A co-kurtosis based dimensionality reduction method for combustion datasets

Anirudh Jonnalagadda\textsuperscript{a}, Shubham P. Kulkarni\textsuperscript{a}, Akash Rodhiya\textsuperscript{a}, Hemanth Kolla\textsuperscript{b}, Konduri Aditya\textsuperscript{a,}\textsuperscript{*}

\textit{\textsuperscript{a}Department of Computational and Data Sciences, Indian Institute of Science, Bangalore, India}

\textit{\textsuperscript{b}Sandia National Laboratories, Livermore, California, USA}

\textbf{Abstract}

Principal Component Analysis (PCA) is a dimensionality reduction technique widely used to reduce the computational cost associated with numerical simulations of combustion phenomena. However, PCA, which transforms the thermo-chemical state space based on eigenvectors of co-variance of the data, could fail to capture information regarding important localized chemical dynamics, such as the formation of ignition kernels, appearing as extreme-valued samples in a dataset. In this paper, we propose an alternate dimensionality reduction procedure, co-kurtosis PCA (CoK-PCA), wherein the required principal vectors are computed from a high-order joint statistical moment, namely the co-kurtosis tensor, which may better identify directions in the state space that represent stiff dynamics. We first demonstrate the potential of the proposed CoK-PCA method using a synthetically generated dataset that is representative of typical combustion simulations. Thereafter, we characterize and contrast the accuracy of CoK-PCA against PCA for datasets representing spontaneous ignition of premixed ethylene-air in a simple homogeneous reactor and ethanol-fueled homogeneous charged compression ignition (HCCI) engine. Specifically, we compare the low-dimensional manifolds in terms of reconstruction errors of the original thermo-chemical state, and species production and heat release rates computed from the reconstructed state. The latter – a comparison of species production and heat release rates – is a more rigorous assessment of the accuracy of dimensionality reduction. We find that, even using a simplistic linear reconstruction, the co-kurtosis based reduced manifold represents the original thermo-chemical state more accurately than PCA, especially in the regions where chemical reactions are important.

\textbf{Keywords:} Dimensionality reduction, Principal component analysis, Co-kurtosis tensor, Independent component analysis

1. \textbf{Introduction}

Direct numerical simulations (DNS) of turbulent reacting flows, which capture detailed chemical kinetics and their interactions with turbulent structures, demand massive computations and are often performed on large supercomputers. Current state-of-the-art simulations use chemical mechanisms that comprise of tens of species participating in tens or hundreds of reactions \cite{1, 2, 3, 4, 5}. A valuable approach in reducing the associated computational expense involves identifying lower-dimensional manifolds, i.e., a smaller subset of representative thermo-chemical scalars to express, and solve the conservation equations. Indeed, based on the modelling strategy being employed, several types of lower-dimensional manifolds can be obtained \cite{6}.
The need to reduce dimensionality has long been held imperative for computations of chemically reacting systems. Approaches based on exploiting the dynamics of reacting systems have preceded the more recent data-driven approaches. Intrinsic low-dimensional manifolds (ILDM) [7] recognizes the broad range of chemical timescales, characteristic of most reacting systems, and identifies an intrinsic lower-dimensional subspace by decoupling the fastest time scales (equivalent to a local equilibrium assumption) from the slower ones. Eigen decomposition of the chemical Jacobian allows identifying the eigenvalues corresponding to the fastest time scales, and truncating the corresponding eigenvectors basis defines the subspace. Computational Singular Perturbation (CSP) [8] also exploits the separation of chemical timescales into fast and slow ones, identified from the eigen decomposition of the Jacobian, but primarily to reduce the stiffness of the ODE system rather than to reduce its dimensionality. To account for diffusion and molecular transport effects in addition to chemical reactions, methods such as flame prolongation of ILDM (FPI) [9] and flamelet-generated manifolds (FGM) [10] use one-dimensional laminar flame simulations parameterized by a few controlling variables that describe the full chemical system over a broader range of conditions.

Among the data-driven approaches, empirical manifolds presented through the application of principal component analysis (PCA) on combustion datasets has received a wide interest, as it provides a rigorous procedure to identify the low-dimensional subspace that consists of orthonormal basis vectors. Several flavours of using PCA for dimensionality reduction are reported in the literature. Sutherland and Parente [11] analysed the application of PCA conditioned on the mixture fraction for DNS datasets, whereas Biglari and Sutherland [12] considered the application of unconditional PCA with a focus on scaling effects, large eddy simulation (LES) filtering and applicability of multivariate adaptive regression splines (MARS) for the forward and reverse projections. Thereafter, Yang et al. [13] applied the analysis presented by Biglari and Sutherland [12] to high-fidelity DNS datasets with more complex chemical mechanisms, used a more sensitive $R^2$ error metric to evaluate the quality of reconstruction, and presented transport equations for the chosen reduced manifolds for future use. Echekki and Mirgolbabaei [14] used PCA to tabulate relevant quantities in terms of reduced principal vectors using artificial neural networks (ANNs) which were previously shown to present a better reconstruction as compared to the linear reconstruction [15]. Ranade and Echekki [16] used PCA to isolate lower dimensional manifolds to develop closure models for turbulent combustion with the reconstruction being performed using a modified version of the pairwise-mixed stirred reactor model. Malik et al. [17] applied the PC-transport approach to 3D LES simulations and used Gaussian Process Regression (GPR) for reconstructing the missing species; the use of GPR reconstruction was motivated by a comparative study performed by Isaac et al. [18].

Apart from the global application of PCA described above, researchers have also applied PCA locally in space/time to capture the non-linear nature of combustion data with greater accuracy. Particularly, Parente et al. [19, 20] compared and contrasted the effectiveness of two local PCA methods, namely vector quantization PCA (VQPCA) and mixture fraction PCA (FPCA) against global PCA. Coussement et al. [21] conducted investigations for representing the non-linear data better by examining the use of a kernel density method based weighted PCA technique coupled with linear reconstruction. Thereafter, Mirgolbabaei and Echekki [15] studied the characteristics of an ANN based reconstruction method for the kernel density weighted PCA.

It is clear that while several methods to transform the thermo-chemical scalars to and from low-dimensional representations have been investigated in the literature, the reduced manifold has been predominantly identified
using PCA. In this work, we investigate the suitability of a co-kurtosis measure based dimensionality reduction approach (called CoK-PCA for brevity) for reduced-order modelling of combustion systems and examine its accuracy, compared to conventional PCA, in representing the thermo-chemical state as well as the chemical kinetics of the system. In particular, we present a detailed and comprehensive analysis by considering reconstruction errors of not only the species mass fractions and temperature but also those of the species production rates and heat release rate. The latter, an analysis of the reconstruction errors of species production and heat release rates, is a more rigorous assessment of accuracy and goes beyond existing PCA literature which typically only report species reconstruction errors. Further, we also compare and contrast the robustness of the PCA and CoK-PCA reduced manifolds in modeling unknown test conditions in the neighbourhood of the original training condition.

The rationale behind our line of investigation stems from the observation that, since PCA principal vectors are not very sensitive to the presence of extreme events in a dataset [22], empirically reduced manifolds constructed through PCA may not be able to capture stiff chemical dynamics associated with localized extreme events such as the formation of ignition kernels. In contrast, since the identification of extreme-valued states (‘anomalous’ events) significantly improves upon using principal vectors from the fourth-order co-kurtosis statistic [22], it can be posited that CoK-PCA reduced manifolds may present improved lower dimensional representations of the local chemical state of the system. We note that the present work focuses on the manner in which the principal vectors are identified and uses the simplest form of reconstruction; an assessment on the performance of the proposed co-kurtosis based dimensionality reduction with non-linear reconstruction methods, which are shown to preserve accuracy better, will be the subject of a future study.

An outline for the remainder of this paper is as follows. In Sec. 2, we first briefly discuss the specifics of PCA and then describe the calculation of the fourth-order co-kurtosis tensor along with its decomposition to yield the co-kurtosis based principal vectors. In Sec. 3, we first demonstrate a proof-of-concept evaluation of the CoK-PCA for a synthetically generated dataset that is representative of combustion datasets. Thereafter, in Sec. 4 we present results relevant to combustion datasets and discuss the merits and demerits of the proposed CoK-PCA relative to the standard PCA. Finally, conclusions from this work are presented in Sec. 5.

2. Fourth-order moment co-kurtosis tensor and decomposition

Consider a dataset \( \mathbf{X} \in \mathbb{R}^{(n_g \times n_v)} \) with \( n_g \) observations each having \( n_v \) real-valued variables or features; in terms of the feature space, \( \mathbf{X} \) can be represented in terms of column vectors as \( \mathbf{X} = \{ x_i \in \mathbb{R}^{(n_g \times 1)} \ \forall \ i \in \{1, \cdots, n_v\} \} \).

For PCA, the principal vectors represent directions of maximal data variance as captured by the second order co-variance matrix \( \mathbf{C} \in \mathbb{R}^{(n_v \times n_v)} \) which is represented using the index notation as:

\[
(\mathbf{C})_{ij} \equiv C_{ij} = \mathbb{E}(x_i x_j), \quad i, j \in \{1, \cdots, n_v\},
\]

(1)

where \( \mathbb{E} \) is the expectation operator. The required principal vectors \( (\mathbf{A}) \) are the eigenvectors of the co-variance matrix obtained through an eigenvalue decomposition, \( \mathbf{C} = \mathbf{A} \mathbf{L} \mathbf{A}^T \). It should be noted that the data used in the definition of joint moments is assumed to be centered around the mean.

Similarly, with the higher order moment of interest, i.e., the fourth-order co-kurtosis tensor, the principal vectors represent the directions of maximal kurtosis or peakiness in the data. The co-kurtosis tensor is defined as:

\[
T_{ijkl} = \mathbb{E}(x_i x_j x_k x_l), \quad i, j, k, l \in \{1, \cdots, n_v\}.
\]

(2)
As alluded to previously, Aditya et al. [22] demonstrated that the principal vectors of the co-kurtosis tensor can be effective in identifying directions of rare or anomalous or extreme-valued samples. The required principal vectors were obtained by decomposing the fourth-order cumulant tensor, which is related to the co-kurtosis tensor by means of an analogy derived from independent component analysis (ICA) which is detailed next.

ICA assumes that the observed features (X) are a linear combination of statistically independent non-Gaussian sources (S). Mathematically, the feature space can be represented as:

$$X = SM^T + N,$$

(3)

where M is the mixing matrix and N is an additive Gaussian noise. The vectors of the mixing matrix M can be computed by realizing that they maximize the non-Gaussian nature of X [23]. Since excess kurtosis is a good measure of the non-Gaussian nature of a dataset, the fourth-order cumulant tensor, K, of X can be factorized in terms of the excess kurtosis of i-th source, $\kappa_s_i$, and the vectors of M as [24, 25]:

$$K = \sum_i \kappa_s_i m_i \otimes m_i \otimes m_i \otimes m_i,$$

(4)

where the cumulant tensor is defined as:

$$K_{ijkl} = T_{ijkl} - C_{ij}C_{kl} - C_{ik}C_{jl} - C_{il}C_{jk}.$$  

(5)

Again, it should be noted that the above equation holds good when the data is centered around the mean.

The mathematical statement of ICA, Eq. 3, is very similar to PCA in that it prescribes a linear transformation from the original features to a transformed basis. However, while PCA yields a set of linearly uncorrelated features, ICA yields statistically independent features which is particularly attractive for turbulent reacting flow simulations, especially when the conservation equations are solved for the transformed variables in the context of large-eddy simulations (LES) or Reynolds-averaged Navier-Stokes (RANS) simulations wherein closure models of the joint PDFs of the variables are required to account for sub-grid terms [12].

To obtain the required CoK-PCA principal vectors, it is necessary to suitably decompose the fourth-order cumulant tensor K. However, as all matrix factorization properties cannot be generalized to higher order tensors, alternate decomposition strategies such as the symmetric canonical polyadic (CP), higher order singular value decomposition (HOSVD), etc. are needed to obtain the required principal vectors and values. Following [25], Aditya et al. [22] recast the fourth-order cumulant tensor as an $n_v \times n_v^3$ matrix (T), which is then decomposed using SVD as $T = USV^T$; the required principal vectors are obtained as the columns of the left singular matrix, i.e., $M = U$. Note that, thus computed, the vectors of M will be ortho-normal to each other, which is an additional constraint not central to ICA.

3. Dimensionality reduction and data reconstruction

In a dimensionality reduction exercise, the most relevant information is retained by choosing the most informative subset of principal vectors that characterize the data, thus representing $X \in \mathbb{R}^{(n_g \times n_v)}$ in a reduced space as $Z_q \in \mathbb{R}^{(n_q \times n_v)}$, with $n_q (< n_v)$ being the number of retained principal vectors. A simple technique to obtain $Z_q$ utilizes the linear transformation:

$$Z_q = XA_q,$$

(6)
where $A_q \in \mathbb{R}^{(n_\nu \times n_q)}$ comprises the chosen subset of principal vectors. The similarities and differences between PCA and CoK-PCA stem from the choice of these principal vectors. For PCA, $A_q$ correspond to the $n_q$ leading singular vectors of the raw data $X$, which are the same as the first $n_q$ eigenvectors of its co-variance matrix $C$. On the other hand, for CoK-PCA the principal vectors are obtained from the HOSVD of the fourth-order cumulant tensor $K$ (Eq. 5), i.e. matricize $K$ to obtain $T \in \mathbb{R}^{(n_\nu \times n_3)}$ and compute $A_q$ as the $n_q$ leading singular vectors of $T$. Because of the connections between the principal vectors of $K$ and ICA noted in the previous section, CoK-PCA is similar to ICA rather than PCA.

To illustrate the dimensionality reduction procedure and demonstrate the potential of the CoK-PCA method, consider a synthetic dataset, $D \in \mathbb{R}^{(625 \times 2)}$, which is constructed by rotating an uncorrelated bi-variate Gaussian distribution by $45^\circ$. The two variables, $v_1$ and $v_2$, in the dataset have a zero mean and a variance of 1.0 and 0.25, respectively. A small, randomly chosen subset of 15 samples are isolated from the main cluster and placed in the neighbourhood of point $(-1,0)$, which is farther away from the normal data, to represent extreme-valued points that characterize an anomalous event, as shown in Fig. 1. To proceed with dimensionality reduction, first the data matrix $D$ is scaled following the procedure described in [22] to obtain the scaled data matrix $X$; specifically, for each feature, the respective mean is subtracted followed by a normalization with the absolute maximum of the feature. Thereafter, the PCA and CoK-PCA principal vectors are computed according to the procedure described in Sec. 2, and are plotted in Fig. 1. It can be clearly seen that while the first principal vector from PCA (solid-red line) aligns in the direction of maximum spread or variance in the data, the first CoK-PCA vector (solid-blue line) captures the direction of extreme-valued samples. The second principal vectors (dashed lines) are, by construction, orthogonal to the respective first principal vectors (solid lines).

For reducing the dimensionality in the synthetic dataset, we construct two low-dimensional manifolds by retaining the first principal vector from PCA and CoK-PCA, i.e., by choosing $n_q = 1$. The two reduced representations of the data are obtained by projecting the original synthetic data onto these manifolds using Eq. 6. Note that a majority of samples in the main cluster are closer, in terms of normal distance, to the first principal PCA vector (solid-red line) and hence, would be represented more accurately in the PCA-reduced manifold. On the other hand, the extreme-valued samples are farther from the first principal PCA vector (solid-red line) when compared to the first principal CoK-PCA vector (solid-blue line). Therefore, these samples would be poorly represented in the PCA based low-dimensional manifold.

To evaluate the quality of the two reduced manifolds, it is imperative to evaluate the accuracy with which the
original state space can be reconstructed. For this, we employ the inverse transformation of Eq. 6 to obtain a linear reconstruction,

$$X_q = Z_q A_q^T,$$

(7)

where $X_q$ is the reconstructed data in the state space, which is the $v_1$-$v_2$ space for the synthetic dataset. Now, a comparison between $X_q$ and $X$ would reveal the quality of the two manifolds. It is important to note that standard reconstruction error metrics such as root mean squared, $R^2$, etc. would typically yield smaller errors with PCA as the first PCA principal vector would align well with the bulk of the data, as explained earlier. Thus, it becomes necessary to identify alternate error metrics which can ascertain the quality of reconstruction for both the extreme-valued as well as the bulk samples. In this work, we use the maximum and average values of the absolute reconstruction error ($\varepsilon = |X - X_q|$), $\varepsilon_M = \max(\varepsilon)$ and $\varepsilon_A = \text{mean}(\varepsilon)$, respectively, to quantify the accuracy in each reconstructed variable. Further, we examine the error ratio,

$$r_i = \ln \left\{ \frac{\varepsilon_{PCA}}{\varepsilon_{CoK-PCA}} \right\},$$

(8)

to establish the performance of CoK-PCA against PCA; the subscript $i$ represents either the maximum ($M$) or the average ($A$) errors. Here, we use the simplifying metric $r_i$ as comparisons in terms of absolute errors are cumbersome due to the fact that variables in combustion datasets take values that range of several orders of magnitude. Note that the absolute value of $r_i$ represents an improvement in the accuracy of one method relative to the other by a factor of $e^{(r_i)}$ with $r_i > 0$ indicating that the error from the PCA reduced manifold is larger than that from the CoK-PCA reduced manifold, and vice-versa for $r_i < 0$; a value of $r_i = 0$ indicates that the reconstruction errors from the two methods are equal.

For the synthetic dataset the values for the error ratio $r_i$ computed based on the maximum and average reconstruction errors are shown in Fig. 2. It can be seen that dimensionality reduction based on CoK-PCA outperforms PCA, both in terms of maximum and average reconstruction errors, in representing $v_1$ data which has significantly greater values for the extreme-valued samples. Conversely, PCA performs better than CoK-PCA in reconstructing $v_2$, however, to a lesser extent. Note that the values $v_2$ are not significant for the extreme-valued samples.

In the context of combustion, the extreme-valued points in the synthetic data can be considered to represent spatial locations corresponding to thin reaction zones associated with the flame propagation or ignition fronts which present stiff chemical dynamics. The variable $v_1$ would represent intermediate species which have greater values...
in the reactions zones and are negligible elsewhere. On the other hand, \( v_2 \) captures the behaviour of reactant or product species. Accordingly, we can speculate that dynamics in the reaction zone would be better captured by the low-dimensional manifold due to CoK-PCA. An assessment of this hypothesis is presented in the next section.

4. Results

4.1. Homogeneous reactor dataset

To investigate the accuracy of the proposed dimensionality reduction procedure for combustion datasets, we first consider a dataset that captures spontaneous ignition in a simple homogeneous (zero-dimensional) reactor. The choice of this particular dataset is motivated by the fact that ignition, being a highly nonlinear exponential process, results in a majority of the reactions, along with the production and consumption of intermediate species, to occur over a small fraction of the simulation time. Thus, the chosen dataset incorporates the kind of extreme events that present stiff dynamics which can potentially be well-captured by CoK-PCA.

The reactants consist of premixed ethylene and air at pressure, temperature, and equivalence ratio of \( P = 1.72 \text{ atm} \), \( T = 1200 \text{ K} \), and \( \phi = 0.4 \), respectively, while a 32-species 206-reactions mechanism [26] was used to represent the chemistry. The homogeneous reactor was evolved for 2.5 ms with a time step of 1 \( \mu \text{s} \) to yield 2501 data samples. Thus, the training data matrix \( D \) (dataset used to construct the reduced manifold) consists of \( n_g = 2501 \) points and \( n_v = 33 \) variables, which includes 32-species and temperature. The data matrix \( D \), as explained in the previous section, is scaled by subtracting with individual feature mean and normalized with feature absolute maximum to obtain the scaled data matrix \( X \). This matrix is then used to compute the co-variance matrix and the co-kurtosis tensor, which are factorized to obtain the PCA and CoK-PCA principal values and vectors.

![Figure 3](image_url)

**Figure 3:** Alignment between PCA and CoK-PCA principal vectors from homogeneous reactor dataset represented as a dot product for each mode.

To check whether PCA and CoK-PCA yield different vectors, we first present the dot-product of the two sets of unit vectors in Fig. 3. It can be seen that due to the inherently different nature of the two methods, the principal vectors are, in general, not aligned and different. This implies that the two methods identify different low-dimensional sub-spaces.

We now move on to characterize the behaviour of PCA and CoK-PCA reduced manifolds. First, we assess the quality of reconstruction of the training dataset itself. Here, we construct two reduced manifolds by first adopting an aggressive dimensionality reduction strategy that retains \( n_q = 5 \) leading principal vectors out of the \( n_v = 33 \) vectors. It should be noted that these \( n_q = 5 \) principal vectors in the PCA and CoK-PCA reduced manifolds correspond to approximately 99% of the variance and 98% of the kurtosis in the dataset, respectively. Thereafter,
we consider a less aggressive reduction with \( n_q = 5 \), i.e., approximately half of the total principal vectors are retained.

In Fig. 4, the accuracy of reduced manifolds with \( n_q = 5 \) is assessed in terms of their ability to represent the full state: thermo-chemical scalars, species production rates, and heat release rate. Note that the reconstructed thermo-chemical scalars, \( \mathbf{D}_q \), are used to compute species production rates and heat release rate, which are compared against those computed with the original state \( \mathbf{D} \). Since production rates depend non-linearly on the species concentrations and temperature, reconstruction errors in the latter are likely to be amplified, thereby providing a more stringent test of accuracy. Further, apart from having a tangible physical meaning, the reconstruction error associated with the heat release rate also provides an overall assessment of the quality of the reduced manifold since the heat release rate represents an aggregate effect of all the quantities of interest. Figure 4 shows the error ratios for these quantities based on the maximum and average errors, \( r_M \) and \( r_A \), respectively. \( N_2 \) being an inert species has not been included here. Figures 4 (a) and (b) show that the error due to CoK-PCA reconstruction is smaller than the PCA reconstruction (demonstrated by the blue colored bars) for 25 and 12 out of the 32 variables for \( r_M \) and \( r_A \) metrics, respectively. It should be noted that among the 25 species reconstructed better with CoK-PCA in \( r_m \) plot, the reconstruction quality for several scalars is much better than that of PCA, as indicated by the taller blue bars. These results indicate that the PCA reduced manifold, while presenting a better representation of the thermo-chemical scalars on an average, demonstrate significantly larger maximum linear reconstruction errors of the thermo-chemical scalars in comparison to the CoK-PCA reduced manifold. As will be shown later in this section, the samples corresponding to the maximum linear reconstruction error are representative of time instants corresponding to the occurrence of the majority of reactions and, thus, significantly contribute to the overall chemical dynamics of
CoK-PCA performs better in the reconstruction of species mass fractions, as is the standard practice, all reconstructed species mass fractions which yielded negative errors were represented by the average and maximum reconstruction errors in the heat release rate shown in Figs. 5 (c) and (b), it can be seen that the error due to CoK-PCA reconstruction is greater than PCA reconstruction for a larger fraction of the system. Figures 4 (c) and (d), show the reconstruction errors for the species production and heat release rates, which are non-linear functions of the thermo-chemical scalars. It can be clearly seen that CoK-PCA outperforms PCA in predicting the production rates for 23 of the 31 species with the $r_M$ metric and 18 of the 31 species for the $r_A$ metric. Furthermore, these figures show that the linear reconstruction error in the heat release rates, for both the $r_M$ and $r_A$ metrics, are smaller for the CoK-PCA thus establishing that the CoK-PCA represents the overall chemical state of the system better than PCA.

We next consider the quality of the reduced manifolds for the less aggressive case, $n_q = 15$. From Figs. 5 (a) and (b), it can be seen that the error due to CoK-PCA reconstruction is greater than PCA reconstruction for a larger number of variables as signified by the greater count of brown-colored bars: 19 and 20 of 32 variables in terms of the $r_M$ and $r_A$ metrics, respectively. However, it should be noted that wherever CoK-PCA performs better for the $r_M$ metric, it does so by a larger extent as compared to PCA (signified by the taller blue bars). Similarly, wherever CoK-PCA performs better in the $r_A$ metric, the performance is just as good as, if not better than, that of PCA. However, as shown in Figs. 5 (c) and (d), CoK-PCA significantly outperforms PCA in terms of reconstruction errors for the production rates of 23 and 21 species out of 31 species for the $r_M$ and $r_A$ metrics, respectively. Also, as in the $n_q = 5$ case, CoK-PCA once again captures the overall chemical state of the system better than PCA as represented by the average and maximum reconstruction errors in the heat release rate shown in Figs. 5 (c) and (d).

It should be noted that in order to ensure that the reconstructed thermo-chemical state results in unit sum of species mass fractions, as is the standard practice, all reconstructed species mass fractions which yielded negative values (that were all slightly smaller than zero) were taken to be zero after which any deviation from the sum...
equalling unity was adjusted for in the non-participating or bath species.

Figure 6: Time evolution of normalised error in heat release rate (HRR) computed from reconstructed data: (a) $n_q = 5$ and (b) $n_q = 15$. The error values are normalised with peak heat release rate value in the original data. Solid-red and dashed-blue lines represent errors from PCA and CoK-PCA, respectively. The progress variable is plotted in grey for reference.

Figure 7: Time evolution of production rate (PR) of (a) $\text{HO}_2$ and (b) $\text{CH}_2\text{O}$ computed from reconstructed data obtained from the $n_q = 15$ reduced manifolds. Solid-green, dashed-red and dashed-blue lines are production rates from original data, PCA and CoK-PCA, respectively. The progress variable is plotted in grey for reference.

Up to this point, the characterization of the CoK-PCA and PCA reduced manifolds that have been presented is agnostic of the temporal evolution of the system. Specifically, it is unclear if the data samples that incur large reconstruction errors for the reduced manifolds with the CoK-PCA and PCA belong to the reactants, the products, or the ignition front. To relate the results presented earlier to the progress of the combustion process, we present, in Fig. 6, the normalized error in heat release rate obtained through linearly reconstructed scalars from both the reduced manifolds for $n_q = 5$ and 15, along with progress variable. The normalized error, $\varepsilon_{\text{HRR}}$, is defined as:

$$
\varepsilon_{\text{HRR}} = \frac{|\text{HRR}_{\text{rm}} - \text{HRR}_{\text{original}}|}{\max(\text{HRR}_{\text{original}})},
$$

where the subscript “$\text{rm}$” refers to reduced manifold. The error profile for $n_q = 5$ presented in Fig. 6(a) shows that the HRR obtained from CoK-PCA incurs greater error in reactants and products as compared to PCA. In contrast, in the reaction zone where significant heat release would be present, CoK-PCA outperforms PCA. Similarly for the $n_q = 15$ reduced manifolds, as evident from Fig. 6(b), CoK-PCA predicts HRR with a greater accuracy in the reaction zone.

The trends in Fig. 6 can be explained as follows. Figs. 4 (a)-(b) and 5 (a)-(b) show that upon reconstructing the thermo-chemical scalars, CoK-PCA and PCA perform better in the $r_M$ and $r_A$ metrics, respectively. It is noteworthy that while $r_A$ captures the average error, including the errors incurred at all the time steps, $r_M$ captures the peak
error. For PCA, this peak error would usually manifest in the flame front where majority of reactions occur (see Fig. 6), and thus would significantly affect the species production rates and the heat release rate. Further, as the flame front is represented by only a fraction of all the data samples, and would typically be considered as extreme-valued samples, this region is expected to be represented better by the CoK-PCA reduced manifold (again see Fig. 6). Accordingly, as demonstrated in Figs. 4 (c)-(d) and 5 (c)-(d), the thermo-chemical scalars reconstructed from CoK-PCA reduced manifold yield a more accurate species production rates and overall heat release rate. Figure 6 presents further evidence that the proposed CoK-PCA method predicts the overall chemical kinetics in the reaction zone to a much better extent than PCA.

| Method  | Cumulative $\varepsilon_{PR}$ | HO$_2$ | CH$_2$O |
|---------|-------------------------------|--------|---------|
| PCA     | 4.9517                        | 3.8152 |         |
| CoK-PCA | 4.3663                        | 1.1834 |         |

Table 1: Cumulative errors in the production rates for the ignition markers in the early reaction zones (up to a progress variable of 0.1) obtained from different reduced manifolds for the homogeneous reactor case.

Further, in order to compare the relative accuracy of the PCA and CoK-PCA reduced manifolds in very-low progress variable region, which characterizes the thermal runaway that is important in accurately predicting the ignition delay, we examine the production rates of two key radicals in these early reaction zones, namely the hydroperoxy (HO$_2$) and formaldehyde (CH$_2$O) radicals, respectively, as ignition markers. For the HO$_2$ radical (Fig. 7 (a)), we qualitatively observe that the CoK-PCA reduced manifold slightly over-predicts the production rate between $t = 0.4$ ms to $t = 0.6$ ms, while the PCA reduced manifold under-predicts to small extent. After $t = 0.6$ ms, the HO$_2$ production rates from both reduced manifolds oscillate around the true production rate values. In contrast, from Fig. 7 (b), the production rates from the CoK-PCA reduced manifold can, qualitatively, be seen to be far better aligned with the true values as compared to those obtained from the PCA reduced manifold. To make a further quantitative assessment, we evaluate the cumulative normalized error in the production rates of the HO$_2$ and CH$_2$O radicals up to a progress variable value of 0.1. The normalized error values are evaluated as:

$$\varepsilon_{PR} = \frac{|PR_{rm} - PR_{original}|}{\max(|PR_{original}|)},$$

and the cumulative values are tabulated in Table 1. It can be seen that the cumulative errors associated with the production rates of the considered ignition markers obtained with the CoK-PCA reduced manifold are smaller than those from the PCA reduced manifold, thus indicating that the proposed method provides a better estimate of the ignition time delay as compared to PCA. It should be noted that the present discussion is based on the reconstructed production rates obtained from the $n_q = 15$ reduced manifolds; similar observations also hold true for the $n_q = 5$ reduced manifolds.

In order to assess the robustness of the dimensionality reduction procedure we test the PCA and CoK-PCA manifolds at conditions different from that of the training dataset. Using the PCA and CoK-PCA manifolds obtained at the training (baseline) conditions of $T_{training} = 1200$ K and $\phi_{training} = 0.4$, we evaluate reconstruction accuracy for datasets at test conditions with different initial temperature and equivalence ratio. We assess the reconstruction
accuracy of the test conditions with the manifolds corresponding to the more aggressive reduction, \( n_q = 5 \) principal vectors. Based on previous experimental studies \([27, 28]\), the test conditions differ from the training conditions by \( \Delta T = \pm 50 \text{K} \) and \( \Delta \phi = \pm 0.025 \), respectively. Reconstruction errors for the species production rates and heat release rate are presented in Fig. 8. It can be seen that the CoK-PCA manifold yields satisfactory predictions on test conditions in both the \( r_M \) and \( r_A \) metric, as evidenced by the larger number of blue bars in comparison to the brown bars. Specifically, when compared to the PCA reduced manifold, the CoK-PCA reduced manifold presents a better reconstruction (averaged across all four test cases) of the species production rates for approximately 73%
and 60% of the considered species in the $r_M$ and $r_A$ metrics, respectively. Further, the CoK-PCA manifold is also able to capture the overall chemical kinetics better than the PCA manifold at the test conditions, as evidenced by the positive or comparable values (for the $r_A$ metric in the $T=1250$ K, $\phi =0.375, 0.425$ cases) of the heat release rate.

4.2. Homogeneous charge compression ignition (HCCI) dataset

Next we consider a dataset that includes the effects of convective and diffusive spatial transport and turbulence, homogeneous charge compression ignition (HCCI) of ethanol representative of internal combustion engines [29]. The simulation corresponds to high pressure, high temperature auto-ignition of turbulent premixed ethanol-air mixed with combustion products, representing “exhaust gas recirculation” (EGR), in a fully periodic domain with a uniform 2D spatial grid comprising $672 \times 672$ grid points. The simulation was initialized with a nominal pressure of 45 atm and a mean temperature of 924 K. The equivalence ratio of the reactants was set to 0.4. To model the uneven mixing caused by the EGR, a spatial temperature fluctuation, along with a separately computed divergence-free turbulent velocity field, was superimposed onto the system. Additionally, compression heating effects induced due to the motion of the piston were also accounted for in the simulation. The chemical mechanism associated with ethanol combustion consisted of 28 chemical species. Thus, at each temporal checkpoint, the dataset $D$ comprises of $n_q = 672 \times 672$ samples with $n_v = 29$ thermo-chemical scalars. The temporal checkpoints of interest correspond to $t = 0.845$ and 1.2 ms [22], as these time instances correspond to the inception of auto-ignition kernels and the propagation of the flame front in the bulk of the domain, respectively, as shown in the heat release rate contours in Fig. 9. It should be noted that the data corresponding to $t = 1.2$ ms can be considered as a “regular” dataset, whereas the $t = 0.845$ ms dataset can be considered to comprise of outlying events due to the presence of localised ignition kernels.

As in the case of the zero-dimensional homogeneous reactor dataset, we obtain the required PCA and CoK-PCA principal vectors from the procedure described in Sec. 2. The entire PCA and CoK-PCA analyses were performed separately on each temporal snapshot, rather than combining them. In Fig. 10, we present the orientation of the principal vectors obtained from the two methods for the considered time instances. It can be observed that, for both the time instances, the PCA and CoK-PCA manifolds are significantly different from one another and thus identify different lower-dimensional sub-spaces.

The required reduced manifolds are constructed by retaining $n_q = 5$ out of the $n_v = 29$ principal vectors as these principal vectors correspond to approximately 99% of the variance and kurtosis in the reduced PCA and CoK-PCA
The heat release rate in terms of the ratio $r$ and average (d) errors of species production rates and heat release rate. The blue and orange lines represent the CoK-PCA drastically deteriorates in terms of its ability to reconstruct the thermo-chemical scalars on average, i.e., $r$.

manifolds. Thereafter, as in the case of the zero-dimensional homogeneous reactor dataset, the thermo-chemical state is linearly reconstructed. In the following, we present a comparison of the quality of the PCA and CoK-PCA reduced manifolds in terms of the reconstruction errors of the thermo-chemical scalars, the species production rates and heat release rate.

![Figure 10: Alignment between PCA and CoK-PCA principal vectors from HCCI dataset represented as a dot product for each mode. The blue and orange lines represent $t = 0.845 \text{ ms}$ and $t = 1.2 \text{ ms}$, respectively.](image)

![Figure 11: Plots of error ratio $r_i$ (i ∈ {M, A}), see Eq. 8) for reconstructed data, $n_q = 5$, from HCCI dataset at $t = 0.845 \text{ ms}$. Error ratio $r_i$ based on maximum (a) and average (b) errors of species mass fractions and temperature. Error ratio $r_i$ based on maximum (c) and average (d) errors of species production rates and heat release rate.](image)

In Figs. 11 (a) and (c), it can be clearly seen that CoK-PCA performs significantly better than PCA for $t = 0.845 \text{ ms}$ data when reconstructing the thermo-chemical scalars along with the species production rates and heat release rate in terms of the $r_M$ metric. However, in Fig. 11(b), we observe that the performance of the CoK-PCA drastically deteriorates in terms of its ability to reconstruct the thermo-chemical scalars on average, i.e., with the $r_A$ metric. Similarly, in Fig. 11(d), we observe that the overall chemical kinetics is poorly represented by CoK-PCA as demonstrated by the poor $r_A$ values of the production rates and heat release rate. Again, it should
be recalled that the excess kurtosis is inherently better suited to capture outlier events, which, in the case of the $t = 0.845 \text{ ms}$ dataset, comprises of a small number of samples corresponding to the formation of ignition kernels (see Fig. 9(a)). We hypothesize that the chemical manifold identified by CoK-PCA better represents the chemical subspace for the smaller number of ignition samples, rather than the overall domain. Thus, in the case of the $r_A$ metric, the non-igniting zones of the simulation domain are not represented by CoK-PCA reduced manifold leading to a higher average error. To test this hypothesis, and to establish an equitable comparison of CoK-PCA against PCA, we evaluate $r_M$ and $r_A$ for the thermo-chemical scalars, species production rates and heat release rate only for those spatial locations that correspond to the reacting zones of the simulation domain. These reacting zones are identified as spatial points having a heat release rate larger than $7 \times 10^8 \text{ Jm}^{-3} \text{s}^{-1}$, which depict the boundaries of matured ignition kernels. For these regions, as shown in Fig. 12, it can be seen that the reduced manifold obtained with the CoK-PCA method provides a much better representation of the thermo-chemical state as well as the chemical kinetics of the system when compared to PCA in terms of both the maximum and average errors.

For the $t = 1.2 \text{ ms}$ time instant it should be noted that the number of samples comprising the reacting and non-reacting zones are comparable (see Fig. 9(b)). Thus, we expect the performance of the PCA and CoK-PCA methods to be largely equivalent to one another in terms of capturing the thermo-chemical scalars. As shown in Figs. 13 (a) and (b), this is indeed observed to be the case. However, Figs. 13 (c) and (d) indicate that CoK-PCA captures the species production rates and the heat release rate, both in terms of both the $r_M$ and $r_A$ metrics, much better than PCA. Similarly, due to the comparable sizes of the reacting and non-reacting zones, the results of the analysis conditioned on the reacting zones (spatial locations with heat release rates larger than $7 \times 10^8 \text{ Jm}^{-3} \text{s}^{-1}$),
Next, we assess the robustness with which the PCA and CoK-PCA manifolds reconstruct the state of datasets that were not included in the training exercise. Here, using the \( t = 0.845 \) \textit{ms} dataset as the training data, we present the reconstruction errors in the production rates and heat release rates obtained from \( n_q = 5 \) reduced manifolds for two test datasets in the neighbourhood of the training data, namely \( t = 0.835 \) \textit{ms} and \( t = 0.855 \) \textit{ms}; as done previously, the analysis comprises a characterization in the whole domain as well as in the reacting zones only. The results are presented in Figs. 15 and 16.

Conforming to the trends observed earlier, in Figs. 15(a), (b) and 16(a), (b) we observe that, for the \( r_M \) metric, the CoK-PCA reduced manifold yields a much better reconstruction, in the entire domain as well as the reacting region, for both test instances considered. Similarly, the results for the \( r_A \) metric for both test datasets, shown in

as shown in Fig. 14, demonstrates similar trends as those for the entire domain.

To summarize, for the HCCI data, the CoK-PCA based lower-dimensional manifold would, at the first glance, appear to present a poor representation of the original chemical state for initial time steps when the number of ignition kernels are small. However, upon careful inspection, it can be appreciated that the advantages of the CoK-PCA, viz. the ability to effectively represent stiff chemical dynamics in the reaction zones of the simulation domain, remain. It is important to note that the accurately represented chemical dynamics in the reaction zones would result in smaller errors being propagated as the reduced-order simulations integrate the lower-dimensional system in time. Further, for the intermediate time instant, CoK-PCA performs just as well as PCA in reconstructing the thermo-chemical scalars due to the equal distribution of the reacting and non-reacting zones. The species production and heat release rates, in this case, are more accurately predicted from CoK-PCA based reconstruction.
such as the formation of localized ignition kernels are present, while also accurately capturing the chemical kinetics the thermo-chemical scalars particularly well, in comparison to PCA, for situations where extreme-valued events are present. It has been found that the reduced manifolds obtained from the CoK-PCA method recover dimensional ignition in a homogeneous reactor, and a two-dimensional DNS of homogeneous charge compression ignition phenomenon. The dimensionality reduction (based on maximum (c) and average (d) errors of species production rates and heat release rate).

Figure 14: Plots of error ratio $r_i$ ($i \in \{ M, A \}$, see Eq. 8) for reconstructed data, $n_q = 5$, from HCCI dataset at $t = 1.2$ ms in the reacting region. Error ratio $r_i$ based on maximum (a) and average (b) errors of species mass fractions and temperature. Error ratio $r_i$ based on maximum (c) and average (d) errors of species production rates and heat release rate.

Figs. 15(c), (d) and 16(c), (d), also follow the trends observed earlier, in that the PCA and CoK-PCA reduced manifold presents a better reconstruction in the entire domain and the reacting zones, respectively.

5. Conclusions and future work

In this paper, we have proposed an alternate dimensionality reduction procedure (CoK-PCA) based on the fourth-order joint moment, i.e., co-kurtosis tensor, as opposed to co-variance which is the basis for the more commonly used Principal Component Analysis (PCA). The rationale for using a higher order joint moment is that chemical reactions represent stiff dynamics and the chemistry-relevant samples in a combustion dataset may typically be extreme-valued (compared to unburnt or fully burnt samples). Accordingly, the directions in thermo-chemical state space that best represent chemical dynamics are captured better by moments higher than co-variance. A brief background on the construction and factorization of the co-kurtosis tensor was presented, along with a discussion on the connection of this method to independent component analysis. The dimensionality reduction (based on PCA and CoK-PCA) and the linear data reconstruction procedures were elucidated by means of a synthetically generated dataset. The potential of the proposed CoK-PCA method has been evaluated using two datasets: zero-dimensional ignition in a homogeneous reactor, and a two-dimensional DNS of homogeneous charge compression ignition phenomenon. It has been found that the reduced manifolds obtained from the CoK-PCA method recover the thermo-chemical scalars particularly well, in comparison to PCA, for situations where extreme-valued events such as the formation of localized ignition kernels are present, while also accurately capturing the chemical kinetics.

17
CoK-PCA in conjunction will be examined in future.
represented by PCA manifolds, and others by a CoK-PCA manifold. Such techniques to consider the PCA and
replacement of PCA in every scenario, the best approach may be a hybrid with certain portions of the dataset
similarly be improved using non-linear reconstruction, which should further improve the accuracy of reconstruction
the accuracy of linear reconstruction is limited and other studies have adopted non-linear reconstruction including
linear reconstruction of thermo-chemical scalars from the identified principal components. It is well-known that
comparison between PCA and CoK-PCA, and to that extent used most straightforward techniques, particularly
methods such as regression splines and neural networks. We posit that the reconstruction accuracy of CoK-PCA can
similarly be improved using non-linear reconstruction, which should further improve the accuracy of reconstruction
of species production rates, and this will be considered in a future study. Similarly, refinements of PCA such
as kernel PCA and local PCA that are widely considered in other studies, can be applied to CoK-PCA. Finally,

Figure 15: Plots of error ratio \( r_i \) (\( i \in \{M, A\} \), see Eq. 8) for reconstructed data for HCCI dataset at \( t = 0.835 \) ms. The data is
reconstructed by retaining from \( n_q = 5 \) principal vectors from manifolds obtained from the training \( t = 0.845 \) ms HCCI dataset. Error
ratio \( r_i \) based on maximum (a) and average (b) errors of species production rates and heat release rate in the entire domain. Error
ratio \( r_i \) based on maximum (c) and average (d) errors of species production rates and heat release rate in reacting regions.

of the reacting zones of system. Further, the present study reveals that PCA demonstrates large reconstruction
errors in the reaction zones, specifically in terms of the species production and heat release rates. In contrast, it is
found that CoK-PCA yields reconstruction errors that are more uniformly distributed across all samples and thus
presents a qualitatively true representation of the overall reaction dynamics.

There are numerous topics of investigation that have been reserved for future work. We have presented an initial
comparison between PCA and CoK-PCA, and to that extent used most straightforward techniques, particularly
linear reconstruction of thermo-chemical scalars from the identified principal components. It is well-known that
the accuracy of linear reconstruction is limited and other studies have adopted non-linear reconstruction including
methods such as regression splines and neural networks. We posit that the reconstruction accuracy of CoK-PCA can
similarly be improved using non-linear reconstruction, which should further improve the accuracy of reconstruction
of species production rates, and this will be considered in a future study. Similarly, refinements of PCA such
as kernel PCA and local PCA that are widely considered in other studies, can be applied to CoK-PCA. Finally,

it should be pointed out that, while it may be unfeasible to use the proposed CoK-PCA method as a drop-in
replacement of PCA in every scenario, the best approach may be a hybrid with certain portions of the dataset
represented by PCA manifolds, and others by a CoK-PCA manifold. Such techniques to consider the PCA and
CoK-PCA in conjunction will be examined in future.
Figure 16: Plots of error ratio $r_i$ ($i \in \{M, A\}$, see Eq. 8) for reconstructed data for HCCI dataset at $t = 0.855$ ms. The data is reconstructed by retaining from $n_q = 5$ principal vectors from manifolds obtained from the training $t = 0.845$ ms HCCI dataset. Error ratio $r_i$ based on maximum (a) and average (b) errors of species production rates and heat release rate in the entire domain. Error ratio $r_i$ based on maximum (c) and average (d) errors of species production rates and heat release rate in reacting regions.

Acknowledgments

The work at IISc was supported under a project from the National Supercomputing Mission, India. AJ was funded by a project from Shell Technology Center, Bengaluru, India. Work by HK was part of the ExaLearn Co-design Center, supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA-0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

[1] K. Aditya, A. Gruber, C. Xu, T. Lu, A. Krisman, M. R. Bothien, J. H. Chen, Direct numerical simulation of flame stabilization assisted by autoignition in a reheat gas turbine combustor, Proc. Combust. Inst. 37 (2019) 2635–2642.

[2] B. Savard, E. R. Hawkes, K. Aditya, H. Wang, J. H. Chen, Regimes of premixed turbulent spontaneous ignition and deflagration under gas-turbine reheat combustion conditions, Combust. Flame 208 (2019) 402–419.
[3] L. Berger, R. Hesse, K. Kleinheinz, M. J. Hegtschweiler, A. Attili, J. Beeckmann, G. T. Linteris, H. Pitsch, A DNS study of the impact of gravity on spherically expanding laminar premixed flames, Combust. Flame 216 (2020) 412–425.

[4] G. Nivarti, S. Cant, Direct numerical simulation of the bending effect in turbulent premixed flames, Proc. Combust. Inst. 36 (2017) 1903–1910.

[5] S. Desai, Y. J. Kim, W. Song, M. B. Luong, F. E. H. Pérez, R. Sankaran, H. G. Im, Direct numerical simulations of turbulent reacting flows with shock waves and stiff chemistry using many-core/gpu acceleration, Comput. Fluids 215 (2021) 104787.

[6] S. B. Pope, Small scales, many species and the manifold challenges of turbulent combustion, Proc. Combust. Inst. 34 (2013) 1–31.

[7] U. Maas, S. Pope, Simplifying chemical kinetics: Intrinsic low-dimensional manifolds in composition space, Combust. and Flame 88 (3) (1992) 239–264.

[8] S. Lam, D. Goussis, Understanding complex chemical kinetics with computational singular perturbation, Symposium (International) on Combustion 22 (1) (1989) 931–941.

[9] O. Gicquel, N. Darabiha, D. Thévenin, Liminar premixed hydrogen/air counterflow flame simulations using flame prolongation of ifldm with differential diffusion, Proc. Combust. Inst. 28 (2) (2000) 1901–1908.

[10] J. van Oijen, F. Lammers, L. de Goey, Modeling of complex premixed burner systems by using flamelet-generated manifolds, Combust. and Flame 127 (3) (2001) 2124–2134.

[11] J. C. Sutherland, A. Parente, Combustion modeling using principal component analysis, Proc. Combust. Inst. 32 (2009) 1563–1570.

[12] A. Biglari, J. C. Sutherland, A filter-independent model identification technique for turbulent combustion modeling, Combust. Flame 159 (2012) 1960–1970.

[13] Y. Yang, S. B. Pope, J. H. Chen, Empirical low-dimensional manifolds in composition space, Combust. Flame 160 (2013) 1967–1980.

[14] T. Echekki, H. Mirgolbabaei, Principal component transport in turbulent combustion: A posteriori analysis, Combust. Flame 162 (2015) 1919–1933.

[15] H. Mirgolbabaei, T. Echekki, Nonlinear reduction of combustion composition space with kernel principal component analysis, Combust. Flame 161 (2014) 118–126.

[16] R. Ranade, T. Echekki, A framework for data-based turbulent combustion closure: A priori validation, Combust. Flame 206 (2019) 490–505.

[17] M. R. Malik, P. O. Vega, A. Coussement, A. Parente, Combustion modeling using principal component analysis: A posteriori validation on sandia flames d, e and f, Proc. Combust. Inst. 38 (2021) 2635–2643.
[18] B. J. Isaac, J. N. Thornock, J. Sutherland, P. J. Smith, A. Parente, Advanced regression methods for combustion modelling using principal components, Combust. Flame 162 (2015) 2592–2601.

[19] A. Parente, J. C. Sutherland, L. Tognotti, P. J. Smith, Identification of low-dimensional manifolds in turbulent flames, Proc. Combust. Inst. 32 (2009) 1579–1586.

[20] A. Parente, J. Sutherland, B. B. Dally, L. Tognotti, P. Smith, Investigation of the mild combustion regime via principal component analysis, Proc. Combust. Inst. 33 (2011) 3333–3341.

[21] A. Coussement, O. Gicquel, A. Parente, Kernel density weighted principal component analysis of combustion processes, Combust. Flame 159 (2012) 2844–2855.

[22] K. Aditya, H. Kolla, W. P. Kegelmeyer, T. M. Shead, J. Ling, W. L. Davis, Anomaly detection in scientific data using joint statistical moments, J. Comput. Phys. 387 (2019) 522–538.

[23] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Networks 13 (2000) 411–430.

[24] P. Comon, Tensor decompositions, Mathematics in Signal Processing V (2002) 1–24.

[25] A. Anandkumar, R. Ge, D. Hsu, S. Kakade, M. Telgarsky, Tensor decompositions for learning latent variable models, J. Mach. Learn. Res. 15 (2014) 2773–2832.

[26] Z. Luo, C. S. Yoo, E. S. Richardson, J. H. Chen, C. K. Law, T. Lu, Chemical explosive mode analysis for a turbulent lifted ethylene jet flame in highly-heated coflow, Combust. Flame 159 (2012) 265–274.

[27] C. M. Geipel, R. D. Rockwell, H. K. Chelliah, A. D. Cutler, Flame structure characterization in a dual-mode scramjet using hydroxyl planar laser-induced fluorescence, AIAA J. 59 (5) (2021) 1824–1835.

[28] D. A. Lieber, C. P. Goyne, R. D. Rockwell, C. M. Geipel, H. K. Chelliah, Design and testing of an additively manufactured flameholder for scramjet flows, Aerospace Science and Technology 126 (2022) 107656.

[29] A. Bhagatwala, J. H. Chen, T. Lu, Direct numerical simulations of hcci/saci with ethanol, Combust. Flame 161 (2014) 1826–1841.