Combustion and Flame 162 (2015) 2592–2601

Contents lists available at ScienceDirect

Combustion and Flame

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

Advanced regression methods for combustion modelling using principal components

Benjamin J. Isaac [a,b,*], Jeremy N. Thornock [a], James Sutherland [a], Philip J. Smith [a], Alessandro Parente [b]

[a] Department of Chemical Engineering, University of Utah, Salt Lake City, UT 84112, USA
[b] Service d'Aéro-Thermo-Mécanique, Université Libre de Bruxelles, Bruxelles, Belgium

ARTICLE INFO

Article history:
Received 4 August 2014
Received in revised form 11 March 2015
Accepted 12 March 2015
Available online 8 April 2015

Keywords:
Combustion
Nonlinear regression
Low-dimensional manifolds
Principal Component Analysis
Reacting flows
Reduced-order modelling

ABSTRACT

Modelling the physics of combustion remains a challenge due to a large range of temporal and physical scales which are important in these systems. Detailed chemical kinetic mechanisms are used to describe the chemistry involved in the combustion process yielding highly coupled partial differential equations for each of the chemical species used in the mechanism. Recently, Principal Components Analysis (PCA) has shown promise in its ability to identify a low-dimensional manifold describing the reacting system. Several PCA-based models have been developed which may be well-suited for combustion problems; however, several challenging aspects of the model must be addressed. In this paper, the parameterization of state-space variables and PC-transport equation source terms are investigated. The ability to achieve highly accurate mapping through various nonlinear regression methods is shown. In addition, the effect of PCA-scaling on the ability to regress the surface is investigated. Finally, the present work demonstrates the capabilities of the model by solving a reduced system represented by several PC-transport equations for a perfectly stirred reactor (PSR) configuration.

1. Introduction

The ability to accurately model a turbulent combustion system remains challenging due to the complex nature of combustion systems. A simple fuel such as CH_4 requires 53 species and 325 chemical reactions [1] to be accurately described. More complex fuels require increasingly complex chemical mechanisms. Each resolved chemical species requires a conservation equation which is a coupled, nonlinear partial differential equation. Such systems are only possible to solve under very limited situations at this time due to computational costs. Current computational expenses result in a need for reduced models which can adequately describe the chemical reactions. Many methods attempt to reduce the complexity of the mechanism by splitting the system into slow and fast variables, using equilibrium assumptions for fast chemical processes, and occupying the computational resources on the more pertinent evolution of species within the reacting system [2,3]. Indeed, in these complex combustion reaction mechanisms many of the species evolve at time-scales much larger than the time-scales of interest, allowing for decoupling of fast and slow processes while maintaining accuracy. Low-dimensional manifolds exist in these systems which can describe the governing characteristics of the flames. Several models take advantage of this, including the steady laminar flamelet model (SLFM) [4–6], flamelet-generated manifolds (FGM) [7,8], or the flame prolongation of ILDM (FPI) [9–11] to name a few. As a fundamental example, the steady laminar flamelet model uses the mixture fraction and mixture fraction variance to describe the flame as an ensemble of steady laminar diffusion flames undergoing various strain rates. In some cases, this provides a good representation of the entire system with a reduced number of variables.

Recently, principal component analysis (PCA) has been investigated for its use in combustion modelling. Several advantages of PCA include: its ability to identify orthogonal variables which are the best linear representation of the system; its ability to reduce in dimensionality requiring fewer coordinates; and the ability to do the analysis on canonical systems, such as the counter diffusion flames or empirical data-sets containing highly complex turbulent chemistry interaction. Parente et al. [12,13] used PCA to identify the low-dimensional manifold in one-dimensional turbulence and experimental data. Biglari and Sutherland [14] and Yang and Pope [15,16] enhanced the capability of the PCA concept by combining the analysis with nonlinear regression, allowing a nonlinear mapping between state-space variables and the linear PCA basis.

* Corresponding author at: 155 South 1452 East, Room 350, Salt Lake City, UT 84112, USA.
E-mail addresses: Benjamin.J.Isaac@utah.edu (B.J. Isaac), J.Thornock@utah.edu (J.N. Thornock), James.Sutherland@utah.edu (J. Sutherland), Philip.Smith@utah.edu (P.J. Smith), Alessandro.Parente@ulb.ac.be (A. Parente).

http://dx.doi.org/10.1016/j.combustflame.2015.03.008
0010-2180/

The work of Biglari and Sutherland showed that the PC parameterization is superior to the standard flamelet parameterization, for the ODT data-set investigated in the study. Mirgolbabaei and Echekki [17] extended the nonlinear mapping concept using artificial neural networks and investigated the potential of kernel PCA [18,19], showing the high compression potential derived by transforming the initial problem into a non-linear featured space where linear PCA is carried out. In addition, several combustion models have been proposed based on the concepts from PCA. Sutherland and Parente [20] derived transport equations for the principal components (PCs), and discussed the feasibility of a model where the PCs are used directly to construct state-space variables. Biglari and Sutherland [14] extended the concept of transporting the PCs by suggesting the nonlinear regression in order to increase the accuracy and reducibility of the model. Coussement et al. [21], Isaac et al. [22] and other groups [23] proposed transporting a reduced set of state-space variables and used the PC basis for reconstructing the variables which are not represented. Najafi-Yazdi et al. [24] used PCA to identify optimal progress variables to use the flamelet-generated manifold framework.

The present work seeks to advance the understanding and application of the PC-transport approach of Sutherland and Parente [20,14] by first analyzing the effect of several scaling methods on the PC basis, and the resultant ability to regress the nonlinear state-space variables to the PC basis. Various regression methods used in previous studies [14,17], as well as several alternative methods are analyzed in their ability to approximate the reacting state-space from the PCs. In order to demonstrate the accuracy of the method within a numerical solver, an unsteady perfectly stirred reactor (PSR) calculation is shown using the PC-transport approach. The PSR provides a validation of the approach by comparing the reduced model to the detailed simulation results. To the authors knowledge all published analysis on the PC-transport concept using nonlinear regression has been carried out on various data-sets using *a priori* analysis [14,17,19,18]. Only recently, *a posteriori* work has begun in this area. Specifically, the work of Mirgolbabaei [25], who provides an *a posteriori* demonstration of the nonlinear PC-transport approach using one-dimensional turbulence (ODT) simulations.

2. Theory

A principal component analysis is done by taking a data-set consisting of n observations and Q independent variables and organizing it as an $n \times Q$ matrix ($\mathbf{X}$). The data $\mathbf{X}$ is centered to zero by its corresponding means $\bar{\mathbf{X}}$, and scaled by the diagonal matrix, $\mathbf{D}$, containing a scaling value for each of the k variables:

$$\mathbf{X^s} = (\mathbf{X} - \bar{\mathbf{X}})\mathbf{D}^{-1} \tag{1}$$

For sake of simplicity, $\mathbf{X^s}$ will be simply indicated as $\mathbf{X}$ in the following. In a PC analysis, the principal components ($\mathbf{Z}$) are identified by performing an eigenvalue decomposition of the covariance matrix of $\mathbf{X}$:

$$\frac{1}{Q-1}\mathbf{X}^T\mathbf{X} = \mathbf{A}^{-1}\mathbf{L}\mathbf{A} \tag{2}$$

The eigenvector matrix $\mathbf{A}$ (referred to here as a 'basis matrix') is then used to project the original state-space into PC space:

$$\mathbf{Z} = \mathbf{X}\mathbf{A} \tag{3}$$

Now given a subset of the basis matrix $\mathbf{A}$, denoted as $\mathbf{A}_q$ and applying the previous equation, an approximation of the original centered and scaled state-space can be made using the following:

$$\mathbf{X} \approx \mathbf{Z_q}\mathbf{A}_q^T. \tag{4}$$

In the PC analysis, the largest eigenvalues correspond to the first columns of $\mathbf{A}$. This means the largest amount of variance in the original variables is described by the first PCs. Accordingly, when one truncates the basis matrix ($\mathbf{A}_q$), the resultant approximation from Eq. (4) may be very accurate, while representing the system with fewer variables.

In the work of Sutherland and Parente [20], a combustion model is proposed where conservation equations for the PCs are derived from the general species transport equation [26]:

$$\frac{\partial}{\partial t}(\rho Y_k) + \frac{\partial}{\partial x_i}(\rho u_i Y_k) = \frac{\partial}{\partial x_i}\left(\rho D_k \frac{\partial Y_k}{\partial x_i}\right) + R_k \tag{5}$$

where R_k is the net production rate of species k. One can easily derive the transport equations for the PCs ($\mathbf{Z}_q$) given the basis matrix $\mathbf{A}_q$, the scaling vector d_k, being the diagonal components of $\mathbf{D}$, and the centering vector $\bar{Y}_k$:

$$\frac{\partial}{\partial t}(\rho Z_q) + \frac{\partial}{\partial x_i}(\rho u_i Z_q) = \frac{\partial}{\partial x_i}\left(\rho D_{Z_q}\frac{\partial}{\partial x_i}(Z_q)\right) + s_{Z_q} \tag{6}$$

$$s_{Z_q} = \sum_{k=1}^{Q}\frac{R_k}{d_k}A_{kq} \tag{7}$$

where s_{Z_q} is simply the net production rate of the principal component. The term $D_{Z_q}\frac{\partial}{\partial x_i}(Z_q)$ is the diffusion flux for the principal component. For a more detailed discussion on the treatment of the PCs diffusive flux, where molecular diffusion is important refer to [27]. According to the proposed formulation, one can theoretically use PCA with its inherent advantages. These advantages include: the ability to represent the system with a reduced number of variables; the option to include a predetermined amount of reconstruction error (dependent on q, the number of retained PCs), and possibly a reduction in stiffness if the selected PCs are highly weighted with reacting species that change more slowly, such as the major species.

In order to use PCA to its fullest potential, several aspects of PCA must be studied. One of these aspects, is how the data is scaled (Eq. (1)). The various effects of scaling have been studied previously in [14,28,22]. The same approach has been followed in the present paper to find the best scaling option for the present application of PCA, using a data-set which exhibits physics of interest. A one-dimensional turbulence (ODT) data-set of a non-premixed synthesis/air jet has been considered here [29,30]. The simulation includes 11 chemical species [31] (H_2, O_2, O, OH, H_2O, H, HO_2, CO, CO_2, HCO, N_2), and 21 chemical reactions and it is initialized with a temperature of 500 K, with air as the oxidizer (0.7241 N_2 and 0.2759 O_2 by mass) and a fuel stream containing 0.0078 H_2, 0.5511 CO, and 0.4411 N_2 by mass. The ODT realizations are saved on a uniform grid of 672 grid points evenly spaced over a 0.01 m domain. The velocity field is initialized with a Reynolds numbers of 2500. The ODT data-set is particularly interesting because of the turbulence/chemistry interaction observed in the data, including physical effects such as extinction and re-ignition. Similarly to previous investigations [14,28,22], the *a priori* analysis showed that pareto scaling has a distinct advantage for major species and source terms reconstruction.

The *a priori* analyses showed, however, that at least 8 PCs were required to accurately reconstruct the ODT data-set and the corresponding source terms, due to the linear nature of the PC-based model. Considering the original 11 degrees of freedom of the system (with differential diffusion, enthalpy and elemental mass fractions are not constant), $q = 8$ implies only a minor problem reduction. An alternative to the direct reconstruction of $\mathbf{X}$ is to use nonlinear regression functions, which can be used to map the nonlinear reaction rates or nonlinear species concentrations to the lower dimensional representation given by the PCs. Biglari and Sutherland [14] suggest applying a nonlinear mapping to the linear