Proceedings of the Combustion Institute, Volume 28, 2000/pp. 75–81

PROPER ORTHOGONAL DECOMPOSITION OF DIRECT NUMERICAL SIMULATION DATA: DATA REDUCTION AND OBSERVER CONSTRUCTION

C. E. FROUZAKIS,[1] Y. G. KEVREKIDIS,[2,3] J. LEE,[4] K. BOULOUCHOS,[4] AND A. A. ALONSO[5]

[1]*Combustion Research Laboratory*
Paul Scherrer Institute
CH-5232 Villigen, Switzerland
[2]*Department of Chemical Engineering*
University of Patras, Patras, GR-26500, Greece
[3]*Princeton University*
Princeton, NJ, USA
[4]*Internal Combustion Engines and Combustion Technology Laboratory*
Swiss Federal Institute of Technology (ETHZ)
CH-8092 Zurich, Switzerland
[5]*Department of Chemical Engineering*
University of Vigo
Apto 874, 36200 Vigo, Spain

In this paper, direct numerical simulation (DNS) data of an opposed-jet hydrogen/air diffusion flame are, in a postprocessing step, analyzed using the proper orthogonal decomposition (POD) technique. The aim of this work is twofold. The first goal is to compute a small number of space-dependent empirical eigenfunctions, so that a low-dimensional representation of the data generated by the large model of the discretized partial differential equations can be obtained using a weighted sum of these few eigenfunctions (POD modes). It is found that only six modes are needed for an accurate representation of the data in an extended range of inflow velocities. This large data reduction takes into account not only chemical kinetics but also transport phenomena in a full two-dimensional context and constitutes the first step toward the construction of low-dimensional dynamic models for the opposed-jet system. It is also found that the PODs have very good interpolatory properties. The second goal is to use part of the available data (i.e., partial measurements), together with the computed modes, to estimate, or, in the terminology of process control, to observe, the "unmeasured" quantities. It is found that only a small number of measurements are needed to obtain accurate estimates of the rest of the data.

Introduction

An accurate, well-resolved discretization of the conservation equations for laminar chemically reactive flows reduces the partial differential equations (PDEs, which are infinite-dimensional dynamical systems) to systems of ordinary differential equations (ODEs). However, the size of the latter model is usually very large: in two dimensions and with detailed chemical kinetics, such models can easily reach (and even exceed) $O(10^5)$ degrees of freedom. While the descriptive and predictive capabilities of such models for a large range of operating parameters can be exceptional, the computational cost can also be prohibitive.

The method of proper orthogonal decomposition (POD) (also known as the method of empirical orthogonal eigenfunctions [EOFs] in the meteorology and oceanography literature [1], as Karhunen-Loève [KL] expansion in pattern recognition [2], and as principal component analysis [PCA] in the statistical literature [3]) has been used in the construction of low-dimensional models from extensive numerical or experimental data from non-reactive turbulent flows (see Ref. [4] for a detailed presentation of the method and its application).

In this paper, we employ the POD technique to obtain low-dimensional representations of a H_2/air opposed-jet diffusion flame over a range of jet-exit velocities. The statistical analysis of the simulation results provides an optimal basis of orthogonal eigenfunctions (modes), in the sense that they minimize the mean square error between the original data and its low-dimensional representation. It turns out that the reactive-flow system with more than 26,000 degrees of freedom can be approximated by the weighted sum of only six modes. This in turn means that one only needs the six coefficients (weights) of the modes to get a good description of the system. We find that this low-dimensional representation approximates well not only the direct numerical simulation (DNS) data used to obtain the

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
04 AUG 2000	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Proceedings of the Combustion Institute (28th). Held in Edinburgh, Scotland, 30 July - 4 Aug 2000, 2000/pp. 75-81, Proper Orthogonal Decompositioin of Direct Numerical Simulation Data: Data Reduction and Observer Construction	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **EORD, PSC 802, BOX 14, FPO 09499-0200**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**7**	

modes but also data for intermediate jet-exit velocities within the range of velocities covered by the data.

In principle, one can substitute the original variables by their weighted sum expansion in the PDEs and perform a Galerkin procedure to obtain a system of, here, only six ODEs for the time derivatives of the coefficients in the expansion (a POD-Galerkin procedure). Time integration of this system could then provide a reasonably accurate approximation to the full (more than 26,000 ODEs) system. For simple forms of the nonlinear terms in the original PDEs (e.g., quadratic), the low-dimensional model can be constructed in explicit form [5,6]. However, for higher-order and/or non-polynomial nonlinearities (as is the case for reactive systems due to the exponential temperature dependence of the reaction rate constants), the evaluation of the right-hand side and the Jacobian for POD-Galerkin models is computationally expensive.

Here we are interested in another problem: the construction of POD-based observers, so that the computed POD modes and a set of partial measurements can be used to estimate the complete field of unmeasured quantities. Typically in experiments, measurements are (or can only be) made for a few variables either at selected positions (e.g., velocity with laser Doppler anemometry, temperature with coherent anti-Raman Stokes or a thermocouple) or, in the best case, over two-dimensional domains (e.g., velocity with particle image velocimetry [PIV], OH concentration with planar laser-induced fluorescence [PLIF], temperature with Rayleigh scattering). In particular, opposed-jet diffusion flames are commonly investigated experimentally by measuring the velocity and temperature at points along the axis of symmetry. In more recent experiments investigating transient processes such as flame-vortex interactions in the same setup, two-dimensional measurement techniques such as PLIF were employed. Recently, Podvin and Lumley [7] used POD modes for the velocity in the wall region of a turbulent boundary layer to reconstruct the flow in the wall region only from wall-shear measurements. We investigate this idea here to find how well partial measurements of a single variable (either in the full domain or along the axis of symmetry) can be used together with the available POD modes to estimate the unmeasured variables; the results are remarkably good.

Methodology and Results

Simulation of the Opposed-Jet H_2/Air Burner

The DNS data were obtained with a spectral-element code that solves the low Mach number form of the conservation equations for momentum, species, and energy of compressible chemically reactive flows [8]. The problem solved was that of a 20% H_2 (by mole) in N_2 and air diffusion flame, stabilized on an opposed-jet burner in a two-dimensional, axisymmetric geometry. The geometry, the boundary conditions, and the grid employed in the simulations are shown in Fig. 1. For fixed geometry inflow-stream compositions and temperature, and for equal velocity magnitudes of the plug-flow profiles at the jet exits, the behavior of the system is parameterized only by the Reynolds number defined at the air side, $Re = u_{air} * D/v_{air}$. Extending the work presented in Ref. [9], simulations were performed in the range from low- to high-flow-rate extinction of the flame, corresponding to Reynolds number values from $Re = 10$ to $Re = 1{,}800$. For each Reynolds number, the simulation starts from an appropriate initial condition, and the whole transient response, until the steady-state flame is established, is computed by time-accurate integration. Resolution is adjusted to accommodate the flame thickness by varying the order of the interpolating polynomial in each element. For the low- to medium-strain-rate simulations ($Re = 20$, 50, 100, 200, and 400), it was found that fourth-order Legendre polynomials (5×5 elemental resolution) in both the radial and axial direction of the 120 spectral-element mesh provide an accurate discretization. The detailed chemical kinetics of Ref. [10] (9 species, 19 reversible elementary reactions) and multicomponent transport based on the mixture-fraction formulation were used in the simulations.

Data Reduction

Forty-nine snapshots (two-dimensional fields at steady state and along the transients) of velocity, pressure, temperature, and species mass fractions (a total of 13 variables each) were extracted from the simulations and interpolated on a uniform 101 $\times$

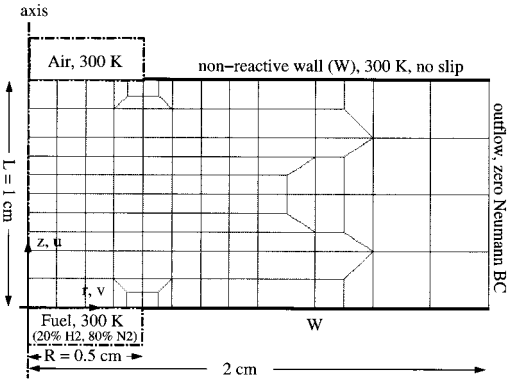

FIG. 1. Spectral element skeleton, coordinate system, and boundary conditions used for the opposed-jet DNS (uniform inflow velocity profiles for both fuel and air).

201 grid. To account for the different units and magnitudes of the variables, they were all normalized between 0 and 1 and then arranged in a matrix, $\boldsymbol{U}' = [\boldsymbol{u}'_1, \boldsymbol{u}'_2, \ldots, \boldsymbol{u}'_M]$, of $M = 49$ one-dimensional column vectors of length $101 \times 201 \times 13$. The latter was subsequently processed with the code KLTOOL [11], which found the orthogonal eigenvectors (the POD modes) using the method of snapshots [12] as follows:

1. Compute the mean vector, $\bar{\boldsymbol{u}}$, with
$$\overline{u_i} = \frac{1}{M}\sum_{j=1}^{M} U'_{ij}$$
2. Compute the vectors with zero mean (variations), $\boldsymbol{u}_i = \boldsymbol{u}'_i - \bar{\boldsymbol{u}}$, and arrange them in a matrix, $\boldsymbol{U}$, as before.
3. Compute the $M \times M$ covariance matrix,
$$\boldsymbol{C} = \frac{1}{M}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}).$$
4. Compute the non-negative eigenvalues, λ_i, $i = 1, \ldots, M$, in the order of decreasing magnitude and the corresponding eigenvectors, ϕ_i, of the covariance matrix, $\boldsymbol{C}$.

The orthogonal eigenfunctions (modes) can then be written as

$$\boldsymbol{\psi_k} = \sum_{i=1}^{M} \phi_{ik}\boldsymbol{U_i}$$

where ϕ_{ik} is the ith component of the kth eigenvector. As an example, the mean field (dimensional) and the six modes (dimensionless) of temperature are shown in Fig. 2. The computed modes describe the variations of the data from the mean. As Re is increased, the flame becomes thinner and longer, as reflected by the isocontours of the temperature modes that show large variations on both sides and close to the tip of the mean field. By projecting the database vectors onto the modes $\psi_k(\boldsymbol{x})$, $k = 1, \ldots, N < M$, the coefficient, a_k, that provide the optimal approximation to the full data,

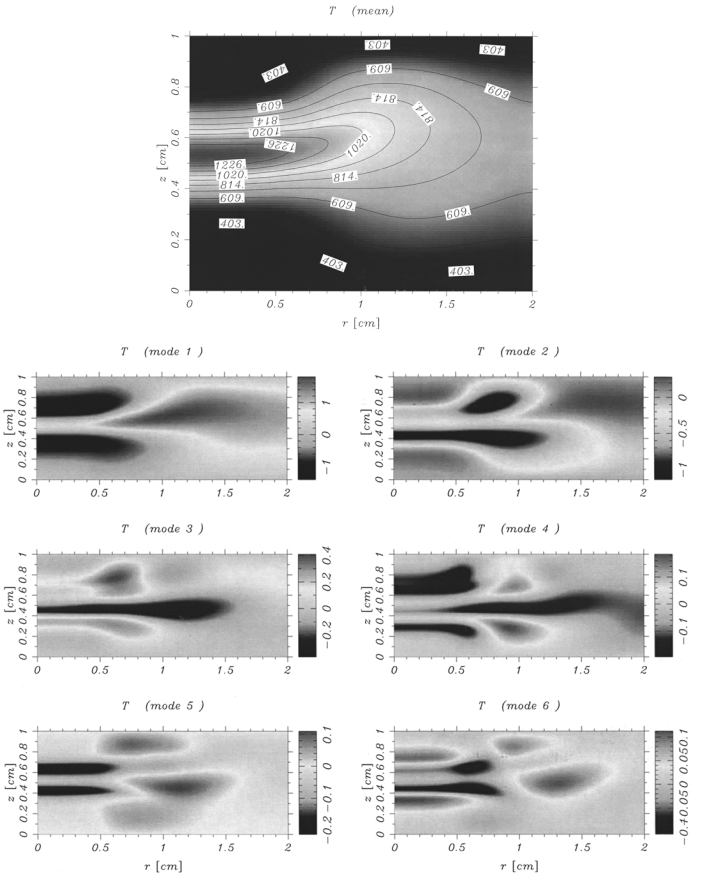

FIG. 2. Mean field and first six POD modes for temperature. The mean field is dimensional, while the modes are dimensionless.

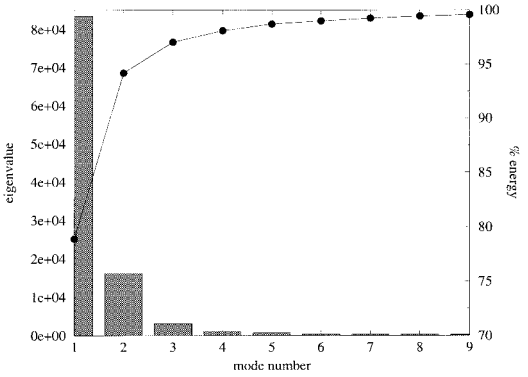

FIG. 3. Histogram of the nine largest eigenvalues (left axis) and percentage of the energy captured by the corresponding modes (line with filled circles, right axis).

$$\boldsymbol{u}' \cong \bar{\boldsymbol{u}} + \sum_{k=1}^{N} a_k \boldsymbol{\psi_k}(\boldsymbol{x})$$

are obtained as

$$a_k = \frac{\boldsymbol{u} \cdot \boldsymbol{\psi}_k}{\psi_k \cdot \boldsymbol{\psi}_k}$$

The sum of the eigenvalues of the covariance matrix is defined as the "energy" of the data. A commonly used criterion for selecting the number of modes to keep in a low-dimensional representation is that of capturing a certain percentage of the energy of the data. As can be seen in Fig. 3, only $N = 6$ and $N = 9$ modes are needed to capture $\Sigma_{n=1}^{N} \lambda_n / \Sigma_{n=1}^{M} \lambda_n =$ 99% and 99.5% of the energy of the DNS data, respectively. The rest of the modes are considered comparatively meaningless and are ignored from the low-dimensional representation. Indeed, velocities, temperature, and species mass fractions (with the exception of HO_2 and H_2O_2, which can be as much as 25% off in some regions) of all the original data fields are reconstructed with less than 13% relative error over the whole two-dimensional domain, with both six and nine modes. The additional three modes offer marginal improvement in the accuracy of the approximation of velocity and mass fraction of major species but have a stronger effect on the mass fractions of minor species.

Accurate approximations were obtained not only for the DNS data used to compute the modes but also for data at intermediate values of the Reynolds number (i.e., for $Re = 20 \div 400$). Fig. 4 compares the DNS fields of temperature and OH mass fraction with those obtained with six modes for the steady-state solution at $Re = 300$ (which was not used to obtain the POD modes). In this case, increasing the number of modes to nine did not reduce the relative error.

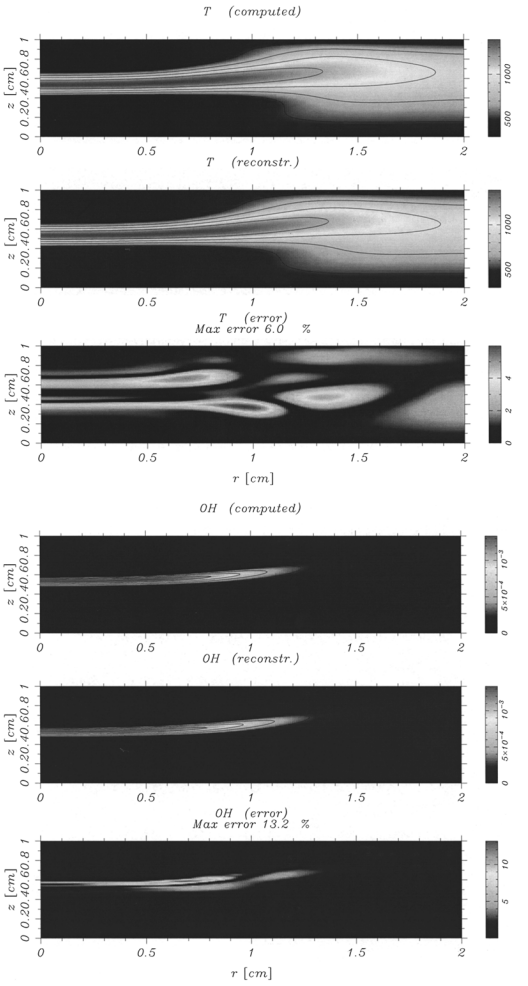

FIG. 4. $Re = 300$: comparison of the DNS two-dimensional fields of temperature (top) and OH mass fractions (bottom) and their approximation via six POD modes. The distribution of the relative error is also plotted.

We can conclude that a linear combination of only six modes provides an accurate low-dimensional representation of the data generated by the full (26,000 degrees of freedom) model for the range of *Re* values used to compute the POD modes. In principle, one can substitute this approximation back into the conservation equation and derive the equations of change for the coefficients, a_k (see, for example, Refs. [5,6] for non-reactive flows), which will effectively reduce the original 26,000 ODEs to only 6 ODEs. In the next subsection, however, we consider the problem of using the available POD modes together with measurements of some of the variables (e.g., temperature) to obtain an estimate of the unmeasured variables.

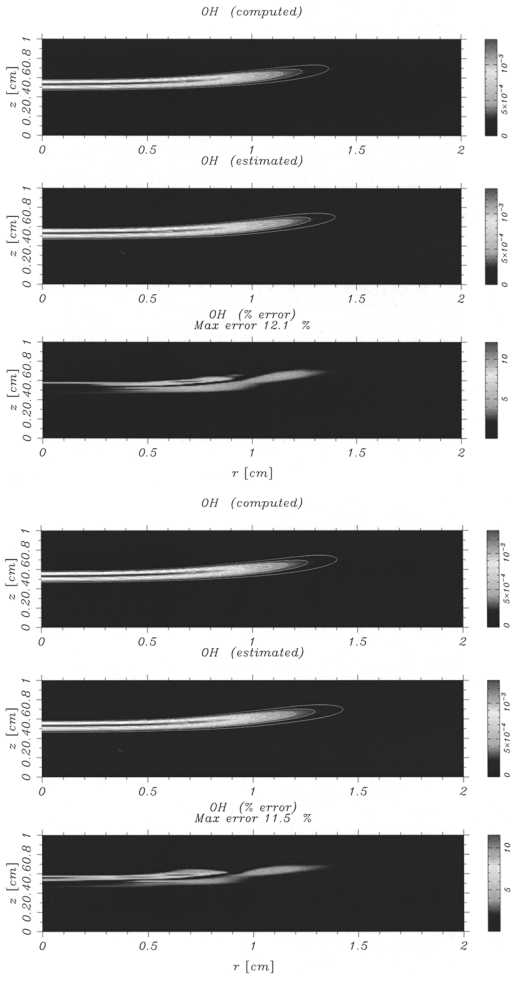

FIG. 5. $Re = 300$: comparison of the DNS OH field with its estimate using six (top) and nine (bottom) POD modes together with temperature two-dimensional measurements. The distribution of the relative error is also plotted.

POD-Based Observers

Since only a small number of POD mode can provide a good approximation to the DNS model, in principle, only the same small number of measurements at the proper locations (i.e., with non-trivial components in the span of the POD modes) is sufficient to evaluate the coefficients, a_k, from which the entire field can be approximated. In fact, one can even find the optimal measurement points, an issue we are currently working on but which will not be considered here. Therefore, a few measurements of easily measured quantities (e.g., temperature or velocity) at selected locations, together with the POD modes, can be used to effectively obtain information on other quantities in the entire domain which cannot be easily measured, such as radical species concentrations. Typically, measurements are performed at many more points than the small number of POD coefficients, resulting in an overdetermined system of linear equations. In this case, the estimate of the POD coefficients are found as a least-square solution to the problem using singular value decomposition [13].

This procedure was applied to our data, substituting the experiment with direct simulation. Two cases commonly used in the experimental study of opposed-jet burners were considered: measurement (1) of two-dimensional fields and (2) along the axis of symmetry. Fig. 5 compares the computed OH mass fraction fields with those obtained using six modes (top half) and nine modes (bottom half) and the temperature two-dimensional field measurement (i.e., at 101×201 points) to estimate the coefficients, a_k, for the $Re = 300$ case. It should be stressed that the $Re = 300$ data were *not* used to compute the POD modes. Again, with the exception of HO_2 and H_2O_2, the "unmeasured" quantities can be estimated with less than 13% maximum relative error. In this case, increasing the number of modes to nine offered only marginal improvement. Good results were obtained even when temperature was measured at only 101 points along the axis of symmetry, as shown in Fig. 6 for the $Re = 300$ case. The behavior and values of the relative errors are practically unaffected in a large region around the axis. This indicates that it is indeed possible to use only a handful of carefully selected measurement points (together with the POD modes) to estimate the remaining quantities over the two-dimensional domain. Higher relative errors are observed in the low OH mass fraction regions away from the axis, where, in the case of axial data only, the observer is not expected to perform well in the first place. When the number of modes is increased from six (Fig. 6, left half) to nine (Fig. 6, right half) close to the axis, the relative error decreases. The relative error is almost doubled, however, in a region one diameter away from the axis, where, again, the OH mass fraction is very low. In passing, we would like to comment on an issue raised by a reviewer with respect to the optimal number of modes for the observers. Different criteria have been proposed [1], with the energy criterion used here being the most popular one. Once the energy threshold is chosen, the optimal representation is fixed. The quality of the observer estimates is then solely determined by the principal angles [13] between the POD and the measurement subspace. It should also be noted that the maximum relative error (L_∞ norm) is just one way to quantify the accuracy of the estimations; any error norm can be selected (in fact, the L_2 norm would give lower error values). We would like to add at this point, as Prof. Gouldin pointed out during the symposium to one of the authors (C.E.F.), that the Karhunen-Loève basis was exploited in a similar way in

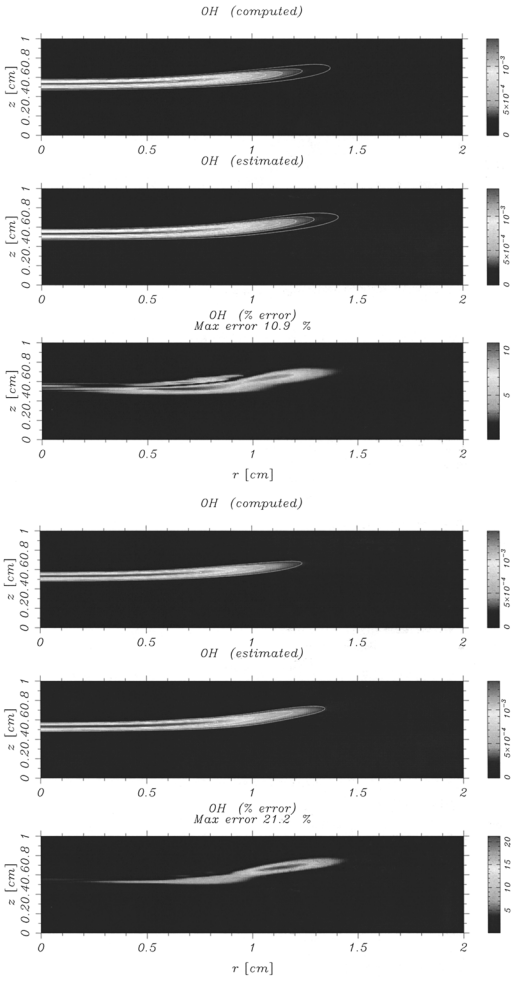

FIG. 6. $Re = 300$: comparison of the DNS OH field with its estimate using six (top) and nine (bottom) POD modes together with temperature measurements along the axis of symmetry. The distribution of the relative error is also plotted.

the evaluation of a tomographic inversion method as a combustion diagnostic tool [14].

Conclusions and Further Work

The method of POD was applied to DNS data for an opposed-jet H_2/air burner. It was found that only six modes capture 99% of the total "energy" of the original system of 26,000 degrees of freedom. In contrast to the traditional reduction techniques applied to chemically reactive systems, this method takes into account not only chemical kinetics but also transport and transient phenomena in a full two-dimensional context. The results suggest that the problem investigated here (26,000 ODEs) can be described effectively with only 6 ODEs. However, the construction of a low-dimensional dynamic model based on the POD modes is complicated by the highly nonlinear form of the reaction rate terms in the conservation equations. It should be possible to overcome this problem either by moving between the (small) POD space and the original (large) space or by fitting a functional form to the right-hand side of the ODEs for the POD coefficients. We plan to explore both approaches, which will allow for the true model reduction to only six ODEs, in the near future.

In addition to data reduction, the POD modes were used in a way which can be of interest in experiments. Since only a few modes are needed to provide a good representation of the full system, only a handful of measurements at selected points can, in principle, be used to obtain estimates of the coefficients, a_i, and therefore of the unmeasured quantities. The problem of selecting the measurement points was not addressed in this paper. Instead, the good performance of the proposed methodology was shown for the case where the number of measurements exceeds the number of modes. The coefficients in this case are obtained as a least-square solution to the corresponding overdetermined system of linear equations. We are currently working on an algorithm for the selection of the minimal number of sensors and their location for the accurate estimation of the complete field. Since well-resolved DNS data were used to obtain the POD modes, we did not address here the effect of noise in the initial data. This will be particularly important for experimental data processing, and we plan to examine this in the future.

Acknowledgments

This work was performed with the financial support of the Swiss National Energy Research Foundation (SNSF), the Swiss Office of Energy (BFE), the Air Force Office of Scientific Research (AFOSR), and the United Technologies Research Center (UTRC).

REFERENCES

1. Preisendorfer, R. W., and Mobley, C. D., *Principal Component Analysis in Meteorology and Oceanography,* Elsevier, Amsterdam, 1988.
2. Fukunaga, K., *Introduction to Statistical Pattern Recognition,* Academic Press, Boston, 1972.
3. Jolliffe, I. T., *Principal Component Analysis,* Springer, New York, 1986.
4. Holmes, P., Lumley, J. L., and Berkooz, G., *Turbulence, Coherent Structures, Dynamical Systems and Symmetry,* Cambridge University Press, Cambridge, 1996.

5. Bangia, A. K., Batcho, P. F., Kevrekidis, I. G., and Karniadakis, G. E., *SIAM J. Sci. Comp.* 18(3):775–805 (1997).
6. Deane, A. E., Kevrekidis, I. G., Karniadakis, G. E., and Orszag, S. A., *Phys. Fluids A* 3(10):2337–2354 (1991).
7. Podvin, B., Lumley, J., *Phys. Fluids* 10(5):1182–1190 (1998)
8. Tomboulides, A. G., Lee, J. C. Y., and Orszag, S. A., *J. Sci. Comput.* 12(2):139–167 (1997).
9. Frouzakis, C. E., Lee, J. C., Tomboulides, A. G., and Boulouchos, K., *Proc. Combust. Inst.* 27:571–577 (1998)
10. Yetter, R. A., Dryer, F. L., and Rabitz, H., *Combust. Sci. Technol.* 79:97 (1991).
11. Heiland, R., "KLTOOL: A Mathematical Tool for Analyzing Spatiotemporal Data," Masters thesis, Dept. of Mathematics, Arizona State University, Tempe, AZ, 1992.
12. Sirovich, L., *Q. Appl. Math., XLV,* 561–571 (1987).
13. Golub, G. H., and van Loan, C. F., *Matrix Computations, 2nd ed.* Johns Hopkins University Press, Baltimore, MD, 1989.
14. Torniainen, E. D., Hinz, A. K., and Gouldin, F. C., *AIAA J.* 36(7):1270–1278 (1998).