Kinetic uncertainty quantification in turbulent flame simulations: A compact approach

Yunyun Wang[a,*], Diederik Coppitters[a], Alessandro Parente[b], Francesco Contino[a]

[a] *Institute of Mechanics, Materials and Civil Engineering, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium*
[b] *Aero-Thermo-Mechanics Laboratory, Université Libre de Bruxelles, 1050 Brussels, Belgium*

Abstract

Uncertainties in the predictions brought about by inherent uncertainties in the kinetic constants need to be assessed to examine the reliability of the simulations. As the volume of the parameter space grows exponentially with dimension, any method that attempts to fully characterise the parameter space suffers from *the curse of dimensionality*. In this work, we present a compact approach to quantifying the kinetic uncertainty in turbulent flame simulations and apply it to Sandia flame D. Firstly, we prescribe the quantities of interest (QoIs) *a priori*. Then, the most influential subsets of sensitivity matrices of the QoIs are isolated. The subset isolation is found to be effective as it has identified the reactions that have significant contributions to the uncertainties given by the polynomial chaos expansion (PCE) *a posteriori*. Lastly, we propagate the uncertainties of the kinetic constants falling in the subset into the turbulent flame simulation predictions, and by constructing ordinary least-squares regression-based sparse PCEs between the predictions and these kinetic constants, uncertainty quantification is performed. The sparse PCEs have demonstrated their ability to identify the most significant region of uncertainties in the computational domain with small sample sizes while preserving lower cross-validation errors in such a region.

Keywords: Uncertainty quantification; Turbulent flame; Subset isolation; Sparse polynomial chaos; Ordinary least squares

1. Introduction

Simulating multi-physics turbulent combustion requires multiple models, such as chemical mechanisms, turbulence models, and combustion models. The multiple parameters of these models have epistemic uncertainties that must be propagated to the predictions of the simulations to assess their reliability [1]. Among the models, kinetic constants represent a very large dimensionality in the uncertain space. This paper seeks the efficient propagation of these uncertainties.

For simulations such as computational fluid dynamics (CFD) that involve high-performance computing, the uncertainty quantification (UQ) with the Monte Carlo (MC) method is almost always infeasible to obtain convergent statistics [2], since it would require several thousands of simulations. Alternatively, approximating the simulation response with a finite series, such as polynomial chaos expansion (PCE) and high-dimensional model representation (HDMR), can significantly reduce the number of simulation realisations up to two orders of magnitude. These two surrogate models have been widely applied in combustion [3]. However, as the volume of a space grows exponentially with dimension, these surrogate models (actually all methods) that attempt to fully characterise the parameter space will suffer from *the curse of dimensionality* [4]. For example, in [5], more than 2000 multi-fidelity large eddy simulations (LESs) were performed to construct the PCE for a 24-dimensional input space. To make matters worse, the scales of the kinetic models also grow rapidly (hundreds of species and thousands of elementary reactions). Despite the rapid growth of computational capability, it does not match the computational demand that comes with the increase in dimensionality. *The curse of dimensionality* is essentially a problem where the system of equations is ill-posed and thus has no unique solution due to our insufficient observations for the high-dimensional space that we are trying to fully characterise.

Nevertheless, the theory of compressed sensing [6] indicates that it is still possible to accurately reconstruct the simulation output when the computational cost results in few observations. In the context of PCE, this is due to the fact that the weights of most basis functions are close to zero [7]. In recent years, a number of stepwise regression greedy algorithms have been emerging [8], which allow the construction of sparse PCE with a very small number of observations [9]. Moreover, a large number of studies on the detailed mechanism reduction tell us that there are many reactions in a kinetic model that have tiny effects on the simulation response or on the so-called quantity of interest (QoI). This implies that the selection of the influential kinetic constants on the QoI before constructing the PCE is necessary since no one pays the computational budget for the reactions that don't have impacts on the simulation response.

In this work, we will present a compact approach to quantifying the kinetic uncertainty in turbulent flame simulations and apply it to the Sandia flame D [10]. This approach has two main parts. In Part I, we first define the QoIs *a priori*, then similar to the principal component analysis [11, 12] or the active subspace [13], the eigendecomposition is applied to isolate the most influential subset of the sensitivity matrices of QoIs. In Part II, we construct the sparse PCEs between QoIs and all kinetic constants falling in the subset to perform UQ.

2. Methods

Let $\boldsymbol{y} = \{y_1, \ldots, y_M\}^{\mathrm{T}}$ denote the QoIs in the thermochemical state space, which can include temperature, concentrations of selected species.

2.1. Subset isolation

In spatially homogeneous systems, at a given time point with given initial conditions, $\boldsymbol{y}$ can be expressed as a function of kinetic rate constants, $\boldsymbol{y}(\boldsymbol{k})\colon \mathbb{R}^Q \to \mathbb{R}^M$, where $\boldsymbol{k} = \{k_1, \ldots, k_Q\}^{\mathrm{T}}$ is the vector of rate constants for Q reactions. Let $\Delta\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{k} + \Delta\boldsymbol{k}) - \boldsymbol{y}(\boldsymbol{k})$ be the variation of $\boldsymbol{y}$ subject to a perturbation in $\boldsymbol{k}$. The Taylor series gives us

$$\boldsymbol{y}(\boldsymbol{k} + \Delta\boldsymbol{k}) \approx \boldsymbol{y}(\boldsymbol{k}) + \frac{\partial \boldsymbol{y}(\boldsymbol{k})}{\partial \boldsymbol{k}}\Delta\boldsymbol{k}, \tag{1}$$

to evaluate the magnitude of fluctuations, which can be written as

$$\sum_i^M \left(y_i\left(\boldsymbol{k} + \Delta\boldsymbol{k}\right) - y_i(\boldsymbol{k})\right)^2 \approx \sum_i^M \left(\frac{\partial y_i(\boldsymbol{k})}{\partial \boldsymbol{k}}\Delta\boldsymbol{k}\right)^2. \tag{2}$$

Note that the right-hand side above is a quadratic form, thus

$$\|\Delta\boldsymbol{y}\|_2^2 \approx (\Delta\boldsymbol{k})^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\Delta\boldsymbol{k}, \tag{3}$$

where $\mathbf{S} = \partial\boldsymbol{y}/\partial\boldsymbol{k}$ is the first-order sensitivity matrix. As the logarithmic transformation leaves parameters additive while preserving monotonicity, it can be normalised as $\tilde{\mathbf{S}} = \partial \ln \boldsymbol{y}/\partial \ln \boldsymbol{k}$.

Since Eq. (3) is positive-definite (unless zero-effect constants exist), the response, $\|\Delta\boldsymbol{y}\|_2^2$, is an elliptic paraboloid, where the most rapid variation occurs along the minor axis of the Q-dimensional elliptic section. To find those axes that characterise the fastest variations, i.e., to identify a subset of the most influential constants, the eigendecomposition can be applied. Let the close-to-zero eigenvalues be zero, then the normalised sensitivity matrix can be approximated

$$\tilde{\mathbf{S}}^{\mathrm{T}}\tilde{\mathbf{S}} \approx \mathbf{A}_q\mathbf{L}\mathbf{A}_q^{\mathrm{T}}, \tag{4}$$

such that

$$\mathrm{rank}(\mathbf{A}_q\mathbf{L}\mathbf{A}_q^{\mathrm{T}}) \ll \mathrm{rank}(\tilde{\mathbf{S}}^{\mathrm{T}}\tilde{\mathbf{S}}) = \mathrm{rank}(\tilde{\mathbf{S}}). \tag{5}$$

In $\mathbf{L}$ and $\mathbf{A}_q^{\mathrm{T}}$ are the truncated largest eigenvalues in descending order, λ_i, and their corresponding eigenvectors, $\boldsymbol{a}_{q,i}^{\mathrm{T}}$, where $i = 1, \ldots, q$, and $q \ll Q$. Similarly, if we separate the first q elements with maximum absolute values in each eigenvector, then the others can be considered as zero. A subset is isolated.

Furthermore, as described at the beginning of this section, $\tilde{\mathbf{S}}$ is observed with given initial conditions and at a given time point. We can obtain the sensitivity matrix at different time points with various initial conditions. After N observations, $\tilde{\mathbf{S}}$ can be enclosed into a larger matrix, $\mathbf{X} = \{\tilde{\mathbf{S}}^1, \ldots, \tilde{\mathbf{S}}^N\}^{\mathrm{T}} \subset \mathbb{R}^{(M\times N)\times Q}$, e.g., if we have only 1 QoI ($M = 1$), then $\mathbf{X}$ contains N observations of Q reactions. Once our collections have traversed all the time points and initial conditions we care about, $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ will provide us with comprehensive kinetic information.

Lastly, we briefly justify the effectiveness of investigating spatially homogeneous systems for turbulent combustion. The production rate of a given QoI is its time derivative, $\dot{\omega}_i = \mathrm{d}y_i/\mathrm{d}t$, and a turbulent combustion model based on finite-rate chemistry for this QoI may be represented as $\bar{\dot{\omega}}_i = \bar{\dot{\omega}}_i(\boldsymbol{u}_t, \dot{\omega}_i)$, where $\bar{\dot{\omega}}_i$ is the production rate considering turbulence-chemistry interaction in each cell, and in $\boldsymbol{u}_t$ are the variables pertaining to turbulence. Assume changes in kinetic constants do not affect the flow field (e.g., incompressible flow with constant viscosity), thus those non-influential subsets are also non-influential on the turbulent combustion model response.[1]

2.2. Regression-based sparse PCE

Let $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_D\}^{\mathrm{T}}$ denote a set of independent random variables that represent the uncertainties in the rate constants, where D is the dimension of the stochastic space. For simplicity, here we consider a single QoI, y, as the following description will hold componentwise. The PCE approximates the model response to the stochastic space by truncating at some finite order

$$y(\boldsymbol{\xi}) \approx \sum_{i=0}^{P} u_i \psi_i(\boldsymbol{\xi}), \tag{6}$$

where in $\boldsymbol{\psi} = \{\psi_0, \ldots, \psi_P\}^{\mathrm{T}}$ are the PCE basis functions and in $\boldsymbol{u} = \{u_0, \ldots, u_P\}^{\mathrm{T}}$ are their corresponding weights. According to the Askey scheme, the multivariate orthogonal polynomials that construct the basis functions are indicated by the probability distributions of $\boldsymbol{\xi}$. Let p be the maximum polynomial order, then the size of the expansion is given by $P + 1 = \frac{(p+D)!}{p!D!}$. For N realisations of the model, a sampling vector of model responses can be written as $\boldsymbol{y} = \{y^1, \ldots, y^N\}^{\mathrm{T}}$, with their corresponding stochastic inputs, $\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^N$. Considering $P+1$ terms of N samples arranged in a basis function matrix, $\boldsymbol{\Psi} = \{\boldsymbol{\psi}(\boldsymbol{\xi}^1), \ldots, \boldsymbol{\psi}(\boldsymbol{\xi}^N)\}^{\mathrm{T}} \subset \mathbb{R}^{N\times(P+1)}$, $\boldsymbol{\Psi}\hat{\boldsymbol{u}} \approx \boldsymbol{y}$ holds if such coefficient estimates ($\hat{\boldsymbol{u}}$) exist. The objective translates into an ordinary least-squares (OLS) problem, i.e., $\min_{\hat{\boldsymbol{u}}} \|\boldsymbol{\Psi}\hat{\boldsymbol{u}} - \boldsymbol{y}\|_2^2$, hence the coefficient estimates

$$\hat{\boldsymbol{u}} = (\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{y}. \tag{7}$$

For a converged solution, oversampling may be required, i.e., $N \geq P + 1$, which further increases the computational cost.

Nevertheless, the theory of compressed sensing [6] indicates that it is still possible to accurately reconstruct the model response when $N \ll P + 1$, in the context of PCE, which is essentially due to the fact that the weights of most basis functions are close to zero. In this work, we demonstrate a forward-backward stepwise regression method to construct sparse PCE, as summarised in Algorithm 1, and details can be found in the work of Abraham et al. [7]. N has become a dimension-independent input parameter based on computational resources. In the forward step, we regress $\boldsymbol{\xi}$ and $\boldsymbol{y}$ on $P + 1$ basis functions individually (line 4). From Eq. (7) we can derive the variance-covariance matrix of $\hat{\boldsymbol{u}}$ and its variance is given by

$$\mathbb{V}(\hat{\boldsymbol{u}}) = \mathrm{diag}\left((\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi})^{-1}\sigma^2\right), \tag{8}$$

where σ^2 is the variance of y and in the j^{th} individual regression it can be estimated by

$$\hat{\sigma}_j^2 = \frac{1}{N-1}\|\boldsymbol{\psi}_j\hat{u}_j - \boldsymbol{y}\|_2^2. \tag{9}$$

Now an indicator of the statistically significant basis function can be given (line 8). Once a basis function has been selected, it will be added to the sparse matrix, $\boldsymbol{\Psi}^*$, and removed from the candidate matrix, $\boldsymbol{\Psi}$. Every time $\boldsymbol{\Psi}^*$ is updated, we perform an OLS to correct its corresponding weights, $\hat{\boldsymbol{u}}^*$. In the backward step, we calculate confidence intervals (CI) of each $\hat{u}_j^*$ to let columns, i.e., basis functions in $\boldsymbol{\Psi}^*$ know whether they should leave or stay (line 14). The forward-backward step will iterate over the residual, $\hat{\boldsymbol{r}}$, until a maximum number of iterations is reached. A sparse PCE through statistical significance testing has been constructed.

Further refinements can be exploited for such an expansion. $\hat{\boldsymbol{u}}^*$ can be rearranged into increasing order with respect to the coefficient of variation of each $\hat{u}_j^*$. We can perform leave-*one*-out (LOO) cross-validation for all possible subsets along this sequence and truncate it when the LOO error reaches its minimum. The LOO error is the error between observation and prediction obtained by extrapolating the N^{th} sample from the other $N - 1$ samples. By averaging the

[1] For flow fields that could be affected by kinetic constants, temperature should be the primary pathway, such as the temperature(energy)-driven density changes in compressible flows, or the temperature-dependent viscosity. In these cases, including temperature in QoIs can be a convenient treatment.

Algorithm 1 The construction of the forward-backward stepwise regression-based sparse PCE.

Input: $\boldsymbol{\xi} \subset \mathbb{R}^{D\times N}, \boldsymbol{y} \subset \mathbb{R}^N$
Output: $\boldsymbol{\Psi}^*, \hat{\boldsymbol{u}}^*$
$\boldsymbol{\Psi}(\boldsymbol{\xi}) = \{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{P+1}\}, \boldsymbol{\Psi}^* \leftarrow 0, \hat{\boldsymbol{u}}^* \leftarrow 0, \hat{\boldsymbol{r}} \leftarrow \boldsymbol{y}$
while $i = 1, 2, \ldots, f(N)$ **do**
  while $j = 1, 2, \ldots,$ columns$(\boldsymbol{\Psi})$ **do** ▷ forward
    $\hat{u}_j = (\boldsymbol{\psi}_j^{\mathrm{T}}\boldsymbol{\psi}_j)^{-1}\boldsymbol{\psi}_j^{\mathrm{T}}\hat{\boldsymbol{r}}$
    $\hat{\boldsymbol{u}} \leftarrow \hat{\boldsymbol{u}} \cup \{\hat{u}_j\}$
    $\mathbb{V}(\hat{\boldsymbol{u}}) \leftarrow \mathbb{V}(\hat{\boldsymbol{u}}) \cup \{\mathbb{V}(\hat{u}_j)\}$
  end while
  $j^* = \arg\max_j\{|\hat{u}_j| \cdot \mathbb{V}(\hat{u}_j)^{-1/2}\}$
  $\boldsymbol{\Psi}^* \leftarrow \boldsymbol{\Psi}^* \cup \{\boldsymbol{\psi}_{j^*}\}$
  $\boldsymbol{\Psi} \leftarrow \boldsymbol{\Psi} \setminus \{\boldsymbol{\psi}_{j^*}\}$
  $\hat{\boldsymbol{u}}^* = (\boldsymbol{\Psi}^{*\mathrm{T}}\boldsymbol{\Psi}^*)^{-1}\boldsymbol{\Psi}^{*\mathrm{T}}\hat{\boldsymbol{r}}$ ▷ OLS
  while $j = 1, 2, \ldots,$ columns$(\boldsymbol{\Psi}^*)$ **do** ▷ backward
    if CI of $\hat{u}_j^*$ includes 0 **then**
      $\boldsymbol{\Psi}^* \leftarrow \boldsymbol{\Psi}^* \setminus \{\boldsymbol{\psi}_j^*\}$
    end if
  end while
  $\hat{\boldsymbol{u}}^* = (\boldsymbol{\Psi}^{*\mathrm{T}}\boldsymbol{\Psi}^*)^{-1}\boldsymbol{\Psi}^{*\mathrm{T}}\hat{\boldsymbol{r}}$ ▷ OLS
  $\hat{\boldsymbol{r}} \leftarrow \hat{\boldsymbol{r}} - \boldsymbol{\Psi}^*\hat{\boldsymbol{u}}^*$
end while

sum of squares of such errors for all samples, it can be derived as

$$E_{\mathrm{LOO}} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{y(\boldsymbol{\xi}^i) - \boldsymbol{\psi}(\boldsymbol{\xi}^i)\hat{\boldsymbol{u}}}{1-h_i}\right)^2, \qquad (10)$$

where h_i is the i^{th} element on the diagonal of the hat matrix, $\mathbf{H} = \boldsymbol{\Psi}(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^{\mathrm{T}}$, of $\boldsymbol{y}$. Similar to the determination coefficient, LOO error can be normalised by the variance of the model response

$$e_{\mathrm{LOO}} = \frac{E_{\mathrm{LOO}}}{\mathbb{V}(y) + \epsilon}, \qquad (11)$$

where in practice we add a very small value, ϵ (e.g., 10^{-30}), to avoid dividing by 0.

3. Application

3.1. Identification of influential subsets

In UQ of CFD combustion, a typical approach can be first to find uncertainties of interest, and then to explore how QoIs are affected by these parameters. In this approach, a QoI or a combination of QoIs is chosen *a priori*, which will bring about different subsets of influential constants. In flame D, we have chosen temperature and the mass fraction of two pollutants with different chemical timescales, $\boldsymbol{y} = \{T, Y_{\mathrm{CO}}, Y_{\mathrm{NO}}\}^{\mathrm{T}}$.

We calculated the sensitivity matrices using the GRI-Mech 3.0 scheme [14] with Cantera©. To make the time scale span the most intense variations up to the quasi-steady state, 36 time points have been selected between 10^{-5} to 0.005 seconds as observations. A consistent initial temperature, T_{initial}= 1880 K, is maintained since this is the pilot flame temperature in flame D. Flame D ($Re = 22400$) has a small degree of local extinction [15], we therefore selected 64 observations in the range of $\phi \in (0, 3.2]$ by taking the equivalence ratio of the main jet as an upper bound ($\phi = 3.17$). Thus, 6912 observations of 325 reactions for three QoIs were enclosed in $\mathbf{X}$.

Table 1: A most influential subset ($q = 7$) of the reactions for the given QoIs, $\boldsymbol{y} = \{T, Y_{\mathrm{CO}}, Y_{\mathrm{NO}}\}^{\mathrm{T}}$, $\phi \in (0, 3.2]$, T_{initial}= 1880 K.

#	Reaction	In which $\boldsymbol{a}_{q,i}^{\mathrm{T}}$	Retain/Drop	U_j
1. R51	H+CH3(+M)⇔CH4(+M)	1, 2, 3, 4, 6, 7	retain	2.0
2. R37	H+O2⇔O+OH	1, 2, 3, 4, 6	retain	1.2
3. R154	CH3+O2⇔O+CH3O	1, 3, 6	retain	3.0
4. R52	H+CH4⇔CH3+H2	1, 6, 7	retain	2.0
5. R158	2CH3⇔H+C2H5	1, 6	retain	5.0
6. R155	CH3+O2⇔OH+CH2O	1, 7	retain	3.0
7. R118	HO2+CH3⇔OH+CH3O	1	retain	10.0
8. R177	N+NO⇔N2+O	2, 3, 4, 5, 7	retain	3.0
9. R10	O+CH4⇔OH+CH3	2, 4, 5, 6, 7	retain	1.5
10. R167	HCO+O2⇔HO2+CO	2, 3, 5	retain	2.0
11. R56	H+CH2O(+M)⇔CH3O(+M)	2, 4, 5	retain	2.0
12. R169	CH3O+O2⇔HO2+CH2O	2, 4, 5	drop	-
13. R239	CH+N2⇔HCN+N	3, 4, 5	retain	3.0
14. R274	CH3+N⇔H2CN+H	3	drop	-
15. R166	HCO+M⇔H+CO+M	5	drop	-
16. R125	CH+H2⇔H+CH2	6	drop	-
17. R98	OH+CO⇔H+CO2	7	drop	-
18. R160	CH3+CH2O⇔HCO+CH4	7	drop	-

The dimensions of the subset, q^2, are arbitrarily specified based on the case of interest, and in this application, we present the case of $q = 7$, as shown in Table 1. A q of 7 indicates that a total of $7^2 = 49$ entries will appear in the subset, some of which belong to the same reaction. We ranked the reactions that fall in the subset according to the importance of the eigenvector ($\boldsymbol{a}_{q,i}^{\mathrm{T}}$) in which they appeared and the frequency. It can be seen that reactions that are recognised as important, such as R37, are captured. Such a subset of $\mathbf{X}$ can be further isolated since it has become tractable. First we removed the reactions that appeared singly in the less significant eigenvectors, i.e., R160, R98, R125, R166, and R274. We then noted that together with R177, the entry magnitudes of R239 dominate the 3rd-5th eigenvectors, namely, R177: -0.603, 0.423, 0.280 and R239: -0.544, -0.460, -0.516, thus we have retained it. Lastly, we arbitrarily dropped one of R56 and R169, say, R169, to maintain a 12-dimensional input space for the next procedure. The final subset has been selected.

Once a subset is identified, its statistical definition must be specified. The kinetic constants are presumed to be independent, and the rate constant of an individual reaction is characterized by a lognormal distribution with a nominal value, k_j^0, and a temperature-independent uncertainty factor U_j, which reads [16, 17]

$$\ln k_j \sim \mathcal{N}\left(\ln k_j^0, (\frac{1}{3}\ln U_j)^2\right). \qquad (12)$$

U_j's can be found in [17, 18], and are listed in Table 1.

3.2. Uncertainty quantification of flame D

We briefly describe the models for the flame D simulations. The simulations were carried out using OpenFOAM® with various corresponding numer-

ical schemes. A k-epsilon Reynolds-averaged turbulence model was chosen to calculate the flow field, which was assumed to be 2-dimensional symmetric for flame D. A Cartesian structured grid was used for spatial discretisation, and after a grid convergence test, a final computational domain with ∼24000 cells was prescribed. For the kinetic model, we used the same GRI-Mech 3.0 scheme as in 3.1. The SEULEX ODE solver [19] in conjunction with the tabulation of dynamic adaptive chemistry [20] were chosen to solve the kinetic mechanism. The kinetic constants following the distribution of (12) in the subset were characterised into the uncertain inputs by the quasi-random Sobol' sequences [21]. The eddy dissipation concept [22] combustion model was used to take turbulence-chemical interactions into account. The main computational domain is in $xz \in [0, 0.15] \times [0, 0.6]$ m^2, where $x = 0$ is the symmetry axis and z is the streamwise direction.

3.2.1. Construction of sparse PCEs

Recalling Algorithm 1 and the cross-validation truncation process, the upper bound of the number of basis functions for a sparse PCE is given by the number of samples, e.g., let $f(N) = N$ in line 2. That is, for any size of samples, the sparse PCE can reconstruct the model responses well, i.e., small LOO error. However, with tiny sample sizes, the statistical moments given by PCE are questioned, even though the cross-validation error is small. Therefore, the convergence of the statistical moments should be tested first.

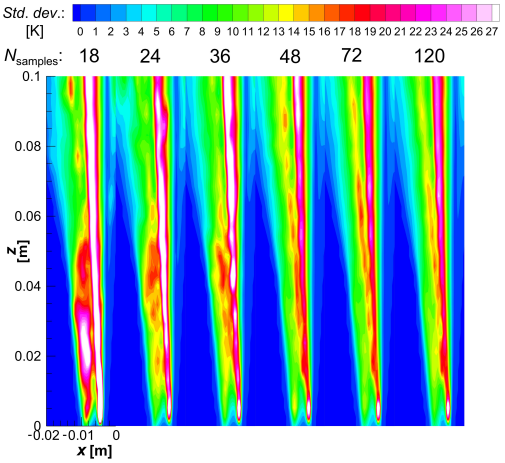

Fig. 1: Standard deviation fields of temperature in flame D given by the sparse PCEs at different sample sizes. In the vicinity of the flame D inlet ($xz \in [-0.02, 0] \times [0, 0.1]$ m^2).

We have constructed sparse PCEs for QoIs of all cells in the computational domain. The calculations of sparse PCEs were performed based on the Python package RHEIA© [23]. At small sample sizes, i.e., 18, 24, and 36 samples, we have found spatial inconsistencies in the standard deviation of temperatures on the co-flow air side at the inlet of flame D, as shown in Fig. 1. The temperature standard deviation field is given by the 4$^{\text{th}}$ order sparse PCEs. The spatial inconsistency here refers to the inconsistency of regions spatially delineated by cells with high standard deviations. In the computational domain outside the region illustrated in Fig. 1, the temperature standard deviation field is almost spatially consistent for all sample sizes, which means that sparse PCEs can give us an asymptotically qualitative result even with a very small sample size. This is not possible with full PCE, for example, the 2$^{\text{nd}}$ order full PCE theoretically requires at least $\binom{12+2}{2} = 91$ samples to obtain a solution for the coefficients for the 12-dimensional uncertainty input space in this work. Assuming that spatial consistency of high standard deviation cells identified by sparse PCEs represents convergence, in this application, 48 samples are sufficient to delineate the region in the computational domain where the uncertainties are most significant (Fig. 1).

It was also found that the region identified as having the most significant uncertainties tended to have smaller cross-validation errors. We can truncate the standard deviation at different LOO errors, i.e., if the e_{LOO} in Eq.(11) of a cell is greater than a given threshold, then let its standard deviation be 0. Fig. 2 shows such truncated standard deviation fields for 48 samples with thresholds of 0.02 and 0.01, respectively. It is clear that the most significant region of uncertainties has not been erased when the error is tiny. Similar behaviours were observed at other sample sizes.

Note that this most significant region of temperature uncertainties identified by the sparse PCEs is not the region with the highest temperature but rather the one just before this region, i.e., where the main jet contacts the pilot flame and begins to burn (Fig. 2). In contrast, the region with the highest temperature shows lower uncertainties relative to its surroundings. But this region is not well captured by the smaller e_{LOO}, probably due to its lower uncertainties.

3.2.2. Global sensitivity analysis

Table 2: Number of occurrences of the reactions with the highest total-order Sobol' indices for T in all cells with non-zero standard deviation. $N_{\text{samples}} = 48$.

#	Occurrences	#	Occurrences
1. R37	1892	7. R158	736
2. R10	1678	8. R56	602
3. R118	1628	9. R155	535
4. R52	1592	10. R154	384
5. R51	1357	11. R167	371
6. R239	1226	12. R177	203

Henceforth, a global sensitivity analysis among the input space can be performed. In the computational domain, we can count the number of occurrences of

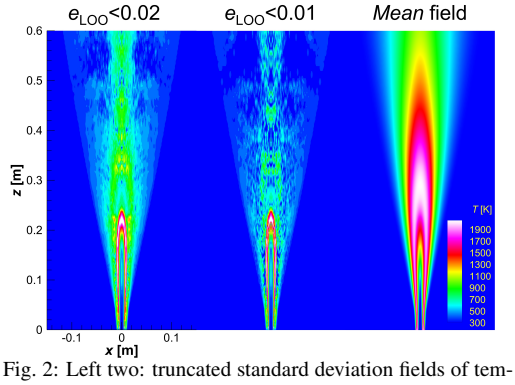

Fig. 2: Left two: truncated standard deviation fields of temperature by different LOO errors. $N_{\text{samples}}= 48$. The same legend as in Fig. 1. Similar behaviours were observed at other sample sizes. Right: the mean field of the temperature.

Table 3: Number of occurrences of the reactions with the highest total-order Sobol' indices for Y_{NO} in all cells with non-zero standard deviation. $N_{\text{samples}}= 48$.

#	Occurrences
1. R239	10731
2. R177	427

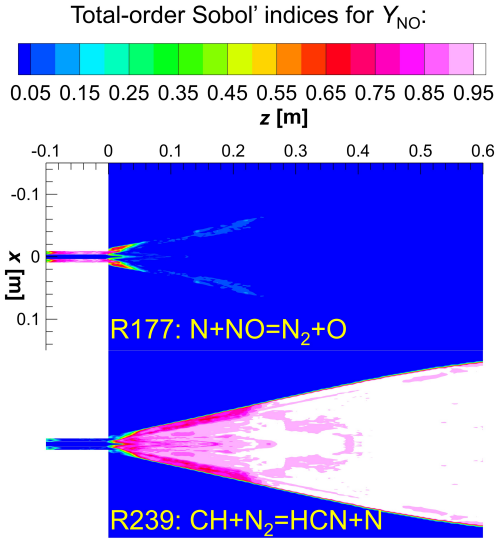

Fig. 3: The total-order Sobol' index fields for Y_{NO} of R177 and R239. $N_{\text{samples}}= 48$. Similar results were observed at other sample sizes.

the reaction with the highest Sobol' index given by the sparse PCEs in the cells whose standard deviations are not zero, i.e. cells where the Sobol' indices exist. For temperature and CO, all reactions in the subset have played their roles of having the highest total-order Sobol' indices in the corresponding cells, which makes it difficult to provide a clear interpretation of this count, e.g., see Table 2. This might be due to the fact that we have chosen a combination of three QoIs with quite distinct properties to identify the subset, making it problematic to target a particular QoI for analysis.

However, for the NO chemistry as a relatively independent module (which can be cut from GRI-Mech), we found that only two reactions dominated the entire computational domain, as listed in Table 3, namely R239: CH+N_2⇔HCN+N and R177: N+NO⇔N_2+O. It is interesting to note that we happen to have retained R239 in 3.1, which confirms the effectiveness of checking the entries in the significant eigenvectors in the subset isolation procedure. Hence, we explored the total-order Sobol' index fields of R177 and R239 regarding the mass fraction of NO, as shown in Fig. 3. The figure clearly demonstrates the transition of the dominant reaction from R177, which dominates the deviations of Y_{NO} in pure pilot flame, to R239 after the pilot flame contact with the main jet. Note that R177 and R239 have an identical uncertainty factor of 3.0 (see Table 1). This simulation behaviour is interesting, as it seems that the effect of uncertainties of the Zeldovich mechanism initially dominates the variations of NO in the pilot flame and is then rapidly replaced by the uncertainties of the Fenimore mechanism. Further research may be required.

4. Conclusions

In this work, we have presented a compact approach to quantifying the kinetic uncertainty in turbulent flame simulations and applied it to flame D. This approach is applicable to arbitrary kinetic models and arbitrary CFD combustion configurations. The main findings are as follows:

- The construction of sparse polynomial chaos expansion (PCE) remarkably alleviates *the curse of dimensionality* in turbulent flame simulations. It has the ability to delineate the region where the uncertainties are most significant with a smaller sample size while preserving a lower cross-validation error in such a region. For extremely small sample sizes, the sparse PCE can still give an asymptotic result.
- The quantities of interest (QoIs)-based subset isolation provides this approach with a flexible selection strategy. We can perform self-validation based on the global sensitivity analysis given by the sparse PCE *a posteriori*. In this application it was found to be effective: it has identified the reactions that contribute significantly to the uncertainties of NO mass fraction.

Finally, we would like to propose two aspects that may merit further exploration.

- In addition to the algorithms in this work, a number of stepwise regression greedy algorithms for constructing sparse PCE have been emerging, such as orthogonal matching pursuit,

least angle regression, etc. [8]. It may be an interesting topic to examine the performances of such algorithms or to develop novel tailored algorithms for numerical combustion.

- The sensitivity matrix is the Jacobian of the QoIs to the kinetic constants, and if the calculation of the system of the Hessian of the QoIs (weighted by uncertainty factor) is possible to be implemented, i.e., a further expansion of Eq. 1, then the subset isolation perhaps can provide the following uncertainty quantification procedure with a more accurate selection scenario.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

D. Coppitters acknowledges the support of the Fonds de la Recherche Scientifique - FNRS [CR 40016260].

References

[1] Some Limitations of Turbomachinery CFD, Vol. Volume 7: Turbomachinery, Parts A, B, and C of Turbo Expo: Power for Land, Sea, and Air. doi:10.1115/GT2010-22540.

[2] R. K. Tripathy, I. Bilionis, Deep uq: Learning deep neural network surrogate models for high dimensional uncertainty quantification, Journal of Computational Physics 375 (2018) 565–588. doi:https://doi.org/10.1016/j.jcp.2018.08.036.

[3] H. Wang, D. A. Sheen, Combustion kinetic model uncertainty quantification, propagation and minimization, Progress in Energy and Combustion Science 47 (2015) 1–31. doi:https://doi.org/10.1016/j.pecs.2014.10.002.

[4] D. L. Donoho, et al., High-dimensional data analysis: The curses and blessings of dimensionality, AMS math challenges lecture 1 (2000) (2000) 32.

[5] X. Huan, C. Safta, K. Sargsyan, G. Geraci, M. S. Eldred, Z. P. Vane, G. Lacaze, J. C. Oefelein, H. N. Najm, Global sensitivity analysis and estimation of model error, toward uncertainty quantification in scramjet computations, AIAA Journal 56 (3) (2018) 1170–1184. doi:10.2514/1.J056278.

[6] E. Candes, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Transactions on Information Theory 52 (2) (2006) 489–509. doi:10.1109/TIT.2005.862083.

[7] S. Abraham, M. Raisee, G. Ghorbaniasl, F. Contino, C. Lacor, A robust and efficient stepwise regression method for building sparse polynomial chaos expansions, Journal of Computational Physics 332 (2017) 461–474. doi:https://doi.org/10.1016/j.jcp.2016.12.015.

[8] N. Lüthen, S. Marelli, B. Sudret, Sparse polynomial chaos expansions: Literature survey and benchmark, SIAM/ASA Journal on Uncertainty Quantification 9 (2) (2021) 593–649. doi:10.1137/20M1315774.

[9] R. Baptista, V. Stolbunov, P. B. Nair, Some greedy algorithms for sparse polynomial chaos expansions, Journal of Computational Physics 387 (2019) 303–325. doi:https://doi.org/10.1016/j.jcp.2019.01.035.

[10] R. Barlow, J. Frank, A. Karpetis, J.-Y. Chen, Piloted methane/air jet flames: Transport effects and aspects of scalar structure, Combustion and Flame 143 (4) (2005) 433–449, special Issue to Honor Professor Robert W. Bilger on the Occasion of His Seventieth Birthday. doi:https://doi.org/10.1016/j.combustflame.2005.08.017.

[11] A. Parente, J. Sutherland, L. Tognotti, P. Smith, Identification of low-dimensional manifolds in turbulent flames, Proceedings of the Combustion Institute 32 (1) (2009) 1579–1586. doi:https://doi.org/10.1016/j.proci.2008.06.177.

[12] S. Vajda, P. Valko, T. Turányi, Principal component analysis of kinetic models, International Journal of Chemical Kinetics 17 (1) (1985) 55–81. doi:https://doi.org/10.1002/kin.550170107.

[13] P. G. Constantine, E. Dow, Q. Wang, Active subspace methods in theory and practice: Applications to kriging surfaces, SIAM Journal on Scientific Computing 36 (4) (2014) A1500–A1524. doi:10.1137/130916138.

[14] G. P. Smith, D. M. Golden, M. Frenklach, N. W. Moriarty, B. Eiteneer, M. Goldenberg, C. T. Bowman, R. K. Hanson, S. Song, W. C. Gardiner, Jr., V. V. Lissianski, Z. Qin, GRI-Mech 3.0, Available at http://www.me.berkeley.edu/gri_mech/, accessed: March 25, 2025(1999).

[15] R. Barlow, J. Frank, Effects of turbulence on species mass fractions in methane/air jet flames, Symposium (International) on Combustion 27 (1) (1998) 1087–1095, twenty-Seventh Sysposium (International) on Combustion Volume One. doi:https://doi.org/10.1016/S0082-0784(98)80510-9.

[16] M. E. Mueller, G. Iaccarino, H. Pitsch, Chemical kinetic uncertainty quantification for large eddy simulation of turbulent nonpremixed combustion, Proceedings of the Combustion Institute 34 (1) (2013) 1299–1306. doi:https://doi.org/10.1016/j.proci.2012.07.054.

[17] D. A. Sheen, X. You, H. Wang, T. Løvås, Spectral uncertainty quantification, propagation and optimization of a detailed kinetic model for ethylene combustion, Proceedings of the Combustion Institute 32 (1) (2009) 535–542. doi:https://doi.org/10.1016/j.proci.2008.05.042.

[18] D. L. Baulch, C. T. Bowman, C. J. Cobos, R. A. Cox, T. Just, J. A. Kerr, M. J. Pilling, D. Stocker, J. Troe, W. Tsang, R. W. Walker, J. Warnatz, Evaluated kinetic data for combustion modeling: Supplement ii, Journal of Physical and Chemical Reference Data 34 (3) (2005) 757–1397. doi:10.1063/1.1748524.

[19] E. Hairer, G. Wanner, Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems, Vol. 14, 1996. doi:10.1007/978-3-662-09947-6.

[20] F. Contino, H. Jeanmart, T. Lucchini, G. D'Errico,

Coupling of in situ adaptive tabulation and dynamic adaptive chemistry: An effective method for solving combustion in engine simulations, Proceedings of the Combustion Institute 33 (2) (2011) 3057–3064. doi:https://doi.org/10.1016/j.proci.2010.08.002.

[21] I. Sobol', On the distribution of points in a cube and the approximate evaluation of integrals, USSR Computational Mathematics and Mathematical Physics 7 (4) (1967) 86–112. doi:https://doi.org/10.1016/0041-5553(67)90144-9.

[22] B. F. Magnussen, The eddy dissipation concept a bridge between science and technology, 2005.
URL https://api.semanticscholar.org/CorpusID:67818474

[23] D. Coppitters, P. Tsirikoglou, W. D. Paepe, K. Kyprianidis, A. Kalfas, F. Contino, RHEIA: Robust design optimization of renewable Hydrogen and dErIved energy cArrier systems, Journal of Open Source Software 7 (75) (2022) 4370. doi:10.21105/joss.04370.