Combustion and Flame 162 (2015) 4025–4035

Contents lists available at ScienceDirect

Combustion and Flame

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

An a-posteriori evaluation of principal component analysis-based models for turbulent combustion simulations

Amir Biglari*, James C. Sutherland

Department of Chemical Engineering, University of Utah, 50 S. Central Campus Dr. Rm 3290, Salt Lake City, UT 84112-9203, United States

ARTICLE INFO

Article history:
Received 11 April 2015
Revised 29 July 2015
Accepted 29 July 2015
Available online 17 August 2015

Keywords:
Principal component analysis (PCA)
Dimension reduction techniques
Turbulent combustion modeling
State-space parameterization

ABSTRACT

Recently, principal component analysis (PCA) has been proposed as a means to identify and parameterize manifolds existing in turbulent reacting flow. PCA provides a way to systematically add dimensionality to state-space parameterizations without resorting to ad-hoc progress variables. This work presents an *a posteriori* study of PCA as a combustion model applied to a non-premixed CO/H2 temporally evolving jet flame with extinction and reignition. As a basis for comparison, results from detailed chemistry calculations are compared with the PCA-transport results to verify the model and evaluate its performance. The effect of increasing the number of retained principal components (PC's) is shown by comparing the results of three different cases retaining 1, 2, and 3 PC's. Invariance of the model's error to a system parameter (Reynolds number) is evaluated over a range of Re numbers from 2500 to 9000.

1. Introduction

There are two general approaches to reduce the state-space dimensionality in combustion problems. The first approach employ reduction techniques such as the quasi steady state approximation (QSSA) [1], or rate controlled constrained-equilibrium (RCCE) [2,3]. The second approach attempts to identify controlling parameters and a function that maps the state of the system to these parameters.

Traditional state-space parameterization methods postulate some of the variables in the system as controlling variables and then parameterize the state variables based on them; therefore, the only equations to be solved are the transport equations for a new smaller set of controlling variables. The Steady Laminar Flamelet Method (SLFM) [4] is an example of these methods using mixture fraction and dissipation rate as the controlling variables. One disadvantage of most parameterization methods is that they are not easily extensible; adding additional parameters are not straightforward. Techniques such as Flamelet-Generated Manifold (FGM) [5–7] and Flamelet-Prolongation of ILDM model (FPI) [8–10] propose ad-hoc progress variables, but suffer from the same problem of identifying the 'best' set of parameters. As additional parameters are added to a model, these should be orthogonal to one another to optimally represent additional manifold dimensions. Additionally, these parameters should represent as much of the variation in the system as possible in order to maximize the effectiveness of each parameterizing variable.

The aforementioned criteria have guided recent work on dimension reduction techniques using principal component analysis (PCA). PCA has been shown as an effective methodology to identify manifolds in turbulent combustion [11–13]. One of the distinguishing features of a PCA-based combustion model is that one can select the number of parameters to achieve a desired accuracy of the model, and that each additional parameter explains the most remaining variance in the original data as possible [11,13,14].

There have been several *a priori* studies on PCA-based models for turbulent combustion simulations. One key outcome of these studies is the recognition that, while PCA provides a basis (a linear rotation of the original state) and a means to truncate this basis (providing the optimal low-dimensional representation), the state variables are still nonlinear functions of the new basis, and nonlinear regression is required [14–19].

Additionally, nonlinear parameterization methods based on PCA have been explored [12,13,20–22]. Most of these, however, result in a reduced representation that effectively precludes incorporation in a CFD framework since the transport equations for the resulting parameters are very complex. A noteworthy exception is the Manifold-Generated Local PCA approach proposed by Coussement et al. [20].

Relatively few *a posteriori* studies of PCA-based models have been undertaken. Several have been done based on variations of the original formulation proposed in [11] including work by Coussement et al. [20] Isaac et al. [23]. Recent work by Mirgolbabaei and Echekki

* Corresponding author.
E-mail addresses: amir.biglari@utah.edu, amir_bgi@yahoo.com (A. Biglari), james.sutherland@utah.edu (J.C. Sutherland).

http://dx.doi.org/10.1016/j.combustflame.2015.07.042
0010-2180/

[24] showed an *a posteriori* demonstration based directly on the originally proposed formulation [11].

In the present work, we use the approaches outlined in [11,14,25] to demonstrate a PCA-based turbulent combustion model which transports the PC's directly. The model is trained on simulations with detailed chemistry involving 11 species, and we explore reductions to 1, 2, and 3 principal components. Multivariate Adaptive Regression Spline (MARS) is used for regressing source terms and regenerating state-variables from PC's to capture nonlinearity in the reduced basis [14,15]. We also present special considerations for PCA models implemented within one-dimensional turbulence codes. Results of *a posteriori* studies on a turbulent nonpremixed simulation involving extinction and reignition are presented, together with a study on the effect of parameters such as the Reynolds number, the scaling method used in training the PCA, the number of retained PC's, and the training data itself. The results show that PCA-based models can effectively predict turbulent reacting with significant finite-rate chemistry effects.

2. Formulation

PCA provides a linear transformation of basis from the original thermodynamic basis[1] $\vec{\phi} = \{T, y_1, \ldots, y_n\}$ to a new basis defined by the principal components (PCs), $\vec{\eta}$. As discussed in the introduction, the state variables can then be regressed onto this basis via any number of nonlinear regression techniques to obtain $\hat{\phi}_i = \mathcal{G}_i(\vec{\eta})$, where $\hat{\phi}_i \approx \phi_i$. This process may be summarized as

$$\vec{\phi} \underset{\text{linear mapping}}{\overset{\text{PCA}}{\longrightarrow}} \vec{\eta} \underset{\text{nonlinear mapping}}{\overset{\text{MARS}}{\longrightarrow}} \phi_i \approx \hat{\phi}_i = \mathcal{G}_i(\vec{\eta}). \tag{1}$$

2.1. Principal component analysis

Here, we briefly summarize the key elements in the PCA transformation and refer the reader elsewhere for further details [11,13,14,25]. We consider a data set $[\phi]$ consisting of m observations of N_ϕ state variables such that $[\phi]$ is a matrix of size $(N_\phi \times m)$. Prior to carrying out a PCA transformation on data, $[\phi]$ is centered and scaled appropriately [13,14,25]. The PCA transformation then defines the principal components (PCs) as

$$[\eta] = [A^\top][S]([\phi] - [M]), \tag{2}$$

where $[M]$ is an $(N_\phi \times m)$ matrix that repeats the mean value of each variable, ϕ_i, in its corresponding row, $[S]$ is an $(N_\phi \times N_\phi)$ diagonal matrix containing the scaling factors for each ϕ_i and $[A]$ are the eigenvectors of the covariance matrix of the centered and scaled data, $[S]([\phi] - [M])$.

The original data can be recovered by simply inverting Eq. (2) and noting that $[A^{-1}] = [A^\top]$,

$$[\phi] = [S^{-1}][A][\eta] + [M]. \tag{3}$$

Eq. (2) is simply a rotation of the original basis $[\phi]$ to the PC basis, $[\eta]$, and does not reduce the dimensionality of the system. However, the eigenvalues of $[S]([\phi] - [M])$ indicate which eigenvectors provide the most information, and form a logical means to select a reduced basis from the eigenvectors. The actual dimension reduction in PCA happens when we choose N_η of the eigenvectors, $[A_s]$, where N_η is the number of PC's that is intended to retain, so that Eq. (2) is written as

$$[\eta] = [A_s^\top][S]([\phi] - [M]). \tag{4}$$

Here $[\eta]$ contains PC vectors, $\vec{\eta}$, of size N_η, where $N_\eta \ll N_\phi$ for strongly attractive manifolds in state space.

[1] Here we omit pressure from the basis as we consider applications in the low-Mach regime. For high speed flows, pressure should be included in the analysis.

In the reduced basis, the reconstruction of $[\phi]$ through simple inversion of Eq. (4) can be problematic due to nonlinearities in ϕ_i in the reduced basis. For this reason, multivariate nonlinear regression techniques are employed to generate $\phi_i \approx \hat{\phi}_i = \mathcal{G}_i(\vec{\eta})$. This approach has been shown to reconstruct state-variables with reasonable accuracy [12,14,16–19].

2.2. Formulation of PC transport equations

As outlined first in [11], one advantage of the linearity of PCA is that transport equations for the PCs are readily obtained. Beginning with the scalar transport equations, we may apply the linear transformation defined in Eq. (4) to obtain the PC transport equations. This is summarized as:

$$\rho \frac{D\vec{\phi}}{Dt} = -\nabla \cdot \vec{\mathbf{j}}_\phi + \vec{s}_\phi, \tag{5}$$

$$\Downarrow [A_s^\top][S]$$

$$\rho \frac{D\vec{\eta}}{Dt} = -\nabla \cdot \vec{\mathbf{j}}_\eta + \vec{s}_\eta. \tag{6}$$

where $\vec{\mathbf{j}}_\eta = [A_s^\top][S]\vec{\mathbf{j}}_\phi$ are the diffusive fluxes of the PCs and $\vec{s}_\eta = [A_s^\top][S]\vec{s}_\phi$ are the PC source terms. We now turn our attention to details on treatment of $\vec{\mathbf{j}}_\eta$ and $\vec{s}_\eta$ appearing in Section 2.2.

Neglecting radiation, the temperature transport equation included in Section 2.2 can be written as

$$\rho c_p \frac{DT}{Dt} = -\nabla \cdot \mathbf{q} + \sum_{i=1}^{n} (h_i \nabla \cdot \mathbf{j}_{y_i} - h_i s_{\rho y_i}), \tag{7}$$

where y_i is the composition of species i, h_i is the enthalpy of species i, c_p is the heat capacity of the mixture, λ is the thermal conductivity of the mixture and $\mathbf{q} = -\lambda \nabla T + \sum_{i=1}^{n} h_i \mathbf{j}_{y_i}$ is the heat flux. Expanding Eq. (7) and simplifying, we find

$$\rho \frac{DT}{Dt} = -\frac{1}{c_p}\left[\nabla \cdot (-\lambda \nabla T) + \sum_{i=1}^{n} (\mathbf{j}_{y_i} \cdot \nabla h_i + h_i s_{\rho y_i})\right]. \tag{8}$$

The first term in the right-hand-side of Eq. (8) is the classical Fourier's law of conduction term, while the second term accounts for the temperature changes due to reaction and species diffusion.

Applying the linear PCA transformation to the species and temperature equations, we arrive at the set of transport equations for the principal components,

$$\rho \frac{D\vec{\eta}}{Dt} = \nabla \cdot [A_s^\top][S] \begin{pmatrix} \frac{\lambda}{c_p}\nabla T \\ \rho D_{y_1} \nabla y_1 \\ \vdots \\ \rho D_{y_{ns-1}} \nabla y_{ns-1} \end{pmatrix} + [A_s^\top][S] \begin{pmatrix} -\frac{1}{c_p}\sum_{i=1}^{n} (\mathbf{j}_{y_i} \nabla h_i + h_i s_{\rho y_i}) \\ s_{\rho y_1} \\ \vdots \\ s_{\rho y_{ns-1}} \end{pmatrix}. \tag{9}$$

The treatment of diffusive fluxes in Eq. (9) will be discussed in the next section. We first consider treatment of the source terms appearing in Eq. (9). The term involving $\mathbf{j}_{y_i} \nabla h_i$ would require special modeling treatment. However, we have found that this term has negligible impact on the evolution of $\vec{\eta}$ and can safely be neglected in turbulent flows. Therefore, we need only consider the remaining terms in the PC sources, which are all point-wise quantities.

As first proposed in [14,15] and later employed in [16,17], nonlinear regression can be employed to construct the nonlinear functions $\phi \approx \hat{\phi}_i = \mathcal{G}_i(\vec{\eta})$ and $s_{\eta_j} \approx \hat{s}_{\eta_j} = \mathcal{S}_j(\vec{\eta})$. Here, we use Multivariate

Adaptive Regression Splines (MARS) [26–28]. Although other regression techniques are also suitable (see the introduction for a summary of techniques), MARS can affordably be used on very large training data sets and is, therefore, our method of choice.

The PCA modeling approach can be succinctly summarized as

1. From training data, the PCA transformation defined in Eq. (4) is obtained.
2. MARS is applied to build models for $\phi \approx \hat{\phi}_i = \mathcal{G}_i(\vec{\eta})$ and $s_{\eta_j} \approx \hat{s}_{\eta_j} = \mathcal{S}_j(\vec{\eta})$ from the training data.
3. The conservative form of the PC transport equations is solved,

$$\frac{\partial \rho \vec{\eta}}{\partial t} = -\nabla \cdot \rho \vec{\eta} \mathbf{u} - \nabla \cdot \vec{\mathbf{j}}_\eta + \vec{s}_\eta. \tag{10}$$

We now turn our attention to the diffusive fluxes appearing in Eq. (9).

2.2.1. PC diffusive fluxes

The diffusion coefficients for $\vec{\phi}$ are given by Eq. (9) as $\vec{D}_\phi = \{\lambda/\rho c_p, D_{y_1}, \ldots, D_{y_{ns-1}}\}$. Diffusive fluxes for the PCs are defined as

$$\vec{\mathbf{j}}_\eta = [A_s^\top][S]\vec{\mathbf{j}}_\phi = -[A_s^\top][S]\rho[D_\phi]\nabla\vec{\phi}, \tag{11}$$

where $[D_\phi] = \text{diag}(\vec{D}_\phi)$ is a diagonal matrix containing state variables diffusion coefficients. A naive approach to computing these would require reconstruction of the full state vector, $\vec{\phi}$, followed by direct application of Eq. (11). Instead of this, we consider two approximations to $\vec{\mathbf{j}}_\eta$:

1. Use the technique proposed in [21], which uses a similarity transform is used to obtain the PC fluxes as

$$\vec{\mathbf{j}}_\eta = -\rho[D_\eta]\nabla\vec{\eta}, \qquad [D_\eta] = [A_s^\top][D_\phi][A_s], \tag{12}$$

   where $[D_\eta]$ is a full non-symmetric matrix that can have negative values.
2. Approximate $[D_\eta]$ by a weighted averaged of state variables based on their weight in each corresponding eigenvector:

$$\vec{\mathbf{j}}_\eta = -\rho[D_\eta]\nabla\vec{\eta}, \qquad [D_\eta] = [D_\phi][A_s][\mathcal{N}], \tag{13}$$

   where $\mathcal{N}$ is a diagonal matrix containing the inverse of the sum of the components of the ith eigenvector to ensure that $[A_s][\mathcal{N}]$ is orthornormal. This method weights each diffusion coefficient by its importance in each principal component estimate the PC diffusion coefficient. The rationale behind this approximation is to obtain averaged PC diffusion coefficients with the same linear combination used to calculate PC's themselves.

Although method 1 may be more accurate in principal, it is more expensive than method 2, and testing has shown little appreciable difference in results. Therefore, we use method 2 in the results presented below.

2.3. One-dimensional turbulence simulation

The PCA model discussed above has been implemented in a one dimensional turbulence (ODT) simulation code [29] to conduct *a posteriori* studies. We select ODT rather than DNS or LES for the following reasons:

- ODT has been shown to capture the essential characteristics of DNS simulations at a fraction of the cost [29]. This allows parametric simulations even when incorporating detailed kinetics which would be cost-prohibitive with DNS.
- ODT does not require closure modeling as with LES. This allows us to evaluate the PCA modeling approach relative to detailed kinetics independently from any closure approximation that would be accompany an LES implementation.

At the core of the ODT modeling approach is the triplet map. The triplet map is the mechanism in ODT by which fields are rearranged to model the effect of three-dimensional turbulent mixing in the one-dimensional domain. In ODT, all of the transported variables must be triplet mapped when an eddy occurs. We require conservation during a triplet mapping. Specifically,

$$\int \rho\vec{\phi} = \int \mathcal{T}(\rho\vec{\phi}), \tag{14}$$

where $\mathcal{T}$ is the triplet map function. Multiplying Eq. (3) by ρ and substituting into Eq. (14), we find

$$\int ([S]^{-1}[A_s](\rho\vec{\eta}) + \rho\vec{M}) = \int \mathcal{T}([S]^{-1}[A_s](\rho\vec{\eta}) + \rho\vec{M}), \tag{15}$$

or in an alternative form,

$$\int (\rho\vec{\eta} + [A_s^\top][S]\rho\vec{M}) = \int \mathcal{T}(\rho\vec{\eta} + [A_s^\top][S]\rho\vec{M}). \tag{16}$$

which is computationally more efficient. Eq. (16) shows that the actual term that should be triplet mapped is $\rho\vec{\eta} + [A_s^\top][S]\rho\vec{M}$ and then $\vec{\eta}$ should be extracted from it after triplet mapping to ensure conservation.

3. Computational configuration

To perform *a posteriori* evaluation of the PCA modeling approach described in Section 2, we consider a temporally evolving non-premixed CO/H_2-air planar jet with 11 species. The non-dimensional time scale of the problem is defined as $\tau = t/t_j$. Where, $t_j \approx 5\mu s$ is the characteristic time scale of the jet. This configuration has been the subject of prior work where the ODT model [29] has been compared to DNS [30,31] to establish the efficacy of the ODT model in its ability to capture the extinction and reignition present in this problem. The reader is referred to [29,31] for details of the computational setup.

This problem represents a challenging modeling problem due to the presence of significant extinction and reignition. There are DNS data available at three Reynolds numbers (cases *L*, *M* and *H* corresponding to low, medium and high Re), which have varying degrees of extinction. Because ODT has previously been shown to capture the essential features of the DNS data sets, and due to the relatively low cost of ODT as a stand-alone modeling approach, our *a posteriori* study will compare ODT with detailed kinetics with ODT using the PCA model.

We consider two methods for obtaining the PCA model:

1. Data from a detailed-kinetics ODT simulation [29].
2. Data from a SLFM model run at a wide range of dissipation rate, including a transient calculation beginning at the steady extinction limit and running through extinction, as previously described in [32].

In either case, the data is aggregated prior to applying PCA, so that the analysis is conducted on the entire data set. A key question is whether a PCA model trained using SLFM can be used to predict the physics present in the test problem.

The principal components are evolved according to Eq. (10) and state-variables are recovered using the MARS mappings. The initial conditions for the PC's are calculated from Eq. (4) using values of $\vec{\phi}$ extracted directly from the DNS data. The time step and grid resolution are the same as used in the DNS: $\Delta y = 15\,\mu$m and $\Delta t = 2$ ns, respectively. All the simulations are run for the duration of $\tau = 50$ and 128 realizations were used to obtain converged statistics.

4. Results

In this section, we present the results of *a posteriori* studies of the cases summarized in Section 3. Unless specified otherwise, results are shown for Pareto scaling with $N_\eta = 2$ and models trained on detailed

Table 1
Reconstruction accuracy, R^2, for cases *L*, *M* and *H* datasets. Note that R^2_L and R^2_H refer to the accuracy by which variables are reconstructed using the PCA obtained for case *L* and *H*, respectively.

	R^2			R^2_L		R^2_H	
	L	*M*	*H*	*M*	*H*	*L*	*M*
T	0.986	0.989	0.985	0.985	0.980	0.986	0.989
Y_{H_2}	0.950	0.964	0.950	0.957	0.934	0.913	0.952
Y_{O_2}	0.997	0.994	0.993	0.993	0.991	0.996	0.995
Y_O	0.861	0.932	0.914	0.887	0.861	0.846	0.934
Y_{OH}	0.927	0.946	0.939	0.933	0.901	0.927	0.953
Y_{H_2O}	0.946	0.945	0.939	0.927	0.898	0.940	0.942
Y_H	0.879	0.915	0.889	0.824	0.757	0.861	0.913
Y_{HO_2}	0.997	0.981	0.970	0.997	0.998	0.948	0.965
Y_{CO}	0.992	0.987	0.985	0.984	0.980	0.987	0.987
Y_{CO_2}	0.985	0.989	0.988	0.980	0.979	0.981	0.989
Y_{HCO}	0.971	0.941	0.907	0.923	0.835	0.958	0.940

kinetics ODT simulations (see Section 3). In the following sections, we consider the effect of N_η, the scaling method (VAST versus Pareto) and issues related to regression.

4.1. Effect of the Reynolds number

In this section, we compare results from cases *L* (Re = 2,510), *M* (Re = 4,478) and *H* (Re = 9,079). The PCA model for all three cases used $N_\eta = 2$ and was trained using data from case *M* with full chemistry.

We first consider an *a priori* analysis where we train the PCA model on each of these cases (*L*, *M* and *H*) and then reconstruct each of the data sets using each of the models. Table 1 shows the result of this *a priori* study and illustrates that there is not a major effect on the reconstruction when training the model on the various datasets.

We now turn our attention to the *a posteriori* studies. Based on the results from the *a priori* study, we initially consider a model trained on case *M* and applied to all three cases (*L*, *M* and *H*). Figure 1 shows the contour plots of these three cases and compares the ensemble average of temperature, H_2O, O_2 and OH, averaged over all of the realizations. In each contour plot, the result of the full chemistry ODT simulation (on the left half-plane) can be compared to the PC-transported ODT simulation results (on the right half plane). From the left, each column represents the results of case *L*, *M* and *H* respectively. The increasing extinction with increasing Re is apparent in both *T* and OH profiles. This figure shows an overall agreement between the full chemistry simulation and the reduced PC-transported simulation for all of the Re. There are quantitative differences, but the two-parameter PCA model does remarkably well at capturing extinction and reignition.

Figure 2 shows the conditional means of the same variables, conditioned on the mixture fraction (Z) over all of the realizations at the beginning of extinction ($\tau = 11$), when the maximum extinction occurs ($\tau = 21$), and at the time that the most of the fuel has reignited ($\tau = 46$) for the three Re. It is worth noting that, in previous *a priori* studies, this two-parameter PCA model has been compared to mixture fraction and dissipation rate parameterizations as a point of reference, and showed a significantly better parameterization of state variables using the two-parameter PCA parameterization [14].

Results shown in Figs. 1 and 2 were obtained using a model trained on Case *M*, based on *a priori* results that suggested there was little sensitivity. We revisit this issue here by considering results of each Re case using models trained various Re. Figures 3 and 4 show the conditional means of the same variables in Fig. 2 in the same format for models trained on Case *L* and *H*, respectively.

In general, the two-parameter PCA model predicts the trend of the conditional mean quantities qualitatively across the range of Re

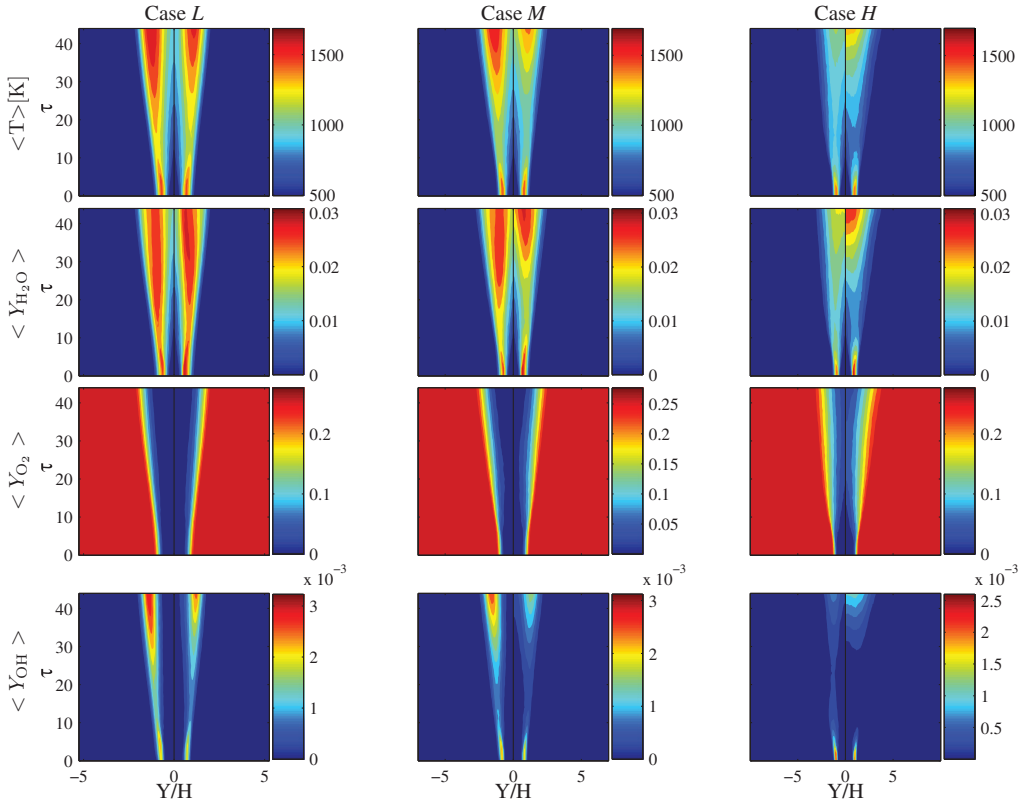


Fig. 1. Contour plots of temperature, Y_{H_2O}, Y_{O_2} and Y_{OH} for different Re. Each contour plot shows the full-chemistry ODT solution on its left half-plane and PC transported ODT solution on the right half-plane. Here, $N_\eta = 2$ and the model is trained on case *M*.

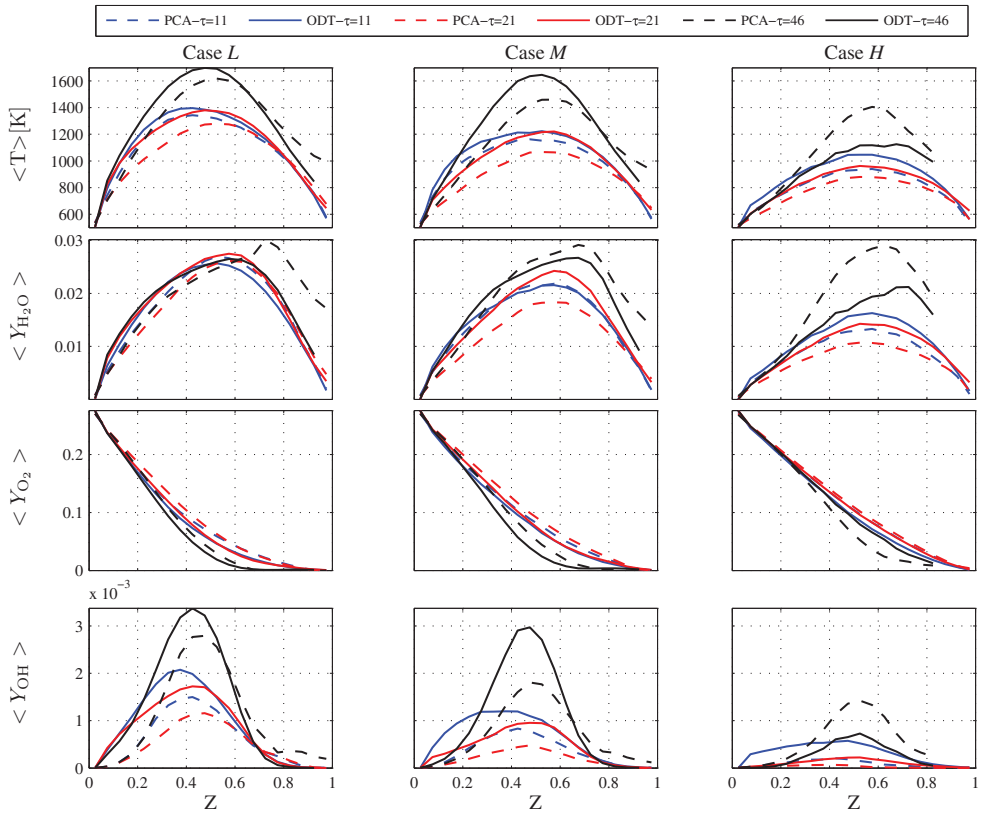


Fig. 2. Mean values of temperature, Y_{H_2O}, Y_{O_2} and Y_{OH} conditioned on the mixture fraction (Z) for different Re. Here the PCA model is trained on Case *M*. Each plot compares full-chemistry solution and PC transported solution at several points in time. See also Fig. 1.

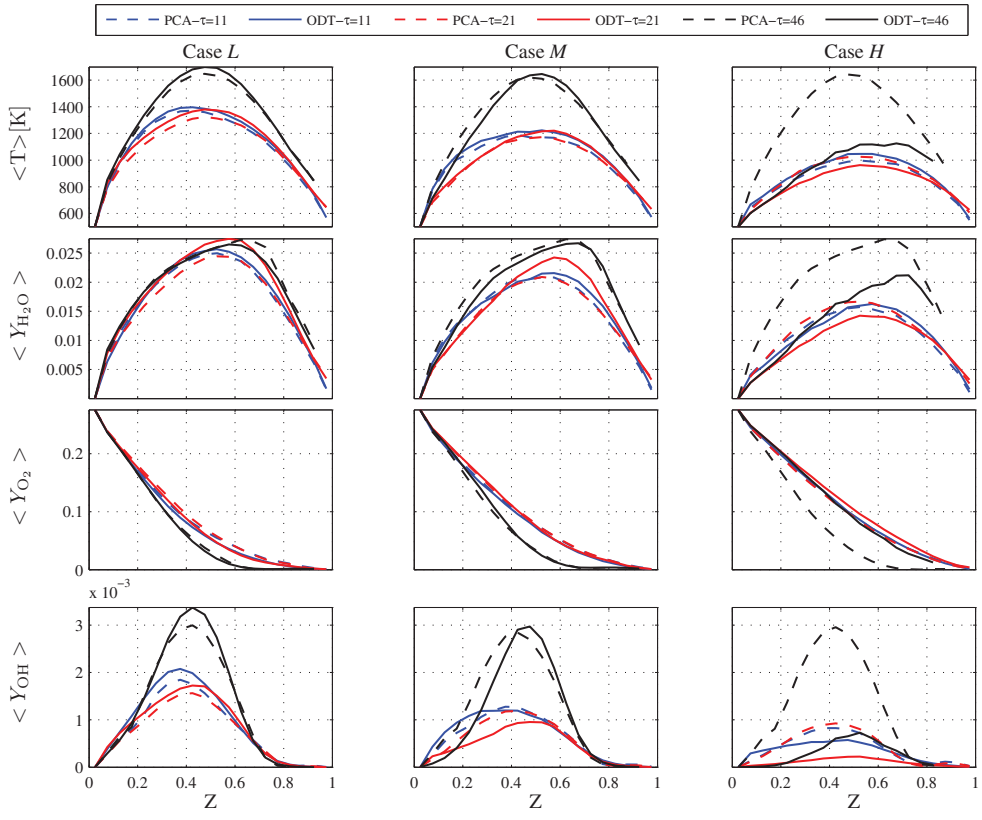


Fig. 3. Mean values of temperature, Y_{H_2O}, Y_{O_2} and Y_{OH} conditioned on the mixture fraction (Z) for different Re. Here the PCA model is trained on a data set with the same Re of Case *L*. Each plot compares full-chemistry solution and PC transported solution at several points in time. See also Figs. 2 and 4. .

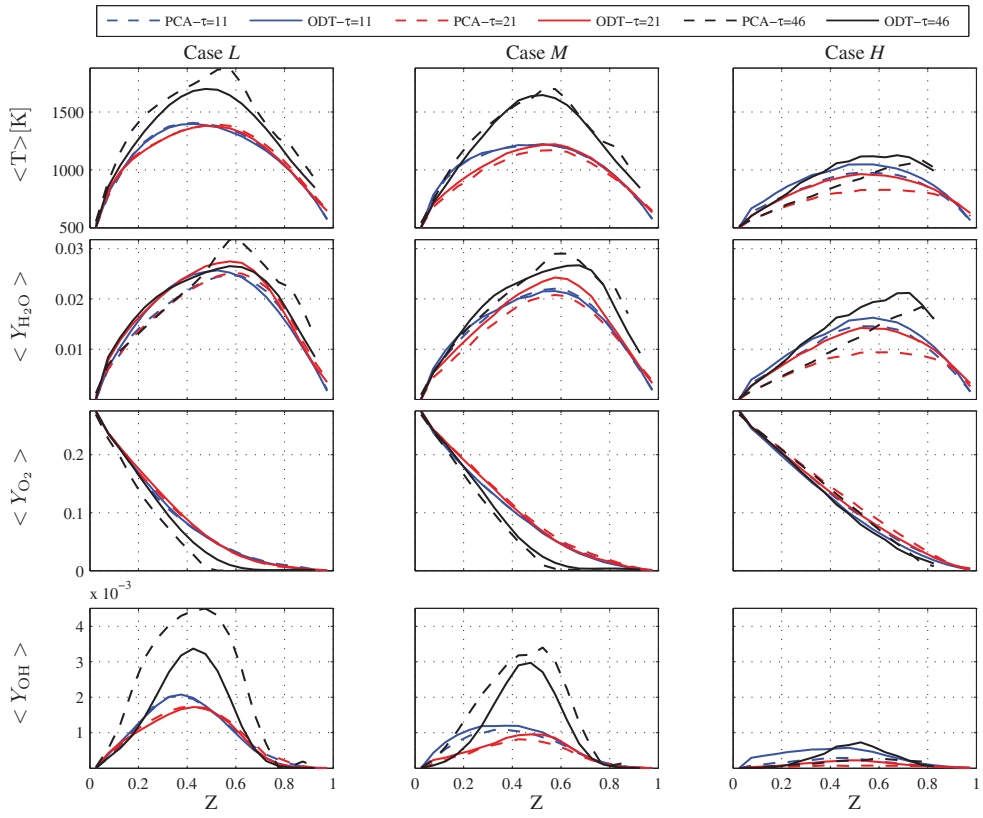


Fig. 4. Mean values of temperature, Y_{H_2O}, Y_{O_2} and Y_{OH} conditioned on the mixture fraction (Z) for different Re. Here the PCA model is trained on a data set with the same Re of Case *H*. Each plot compares full-chemistry solution and PC transported solution at several points in time. See also Figs. 2 and 3.

considered, but provide much better results on Cases *L* and *M* than on *H*, where the PCA model struggles to properly capture reignition. Contrary to the *a priori* results (see Table 1), training the PCA model on a data set with higher Re gives more accurate *a posteriori* results. For example, the results for temperature and Y_{H_2O} (as a major product) in case *H* at $\tau = 46$ is significantly better when using a model trained on Case *H* (see Fig. 4) than for models trained on Cases *L* or *M* (see Figs. 2 and 3). Also note that Case *H* nears global extinction and is, thus, a challenging modeling case - particularly for a two-parameter model as considered here.

4.2. Effect of the number of retained PC's

The results in the previous section were for a two-dimensional model ($N_\eta = 2$). Previous *a priori* studies have quantified the accuracy with which the thermochemical state variables can be reconstructed with varying N_η [14]. Here, we compare the results obtained with $N_\eta = 1$, 2 and 3.

Figure 5 shows contour plots of temperature, H_2O, O_2 and OH for simulation of Case *M* where $N_\eta = 1$ (top row) and $N_\eta = 3$. The left side of each contour represents the full chemistry solution, while the right side shows the PC transported solution. Comparing these results to the middle column (Case *M*) of Fig. 1, we see higher accuracy of reconstruction when increasing N_η.

Interestingly, temperature is still represented quite well with the one-parameter model, despite the fact that the species compositions are predicted poorly. This is because Pareto scaling was used, which gives more importance to variables with larger magnitude (*e.g.*, temperature). This is an interesting characteristic of PCA models: they can be 'tuned' to capture different state variables based on the choice of scaling parameters. This aspect of PCA models is something that merits further investigation.

Figure 6 shows the mean *T*, H_2O, O_2 and OH conditioned on the mixture fraction for PCA models with $N_\eta = 2$ and 3. Large errors in the species compositions predicted result in the case of $N_\eta = 1$ prevents us from presenting a meaningful comparison of conditional statistics since the mixture fraction cannot reliably be calculated due to large errors in some of the species compositions. Figure 6 shows that $N_\eta = 2$ and 3 both capture the species and temperature profiles well, and that there is an added benefit when going from $N_\eta = 2$ to 3. These observations are consistent with the *a priori* results shown in Table 2, which indicates that we should expect a significant improvement in species prediction when moving from $N_\eta = 1$ to 2 and a smaller improvement from 2 to 3. As previously discussed in *a priori* studies, the error decreases exponentially when adding additional PCs [11,13,14]. The *a posteriori* analysis here confirms the observations from *a priori* studies.

When a simulation accesses regions outside where the model was trained, the model behavior is not guaranteed to be physically realistic. This problem is compounded as the model dimensionality (N_η) increases. Figure 7 shows the structure of the manifolds from $N_\eta = 1$ to 3. For $N_\eta = 1$, the only way for a data point to go out of bound is to go beyond the limits of η_1. In the 2D case, this limiting boundaries grow to a non-trivial region, and for $N_\eta = 3$ the data cluster near a 2D surface in 3D space. As the number of retained PCs increases, the resulting structure more closely approaches a surface in high-dimensional space, and it becomes increasingly possible to drift off of the manifold. Once the simulation accesses points outside the trained manifold, the regressed values are not reliable, and this can accelerate divergence from the manifold.

One possible technique that one could consider to avoid divergence from the manifold in higher dimensions is to design the regression technique so that outside the training data space, PC source terms are either zeroed or designed to move the trajectory back toward the manifold. A key challenge here is to clearly define the

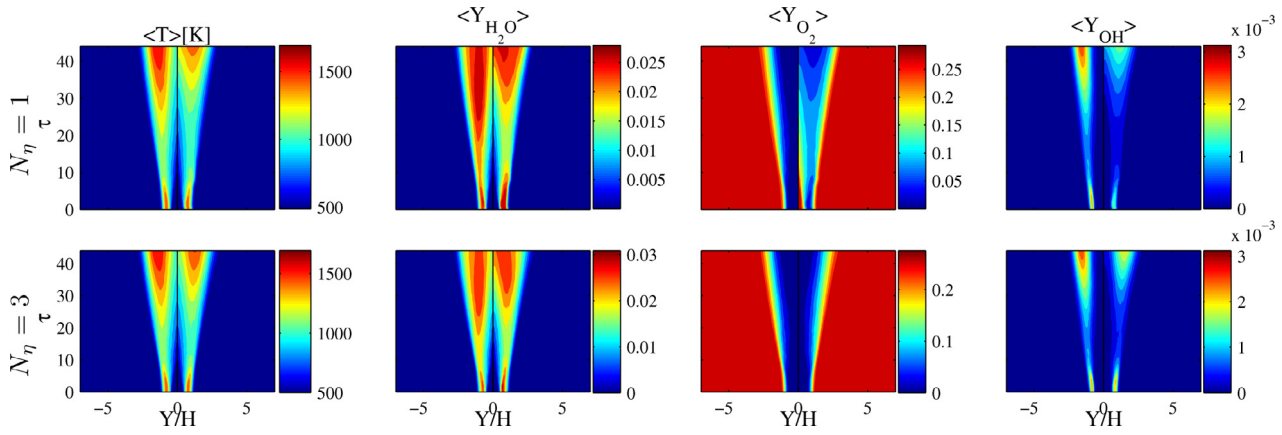


Fig. 5. Contour plots for case *M* using $N_\eta = 1$ (top row) and $N_\eta = 3$ (bottom row). These are comparable to the middle column (Case *M*) of Fig. 1. The left half-plane in each contour plot shows the full-chemistry while the PC transported solution is on the right side.

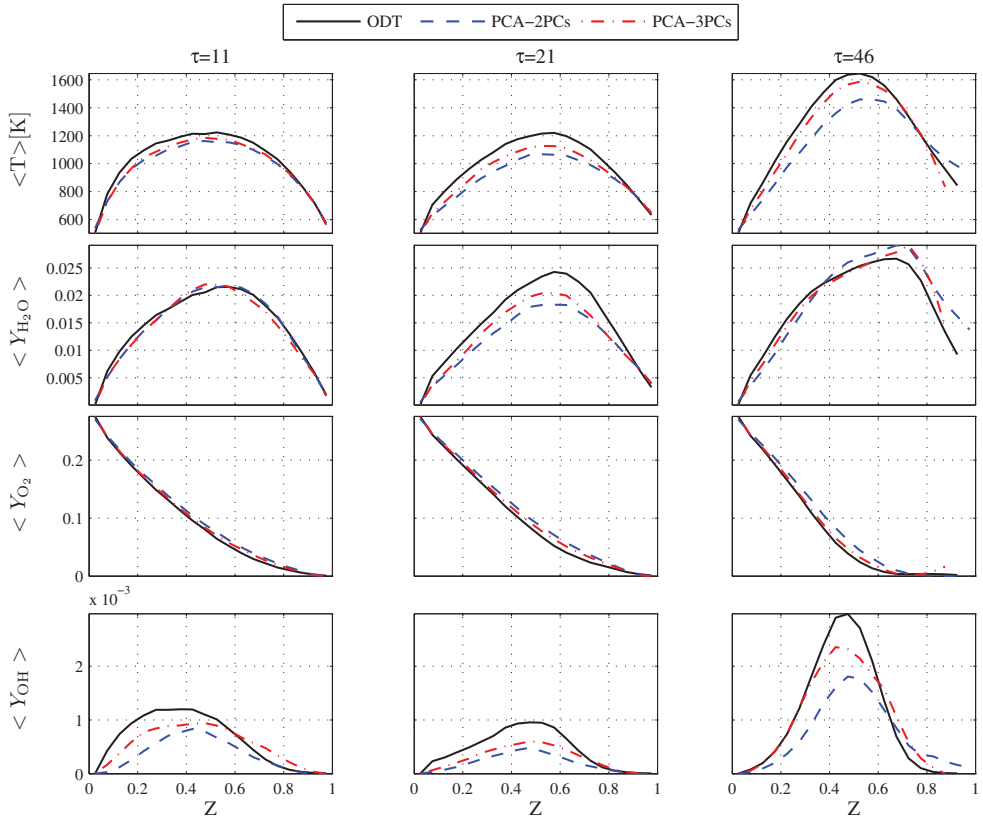


Fig. 6. Mean temperature, Y_{H_2O}, Y_{O_2} and Y_{OH} conditioned on the mixture fraction (Z) at three different key points in time (at the beginning, $\tau = 11$, and the end, $\tau = 21$, of extinction and after reignition is completed, $\tau = 46$).

Table 2
R^2 values for MARS regression of state-variables for $N_\eta = 1$ to $N_\eta = 3$ with the Pareto scaling.

N_η	*T*	H_2	O_2	*O*	OH	H_2O	*H*	HO_2	CO	CO_2	*HCO*	Average
1	1.000	0.323	0.839	0.767	0.898	0.984	0.873	0.710	0.592	0.996	0.312	0.754
2	1.000	0.992	0.9997	0.986	0.984	0.996	0.985	0.920	1.000	0.997	0.953	0.983
3	1.000	0.993	0.9999	0.988	0.986	0.997	0.986	0.921	1.000	1.000	0.954	0.984

Table 3
R^2 values for MARS regression of state-variables with VAST and Pareto scaling for $N_\eta = 2$.

Scaling	*T*	H_2	O_2	*O*	OH	H_2O	*H*	HO_2	CO	CO_2	*HCO*	Average
VAST	0.9997	0.995	0.9999	0.995	0.993	0.996	0.987	0.936	0.999	0.997	0.997	0.990
Pareto	1.0000	0.992	0.9997	0.986	0.984	0.996	0.985	0.920	1.000	0.997	0.953	0.983

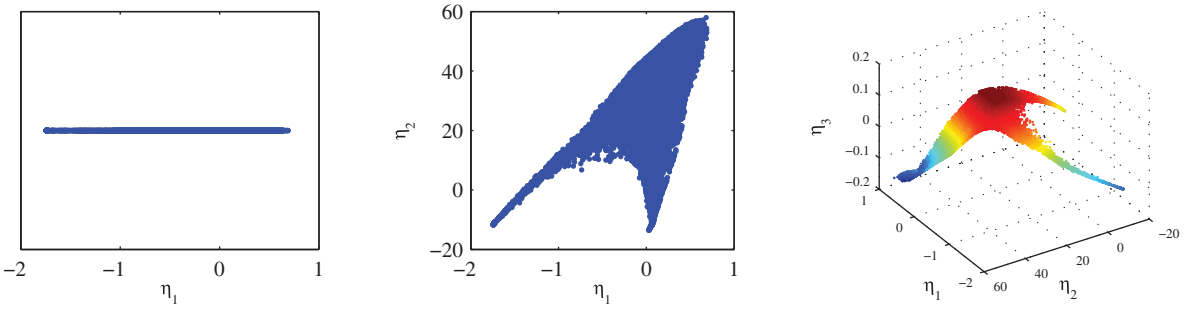


Fig. 7. Evolution of the PC manifolds from 1 retained PC to 3 retained PC's.

Table 4
R^2 values for MARS regression of PC source terms with VAST and Pareto scaling for $N_\eta = 1$ and 2.

Scaling	$N_\eta = 1$		$N_\eta = 2$	
	S_{η_1}	S_{η_1}	S_{η_2}	Average
VAST	0.67	0.819	0.699	0.76
Pareto	0.84	0.898	0.911	0.90

manifold (in high-dimensional space)and distinguish areas with sparse sampling (poorly characterized manifold) from areas with no realizations (off the manifold). These challenge increase in higher dimensional space. This facet of PCA-based modeling is something that requires further research effort.

4.3. Effect of scaling

The scaling method used in PCA (see Section 2.1) can have significant effect on the *a priori* evaluation of models. Based on previous results, we consider two different scaling options: Pareto and VAST. The reader is referred to [14] and [25] for detailed discussions of scaling methods. Tables 3 and 4 show *a priori* calculation of R^2 values for reconstruction of state-variables and PC source terms for the detailed chemistry ODT data set (see [14] for details). VAST scaling results in a more consistent R^2 value across the state variables, but also shows a lower R^2 for source terms. Pareto, on the other hand, weights temperature more heavily in the PCA basis, resulting in higher R^2 on temperature and also better reconstruction of the resulting PC source terms.

Figure 8 shows the mean T, H_2O, O_2 and OH profiles conditioned on the mixture fraction in an *a posteriori* case where the results

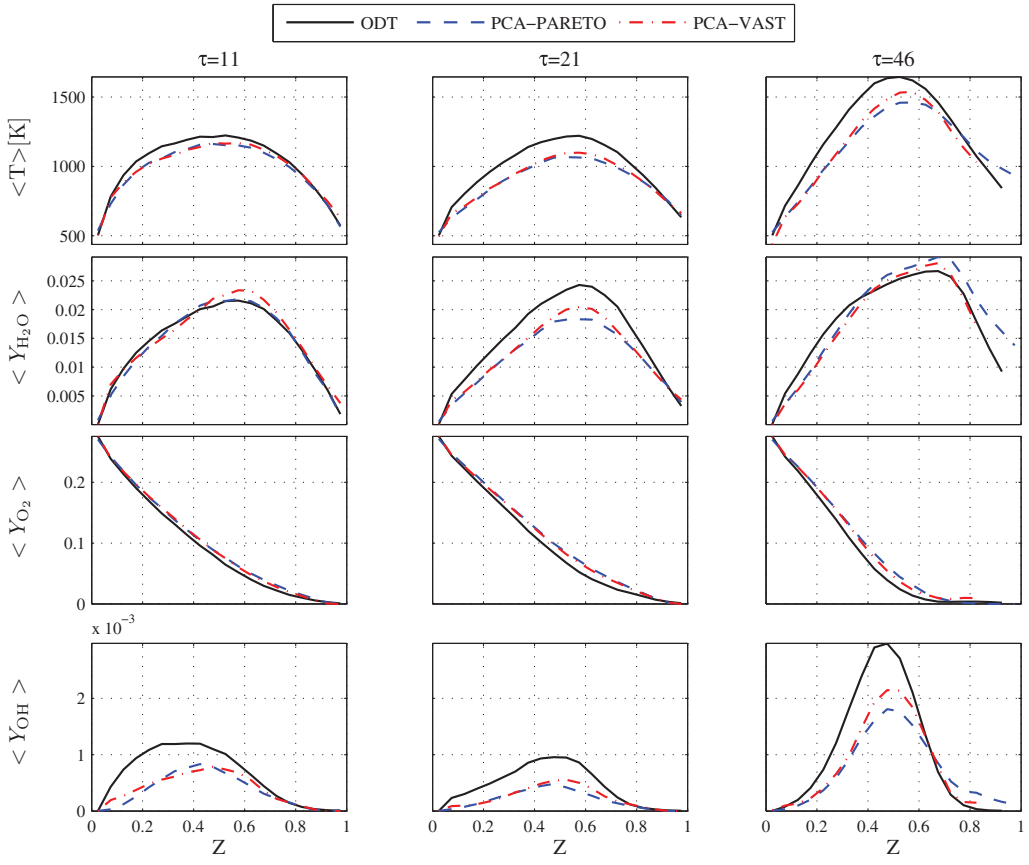


Fig. 8. Mean temperature, Y_{H_2O}, Y_{O_2} and Y_{OH} conditioned on the mixture fraction. Each plot shows the statistical means of these variable solutions using VAST and Pareto scaling in the PC transported simulation and the full-chemistry ODT simulation at a specific time (at the beginning, $\tau = 11$, and the end, $\tau = 21$, of extinction and after reignition is completed, $\tau = 46$).

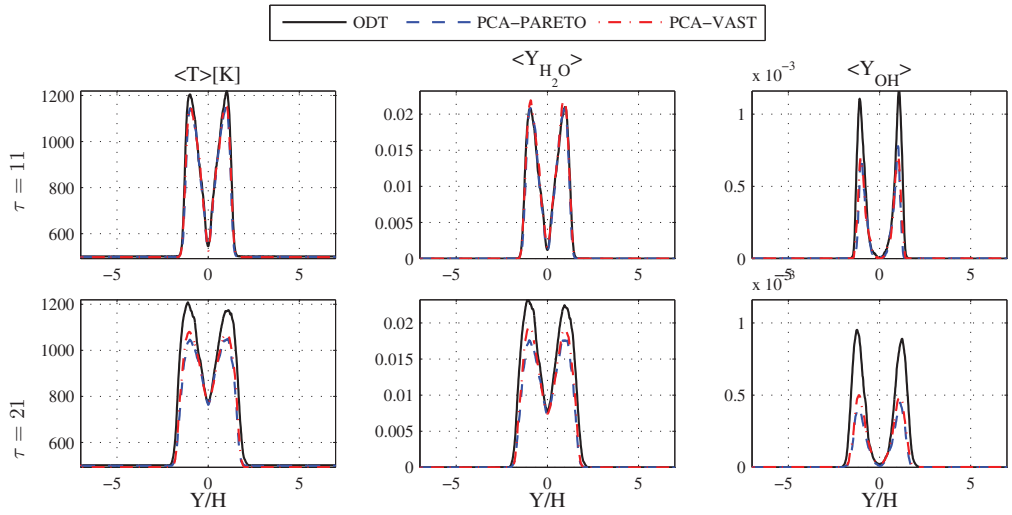


Fig. 9. Profiles of T, Y_{H_2O} and Y_{OH} at the onset of extinction ($\tau = 11$) and reignition ($\tau = 21$). See also Fig. 1.

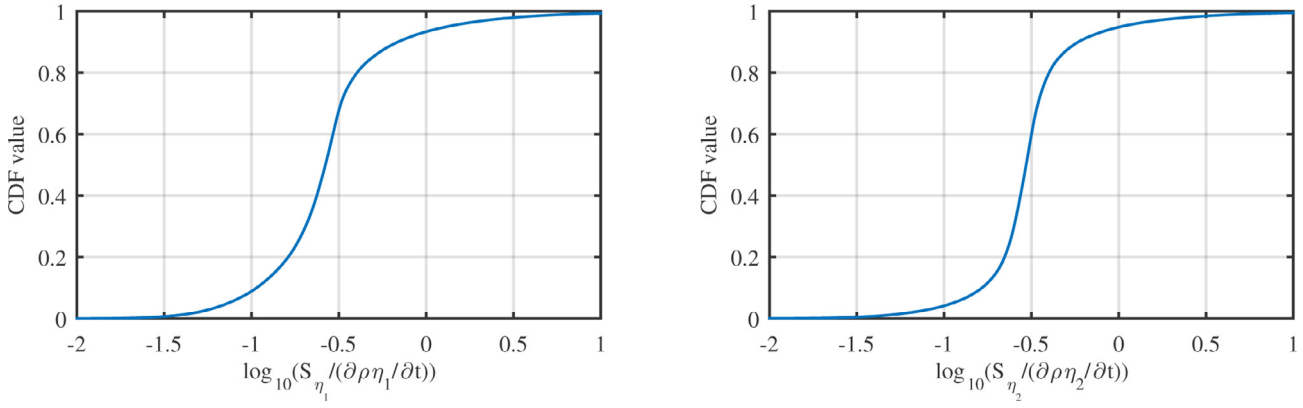


Fig. 10. Cumulative distribution function of the relative contributions of the PC source term to the RHS of the PC transport equation, $\frac{S_{\eta_1}}{\partial\rho\eta_1/\partial t}$ (left frame) and $\frac{S_{\eta_2}}{\partial\rho\eta_2/\partial t}$ (right frame), conditioned on $Z = [0.35, 0.5]$ ($Z_{st} = 0.42$ [29]).

of VAST scaling are compared to the Pareto scaling for training the PCA mapping. The results show that both of these scaling methods yield comparable results (with VAST scaling providing slightly better agreement with the full chemistry calculations) despite the fact that the PC source terms are parameterized better with Pareto than VAST scaling.

Figure 9 shows the corresponding ensemble-averaged spatial profiles at key points in time for VAST and Pareto scaling (see Fig. 1 for contour plots over space and time). Again, there is a good agreement between the PC transported results and full-chemistry results for both scaling methods. This result is encouraging because, despite the relatively low R^2 value on the source terms shown in Table 4, the models are able to capture the evolution of the system with reasonable accuracy.

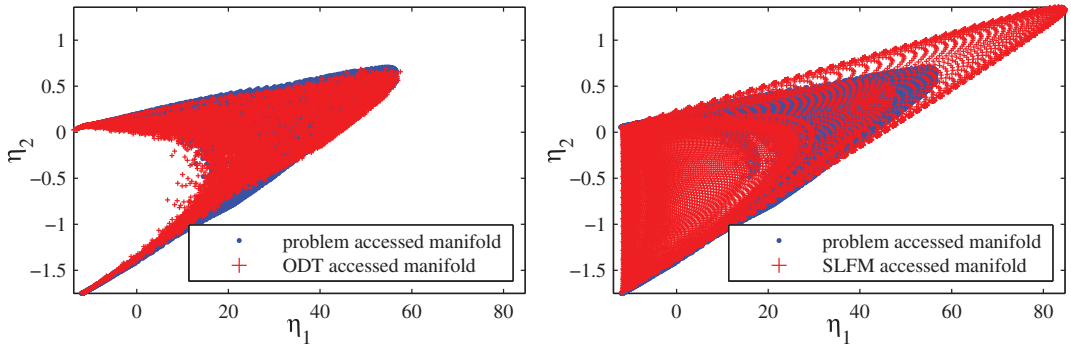
To determine why the relatively inaccurate regression of PC source terms does not have a major adverse impact on the model's performance, we consider the relative contribution of the source terms to the overall rate of change of the η_1 and η_2 as defined by (10). Figure 10 shows the relative contribution of the source term in the PC transport equations for η_1 and η_2: $\frac{s_{\eta_1}}{\partial\rho\eta_1/\partial t}$ and $\frac{s_{\eta_2}}{\partial\rho\eta_2/\partial t}$, respectively. These are conditioned on areas near stoichiometric, $Z = [0.3, 0.5]$. Figure 10 indicates that more than 90% of the observations near stoichiometric had source terms whose contribution was less than the transport. While the effect of these source terms is non-negligible, it does not dominate the PC transport even near stoichiometric. This likely explains why the relatively inaccurate regression of PC source terms still yielded reasonable agreement for the state variables. However, improvement in the representation of PC source terms is certainly an area where further research is warranted.

4.4. Effect of training data

If the PC transport drifts outside of the training manifold, the regression will return a source term which may not be physically correct or even realizable. An ideal training data set should cover the whole range of parameter space accessed by the application to avoid 'extrapolation' into regions where the state space was not covered by the training data.

To investigate the effect of the training data set used to train the PCA/MARS model, two different data sets, explained in Section 3, are considered. The first training data set is obtained from ODT calculations with detailed chemistry and the second (SLFM) is obtained from a flamelet calculation using the procedure outlined in [32], where a steady flamelet library is augmented by the transient extinction solution obtained at χ_{max}. This data set is then re-parameterized using PCA.

Figure 11 shows realizations of the two different training data sets in PC space. Also shown are a down-sampled subset of the points accessed by the *a posteriori* simulation. The empty region on the left side of the ODT training data plot in Fig. 11a, corresponds to extinction, which is well-covered by the flamelet training data set in Fig. 11b. Because complete extinction is not observed in the calcula-



(a) ODT full chemistry training data.

(b) SLFM training data augmented with transient extinction.

Fig. 11. The manifold coverage of the two data sets considered in this work. Note that the full-chemistry ODT data set size is reduced 25 times and SLFM data set size is reduced 5 times for better visibility.

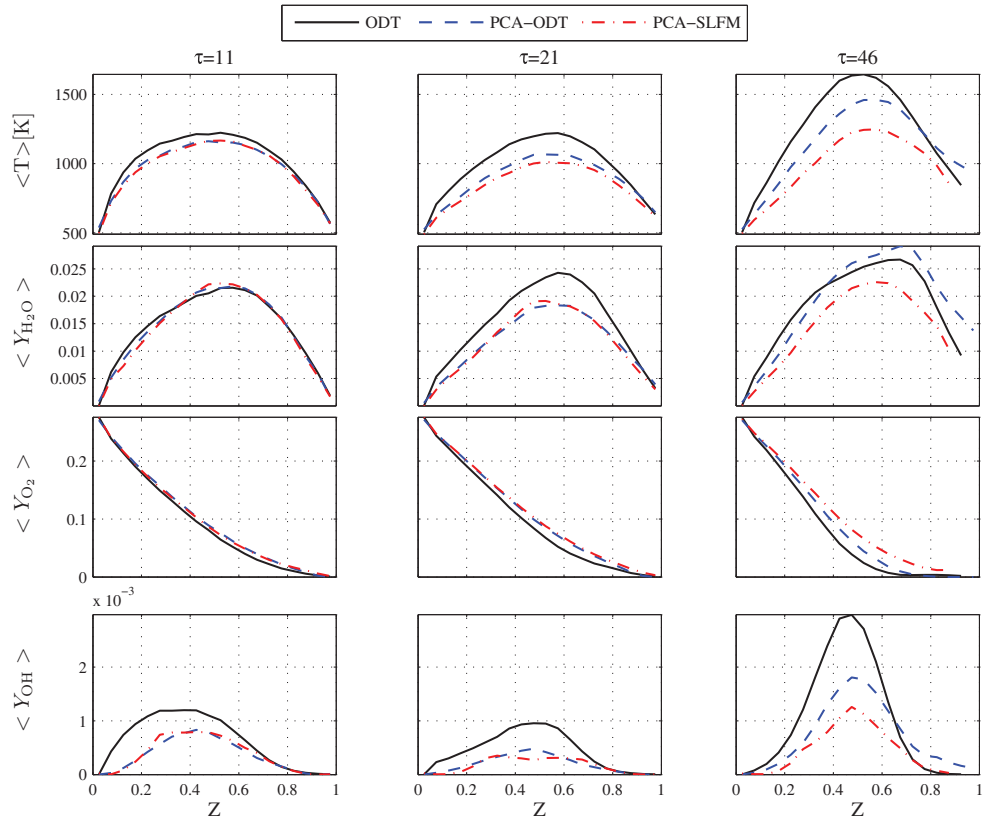


Fig. 12. Conditional means with respect to the mixture fraction. Each plot shows the statistical means of the variable solutions using the SLFM data set and the full-chemistry ODT data set Section 3 from the PC transported simulation and compares them against the full-chemistry ODT simulation at three different times (at the beginning, $\tau = 11$, and the end, $\tau = 21$, of the extinction and after reignition is completed, $\tau = 46$).

tions, the extinction region is not accessed by either the ODT training data or the *a posteriori* calculation. More concerning, however, is the fact that some areas on the boundary of the manifold accessed by the *a posteriori* calculation are not covered by the ODT training data.

Figure 11 b compares the SLFM generated manifold against the accessed manifold. The SLFM data set covers the whole accessed manifold of the problem plus the whole extinction area. Note that, in this case, the flamelet generated manifold is very similar to the ODT-generated manifold. This is likely due to the fact that, for this problem, the underlying physical manifold is nearly two-dimensional.

It is worth noting that the SLFM data set implies a manifold by construction since the dissipation rate (the key varying parameter in the model construction) does not vary according to random turbulent fluctuations and time-history effects are not incorporated. This manifold is, thus, implicitly embedded into the PCA model. However, the effects of extinction and reignition are captured due to the PCA rotation of the basis from mixture fraction and dissipation rate to PCs which act more as progress variables.

Figure 12 shows the conditional mean values of temperature, H_2O, O_2 and OH for *a posteriori* studies using the SLFM–PCA model as well as the full-chemistry ODT–PCA model. The SLFM model does not predict the scalar profiles as well as the ODT model, but generating the SLFM model is an order of magnitude less costly than generating the ODT model.

5. Conclusions

In this work the PCA/MARS modeling is examined as a framework to simulate a challenging problem including local extinction and reignition where PC's are transported independently and state-variables are calculated from PC's and MARS. We present results from *a posteriori* studies that compare PCA-based models with detailed kinetics models.

Comparisons are made for prediction of turbulent nonpremixed CO/H_2 data sets involving extinction and reignition. Results demonstrate that PCA-based models can match profiles and conditional statistics of the problem when retaining only two dimensions (*i.e.*, a 2-parameter model). The effect of scaling method on the resulting model accuracy is not large for $N_\eta = 2$, which is consistent with previous *a priori* studies.

The applicability of models across Re is studied *a priori* and *a posteriori*. The *a priori* study shows little sensitivity of model accuracy when trained at one Re and applied at another. On the other hand, *a posteriori* results show that, as the Reynolds number changes enough to cause change the balance between mixing and reaction timescales, choosing a training data set appropriate to each regime provides more accurate results. Interestingly, relatively simple models such as the steady flamelet model augmented with transient extinction results can also yield reasonable results, provided that the training data covers the essential state space accessed by the application. This can reduce the cost of a PCA modeling strategy since flamelet calculations are less costly than ODT calculations for training PCA models.

As expected from *a priori* studies, increasing the number of retained PC's thereby increasing the dimensionality of the model clearly improves the state-variable's reconstruction results.

We also identify a challenge facing PCA-based models: as N_η increases, the likelihood that the simulation will depart from the region over which the manifold was trained increases. Care must be taken in these situations to avoid divergent departure from the manifold, and this is an important future research direction.

Finally, the results from this *a posteriori* study show that the magnitude of the PC source terms is usually smaller than the transport terms in the PC transport equation. Thus, while the PC source terms are not represented as accurately as the state variables, this effect is diminished due to the dominance of transport relative to source terms in the PC transport. Further work to improve representation of the PC source terms would be beneficial to improve PCA as a modeling strategy, however.

In summary, PCA models provide a powerful and viable technique to accurately model challenging combustion problems. They provide a systematic means to identify a basis on which to effectively collapse the thermochemical state of the system, and a means to expand that basis to achieve desired accuracy. The *a posteriori* study conducted here demonstrates the efficacy of PCA and also presents some future research challenges.

Acknowledgments

This work was supported by the National Science Foundation PetaApps award 0904631.

References

[1] K.K. Kuo, Principles of Combustion, Wiley, 1986.
[2] W.P. Jones, S. Rigopoulos, Combust. Flame 142 (3) (2005) 223–234.
[3] S. Rigopoulos, Int. J. Multiscale Comput. Eng. 5 (1) (2007) 11–18.
[4] N. Peters, Prog. Energy Combust. Sci. 10 (1984) 319–339.
[5] J.A. van Oijen, Flamelet-generated manifolds: development and application to premixed flames, Eindhoven University of Technology, 2002 Ph.D. thesis.
[6] J.A. van Oijen, L.P.H. de Goey, Combust. Sci. Technol. 161 (2000) 113–137.
[7] J.A. van Oijen, L.P.H. de Goey, Combust. Theory Model. 6 (2002) 463–478.
[8] O. Gicquel, N. Darabiha, D. Thévenin, Proc. Combust. Inst. 28 (2000) 1901–1908.
[9] B. Fiorina, R. Baron, O. Gicquel, D. Thévenin, S. Carpentier, N. Darabiha, Combust. Theory Model. 7 (2003) 449–470.
[10] B. Fiorina, O. Gicquel, S. Carpentier, N. Darabiha, Combust. Sci. Technol. 176 (5) (2004) 785–797.
[11] J.C. Sutherland, A. Parente, Proc. Combust. Inst. 32 (1) (2009) 1563–1570.
[12] A. Parente, J.C. Sutherland, B.B. Dally, L. Tognotti, P.J. Smith, Proc. Combust. Inst. 33 (2) (2011) 3333–3341.
[13] A. Parente, J.C. Sutherland, L. Tognotti, P.J. Smith, Proc. Combust. Inst. 32 (1) (2009) 1579–1586.
[14] A. Biglari, J.C. Sutherland, Combust. Flame 159 (5) (2012) 1960–1970.
[15] J.C. Sutherland, A. Parente, 5th US Natl. Combust. Meet. (2009).
[16] S.B. Pope, Proc. Combust. Inst. 34 (2013) 1–31.
[17] Y. Yang, S.B. Pope, J.H. Chen, Combust. Flame 160 (2013) 1967–1980.
[18] H. Mirgolbabaei, T. Echekki, Combust. Flame 160 (2013) 908–989.
[19] B.J. Isaac, A. Parente, P.J. Smith, G. Fru, D. Thévenin, Proc. Combust. Inst. (2013).
[20] A. Coussement, O. Gicquel, A. Parente, Proc. Combust. Inst. 34 (1) (2013) 1117–1123.
[21] H. Mirgolbabaei, T. Echekki, Combust. Flame 161 (2013) 118–126.
[22] H. Mirgolbabaei, T. Echekki, N. Smaoui, Int. J. Hydrogen Energy 39 (2014) 4622–4633.
[23] B.J. Isaac, A. Coussement, O. Gicquel, P.J. Smith, A. Parente, Combust. Flame 161 (2014) 2785–2800.
[24] T. Echekki, H. Mirgolbabaei, Combust. Flame, (2015).
[25] A. Parente, J.C. Sutherland, Combust. Flame 160 (2) (2013) 340–350, doi:10.1016/j.combustflame.2012.09.016.
[26] J.H. Friedman, Ann. Stat. 19 (1) (1991) 1–67.
[27] J.H. Friedman, Fast MARS, Technical Report 110, Stanford University Department of Statistics, 1993.
[28] J.H. Friedman, C.B. Roosen, Stat. Methods Med. Res. 4 (1995) 197–217.
[29] N. Punati, J.C. Sutherland, A.R. Kerstein, E.R. Hawkes, J.H. Chen, Proc. Combust. Inst. 33 (2011) 1515–1522.
[30] E.R. Hawkes, R. Sankaran, J.C. Sutherland, J.H. Chen, Proc. Combust. Inst. 31 (2007) 1633–1640.
[31] E.R. Hawkes, R. Sankaran, J.H. Chen, S.A. Kaiser, J.H. Frank, Proc. Combust. Inst. 32 (1) (2009) 1455–1463.
[32] J.C. Sutherland, P.J. Smith, J.H. Chen, Combust. Theory Model. 11 (2) (2007) 287–303.