Combustion and Flame 161 (2014) 118–126

Contents lists available at ScienceDirect

Combustion and Flame

journal homepage: www.elsevier.com/locate/combustflame

Nonlinear reduction of combustion composition space with kernel principal component analysis

Hessam Mirgolbabaei, Tarek Echekki *

Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC, USA

ARTICLE INFO

Article history:
Received 1 February 2013
Received in revised form 17 April 2013
Accepted 21 August 2013
Available online 14 September 2013

Keywords:
Kernel principal component analysis
Principal component analysis
Turbulent nonpremixed flames

ABSTRACT

Kernel principal component analysis (KPCA) as a nonlinear alternative to classical principal component analysis (PCA) of combustion composition space is investigated. With the proposed approach, thermo-chemical scalar's statistics are reconstructed from the KPCA derived moments. The tabulation of the scalars is then implemented using artificial neural networks (ANN). Excellent agreement with the original data is obtained with only 2 principal components (PCs) from numerical simulations of the Sandia Flame F flame for major species and temperature. A formulation for the source and diffusion coefficient matrix for the PCs is proposed. This formulation enables the tabulation of these key transport terms in terms of the PCs and their potential implementation for the numerical solution of the PCs' transport equations.

1. Introduction

Moment-based methods have been widely used in turbulent combustion modeling. These methods are based on the reconstruction of thermo-chemical scalars' statistics from a set of transported moments. Instead of *ad hoc* strategies to select moments, more optimally derived parameters have been proposed recently using principal component analysis (PCA) [1–7]. However, it is not evident that the linear PCA alone can adequately represent the nonlinear nature of the space representing thermo-chemical scalars' statistics.

A number of approaches have been adopted for combustion to address the non-linearity of the combustion composition space using PCA-based methods. Parente et al. [3] introduced two local PCA methodologies that partition the composition space into clusters where the local composition space exhibits linearity. They find that their proposed approach does improve the reconstruction of experimental data sets that exhibit inherent non-linearity in their composition space. The methods proposed in Ref. [3], although convenient for reconstruction purposes, may not be readily suitable for the purposes of the proposed effort here, which is concerned with the ability to develop transport equations for principal components (PCs). The principal constraints of the method in Ref. [3] is that, if implemented, it requires the transport of different PCs appropriate for the different regions of composition space, which may not be easily translatable to solutions in physical space.

Coussement et al. [6] propose an extension of the local PCA approach called the manifold generated by a local PCA or MG-L-PCA. The method is designed to couple the manifolds identified by the local PCA approach and implement them within the context of direct numerical simulation (DNS). In the MG-L-PCA approach, a variable, which evolves monotonically across the flame (e.g. temperature), is used to identify continuous clusters. A fixed number of thermo-chemical scalars is chosen based on the ability to reconstruct all thermo-chemical scalars such that reconstruction of the original scalars is achieved in all clusters. Therefore, the MG-L-PCA method is based on the transport of a subset of the original thermo-chemical scalars' vector and not on the transport of PCs. Comparison with DNS of premixed flame-vortex interactions shows that MG-L-PCA reproduces some salient features of the composition space with only 5 transported thermo-chemical scalars retained. With the proposed procedure, it is not evident that the MG-L-PCA can achieve a significant reduction of the number of transported variables as compared to the non-linear implementation of PCA.

In the present work, we investigate the use of a non-linear alternative, kernel PCA (KPCA) [8], to reconstruct turbulent combustion thermo-chemical space. Kernel principal components (PCs) are derived based on stand-alone one-dimensional turbulence (ODT) solutions of the Sandia Flame F flame [9], which is characterized by different regimes of combustion starting with pilot stabilization, to extinction and re-ignition and self-stabilized combustion. Then, the composition space is reconstructed using artificial neural networks (ANN). The tabulated and the original data are compared to demonstrate the validity of the approach and to investigate the number of PCs required to reproduce the original thermo-chemical scalars' vector. Recently, a similar procedure based on

* Corresponding author.
E-mail address: techekk@ncsu.edu (T. Echekki).

0010-2180/$ - see front matter © 2013 The Combustion Institute. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.combustflame.2013.08.016

linear PCA and ANN has been proposed by the authors [7]. Another important objective of this study is the formulation of transport terms for the PCs. Such terms tabulated in terms of the PCs will enable the numerical solution of the PCs transport equations.

2. PCA parameterization and ANN Tabulation

The PCA or KPCA analysis starts with detailed experimental and computational data of thermo-chemical scalars' vector, such as what is adopted here based on temperature and composition $\boldsymbol{\theta} = (T, Y_1, \cdots, Y_{N-1})^T$ where T are the mixture temperature and the mass fractions of $N-1$ species in the mixture denoted by Y_1, Y_2, …, Y_{N-1}. The original thermo-chemical scalars are subject to a set of transport equations:

$$\frac{D\boldsymbol{\theta}}{Dt} = \nabla \cdot \mathbf{J}_{\theta} + \mathbf{S}_{\theta} \tag{1}$$

where $\mathbf{J}_{\boldsymbol{\theta}}$ and $\mathbf{S}_{\boldsymbol{\theta}}$ correspond to the diffusive fluxes and the chemical source terms for the original thermo-chemical scalars' vector, θ, respectively. PCA or KPCA analysis results in a smaller vector of principal components: $\boldsymbol{\Psi} = (\Psi_1, \Psi_2, \cdots, \Psi_{N_{PC}})$ where the Ψ's are the principal components and N_{PC} is the dimension of the principal components' vector. We expect N_{PC} to be much smaller than N for the reduction process to be meaningful.

Similarly, we would like to investigate if we can model the transport equations for the PCs in a similar form to that of the thermo-chemical scalars:

$$\frac{D\psi}{Dt} = \nabla \cdot J_{\psi} + S_{\psi} \tag{2}$$

where $\mathbf{J}_{\boldsymbol{\psi}}$ and $\mathbf{S}_{\boldsymbol{\psi}}$ correspond to the diffusive fluxes and the chemical source terms for the PCs, respectively. We would like to reconstruct these terms as well; but, the scope of the present contribution is to address whether KPCA can result in a more effective reduction of the original data compared to PCA, which translates into less PCs required for representing the original thermo-chemical scalars.

The steps used to derive the moments and the terms in their governing equations, Eq. (2), are as follows: (i) A subset of the thermo-chemical scalars' vector is selected, which may mark the important measure of the mixing and reaction in the composition space [7], (ii) the KPCA is performed on the multiple observations of the selected variables, and (iii) ANN is used to reconstruct the full set of the original thermo-chemical scalars' vector, including a subset of variables upon which the moments, PCs, are constructed and the remaining variables from the original set.

As in the linear PCA analysis, KPCA requires a dataset. In our present analysis, the dataset corresponds to spatial and temporal data at n discrete grid points of ODT solutions for temperature and a subset of the computed species mass fractions. According to Cover's theorem, the non-linear data structure in the input space is more likely to be linear after high-dimensional non-linear mapping [8]. The basic idea of KPCA is to non-linearly map the original data, $\boldsymbol{\theta}$, into a higher dimensional space, with an arbitrary large, possibly infinite dimensionality, $\boldsymbol{\theta} \rightarrow \boldsymbol{\phi}$, called *feature space*. The new mapped data, $\boldsymbol{\phi}$, have linear dependencies in this feature space. In this space, similar to classical PCA, a covariance matrix of the centered mapped data, $\tilde{\boldsymbol{\phi}}(\boldsymbol{\theta}_i)$, is defined as:

$$\mathbf{C} = \frac{1}{n}\tilde{\phi}(\boldsymbol{\theta}) \cdot \tilde{\phi}(\boldsymbol{\theta})^T \text{ where } \tilde{\phi}(\boldsymbol{\theta}) = \phi(\theta) - \bar{\phi} \tag{3}$$

where $\bar{\phi}$ is the mean of the feature space variable. An eigenvalue problem for the $n \times n$ covariance matrix $\mathbf{C}$, where n corresponds to the number of available observations, may be set up as follows:

$$\lambda \mathbf{V} = \mathbf{C}\mathbf{V}. \tag{4}$$

In this expression, λ is a diagonal matrix of the eigenvalues of $\mathbf{C}$, and $\mathbf{V}$, is the matrix of its eigenvectors. Solving the eigenvalue problem (Eq. (4)) in feature space is expensive computationally due to its high dimension and requires an explicit form for the mapping. Instead, KPCA uses an alternative eigenvalue problem based on the so-called kernel matrix:

$$\widetilde{\mathbf{K}}\alpha_m = n\lambda_m\alpha_m. \tag{5}$$

In the above expression, the matrix $\widetilde{K}$ is the $n \times n$ kernel matrix based on the centered mapped data. Its components are defined as $\widetilde{\mathbf{K}}_{ij} = \tilde{\phi}(\boldsymbol{\theta}_i) \cdot \tilde{\phi}(\boldsymbol{\theta}_j)$. The kernel matrix based on the non-centered feature space variables is expressed as $\mathbf{K}_{ij} = \phi(\boldsymbol{\theta}_i) \cdot \phi(\boldsymbol{\theta}_j)$. In Eq. (5), $n\lambda_m$ and α_m correspond to the mth eigenvalue and the $n \times 1$ eigenvector of $\widetilde{\mathbf{K}}$, respectively. Again, to avoid having to address variables in feature space, an alternative strategy is adopted to determine the components of the kernel matrix based on the so-called kernel trick [8]. These components are expressed as:

$$\mathbf{K}_{ij} = k(\theta_i, \theta_j) \tag{6}$$

where k is a kernel function. Different kernel functions have been adopted in the literature, including the polynomial, the Gaussian, the radial basis and the hyperbolic tangent. In this study, we adopt the polynomial with coefficient 2, $k(\theta_i, \theta_j) = (\theta_i^T\theta_j)^2$. A simple conversion relation, which is not derived here, can be used to relate the kernel matrix of the centered mapped data, $\widetilde{\mathbf{K}}$, and the kernel matrix of the non-centered mapped data, $\mathbf{K}$, based on Eq. (6) (see Ref. [8]). The mth principal component based on the training data is calculated as:

$$\psi_m = \frac{1}{\sqrt{n\lambda_m}}\alpha_m^T\widetilde{\mathbf{K}}. \tag{7}$$

We can reconstruct any pairs of original thermal-chemical scalars, θ_{test}, and their corresponding first N_{PC} kernel principal components using:

$$\psi_{m,\text{test}} = \frac{1}{\sqrt{n\lambda_m}}\alpha_m^T\widetilde{\mathbf{K}}_{test} \tag{8}$$

where $\widetilde{\mathbf{K}}_{test}$ is the kernel matrix correlating the new thermo-chemical scalars and the original data on which the eigenvalues and the eigenvectors are constructed, $\widetilde{\mathbf{K}}_{test} = k(\theta_{test}, \theta_{data})$. The MATLAB Toolbox for Dimensionality Reduction is used for the KPCA [10].

ANN provides the flexibility needed for the choice of input parameters (the principal components) and output parameters (thermo-chemical scalars and, subsequently, transport and source terms for the principal components). In the tabulation used here, the same ANN layer structure is used for all species and temperature; however, the various thermo-chemical scalars are tabulated independently of each other. They include the input layer containing the PCs, 1 hidden layer composed of 20 neurons, and an output layer with one neuron containing any one of the variables to be tabulated (thermo-chemical scalars in this paper). Bias connections for the neurons in each hidden layer and output layer are also used. The inputs, principal components, and output of a network are re-scaled so they fall in the range from −1 to 1.

The activation function for neurons in the hidden layer is the tangent sigmoid function. The sigmoid function yields values that fall in the range of −1 to 1. The choice of a sigmoid function would have been, in theory, ideal for the output neurons, as well, because it keeps the thermo-chemical scalars bounded. Accordingly, mass fractions would be bounded by the full range of the original data yielding, for example, mass fractions that fall between 0 and 1. However, we have found that a pure linear function between the hidden layer and the output neuron to yield better overall predictions of the species and temperature profiles. As a result, an insignificant portion of the mass fractions may fall below zero.

Therefore, further optimization of the ANN activation function is needed to obtain both improved predictions of the tabulation and bounded scalars.

3. Evaluation of transport terms

In Ref. [1], Sutherland and Parente derived expressions for the two key transport terms for the principal components: (1) their source terms and (2) their diffusive flux terms. Their analysis is based on the linear relation between the PC's and the original thermo-chemical scalars vector: $\psi = \mathbf{Q}^T\theta$ and $\psi^{\text{red}} = \mathbf{A}^T\theta$, where $\mathbf{Q}$ is the NxN matrix made up of the N orthogonal eigenvectors. $\mathbf{A}$ corresponds to the matrix made up of the leading N_{PC} eigenvectors of $\mathbf{Q}$. The superscript "red" corresponds to the selected reduced set of PCs chosen to represent the data. If a single PCA analysis is adopted for the entire data, the matrices $\mathbf{Q}$ and $\mathbf{A}$ are constant matrices. From the linear relation between PCs and the thermo-chemical scalars, the following two principal relations are obtained for the PCs source and diffusive flux terms: $\mathbf{S}_\psi = \mathbf{Q}^T\mathbf{S}_\theta$ ($\mathbf{S}_\psi^{\text{red}} = \mathbf{A}^T\mathbf{S}_\theta$) and $\mathbf{J}_\psi = \mathbf{Q}^T\mathbf{J}_\theta$ ($\mathbf{J}_\psi^{\text{red}} = \mathbf{A}^T\mathbf{J}_\theta$) [1]. They indicate that the source terms and the diffusive flux terms for the PCs, too, are linearly related to the source and chemical flux terms of the original thermo-chemical scalars. For non-linear PCA applications, the linear relation, which simplifies the analysis of the transport terms, is not available. In the present analysis, we attempt to provide a similar formulation that can apply to non-linear PCA analysis, such as the case adopted here.

Although, Sutherland and Parente [1] proposed a simple relation for the diffusive fluxes of the PCs, such a relation may not be convenient for the purposes of tabulation of the diffusion properties of these PCs in terms of the PCs as is intended in our analysis. Indeed, diffusive fluxes depend on the spatial gradients of scalars as well as their corresponding diffusion coefficients. Here, we propose a more convenient tabulation by deriving expressions for the "diffusion coefficients" of PCs.

We write the diffusive fluxes for θ and Ψ as: $J_\psi = \rho D_\psi \nabla\psi$ and $\mathbf{J}_\theta = \rho\mathbf{D}_\theta\nabla\theta$ and substitute them into the expressions for the diffusive fluxes for the full matrix $\mathbf{Q}$: $\mathbf{J}_\psi = \mathbf{Q}^T\mathbf{J}_\theta$. Here, $\mathbf{D}_\theta$ is the diagonal matrix of diffusion coefficients for thermo-chemical scalars (mass diffusivities for species and thermal diffusivity for temperature). D_ψ is the matrix of diffusion coefficients for the PCs. We get after manipulation:

$$\mathbf{D}_\psi = Q^T\mathbf{D}_\theta(Q^T)^{-1}. \tag{9}$$

This expression is an important extension of the analysis in Ref. [1] and provides a direct method of determining the diffusion coefficient of the PCs, a "property" of the PCs, given the diffusion coefficients of the thermo-chemical scalars. Therefore, such a formulation eliminates the role of spatial gradients for both PCs and thermo-chemical scalars. If only a reduced set of PCs is used, then the matrix of diffusion coefficients, only the first N_{PC} columns of $\mathbf{D}_\psi$ is used.

It is important to note a few characteristics of the diffusion coefficients matrix $\mathbf{D}_\psi$ based on Eq. (9). First, while $\mathbf{D}_\theta$ is a diagonal matrix, we expect $\mathbf{D}_\psi$ to be a full matrix. Moreover, we expect $\mathbf{D}_\psi$ to be a symmetric matrix given that the eigenvectors matrix $\mathbf{Q}$ is made up by choice by a set of orthonormal vectors. More importantly, there is no guarantee that the coefficients of $\mathbf{D}_\psi$ are positive and their sign depends on the weighting of the components of $\mathbf{Q}$. Therefore, $\mathbf{D}_\psi$ may fulfill some of the characteristics that make it, mathematically, a matrix of binary diffusion coefficients. The possibility that these coefficients may be negative does degrade their physical meaning. Nonetheless, they can be tabulated as properties of the PCs in contrast to the diffusive fluxes, which depend on spatial gradients as well.

Next, we propose an extension of the above analysis for non-linear PCA, such as in KPCA. We do not expect the linear relations between PCs and the original scalars as in linear PCA. However, this linear relation suggests that matrices $\mathbf{A}$ and $\mathbf{Q}$ serve as Jacobian matrices:

$$\mathbf{Q}^T \equiv \frac{\partial\boldsymbol{\psi}}{\partial\boldsymbol{\theta}} \quad \text{and} \quad \mathbf{A}^T \equiv \frac{\partial\boldsymbol{\psi}^{\text{red}}}{\partial\boldsymbol{\theta}}; \tag{10}$$

and this analogy is adopted for non-linear PCA, including KPCA. We define similar Jacobian matrices for non-linear PCs:

$$\widehat{\mathbf{Q}}^T \equiv \frac{\partial\boldsymbol{\psi}}{\partial\boldsymbol{\theta}} \quad \text{and} \quad \widehat{A}^T \equiv \frac{\partial\boldsymbol{\psi}^{\text{red}}}{\partial\boldsymbol{\theta}}. \tag{11}$$

Although, theoretically, one may construct as many PCs as data in KPCA, going beyond the prescribed number of thermal-chemicals scalars does defeat the purpose of KPCA. Therefore, we will take the matrix $\widehat{\mathbf{Q}}$ as having a dimension N by N (i.e. a square matrix). Matrix $\widehat{\mathbf{Q}}$ may be evaluated numerically using small perturbations of the thermo-chemical scalars vector. Matrix $\widehat{\mathbf{A}}$ is a truncation corresponding to the first N_{PC} rows of $\widehat{\mathbf{Q}}^T$.

Next we derive an expression for the PCs source term by considering the analogy of a homogeneous system, such that:

$$\frac{\partial\boldsymbol{\theta}}{\partial t} = \mathbf{S}_{\boldsymbol{\theta}} \quad \text{and} \quad \frac{\partial\boldsymbol{\psi}}{\partial t} = \mathbf{S}_{\boldsymbol{\psi}}. \tag{12}$$

We can expand the equation for the PC by writing it as follows: $\mathbf{S}_\psi = \frac{d\psi}{dt} = \frac{\partial\psi}{\partial\theta}\frac{\partial\theta}{\partial t} = \mathbf{Q}^T\mathbf{S}_\theta$. Therefore, the source term for the PCs can be determined as follows:

$$\mathbf{S}_\psi = \widehat{\mathbf{Q}}^T\mathbf{S}_{\boldsymbol{\theta}} \quad \text{and} \quad \mathbf{S}_\psi^{\text{red}} = \widehat{\mathbf{A}}^T\mathbf{S}_{\boldsymbol{\theta}}. \tag{13}$$

Although Eq. (13) is derived for a homogeneous system, the same matrix is valid for the full transport equation for PCs. More importantly, the expression for the PCs source terms is very similar to the one derived in Ref. [1], except that the matrices $\widehat{\mathbf{Q}}$ and $\widehat{\mathbf{A}}$ are not constant. Nonetheless, they can be tabulated the same way that the PCs can be tabulated as functions of the thermo-chemical scalars vector or *vice versa*.

Similarly, we proceed to determine an expression for an effective diffusion coefficients' matrix in the PCs transport equation. For ease of analysis, we also consider only a mixture subject to diffusion (i.e. reaction) and multiply the transport equation for the original thermo-scalars vector by $\widehat{\mathbf{Q}}$: $\widehat{\mathbf{Q}}^T\frac{\partial\boldsymbol{\theta}}{\partial t} = \widehat{\mathbf{Q}}^T\nabla\cdot\mathbf{J}_{\boldsymbol{\theta}}$. The left-hand side corresponds exactly to $\frac{\partial\boldsymbol{\Psi}}{\partial t}$. We also explicitly express the diffusion coefficient in the diffusive flux for $\theta : \mathbf{J}_\theta = \rho\mathbf{D}_\theta\nabla\theta$ to obtain:

$$\frac{\partial\boldsymbol{\Psi}}{\partial t} = \widehat{\mathbf{Q}}^T\nabla\cdot\mathbf{J}_{\boldsymbol{\theta}} = \widehat{\mathbf{Q}}^T\nabla\cdot[\rho\mathbf{D}_\theta(\widehat{\mathbf{Q}}^T)^{-1}\nabla\boldsymbol{\Psi}]. \tag{14}$$

For linear PCA, we can bring $\widehat{\mathbf{Q}}$ inside the divergence term on the LHS expression of Eq. (14) and recover Eq. (9) for the diffusion coefficient matrix and Eq. (2) for the PC's transport equation.

However, for non-linear PCA methods, $\widehat{\mathbf{Q}}$ is expected to vary in physical space, since it also varies in composition space. Accordingly, Eq. (2) will have to be augmented by an extra transport term. Nonetheless, we may define an effective diffusion coefficients' matrix for the PCs:

$$\mathbf{D}_{\boldsymbol{\Psi}} = \widehat{\mathbf{Q}}^T\mathbf{D}_\theta(\widehat{\mathbf{Q}}^T)^{-1}. \tag{15}$$

This matrix may (1) be full, (2) have negative diffusion coefficients, and (3) be non-symmetric.

4. Results

4.1. Computational data and reduction process

The present results provide a validation of the proposed KPCA-ANN parameterization approach. The "training" data on which the KPCA is implemented is generated for the Sandia piloted methane-air Flame F [9] using the stand-alone ODT model. Details of the model formulation and the run conditions can be found in Refs. [11] and [12]. The ODT model is based on a deterministic implementation of reaction and diffusion and a stochastic implementation of turbulent advection in a space- and time-resolved simulation on a 1D domain [11]. For jet flames, the 1D domain corresponds to the transverse direction of the mean flow; while, the temporal evolution of the 1D profile for the streamwise momentum, energy and the species equations is interpreted as a downstream evolution of the jet transverse profiles. The governing equations are:

- The streamwise momentum equation:

$$\frac{\partial u}{\partial t} = \frac{1}{\rho}\frac{\partial}{\partial y}\left(\mu\frac{\partial u}{\partial y}\right) + \Omega_u, \tag{16}$$

- The species equation:

$$\frac{\partial Y_k}{\partial t} = -\frac{1}{\rho}\frac{\partial}{\partial y}(\rho V_k Y_k) + \frac{\dot{\omega}_k}{\rho} + \Omega_k, \tag{17}$$

- The temperature equation:

$$\frac{\partial T}{\partial t} = \frac{1}{\rho\bar{c}_p}\sum_{k=1}^{N} c_{p,k} Y_k V_k \frac{\partial T}{\partial y} + \frac{1}{\rho\bar{c}_p}\frac{\partial}{\partial y}\left(\lambda\frac{\partial T}{\partial y}\right) - \frac{1}{\rho\bar{c}_p}\sum_{k=1}^{N} h_k\dot{\omega}_k + \Omega_T \tag{18}$$

In the above equations, all the symbols have their usual meaning. The thermodynamic pressure, p, is assumed to be spatially uniform, and the equation of state:

$$p = \rho R_u T\sum_{k=1}^{N}(Y_k/W_k), \tag{19}$$

is used to compute the mixture density, ρ. Eqs. (16)–(18) represent a temporal solution of a turbulent jet flame. The temporal evolution of the solution represents a downstream evolution of the spatial profiles of the solution vector. The temporal evolution is interpreted as a downstream spatial evolution (in x) of the 1D velocity and scalar profiles using the following equations [11]:

$$\bar{u} - u_\infty = \frac{\int_{-\infty}^{+\infty}\rho(u-u_\infty)^2 dy}{\int_{-\infty}^{+\infty}\rho(u-u_\infty)dy} \quad \text{and} \quad x(t) = \int_0^t \bar{u}(t')dt' \tag{20}$$

where $\bar{u}$ and u_∞ correspond, respectively, to a bulk velocity and the co-flow velocity.

Turbulent advection, represented through the terms Ω_u, Ω_k and Ω_T in Eqs. (16)–(18), is implemented stochastically using stirring events, each involving the application of a 'triplet map' [11]. The frequency of stirring events is governed by the spatially-resolved evolving rate of shear in the jet. Two adjustable parameters, the so-called A and β [11], are identical to previous values used in jet configurations with ODT.

1D profiles for species and temperature are collected for different times (interpreted as downstream evolutions) and for 50 realizations. The different realizations are obtained using different sequences for samplings associated with the stochastic stirring events.

The ODT data was designed to mirror experimental data by Barlow and Frank [9], which corresponds to statistics at downstream distances of x/d = 1, 2, 3, 7.5, 15, and, 30, where d is the jet diameter. This range covers conditions of piloted stabilization, extinction and reignition. The fuel jet is made up of 50/50 by volume methane and nitrogen. Different stages of the jet flame evolution are present at different downstream distances. The near-field data corresponds to pilot stabilization at x/d = 1, 2, 3 and 7.5. At x/d = 15, there is significant extinction, while at x/d = 30 reignition has already occurred. Therefore, the challenge is to capture all these conditions within a computed set of PCs.

The chemistry is based on the 12-step augmented reduced mechanism (ARM) [13], consisting of 16 species, H_2, H, O_2, OH, H_2O, HO_2, H_2O_2, CH_4, CO, CO_2, CH_2O, C_2H_2, C_2H_4, C_2H_6 and N_2. Transport properties and the reaction mechanism are computed using the CHEMKIN library [14–16]. Therefore, the ODT stand-alone 1D solutions are based on the full set of the above species, temperature and streamwise velocity.

The original manifold from the ODT solution is spanned by the 17 dimensional thermo-chemical scalars' vector, including, temperature and the mass fractions of the species in the 12-step reduced mechanism. For the present study, a subset of 7 variables of the original manifold is selected, $\theta = (T, Y_{H_2}, Y_{O_2}, Y_{H_2O}, Y_{CH_4}, Y_{CO}, Y_{CO_2})$ [7]. The remaining scalars are not used for the reduction process; however, they are tabulated during the reconstruction process to validate the KPCA-ANN approach. The selection of a subset of the original thermo-chemical scalars' vector will help in reducing some of the complexity related to inter-dependencies of the various scalars. However, it is possible to develop a more quantitative reasoning for the reduction process and the choice of the subset variables. For example, the idea of splitting the thermo-chemical scalars' vector into "represented" and "unrepresented" variables has been proposed in Refs. [17,18]. Another approach to identify a reduced set of represented variables has been proposed in Ref. [6]. Both methods may use PCA directly to obtain a subset of the original thermo-chemical scalars' vector. Such approaches may yield either a smaller or larger subset based potentially on criteria for optimum representation of the composition space variables. However, neither approach is attempted here. In our present study, the choice of the representative variables is made based on their being major species or temperature and a choice of representative variables on the fuel, oxidizer, products and intermediates species.

The ODT data corresponds to 50 realizations corresponding to "radial" profiles of the reduced set of thermo-chemical scalars at 4 downstream distances, x/d = 3, 7.5, 15, and 30, where d is the fuel jet diameter. The different realizations are obtained due to the stochastic implementation of turbulent transport. Before carrying out the KPCA, the original data variables is normalized to yield a range between -1 and 1 using the minimum and maximum values for each variables. A different choice of normalization will yield different PCs and values for the transport terms. We believe that the normalization is one potential option for optimizing the KPCA procedure; and this optimization will be explored further in future work.

Although not presented here, the use of a single realization out of the 50 available for the determination of the eigenvalues and eigenvectors of the kernel matrix yields the same results as the use of all realizations. This observation is clearly convenient, because the size of the eigenvalues problem increases with the data size; however, there are other strategies that can be explored to parse the data as well. The data based on one realization corresponds to 3604 discrete grid points, with each grid point containing the values of temperature and the represented species mass fractions of the reduced set. The eigenvalues and eigenvectors of the kernel matrix are determined based on this data set. However, we construct the ANN tabulation using a larger subset of the computational data (17 realizations) using the evaluation procedure given in Eq. (8). Validation of the KPCA-ANN procedure is implemented on the remaining realizations (30 realizations). However, the results presented below only show one realization comparisons, since similar results are obtained for other

realizations. The ANN tabulation is implemented on this data set as well using the ANN Toolbox in MATLAB.

In the present work and similar to Ref. [7], the same ANN topology is used for all thermo-chemical scalars of the reduced set. It has a number of inputs equal to the number of PCs (1 or 2 in the present study), 1 hidden layer composed of 20 neurons with a tangent sigmoid activation function, and an output layer of one neuron with a pure linear function.

4.2. Thermo-chemical scalars' profiles

In this section, we show comparisons of reconstructed and original ODT profiles of 4 thermo-chemical scalars, which are representative of the remaining scalars, O_2 mass fraction and temperature, which are present in the reduced set and OH and C_2H_6 mass fractions, which are part of the extended set of thermo-chemical scalars. Also, we illustrate the extent of predictions of O_2 and temperature profiles for 1 and 2 PCs to assess the extent at which the number of PCs can improve the predictions.

Figures 1 and 2 show profiles of Y_{O_2} temperature, respectively, based on the KPCA-ANN analysis (dashed) and the original profiles (solid) corresponding to a single realization of ODT simulations at 4 different downstream conditions. The profiles are based on 1 PC. The two profiles for O_2 are in reasonable qualitative agreement; but, there are some obvious quantitative agreements as well. Nonetheless, some peak values and their spatial locations are reasonably captured. A much larger discrepancy is present for the temperature profiles.

Now, we implement the same comparisons shown in Figs. 1 and 2, but using 2 PCs. The results of such comparisons are shown in Figs. 3 and 4. Figs. 3 and 4 clearly show an excellent agreement between the KPCA-ANN and original O_2 mass fraction and temperature profiles, indicating that 2 PCs are adequate to reproduce the original thermo-chemical scalars. Using linear PCA, we have found that 3 PCs were necessary to reproduce O_2 mass fraction profiles [7].

Next, we investigate if species not included in the reduced subset are adequately represented by 2 PCs. Figs. 5 and 6 show profiles of OH and C_2H_6 mass fraction, based on the KPCA-ANN analysis (dashed) and the original profiles (solid) corresponding to a single realization of ODT simulations at 4 different downstream conditions. The profiles are based on 2 PCs. The KPCA-ANN and original profiles are again in excellent agreement for OH, suggesting that 2 PCs are equally adequate to represent OH mass fraction. For C_2H_6 and other "minor species" (not shown here), there is still some discrepancy between the KPCA-ANN profiles similar to the major species observations with 1 PC. Further refinement of the implemented KPCA-ANN procedure may be considered in the future. They include: (1) considerations of more PC's for minor species, (2) refinement of the list of the subset of scalars used for KPCA-ANN reduction, and (3) fine-tuning of the size of the sample used for the KPCA eigenvalue problem and for ANN tabulation.

Nonetheless, it is important to re-iterate that similar predictions for major species using linear PCA required 3 principal components instead of 2. Moreover, with linear PCA, a single realization was not adequate to reduce the composition space with the PCA-ANN approach. Invariably, KPCA-ANN is more expensive; but, the cost of both procedures is insignificant and can be completed within one hour using Matlab software on a stand-alone workstation.

4.3. Transport terms

In this section, we present results on the PCs source terms and diffusion coefficients' matrix $\mathbf{D}_{\boldsymbol{\Psi}}$. As discussed earlier, both contributions depend on the coefficients of the Jacobian matrix, $\widehat{\mathbf{Q}}^T$. This matrix is determined numerically, by perturbing variables in the thermo-chemical scalars' vectors and evaluating the corresponding changes in the PCs. For species mass fractions, the remaining mass fractions are normalized after perturbation of one of the species to maintain a sum of mass fractions equal to unity. Perturbations of the thermo-chemical scalars 0.1–0.5% of their means yield reasonably converged values for the Jacobian matrix coefficients.

Figure 7 shows profiles of the source term for the first PC, S_{ψ_1}, based on the KPCA-ANN analysis corresponding to a single realization of ODT simulations at 4 different downstream conditions. At

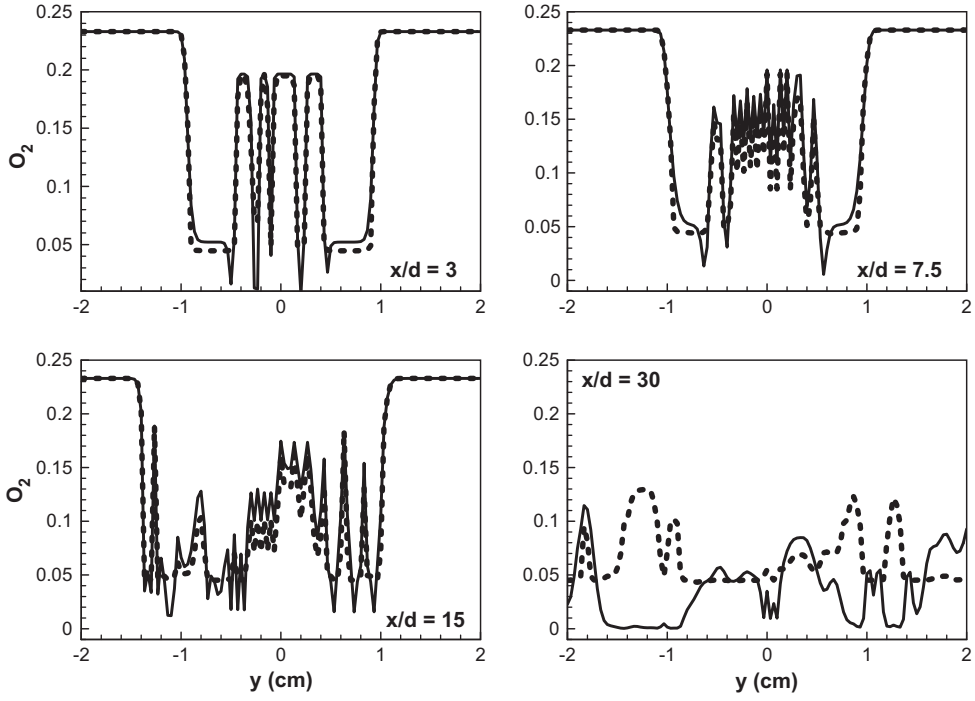

Fig. 1. Comparison of Y_{O_2} profiles between KPCA-ANN (dashed) and ODT data (solid), based on 1 PC.

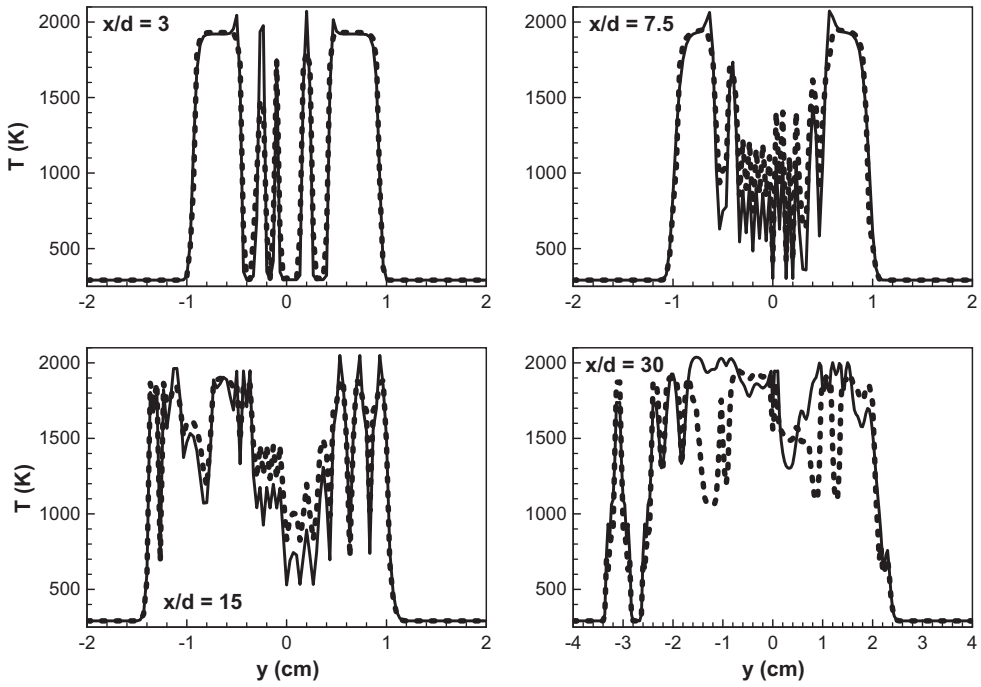

Fig. 2. Comparison of T profiles between KPCA-ANN (dashed) and ODT data (solid), based on 1 PC.

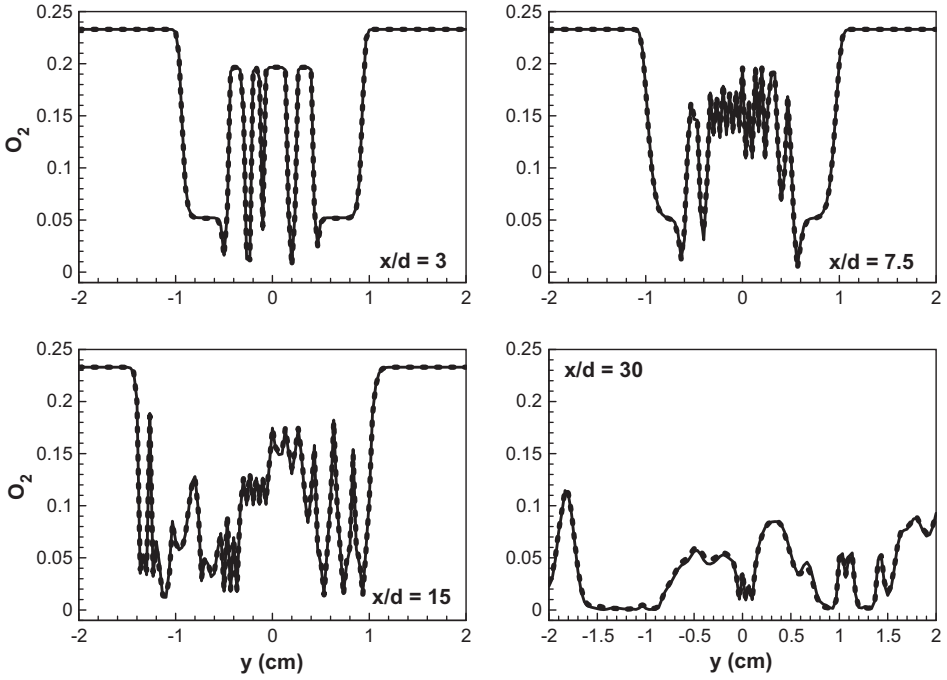

Fig. 3. Comparison of Y_{O_2} profiles between KPCA-ANN (dashed) and ODT data (solid), based on 2 PCs.

present, there is no independent validation for this PC. Therefore, they are shown to illustrate the overall shapes of the PC sources. As can be seen in the profiles, the PC's source term does mirror variations in PC's profiles in Figs. 3–6. More importantly, and, although not shown here, the spatial variations of the PC's source terms are not much different from those of the original thermo-chemical scalars' source term profiles. This may be largely attributed to the fact that chemical reaction, which is responsible for both source terms, are spatially localized phenomena.

Eqs. (21) and (22) below illustrate the values of the diffusion coefficient matrix for the thermo-chemical scalars, $\mathbf{D_0}$, and the "effective diffusion coefficients" matrix, $\mathbf{D_\Psi}$ in the oxidizer stream (i.e. valid at any realization and any downstream conditions). Both matrices can have different values at different conditions of the composition space. More importantly, the values of the coefficients of $\mathbf{D_\Psi}$ can be different if the original thermo-chemical scalars are weighted differently, say by adopting a different normalization for each variable. This normalization, also, can be a way to

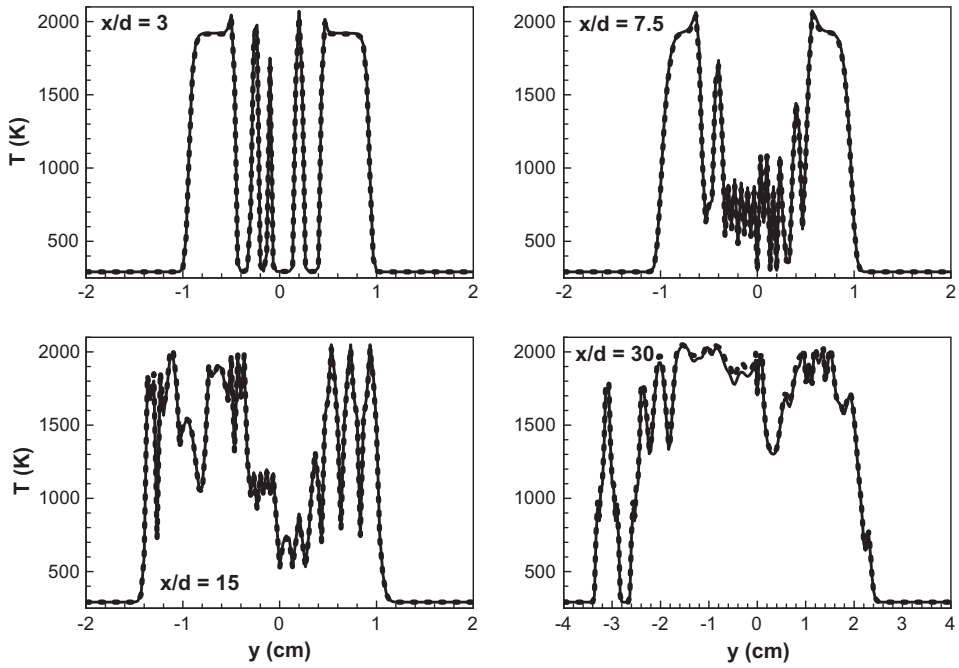

Fig. 4. Comparison of T profiles between KPCA-ANN (dashed) and ODT data (solid), based on 2 PCs.

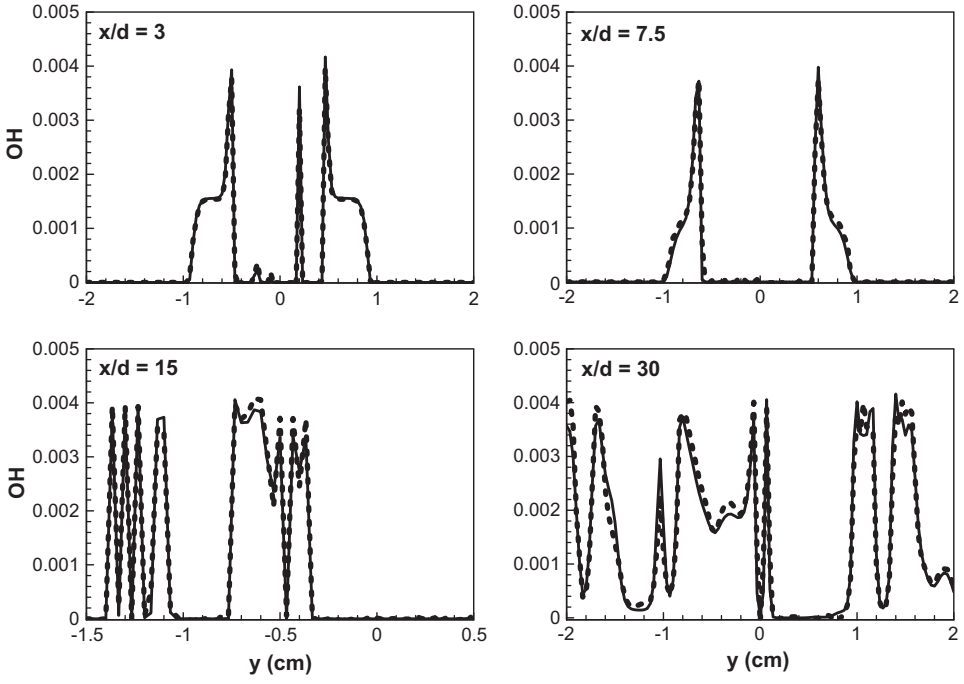

Fig. 5. Comparison of Y_{OH} profiles between KPCA-ANN (dashed) and ODT data (solid), based on 2 PCs.

optimize the KPCA procedure and improve the tabulation of the PCs source terms and diffusion coefficients.

$$\mathbf{D}_\theta = \begin{bmatrix} 0.212273 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.762568 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.192912 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.213336 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.213439 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.196959 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.149026 \end{bmatrix} \tag{21}$$

and

$$\mathbf{D}_\Psi = \begin{bmatrix} -46.8386 & -57.5962 & 151.1541 & 50.26939 & 27.6615 & -13.4559 & -9.59522 \\ -4.14429 & -4.82771 & 13.30394 & 4.426652 & 2.393772 & -1.15204 & -0.83706 \\ 10.34534 & 12.82217 & -33.0445 & -11.0353 & -6.28262 & 3.142959 & 2.124687 \\ -69.1159 & -85.0233 & 222.1006 & 74.00845 & 41.16221 & -20.2494 & -14.1421 \\ -5.61435 & -6.86064 & 18.01682 & 5.994048 & 3.496475 & -1.60707 & -1.11791 \\ -5.19827 & -6.40631 & 16.69224 & 5.55042 & 3.1112 & -1.33581 & -1.0405 \\ 50.39791 & 61.91443 & -161.917 & -53.8158 & -29.9096 & 14.68809 & 10.4822 \end{bmatrix} \tag{22}$$

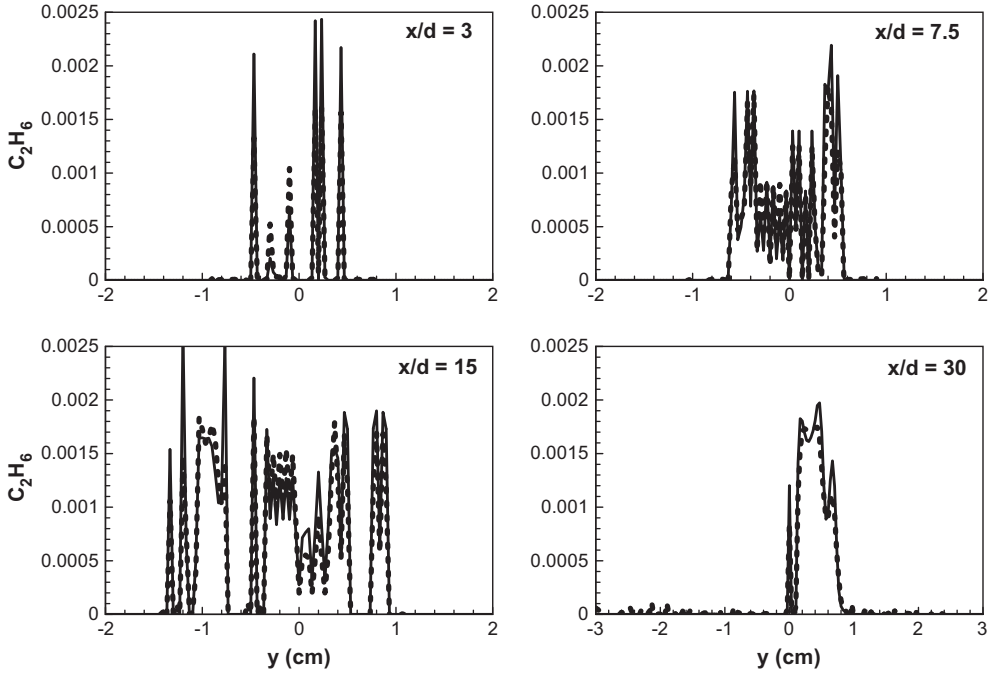

Fig. 6. Comparison of $Y_{C_2H_6}$ profiles between KPCA-ANN (dashed) and ODT data (solid), based on 2 PCs.

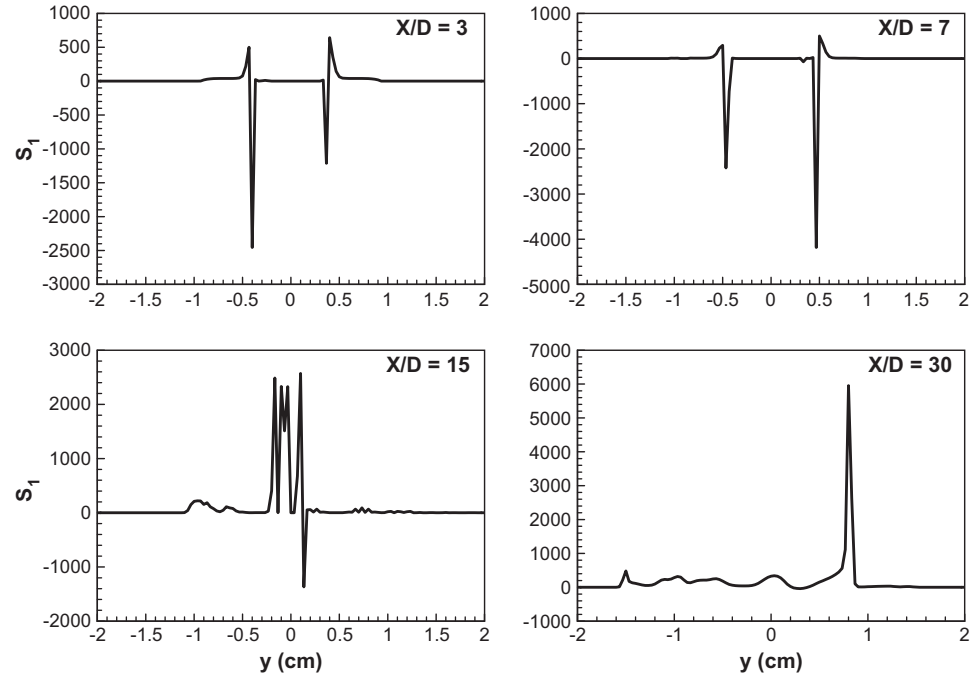

Fig. 7. 1st PC's source term profile at different downstream distances.

By design, $\mathbf{D_\theta}$ is a diagonal matrix. The diffusion coefficients' values also indicate that for the mixture, most diffusion coefficients are similar in value, except for the value of the diffusion coefficient of hydrogen, H_2 (the second row). In contrast with $\mathbf{D_\theta}$, $\mathbf{D_\Psi}$ is a full matrix and includes negative coefficients. The matrix, also, is not symmetric in contrast to the common binary diffusion coefficients' matrices. More importantly, the magnitudes of the effective diffusion coefficients can exceed those of the thermo-chemical scalars. The minor differences in the diffusion coefficients of the thermo-chemical scalars, except for H_2, can be magnified by the operation in Eq. (15). This magnification reflects the inherent complexity related to the transport of scalars that are functions of the transport of a combination of scalars. Another artifact of this complexity is the generation of negative diffusion coefficients. Although, the effective diffusion coefficients $\mathbf{D_\Psi}$ are meaningful in the mathematical sense for the transport of the prin-

cipal components, they are devoid of any simple physical meaning that characterizes the diffusion coefficients of the original thermo-chemical scalars. Nonetheless, they can be tabulated and used within the context of non-linear PCA methods.

5. Concluding remarks

A KPCA-ANN approach is proposed as a non-linear reduction algorithm for turbulent combustion composition space. The procedure, in principle, is better adapted than PCA-ANN to address non-linear manifolds for this composition space. The use of KPCA has resulted in a lower number of PCs than linear PCA for the reconstruction of data from turbulent non-premixed flames subject to extinction and re-ignition. Strategies to reconstruct source and transport terms for the resulting PCs are needed to make the proposed KPCA-ANN approach viable for turbulent combustion simulations. Moreover, while the KPCA-ANN framework yields satisfactory predictions of major species, further refinement of the algorithm may be needed to improve the predictions of minor species.Another potential shortfall of linear or non-linear variations of PCA is the potential loss of physical meaning associated with principal components. The methods offered in Refs. [3,6,17,18] provide a possible path to relate PCs, but not individually, to a reduced subset of the original thermo-chemical vector. Contributions of the various thermo-chemical scalars to a particular PC can, at certain conditions, also identify the dominance of the contribution of a particular scalar or a subset of the thermo-chemical scalars' vector. Despite the above observations, KPCA (and other manifold-based methods for turbulent combustion composition space) can offer a potential strategy for accelerating combustion calculations due to at least 2 potential advantages. The first is the obvious one, which corresponds to a reduction in the number of transported scalars from the total set of thermo-chemical scalars to the number of principal components. However, perhaps the most important saving may come from our choice of a subset of scalars that can eliminate species involved in the fastest reactions at the origin of the mechanism stiffness.The study also includes a formulation for the PCs source terms and an effective diffusion coefficient matrix. Both contributions can be tabulated as functions of the PCs potentially enabling the solution of transport equations for PCs instead of thermo-chemical scalars. Some attributes of the coefficients of the PCs effective diffusion matrix are fundamentally different from the common attributes for thermo-chemical scalars. They include: (1) negative diffusion coefficients (i.e. counter-gradient diffusion), (2) the non-symmetric nature of the matrix, and (3) the amplification of the diffusion coefficients relative to that of the thermo-chemical scalars. Nonetheless, the matrix remains useful for the transport of PCs within the context of linear or non-linear PCA.

Acknowledgments

The work was supported by the National Science Foundation Computational Mathematics Program under grant DMS- 1217200.

References

[1] J.C. Sutherland, A. Parente, Proc. Combust. Inst. 32 (2009) 1563–1570.
[2] A. Parente, J.C. Sutherland, L. Tognotti, P.J. Smith, Proc. Combust. Inst. 32 (2009) 1579–1586.
[3] A. Parente, J.C. Sutherland, B.B. Dally, L. Tognotti, P.J. Smith, Proc. Combust. Inst. 33 (2011) 3333–3341.
[4] A. Bilgari, J.C. Sutherland, Combust. Flame 159 (2012) 1960–1970.
[5] A. Coussement, O. Gicquel, A. Parente, Combust. Flame 159 (2012) 2844–2855.
[6] A. Coussement, O. Gicquel, A. Parente, Proc. Combust. Inst. 34 (2013) 1117–1123.
[7] H. Mirgolbabaei, T. Echekki, Combust. Flame 160 (2013) 898–908.
[8] B. Scholkopf, A.J. Smola, K. Muller, Neural Comput. 10 (1998) 1299–1399.
[9] R.S. Barlow, J.H. Frank, Proc. Combust. Inst. 27 (1998) 1087–1095.
[10] L.J.P. van der Maaten, E.O. Postma, H.J. van den Herik, TiCC-TR 2009-005, 2009.
[11] T. Echekki, A.R. Kerstein, T.D. Dreeben, Combust. Flame 125 (2001) 1083–1105.
[12] B. Ranganath, T. Echekki, Combust. Sci. Technol. 181 (2009) 570–596.
[13] C.J. Sung, C.K. Law, J.-Y. Chen, Combust. Flame 125 (2001) 906–919.
[14] R.J. Kee, J.A. Miller, T.H. Jefferson, CHEMKIN: A General-Purpose, Problem-Independent, Transportable, Fortran Chemical Kinetics Code Package. Sandia National Laboratories, Report SAND80-8003, 1980.
[15] R.J. Kee, J. Warnatz, J.A. Miller, A FORTRAN Computer Code Package for the Evaluation of Gas-Phase Viscosities, Conductivities, and Diffusion Coefficients. Sandia National Laboratories, Report SAND83-8209, 1983.
[16] R.J. Kee, J.F. Grcar, M.D. Smooke, J.A. Miller, A FORTRAN Program for Modeling Steady Laminar One-Dimensional Premixed Flames. Sandia National Laboratories, Report SAND85-8240, 1985.
[17] Z. Ren, G.M. Goldin, V. Hiremath, S.B. Pope, Combust. Theor. Model. 15 (2011) 827–848.
[18] S.B. Pope, Proc. Combust. Inst. 34 (2013) 1–31.