Lawrence Berkeley National Laboratory

Title
Dimensionality estimate of the manifold in chemical composition space
for a turbulent premixed H2+air flame

Permalink
https://escholarship.org/uc/item/5w95q7pr

Authors
Tonse, Shaheen R.
Brown, Nancy J.

Publication Date
2003-02-26

Dimensionality Estimate of the Manifold in

Chemical Composition Space for a Turbulent Premixed

H$_2$+air flame

Shaheen R. Tonse and Nancy J. Brown

Environmental Energy Technologies Division,

Lawrence Berkeley National Laboratory, Berkeley, CA 94270, USA.

* Author for correspondence.

Telephone: Int + 1(510)486-4556

Fax: Int + 1(510)486-5928

Email: tonse@lbl.gov

Abstract

The dimensionality ($\mathcal{D}$) of manifolds of active chemical composition space has been measured using three different approaches: the Hausdorff geometrical binning method, Principal Component Analysis, and the Grassberger-Procaccia cumulative distribution method. A series of artificial manifolds is also generated using a Monte Carlo approach to discern the advantages and limitations of the three methods. Dimensionality is quantified for different levels of turbulent intensity in a simulation of the interactions of a 2D premixed hydrogen flame with a localized region of turbulence superimposed over the cold region upstream of the flame front. The simulations are conducted using an adaptive mesh refinement code for low Mach number reacting flows. By treating the N_s species and temperature of the local thermo-chemical state as a point in multi-dimensional chemical composition space, a snapshot of a flame region is mapped into chemical composition space to generate the manifold associated with the 2-D flame system. An increase in $\mathcal{D}$ was observed with increasing turbulent intensity for all three methods. Although each method provides useful information, the Grassberger-Procaccia method is subject to fewer artifacts than the other two thereby providing the most reliable quantification of $\mathcal{D}$.

1 Introduction

Reactive computational fluid dynamics (CFD) calculations with large (10^5–10^6) 3D grids and large (≥ 50 species) chemical reaction sets require large parallel computers. A substantial fraction of the computing resources is used for integrating the chemical kinetic rate equations, which are ordinary differential equations (ODE's), to calculate the changes in concentration of the chemical species and enthalpy. A 2-dimensional calculation with the chemistry of a H_2-air mixture, using a reaction set of 9 chemical species and 29 reactions, spends 85 to 90% of the CPU time on the chemistry. The hydrogen reaction set is a comparatively simple one. If larger reaction sets, for example, with 50 chemical species are used, the fraction of CPU time spent on chemistry is even greater. Considerable effort has been devoted to reducing the complexity of the chemistry so that the computational burden it imposes is reduced. Some of the approaches that have been pursued are:

- Reduction of the reaction set, both in the number of chemical species and the number of reactions, in a systematic way after examining sensitivities and reaction fluxes. [1–3]

- Steady-state and partial equilibrium approximations. [4, 5]

- Computational Singular Perturbation [6], based upon classifying reactions according to their time scales. During the various stages of combustion, fast reaction groups are excluded after they equilibrate.

- Using mathematical analysis methods to upgrade the performance of existing chemical kinetic numerical solvers, through implementation of DAEPACK [7], which is especially useful for sparse chemical mechanisms of several hundred species.

- Optimization Approach [8] where a global reduction of the original system of chemical rate equations occurs by replacing them with a reduced set which inherits the stability and non-linear behavior of the original set.

- Principal Component Analysis [9], an eigenvalue-eigenvector analysis of the sensitivity matrix of the normalized sensitivity coefficients in which the magnitude of the coefficients in the eigenvectors measure the importance of the reactions for a given eigenvalue.

- The Intrinsic Low-dimensional Manifold Method [10–12] based upon a reaction trajectory in chemical composition space quickly converging to a low-dimensional manifold of dimension n_c, where $n_c < n_s$, the total number of species. The approach depends on the existence of many different reaction time scales during the combustion process.

All these methods reduce the severity of the computational problem, but still require the solution of differential equations.

Another family of approaches reduces the cost of complex chemistry by developing models in which the evolution of the chemical kinetics over the time interval Δt is viewed as a mapping from one chemical composition (and temperature) to another: $\mathcal{F}(C_1^t, C_2^t, \ldots, T^t, \Delta t) \rightarrow (C_1^{t+\Delta t}, C_2^{t+\Delta t}, \ldots, T^{t+\Delta t})$, where C_i^t is concentration of species i at time t. An inexpensive approximate mapping Φ is constructed to replace $\mathcal{F}$. Approaches within this class include: Solution Mapping [13]; Piecewise Reusable Implementation of Solution Mapping (PRISM) [14]; Fifth- to eighth-order polynomial parameterizations [15]; *In situ* adaptive tabulation (ISAT) [16,17]; and laminar flamelet libraries [18]. Typically the construction of the approximate mapping Φ requires the solution of the chemical kinetics rate equations at several discrete chemical compositions followed by a parametrization

procedure to evaluate the solution over a continuous range of input chemical compositions. Further utilization of the model does not require additional ODE integration, and is far less expensive. As a result of the nonlinear character of kinetic rate equations, Φ is only valid in some localized neighborhood. Hence we compile libraries of such localized neighborhoods, each with its own slightly different parameterization. Both the extent of this neighborhood and the order of the parametrization control the model accuracy. In PRISM, Φ is a set of quadratic polynomials defined within a hypercube and in ISAT, it is a set of linear polynomials defined on ellipsoidal regions.

The purpose of this study is to address a key question important to these approaches, upon which their success hinges: *Is there sufficient reuse of these localized mappings to offset the cost of their construction?* The answer is related to the geometry of the solution in chemical composition space. For a fluid having N_s species and temperature at any instant of time, we describe the local thermo-chemical state as a point in an N_s+1 dimensional chemical composition space, denoted hereafter by $\mathcal{C}$. The portion of this space accessed over the course of a simulation will be referred as "active" $\mathcal{C}$. It covers only a small portion of $\mathcal{C}$, and also has lower dimensionality. This is partly because of physical constraints. In a simulation conducted at constant pressure, a local portion of "active" $\mathcal{C}$ in a small temperature range would be restricted by Gay-Lussac's Law (equal volumes of gas at equal pressure and temperature contain equal numbers of molecules) to constrain the sum of molar concentrations of all species to be nearly constant. Further, during chemical reactions, while elements are exchanged between species, the total amount of each element is conserved, e.g., in hydrocarbon combustion, conservation of O,H, and C provides three additional constraints. Another less obvious restriction arises from the stiff nature of the chemical kinetic rate equations, where often the concentrations of species with fast reaction times appear to be functions of species

with slow reaction times. This is not a strict constraint, but it further reduces the degrees of freedom that "active" $\mathcal{C}$ exhibits, as demonstrated by Pope and Maas' ILDM method [10]. Finally, in CFD simulations, as in nature, the chemical composition between neighboring spatial points varies without discontinuities, a result of molecular diffusion. When compositions are mapped into $\mathcal{C}$, points that are neighbors in physical space tend to be neighbors in $\mathcal{C}$ and the dimensionalities in the two spaces tend to be nearly equal. These considerations indicate that the dimensionality of "active" $\mathcal{C}$ will be much less than N_s+1. It is possible that as one examines "active" $\mathcal{C}$ at coarser, finite length-scales, the observed dimensionality may change, e.g., if sufficiently convoluted it may appear to have a higher dimension, whereas if confined and localized it may appear to have a lower dimension. The "observed" dimensionality commonly changes with the length-scale of the observation. To illustrate this, consider a very loosely bunched, convoluted ball of string. At the lowest length-scale we would see its true dimension, $\mathcal{D}=1$. As we increase the lengthscale of observation it will appear to be more convoluted with a higher dimension of $\mathcal{D}=3$. As the lengthscale of observation continues to increase eventually we would observe the ball of string to be a point, with $\mathcal{D}=0$.

Returning to PRISM, in which $\mathcal{C}$ is contained in many hypercubes, it is sometimes necessary to more finely resolve $\mathcal{C}$ into smaller hypercubes to achieve better accuracy. It is here that a high value of dimensionality of "active" $\mathcal{C}$ could potentially lead to a large degradation of performance. Bisecting every side of a hypercube results in 2^{N_s+1} daughter hypercubes, viz. 1024 if $N_s=9$. Fortunately PRISM is only concerned with non-empty hypercubes, so were the resolution to be increased by a factor 2, the number of unoccupied hypercubes would increase as $2^{\mathcal{D}}$, where $\mathcal{D}$ is the dimensionality of $\mathcal{C}$. (How this comes about is explained later under the Hausdorff method [19, 20]

in Sec. 2.2.) PRISM is economically viable if the mean reuse of hypercubes is sufficiently high; any increase in the number of required hypercubes results in a corresponding decrease in mean reuse. Mean values of hypercube reuse that we have observed for common simulations suggest that an observed $\mathcal{D}$ of 2 would be acceptable. If $\mathcal{D}$ were four, the resultant penalty factor of $2^4 = 16$ would certainly impact computational savings. We are interested in $\mathcal{D}$ at the lengthscales of approximate PRISM hypercube sizes to assess the computational penalty.

In a previous study [21] we determined $\mathcal{D}$ for a similar system using only the Hausdorff method, and using a manifold obtained by summing over many time snapshots. While this provided us with useful information, it did not really allow an exploration of the impact of turbulent intensity on $\mathcal{D}$. Additionally the Hausdorff method is subject to artifacts that are discussed later. To address the limitations of the previous study, the current study was conducted to explore the impact of increasing turbulent intensity on the performance of PRISM and comparable methods. We numerically investigate the dimensionality of "active" $\mathcal{C}$ at several instantaneous values. Computer simulations of a premixed hydrogen flame propagating through an unburnt turbulent mixture in 2D are conducted, and the portion of $\mathcal{C}$ being accessed at any instant is obtained. To further probe the dimensionality without the inherent limitations of the Hausdorff method, we added two other independent measures of $\mathcal{D}$: Principal Component Analysis (PCA) [22], and Grassberger & Procaccia's r^ν method [23, 24]. A description of the three methods and an investigation of their systematic biases, determined by applying them to Monte Carlo-generated manifolds of known dimension, is presented in Sec. 2. We also discuss the relative advantages and disadvantages of each method. In Sec. 3 , the dimensionality of manifolds associated with differing levels of turbulent intensity is presented and discussed.

2 Method

2.1 Generation of the Manifold

We generate manifolds of known dimensionality using a Monte Carlo method, as an approach to determine the effectiveness and biases inherent in our methods. This has the added advantage of serving as a debugging tool. Manifolds associated with the reactive flow calculation are generated by first conducting the simulations from which we extract instantaneous spatial snapshots of the reacting system species' concentrations and temperature. We obtain a geometrical representation of "active" $\mathcal{C}$ by plotting the concentration and temperature coordinates of every CFD grid cell in an N_s+1-dimensional space. $\mathcal{C}$ is then divided into hypercubes, with each occupied hypercube containing a portion of the manifold, upon which we will perform measurements of $\mathcal{D}$.

Monte Carlo Manifold: In order to understand better the systematic biases of the three methods, each was tested on Monte Carlo'd manifolds of known $\mathcal{D}$ ranging from 1 to 4, embedded in a space of 9 dimensions. To evaluate the three methods we generated an ensemble of manifolds with varying orientations and placement within hypercubes, and composed of varying numbers of points. Most evaluations were conducted with purely linear manifolds (lines, planes, etc.) embedded in a 9D space, although we occasionally added quadratic terms. The Monte Carlo method uses simple analytical geometry and a random number generator (RNG).

We illustrate the procedure with a simple example of a linear 2D manifold (composed of points) embedded in a 9D space. For this we require the analytical geometry equations for a 2D plane

given by:

$$y_i = \sum_{j=1}^{2}(m_{i,j}x_j) + c_i \ldots i = 1, 2 \ldots 7 \tag{1}$$

The RNG is used to choose random values for the slopes $m_{i,j}$ and intercepts c_i selected from a uniform distribution, from which the equation for the plane is constructed. Since all 9 coordinates are equivalent in the beginning by symmetry, the first two are treated as "independent" and the last seven as "dependent" with no loss of generality. To add pure quadratic terms to Eqn. 1 the form of the equation will be:

$$y_i = \sum_{j=1}^{2}(m_{i,j}x_j + q_{i,j}x_j^2) + c_i \ldots i = 1, 2 \ldots 7 \tag{2}$$

once again with $q_{i,j}$ chosen by the RNG. The next step is the generation of discrete points on this plane. To generate each point, x_1 and x_2 are chosen with the RNG from a uniform distribution, and then Eqns. 1 are evaluated to give a point with coordinate in 9D: $[x_1, x_2, y_1, y_2 \ldots y_7]$.

Depending on the case, between 100 and 1000 manifolds were constructed for each dimensionality , with the slope and intercept values allowed to vary between specified limits so that the final ensemble of manifolds has a variety of positions and orientations within the hypercubes. Random points (varying from 100's to 1000's) were generated to create each manifold, and these were additionally scaled and shifted to resemble the chemical data. This is a good practice, useful for uncovering possible bugs in the $\mathcal{D}$-measuring code. The three $\mathcal{D}$-measuring methods were applied to the manifolds generated. Fig. 1 depicts such 2,3 and 4D linear manifolds in a 9D space.

[Figure 1 about here.]

The figure illustrates that it is easy to mistakenly identify the 3D manifold as being a fuzzy 2D manifold and the 4D manifold as being a fuzzy 3D manifold when visual examination is used.

Visual examination should not replace mathematical measures of $\mathcal{D}$. The main uses for the Monte Carlo generated manifolds were:

- to measure systematic biases and errors associated with each method.

- to determine reasonable values for method parameters. In the PCA method (Sec. 2.2), to be counted as a principle component, an eigenvalue needed to be > 0.1, and the minimum number of such eigenvalues that could account for 90% of the variance would be interpreted as $\mathcal{D}_2$.

- to examine the effect of quadratic curvature on the result, particularly for PCA, which we expect to be most susceptible to it.

- to reduce the number of manifold points and to determine the sample size where the method becomes undependable.

- to use for debugging the algorithm and code.

Reactive Flow Manifolds: The reacting flow simulation is based on a conservative form of the low Mach number model introduced by Rehm and Baum [25] and developed by Bell and co-workers [26]. The system is evolved numerically, using a modified version of the adaptive algorithm for low Mach number reacting flows as discussed by Day and Bell [26]. The method employs an approximate projection algorithm incorporated into a hierarchical block-structured adaptive framework. The underlying algorithm consists of an explicit Godunov scheme for convective transport, and a time-centered Crank-Nicolson discretization for diffusion coupled to a time-implicit integration for the stiff chemistry production terms.

The above methodology is applied to the interactions of 2D premixed hydrogen flames with a localized region of turbulence [27] superimposed over the cold region upstream of the flame front (see Fig. 2). Inflow occurs at the bottom of the domain, while outflow occurs at the top. The inflow velocity is set to the laminar flame speed so that in the absence of turbulence, the flame is stationary. The sides of the domain are periodic. The turbulence is present as a superimposed patch, which decays a small amount before it reaches the flame; we report the actual turbulent intensity that we measure at the flame. The inflow velocity convects the turbulent region toward the flame front. The chemistry mechanism contains 9 species and 27 reactions and is derived by removing carbon-containing species from the GRI-Mech 1.2 reaction model [28]. We compute the solution with turbulent intensities in the range 0.2 m/s to 10 m/s. These intensities represent a broad range of wrinkled flame phenomena.

[Figure 2 about here.]

2.2 Measurement of $\mathcal{D}$

Three methods are used to measure $\mathcal{D}$, the Hausdorff method, Principal Component Analysis (PCA), and Grassberger & Procaccia's r^{ν} method. For purposes of comparison the values of $\mathcal{D}$ obtained using them are denoted $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$, respectively. The first and third methods have been used in the field of fractal dimensionality where it is frequently necessary to measure the dimensionality of a manifold, often called a strange attractor. The attractors consist of discrete points that are generated by reiterating a recursive equation, in which the outputs from the nth iteration become the inputs for the $n + 1^{th}$ iteration.

11

Hausdorff method ($\mathcal{D}_1$): This method is applied by binning $\mathcal{C}$ into hypercubes and counting the number of hypercubes that contain a portion of "active" $\mathcal{C}$. Next every side of each hypercube is halved, so that each hypercube is divided into 2^{N_s+1} daughter hypercubes. Not all of the daughters contain a part of "active" $\mathcal{C}$. For example, for a one-dimensional line in a multi-dimensional space, halving the edges of the hypercubes roughly doubles the number of occupied hypercubes, while in the case of a 2D plane, the number of occupied daughters roughly quadruples. In general, the number of occupied daughter hypercubes increases as $2^{\mathcal{D}}$, where $\mathcal{D}$ is the dimension of "active" $\mathcal{C}$. From a comparison of occupied hypercubes before (M_1) and after (M_2) bisection, $\mathcal{D}_1$ is obtained straightforwardly from $M_2 = M_1 \cdot 2^{\mathcal{D}_1}$. Limitations of this method are that it gives poor results when there are few points in a hypercube, since even a single count results in an occupied daughter. The position and orientation of the manifold within the hypercube also influences the number of occupied daughters.

Principal Component Analysis ($\mathcal{D}_2$): Principal Component Analysis (PCA) [22] is commonly used to determine the most significant variables or linear combinations of variables in a multi-variate data set by recasting it in terms of a new linear basis, so that most of the variance can be captured by as few of the new variables as possible. PCA works by formulating the problem as an eigenvalue problem, in which the resulting eigenfunction that corresponds to the largest eigenvalue identifies the axis along which the variance of the original data is maximized. As a simple example, consider an arbitrary line of finite extent in a 2D space. The points on this line generally exhibit variance along both dimensions. A PCA operation would return an eigenfunction aligned parallel to the line. Variance along the other (perpendicular) component would be zero. In general the orientations of

the principal axes in terms of the old axis vectors are the eigenfunctions returned by the PCA calculation. In our case PCA is applied to the portion of manifold contained within a single hypercube. The chemical composition points of the manifold are input to NAG library's g03aae PCA routine which returns the eigenvalues/functions. We rank these eigenvalues by decreasing size and define the number of principal components to be the number of eigenvalues sufficient to account for 90% of the variance. We interpret this as being the dimensionality $\mathcal{D}_2$. PCA is independent of the orientation of the manifold, and of its absolute size. A limitation of PCA is that the number of principal components should be small compared to the number of original axes. Non-linearity of the manifold can also be a problem, as the variance introduced by curvature in a particular direction can be interpreted as an extra dimension. Finally, for a manifold that consists of a high density cluster of points, a case we encounter as the initial condition of our simulation, the PCA method does not give a dependable result.

Grassberger & Procaccia's r^ν method ($\mathcal{D}_3$): This method is based on the purely geometrical consideration that for a manifold (composed of fairly uniformly distributed discrete points) of dimension ν, embedded in a space of larger dimension, the number of points within a hypersphere of radius r increases as $\approx r^\nu$ as r increases. To determine ν the pairwise distance $r = \mid \mathbf{r_i} - \mathbf{r_j} \mid$ between every pair (i,j) of points on the manifold is determined and the distribution of r, $N(r)$, is obtained. The cumulative distribution $C(r)$ is obtained by integrating this distribution.

$$C(r) \ = \int_0^r N(r')dr' \tag{3}$$

The rate at which $C(r)$ increases with r depends on the dimensionality of the manifold to which the points belong, such that $C(r) \ \alpha \ r^\nu$. We extract ν simply by plotting $C(r)$ in log-log form

and measuring the slope ν, which is interpreted as being the dimensionality $\mathcal{D}_3$. As mentioned earlier the r^ν method works under the assumption that the distribution of individual points on the manifold is fairly uniform. For a manifold that consists of several high density clusters of points, a case we encounter as the initial condition of our simulation, the $C(r)$ distribution will increase through discontinuous jumps followed by flat intervals. Under these conditions the method will not give a dependable result for ν.

3 Results and Discussion

3.1 $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{D}_3$ from Monte Carlo'd manifolds

As described earlier in Sec. 2.2 we constructed randomly oriented manifolds with known dimensionalities from $\mathcal{D}=1$ to 4 and placed them into hypercubes. All three methods of measuring $\mathcal{D}$ were applied to these manifolds. The results are summarized in Table 1 and are quite compatible with each other.

[Table 1 about here.]

In the Hausdorff method, the mean of the "# daughters" column associated with $\mathcal{D}_1$ follows the expected $M_2 = M_1 \cdot 2^\mathcal{D}$ trend fairly well, where M_1 and M_2 are number of occupied hypercubes before and after bisection, respectively. The large width (1σ) associated with the "# daughters", shows that $\mathcal{D}_1$ taken one hypercube at a time, differs from the true $\mathcal{D}$. This difference is caused by positioning and orientation of the manifold within the hypercube, which often results in very few points being left inside and it has a significant effect on the number of daughters. This can be

illustrated by drawing an oblique line through a square and then halving the sides of the square. The line passes through 1,2 or 3 daughter squares depending on its orientation and placement in the square. This variability in the number of daughters translates to a large error bar for $\mathcal{D}_1$. By contrast, $\mathcal{D}_2$ and $\mathcal{D}_3$ have smaller errors. The Hausdorff method also suffers when the number of points is low, (< 400) which happens if the manifold is tucked away in a corner of the hypercube. We retain the Hausdorff method because $\mathcal{D}_1$ indicates how exactly PRISM would experience an efficiency penalty, were an increase in resolution necessary. $\mathcal{D}_1$ is more useful to us as a measure of PRISM's computational penalty than as a measure of the observed dimensionality of the manifold.

The PCA method returned very accurate estimates of the dimensionality, eg. for $\mathcal{D}{=}3$ manifolds 92/92 hypercubes returned $\mathcal{D}_2{=}3$. For $\mathcal{D}{=}4$ manifolds 452/475 hypercubes returned $\mathcal{D}_2{=}4$, while the remaining 23 returned $\mathcal{D}_2{=}3$. We also tested the PCA method with quadratically curving manifolds that have a quadratic coefficient equal to the linear coefficient. In this case, for $\mathcal{D}{=}1$ only 2/50 hypercubes returned $\mathcal{D}_2{=}2$, while for $\mathcal{D}{=}3$ all 50/50 returned $\mathcal{D}_2{=}3$. If the quadratic coefficient is increased to be ten times the linear coefficient, then a $\mathcal{D}{=}1$ manifold is measured as having $\mathcal{D}_2{=}2$ about one quarter of the time, and a $\mathcal{D}{=}3$ manifold measured 50/50 times as $\mathcal{D}_2{=}3$. If the curvature of a $\mathcal{D}{=}3$ manifold is further increased by adding quadratic cross terms, then $\mathcal{D}_2{=}4$ is measured about one third of the time. The PCA method is usually reliable, and works well even with a sample as small as 200 points. As expected, it begins to break down when the manifold being measured deviates significantly from linearity. One other property of the PCA method is that the result is independent of the absolute size of the manifold being measured because it works through a relative comparison of the variances along different axes,

The r^ν method gives a dependable result even with a sample as small as 200 points. Other advantages of this method are that $\mathcal{D}_3$ is weighted by number of points of a particular dimensionality, so that if there were a variation in $\mathcal{D}$ within a hypercube, the measurement of $\mathcal{D}_3$ would be a weighted mean reflecting the number of points that contribute from each region. Another advantage of the approach is that by visually examining the slope of the $C(r)$ distribution computed as part of this method, one can see the length-scale dependence of $\mathcal{D}$. At length-scales approaching the hypercube size the points near the hypercube edge have no neighbors in the outer direction and thus contribute fewer counts to $C(r)$. This implies that the r^ν method is not dependable at this lengthscale.

3.2 $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{D}_3$ from reaction simulation

During our 2D case study in which a 2D turbulence field impinges on a laminar flame, we made 4 snapshots at different simulation times, during which the level of turbulent intensity rose as the turbulence convected into the flame. Table 2 shows the turbulent intensities used. Fig. 3 shows the spatial temperature field for snapshots 2–4 (snapshot 1 was similar in appearance to snapshot 2) and Fig. 4 shows the manifold in $\mathcal{C}$ for snapshots 1–3 (snapshot 4 was similar in appearance to snapshot 3). The three methods were applied to the manifold from each snapshot, with the results given in Table 2.

[Table 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

16

The first snapshot is at a time very shortly after the start of the simulation, before any turbulence-flame interaction. The initial condition was defined with uniform concentrations for the entire unburnt and burnt regions upstream and downstream of the flame respectively, and these appear as single points in $\mathcal{C}$. As a result $\mathcal{D}$ is very low. The initial condition across the flame region should be appear close to 1D, as the initial conditions of the posed problem are translationally symmetric across the grid from left to right. Our first time snapshot is made 10 timesteps into the simulation, and the symmetry has already been broken slightly as we see from the value of $\mathcal{D}_3$. A visual inspection of the manifold showed that it consisted of a few clusters, each with a large number of points (top panel Fig. 4). The PCA and r^ν methods do not work well on manifolds that are composed of a few widely separated clusters each consisting of many points and so the entries are omitted from the table. As the simulation time increases, $\mathcal{D}$ also rises to 2 as fluctuations, caused by molecular and turbulent diffusion break the symmetry. It further rises above 2 as the turbulence interacts with the flame and begins to convolute the manifold.

3.3   Discussion

We have made independent measurements of dimensionality of a chemical manifold in 2D (spatial) simulations of a turbulent flow field interacting with a premixed H_2+air laminar flame using three methods: the Hausdorff geometrical binning method ($\mathcal{D}_1$), principal component analysis ($\mathcal{D}_2$), and Grassberger & Procaccia's r^ν method ($\mathcal{D}_3$). The use of the three methods in conjunction is useful. The Hausdorff method is most susceptible of the three to low count statistics and to orientation and positioning of the manifold being measured. If the manifold being measured were translated or rotated sufficiently, we would most likely obtain the same result for the mean value of $\mathcal{D}_1$, but the

individual $\mathcal{D}_1$'s of each hypercube would change significantly whereas the other two methods would not do this. Additionally, while the Hausdorff method is simple to implement for regular shapes like hypercubes, it would not be simple for irregularly-shaped volumes or even for ellipsoids. We retain the Hausdorff method since it illustrates the increase in number of hypercubes that would be seen by PRISM were an increase in resolution necessary; consequently, it is a direct measure of penalty. PCA and r^ν work even with low count statistics. Additionally they are coordinate-independent, and so are less influenced by the orientation and positioning of the manifold being measured.

Fig. 5 allows us to visualize $\mathcal{D}_3$ as a spatial field. The figure is constructed by determining the value of $\mathcal{D}_3$ of the hypercube to which the concentration in a CFD cell belongs, and plotting $\mathcal{D}_3$ at the (x, y) position of each CFD cell. Fig. 5b and c show the changes in $\mathcal{D}$ as the turbulent vortices upstream of (below) the flame have begun to cause it to wrinkle. The panels in Fig. 5 correspond to those of Fig. 3. Note that the regions of highest $\mathcal{D}$ are toward the downstream side of the flame away from the vortices. In this study we have focused on measuring $\mathcal{D}$; however plots of this sort could be used to understand some of the underlying causes responsible for the increase in $\mathcal{D}$ in specific locations. This understanding could be achieved by comparing and analyzing plots like those shown in Fig. 5 with field plots of concentrations, temperatures or other derived variables such as vorticity.

[Figure 5 about here.]

In a previous work [21] we determined $\mathcal{D}$ using the Hausdorff method. In that study we combined the manifolds from all timesteps of the CFD simulation into one and made measurements upon it. We thereby lost important information regarding the influence of increasing turbulence intensity on $\mathcal{D}$. In this study the manifolds have been constructed as a function of the turbulence level

at the flame. The measurement of $\mathcal{D}$ has been conducted with the old approach and with two better methods that have much smaller uncertainties, and improved accuracy even under low count statistics, as well as coordinate-independent properties.

Under laminar conditions, where the flame is still relatively undisturbed (snapshot 2 in Table 2), $\mathcal{D}_2$ and $\mathcal{D}_3 \approx 2$. $\mathcal{D}_1$ indicates that the computational penalty of reducing hypercube sides by 2 would be $2^{1.6} \approx 3$. As the turbulence level increases, $\mathcal{D}_1$ increases to 2.6. $\mathcal{D}_2$ and $\mathcal{D}_3$ also give results close to this. Here the computational penalty for going to higher resolution would be $2^{2.6} \approx 6$, i.e. mean hypercube reuse would decrease by a factor 6. The large values of hypercube reuse that have been observed in previous simulations of H_2+air combustion [29] with the Coyote CFD code [30] interfaced with PRISM: 20,000 for a premixed 1D laminar flame, 12,000 for a premixed 2D flame, and 7,000 for a non-premixed 2D turbulent jet indicate that this would still be an acceptable penalty. Additionally in these previous studies, like the current one, we observe a trend toward reduced hypercube reuse and hence reduction in efficiency as the turbulence level increases.

There is some convolution of the manifold as turbulent intensity increases from 0.2 to 10 m/s, but only enough to increase $\mathcal{D}$ by approximately 1. A comparison of timescales of chemistry and turbulence is useful in showing why this is not surprising: The simulation uses cell sizes of 10^{-4} m, but the size of the turbulent features, i.e., vortices, is about 10^{-3} m. At turbulent intensities of 10 m/s, a timescale for changes brought about by turbulence, is 10^{-4} s. The chemical timescales are still faster than this, with the faster reactions occurring at $<10^{-7}$ s and the slower ones at 10^{-5} s. Thus we are still in the wrinkled laminar flame regime of turbulence-flame interaction. Increasing turbulent intensity by a further factor of 10 would make the timescales of turbulence and slow-chemistry comparable, and this could cause the manifold to convolve in more directions,

increasing $\mathcal{D}$ further.

For the dimensionalities measured, the possible computational penalties are still acceptable, but are borderline. At the very least, when designing a code, one should take into consideration that a required increase in resolution could have a severe economic penalty. It would be wise to design the code so that variable resolution could be used, such that hypercubes with increased resolution would only be used where necessary.

Acknowledgments

Appendix A   Notation

N_s: Number of chemical species

Δt: Length of timestep

$\mathcal{C}$: Chemical composition space of $N_\mathrm{s}+1$ dimensions. The term "active" $\mathcal{C}$ refers to the subspace of $\mathcal{C}$ that is accessed during the course of a reactive flow simulation.

$\mathcal{D}$: Dimensionality of a manifold

$\mathcal{D}_1$: Dimensionality measured by Hausdorff method

$\mathcal{D}_2$: Dimensionality measured by Principal Component Analysis method

$\mathcal{D}_3$: Dimensionality measured by Grassberger & Procaccia r^ν method

References

[1] Wang, H.; Frenklach, M. Combust Flame 1991, 87, 365–370.

[2] Frenklach, M. Reduction of chemical reaction models. In Numerical Approaches to Combustion Modeling; Oran, E. S.; Boris, J. P., Eds.; American Institute of Aeronautics and Astronautics: Washington, D.C., 1991.

[3] Hewson, J. C.; Bollig, M. Reduced mechanisms for NO_x emissions from hydrocarbon diffusion flames. In Twenty Sixth (International) Symposium on Combustion; The Combustion Institute: Pittsburgh, PA, 1996.

[4] Peters, N.; Williams, F. A. The structure of methane flames. In Complex Chemical Reaction Systems, Mathematical Modelling and Simulation, Vol. 47; Warnatz, J.; Jäger, W., Eds.; Springer-Verlag: Berlin, 1987.

[5] Ramshaw, J. D. Phys Fluid 1980, 23, 675.

[6] Lam, S. H.; Goussis, D. A. Int J Chem Kinet 1994, 26, 461–486.

[7] Schwer, D. A.; Tolsma, J. E.; Green, W. H.; Barton, P. I. Combust Flame 2002, 128, 270–291.

[8] Petzold, L. R.; Zhu, W. A I Ch E Jou 1999, 45, 869.

[9] Brown, N. J.; Li, G.; Koszykowski, M. L. Int J Chem Kinet 1997, 29, 393–414.

[10] Maas, U.; Pope, S. B. Combust Flame 1992, 88, 239–264.

[11] Blasenbrey, T.; Maas, U. Proc Comb Inst 2000, 28, 1623.

[12] Bongers, H.; van Oijen, J. A.; de Goey, L. P. H. Proc Comb Inst 2002, 29, sub.

[13] A. R. Marsden Jr.,; Frenklach, M.; Reible, D. D. J Air Pollut Control Assoc 1987, 37, 370–376.

[14] Tonse, S. R.; Moriarty, N. W.; Brown, N. J.; Frenklach, M. Israel J Chem 1999, 39, 97–106 (also LBNL # 42576).

[15] Turanyi, T. Comp Chem 1994, 18, 45–54.

[16] Pope, S. B. Combust. Theory Modelling 1997, 1, 41–63.

[17] Yang, B.; Pope, S. B. Combust Flame 1998, 112, 85–112.

[18] Bray, K. N. C.; Peters, N. Laminar Flamelets in Turbulent Flames. In Turbulent Reacting Flows; Libby, P. A.; Williams, F. A., Eds.; Academic Press: San Diego, CA, USA. 92101-4311, 1994.

[19] Mandelbrot, B. B. Fractals: Form and Chance Freeman: San Francisco, CA, 1977.

[20] Russell, D. A. Phys Rev Letters 1980, 45, 1175.

[21] Bell, J. B.; Brown, N. J.; Day, M. S.; Frenklach, M.; Grcar, J. F.; Propp, R. M.; Tonse, S. R. Proc Comb Inst 2000, 28, (also LBNL # 44732).

[22] Everitt, B. S.; Dunn, G. Applied Multivariable Data Analysis Oxford University Press: New York, NY, 1992.

[23] Grassberger, P.; Procaccia, I. Phys Rev Letters 1983, 50, 346.

[24] Grassberger, P.; Procaccia, I. Physica 1983, 9D, 189-208.

[25] Rehm, R.; Baum, H. N B S J Res 1978, 83, 297-308.

[26] Day, M. S.; Bell, J. B. Combust. Theory Modelling 1999, 4, 535–556 (also LBNL Report # 44682).

[27] Baum, M.; Poinsot, T. J.; Haworth, D. C.; Darabiha, N. J Fluid Mech 1994, 281, 1-32.

[28] Frenklach, M.; Wang, H.; Goldenberg, M.; Smith, G. P.; Golden, D. M.; Bowman, C. T.; Hanson, R. K.; Gardiner, W. C.; Lissianski, V. "GRI-Mech—An Optimized Detailed Chemical Reaction Mechanism for Methane Combustion", Technical Report GRI-95/0058, Gas Research Institute, 1995 http://www.me.berkeley.edu/gri_mech/.

[29] Tonse, S. R.; Moriarty, N. W.; Brown, N. J.; Frenklach, M. "Computational Economy Improvements in PRISM", Technical Report LBNL # 48858, Lawrence Berkeley National Laboratory, 2002 (submitted to Int. J. Chem. Kinet.).

[30] Cloutman, L. D. "Coyote: A computer program for 2D reactive flow simulation", Technical Report UCRL-ID-103611, Lawrence Livermore National Laboratory, 1990.

List of Tables

Table 1: Measurement of $\mathcal{D}$ on Monte Carlo'd manifolds of known dimensionality embedded in 9D space. The three methods are each applied to a large ensemble of manifolds differing in orientation, placement and number of points. The mean value of the distribution is shown in the table and the error signifies the standard deviation of $\mathcal{D}$ over the ensemble. Additionally the mean number of occupied daughter hypercubes per parent hypercube is shown for the Hausdorff method.

Manifold	Hausdorff		PCA	r^ν
$\mathcal{D}$	# daughters	$\mathcal{D}_1$	$\mathcal{D}_2$	$\mathcal{D}_3$
1	2.1 ± 1.1	1.1	1	$1.1 \pm .04$
2	3.5 ± 3.2	1.8	2	$2.0 \pm .1$
3	7.5 ± 7.1	2.9	3	$2.9 \pm .06$
4	12.7 ± 14.0	3.7	4	$3.7 \pm .05$

Table 2: Measurement of $\mathcal{D}$, using the three methods, on manifolds obtained from reactive flow simulation, as a function of turbulent intensity seen in the flame region, going from a quiescent laminar state to a turbulent state. For the case in which the manifold consists of a high-density cluster of points, i.e., $\mathcal{D} \approx 0$, PCA and r^ν have not been applied.

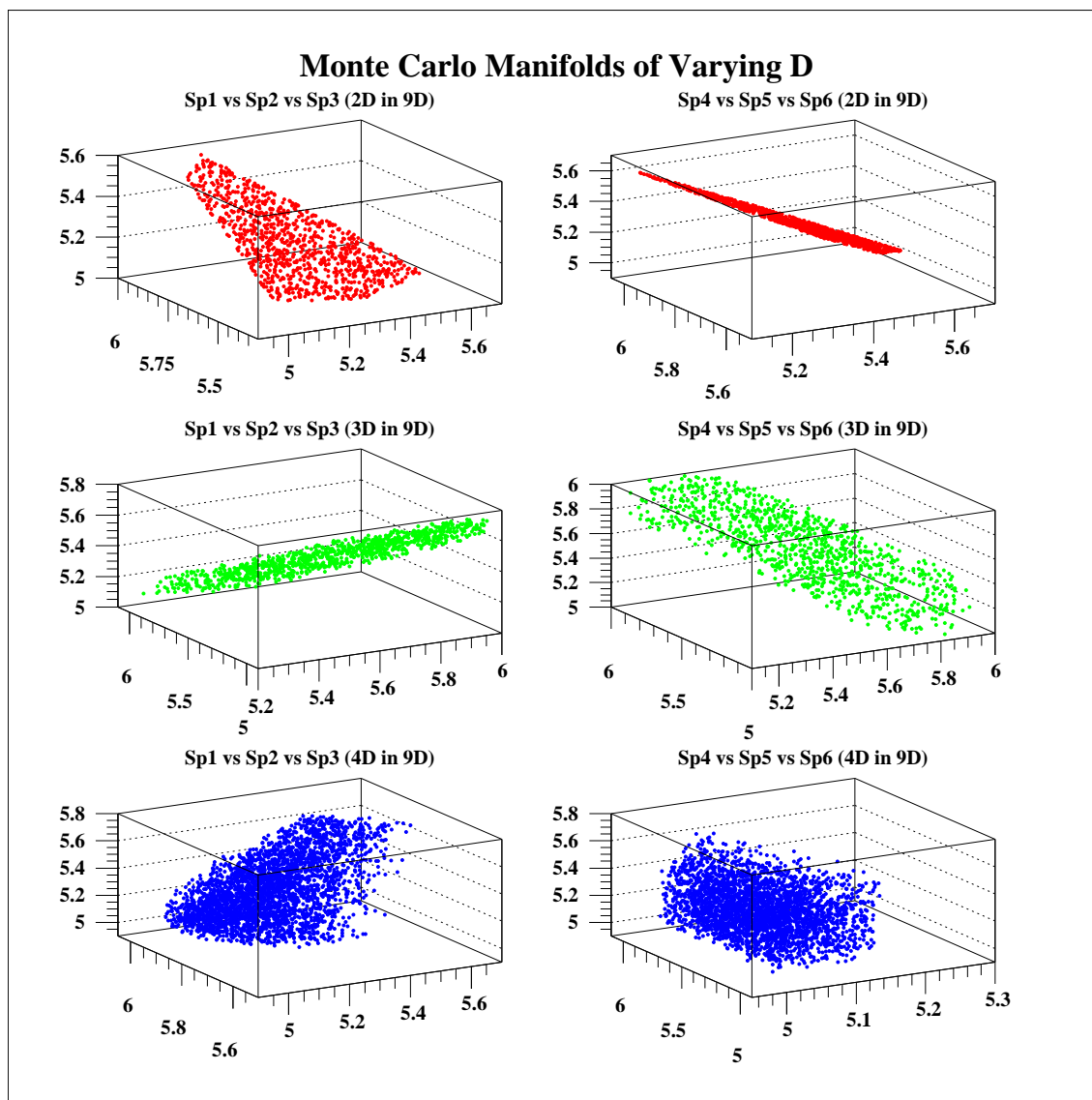| Snapshot #	Visual description	$(v_x)_{rms}$ in m/s	Simulation time(ms)	$\mathcal{D}_1$	$\mathcal{D}_2$	$\mathcal{D}_3$
1	laminar undisturbed	0.2	0.002	0.25	–	–
2	slightly deformed	2	0.05	1.6	1.9	2.1
3	developing turbulence	10	0.1	1.9	2.5	2.4
4	developed turbulence	10	1.1	2.6	2.9	2.4

List of Figures

Figure 1: 3D depictions of randomly oriented 2,3, & 4D manifolds embedded in a 9D space. The figures on the right depict the same manifolds as those on the left, but viewed using three different axis variables. The axis values reflect that the generated points were scaled and shifted to resemble chemical data and are -$log_{10}(concentration.)$

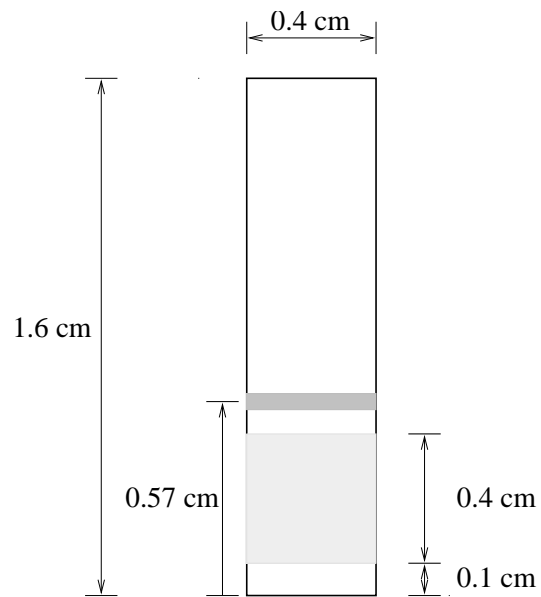Figure 2: Schematic of the sample problem. Inflow is at the bottom boundary and outflow at the top. The flame front (darker shading) is initially 0.57 cm from the bottom of the domain. Below it (lighter shading) is a 0.4 cm × 0.4 cm patch of turbulence which will be convected into the flame.
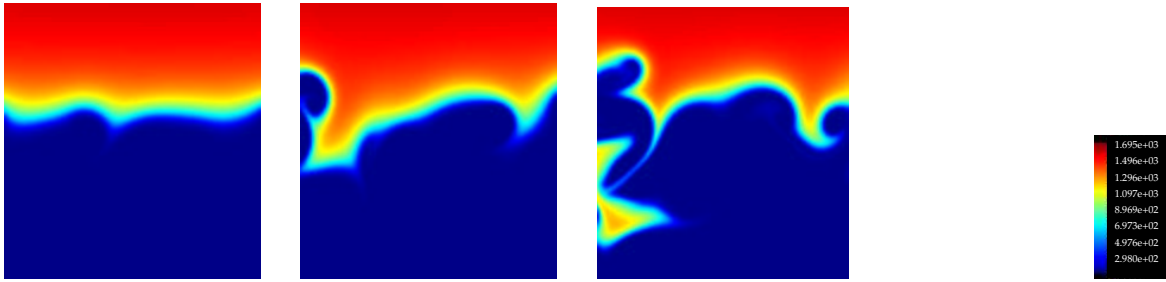
Figure 3: Temperature field near the flame front for turbulent intensity cases 2–4 in Table 2. The domains shown are about 0.4 cm. by 0.4 cm.
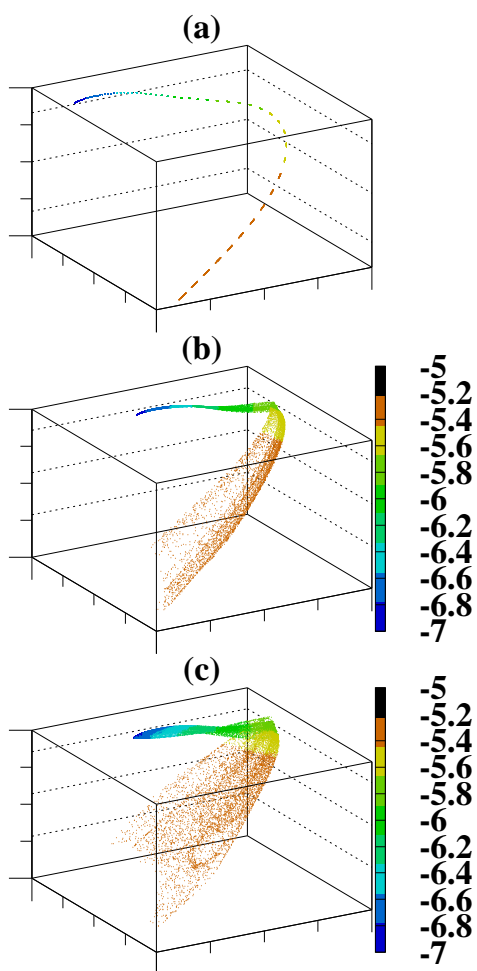
Figure 4: 4D projections of the manifolds in $\mathcal{C}$ for turbulent intensity cases 1–3 in Table 2. The axes have been left unlabeled intentionally, but represent the $log(concentration)$ of three chemical species. A fourth species is mapped as a color.
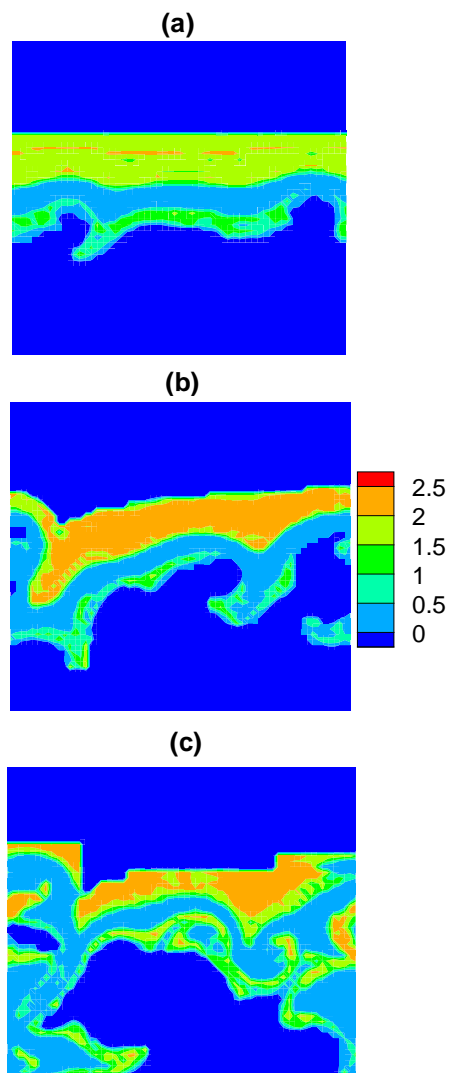
Figure 5: Visualization of $\mathcal{D}_3$ as a spatial field for turbulent intensity cases 2–4 in Table 2. The figure is constructed by plotting the value of $\mathcal{D}_3$ of the hypercube to which the concentration in a CFD cell belongs, at the (x, y) position of that CFD cell, and is plotted for every cell.