ELSEVIER

Combustion modeling using principal component analysis

James C. Sutherland [a,*], Alessandro Parente [b]

[a] Department of Chemical Engineering, University of Utah, 50 S. Central Campus Drive,
3290 MEB, Salt Lake City, UT 84112-9203, USA
[b] Department of Chemical Engineering, University of Pisa, Pisa, Italy

Abstract

The thermochemical state of a single-phase reacting system containing n_s species is $n_s + 1$ dimensional. However, it is widely recognized that low-dimensional manifolds exist in this space. The question then arises as to the best way to approximate these manifolds. Common chemistry reduction approaches for combustion simulation (such as laminar flamelet, equilibrium, and flame-sheet (Burke–Schuman) chemistry) specify a priori the manifold dimensionality and the parameters that form the basis for representing the thermochemical state, which is then prescribed as a unique function of these parameters. If high-fidelity data are available from experiment or direct numerical simulation (DNS), manifolds may be determined via a principal component analysis (PCA), with the principal components (PCs) forming a new basis for describing the thermochemical state. By considering a truncated set of PCs, one may approximate the thermochemical state with rigorous error bounds, extract a parameterized representation of the thermochemical state, and derive transport equations for the principal components. This paper outlines a methodology for constructing a reduced model for the thermochemical state from high-fidelity data with particular focus on the ability to parameterize source terms appearing in the transport equations for the principal components. The modeling approach allows quantitative a priori control of the error in the state-space parameterization, and can be applied to both laminar and turbulent simulations in all combustion regimes.
© 2009 The Combustion Institute. Published by Elsevier Inc. All rights reserved.

Keywords: Turbulent flames; DNS; Modeling; PCA; Manifold

1. Introduction

Direct numerical simulation (DNS) of practical combustion systems is impossible due to two general complications:

1. *Resolution requirements*. DNS requires that all length and time scales physically realizable by the system of interest be fully resolved. The separation of length scales in turbulent flow implies that resolution requirements scale roughly with the turbulent Reynolds number as Re^3 [1].

2. *Number of equations to be solved*. Detailed combustion mechanisms for simple fuels such as methane involve 53 species and 325 reactions [2]. Hence, the direct solution of a reacting system involving only methane in air would require solution of 57 strongly coupled PDEs (52 species, continuity, three momentum, and energy). For higher hydrocarbons, the number of equations increases significantly.

---
* Corresponding author. Fax: +1 801 585 9292.
  E-mail address: James.Sutherland@utah.edu (J.C. Sutherland).

Models are commonly introduced to alleviate the cost associated with both of these issues. Resolution requirements (issue 1 above) are commonly addressed by filtering the governing equations spatially and/or temporally to produce the LES and RANS equations. This leads to turbulent closure problems with convective and source terms in particular.

Models to reduce the number of equations to be solved (issue 2 above) vary widely. In general, the thermochemical state of a single-phase reacting system with n_s chemical species is uniquely determined by $n_s + 1$ parameters (e.g. T, p, and $n_s - 1$ mass fractions, Y_i). However, it is widely recognized that in turbulent combustion, lower dimensional manifolds exist in this high-dimensional space [3,4]. All of the models which reduce the number of thermochemical degrees of freedom rely, either directly or indirectly, on the assumption of a low-dimensional attractive manifold to which the thermochemical dynamics quickly relax.

Among the most common approaches to reducing the number of thermochemical degrees of freedom are mechanism reduction and state-space parameterization. Mechanism reduction leads to a reduced set of species equations, typically by analysis of the reaction rates [5–9]. Reparameterization of the thermochemical state space by a small number of parameters is perhaps the most common combustion modeling approach. The most common choice of a parameter for nonpremixed combustion is the mixture fraction (f), which is a particularly convenient choice since if all species diffusivities are equal, f is a conserved scalar [10,11] and no source terms require closure. However, properties such as density, viscosity, etc. must still be calculated from f.

This paper focuses on the second of these approaches, re-parameterizing the thermochemical state by a small number of parameters based on the existence of a low-dimensional manifold. Most successful attempts at exploiting these manifolds have been to prescribe a priori the parameters that characterize the manifold [4,12–14]. However, such an approach restricts the subspace that the thermochemistry may access without providing any quantitative error analysis a priori. Indeed, as mixing and reaction timescales increasingly overlap, the dimensionality of a manifold increases, as does the error associated with a parameterization of fixed dimensionality [4,15].

Principal component analysis (PCA) offers the potential to automate the selection of an optimal basis for representing the manifolds which exist in turbulent combustion. The central idea of PCA is to reduce the dimensionality of a data set consisting of a large number of correlated variables while retaining most of the variation present in the original data. The dimension reduction is achieved by recasting the data into an n_η-dimensional space identified by the principal components (PCs). In many cases (particularly those where manifolds exist), the number of dimensions, n_η, required to accurately approximate the original data are much smaller than the original dimension of the data.

PCA provides an optimal representation of the system based on $n_\eta \ll n_s + 1$ "optimal" variables, η, which are linear combination of the $n_s + 1$ primitive variables T, p and Y_i. PCA provides a linear mapping from the original variables (T, p, Y_i) to a set of PCs, η. Using this mapping, transport equations with associated initial and boundary conditions may be derived for the PCs, as discussed in Section 2.1.1.

The primary advantage of the PCA modeling approach presented herein is that it provides a rigorous mathematical formalism to select parameters which optimally represent the thermochemical state of the system. Thus, models may be created which satisfy a given error threshold on any thermochemical variable (e.g. ρ, μ, T, λ, Y_i, etc.). However, for PCA to be effectively exploited as a modeling technique, two conditions must hold. First the thermochemical variables must be parameterized within acceptable error bounds. Second, the source terms for the PCs must be parameterized within acceptable error bounds.

2. Principal component analysis

2.1. PCA formulation

Consider m observations of n variables arranged in an $n \times m$ matrix $\mathbf{X}$ whose columns represent individual observations and rows represent different variables. The basic idea of PCA is to find a basis for representing the data $\mathbf{X}$ such that the data are well represented by a truncated basis [16,17]. We define the covariance matrix[1] $\mathbf{R} = 1/(n-1)\mathbf{X}^\mathsf{T}\mathbf{X}$ and perform an eigenvector decomposition of $\mathbf{R}$ to obtain $\mathbf{\Lambda} = \mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}$, where $\mathbf{Q}$ are the orthonormal eigenvectors of $\mathbf{R}$, which implies that $\mathbf{Q}^{-1} = \mathbf{Q}^\mathsf{T}$. The eigenvectors (columns of $\mathbf{Q}$) form a new basis, and the PCs of the data represented by $\mathbf{X}$ are defined as

$$\eta = \mathbf{QX}. \tag{1}$$

The full set of PCs exactly reproduces all observations in the original data, by definition.

The PCA analysis recasts the observations in a rotated basis that has some desirable properties:

- The basis is orthonormal.

---

[1] Here we have assumed that the data are centered (its mean is zero) and scaled by constant factors γ_i. These are common procedures that can strongly influence the results of the PCA. For results here, we chose vast scaling (see [17]).

- The basis optimally represents the variance in the original data.
- The eigenvalues, Λ, indicate which basis vectors (eigenvectors) are of greatest importance in representing the data.
- Since PCA is a linear transformation, if we have governing equations for the original basis, we may easily obtain governing equations in the new (PC) basis, as discussed in Section 2.1.1.

The real utility in PCA comes by exploiting the fact that PCA maximizes the variance of the data in each PC direction. Hence, the rotated coordinate system has the property that the first dimension (corresponding to the largest eigenvalue) is selected to best represent the variance in the data. Subsequent directions each represent the next-largest variance in the data. Therefore, a truncated basis, i.e. a subset of the columns in $\mathbf{Q}$, can approximate the original data remarkably well. We define a transformation matrix $\mathbf{A}$ as a rank-deficient subset of the $\mathbf{Q}$ matrix with n rows and n_η columns. The columns of $\mathbf{A}$ correspond to the columns of $\mathbf{Q}$ with the n_η largest eigenvalues. We may then approximate $\mathbf{X}$ as

$$\mathbf{X} \approx \boldsymbol{\eta}\mathbf{A}^{\mathsf{T}}. \qquad (2)$$

In the context of combustion applications, the n variables comprising the rows of $\mathbf{X}$ are the $n_s + 1$ primitive variables $[T, p, Y_1, Y_2, \ldots, Y_{n_s-1}]$. Performing a PCA on this set of variables yields a new $(n_s + 1)$-dimensional basis, $\boldsymbol{\eta}$, which is a rotation of the original basis. Retaining the $n_\eta < (n_s + 1)$ columns of $\mathbf{Q}$ with the largest eigenvalues defines a basis for a n_η-dimensional parameterization of the thermochemical state of the system.

2.1.1. Transport equations for the PCs

The transport equations for a set of variables $\boldsymbol{\Phi} = [T, p, Y_1, Y_2, \ldots, Y_{n_s-1}]^{\mathsf{T}}$ may be written as[2]

$$\rho\frac{\mathrm{D}(\boldsymbol{\Phi})}{\mathrm{D}t} = -\nabla \cdot (\mathbf{j}_{\boldsymbol{\Phi}}) + (s_{\boldsymbol{\Phi}}), \qquad (3)$$

where $\frac{\mathrm{D}}{\mathrm{D}t} \equiv \rho\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla$ is the material derivative operator, $\mathbf{u}$ is the mass-averaged velocity of the system, $\mathbf{j}_{\boldsymbol{\Phi}}$ is the mass-diffusive flux of $\boldsymbol{\Phi}$ relative to the mass-averaged velocity, and $s_{\boldsymbol{\Phi}}$ is the volumetric rate of production of $\boldsymbol{\Phi}$. For the analysis presented herein, data were considered from a multitude of observations in both space and time, and the PCA was performed on the combined spatial-temporal dataset. Therefore, $\mathbf{A}$ is constant in space and time by construction, and commutes with differential operators such as ∇, $\nabla\cdot$, and $\frac{\mathrm{D}}{\mathrm{D}t}$. Given that $\mathbf{A}$ is con-stant and since PCA represents a linear transformation, we may multiply Eq. (3) by $\mathbf{A}$, with $\mathbf{X} = \boldsymbol{\Phi}$, to find transport equations for the PCs

$$\rho\frac{\mathrm{D}}{\mathrm{D}t}(\boldsymbol{\eta}) = -\nabla \cdot (\mathbf{j}_{\boldsymbol{\eta}}) + (s_{\boldsymbol{\eta}}), \qquad (4)$$

where $(\boldsymbol{\eta}) = [\mathbf{A}](\boldsymbol{\Phi})$, $(\mathbf{j}_{\boldsymbol{\eta}}) = [\mathbf{A}](\mathbf{j}_{\boldsymbol{\Phi}})$ and $(s_{\boldsymbol{\eta}}) = [\mathbf{A}](s_{\boldsymbol{\Phi}})$. In Eq. (4), the source terms of temperature, pressure, and *all* species contribute to the source term for each PC (assuming nonzero A_{ij}).[3] Therefore, the truncated basis (where $n_\eta < n_s + 1$) must not only parameterize the thermochemical state but the source terms as well. This paper considers both of these issues quantitatively by examining data obtained from DNS.

2.2. PCA modeling approach

A complete PCA modeling approach requires several ingredients. First, the PCs must be identified using the procedure outlined in Section 2.1. This identification requires high-fidelity, fully-resolved data including source terms. Once the PCs are selected, transport equations may be derived for each PC as described in Section 2.1.1.

Second, the initial conditions (ICs) and boundary conditions (BCs) on the PCs must be defined using the transformation matrix $\mathbf{A}$. For Dirichlet BCs on all $\boldsymbol{\Phi}$, we obtain Dirichlet conditions on the PCs, $\boldsymbol{\Phi} = \boldsymbol{\Phi}^* \Rightarrow \boldsymbol{\eta} = \mathbf{A}^{\mathsf{T}}\boldsymbol{\Phi}^*$ (ICs are analogously defined). Likewise, Neumann conditions on $\boldsymbol{\Phi}$ yield Neumann conditions on $\boldsymbol{\eta}$, $\mathbf{n} \cdot \nabla\boldsymbol{\Phi} = \boldsymbol{\Theta} \Rightarrow \mathbf{n} \cdot \nabla\boldsymbol{\eta} = \mathbf{A}^{\mathsf{T}}\boldsymbol{\Theta}$. Mixed conditions on $\boldsymbol{\Phi}$, e.g. Neumann conditions on ϕ_1 and Dirichlet conditions on $\phi_2 \ldots \phi_n$ yield Robin boundary conditions on $\boldsymbol{\eta}$.

Diffusion terms in the transport equations for $\boldsymbol{\eta}$ require evaluation of the diffusive fluxes for each component of $\boldsymbol{\Phi}$. In turbulent flow calculations, the molecular diffusion term is typically augmented by a "turbulent diffusion" term arising from closure of the convective term. In many cases, and particularly at high Reynolds number, the molecular diffusion term is small relative to the turbulent diffusion term and is neglected. However, even when one wishes to retain the full description of molecular diffusion, the treatment with PCA is straightforward. First, $\boldsymbol{\Phi}$ is approximated from $\boldsymbol{\eta}$. Next, the diffusive terms for $\boldsymbol{\Phi}$, $\mathbf{j}_{\boldsymbol{\Phi}}$, are constructed. Finally, the diffusive fluxes for the PCs are calculated as $(\mathbf{j}_{\boldsymbol{\eta}}) = [\mathbf{A}]^{\mathsf{T}}(\mathbf{j}_{\boldsymbol{\Phi}})$.

---

[2] This analysis applies equally to conserved variables $\rho, \rho h, \rho Y_i$.

[3] For open domains in the absence of strong pressure gradients (e.g. shock waves), the thermochemical state is relatively insensitive to small pressure changes. Therefore, the coefficient scaling pressure in the PCA will be negligible. In the analysis conducted here, pressure was not considered.

Source terms for the PCs, s_η, can be parameterized by η and tabulated a priori to avoid run-time calculation. Accurate representation of s_η is crucial to this modeling strategy and will be addressed in Section 3.2.

In turbulent flow, closure is required for the source terms (convective terms can be treated in a manner consistent with other equations). Depending on the number of PCs and their spatial correlation, one may consider presumed PDF approaches, quadrature method of moments [18–20], or transported PDF approaches [21,22]. This paper will not address specific issues related to turbulent closure.

2.2.1. Obtaining the PCs

As discussed above, a successful application of PCA as a modeling approach requires parameterization of the source terms for each PC. This, in turn, requires that the data that PCA is applied to must have source terms for all Φ, which is currently impossible to obtain from experimental data. Therefore, this modeling approach will require availability of computational data generated from reliable chemical mechanisms using methods such as DNS. Furthermore, the reliability of PCA as a modeling approach also hinges on the relative invariance of PCs from one dataset to another that is nearby in parameter space.

3. Results

In this section, we present results of PCA applied to two DNS datasets of nonpremixed CO/H_2 combustion. The DNS datasets were calculated using a code with eighth-order spatial and fourth-order temporal discretization. Detailed kinetics of CO/H_2 oxidation were used [23,24], along with mixture-averaged transport approximations. The fuel stream was 0.45% CO, 0.05% H_2, and 0.5% N_2, giving a stoichiometric mixture fraction of $f_{st} = 0.4375$, and both fuel and air streams are at 300 K.

Case A is a spatially evolving jet with an initial $\chi_{max} = 25\ s^{-1}$, while case B is a temporally evolving jet with an initial $\chi_{max} = 125\ s^{-1}$. The primary difference between the two datasets is the initial scalar dissipation rate (χ) and turbulence intensity, which affects the degree of extinction observed; case A exhibits virtually no extinction, while case B exhibits moderate extinction. The existance of moderate extinction in case B is shown qualitatively in Fig. 1a, which shows T versus χ at f_{st}. Additional details of the DNS code and simulation configuration may be found elsewhere [4,15].

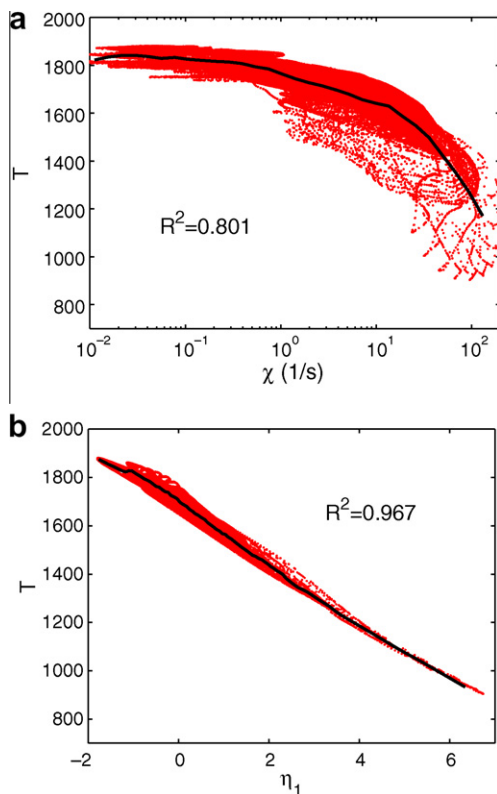To quantify the error in representing the data in a low-dimensional space parameterized by η, we calculate the R^2 value

Fig. 1. Parameterization of T at f_{st} by χ (a) and η_1 (b) for case B. Solid lines are the doubly-conditional mean temperature, $\langle T|_{f_{st},\eta_1}\rangle$. R^2 is calculated from Eq. (5). See the corresponding entries in Table 2.

$$R^2 = 1 - \left[\sum_{i=1}^{m}(\varphi_i - \varphi_i^*)^2\right]\left[\sum_{i=1}^{m}(\varphi_i - \bar{\varphi})^2\right]^{-1}, \quad (5)$$

where φ_i is the ith observation, φ_i^* is the parameterized approximation to φ_i, and $\bar{\varphi}$ is the mean of φ_i. In this sense, the R^2 value quantifies the ability of the PCs to parameterize a quantity φ.

3.1. Parameterizing the state variables

Here we present results of parameterization of the state variables. We choose to first condition the data on mixture fraction, f, since this is a convenient variable to "force" as the first component.

Figure 1a and b shows the parameterization (at f_{st}) of T by χ and the first PC, η_1, respectively, for case B. Examining Fig. 1b, we see that η_1 acts as a progress variable, capturing the extinction process remarkably well. This has also been observed for other choices of progress variables such as CO_2 (see, e.g. [4,14]). Comparing the two-parameter PCA approach with the (f, χ) parameterization is reasonable since both are two-parameter models, although the second parameter (χ versus η_1) repre-

sents different physical phenomena (gradient versus chemical state). Although not presented here, the analogous plot for case A shows virtually identical behavior, with the range on η_1 decreased and the minimum temperature increased as case A does not exhibit extinction. Figure 2 shows the parameterization of the OH mass fraction by the common (f, χ) and the proposed (f, η_1) parameterizations. This demonstrates that the PCA approach can be used to represent a wide range of the state variables, not temperature alone. Table 1 gives the weightings (a_i) on the state variables (ϕ_i) that define η_1, η_2, and η_3, where $\eta = \sum_{i=1}^{n} a_i \phi_i$, for the PCA applied at f_{st} on case B. From Table 1, it is clear that the PCs do not have a simple physical interpretation other than a linear combination of each of the original variables.

Also shown on Fig. 1a and b is the R^2 value as calculated by Eq. (5), which will be used as a quantitative measure for the quality of the parameterization. Table 2 lists R^2 values for reconstruction of the temperature and all species mass fractions in the detailed mechanism employed in the DNS for various n_η at f_{st}. These values are a concise, quantita-

Table 1
Definition of η_1 and η_2 in terms of the original state variables at f_{st} for case B. $\eta = \sum_{i=1}^{n_\phi} a_i \phi_i$

i	ϕ_i	a_i for η_1	a_i for η_2	a_i for η_3
1	T	−0.793	−0.2897	0.0503
2	H_2	0.095	0.2226	0.1509
3	O_2	0.222	0.1399	−0.1000
4	O	0.121	0.1438	0.5592
5	OH	−0.298	0.1885	0.4522
6	H_2O	−0.353	0.8607	−0.2476
7	H	0.093	0.1374	0.4124
8	HO_2	0.080	0.0432	−0.0703
9	H_2O_2	0.092	0.0707	0.0116
10	CO	0.232	0.1458	−0.0913
11	CO_2	−0.030	−0.0007	−0.4500
12	HCO	0.078	0.0540	0.0173

tive representation of the information presented graphically in Figs. 1 and 2. For example, for case B with $n_\eta = 1$ (i.e. a parameterization based on mixture fraction and a single PC), we obtain $R_T^2 = 0.967$, corresponding to Fig. 1b. For comparison, Table 2 also lists R^2 values for an (f, χ) parameterization, e.g. $R_T^2 = 0.801$ (see also Fig. 1a). Clearly, the two-parameter (f, η_1) parameterization reconstructs the temperature and most other state variables with much more accuracy than the (f, χ) parameterization. It should be noted that the results for the (f, χ) parameterization represent the best possible performance of a model based on (f, χ); the steady laminar flamelet model typically does not perform ideally [4].

Table 2 also demonstrates that increasing the number of retained PCs increases the accuracy with which the state variables are represented. This illustrates the point that one may select a desired error threshold and then determine the minimum number of PCs required to achieve that accuracy. Conversely, one may choose the number of PCs and estimate a priori the associated error.

3.2. Parameterizing source terms

As discussed in Section 2.1.1, the PCs are not conserved variables, and their source terms must also be parameterized. In this section, we explore the ability of PCA to parameterize source terms. Any function of $\mathbf{X}$ may be approximated by $\mathscr{F}(\mathbf{X}) \approx \mathscr{F}(\mathbf{A}\boldsymbol{\eta})$. However, it is more accurate to calculate $\mathscr{F}(\mathbf{X})$ directly from the data in n-dimensional space and then project it onto $\boldsymbol{\eta}$ by calculating the conditional mean $\langle \mathscr{F}(\mathbf{X})|_{\boldsymbol{\eta}} \rangle$ in n_η-dimensional space than to calculate $\mathscr{F}(\mathbf{A}\boldsymbol{\eta})$. Thus, source terms $(s_{\boldsymbol{\eta}})$ are calculated directly from the original observables, $\mathbf{X}$, and their conditional means (in n_η-dimensional space) are projected onto $\boldsymbol{\eta}$. Figure 3 illustrates this for the two-dimensional (f, η_1) parameterization where we require s_{η_1}. The realizations of s_{η_1} along with the conditional mean $\langle s_{\eta_1}|_{f_{st}, \eta_1} \rangle$ are shown as functions of η_1 for case B.
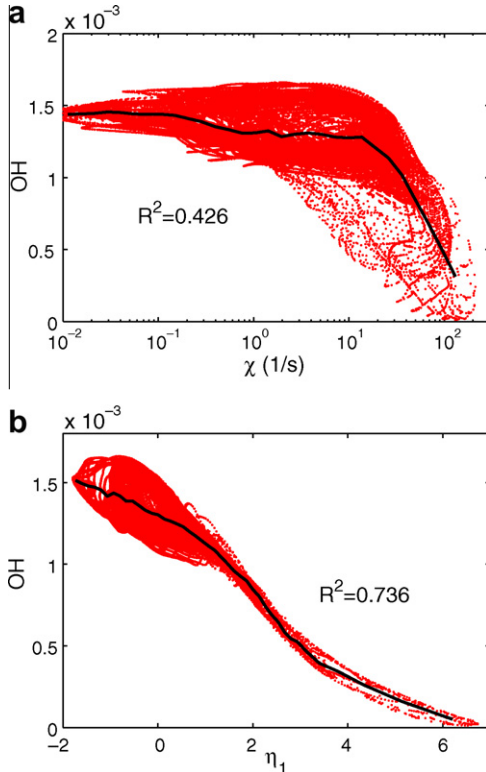

Fig. 2. Parameterization of OH at f_{st} by χ (a) and η_1 (b) for case B. Solid lines are the doubly-conditional mean OH value, $\langle Y_{OH}|_{f_{st}, \eta_1} \rangle$. R^2 is calculated from Eq. (5). See the corresponding entries in Table 2.

Table 2
R^2 values defined by Eq. (5). Also shown are results for the χ parameterization. All results are at $f = f_{st} = 0.4375$

|	n_η	T	H_2	O_2	O	OH	H_2O	H	HO_2	H_2O_2	CO	CO_2	HCO
Case A	χ	0.789	0.344	0.811	0.718	0.165	0.085	0.695	0.839	0.816	0.803	0.827	0.828
	1	0.983	0.259	0.976	0.930	0.240	0.178	0.823	0.986	0.916	0.978	0.956	0.980
	2	0.983	0.936	0.968	0.958	0.963	0.924	0.964	0.980	0.985	0.969	0.976	0.980
Case B	χ	0.801	0.509	0.807	0.697	0.426	0.186	0.648	0.665	0.729	0.810	0.058	0.817
	1	0.967	0.370	0.910	0.614	0.736	0.531	0.524	0.940	0.849	0.907	0.094	0.901
	2	0.996	0.845	0.982	0.882	0.931	0.990	0.858	0.974	0.941	0.981	0.378	0.984
	3	0.990	0.904	0.982	0.984	0.979	0.991	0.985	0.977	0.933	0.981	0.854	0.980
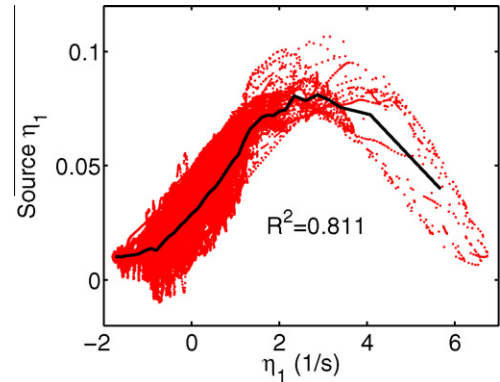
Fig. 3. Parameterization of s_{η_1} at f_{st} by χ (a) and η_1 (b) for case B. Solid lines are $\langle s_{\eta_1}|_{f_{st},\eta_1}\rangle$, the doubly-conditional mean value of s_{η_1}. R^2 is calculated from Eq. (5). See the corresponding entries in Table 2.

Table 3 summarizes the ability of an n_η-dimensional PCA to parameterize the source terms of the PCs. We first consider the column describing the results at f_{st}. For case A, a two-dimensional parameterization (f, η_1) captures s_{η_1} with $R^2_{s_{\eta_1}} = 0.993$. For case B, we require three PCs to parameterize s_{η_1} to a similar degree of accuracy. Comparing the dimensionality requirements for parameterizing s_η with those for parameterizing the state variables (see Table 2), we see that parameterizing the source terms does not require more PCs than parameterization of the state variables themselves, an encouraging result.

3.3. Global versus semi-local PCA

The results presented thus far have been obtained "locally" at f_{st}. One may consider whether a PCA performed at f_{st} is applicable at other f. We term this a "semi-local" PCA. If the PCA is highly dependent on mixture fraction, then one of two options must be considered:

- Eliminate the mixture fraction as a parameter and seek a global PCA on the entire dataset. This approach typically requires more PCs than a PCA obtained at f_{st} [25].
- Perform a local PCA in f-space and derive transport equations for $\eta|_f$. These equations would have exchange terms representing transport in mixture fraction space. This approach is further complicated by the fact that the very definition of the PCs would vary with f.

If the PCA obtained at f_{st} reasonably represents the data at other f, then the transport equations derived in Section 2.1.1 may be used directly at all f, eliminating the need for conditional equations in f-space.

Tables 4 and 5 provide parameterization errors for the state variables at $f = 0.2$ and $f = 0.6$, respectively. Table 3 shows the parameterization errors for s_η at $f = 0.2$ and $f = 0.6$. Interestingly, the parameterizations do not perform well at lean conditions; the same is true for the (f, χ) parameterization. A posteriori testing is necessary to fully determine the parameterization accuracy required. However, these results show promise for the ability to use a PCA obtained at f_{st} globally.

Table 3
R^2 values defined by Eq. (5) for PC source terms, s_η

Case	s_η	$f = 0.2$			$f = f_{st} = 0.4375$			$f = 0.6$		
		$n_\eta = 1$	$n_\eta = 2$	$n_\eta = 3$	$n_\eta = 1$	$n_\eta = 2$	$n_\eta = 3$	$n_\eta = 1$	$n_\eta = 2$	$n_\eta = 3$
A	s_{η_1}	0.993	0.985	–	0.978	0.985	–	0.923	0.934	–
	s_{η_2}	–	0.996	–	–	0.922	–	–	0.876	–
B	s_{η_1}	0.270	0.844	0.967	0.815	0.932	0.958	0.809	0.852	0.902
	s_{η_2}	–	0.835	0.955	–	0.951	0.961	–	0.883	0.909
	s_{η_3}	–	–	0.976	–	–	0.731	–	–	0.831

Table 4
R^2 values at $f = 0.2$ using the PCA obtained at f_{st}. Also shown are results for the χ parameterization

	n_η	T	H_2	O_2	O	OH	H_2O	H	HO_2	H_2O_2	CO	CO_2	HCO
Case A	χ	0.097	0.798	0.169	0.774	0.736	0.245	0.827	0.812	0.811	0.580	0.432	0.881
	1	0.500	0.413	0.816	0.212	0.188	0.134	0.319	0.433	0.398	0.666	0.555	0.619
	2	0.968	0.910	0.881	0.868	0.859	0.940	0.888	0.838	0.855	0.867	0.940	0.934
Case B	χ	0.497	0.542	0.390	0.303	0.329	0.269	0.558	0.537	0.390	0.417	0.206	0.689
	1	0.979	0.741	0.866	0.337	0.219	0.749	0.127	0.805	0.858	0.859	0.513	0.403
	2	0.996	0.877	0.945	0.819	0.822	0.994	0.806	0.970	0.960	0.958	0.737	0.860
	3	0.990	0.963	0.958	0.989	0.977	0.994	0.978	0.984	0.982	0.968	0.808	0.955

Table 5
R^2 values at $f = 0.6$ using the PCA obtained at f_{st}. Also shown are results for the χ parameterization

	n_η	T	H_2	O_2	O	OH	H_2O	H	HO_2	H_2O_2	CO	CO_2	HCO
Case A	χ	0.676	0.190	0.740	0.642	0.548	0.073	0.434	0.741	0.727	0.467	0.572	0.555
	1	0.956	0.287	0.958	0.887	0.587	0.076	0.542	0.966	0.867	0.836	0.868	0.751
	2	0.959	0.962	0.949	0.923	0.775	0.826	0.768	0.955	0.898	0.911	0.919	0.889
Case B	χ	0.628	0.081	0.593	0.662	0.112	0.246	0.365	0.508	0.616	0.521	0.268	0.570
	1	0.964	0.134	0.904	0.804	0.197	0.755	0.721	0.938	0.650	0.844	0.442	0.896
	2	0.984	0.612	0.928	0.836	0.373	0.986	0.822	0.960	0.791	0.873	0.542	0.930
	3	0.986	0.769	0.948	0.888	0.543	0.991	0.913	0.967	0.839	0.909	0.841	0.941

4. Conclusions

A modeling approach based on PCA has been proposed and tested a priori using DNS data. This modeling approach is complete, with the exception of a turbulent closure model which would be required if this model were used in a LES or RANS context. The model is based on a rotation of the thermochemical state basis from one based on temperature, pressure, and $n_s - 1$ species mass fractions to one which best represents the variance in the data. PCA has the convenient property that a truncated basis provides an optimal representation of the data. Implementation of the model requires transport equations for the principal components, which are reacting scalars.

Results from a quantitative a priori analysis of this approach using DNS data show great promise. State variables and source terms both are parameterized well by a two-parameter (f, η_1) model, and adding additional parameters provides a significant increase in accuracy for all state variables and their reaction rates. Results also indicate a uniformly better representation of the DNS data using an (f, η_1) parameterization over the commonly used (f, χ).

There are many potential applications of this modeling approach. For example, laminar flame calculations could benefit from PCA modeling approaches to provide rapid solutions using a reduced set of equations. Once a full calculation has been performed, subsequent calculations may be performed using PCs rather than the full set of species and energy equations. The number of PCs retained can be chosen by the desired accuracy. This could be particularly useful for stochas-

tic models such as the linear eddy model (LEM) [26,27] and the one-dimensional turbulence (ODT) model [28,29], which require many realizations of a flow field. The first realization could employ full chemistry while subsequent realizations utilize a reduced set of equations defined by PCA. Another application is in modeling turbulent flows, where a compact parameterization of the thermochemical state is key to achieving affordable simulations. Additionally, while the analysis presented herein has been applied to nonpremixed combustion, the PCA approach applies in principle to all combustion regimes from premixed to nonpremixed.

For application to turbulent flows, additional closure models are required for the unresolved convective and source terms. In the context of transported PDF methods, a PCA modeling approach could drastically reduce the computational cost by significantly reducing the thermochemical dimensionality while maintaining a quantified error bound on the thermochemical reduction.

Future work will focus on examining the feasibility of PCA with various fuels and exploring the universality of the PCA, i.e. the applicability of a PCA obtained under one set of conditions to be applied to a simulation at another set of conditions. Also, a posteriori tests will be conducted to determine the effect of nonlinear propagation of errors in source term parameterization.

The PCA modeling approach proposed herein requires high-fidelity, fully-resolved data including reaction rates. For this reason, we plan to consider LEM and ODT as methods to generate surrogate DNS data. These methods are signifi-

cantly less expensive than DNS, but it remains to be seen in reacting flow whether such methods can be used to reliably reproduce statistics obtained from DNS data.

References

[1] S.B. Pope, *Turbulent Flows*, Cambridge University Press, New York, 2000.

[2] G.P. Smith, D.M. Golden, M. Frenklach, et al., W.C.Gardiner Jr., V.V. Lissianski, Z. Qin, available at http://www.me.berkeley.edu/gri_mech.

[3] U. Maas, S.B. Pope, *Proc. Combust. Inst.* 24 (1992) 103–112.

[4] J.C. Sutherland, P.J. Smith, J.H. Chen, *Combust. Theor. Model.* 11 (2) (2007) 287–303.

[5] T. Lu, C.K. Law, *Combust. Flame* 126 (2001) 1445–1455.

[6] B. Bhattacharjee, D.A. Schwer, P.I. Barton, W.H. Green, *Combust. Flame* 135 (3) (2003) 191–208.

[7] M. Valorani, H.N. Najm, D.A. Goussis, *Combust. Flame* 134 (1–2) (2003) 35–53.

[8] T.F. Lu, C.K. Law, *Proc. Combust. Inst.* 30 (2004) 1333–1341.

[9] Y. Rastigejev, M.P. Brenner, D.J. Jacob, *Proc. Natl. Acad. Sci. USA* 104 (35) (2007) 13875–13880.

[10] F.A. Williams, *Combustion Theory*, second ed., Perseus Books, Cambridge, 1985.

[11] J.C. Sutherland, P.J. Smith, J.H. Chen, *Combust. Theor. Model.* 9 (2) (2005) 365–383.

[12] N. Peters, *Prog. Energy Combust. Sci.* 10 (1984) 319–339.

[13] J.A. van Oijen, L. de Goey, *Combust. Theor. Model.* 6 (2002) 463–478.

[14] C.D. Pierce, P. Moin, *J. Fluid Mech.* 504 (2004) 73–97.

[15] J.C. Sutherland, *Evaluation of Mixing and Reaction Models for Large-Eddy Simulation of Nonpremixed Combustion Using Direct Numerical Simulation*, Ph.D. thesis, University of Utah, Salt Lake City, UT, May 2004.

[16] J.E. Jackson, *A User's Guide to Principal Component Analysis*, Wiley Series in Probability and Statistics, 1991.

[17] I.T. Jolliffe, *Principal Component Analysis*, second ed., Springer, New York, 2002.

[18] K. Tsai, P. Gillis, S. Sen, R.O. Fox, *Trans. ASME* 124 (2002) 102–107.

[19] R.O. Fox, *Computational Models for Turbulent Reacting Flows*, Cambridge University Press, Cambridge, 2003.

[20] D.L. Marchisio, R.O. Fox, *J. Aerosol Sci.* 36 (1) (2005) 43–73.

[21] S.B. Pope, *Prog. Energy Combust. Sci.* 11 (1984) 119–195.

[22] W. Kollmann, *Theor. Comput. Fluid Dyn.* 1 (1990) 285–349.

[23] R.A. Yetter, F.L. Dryer, H. Rabitz, *Combust. Sci. Technol.* 79 (1–3) (1991) 97–128.

[24] M.A. Mueller, T.J. Kim, R.A. Yetter, F.L. Dreyer, *Int. J. Chem. Kinet.* 31 (2) (1999) 113–125.

[25] A. Parente, J.C. Sutherland, L. Tognotti, P.J. Smith, *Identification and Parameterization of Low-Dimensional Manifolds in Turbulent Flames*, Paper 07F-06, WSSCI Fall Meeting, Livermore, CA, October 2007, available at http://www.che.utah.edu/~sutherland/papers/Parente_et_al_WSS-CI_07_#07F-06.pdf.

[26] P.A. McMurtry, S. Menon, A.R. Kerstein, *Proc. Combust. Inst.* 24 (1992) 271–278.

[27] S. Menon, P. McMurtry, A.R. Kerstein, in: B. Galperin, S. Orszag (Eds.), *LES of Complex Engineering and Geophysical Flows*, Cambridge University Press, Cambridge, 1993, pp. 287–314.

[28] T. Echekki, A.R. Kerstein, T.D. Dreeben, J. Chen, *Combust. Flame* 125 (2001) 1083–1105.

[29] J.C. Hewson, A.R. Kerstein, *Combust. Sci. Technol.* 174 (2002) 35–66.