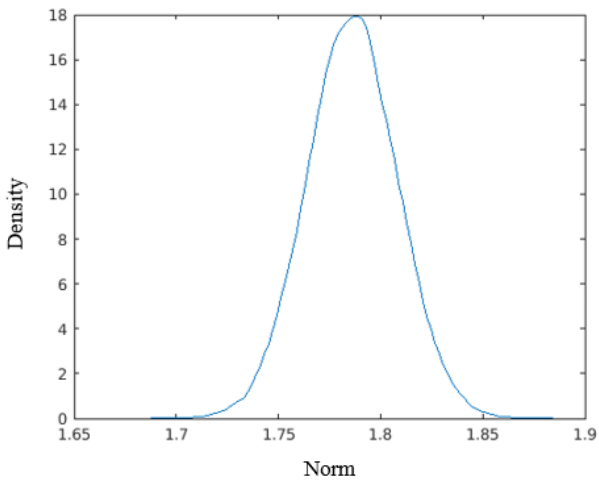

Figure 11: PDF of the norm of normalized scaling vector, $\left(\sum_{i=1}^{d}\left(UF_i^{\xi_i}\right)^2\right)^{1/2}/\sqrt{d}$.

Appendix A. Global sensitivity analysis

Appendix A.1. Sobol indices from PC expansion

In this work, we consider global sensitivity analyses based on the Sobol decomposition of the variance of the QoI [36, 35, 38, 37, 83, 39]. When the PC representation is expressed in terms of an orthogonal basis, as in LASSO constructions outlined above, the PC coefficients can be readily exploited to quantify the sensitivity of the solution with respect to the uncertain inputs $\boldsymbol{\xi}$ of the expansion. The Sobol decomposition is used to estimate first-order sensitivity and total order sensitivity of the QoI to individual components ξ_i of the germ. Consider a QoI $f(\boldsymbol{\xi})$ with PC expansion

$$f(\boldsymbol{\xi}) = \sum_{\alpha \in \mathcal{A}} c_\alpha \Psi_\alpha(\boldsymbol{\xi}).$$

The variance of f is given by

$$\mathbb{V}(f) = \sum_{\alpha \in \mathcal{A} \setminus \{(0 \ldots 0)\}} c_\alpha^2 \|\Psi_\alpha\|^2.$$

In addition, the first-order sensitivity indices, $\mathbb{S}(i)$, and total sensitivity indices, $\mathbb{T}(i)$, of a QoI can be obtained according to [37]:

$$\mathbb{S}_i = \frac{\sum_{\alpha \in \mathcal{S}_i} c_\alpha^2 \|\Psi_\alpha\|^2}{\mathbb{V}(f)}, \tag{A.1}$$

and

$$\mathbb{T}_i = \frac{\sum_{\alpha \in \mathcal{T}_i} c_\alpha^2 \|\Psi_\alpha\|^2}{\mathbb{V}(f)}, \tag{A.2}$$

where $i \in \{1, ..., d\}$, $\mathcal{S}_i = \{\alpha \in \mathcal{A} \mid \alpha_i > 0 \text{ and } \alpha_j = 0 \text{ for } j \neq i\}$, whereas $\mathcal{T}_i = \{\alpha \in \mathcal{A} \mid \alpha_i > 0\}$.

Appendix A.2. Sobol indices for CT

The computation of the Sobol indices is much more challenging for the coordinate transformation method because the polynomial expansion does not involve orthogonal polynomials, and the transformed coordinates are not independent. We initially explored the possibility of applying Monte-Carlo sampling [84, 38, 37] of the CT representations; this approach proved unsuitable because of the high dimensionality of the input space. Specifically, we experienced substantial statistical convergence hurdles that manifested in the generation of a large number of negative indices with small magnitudes. To overcome this challenge, we developed an analytical mapping technique that enabled the projection of the CT representations' directions onto individual elements of the original orthogonal basis in $\boldsymbol{\xi}$. Using this projection, one can, on-the-fly, compute the corresponding contributions to different Sobol indices (with respect to the dimensions of $\boldsymbol{\xi}$).

We illustrate this procedure for the case of a second-order polynomial expansion (with Legendre polynomials) in a single transformed variable, $\eta \equiv \boldsymbol{W}^T \boldsymbol{\xi}$. We start by recalling that the zero-th, first and second order Legendre polynomials are $\psi_0 \equiv 1$, $\psi_1(x) \equiv x$, and $\psi_2(x) = (3x^2 - 1)/2$ respectively. In this case, the functional representation of $f(\boldsymbol{\xi})$ is of the form:

$$\begin{aligned} f(\boldsymbol{\xi}) &= h(\eta = \boldsymbol{W}^T \boldsymbol{\xi}) = a_0 \psi_0(\eta) + a_1 \psi_1(\eta) + a_2 \psi_2(\eta) = a_0 + a_1 \eta + a_2 \left(\frac{3}{2} \eta^2 - \frac{1}{2} \right) \\ &= a_0 - \frac{a_2}{2} + a_1 \eta + \frac{3a_2}{2} \eta^2. \end{aligned}$$

Inserting the coordinate transformation, $\eta = \boldsymbol{W}^T \boldsymbol{\xi}$, it comes

$$f(\boldsymbol{\xi}) = h(\eta = \boldsymbol{W}^T \boldsymbol{\xi}) = a_0 - \frac{a_2}{2} + a_1 \sum_{i=1}^{d} W_i \xi_i + \frac{3a_2}{2} \left(\sum_{i=1}^{d} W_i \xi_j \right)^2.$$

Similarly, the second-order PC expansion of $h(\boldsymbol{\xi})$ writes as:

$$\begin{aligned} f(\boldsymbol{\xi}) &= c_0\psi_0 + \sum_{i=1}^{d} c_i\psi_1(\xi_i) + \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} c_{ij}\psi_1(\xi_i)\psi_1(\xi_j) + \sum_{i=1}^{d} c_{ii}\psi_2(\xi_i) \\ &= c_0 + \sum_{i=1}^{d} c_i\xi_i + \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} c_{ij}\xi_i\xi_j + \frac{1}{2}\sum_{i=1}^{d} c_{ii}(3\xi_i^2 - 1). \end{aligned} \tag{A.3}$$

By identification, and using that $\|\boldsymbol{W}\|^2 = 1$, we find that

$$c_0 = a_0 \tag{A.4}$$

$$c_i = a_1 W_i, \quad i = 1, \ldots, d \tag{A.5}$$

$$c_{ii} = a_2 W_i^2, \quad i = 1, \ldots, d \tag{A.6}$$

$$c_{ij} = 3a_2 W_i W_j, \quad 1 \le i, j \le d, i \ne j. \tag{A.7}$$

Note that (A.3) is an orthogonal expansion, and so it is straightforward to estimate the total sensitivities. In particular, the variance is

$$\mathbb{V}(f) = \sum_{i=1}^{d} c_i^2 \|\psi_1\|^2 + \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} c_{ij}^2 \|\psi_1\|^4 + \sum_{i=1}^{d} c_{ii}^2 \|\psi_2\|^2, \tag{A.8}$$

so the first-order sensitivity indices are given by

$$\mathbb{S}_i = \frac{a_1^2 W_i^2 \|\psi_1\|^2 + a_2^2 W_{ii}^4 \|\psi_2\|^2}{\mathbb{V}(f)} \tag{A.9}$$

and the total order sensitivity indices

$$\mathbb{T}_i = \frac{a_1 W_i^2 \|\psi_1\|^2 + 9a_2 W_i^2 \sum_{j\ne i} W_j^2 \|\psi_1\|^4 + a_2^2 W_i^4 \|\psi_2\|^2}{\mathbb{V}(f)} \tag{A.10}$$

Note that the coefficients in (A.5–A.7) can be individually evaluated. Consequently, their contributions to the sensitivity indices may be accounted for without the need to define the multi-indices for the d-dimensional PCE. Besides, the above analytical transformation mapping the coefficients of the CT representation η to the PCE in the canonical coordinates $\boldsymbol{\xi}$ may be generalized to higher-order expansions. Further, a similar identification remains feasible when the CT construction involves several directions $\boldsymbol{W}_i$ with multiple η_i's. This procedure enables us to estimate the Sobol indices as soon as the CT representation's coefficients are known.

Acknowledgments

Research reported in this publication was supported by research funding from King Abdullah University of Science and Technology (KAUST). The authors are grateful to Prof. Aamir Farooq for helpful discussions.

References

[1] D. Davidson, M. Oehlschlaeger, J. Herbon, R. Hanson, Shock tube measurements of iso-octane ignition times and oh concentration time histories, Proc. Combust. Inst. 29 (2002) 1295 – 1301. doi:10.1016/S1540-7489(02)80159-6.

[2] D. Davidson, R. Hanson, Interpreting shock tube ignition data, International Journal of Chemical Kinetics 36 (9) (2004) 510–523. doi:10.1002/kin.20024.

[3] M. Fikri, J. Herzler, R. Starke, C. Schulz, P. Roth, G. Kalghatgi, Autoignition of gasoline surrogates mixtures at intermediate temperatures and high pressures, Combustion and Flame 152 (1-2) (2008) 276–281. doi:10.1016/j.combustflame.2007.07.010.

[4] H. Li, L. Yu, S. Sun, S. Wang, X. Lu, Z. Huang, A shock tube experimental and modeling study of multicomponent gasoline surrogates diluted with exhaust gas recirculation, Energy and Fuels 32 (3) (2018) 3800–3813. doi:10.1021/acs.energyfuels.7b01028.

[5] S. S. Goldsborough, A chemical kinetically based ignition delay correlation for iso-octane covering a wide range of conditions including the ntc region, Combustion and Flame 156 (6) (2009) 1248–1262.
[6] R. Minetti, M. Carlier, M. Ribaucour, E. Therssen, L. Sochet, Comparison of oxidation and autoignition of the two primary reference fuels by rapid compression, Proc. Combust. Inst. 26 (1) (1996) 747–753.
[7] R. Minetti, M. Ribaucour, M. Carlier, L. R. Sochet, Autoignition delays of a series of linear and branched chain alkanes in the intermediate range of temperature, Combustion science and technology 113 (1) (1996) 179–192.
[8] G. Vanhove, G. Petit, R. Minetti, Experimental study of the kinetic interactions in the low-temperature autoignition of hydrocarbon binary mixtures and a surrogate fuel, Combustion and Flame 145 (3) (2006) 521–532.
[9] S. Tanaka, F. Ayala, J. C. Keck, A reduced chemical kinetic model for hcci combustion of primary reference fuels in a rapid compression machine, Combustion and flame 133 (4) (2003) 467–481.
[10] D. Healy, H. Curran, S. Dooley, J. Simmie, D. Kalitan, E. Petersen, G. Bourque, Methane/propane mixture oxidation at high pressures and at high, intermediate and low temperatures, Combustion and Flame 155 (3) (2008) 451–461.
[11] E. Petersen, M. Lamnaouer, J. De Vries, H. Curran, J. Simmie, M. Fikri, C. Schulz, G. Bourque, Discrepancies between shock tube and rapid compression machine ignition at low temperatures and high pressures, in: Shock Waves, Springer, 2009, pp. 739–744.
[12] X. He, S. Walton, B. Zigler, M. Wooldridge, A. Atreya, Experimental investigation of the intermediates of isooctane during ignition, International Journal of Chemical Kinetics 39 (9) (2007) 498–517.
[13] S. Walton, X. He, B. Zigler, M. Wooldridge, A. Atreya, An experimental investigation of iso-octane ignition phenomena, Combustion and Flame 150 (3) (2007) 246–262.
[14] X. He, M. Donovan, B. Zigler, T. Palmer, S. Walton, M. Wooldridge, A. Atreya, An experimental and modeling study of iso-octane ignition delay times under homogeneous charge compression ignition conditions, Combustion and Flame 142 (3) (2005) 266–275.
[15] X. He, B. Zigler, S. Walton, M. Wooldridge, A. Atreya, A rapid compression facility study of oh time histories during iso-octane ignition, Combustion and flame 145 (3) (2006) 552–570.
[16] G. Mittal, C.-J. SUNG*, A rapid compression machine for chemical kinetics studies at elevated pressures and temperatures, Combustion Science and Technology 179 (3) (2007) 497–530.
[17] G. Mittal, C.-J. Sung, Homogeneous charge compression ignition of binary fuel blends, Combustion and Flame 155 (3) (2008) 431–439.
[18] N. Wiener, The homogeneous chaos, Am. J. Math. 60 (1938) 897–936. doi:10.2307/2371268.
[19] R. G. Ghanem, P. D. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer New York, 1991. doi:10.1007/978-1-4612-3094-6.
[20] O. L. Maître, O. Knio, Spectral Methods for Uncertainty Quantification, Springer, 2010.
[21] M. T. Reagan, H. N. Najm, R. G. Ghanem, O. M. Knio, Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection, Combust. Flame 132 (3) (2003) 545–555.
[22] M. Reagan, H. Najm, B. Debusschere, O. L. Maître, O. Knio, R. Ghanem, Spectral stochastic uncertainty quantification in chemical systems, Combust. Theory Model. 8 (2004) 607–632.
[23] M. Reagan, H. Najm, P. Pébay, O. Knio, R. Ghanem, Quantifying uncertainty in chemical systems modeling, Int. J. Chem. Kinet. 37 (2005) 368–382.
[24] H. N. Najm, B. J. Debusschere, Y. M. Marzouk, S. Widmer, O. Le Maître, Uncertainty quantification in chemical systems, Int. J. Numer. Meth. Eng. 80 (6-7) (2009) 789–814.
[25] O. P. Le Maître, H. N. Najm, P. P. Pebay, R. G. Ghanem, O. M. Knio, Multi-resolution-analysis scheme for uncertainty quantification in chemical systems, SIAM J. Sci. Comput. 29 (2) (2007) 864–889. doi:Doi10.1137/05064318.
[26] D. A. Sheen, H. Wang, The method of uncertainty quantification and minimization using polynomial chaos expansions, Combust. Flame 158 (12) (2011) 2358–2374. doi:Doi10.1016/J.Combustflame.2011.05.010.
[27] D. A. Sheen, H. Wang, Combustion kinetic modeling using multispecies time histories in shock-tube oxidation of heptane, Combust. Flame 158 (4) (2011) 645–656.
[28] M. Villegas, F. Augustin, A. Gilg, A. Hmaidi, U. Wever, Application of the polynomial chaos expansion to the simulation of chemical reactors with uncertainties, Math. Comput. Simulat. 82 (5) (2012) 805–817. doi:10.1016/j.matcom.2011.12.001.
[29] D. Kim, F. Rizzi, K. W. Cheng, J. Han, F. Bisetti, O. M. Knio, Uncertainty quantification of ion chemistry in lean and stoichiometric homogenous mixtures of methane, oxygen, and argon, Combust. Flame 162 (7) (2015) 2904–2915.
[30] J. Winokur, D. Kim, F. Bisetti, O. P. Le Maître, O. M. Knio, Sparse pseudo spectral projection methods with directional adaptation for uncertainty quantification, J. Sci. Comput. 68 (2) (2016) 596–623.
[31] F. Bisetti, D. Kim, O. Knio, Q. Long, R. Tempone, Optimal bayesian experimental design for priors of compact support with application to shock-tube experiments for combustion kinetics, Int. J. Numer. Meth. Eng. 108 (2016) 136–155.
[32] D. Kim, I. El Gharamti, M. Hantouche, A. E. Elwardany, A. Farooq, F. Bisetti, O. Knio, A hierarchical method for bayesian inference of rate parameters from shock tube data: Application to the study of the reaction of hydroxyl with 2-methylfuran, Combust. Flame 184 (2017) 55–67.
[33] M. Hantouche, S. Sarathy, O. Knio, Global sensitivity analysis of n-butanol reaction kinetics using rate rules, Combust. Flame 96 (2018) 452–465.
[34] R. Malpica Galassi, M. Valorani, H. Najm, C. Safta, M. Khalil, P. Ciottoli, Chemical model reduction under uncertainty, Combustion and Flame 179 (2017) 242–252. doi:10.1016/j.combustflame.2017.02.018.
[35] I. M. Sobol', On sensitivity estimation for nonlinear mathematical models, Matematicheskoe Modelirovanie 2 (1) (1990) 112–118.
[36] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates, Math. Comput. Simulat. 55 (1) (2001) 271–280.

[37] T. Crestaux, O. Le Maître, J.-M. Martinez, Polynomial chaos expansion for sensitivity analysis, Reliab. Eng. Syst. Safe. 94 (7) (2009) 1161–1172.
[38] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliab. Eng. Syst. Safe. 93 (7) (2008) 964–979.
[39] A. Alexanderian, J. Winokur, I. Sraj, A. Srinivasan, M. Iskandarani, W. C. Thacker, O. M. Knio, Global sensitivity analysis in an ocean general circulation model: a sparse spectral projection approach, Computat. Geosci. 16 (3) (2012) 757–778.
[40] P. G. Constantine, M. S. Eldred, E. T. Phipps, Sparse pseudospectral approximation method, Comput. Method. Appl. M. 229 (2012) 1–12. doi:Doi10.1016/J.Cma.2012.03.019.
[41] P. R. Conrad, Y. M. Marzouk, Adaptive smolyak pseudospectral approximations, SIAM J. Sci. Comput. 35 (6) (2013) A2643–A2670. doi:Doi10.1137/120890715.
[42] J. D. Jakeman, M. S. Eldred, K. Sargsyan, Enhancing $l1$-minimization estimates of polynomial chaos expansions using basis selection, Journal of Computational Physics 289 (2015) 18–34.
[43] G. Blatman, B. Sudret, Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach, Comptes Rendus Mécanique 336 (6) (2008) 518–523.
[44] R. Tibshirani, Regression shrinkage and selection via the lasso: a retrospective, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73 (3) (2011) 273–282.
[45] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, Sparsity and smoothness via the fused lasso, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67 (1) (2005) 91–108.
[46] S. L. Kukreja, J. Lofberg, M. J. Brenner, A least absolute shrinkage and selection operator (lasso) for nonlinear system identification, IFAC Proceedings Volumes 39 (2006) 814 – 819. doi:10.3182/20060329-3-AU-2901.00128.
[47] H. Cheng, A. Sandu, Collocation least-squares polynomial chaos method, in: Proceedings of the 2010 Spring Simulation Multiconference, Society for Computer Simulation International, 2010, p. 80.
[48] M. Berveiller, B. Sudret, M. Lemaire, Stochastic finite element: a non intrusive approach by regression, European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique 15 (1-3) (2006) 81–92.
[49] D. W. Marquardt, R. D. Snee, Ridge regression in practice, The American Statistician 29 (1) (1975) 3–20.
[50] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, et al., Least angle regression, The Annals of statistics 32 (2) (2004) 407–499.
[51] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, Journal of Computational Physics 230 (6) (2011) 2345–2367.
[52] R. Tipireddy, R. Ghanem, Basis adaptation in homogeneous chaos spaces, Journal of Computational Physics 259 (2014) 304–317.
[53] P. Tsilifis, R. G. Ghanem, Reduced wiener chaos representation of random fields via basis adaptation and projection, Journal of Computational Physics 341 (2017) 102–120.
[54] P. Tsilifis, X. Huan, C. Safta, K. Sargsyan, G. Lacaze, J. Oefelein, H. Najm, R. Ghanem, Compressive sensing adaptation for polynomial chaos expansions, Journal of Computational Physics 380 (2019) 29–47. doi:10.1016/j.jcp.2018.12.010.
[55] P. Constantine, Active Subspaces, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2015. doi:10.1137/1.9781611973860.
[56] P. Constantine, P. Diaz, Global sensitivity metrics from active subspaces, Reliab. Eng. Syst. Saf. 162 (2017) 1–13.
[57] M. Vohra, A. Alexanderian, C. Safta, S. Mahadevan, Sensitivity-driven adaptive construction of reduced-space surrogates, Journal of Scientific Computing 79 (2) (2019) 1335–1359. doi:10.1007/s10915-018-0894-4.
[58] M. Vohra, A. Alexanderian, H. Guy, S. Mahadevan, Active subspace-based dimension reduction for chemical kinetics applications with epistemic uncertainty, Combustion and Flame (2019) 152–161doi:10.1016/j.combustflame.2019.03.006.
[59] A. Alexanderian, O. Le Maître, H. Najm, M. Iskandarani, O. Knio, Multiscale stochastic preconditioners in non-intrusive spectral projection, SIAM J. Sci. Comput. 50 (2012) 306–340.
[60] A. Alexanderian, F. Rizzi, M. Rathinam, O. P. Le Maître, O. M. Knio, Preconditioned bayesian regression for stochastic chemical kinetics, J. Sci. Comput. 58 (3) (2014) 592–626.
[61] M. Mehl, H. Curran, W. Pitz, C. Westbrook, iso-octane, version 3 (2009).
[62] M. Mehl, H. Curran, W. Pitz, C. Westbrook, Chemical kinetic modeling of component mixtures relevant to gasoline, in: 4th European Combustion Meeting, Vienna, Austria, April 14–17, 2009.
[63] M. Mehl, W. Pitz, M. Sjöberg, J. E. Dec, Detailed kinetic modeling of low-temperature heat release for prf fuels in an hcci engine, in: SAE Technical Paper 2009-01-1806, 2009. doi:10.4271/2009-01-1806.
[64] T. Nagy, T. Turányi, Uncertainty of arrhenius parameters, Int. J. Chem. Kinet. 43 (7) (2011) 359–378.
[65] M. Mendes, J. Pereira, J. Pereira, Calculation of premixed combustion within inert porous media with model parametric uncertainty quantification, Combust. Flame 158 (3) (2011) 466–476.
[66] D. Baulch, C. Bowman, C. Cobos, R. Cox, T. Just, J. Kerr, M. Pilling, D. Stocker, J. Troe, W. Tsang, et al., Evaluated kinetic data for combustion modeling: supplement ii, J. Phys. Chem. Ref. Data 34 (3) (2005) 757–1397.
[67] A. A. Konnov, Remaining uncertainties in the kinetic mechanism of hydrogen combustion, Combust. Flame 152 (4) (2008) 507–528.
[68] T. Turányi, L. Zalotai, S. Dóbé, T. Bérces, Effect of the uncertainty of kinetic and thermodynamic data on methane flame simulation results, Phys. Chem. Chem. Phys. 4 (12) (2002) 2568–2578.
[69] C. K. Westbrook, F. L. Dryer, Chemical kinetic modeling of hydrocarbon combustion, Prog. Energy Combust. Sci. 10 (1) (1984) 1–57.
[70] H. J. Curran, P. Gaffuri, W. J. Pitz, C. K. Westbrook, A comprehensive modeling study of n-heptane oxidation, Combust. Flame 114 (1) (1998) 149–177.

[71] H. J. Curran, P. Gaffuri, W. Pitz, C. Westbrook, A comprehensive modeling study of iso-octane oxidation, Combust. Flame 129 (3) (2002) 253–280.
[72] C. K. Westbrook, W. J. Pitz, O. Herbinet, H. J. Curran, E. J. Silke, A comprehensive detailed chemical kinetic reaction mechanism for combustion of n-alkane hydrocarbons from n-octane to n-hexadecane, Combust. Flame 156 (1) (2009) 181–199.
[73] S. M. Sarathy, C. K. Westbrook, M. Mehl, W. J. Pitz, C. Togbe, P. Dagaut, H. Wang, M. A. Oehlschlaeger, U. Niemann, K. Seshadri, P. S. Veloo, C. Ji, F. N. Egolfopoulos, T. Lu, Comprehensive chemical kinetic modeling of the oxidation of 2-methylalkanes from c 7 to c 20, Combust. Flame 158 (12) (2011) 2338–2357.
[74] S. M. Sarathy, S. Vranckx, K. Yasunaga, M. Mehl, P. Oßwald, W. K. Metcalfe, C. K. Westbrook, W. J. Pitz, K. Kohse-Höinghaus, R. X. Fernandes, H. J. Curran, A comprehensive chemical kinetic combustion model for the four butanol isomers, Combust. Flame 159 (6) (2012) 2028–2055.
[75] C. Safta, H. Najm, O. Knio, Tchem - a software toolkit for the analysis of complex kinetic models, Tech. Rep. Report No. SAND2011-3282, Sandia National Laboratories, Livermore, CA, USA (2011).
[76] S. Almohammadi, O. Knio, O. L. Maître, Ignition delay data set, doi:10.25781/KAUST-U92E2 (2020).
[77] D. Xiu, G. E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.
[78] O. Knio, O. L. Maître, Uncertainty propagation in CFD using polynomial chaos decom-position, Fluid Dynamics Research 38 (2006) 616–640.
[79] H. Najm, Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, Annual Review of Fluid Mechanics 41 (2009) 35–52. doi:10.1146/annurev.fluid.010908.165248.
[80] A. Kaintura, T. Dhaene, D. Spina, Review of polynomial chaos-based methods for uncertainty quantification in modern integrated circuits, Electronics 7 (3) (2018) 30.
[81] P. Boufounos, M. F. Duarte, R. G. Baraniuk, Sparse signal reconstruction from noisy compressive measurements using cross validation, in: Statistical Signal Processing, 2007. SSP'07. IEEE/SP 14th Workshop on, IEEE, 2007, pp. 299–303.
[82] H. N. Najm, M. Valorani, Enforcing positivity in intrusive PC-UQ methods for reactive ODE systems, J. Comput. Phys. 270 (2014) 544–569. doi:10.1016/j.jcp.2014.03.061.
[83] G. Blatman, B. Sudret, Efficient computation of global sensitivity indices using sparse polynomial chaos expansions, Reliab. Eng. Syst. Safe. 95 (11) (2010) 1216–1229.
[84] A. Saltelli, Making best use of model evaluations to compute sensitivity indices, Computer physics communications 145 (2) (2002) 280–297.