Combustion and Flame 160 (2013) 1967–1980

Contents lists available at SciVerse ScienceDirect

Combustion and Flame

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

Empirical low-dimensional manifolds in composition space

Yue Yang [a,b,*], Stephen B. Pope [a], Jacqueline H. Chen [b]

[a] Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14853, USA
[b] Combustion Research Facility, Sandia National Laboratories, Livermore, CA 94551-0969, USA

ARTICLE INFO

Article history:
Received 30 October 2012
Received in revised form 2 April 2013
Accepted 13 April 2013
Available online 11 May 2013

Keywords:
Turbulent combustion
Empirical low-dimensional manifold
Principal component analysis
Multivariate adaptive spline regression

ABSTRACT

To reduce the computational cost of turbulent combustion simulations with a detailed chemical mechanism, it is useful to find a low-dimensional manifold in composition space. Most previous low-dimensional manifolds in turbulent combustion are based on the governing conservation equations or thermochemistry and their application involves certain assumptions. On the other hand, empirical low-dimensional manifolds (ELDMs) are constructed based on samples of the compositions observed in experiments or in direct numerical simulation (DNS). Plane and curved ELDMs can be obtained using principal component analysis (PCA) and multivariate adaptive spline regression (MARS), respectively. The framework for ELDMs based on the represented compositions and principal components is considered in this study, where the represented compositions are selected from the PCA results.

Both PCA and MARS are applied to the DNS databases of a non-premixed CO/H_2 temporally evolving jet flame and of an ethylene lifted jet flame. It is more accurate to represent the species mass fractions by curved MARS ELDMs than by plane PCA ELDMs. To achieve a overall departure less than a given level, more dimensions are required for the ethylene case (which involves 22 species) than for the CO/H_2 case (which involves 11 species). For MARS to achieve less than the 5% departure level, seven dimensions are required for the ethylene case, and just two dimensions are required for the CO/H_2 case. However, it is much more challenging to obtain the ELDMs with high accuracy for the chemical source terms using a small number of dimensions. In addition, the effects on the departure from ELDMs of the scaling method in PCA, local extinction, and the Reynolds number are discussed. Two different approximations for the chemical source term are compared with discussions for further *a posteriori* simulations.

1. Introduction

With the development of the detailed description of combustion chemistry for hydrocarbon fuels, detailed or reduced chemical mechanisms have been used within turbulent combustion computations [1]. However, the computational cost of such simulations is usually very high, e.g., a direct numerical simulation (DNS) of a high-Reynolds-number jet flame with a reduced mechanism of order 20 species may take millions of CPU hours on a supercomputer [2]. The large number of species in the detailed chemical mechanism imply that the number of dimensions in composition space is very high. Hence, to reduce the computational cost, it is useful to find a low-dimensional manifold in the composition space that can approximate the full system dynamics of the reacting flow to reduce the number of transport equations to be solved. Let the composition variables be partitioned into a set of *represented compositions* and the remaining *unrepresented compositions*. The low-dimensional manifold can be considered as a smooth non-linear multivariate function of a small number of selected, represented compositions.

A general classification of these low-dimensional manifolds is given in [3]. Most previous low-dimensional manifolds in turbulent combustion are based on the governing conservation equations or thermochemistry. The six classes of manifolds identified in [3] are as follows: (1) *Skeletal manifolds* are based on skeletal mechanisms which neglect some species and reactions. (2) *Thermodynamic manifolds* are determined by the thermodynamic properties of the system, which are known functions of species mass fractions, e.g., the constrained equilibrium manifold (CEM) used in the rate-controlled constrained equilibrium (RCCE) method [4,5]. (3) *Reaction manifolds* are based on the autonomous set of ordinary differential equations describing the change of composition due to reaction in isobaric, adiabatic autoignition, e.g., intrinsic low-dimensional manifold (ILDM) [6] and quasi steady-state (QSS) manifolds [7], and those implied by computational singular perturbation (CSP) [8]. (4) *Diffusion manifolds* are parameterized by mixture fractions or linearly independent species in the inert mixing [9]. (5) *Reaction–diffusion manifolds* are obtained as the solution of ordinary or partial differential equations which contain both diffusion and chemical source terms, e.g., the steady flamelet

* Corresponding author. Address: 138 Upson Hall, Cornell University, Ithaca, NY 14853, USA. Fax: +1 607 255 1222.
E-mail address: yy463@cornell.edu (Y. Yang).

0010-2180/$ - see front matter © 2013 The Combustion Institute. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.combustflame.2013.04.006

model [10], the flamelet/progress variable model (FPV) [11], flamelet-generated manifolds (FGM) [12], and the REDIM method [13]. (6) *Conditional manifolds* are defined as the expectation of unrepresented species conditional on represented compositions, e.g., conditional moments closures (CMC) [14,15] and in multiple mapping conditioning (MMC) [16].

The manifolds described above are all based on theoretical analyses and their application involves certain assumptions. In contrast, *empirical low-dimensional manifolds* (ELDMs) are constructed from samples of the compositions observed in experiments or in DNS. The simplest ELDM is the plane manifold obtained from the principal component analysis (PCA) [17,18]. On the other hand, representing and determining "best-fit" curved manifolds (with non-zero curvature) is much more challenging than PCA. The multivariate adaptive spline regression (MARS) [19,20] has been shown as a promising approach to fitting high-dimensional non-linear functions. In the present context, MARS can be used to approximate non-linear functions such as the unrepresented species on a curved manifold and the chemical source term. It is noted that the curved ELDM obtained from MARS is an estimate of the conditional manifold [14–16]. Sutherland et al. [21–25] have examined PCA ELDMs and MARS ELDMs based on data from experiments, DNS, and from one-dimensional turbulence (ODT) simulations. The PCA of species mass fractions can be also used as a systematic approach to define progress variables in the tabulation of low-dimensional manifolds in composition space [26].

Besides using PCA and MARS, several other dimension reduction methods have been previously used to find ELDMs. From detailed chemical calculations, the ELDMs (the term "repro-models" are used instead in [27,28]) were extracted and stored in the form of high-order multivariate polynomials, and the simulations of the combustion of wet CO with two- or three-dimensional ELDMs have been shown to be in the order of 10^4 faster than the calculation with the full chemistry model [27]. The algorithm for the ELDM based generation of ILDMs has been developed to produce the low-dimensional manifold using the spline fitting, which is applied to the simulation of the combustion of hydrogen in air [28]. A correlation analysis of two-dimensional DNS data of a turbulent non-premixed H_2/air flame with detailed chemistry has been used to find the ELDMs [29]. The proper orthogonal decomposition (POD) analysis has been applied to obtain the low-dimensional representations of the DNS data for H_2/air flames [30,31] and to simplify atmospheric chemistry [32] using a small number of POD modes. Here, the POD is suggested to be applied locally by dividing the composition space into sub-domains and then different representations can be used in different sub-domains [32]. In addition, an isomap based technique [33] has been applied to the DNS data of autoignition and front propagation in mixtures of dimethyl-ether (DME)/air to identify curved manifolds [34].

Several low-dimensional manifolds have been applied to the numerical simulation of turbulent combustion, for example, the reduced DNS with ELDMs [27,35], the large-eddy simulation (LES) with FPV [11,36], and the LES/probability density function (PDF) modeling with RCCE [37]. However, the assumption of the low-dimensional manifolds in composition space and the accuracy of the approximation they provide still need to be examined.

For *a posteriori* tests, the number of transport equations of compositions can be reduced via the application of PCA. The ELDMs identified by the PCA-based model are insensitive to filtering, which suggests that it can be suitable for use in LES [24]. Based on the ELDMs with an appropriate approximation of the chemical source term, it is useful to select a small number of represented compositions so that their transport equations can be solved without the transformation of PCA.

In this study, we perform *a priori* examinations based on the DNS data of a CO/H_2 non-premixed jet flame (which involves 11 species) [38] and of an ethylene lifted jet flame (which involves 22 species) [39] at moderate Reynolds numbers. The main objective to understand is that, for a given dataset of N samples of species mass fractions, thermodynamic properties, and chemical source terms, how closely do the PCA and MARS manifolds of dimension n_r approximate the data. The concept of ELDMs is reviewed in Section 2. The two DNS datasets used in the present study are described in Section 3. The methodologies of the construction of ELDMs and corresponding results are given in Sections 4 and 5, respectively. Some conclusions are drawn in Section 6.

Compared to the prior work investigating ELDMs using PCA and MARS [21,24], this study makes original contributions as follows: (1) A framework of the ELDM based on represented compositions which are selected from the PCA results is developed. The transport equations of represented compositions can be applied readily in *a posteriori* simulations without the PCA transformation. (2) The high-fidelity DNS datasets used in this study contain all of the composition variables and chemical source terms, and they are more accurate and complete than the one-dimensional turbulence data and the experimental results without species reaction rates used in [24]. The reduced ethylene–air mechanism [40] in the DNS dataset [39] involves 22 species, which is greater than the 11 species in the skeletal CO/H_2 mechanism used in [21,24] and this study. The results based on the two different mechanisms can show the effect of the chemical mechanisms with different numbers of species on the accuracy of ELDMs. (3) A more sensitive measure to examine the root-mean-squares (r.m.s.) departure of the ELDM is adopted. The departure is a function of the dimension of the ELDM. The number of the dimension of the ELDM investigated in this work is up to 10 compared to two and three dimensions considered in [21,24], which can clarify the decaying behavior of the departure more clearly.

2. ELDMs in composition space

2.1. Represented and unrepresented compositions

In general, an n_ϕ-dimensional composition space is spanned by a set of composition vectors $\boldsymbol{\phi} = \{\phi_1, \ldots, \phi_{n_\phi}\}$, which contain n_s species mass fractions and $n_\phi - n_s$ thermodynamic properties. We partition the compositions into a set of n_r represented compositions and the remaining $n_u \equiv n_\phi - n_r$ unrepresented compositions. By suitably ordering the represented and unrepresented compositions, the composition vector can be written as

$$\boldsymbol{\phi} = \begin{bmatrix} \boldsymbol{\phi}^r \\ \boldsymbol{\phi}^u \end{bmatrix}, \tag{1}$$

where $\boldsymbol{\phi}^r$ and $\boldsymbol{\phi}^u$ are n_r- and n_u-vectors in the represented and unrepresented subspaces, respectively.

An n_r-dimensional manifold in the n_ϕ-dimensional composition space (with $1 \leqslant n_r < n_\phi$) can be described as a function $\boldsymbol{\Phi}^M(\boldsymbol{\theta})$ with n_r variable $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_{n_r}\}$ which are the suitable linear combinations of $\boldsymbol{\phi}$ as

$$\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \boldsymbol{R}\boldsymbol{\phi}, \tag{2}$$

where $\boldsymbol{\theta}_0$ is a constant n_r-vector and $\boldsymbol{R}$ is an $n_r \times n_\phi$ matrix. In particular, we consider $\boldsymbol{\theta} = \boldsymbol{\phi}_r$. This implies that compositions occurring in turbulent combustion lie close to a low-dimensional manifold $\boldsymbol{\Phi}^M(\boldsymbol{\phi}^r)$, i.e., with $\boldsymbol{\phi}^r$ being used as the parameters, the manifold can be expressed as a function

$$\boldsymbol{\phi} \approx \boldsymbol{\Phi}^M(\boldsymbol{\phi}^r) = \begin{bmatrix} \boldsymbol{\phi}^r \\ \boldsymbol{\Phi}^m(\boldsymbol{\phi}^r) \end{bmatrix}, \tag{3}$$

where $\boldsymbol{\Phi}^m(\boldsymbol{\phi}^r)$ is a mapping from the represented subspace to the unrepresented subspace. The manifold such as $\boldsymbol{\Phi}^M(\boldsymbol{\phi}^r)$ is defined as an ELDM if it is constructed from observations of the composition in experiments or in DNS. We remark that it is not true that any n_r-dimensional single-valued and realizable manifold can be expressed as $\boldsymbol{\Phi}^M(\boldsymbol{\phi}^r)$ for any choice of $\boldsymbol{\phi}^r$ [3].

2.2. Approximations for the chemical source term

Similar to Eq. (1), the chemical source term $\boldsymbol{S}$ (i.e., the rate of change of $\boldsymbol{Y}$ due to chemical reactions) can be partitioned as

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}^r \\ \boldsymbol{S}^u \end{bmatrix}, \tag{4}$$

and $\boldsymbol{S}$ can be approximated by a manifold parameterized by represented compositions $\boldsymbol{\phi}^r$ as

$$\boldsymbol{S} \approx \widehat{\boldsymbol{S}}^M(\boldsymbol{\phi}^r) = \begin{bmatrix} \widehat{\boldsymbol{S}}^{r,m}(\boldsymbol{\phi}^r) \\ \widehat{\boldsymbol{S}}^{u,m}(\boldsymbol{\phi}^r) \end{bmatrix}, \tag{5}$$

or the linear combinations θ of $\boldsymbol{\phi}$, where $\widehat{\boldsymbol{S}}^{r,m}(\boldsymbol{\phi}^r)$ and $\widehat{\boldsymbol{S}}^{u,m}(\boldsymbol{\phi}^r)$ are mappings from the subspace for represented compositions to the subspace for chemical source terms for represented and for unrepresented compositions, respectively. Thus, the chemical source term $\boldsymbol{S}^r$ for represented compositions is approximated by a non-linear function in terms of $\boldsymbol{\phi}^r$ as

$$\boldsymbol{S}^r \approx \widehat{\boldsymbol{S}}^{r,m}(\boldsymbol{\phi}^r). \tag{6}$$

Alternatively, since $\boldsymbol{S} = \widehat{\boldsymbol{S}}(\boldsymbol{\phi})$ is a non-linear function $\widehat{\boldsymbol{S}}$ of compositions, which is determined by a specific chemical mechanism and thermochemistry properties, we can approximate the chemical source term $\boldsymbol{S}^r$ for represented compositions as

$$\boldsymbol{S}^r \approx \widehat{\boldsymbol{S}}^r(\boldsymbol{\Phi}^M(\boldsymbol{\phi}^r)) = \widehat{\boldsymbol{S}}^r\left(\begin{bmatrix} \boldsymbol{\phi}^r \\ \boldsymbol{\Phi}^m(\boldsymbol{\phi}^r) \end{bmatrix}\right), \tag{7}$$

where $\boldsymbol{S}^r$ is determined by the non-linear function $\widehat{\boldsymbol{S}}^r$ of all the compositions lying on the manifold $\boldsymbol{\Phi}^M(\boldsymbol{\phi}^r)$.

2.3. Measures of the departure for ELDMs

In this study, we consider the low-Mach-number flow of a reactive ideal gas mixture. The composition vector $\boldsymbol{\phi} = \{\boldsymbol{Y}, T\}$ contains n_s species mass fractions $\boldsymbol{Y} = \{Y_\alpha, \alpha = 1, \ldots, n_s\}$ and temperature T. For a given dataset, an ELDM $\boldsymbol{\Phi}^M(\boldsymbol{\phi}^r)$ for

$$\boldsymbol{\phi} = \begin{bmatrix} \boldsymbol{\phi}^r \\ \boldsymbol{\phi}^u \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}^r \\ \boldsymbol{\Phi}^m(\boldsymbol{\phi}^r) + \boldsymbol{\phi}'^u \end{bmatrix}, \tag{8}$$

can be constructed from an ensemble of N observations of $\boldsymbol{\phi}$ by minimizing some measure of the departures of the observations from the manifold, where

$$\boldsymbol{\phi}' \equiv \boldsymbol{\phi} - \boldsymbol{\Phi}^M(\boldsymbol{\phi}^r) = \begin{bmatrix} 0 \\ \boldsymbol{\phi}'^u \end{bmatrix} \tag{9}$$

is the departure from the n_r-dimensional manifold with an n_u-vector $\boldsymbol{\phi}'^u$, and the unrepresented composition is decomposed as

$$\boldsymbol{\phi}^u = \boldsymbol{\Phi}^m(\boldsymbol{\phi}^r) + \boldsymbol{\phi}'^u. \tag{10}$$

Similarly, we can define the departure from a manifold in $\boldsymbol{S}$-space parameterized by represented compositions $\boldsymbol{\phi}^r$ as

$$\boldsymbol{s}' \equiv \boldsymbol{S} - \widehat{\boldsymbol{S}}^M(\boldsymbol{\phi}^r). \tag{11}$$

To examine how closely does the ELDM of dimension n_r approximate the data, we use the normalized r.m.s. departure of the data from the manifold. Specifically, for a composition i we define

$$\epsilon_i = \sqrt{\langle v_i'^2 \rangle}/\sigma_i, \tag{12}$$

where $\langle \cdot \rangle$ denotes the ensemble average over N observations, v_i is the i-component of a vector for the departure, and $\sigma_i \equiv \sqrt{\langle (v_i' - \langle v_i' \rangle)^2 \rangle}$ is the standard deviation of v_i'. The choice of v_i' in ϵ_i can be ϕ_i' from Eq. (9) and s_i' from Eq. (11) to measure the departure of the species mass fractions and of the chemical source terms, respectively. In addition, the maximum departure is defined as

$$\epsilon_{i,max} = \max\left(|v_i'|\right)/\sigma_i \tag{13}$$

to measure the maximum departure from an ELDM in N observations.

Sutherland and co-workers [21,24] use another measure, the R^2 values to measure the accuracy of an ELDM, where

$$R_i^2 = 1 - \epsilon_i^2. \tag{14}$$

It is noted that ϵ_i is more sensitive than R_i^2 to measure the level of departure, e.g., $\epsilon_i = 5\%$ is equivalent to $R_i^2 = 0.9975$, and the decaying behavior of ϵ_i with the dimension n_r can be examined in the logarithm scale.

2.4. Transport equations

The transport equations for $\boldsymbol{\phi}$[1] can be written as

$$\rho \frac{\mathrm{D}\boldsymbol{\phi}}{\mathrm{D}t} = -\nabla \cdot (\boldsymbol{j}) + \boldsymbol{S}, \tag{15}$$

where $\frac{\mathrm{D}}{\mathrm{D}t} \equiv \rho \frac{\partial}{\partial t} + \boldsymbol{u} \cdot \nabla$ is the material derivative, $\boldsymbol{u}$ is the mass-averaged velocity, $\boldsymbol{j}$ is the diffusive flux of $\rho\boldsymbol{\phi}$. The transport equations for represented compositions are simply

$$\rho \frac{\mathrm{D}\boldsymbol{\phi}^r}{\mathrm{D}t} = -\nabla \cdot (\boldsymbol{j}^r) + \boldsymbol{S}^r, \tag{16}$$

where $\boldsymbol{j}^r$ is the diffusive flux of $\rho\boldsymbol{\phi}^r$. To apply the ELDMs to *a posteriori* numerical simulations and obtain all the mass fractions, we need to approximate unrepresented compositions $\boldsymbol{\phi}^u \approx \boldsymbol{\Phi}^m(\boldsymbol{\phi}^r)$ from Eq. (10) and chemical source terms for represented compositions $\boldsymbol{S}^r$ by Eq. (6) or Eq. (7).

3. DNS databases

3.1. CO/H_2 temporally evolving jet flame

The DNS of a non-premixed CO/H_2 temporally evolving planar jet flame was preformed by Hawkes et al. [38]. The fuel is composed of 50% CO, 10% H_2 and 40% N_2 by volume, which represents syngas. The co-flow is composed of 25% O_2 and 75% N_2. The temperature of the co-flow is 500 K, and the pressure is set to atmospheric. The mixture-averaged transport properties for 11 species (H_2, O_2, O, OH, H_2O, H, HO_2, CO, CO_2, HCO and N_2) and the skeletal CO/H_2 mechanism with 21 reaction steps [38] based on a detailed C1-kinetics [41] were used.

The flow is statistically one-dimensional, with statistics varying only in the cross-stream direction and time. At $t = 0$, the flow consists of a slab of thickness H of fuel. The characteristic time scale is defined as $t_j = H/U$, where U is the bulk jet velocity. The evolution of the turbulent jet flame is triggered by small velocity fluctuations near the central jet. The flow Reynolds number is defined by $Re = UH/\nu_f$ where ν_f is the kinematic viscosity of the pure fuel stream. We use two DNS datasets at $Re = 9079$ and $Re = 2510$. The DNS dataset at the high Re is henceforth referred to as

[1] The total enthalpy may be used instead of temperature in $\boldsymbol{\phi}$ to naturally satisfy a conservation equation in the form of Eq. (15).

Table 1
Scaling methods for PCA.

Method	Scaling parameter s_i
No scaling	1
STD	σ_i
VAST	$\sigma_i/\langle\phi_i\rangle$
Range	$\max(\phi_i) - \min(\phi_i)$
Level	$\langle\phi_i\rangle$
Max	$\max(\phi_i)$
Pareto	$\sqrt{\sigma_i}$

"DNS-syngas" in this study. The Damköhler number is $Da = \chi_q t_j = 0.011$ for both cases, which is low enough to result in local extinction, where χ_q is the steady extinction scalar dissipation rate of a laminar flame.

3.2. Ethylene lifted jet flame

The DNS of a turbulent lifted ethylene planar jet flame (referred to as "DNS-ethylene") was performed by Yoo et al. [39]. The fuel is composed of 18% C_2H_4 and 82% N_2 by volume, and the co-flow is air. The temperature of the co-flow is 1550 K, and the pressure is set to atmospheric. The central C_2H_4/N_2 jet is self-ignited by the highly heated coflow, and then the lifted jet flame base stabilization reaches a statistically stationary state. Based on the detailed mechanism for ethylene oxidation with 75 species and 529 elementary reactions [42], a skeletal mechanism with 32 species and 206 reactions was obtained. The quasi steady-state approximations (QSSA) were applied to 10 fast depleting radicals to further reduce the number of transported variables. Then a reduced ethylene–air mechanism was derived [40] and used in the DNS, which comprises 22 species (H_2, H, O, 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, HCCO, CH_2CO, CH_3CHO, α-C_3H_5, C_3H_6 and N_2) and 18 global reaction steps. It is noted that both mechanisms used in DNS-syngas and DNS-ethylene are non-stiff for the respective DNS [1,40].

4. Construction of ELDMs

4.1. PCA

A plane-ELDM can be obtained by applying PCA to a DNS or experimental dataset. The PCA is a general methodology to reduce a complex dataset to a lower dimension, which can retain as much as possible of the variation to reveal the simplified underlying structure of the data [17,18]. Let $\boldsymbol{A}$ be an $n_\phi \times N$ matrix for the centered original data with appropriate scaling

$$\boldsymbol{\phi}^* \equiv \boldsymbol{D}_s^{-1}(\boldsymbol{\phi} - \langle\boldsymbol{\phi}\rangle), \tag{17}$$

where N is the total number of observations, $\boldsymbol{D}_s = \mathrm{diag}(s_1, \ldots, s_{n_\phi})$ is a diagonal matrix with the scaling parameter s_i, and $\langle\boldsymbol{\phi}\rangle$ is the mean composition vector, where each component $\langle\phi_i\rangle$ is the ensemble average over N observations of ϕ_i. The N observations can be sampled from the DNS database at a given time or several given times by selecting every n_dth data point starting with the first data point, where the integer n_d denotes the offset. In PCA, the number of observations N should be sufficiently large so that the PCA result is not sensitive to the change in N.

The scaling methods considered in [24] and the present study are listed in Table 1. The PCA results can be significantly influenced by the scaling method [17], which is discussed in Section 5.3. Instead of using Eq. (17), non-linear scalings could be used especially for minor species; but only linear scalings are used in the present work. Unless otherwise specified, the PCA results in this study are based on the data without scaling (i.e., $s_i = 1$).

A PCA re-representation $\boldsymbol{B}$ of the data $\boldsymbol{\phi}^*$ can be written as

$$\boldsymbol{B} = \boldsymbol{P}\boldsymbol{A}. \tag{18}$$

The unit row vectors $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{n_\phi}\}$ of the orthonormal $n_\phi \times n_\phi$ matrix $\boldsymbol{P}$ are principal components (PCs) of $\boldsymbol{A}$ and they are normal to each other. The matrix $\boldsymbol{P}$ for PCs is obtained by performing an eigenvector decomposition $\boldsymbol{C}_A = \boldsymbol{P}^T\boldsymbol{\Lambda}\boldsymbol{P}$ for the covariance matrix

$$\boldsymbol{C}_A \equiv \frac{1}{N-1}\boldsymbol{A}\boldsymbol{A}^T, \tag{19}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n_\phi})$ is a diagonal matrix with the eigenvalues λ_i of $\boldsymbol{C}_A$ and $\lambda_1 \geqslant \cdots \geqslant \lambda_{n_\phi}$. Besides the eigenvalue decomposition for $\boldsymbol{C}_A$, we can also perform the singular value decomposition (SVD) of $\boldsymbol{A}$ to obtain $\boldsymbol{P}$ in the implementation [17,18].

The original composition vector can be projected onto the PC basis vectors as

$$\boldsymbol{\eta} = \boldsymbol{P}\boldsymbol{\phi}^* = \boldsymbol{P}\boldsymbol{D}_s^{-1}(\boldsymbol{\phi} - \langle\boldsymbol{\phi}\rangle), \tag{20}$$

where $\boldsymbol{\eta} = \{\eta_1, \ldots, \eta_{n_\phi}\}$, which is a suitable choice of $\boldsymbol{\theta}$ in Eq. (2). The eigenvalue λ_i represents the variance of η_i along the PC direction p_i, so most of the variation in the data can be retained in the first several PCs.

The matrix $\boldsymbol{P}$ can be partitioned as

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{P}^r \\ \boldsymbol{P}^u \end{bmatrix} \tag{21}$$

where $\boldsymbol{P}^r$ contains the first n_r PCs $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{n_r}\}$, and correspondingly the matrix $\boldsymbol{B}$ is partitioned as

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}^r \\ \boldsymbol{B}^u \end{bmatrix}, \tag{22}$$

where $\boldsymbol{P}^r$ and $\boldsymbol{P}^u$ are $n_r \times n_\phi$ and $n_u \times n_\phi$ matrices, respectively, and $\boldsymbol{B}^r$ and $\boldsymbol{B}^u$ are $n_r \times N$ and $n_u \times N$ matrices, respectively.

Using n_r PCs, the approximation of the original data $\boldsymbol{A}$ can be expressed as

$$\boldsymbol{A}' = \boldsymbol{P}^{rT}\boldsymbol{B}^r. \tag{23}$$

This approximation corresponds to a plane PCA manifold

$$\boldsymbol{\phi} \approx \boldsymbol{\Phi}^M(\boldsymbol{\eta}^r) = \langle\boldsymbol{\phi}\rangle + \boldsymbol{D}_s\boldsymbol{P}^{rT}\boldsymbol{\eta}^r \tag{24}$$

in terms of the coordinates along the PCs $\boldsymbol{\eta}^r = \{\eta_1, \ldots, \eta_{n_r}\}$ that are linear combinations of $\boldsymbol{\phi}^*$ by $\boldsymbol{\eta}^r = \boldsymbol{P}^r\boldsymbol{\phi}^*$. Similar to Eq. (9), the departure from the manifold parameterized by PCs is

$$\boldsymbol{\phi}'_\eta = \boldsymbol{\phi} - \boldsymbol{\Phi}^M(\boldsymbol{\eta}^r). \tag{25}$$

The transport equations Eq. (15) for $\boldsymbol{\phi}$ in reacting flows can be projected onto the PC basis to obtain the transport equations for $\boldsymbol{\eta}$ [21]

$$\rho\frac{\mathrm{D}\boldsymbol{\eta}}{\mathrm{D}t} = -\nabla\cdot\boldsymbol{j}_\eta + \boldsymbol{S}_\eta \tag{26}$$

where $\boldsymbol{j}_\eta(\boldsymbol{\eta}) = \boldsymbol{P}\boldsymbol{j}$ and the chemical source term based on the PCs

$$\boldsymbol{S}_\eta(\boldsymbol{\eta}) = \boldsymbol{P}\boldsymbol{S}. \tag{27}$$

The initial and boundary conditions on the PCs are also required to be transformed from the original ones using $\boldsymbol{P}$ [21]. Similar to Eq. (24), the ELDMs for $\boldsymbol{S}$ and $\boldsymbol{S}_\eta$ can be parameterized by $\boldsymbol{\eta}^r$ as $\widehat{\boldsymbol{S}}^M(\boldsymbol{\eta}^r)$ and $\widehat{\boldsymbol{S}}^M_\eta(\boldsymbol{\eta}^r)$, respectively. To reduce the number of transport equations in *a posteriori* simulations with PCA, only the equations for $\boldsymbol{\eta}^r$ in the form of Eq. (26) with the ELDM $\widehat{\boldsymbol{S}}^M_\eta(\boldsymbol{\eta}^r)$ are solved.

4.2. Selection of represented compositions

For the straightforward implementation of *a posteriori* simulations with ELDMs, it is useful to select the represented

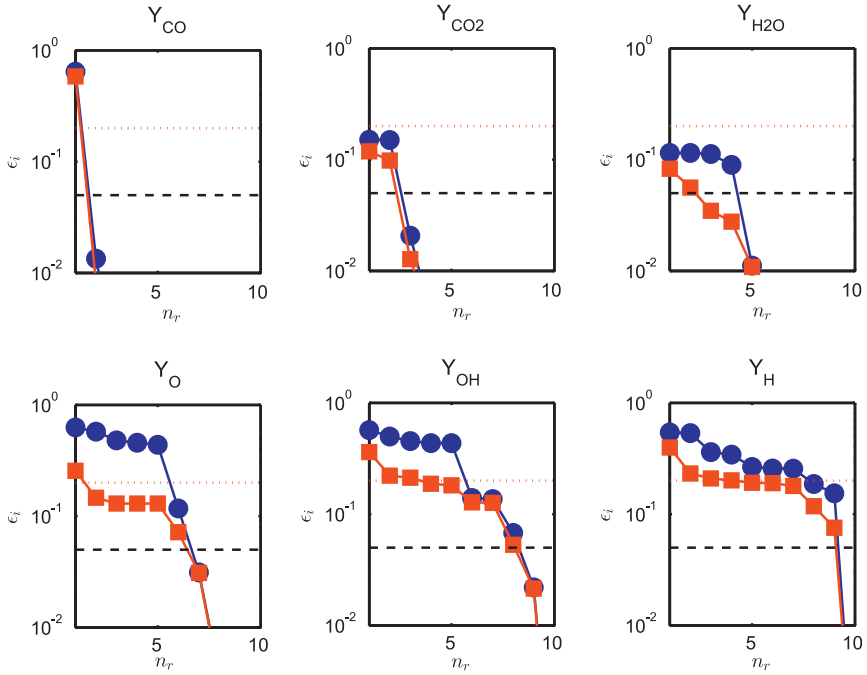


Fig. 1. Normalized r.m.s. departure ϵ_i of different species mass fractions from PCA and MARS for DNS-syngas (circles: PCA; squares: MARS). The dashed line shows the 5% level and the dotted line shows the 20% level.

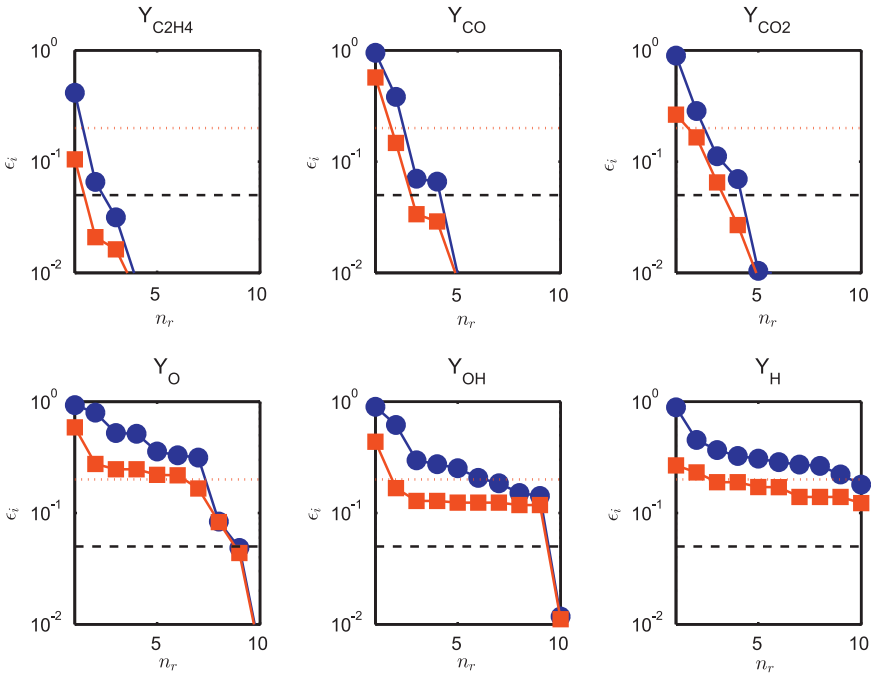


Fig. 2. Normalized r.m.s. departure ϵ_i of different species mass fractions from PCA and MARS for DNS-ethylene (circles: PCA; squares: MARS). The dashed line shows the 5% level and the dotted line shows the 20% level.

compositions $\boldsymbol{\phi}^r$, so we can solve Eq. (16) without the PCA transformation. In this subsection, we show how the PCA can be used to determine a good set of n_r represented compositions. For any choice of represented compositions, provided that each component is not constant on the manifold, there is an invertible mapping $\boldsymbol{\phi}^r = \langle \boldsymbol{\phi}^r \rangle + \boldsymbol{Q}\boldsymbol{\eta}^r$, where $\boldsymbol{Q}$ is an $n_r \times n_r$ matrix. Hence we can also express the PCA manifold Eq. (24) as

$$\boldsymbol{\Phi}^M(\boldsymbol{\phi}^r) = \langle \boldsymbol{\phi}^r \rangle + (\boldsymbol{D}_s \boldsymbol{P}^{rT} \boldsymbol{Q}^{-1})(\boldsymbol{\phi}^r - \langle \boldsymbol{\phi}^r \rangle). \tag{28}$$

Thus, a "good" choice of represented compositions is such that the principal angles between the $\boldsymbol{\phi}^r$-subspace and the $\boldsymbol{\eta}^r$-subspace are minimized (or at least not large).

Given an n_ϕ-composition vector $\boldsymbol{\phi}$, we define an $n_\phi \times n_\phi$ permutation matrix $\boldsymbol{W}$. The row index of a unity element in $\boldsymbol{W}$ corresponds to the index i of ϕ_i. The orthogonal matrix $\boldsymbol{W}$ is partitioned as $[\boldsymbol{W}^r \boldsymbol{W}^u]$, where $\boldsymbol{W}^r$ is an $n_\phi \times n_r$ matrix and $\boldsymbol{W}^u$ is an $n_\phi \times n_u$ matrix. The row indices of the unity elements in $\boldsymbol{W}^r$ and $\boldsymbol{W}^u$ represent the indices of represented compositions and unrepresented compositions in $\boldsymbol{\phi}$, respectively.

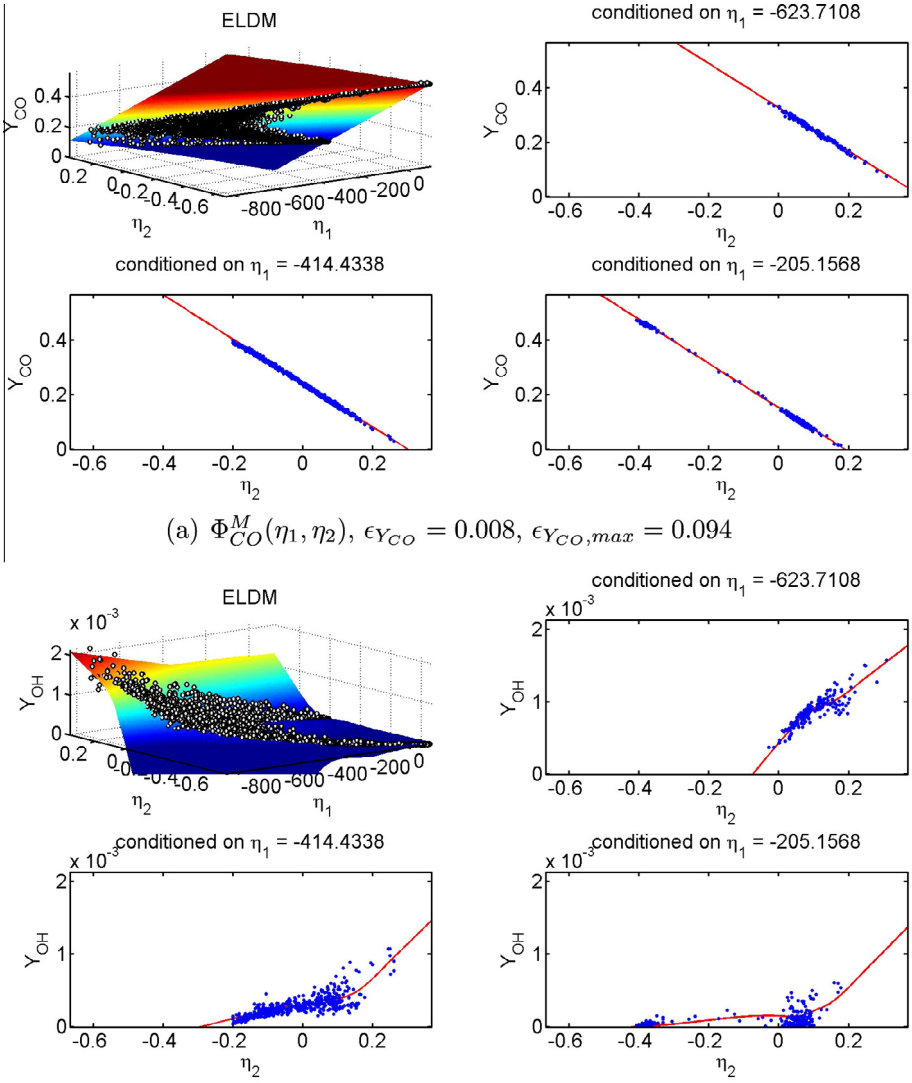


Fig. 3. The ELDMs $\Phi^M_i(\eta_1,\eta_2)$ (top-left in each subfigure) and $\Phi^M_i(\eta_1,\eta_2)$ conditioned on η_1 from MARS with corresponding sample points $Y_i(\eta_1,\eta_2)$, the r.m.s. departures ϵ_i, and the maximum departures $\epsilon_{i,max}$ for DNS-syngas. The color on the surface is coded by Y_i. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

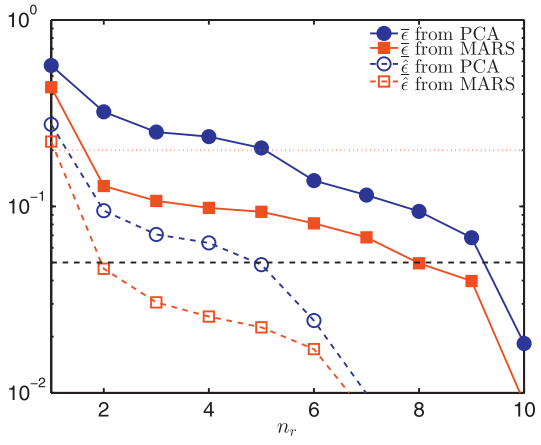


Fig. 4. Unscaled mean r.m.s. departure $\bar{\epsilon}$ and scaled mean r.m.s. departure $\bar{\bar{\epsilon}}$ over all the species mass fractions from PCA and MARS for DNS-syngas. The dashed line shows the 5% level and the dotted line shows the 20% level.

For a given n_r and the orthogonal matrix $\boldsymbol{P}^T = [\boldsymbol{P}^{rT}\ \boldsymbol{P}^{uT}]$ from PCA, the choice of $\boldsymbol{W}^r$ can be determined by minimizing the distance [43]

$$\mathrm{dist}(S_1,S_2) = \|\boldsymbol{W}^{rT}\boldsymbol{P}^{uT}\|_2 = \|\boldsymbol{P}^{rT}\boldsymbol{W}^{uT}\|_2 \tag{29}$$

between subspaces $S_1 = \mathrm{ran}\,(\boldsymbol{W}^r)$ and $S_2 = \mathrm{ran}\,(\boldsymbol{P}^{rT})$, where ran $(\cdot)$ denotes the range of a matrix and $\|\cdot\|_2$ denotes the L_2 norm.

The ordering of represented compositions can be obtained from a series of row indices for the unity elements in $\boldsymbol{W}^r$ by minimizing Eq. (29) with n_r from 1 to $n_\phi - 1$. Using the PCA results from two DNS datasets, we obtain the ordering of represented compositions for DNS-syngas:

T, **CO**, **CO_2**, **O_2**, **H_2O**, **N_2**, H_2, O, OH, HO_2, H, HCO,

and the ordering of represented compositions for DNS-ethylene:

T, **O_2**, **CO**, **C_2H_4**, **CO_2**, **C_2H_2**, N_2, O, H_2O, OH, C_2H_6, CH_3, CH_2O, CH_4, HO_2, HCCO, C_3H_6, H_2, H, CH_3CHO, CH_2CO, α-C_3H_5, H_2O_2.

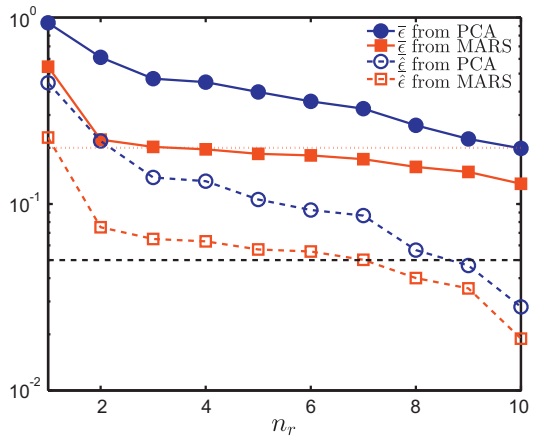


Fig. 5. Unscaled mean r.m.s. departure $\bar{\epsilon}$ and scaled mean r.m.s. departure $\tilde{\epsilon}$ over all the species mass fractions from PCA and MARS for DNS-ethylene. The dashed line shows the 5% level and the dotted line shows the 20% level.

Here, the analyzed data are sampled from the DNS database at $t/t_j = 10$, 20, 30, and 40 for DNS-syngas and from the database at the stationary state for DNS-ethylene. In the following discussions, the first five species in the ordering of represented compositions excluding temperature will be referred to as "major species" (and shown above in boldface) and the remaining will be referred to as "minor species" for each flame.

4.3. MARS

Since the PCA provides a first approximation to the ELDM, it is natural to find a curved manifold that can minimize the departures normal to the PCA plane manifold. As an extension of linear models, MARS [19,20] is a non-parametric regression technique and can automatically model non-linearities and interactions in a dataset. Hence, it is promising to represent and determine the curved manifolds, which can be used to approximate non-linear functions such as the chemical source term on a plane manifold and to approximate a curved manifold by representing $\mathbf{\Phi}^M(\boldsymbol{\phi}^r)$ or $\mathbf{\Phi}^M(\boldsymbol{\eta}^r)$. We note that the MARS manifold is more suitable to be constructed in $\boldsymbol{\eta}^r$-space than in $\boldsymbol{\phi}^r$-space. Since PCA identifies that a majority of variation exists along the directions of a few of PCs, the sample points in the $\boldsymbol{\eta}^r$-coordinates tend close to a manifold which is single-valued when projected onto $\boldsymbol{\eta}^r$-space, and hence can be parameterized by $\boldsymbol{\eta}^r$.

The MARS model is a weighted sum of basis functions. The model building process involves two phases. In the forward phase, it starts with a model consisting of the mean of the observations and adaptively adds pairs of piecewise-linear or piecewise-cubic basis functions, which gives the maximum of reduction in residual error. At the end of the forward phase, the resulting model usually overfits the data. Then, in the backward phase, the model is simplified by deleting the least effective basis functions according to generalized cross-validation.

In the present work, MARS is applied using the Matlab toolbox ARESLab [44]. The ELDMs such as $\mathbf{\Phi}^M(\boldsymbol{\eta}^r)$, $\widehat{\boldsymbol{S}}^M(\boldsymbol{\eta}^r)$, and $\widehat{\boldsymbol{S}}^M_\eta(\boldsymbol{\eta}^r)$ are fitted as n_r-dimensional non-linear functions by piecewise-cubic splines. The implementation of MARS involves several nontrivial parameters: the maximal number of basis functions N_b included in the model in the forward model building phase, and the maximum degree of interactions N_{it}. The departure ϵ_i from MARS manifolds converges with increasing the number of samples N and N_b. The scalings for the growth of the computational cost with n_r, N, and N_b are between linear and quadratic. Considering the accuracy of the MARS model and the computational cost, typically we choose $N \approx 20{,}000$, $N_b = 15$, and $N_{it} = 2$ in this study. The reduction of the normalized r.m.s departure ϵ_i Eq. (12) with increasing N, N_b, and N_{it} from those shown above is less than 1%. The computational cost of constructing a typical MARS ELDM $\Phi^M_{OH}(\boldsymbol{\eta})$ of 10 dimensions for DNS-syngas on a machine with a 1.6 GHz CPU is about 25 min.

5. Results

5.1. PCA and MARS results for species mass fractions

Using PCA and MARS, plane and curved ELDMs are obtained respectively. The normalized r.m.s. departures ϵ_i defined in Eq. (12) for different species mass fractions as functions of the dimension n_r of the ELDM $\mathbf{\Phi}^M(\boldsymbol{\eta}^r)$ in the high-*Re* CO/H_2 temporally evolving flame and the ethylene lifted flame are shown in Figs. 1 and 2, respectively. As may be seen from the figures, for both DNS databases, ϵ_i decays rapidly with n_r for most of major species such as Y_{CO} and Y_{CO_2} (upper rows). Typically less than five dimensions are needed to reduce these major-species departures from the PCA ELDM to below 5%, while for minor species such as Y_{OH} and Y_O (lower rows), about 8–10 dimensions are needed to achieve the 5% departure level shown in the dashed line. In general, the maximum departure $\epsilon_{i,max}$ defined in Eq. (13) is an order larger than ϵ_i of the same dimension, and the decay of $\epsilon_{i,max}$ agrees qualitatively with ϵ_i (not shown).

Compared to the PCA results, better accuracy is expected to be achieved by curved MARS ELDMs, because MARS can minimize the departures normal to the plane PCA manifold if the ELDM is curved. The illustrations of the curved MARS ELDMs $\Phi^M_{CO}(\eta_1, \eta_2)$

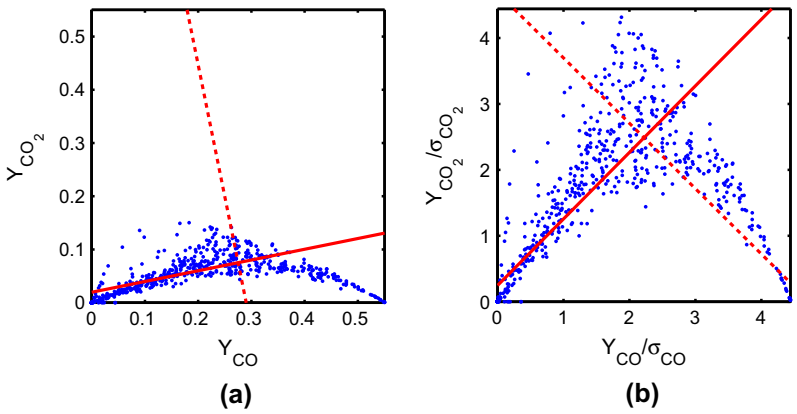


Fig. 6. Scatter plots for the samples in a reduced composition space $\{Y_{CO}, Y_{CO_2}\}$ with the directions of the PCs (solid line: first PC; dashed line: second PC).

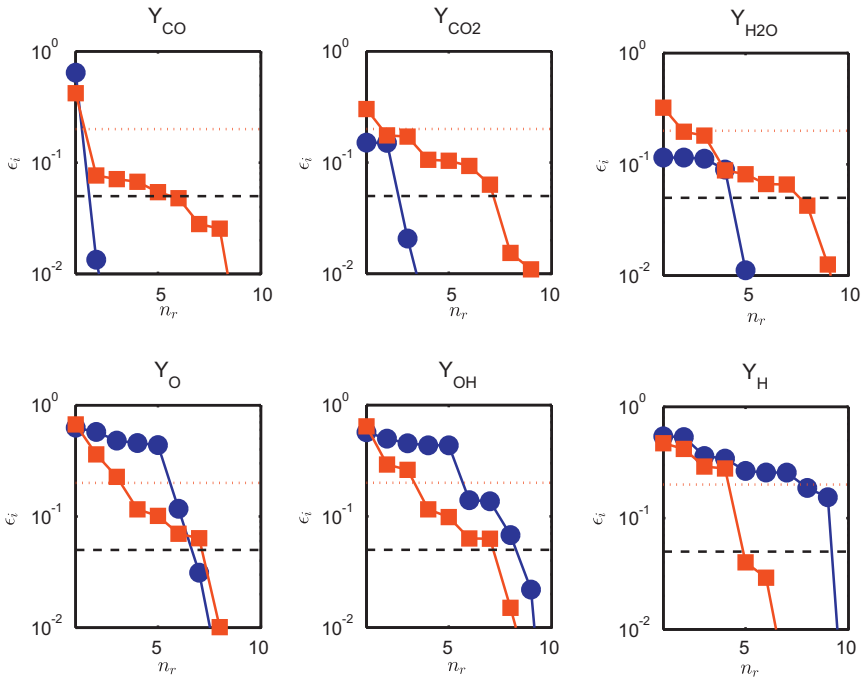


Fig. 7. Normalized r.m.s. departure ϵ_i of different species mass fractions from PCA with different scalings (circles: no scaling; squares: STD scaling). The dashed line shows the 5% level and the dotted line shows the 20% level.

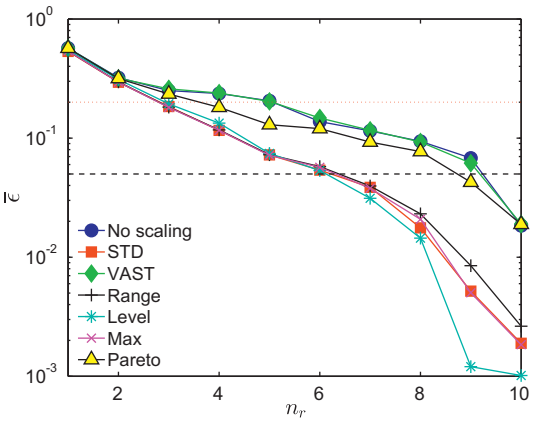


Fig. 8. Mean r.m.s. departure $\bar{\epsilon}$ over all the species mass fractions from PCA with different scalings. The dashed line shows the 5% level and the dotted line shows the 20% level.

and $\Phi^M_{OH}(\eta_1,\eta_2)$ with sample points, the r.m.s. departures ϵ_i Eq. (12), and the maximum departures $\epsilon_{i,max}$ Eq. (13) for DNS-syngas are shown in Fig. 3. Besides showing $\Phi^M_i(\eta_1,\eta_2)$ as a two-dimensional surface in a three-dimensional space (top-left in each subfigure), $\Phi^M_i(\eta_1,\eta_2)$ conditioned on different η_1 are also shown as the red curves to demonstrate the departure of the sample points from the manifold. We can see that the manifold $\Phi^M_{CO}(\eta_1,\eta_2)$ is close to a plane manifold with a very small r.m.s. departure. For comparison, the manifold $\Phi^M_{OH}(\eta_1,\eta_2)$ is much more curved and the sample points have large departures from the manifold.

In Figs. 1 and 2, for a given dimension n_r, the departures are smaller for MARS than for PCA. The significant improvements of MARS over PCA appear in the first several dimensions for minor species such as Y_{OH} and Y_O and the minority of major species such as Y_{H_2O}. This shows that the non-linear fitting in MARS can approximate the curved manifolds in low dimensions to obtain higher accuracy than PCA. In general, less than five dimensions are needed to reduce these minor-species departures from MARS ELDMs to below the 20% departure level shown in the dotted line, compared to nine dimensions for PCA. In contrast, for the ELDMs close to plane manifolds, e.g., $\Phi^M_{CO}(\eta_1,\eta_2)$ in Fig. 3a and the ELDMs in high dimensions, the difference of ϵ_i from PCA and MARS is relatively small.

5.2. Mean r.m.s. departures from PCA and MARS

To compare overall r.m.s. departures from PCA and MARS, we define the mean r.m.s. departure $\bar{\epsilon}$ which is the average of the normalized r.m.s. departures ϵ_i over n_s species. To prevent small departures in minor species from making a disproportional contribution in $\bar{\epsilon}$ and overpredicting the required dimensions of ELDMs to below a prescribed departure level, we also define a scaled normalized r.m.s. departure for composition i

$$\hat{\epsilon}_i = \frac{\sqrt{\langle \phi_i'^2 \rangle}}{\max(\sigma_i, \phi_{tol})}, \tag{30}$$

where the error tolerance $\phi_{tol} = 10^{-3}$ is chosen to eliminate ϵ_i with very small r.m.s. departure $\sqrt{\langle \phi_i'^2 \rangle}$ from minor species. Then the scaled average r.m.s. departure $\bar{\hat{\epsilon}}$ is obtained by the average $\hat{\epsilon}_i$ over n_s species. It is noted that $\hat{\epsilon}_i = \epsilon_i$ with $\phi_{tol} = 0$, and $\bar{\hat{\epsilon}}$ converges to $\bar{\epsilon}$ with decreasing ϕ_{tol}.

The unscaled and scaled mean r.m.s. departures for both DNS flames are shown in Figs. 4 and 5. In order to reduce the scaled mean r.m.s. departure $\bar{\hat{\epsilon}}$ to below 5%, for DNS-syngas, five dimensions for PCA are required compared to two dimensions for MARS. For DNS-ethylene, nine dimensions and seven dimensions are required for PCA and MARS, respectively. To achieve the 5% departure level, the number of dimensions for MARS is two to three less than that for PCA. In addition, for comparison, the unscaled average r.m.s. departure $\bar{\epsilon}$ is much larger for a given dimension. Since the chemical mechanism contains 11 species for DNS-syngas and 22 species for DNS-ethylene, the relatively small r.m.s. departures of minor species have more weight in the calculation of $\bar{\epsilon}$. In general, higher dimensions are required for the chemical

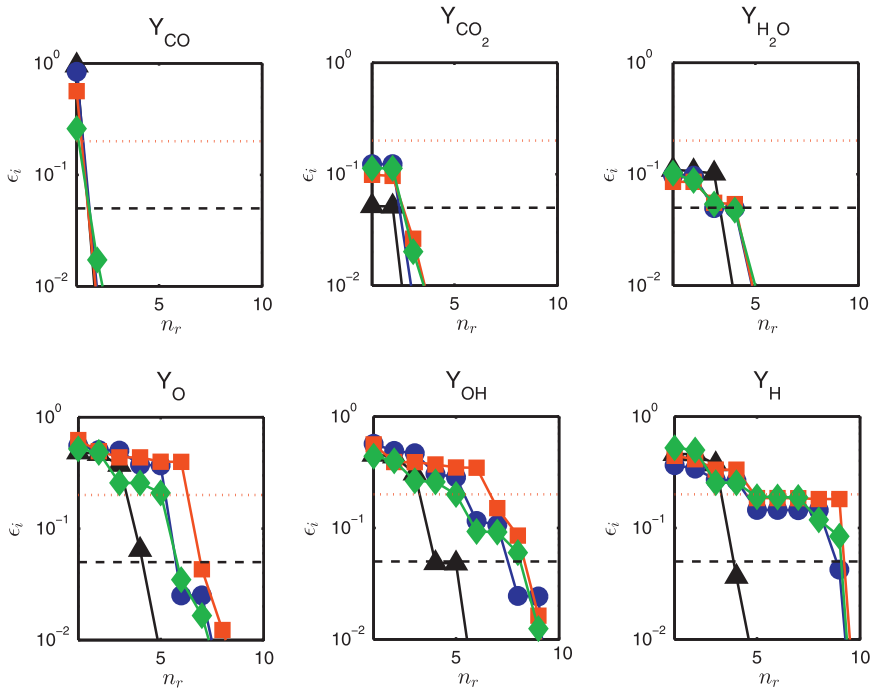


Fig. 9. Normalized r.m.s. departure ϵ_i of different species mass fractions for DNS-syngas at different times (triangles: $t = 0$; circles: $t/t_j = 10$; squares: $t/t_j = 20$; diamonds: $t/t_j = 40$). The dashed line shows the 5% level and the dotted line shows the 20% level.

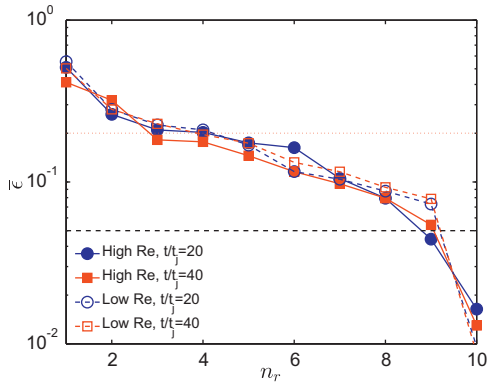


Fig. 10. Mean r.m.s. departure over all the species mass fractions from PCA for DNS of the CO/H_2 jet flame at different times and *Re*. The dashed line shows the 5% level and the dotted line shows the 20% level.

mechanisms that involve more species to obtain an accurate ELDM. In terms of $\bar{\epsilon}$, the ELDM in DNS-ethylene needs five additional dimensions over DNS-syngas to achieve the 5% level.

5.3. Effects of scalings in PCA

The PCA results are sensitive to the method used to scale the original dataset [17]. For example, Fig. 6 gives scatter plots of samples from DNS-syngas in the two-dimensional composition spaces $\{Y_{CO}, Y_{CO_2}\}$ and $\{Y_{CO}/\sigma_{CO}, Y_{CO_2}/\sigma_{CO_2}\}$ with STD scaling. The directions of the first and second PCs are shown in the solid line and the dashed line, respectively. For the PCA result without scaling in Fig. 6a, the variation in Y_{CO} is greater than that in Y_{CO_2}. Since the first PC describes the direction of maximum variation, it is more aligned with the direction of Y_{CO} than Y_{CO_2}. With STD scaling as shown in Fig. 6b, the two variables have the same degree of variation, such that each PC assigns similar weight to Y_{CO} and Y_{CO_2}.

For DNS data without scaling in the full composition space $\{\boldsymbol{Y}, T\}$, given $0 \leqslant Y_\alpha \leqslant 1$ and T ranging between 300 K and 2500 K, the first PC is entirely dominated by T, the second PC is dominated by the species mass fraction with the largest variance (e.g., Y_{CO} for DNS-syngas), the third PC is dominated by the species mass fraction with the second largest variance (if it is not perfectly correlated with species with the largest variance), and so on. This is also revealed in the ordering of represented compositions in Section 4.2. For DNS-syngas, the normalized r.m.s. departures ϵ_i for different species mass fractions from PCA without scaling and with STD scaling are shown in Fig. 7. For a given n_r, the departure for major species (upper row) is much less than that for minor species (lower row). Considering the importance of temperature and mass fraction of major species in reactions in turbulent combustion, the choice of PCs from the data without scaling or only with the normalization of the measurements with different units to mass fraction (e.g., temperature) may be appropriate. On the other hand, since each PC has similar weight for all of the composition variables with STD scaling, the decaying behaviors of the departures for different species mass fractions appear to be similar in Fig. 7.

The effect of various methods listed in Table 1 to scale the original data for PCA are examined in [24], but the results are in terms of the R^2 error and n_r is limited to two and three dimensions. To clearly demonstrate the effect of scalings in PCA, the mean r.m.s. departures $\bar{\epsilon}$ from PCA with different scalings for DNS-syngas are shown in Fig. 8. Since the PCA ELDM $\boldsymbol{\Phi}^M(\boldsymbol{\eta}^r)$ minimizes $\sum_{i=1}^{n_\phi}\langle\phi_i'^2\rangle/s_i^2$ using the first n_r PCs (see Property A5 in [17]), the overall departure of the PCA result with the scaling parameter s_i actually depends on the definition of the overall departure itself. For example, the PCA result with STD scaling gives the minimum overall departure $\frac{1}{n_\phi}\sum_{i=1}^{n_\phi}\langle\phi_i'^2\rangle/\sigma_i^2$ in all the scalings. It is noted that each variable in the centered original data is scaled by a measure of its variation, and the values of the scaling parameter s_i in STD, Range, Level, Max scalings are of the same order and can be much more than that in VAST and Pareto scalings. Therefore, with the current definition of $\bar{\epsilon}$, for a given n_r, the results without scaling and with VAST scaling show relative large values of $\bar{\epsilon}$, whereas

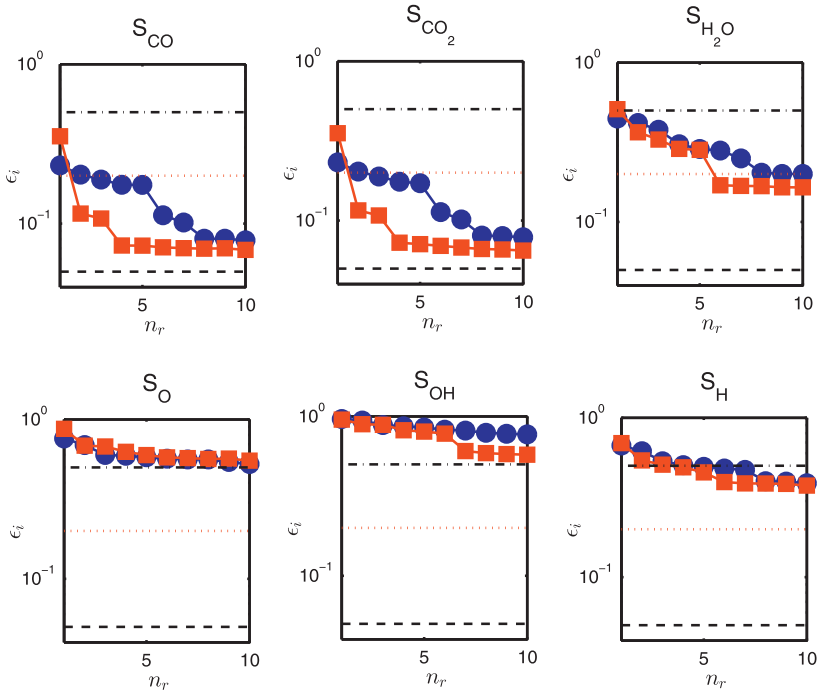


Fig. 11. Normalized r.m.s. departure ϵ_i of different chemical source terms from MARS for DNS-syngas (circles: no scaling; squares: STD scaling). The dashed line shows the 5% level, the dotted line shows the 20% level, and the dash-dotted line shows the 50% level.

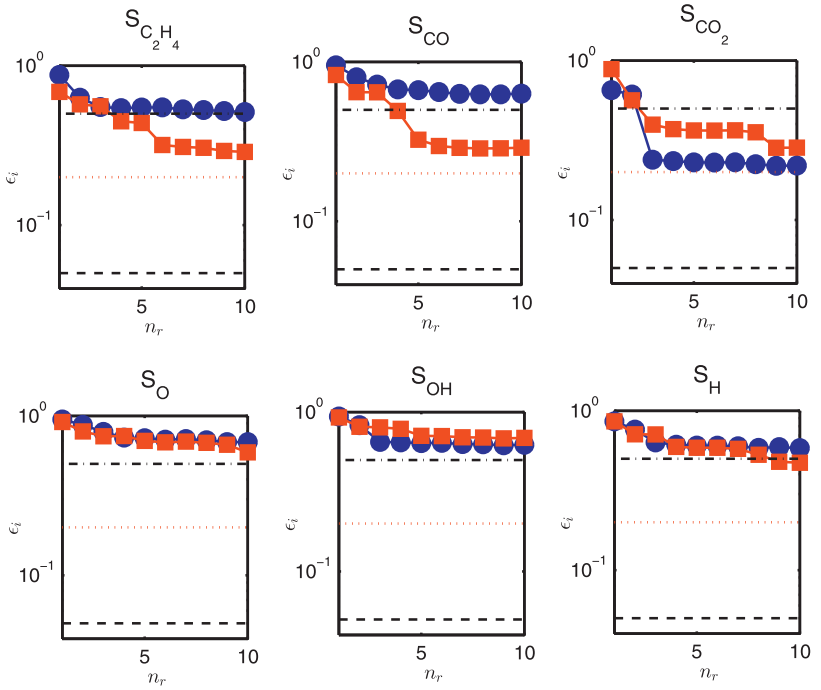
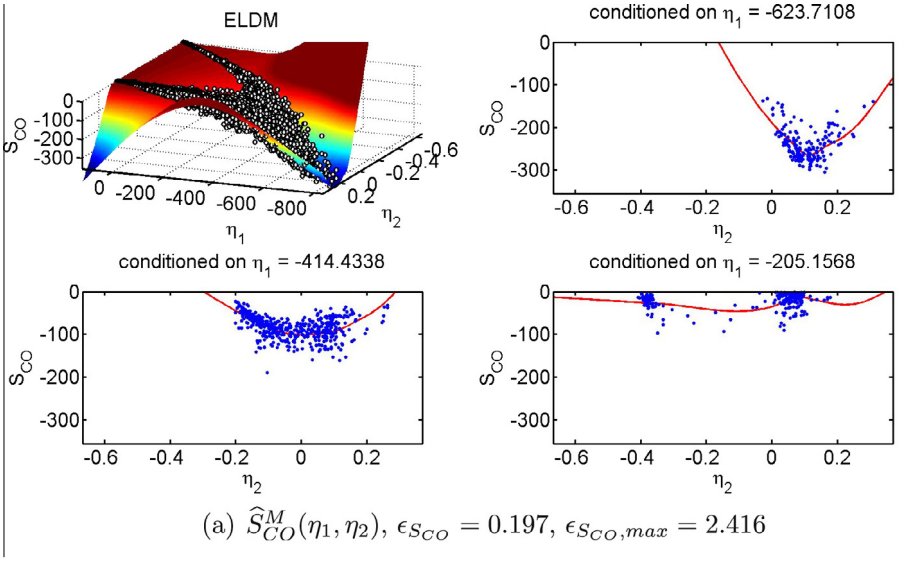


Fig. 12. Normalized r.m.s. departure ϵ_i of different chemical source terms from MARS for DNS-ethylene (circles: no scaling; squares: STD scaling). The dashed line shows the 5% level, the dotted line shows the 20% level, and the dash-dotted line shows the 50% level.

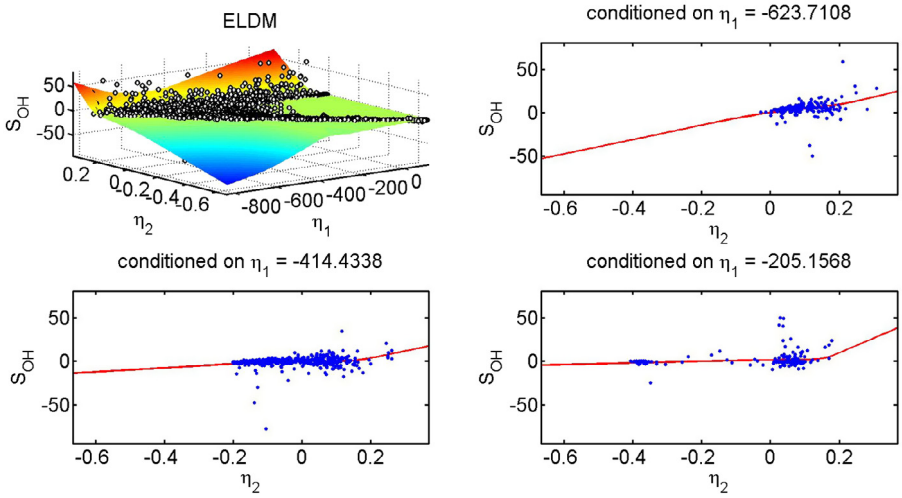
the results with STD, Range, Level, and Max scalings are similar and show smaller values of $\bar{\epsilon}$.

5.4. Effects of local extinction and Reynolds numbers

The effect of turbulent mixing and reaction on the accuracy of ELDMs is investigated by applying PCA to DNS-syngas at different times at t/t_j = 10, 20, 30, and 40. For DNS-syngas, mixing is initially sufficiently rapid relative to reaction at the low value of *Da* in this flame, so that strong turbulence–chemistry interactions can result in local extinction. The DNS results show that the maximum local extinction occurs near t/t_j = 20 and the flame is fully re-ignited near t/t_j = 40 [38]. The normalized r.m.s. departures ϵ_i for different species mass fractions are shown in Fig. 9. The number of dimensions for the prescribed 5% departure level for DNS-syngas at different times is very close for major species. Some noticeable differences,



(a) $\widehat{S}^M_{CO}(\eta_1,\eta_2)$, $\epsilon_{S_{CO}} = 0.197$, $\epsilon_{S_{CO},max} = 2.416$



(b) $\widehat{S}^M_{OH}(\eta_1,\eta_2)$, $\epsilon_{S_{OH}} = 0.935$, $\epsilon_{S_{OH},max} = 42.08$

Fig. 13. The ELDMs $\widehat{S}^M_i(\eta_1,\eta_2)$ (top-left in each subfigure) and $\widehat{S}^M_i(\eta_1,\eta_2)$ conditioned on η_1 from MARS with corresponding sample points $S_i(\eta_1,\eta_2)$, the r.m.s. departures ϵ_i, and the maximum departures $\epsilon_{i,max}$ for DNS-syngas. The color on the surface is coded by S_i. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

however, are observed for the large departures of minor-species mass fractions, e.g., Y_O and Y_{OH}, at $t/t_j = 20$. This implies that, on a very curved ELDM, the minor species vary non-linearly with represented compositions when the maximum local extinction occurs, such that their mass fractions predicted by the models based on very-low-dimensional manifolds may give relatively large departures.

The mean r.m.s. departure $\bar{\epsilon}$ from PCA for DNS of the CO/H_2 jet flame at $t/t_j = 20$ and $t/t_j = 40$ for both low and high Reynolds numbers is shown in Fig. 10. The observed similar departures show that the Reynolds number has a minor effect on the PCA results in terms of the overall error. This suggests that the ELDMs obtained from a DNS flame database at a particular *Re* may be applicable over a certain range of *Re* with sufficient accuracy.

5.5. PCA and MARS results for chemical source terms

Using MARS, the normalized r.m.s. departure ϵ_i from the ELDM $\widehat{\boldsymbol{S}}^M(\boldsymbol{\eta}^r)$ for chemical source terms $\boldsymbol{S}$ as a function of the dimension n_r in DNS-syngas and DNS-ethylene are shown in Figs. 11 and 12, where the chemical source terms are obtained from the DNS data. The results are based on $\boldsymbol{\eta}^r$ obtained from PCA without scaling and with STD scaling. Different from the species mass fractions, the PCA does not necessarily identify the optimal basis for representing the source term [24], and the departure for $\widehat{\boldsymbol{S}}^M(\boldsymbol{\eta}^r)$ has no explicit relation to the PCA scaling method. The departure from the ELDM parameterized by $\boldsymbol{\eta}^r$ with STD scaling appears to be smaller than that without scaling. Compared to the approximations of species mass fractions shown in Figs. 1 and 2, the departures for chemical source terms are much larger, which qualitatively agrees with the finding in [24]. For DNS-syngas, five dimensions are required for major species to achieve the 20% departure level, and four dimensions are needed to achieve the 50% departure level for DNS-ethylene.

The large departure for the source term can be interpreted by the geometry of $\widehat{\boldsymbol{S}}^M(\boldsymbol{\eta}^r)$ in composition space. The MARS ELDMs $\widehat{S}^M_{CO}(\eta_1,\eta_2)$ and $\widehat{S}^M_{OH}(\eta_1,\eta_2)$ without scaling are shown in Fig. 13. Compared to the ELDMs of species mass fractions in Fig. 3, the manifold $\widehat{S}^M_{CO}$ exhibits very curved geometry, and the scattered sample points with very large departures away from the ELDM for S_{OH} exist as indicated by the large maximum departure $\epsilon_{S_{OH},max}$. Figure 13 shows that the chemical source terms,

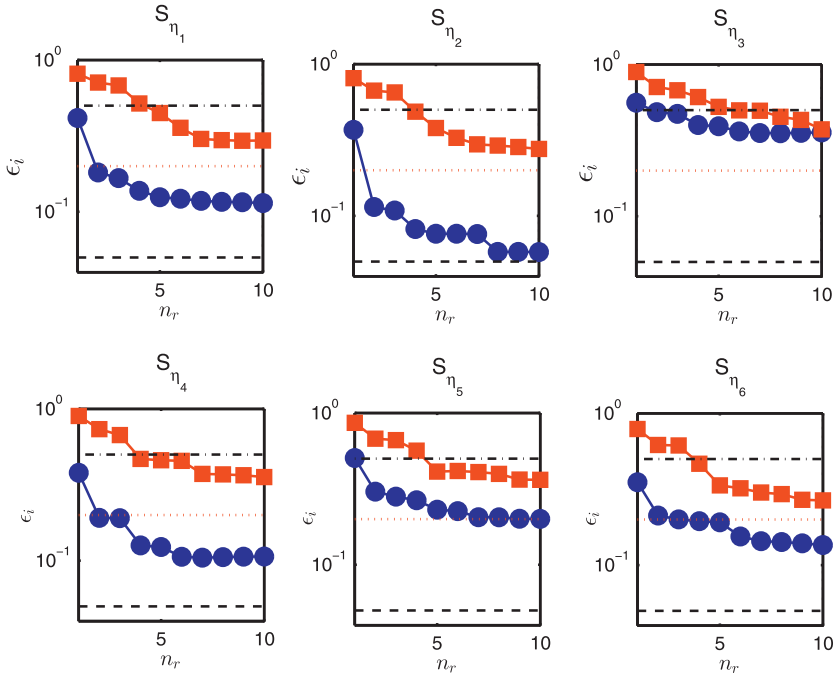


Fig. 14. Normalized r.m.s. departure ϵ_i of different chemical source terms based on PCs from MARS (circles: DNS-syngas; squares: DNS-ethylene). The PCs are obtained from the data with the STD scaling. The dashed line shows the 5% level, the dotted line shows the 20% level, and the dash-dotted line shows the 50% level.

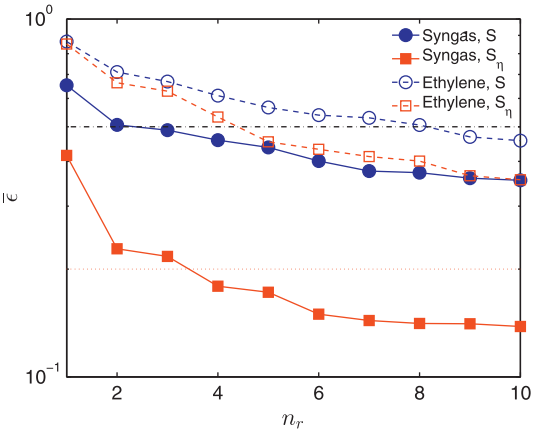


Fig. 15. Mean r.m.s. departure $\bar{\epsilon}$ over all the chemical source terms from MARS for DNS-syngas and DNS-ethylene. The dotted line shows the 20% level and the dash-dotted line shows the 50% level.

particularly for radicals, can span several orders of magnitude and are highly sensitive to small changes in the PCs, which is contrary to the assumption of the low-dimensional manifold and poses a challenge for fitting the source terms for radicals using a series of smooth spline basis functions in MARS. These difficulties appear to cause the slow convergence of the r.m.s. departures for the source terms shown in Figs. 11 and 12.

The source term based on the PCs $\boldsymbol{S}_\eta(\boldsymbol{\eta})$ defined in Eq. (27) is a linear combination of species source terms. It is required to close the transport equation Eq. (26) for $\boldsymbol{\eta}$ [24]. For DNS-syngas and DNS-ethylene, the normalized r.m.s. departure ϵ_i of the ELDMs for $S_{\eta_1}(\boldsymbol{\eta}^r)$ to $S_{\eta_6}(\boldsymbol{\eta}^r)$ is shown in Fig. 14. It is evident that ϵ_i for DNS-syngas is much less than DNS-ethylene, which shows that the accuracy of the ELDM for source terms strongly depends on the chemical mechanism and the number of species involved in the mechanism. The mean r.m.s. departures $\bar{\epsilon}$ of $\boldsymbol{S}(\boldsymbol{\eta}^r)$ and $\boldsymbol{S}_\eta(\boldsymbol{\eta}^r)$ over n_s species are shown in Fig. 15. We find that $\bar{\epsilon}$ from the ELDM $\widehat{\boldsymbol{S}}^M_\eta(\boldsymbol{\eta}^r)$ is less than that from $\widehat{\boldsymbol{S}}^M(\boldsymbol{\eta}^r)$. To reduce $\bar{\epsilon}$ to below the 50% departure level shown in the dash-dotted line, for DNS-syngas, one dimension is required for $\boldsymbol{S}_\eta(\boldsymbol{\eta}^r)$ compared to three dimensions for $\boldsymbol{S}(\boldsymbol{\eta}^r)$. For DNS-ethylene, five and nine dimensions are required to reduce the departures to lower than 50% for $\boldsymbol{S}_\eta(\boldsymbol{\eta}^r)$ and $\boldsymbol{S}(\boldsymbol{\eta}^r)$, respectively.

5.6. Discussions on a posteriori tests using ELDMs

Although the ELDMs based on the PCs show great advantages to reduce the dimensionality of the original composition space, it is still worthwhile to investigate the ELDMs based on the primitive represented composition variables $\boldsymbol{\phi}^r$. The transport equation Eq. (26) for $\boldsymbol{\phi}^r$ with $\boldsymbol{S}^r(\boldsymbol{\phi}^r)$ can be solved without the PCA transformation.

Two different approaches are introduced in Section 2.2 to approximate $\boldsymbol{S}^r(\boldsymbol{\phi}^r)$ using n_r represented compositions, where the ordering of represented compositions can be determined by the method described in Section 4.2. For Eq. (6), the non-linear function $\widehat{\boldsymbol{S}}^{r,m}(\boldsymbol{\phi}^r)$ in terms of represented compositions can be fitted by MARS directly. For the other approximation Eq. (7), first we need to reconstruct the unrepresented compositions from the ELDM $\boldsymbol{\Phi}^M(\boldsymbol{\phi}^r)$ that can be obtained *a priori* from the DNS data. The realizability of species mass fractions in $\boldsymbol{\Phi}^M(\boldsymbol{\phi}^r)$ can be satisfied by clipping the unrepresented species mass fractions to lie between zero and unity and normalizing the mass fractions so that their sum is equal to unity. Then, the chemical source term $\widehat{\boldsymbol{S}}(\boldsymbol{\Phi}^M(\boldsymbol{\phi}^r))$ is evaluated by CHEMKIN [45] with a specific mechanism and necessary thermochemistry data.

For DNS-syngas, both approximations of the source terms for major species are compared in Fig. 16, and their corresponding normalized r.m.s. departures ϵ_i are very similar. For *a posteriori* tests, the advantage of the second approximation is that, only the *a priori* knowledge of $\boldsymbol{\Phi}^M(\boldsymbol{\phi}^r)$ is needed without knowing $\widehat{\boldsymbol{S}}^M(\boldsymbol{\phi}^r)$ since it is usually not available in the experimental data. Thus, only the ELDM for $\boldsymbol{\Phi}^m(\boldsymbol{\phi}^r)$ (without $\widehat{\boldsymbol{S}}^{r,m}(\boldsymbol{\phi}^r)$) needs to be tabulated for subsequent approximations of $\boldsymbol{S}^r$ in Eq. (16) by Eq. (7) in *a posteriori*

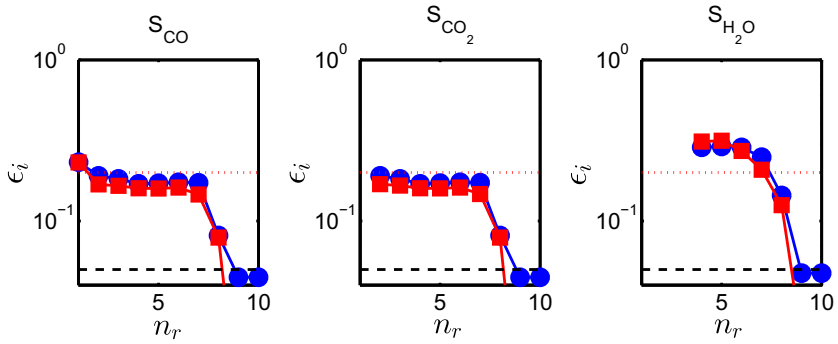


Fig. 16. Normalized r.m.s. departure ϵ_i of the approximation of chemical source terms using MARS for DNS-syngas (circles: $\widehat{S}_i^m(\boldsymbol{\phi}^r)$; squares: $\widehat{S}_i(\boldsymbol{\Phi}^M(\boldsymbol{\phi}^r))$). The dashed line shows the 5% level and the dotted line shows the 20% level.

simulations. The implementation of this method with reduced DNS or LES is feasible, and it is similar to the LES/PDF modeling with RCCE [37] and to the LES with flamelet-like models [10,36].

6. Conclusions

With the assumption that compositions occurring in turbulent combustion lie close to a low-dimensional manifold, an ELDM is constructed from an ensemble of observations in DNS or experiments by minimizing the r.m.s. departures of the observations from the manifold. The plane and curved ELDMs can be obtained using the dimension–reduction methods PCA and MARS, respectively. Both PCA and MARS are applied to the DNS databases of a non-premixed CO/H_2 temporally evolving jet flame and of an ethylene lifted jet flame. In terms of the scaled average r.m.s. departure $\bar{\bar{\epsilon}}$ for the data from DNS-syngas and DNS-ethylene, approximately five and nine dimensions are needed, respectively, to reduce these departures from the PCA ELDM to below the 5% departure level. Although representing and determining curved manifolds is more challenging than PCA, better accuracy can be achieved by curved MARS ELDMs with lower dimensions. In order to reduce $\bar{\bar{\epsilon}}$ to below the 5% departure level using MARS, on average two and five dimensions are required for DNS-syngas and DNS-ethylene, respectively. These findings are qualitatively similar to those in [24], but the current results are based on the normalized r.m.s. departure ϵ_i, which is more sensitive and suitable than the R^2 error used in [21,24] to examine the departures of the ELDM with more than three dimensions.

The PCA results can be significantly affected by the scaling method. The ELDMs with different PCA scalings are compared and fully discussed. We find that the results without scaling are suitable to predict mass fractions of major species using a small number of dimensions, e.g., less than five for both flames, and they are similar to the results with the preferred VAST scaling used in [21,24]. The results with STD scaling give the relatively small mean r.m.s. departure $\bar{\epsilon}$, but similar dimensions are needed to achieve a prescribed departure level for both major and minor species. We clarify that the mean departure based on PCA models depends on the relation between its own definition and the PCA scaling parameter. In the present study, only linear scalings are discussed; nevertheless non-linear scaling methods could be considered in future work.

The effects of local extinction and Reynolds numbers are investigated for DNS-syngas. The normalized r.m.s. departure ϵ_i for the mass fraction of minor species is found to be maximum when strong local extinction occurs. The mean r.m.s. departures $\bar{\epsilon}$ from ELDMs are insensitive to the Reynolds number.

Compared to the ELDMs of species mass fractions, it is much more challenging to obtain the ELDMs $\widehat{\boldsymbol{S}}^M(\boldsymbol{\eta}^r)$ for the source terms with high accuracy. In terms of the normalized r.m.s. departure ϵ_i, about five dimensions are required for major species to achieve the 20% departure level for DNS-syngas, and more than 10 dimensions are needed for DNS-ethylene. The mean r.m.s. departure $\bar{\epsilon}$ from the ELDM based on the PCs $\widehat{\boldsymbol{S}}_\eta^M(\boldsymbol{\eta}^r)$ is less than that from $\widehat{\boldsymbol{S}}^M(\boldsymbol{\eta}^r)$. To reduce $\bar{\epsilon}$ from $\widehat{\boldsymbol{S}}_\eta^M(\boldsymbol{\eta}^r)$ to below 20%, one dimension is required for DNS-syngas and five dimensions are required for DNS-ethylene.

For *a posteriori* tests, two different approaches to construct the ELDMs for chemical source terms are discussed and their accuracies to approximate the source term $\boldsymbol{S}^r(\boldsymbol{\phi}^r)$ for represented compositions are comparable. Only the ELDM for $\boldsymbol{\Phi}^m(\boldsymbol{\phi}^r)$ needs to be tabulated for further use in *a posteriori* simulations without the transformation of PCA on the transport equations, where the ordering of represented compositions can be determined from the PCA result by minimizing the distance between subspaces of $\boldsymbol{\eta}^r$ and $\boldsymbol{\phi}^r$. Compared to solving the transport equations Eq. (26) for the PC basis $\boldsymbol{\eta}$ considered in [21,24], it is more feasible to add the tabulation of the ELDM for $\boldsymbol{\Phi}^m(\boldsymbol{\phi}^r)$ to most of the current implementations of combustion LES or LES/PDF based on composition variables.

The current investigations on ELDMs are based on *a priori* tests. The applicability and accuracy of the ELDMs should be examined in *a posteriori* tests. The available DNS databases [38,39] can be used to construct the ELDM $\boldsymbol{\Phi}^m(\boldsymbol{\phi}^r)$ which can be incorporated in LES/PDF [46] for these flames, and the implementation of LES/PDF/ELDM would be similar to the current implementation of LES/PDF/RCCE [37]. Although we have shown that the ELDM is not very sensitive to the local flow condition and Reynolds numbers, the chemical mechanism has an impact on the accuracy of ELDMs and it might be challenging to obtain an accurate ELDM within several dimensions for a complex mechanism with many species. Therefore, with limited high-fidelity combustion DNS for simple fuels [2], the application of ELDM to more practical *a posteriori* simulations in engineering problems remains an open question.

Acknowledgments

We are grateful to C.F. Van Loan for the algorithm of minimizing the distance between subspaces, E.R. Hawkes and H. Kolla for the postprocessing of DNS data, and J.C. Sutherland for valuable discussions. This research is supported by the Combustion Energy Frontier Research Center, an Energy Frontier Research Center funded by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences under Award No. DE-SC0001198. Computer allocations were awarded by DOEs Innovative and Novel Computational Impact on Theory and Experiments (INCITE) program. This research used resources of the National Center for Computational Sciences at Oak Ridge National Laboratory (NCCS/ORNL) which is supported by the Office of Science of the US DOE under Contract No. DE-AC05-00OR22725.

References

[1] T.F. Lu, C.K. Law, Prog. Energy Combust. Sci. 35 (2009) 192–215.
[2] J.H. Chen, Proc. Combust. Inst. 33 (2011) 99–123.
[3] S.B. Pope, Proc. Combust. Inst. 34 (2013) 1–31.
[4] J.C. Keck, D. Gillespie, Combust. Flame 17 (1971) 237–241.
[5] D. Hamiroune, P. Bishnu, M. Metghalchi, J.C. Keck, Combust. Theory Model. 2 (1998) 81–94.
[6] U. Maas, S.B. Pope, Combust. Flame 88 (1992) 239–264.
[7] M. Bodenstein, S.C. Lind, Z. Phys. Chem. 57 (1906) 168.
[8] S.H. Lam, D.A. Goussis, Int. J. Chem. Kinet. 4 (1994) 461–486.
[9] S.B. Pope, Flow Turbul. Combust. 72 (2004) 219–243.
[10] N. Peters, Turbulent Combustion, Cambridge University Press, 2000.
[11] C.D. Pierce, P. Moin, J. Fluid Mech. 504 (2004) 73–97.
[12] J.A. van Oijen, L.P.H. de Goey, Combust. Sci. Technol. 161 (2000) 113–137.
[13] V. Bykov, U. Maas, Combust. Theory Model. 11 (2007) 839–862.
[14] R.W. Bilger, Phys. Fluids A 5 (1993) 436–444.
[15] A.Y. Klimenko, R.W. Bilger, Prog. Energy Combust. Sci. 25 (1999) 595–687.
[16] A.Y. Klimenko, S.B. Pope, Phys. Fluids 15 (2003) 1907–1925.
[17] I.T. Jolliffe, Principal Component Analysis, Springer, 2002.
[18] J. Shlens, A Tutorial on Principal Component Analysis, 2009 <www.snl.salk.edu/~shlens/pca.pdf>.
[19] J.H. Friedman, Ann. Stat. 19 (1991) 1–67.
[20] J.H. Friedman, C.B. Roosen, Stat. Meth. Med. Res. 4 (1995) 197–217.
[21] J.C. Sutherland, A. Parente, Proc. Combust. Inst. 32 (2009) 1563–1570.
[22] A. Parente, J.C. Sutherland, L. Tognotti, P.J. Smith, Proc. Combust. Inst. 32 (2009) 1579–1586.
[23] A. Parente, J.C. Sutherland, B.B. Dally, L. Tognotti, P.J. Smith, Proc. Combust. Inst. 33 (2011) 3333–3341.
[24] A. Biglari, J.C. Sutherland, Combust. Flame 159 (2012) 1960–1970.
[25] H. Mirgolbabaei, T. Echekki, Combust. Flame 160 (2013) 898–908.
[26] A. Najafi-Yazdi, B. Cuenot, L. Mongeau, Combust. Flame 159 (2012) 1197–1204.
[27] T. Turányi, Proc. Combust. Inst. 25 (1994) 949–955.
[28] A. Buki, T. Perger, T. Turányi, U. Maas, J. Math. Chem. 31 (2002) 345–362.
[29] U. Maas, D. Thévenin, Proc. Combust. Inst. 27 (1998) 1183–1189.
[30] C.E. Frouzakis, Y.G. Kevrekidis, J. Lee, K. Boulouchos, A.A. Alonso, Proc. Combust. Inst. 28 (2000) 75–81.
[31] S.J. Danby, T. Echekki, Combust. Flame 144 (2006) 126–138.
[32] B. Sportisse, R. Djouad, J. Geophys. Res. – Atmos. 112 (2007) D6.
[33] J.B. Tenenbaum, V. de Silva, J.C. Langford, Science 22 (2000) 2319–2323.
[34] G. Bansal, A. Mascarenhas, J.H. Chen, in: 7th US National Technical Meeting of the Combustion Institute, Atlanta, GA, USA, March, 2011, p. 1C08.
[35] A. Coussement, O. Gicquel, A. Parente, Proc. Combust. Inst. 34 (2013) 1117–1123.
[36] H. Pitsch, Annu. Rev. Fluid Mech. 38 (2006) 453–482.
[37] V. Hiremath, Z. Ren, S.B. Pope, Combust. Flame 158 (2011) 2113–2127.
[38] E.R. Hawkes, R. Sankaran, J.C. Sutherland, J.H. Chen, Proc. Combust. Inst. 31 (2007) 1633–1640.
[39] C.S. Yoo, E.S. Richardson, R. Sankaran, J.H. Chen, Proc. Combust. Inst. 33 (2011) 1619–1627.
[40] Z. Luo, C.S. Yoo, E.S. Richardson, J.H. Chen, C.K. Law, T.F. Lu, Combust. Flame 159 (2012) 265–274.
[41] J. Li, Ph.D. thesis, Mechanical and Aerospace Engineering Department, Princeton University, 2004.
[42] H. Wang, A. Laskin, A comprehensive reaction model of ethylene and acetylene combustion, 2007 <http://ignis.usc.edu/Mechanisms/C2-C4/c2.html>.
[43] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., The Johns Hopkins University Press, 1996.
[44] G. Jekabsons, 2011 <http://www.cs.rtu.lv/jekabsons>.
[45] R.J. Kee, F.M. Rupley, E. Meeks, J.A. Miller, CHEMKIN-III: A Fortran Chemical Kinetic Package for the Analysis of Gas-Phase Chemical and Plasma Kinetics, Tech. Rep. SAND96-8216, Sandia National Laboratories, 1996.
[46] Y. Yang, H. Wang, S.B. Pope, J.H. Chen, Proc. Combust. Inst. 34 (2013) 1241–1249.