A Data-Fitting Procedure For Chaotic Time Series

J. M. McDonough, S. Mukerji
Department of Mechanical Engineering
University of Kentucky, Lexington, KY 40506-0108.

and

S. Chung
Department of Chemical Engineering
University of Illinois, Urbana-Champaign, IL 61801-3792.

Keywords: Time series analysis, subgrid-scale turbulence models, chaotic dynamical systems.

DISCLAIMER

Abstract

In this paper we introduce data characterizations for fitting chaotic data to linear combinations of one-dimensional maps (say, of the unit interval) for use in subgrid-scale turbulence models. We test the efficacy of these characterizations on data generated by a chaotically-forced Burgers' equation and demonstrate very satisfactory results in terms of modeled time series, power spectra and delay maps.

1. Introduction

Techniques have recently been proposed by Hylin and McDonough [1] and McDonough [2] for constructing models of hydrodynamic turbulence based on discrete dynamical systems (cf. Collet and Eckmann [3]). This approach has been motivated in part by earlier successful modeling of transitional pipe flow by Sreenivasan and Ramshankar [4], by the demonstration by Pulliam and Vastano [5] that direct numerical simulation (DNS) of flow over an airfoil exhibits a bifurcation sequence similar to that of the logistic map (see Feigenbaum [6]), and the analogy drawn between a similar quadratic map and the Navier-Stokes (N.-S.) equations due to Frisch [7].

While several of the abovementioned studies merely sought to demonstrate that specific experimental or computational results could be quite adequately represented by a nonlinear algebraic map, the work presented in [1, 2] served to demonstrate the potential for constructing models of turbulent fluctuations to be used as subgrid-scale behavior in computational techniques similar to large-eddy simulation (LES). These models have been based on the simple representation

$$u^* = A\zeta M\,, \tag{1}$$

where u^* is a fluctuating quantity (e.g., a velocity component), A is an amplitude factor derived from isotropic turbulence theory (cf. Tennekes and Lumley [8]), ζ is an anisotropy

correction factor constructed via structure functions of high-pass filtered large-scale portions, $\overline{u}$, of the complete velocity u, and M is obtained from a nonlinear chaotic algebraic map.

It should be noted that u^* is very inexpensive to compute, and the details of formulas for A and ζ (see [1, 2]) are well founded in theory. On the other hand, little has been done to date to establish the appropriate map(s) M to be used in this formula, or more important, how candidate maps might be identified from experimental data. Exceptions are the recent work of Mengüç et al. [9] associated with the very specific phenomenon of soot volume fraction fluctuations in a turbulent ethylene diffusion flame, and limited theoretical analysis in [1].

The goal of the present study is to provide a systematic, and ultimately automatic, approach for determining the best fit of linear combinations of discrete dynamical systems to experimental and/or DNS data. We note at the outset that there have been numerous previous works devoted to curve fitting of time series of chaotic phenomena (cf. Casdagli and Eubank [10], and references therein). These have typically focused on short-term prediction based on past data, i.e., extrapolation of polynomial fits of chaotic time series, such as described in Hunter [11]. Our goal in the present case differs from this in that we are mainly interested in constructing maps that approximately exhibit the long-time stationary behavior, as quantified by certain statistical properties, of the data but in addition yield the same physical bifurcation sequence. In one sense, our requirements are less rigid than those of many of the previous studies such as in [11], but in another sense they are far more encompassing of the overall underlying physical phenomena.

In this paper we will study time series obtained from a chaotically-forced Burgers' equation and test various quantifications of these data for producing optimal fits to a series of modified logistic maps. We will concentrate on one particular time series obtained from a single computer solution of the governing equation in order to emphasize the steps needed to carry out the curve-fitting procedure. Our results show that a linear combination of logistic

maps very effectively models the temporal behavior of the Burgers' equation solution in the sense of a least squares fit of various quantifying features, and in the somewhat qualitative sense of apparent strong similarities between time series, power spectra and delay maps of modeled and actual (directly simulated) data. We note, however, that we have not addressed the bifurcation sequence problem in this initial paper.

The remainder of the paper is organized as follows. In Sec. 2 we introduce the tools for analysis of chaotic time series, and we briefly describe the model problem used to produce our "experimental" data. Section 3 is devoted to presentation and discussion of results from this modeling process, and we provide a summary and conclusions in a final section.

2. Data Analysis Techniques

In this section we will discuss the form of chaotic map to which we will attempt to fit data. We then describe the methods employed to do this, and end by introducing a model problem on which these techniques will be tested.

2.1. Linear combinations of chaotic maps

The general form of the linear combination of chaotic maps which will be used in data fitting is constructed as

$$M^* = \sum_{k=1}^{K} \alpha_k \, S_k^{(n+1)}(\omega_k, d_k, m_k(b_k)) \,, \tag{2}$$

where the superscript $(n+1)$ indicates an advanced iteration level, and the index k is used to distinguish between different maps in the combination. The final value of the map combination at the advanced iteration is constructed by adding a history term

$$M^{(n+1)} = \theta M^* + (1-\theta) M^{(n)} \,. \tag{3}$$

Here θ is an implicitness factor which determines the weight assigned to the history term.

In equation (2) α_k is the amplitude, and each m_k is a (usually) normalized chaotic map which depends on a bifurcation parameter b_k. In general, the m_ks could all be different maps, but in the present study the modified logistic map,

$$m_k^{(n+1)} = b_k m_k^{(n)} \left(1 - \left|m_k^{(n)}\right|\right) , \tag{4}$$

will always be employed. Figure 1 presents the bifurcation diagram for this map for positive b_k values. A mirror image of this diagram is obtained for negative b_k values.

The coefficient $S_k^{(n+1)}$ in equation (2) takes on only two values, 0 or 1, and it determines if the k^{th} map makes any contribution to the map combination at the $(n+1)^{\text{th}}$ iteration. It depends on the frequency of evaluation, ω_k, and the duration, d_k, over which each map remains active once it is switched on. This coefficient depends on a switching function, $F_k^{(n)}$, and a duration function, $D_k^{(n)}$. The first of these is defined as

$$F_k^{(n)} \equiv \begin{cases} 1 & \text{if } (n \bmod \omega_k) = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

This determines whether the k^{th} map is to be activated at the n^{th} iteration. The duration function is defined as

$$D_k^{(n)} \equiv \begin{cases} 1 & \text{if } F_k^{(n)} = 1, \\ D_k^{(n-1)} + 1 & \text{if } F_k^{(n)} = 0 \text{ and } 0 < D_k^{(n-1)} < d_k, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

This function determines how long a map remains active. It is initialized to 1 when the map is activated according to equation (5), is incremented by 1 each time the map is evaluated and is set to 0 if it exceeds d_k. The switching function, $F_k^{(n)}$, is evaluated at every iteration independent of $D_k^{(n)}$. Whenever $F_k^{(n)}$ is unity $D_k^{(n)}$ is initialized to 1 and the k^{th} map is (re-)activated. This happens irrespective of the current value of $D_k^{(n)}$. Based on these

function definitions, the coefficient, $S_k^{(n)}$, is computed as

$$S_k^{(n)} = \begin{cases} m_k^{(n-1)} & \text{if } D_k^{(n)} = 0, \\ m_k^{(n)} & \text{otherwise.} \end{cases} \tag{7}$$

We comment that the representation (2) is somewhat suggestive of a Fourier, or a wavelet, decomposition with the m_ks serving as basis functions. Furthermore, it should be noted that not only are α_k, ω_k, d_k and b_k unknown parameters, but in actuality not even the m_ks or K are known. All of these must be determined as part of the data-fitting process, and as mentioned earlier our goal is to make this process as systematic and automatic as possible. The approach to doing this will be described in the next subsection.

2.2. Approach to data fitting

In constructing maps of the form (2) for arbitrary time series we must first establish a means by which a good fit can be distinguished from a bad one, and this in turn requires quantification of various properties associated with the time series. We wish to again emphasize that our goal here is not to produce a discrete dynamical system whose time series exactly coincides with the data for some period of time, and then ultimately deviates, as is typically the case in most studies contained in [10]. Rather, we will be attempting to construct chaotic maps possessing statistics and qualitative "appearance" as close as possible to those of the experimental data. We comment that in the context of the intended use of these maps (construction of subgrid-scale turbulence models) this is the more appropriate approach because the experimental data themselves would not be reproducible in detail – only their statistics can be expected to be relatively invariant from one realization to the next. We are thus faced with the task of statistically quantifying time series of modeled and measured dynamical systems, and then changing model details so the statistics coincide with those of measured results, but at the same time reproducing the qualitative appearance of the data.

We begin by noting that there are several easily constructed figures that can aid in comparing dynamical behavior of time series. We present examples of these in Figure 2 to aid in motivating our analysis approach. These were generated from the modified logistic map forced Burgers' equation to be discussed later. Figure 2 depicts the time series corresponding to 4096 time steps, while the inset displays a limited portion of this time series. Figures 3a and 3b show the power spectrum and two-dimensional delay map (with delay=4), respectively.

The last of these seldom provides any more than a qualitative indication of topology of a dynamical system, but it is a simple (though not completely fool-proof) indicator of deterministic, as opposed to random, behavior. The power spectrum can provide important quantitative information on the frequency distribution in both periodic and aperiodic phenomena and thus, the physical time scales represented in the data, but in the present case there appears to be only limited useful information beyond the general feature of broad-band noise and rate of decay of power with increasing frequency, with a hint of "noisy periodicity" at low frequencies.

The time series in this case contains a significant amount of information. One of the first things we can estimate directly is the number of terms one might use in the model, i.e., the value of K. In particular, since each term in (2) has an associated amplitude, we can estimate the required number of terms by looking for "structures" of various sizes in the time series. We have indicated several of these in Figure 2. In addition, the "size" of each such structure (i.e., amplitude of peak-to-peak oscillation) provides an estimate of the corresponding α_k. Finally, the frequency of occurrence of structures of various sizes allows us to specify values of ω_k for each of these. It must be emphasized, however, that all of the analysis to this point has been rather qualitative, and more quantitative measures of behavior are needed to provide a good fit.

All of the properties of the time series to be considered here are statistical in nature,

but some go beyond what are typically considered in usual data analysis. The need for more detailed characterizations arises from the chaotic nature of the data and the desire to retain "appearance" in modeled results. The most straightforward quantification is the time average, which for a discrete dynamical system is simply

$$\overline{u} = \frac{1}{N-1}\sum_{n=1}^{N} u^{(n)} \simeq \frac{1}{T}\int_0^T u(t)dt\,, \tag{8}$$

where N is the total number of discrete values of $u^{(n)}$. It should be clear from the nature of the time series shown in Figure 2 that $\overline{u}$ does not provide very definitive information. In particular, we would not be able to distinguish the complicated behavior shown in the figure from even a constant time series with a mean value equal to that constant if only the mean value were used.

A partial remedy for this is to consider the mean or average variation of the time series; that is, define the variation of u to be

$$V(u) \equiv \sum_{n=1}^{N-1} \left| u^{(n+1)} - u^{(n)} \right|, \tag{9}$$

and the average variation as

$$\overline{V}(u) = \frac{1}{N-1}\sum_{n=1}^{N-1} \left| u^{(n+1)} - u^{(n)} \right|. \tag{10}$$

This obviously provides some distinctions, and in particular a constant time series has $V(u) = \overline{V}(u) = 0$, whereas $V(u) > 0$ for all nonconstant time series.

This still is not nearly adequate for distinguishing between very complicated time series such as shown in Figure 2. A next step is to consider the separate averages of positive values of u, denoted $\overline{u}^+$ and of negative values, $\overline{u}^-$ given as

$$\overline{u}^+ = \frac{1}{N^+-1}\sum_{n=n_1^+}^{N^+} (u^+)^n\,, \tag{11}$$

and

$$\overline{u}^- = \frac{1}{N^--1}\sum_{n=n_1^-}^{N^-} (u^-)^n\,. \tag{12}$$

Associated with these average values are the times T^+, T^-(or N^+, N^- the number of discrete points) during which u is positive or negative. While matching of these quantities between modeled and measured results brings us a step closer to a unique specification of the α_k, ω_k, d_k and b_k in (2), it still is not sufficient. For example, it would be possible to construct a time series taking on only a few values and match all of the quantities discussed so far. We can refine the analysis further by monitoring u^+_{max}, u^+_{min}, u^-_{max} and u^-_{min} the maximum and minimum of positive and negative values, but this still is not sufficient.

Beyond the average variation we have as yet done nothing to require a matching of oscillation frequencies. There are a number of quantities that can be used in this regard. We first mention several properties derived from the power spectrum; these include maximum and minimum power, P_{max}, P_{min}, and the average decay rate of this power. But it must be noted that many different time series can produce the same power spectrum. Because of this in the present study we propose to use additional measures taken directly from the time series and not employ the power spectrum in the fitting procedure (but see Mengüç et al. [9]). These include the following: number of zero crossings (possibly per unit time), N_0, number of positive mean crossings, N_{mean+} and number of negative mean crossings, N_{mean-}. In addition, we will employ separate positive and negative variations. To obtain these we define the set $\mathcal{N}_{V+}$to be the set of indices n such that if $u^{(n)} > 0$, then $n \in \mathcal{N}_{V+}$ if and only if $u^{(n-1)} > 0$ and $u^{(n+1)} > 0$. We can then define the positive average variation of u as

$$\overline{V}^+(u) \equiv \frac{1}{N_{V+} - 1} \sum_{n \in \mathcal{N}_{V+}} \left| u^{(n+1)} - u^{(n)} \right| , \tag{13}$$

where N_{V+} is the number of indices in the set $\mathcal{N}_{V+}$. We can define the negative average variation in an analogous fashion:

$$\overline{V}^-(u) \equiv \frac{1}{N_{V-} - 1} \sum_{n \in \mathcal{N}_{V-}} \left| u^{(n+1)} - u^{(n)} \right| . \tag{14}$$

It should be clear that formulas (10), (13) and (14) provide measures of the average rate of oscillation of the time series. However, this can also be quantified by the number of

changes in sign of the slope between two successive points in the time series. We denote this by $N_{S\pm}$. It is more convenient to express this as a fraction of the total number of points, N:

$$f_{s\pm} \equiv N_{S\pm}/N\,. \tag{15}$$

Just as was done for variation of the time series, we can separately consider this fraction for the positive and negative parts of the time series. Thus we define

$$f^{+}_{s\pm} \equiv N^{+}_{S\pm}/N^{+}\,, \tag{16}$$

and

$$f^{-}_{s\pm} \equiv N^{-}_{S\pm}/N^{-}\,. \tag{17}$$

The final property of the time series to be considered here that directly involves detailed behavior is what we will term the intermittency distribution. Recall that the intermittency factor, or simply intermittency, at a point in a turbulent flow is the fraction of time during which the flow is turbulent at that point. Here we will define intermittency levels by first requiring that the segment of the time series under consideration be oscillatory, and then establishing the amplitude range as a fraction of the interval $[u_{min},u_{max}]$. The number of intermittency levels employed will be specified at the time analysis of a given data set is initiated, and can be expected to change somewhat from one time series to the next. The means by which this can be done is similar to the qualitative analysis of Figure 2, discussed earlier.

To check whether a given segment of data is oscillatory we carry out the following test. At the n^{th} discrete point we calculate the forward and backward differences,

$$\begin{aligned} \Delta_{+}u^{(n)} &\equiv u^{(n+1)} - u^{(n)}\,, \\ \Delta_{-}u^{(n)} &\equiv u^{(n)} - u^{(n-1)}\,. \end{aligned} \tag{18}$$

If either of the following inequalities is satisfied we consider the time series to be oscillatory

in a neighborhood of the n^{th} point:

$$\begin{aligned}(\Delta_+ u^{(n)})(\Delta_- u^{(n)}) < 0\,,\\ (\Delta_+ u^{(n)})(\Delta_- u^{(n-1)}) < 0\,.\end{aligned} \tag{19}$$

Figure 4 displays the types of oscillations that can be detected with these tests. We comment that the notions underlying equations (18,19) can be easily extended to treat larger numbers of points, but we feel that this would be necessary only for extremely highly resolved turbulent data.

Once it has been established that a point is part of an oscillation it is then possible to assign it to an intermittency level. We denote the intermittency levels by I_i and the end points of these levels by g_i. Then for a total of M intermittency levels $(I_1, \ldots, I_M)$ there are $M+1$ end points $(g_0, \ldots, g_M)$ with $g_0 = u_{min}$ and $g_M = u_{max}$. The intermittency levels can now be defined in terms of g_i's as:

$$\begin{aligned} I_1 &= [g_0, g_1) \\ I_2 &= [g_1, g_2) \\ &\vdots \\ I_M &= [g_{M-1}, g_M]\,. \end{aligned} \tag{20}$$

We will denote the number of points of the time series in an intermittency level i by N_{I_i}. We can then define the intermittency factors for each level as

$$f_{I_i} = N_{I_i}/N\,. \tag{21}$$

The choice of M and the spacing between the g_i's depends on the type of intermittency information that is critical to the analysis and has to be matched by the modeled time series. It will clearly be somewhat difficult to automate this portion of the analysis procedure, and this has not been attempted in the present study.

There are several things to note at this point. The first is that although the preceding properties appear to provide a fairly comprehensive quantification of a time series, they do

not specifically address periodic (or quasiperiodic) behavior. These cases are best handled via power spectral analysis. The second is that even in conjunction with power spectral analysis it is not clear that the above quantifications are sufficient (or whether all are necessary) to uniquely determine the α_k, b_k, d_k and ω_k in equation (2) – even if K and the form of the $m_k^{(n)}$ are specified. In addition, there are several other global properties that may be useful, possibly necessary, for production of accurate models. We discuss some of these in the following paragraphs.

A global mathematical quantity that is widely used to characterize the "size" of functions is the norm, $\|\cdot\|$. For physical phenomena for which we expect the energy to be finite for finite time, the L^2 norm is appropriate. For a discrete time series consisting of N (equally-spaced) values $\{u^{(n)}\}_{n=1}^N$, this can be approximated using

$$\|u\|_2 \simeq \left[\frac{T}{N-1}\sum_{n=1}^{N}\left|u^{(n)}\right|^2\right]^{1/2},$$

where T is the length of the "physical" time interval. The time average of this quantity is often more useful; this is given by

$$\frac{1}{T}\left(\int_0^T |u|^2\,dt\right)^{1/2} = \frac{1}{T}\|u\|_2 \simeq \frac{1}{\sqrt{(N-1)T}}\left(\sum_{n=1}^{N}\left|u^{(n)}\right|^2\right)^{1/2}. \tag{22}$$

From a purely mathematical standpoint, the L^1 norm,

$$\|u\|_1 \equiv \int_0^T |u|\,dt\,,$$

is sometimes more appropriate, and we calculate the time average of this as

$$\frac{1}{T}\int_0^T |u|\,dt = \frac{1}{T}\|u\|_1 \simeq \frac{1}{N-1}\sum_{n=1}^{N}\left|u^{(n)}\right|. \tag{23}$$

There are additional statistical properties that are widely used in studies of turbulence, and which may be appropriate quantifiers of the time series we study here. These include the autocorrelation function, the probability density function (pdf), the flatness and the skewness. Each of these quantities can be constructed for various derived properties of the

time series, but we will not go into these details here and instead restrict attention to the basic time series, itself. The autocorrelation of $u(t)$ for any fixed location x is

$$C(u,\tau) = \frac{\langle u(t), u(t+\tau)\rangle}{\|u\|_2^2} = \frac{1}{\|u\|_2^2}\int_0^T u(t)u(t+\tau)dt\,.$$

For a discrete data set this can be approximated as

$$C(u,\tau_j) = \frac{1}{\|u\|_2^2}\sum_{n=1}^{N-j} u^{(n)}u^{(n+j)}\,, \tag{24}$$

where $\tau_j = j\Delta t$; Δt is the time difference between discrete data points, and j is a prescribed number of points.

The pdf provides information similar to, but actually less detailed than, the intermittency distribution described earlier. It is widely used in experimental studies since it demonstrates whether data have a Gaussian distribution. It is also used to construct the various moments of the data. Two quantities widely used in turbulence studies that can be obtained in this manner are the skewness (third moment) and flatness (fourth moment). For a velocity component these can also be directly calculated as

$$S = \frac{\overline{u^3}}{\left(\overline{u^2}\right)^{3/2}} = \frac{\frac{1}{N}\sum_{n=1}^{N}(u^{(n)})^3}{\left[\frac{1}{N}\sum_{n=1}^{N}(u^{(n)})^2\right]^{3/2}} \tag{25}$$

for skewness, and

$$F = \frac{\overline{u^4}}{\left(\overline{u^2}\right)^2} = \frac{\frac{1}{N}\sum_{n=1}^{N}(u^{(n)})^4}{\left[\frac{1}{N}\sum_{n=1}^{N}(u^{(n)})^2\right]^2} \tag{26}$$

for flatness (also known as kurtosis), without using the pdf.

We close this section on tools employed in the analysis of time series by mentioning several more modern techniques developed specifically in the context of dynamical systems. These are the Lyapunov exponents, Kolmogorov entropy, correlation dimension and $f(\alpha)$ singularity spectrum. The first of these is considered to be the surest test for deterministic chaos because it shows the rate of divergence of nearby trajectories, and hence degree of

sensivity to initial conditions; but the extant computational algorithms are not completely reliable. To some extent this is true also for correlation dimension and entropy, particularly if the dimension happens to be fairly high, say approaching 10 or greater. We will not employ these approaches here.

The various measures we have discussed for quantifying data in the form of time series will be used in the model construction process. In particular, if we let p_i denote the numerical result of computing property i, and define

$$\delta p_i \equiv p_i^{meas} - p_i^{model}\,, \tag{27}$$

then we can define a weighted least-squares functional corresponding to N_p properties as

$$Q(\alpha, b, \omega, d) = \sum_{i=1}^{N_p} \phi_i (\delta p_i)^2\,, \tag{28}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)^T$, etc., and ϕ_i are the weights associated with different properties. We can then determine the unknown amplitudes, bifurcation parameters, frequencies and durations of map evaluations by minimizing Q. Conceptually this is straightforward, but in practice it is difficult in the present case because the δp_i are generally not differentiable with respect to b, ω and d. Thus, direct search techniques have been used in the present study. We comment that the use of neural networks or simulated annealing techniques could be very effective in these searches, and these will be investigated in future studies.

2.3. Burgers' equation model problem

In this section we introduce a model problem employed to generate chaotic time series with which to test the chaotic map construction process outlined above. The advantages of using results from a model problem are that they are completely controllable with respect to degree of complexity, and there are no measurement errors to raise issues regarding interpretation of results. Moreover, if effects of "experimental error" are to be considered, such errors can easily be introduced at known prescribed levels via random perturbations.

The model problem consists of the widely-studied Burgers' equation with a forcing term in the form of a nonlinear algebraic map. Complete details of the problem and solution algorithm can be found in Peng [12]; here we merely present the main equations and sketch the highlights of the numerical methods used in their solution.

The form of Burgers' equation to be employed here is

$$u_t + \frac{1}{2}(u^2)_x = -p_x + \nu u_{xx}\,, \quad x \in (0,1), \quad t \in (0,T] \tag{29}$$

with initial condition

$$u(x,0) = 1 + x\,, \quad x \in (0,1)\,, \tag{30}$$

and consistent boundary conditions

$$u(0,t) = 1, \quad u(1,t) = 2 \quad \forall\, t \in (0,T]\,. \tag{31}$$

In equation (29) subscripts denote partial differentiation, ν is analogous to kinematic viscosity, and p_x is a prescribed forcing function of the form

$$p_x(x,t) = -(1+x) + \gamma m(t)\,. \tag{32}$$

Here, γ is a given constant, and m is the modified logistic map, equation (4). The discrete form of (32) is

$$p_x(x_i, t^n) = -(1 + x_i) + \gamma m_i^{(n)}\,. \tag{33}$$

In the present study (29) has been discretized on a grid of 256 points using standard second-order centered differencing. Time integration is via the trapezoidal method for linear terms, and midpoint method for the nonlinear term. Care has been taken to use values of ν sufficiently large that the solution contains no spurious cell-Re oscillations, since these would be difficult to distinguish from the chaotic map oscillations. Details of all parameter values employed to obtain the results of this study will be given in the next section.

3. Results and Discussion

In this section we present results from applying the analysis techniques discussed in the preceding section to a numerical time series obtained as the approximate solution to equations (29-32). All results to be presented have been computed in double precision (64-bit) FORTRAN on the HP-Convex Exemplar SPP 1200 at the University of Kentucky Computing Center.

The time series we consider is displayed in Figure 5. This corresponds to the portion of the solution to (29) at $x \cong 0.502$. The calculations were performed with kinematic viscosity ν=0.01, γ=10 in equation (33), b=4.25 in equation (4), with 256 equally spaced grid points in the unit x-interval, and Δt=0.01 for the time integration. A total of 4096 time steps were computed to guarantee a sufficiently long stationary state to permit reliable analyses of the sort being performed here. The previously discussed measures for quantifying time series data, were computed for this time series and are shown in the third column of Table 2. These measures include 10 different autocorrelations computed via equation (24) with $j = 1$ through $j = 10$, and also intermittencies from equation (21) in 10 equally spaced intervals, i.e. $M = 10$ in equation (20). The corresponding power spectrum and delay map (delay=4) are presented in Figure 8. The histograms of the intermittency distributions and autocorrelations are shown in Figure 10. Quite similar results are obtained at all other x locations in (0,1), and only those shown in Figure 5 will be analyzed.

As was the case for a similar time series displayed in Figure 2, the present one shows evidence of at least three different scales of structure in both space and time, suggesting that initially we should attempt a fit with three terms (K=3) in the linear combination (2). We have also chosen to initially take all b_k=4.25, k=1,2,3. It is then necessary to determine the amplitudes α_k, evaluation frequencies ω_k, duration intervals d_k and the implicitness factor θ. Figure 6 displays a first attempt at this with α_1=1.0, α_2=1.0, α_3=1.0, ω_1=1, ω_2=1, ω_3=1,

$d_1 = 1$, $d_2 = 1$, $d_3 = 1$ and $\theta = 1.0$. To show the details of the oscillations, the time scale in Figure 6 is only a tenth of that of the computed result shown in Figure 5. It is obvious that the frequency of oscillations in the present case is very high, and these results are not accurate even in a qualitative sense; it is clear that more detailed analyses are needed.

The first step in the detailed analyses is to set up a parameter space by choosing a wide range of combinations of values of α_k, b_k, d_k and ω_k, $k = 1, 2, 3$. Each chaotic map combination is characterized by 12 parameters which are selected as unique permutations of the values in the parameter space. The time series of the chaotic map combination is generated for each of these permutations using equations (2) and (3). The same quantifying statistics that were calculated for the Burgers' equation data are now computed for each model time series, and the least-squares functional, equation (28), is then constructed for each case. The computations for each case are independent of all others and are done in parallel. The cases with the smallest value of the functional, Q, are considered to be the 'best' cases and ten of these cases are stored in each run. These results are used to manipulate the parameter space for the subsequent runs. Initially, coarse-grained searches are made on slightly overlapping segments of the parameter space. Once the best segment has been identified, fine-grained searches are made on smaller subsets of this segment. This is repeated until a minimum value of Q, Q_{min}, is obtained.

The direct search was carried out according to the strategy described above. The map parameters corresponding to the best case are given in Table 1. The implicitness factor, θ, was not a part of the direct search but was manually modified as the search progressed, and its optimal value is also shown in Table 1. The search procedure produced $Q_{min} = 3.766 \times 10^{-6}$. To put this into context, we note that the maximum value was $Q_{max} = \mathcal{O}(10^{-2})$. The time series for the best case is shown in Figure 7; the corresponding power spectrum and delay map (delay=4) are shown in Figure 9, and the histograms of the intermittency distributions and autocorrelations are presented in Figure 11. These results have to be compared with

those obtained from the Burgers' equation solution, Figures 5, 8 & 10 respectively.

From the experience gained during this research, it appears that the skewness and flatness properties are not important in obtaining a good match between modeled and computed data in the present case. Hence these properties were not calculated. Also, it was found that the intermittency factors and autocorrelations were critical for a good data fit. The normalized weights shown in fifth column of Table 2 reflect this, with heavier weights assigned to these properties.

The time series, (Figures 5,7), are in excellent qualitative agreement. The three scales of "structures" described in Figure 2 are replicated in detail in the modeled time series. The distinct amplitudes α_k in Table 1 give quantitative information about the relative sizes of the different scales. The two power spectra, (Figures 8a,9a), agree well in terms of power levels and decay rates, despite the fact that no explicit attempt was made to match them between model and the direct solution. The power levels at higher frequencies appear to be lower in the modeled case possibly because of additional dissipation introduced by using sums of maps (see [1] for a discussion of this effect). Evidence of this extra dissipation can also be detected in the delay maps, (Figures 8b, 9b), where the modeled delay map appears to be slightly more tightly wound than the map for the direct solution data. The general agreement in the intermittency distribution, (Figures 10a,11a), and the autocorrelations, (Figures 10b,11b), is also very good. There is generally good quantitative agreement too, as can be seen from Table 2 where the results for the various statistical properties discussed earlier are presented for both the modeled and directly computed cases.

4. Summary and Conclusions

In this paper we have presented an approach to fitting experimental and/or DNS data to a linear combination of chaotic maps. We have emphasized the goal of finding sufficient

statistical characterizations of data to essentially guarantee that the model exhibits qualitative features in good agreement with the data and at the same time provides satisfactory matches (in a quantitative sense) of important statistical properties. We conjecture that this will allow phase space trajectories of the model time series to be associated with an attractor similar to that of the data, at least in a spatially local sense. To provide tight control of the overall exercise we have employed a DNS time series obtained by solving a chaotically-forced Burgers' equation in lieu of actual experimental data. Application to experimental data is reported elsewhere (see Mengüç et al. [9]).

We have described a long list of data characterizations and demonstrated specific effects on the modeling process of the more important of these. Results obtained for the specific case considered in this work are generally quite satisfactory as is evident from the data presented in Sec. 3., thus indicating the feasibility of employing linear combinations of a single discrete dynamical system to match the long time stationary features of the evolution of an infinite-dimensional dynamical system (albeit, one probably possessing a finite-dimensional global attractor, or at least inertial manifold; cf. Temam [13] and Constantin et al. [14]). The intrinsic importance of this type of result is that, in fact, it is believed that the situation is quite analogous to the Navier-Stokes equations, so our results should indeed be valuable in utilizing both experimental and DNS data for chaotic map construction in the context of the Hylin and McDonough [1] additive turbulent decomposition subgrid-scale model formalism represented in equation (1).

Despite the success of our fitting process, it must be emphasized that only a single data set has been employed here, and several issues have not been addressed. With regard to the intended use of these data fits in subgrid-scale models such as equation (1), the key one of these is replication of bifurcation sequences. Clearly this is a crucial aspect of models of the form of (1), but we have not considered it here. (We remark, however, that this has been studied in a limited way in Mengüç et al. [9], and in a related work by McDonough et al.

[15].) A second issue is that of establishing necessary and sufficient conditions (i.e., set of statistical properties) to guarantee a reliable fit of data. This was one of our stated goals of the present work, and we must conclude that it has been only partially accomplished. In particular, our results imply sufficiency of the set of statistical parameters employed, but we have not unequivocally demonstrated their necessity. In fact, we did find that two properties, skewness and flatness, were not necessary; we conjecture that they possibly represent a degree of redundancy in light of the fact that our intermittency distribution is closely related to a pdf from which skewness and flatness can be extracted. Moreover, while autocorrelation proved to be quite important in fitting the data herein, it was insignificant in the context of a particular laboratory data set considered by Mengüç et al. [9] and McDonough et al. [15]. We must evidently conclude that the specific properties needed to adequately fit a given data set may be data dependent, and thus probably must be determined as part of the data fitting process.

Finally, we wish to comment on a few general related issues. First, we have restricted attention to linear combinations of a single chaotic map. While our approach appears to have been fairly successful, clearly linear combinations of different maps might be considered, and in fact this may be necessary to induce quasiperiodic behavior (which does not occur in our present model). Second, the next step up in complexity would be to consider nonlinear compositions of different maps, and then linear or even nonlinear combinations of these. Third, all of the chaotic maps considered to date have been one-dimensional. An open question is whether multi-dimensional versions of equation (1), and, correspondingly, the maps M might be more appropriate. Finally, as we hinted earlier, the fitting process itself could probably be made significantly more efficient by use of neural networks and/or simulated annealing. One of our ultimate goals for the use of chaotic maps is in real-time control of physical systems, and neural networks are already seeing applications in this area.

It is clear that many interesting problems and applications associated with the analysis

technique reported here remain, and we hope these initial results will provide a first step in fitting chaotic data in a global sense, i.e., in terms of the underlying attractor, instead of individual trajectories as has been considered in numerous earlier investigations (cf. [10]), and stimulate further studies.

Acknowledgements- The first two authors gratefully acknowledge the support of this work under DOE Grant #DE-FG22-93PC93210. The first author was also partially supported under NASA/EPSCOR Grant WKU #522761-95-04, while the second author received partial support from the University of Kentucky Center for Computational Sciences. The third author was funded under the NSF REU Program.

References

[1] E.C. Hylin, and J.M. McDonough, Theoretical development of a stochastic model for small-scale turbulence in an additive decomposition of the Navier-Stokes equations. Mechanical Engineering Report No. CFD-02-96, University of Kentucky, Lexington, KY, 1996. Submitted to *J. of Fluid Mech.*.

[2] J.M. McDonough, Parallelization of ATD/chaotic-map turbulence models, in *Parallel Computational Fluid Dynamics: Algorithms and Results Using Advanced Computers.* (P. Schiano et al. Eds.), Elsevier Science B.V. Amsterdam, 1997.

[3] P. Collet, and J.P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems.* Birkhäuser-Boston, Boston, 1980.

[4] K.R. Sreenivasan, and R. Ramshankar, Transition intermittency in open flows, and intermittency routes to chaos. *Physica* 23D:246-258 (1986).

[5] T.H. Pulliam, and J.A. Vastano, Transition to chaos in an open unforced 2D flow. *J. Comput. Phys.* 105:133-149 (1993).

[6] M.J. Feigenbaum, Quantitative universality for a class of nonlinear transformations. *J. of Stat. Phys.* 19:25-52 (1978).

[7] U. Frisch, *Turbulence.* Cambridge University Press, Cambridge, 1995.

[8] H. Tennekes, and J.L. Lumley, *A First Course in Turbulence.* MIT Press, Cambridge, MA, 1983.

[9] M.P. Mengüç, J.M. McDonough, S. Manickavasagam, S. Mukerji, and S. Ghosal, Fluctuations of soot in turbulent diffusion flames: Experimental data and chaotic map models. Presented at the International Mechanical Engineering Congress and Exposition (IMECE) in Atlanta, November 1996. Submitted to *ASME J. of Heat Transfer.*

[10] M. Casdagli, and S. Eubank, (Eds.), *Nonlinear Modeling and Forecasting.* Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XII, Addison-Wesley, 1992.

[11] N.F. Hunter, Application of nonlinear time-series models to driven systems, in *Nonlinear Modeling and Forecasting.* Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XII, (M. Casdagli and S. Eubank, Eds.), Addison-Wesley, 1992.

[12] J.-S. Peng, An Analysis of Reynolds Averaging via One-dimensional Burgers' Equation. M.S. Thesis, University of California, Los Angeles, CA, 1990.

[13] R. Temam, *Infinite-dimensional Dynamical Systems in Mechanics and Physics.* Applied Mathematical Sciences, v. 68, Springer-Verlag, New York, 1988.

[14] P. Constantin, C. Foias, B. Nicolaenko, and R. Temam, *Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations.* Applied Mathematical Sciences, v. 70, Springer-Verlag, New York, 1989.

[15] J.M. McDonough, M.P. Mengüç, S. Mukerji, S. Manickavasagam, and S. Chung, Chaotic map models of soot fluctuations in turbulent diffusion flames. To be submitted to *Int. J. of Heat and Mass Transfer* (1997).

List of Figures

List of Tables

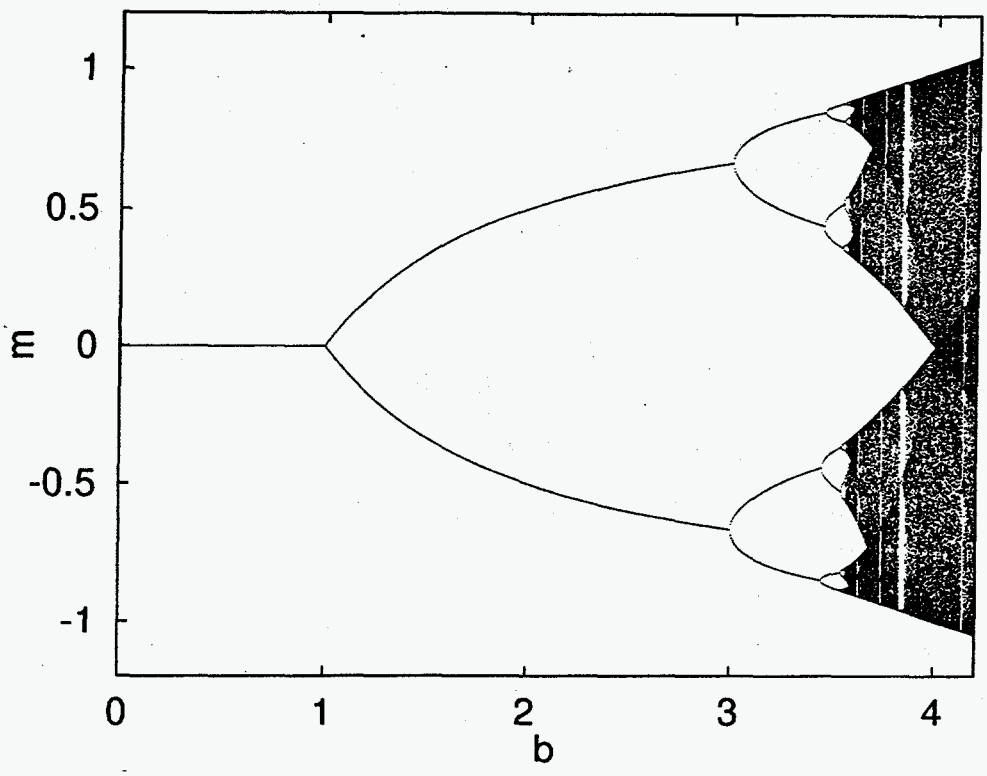


FIG. 1. Bifurcation diagram of the modified logistic map.

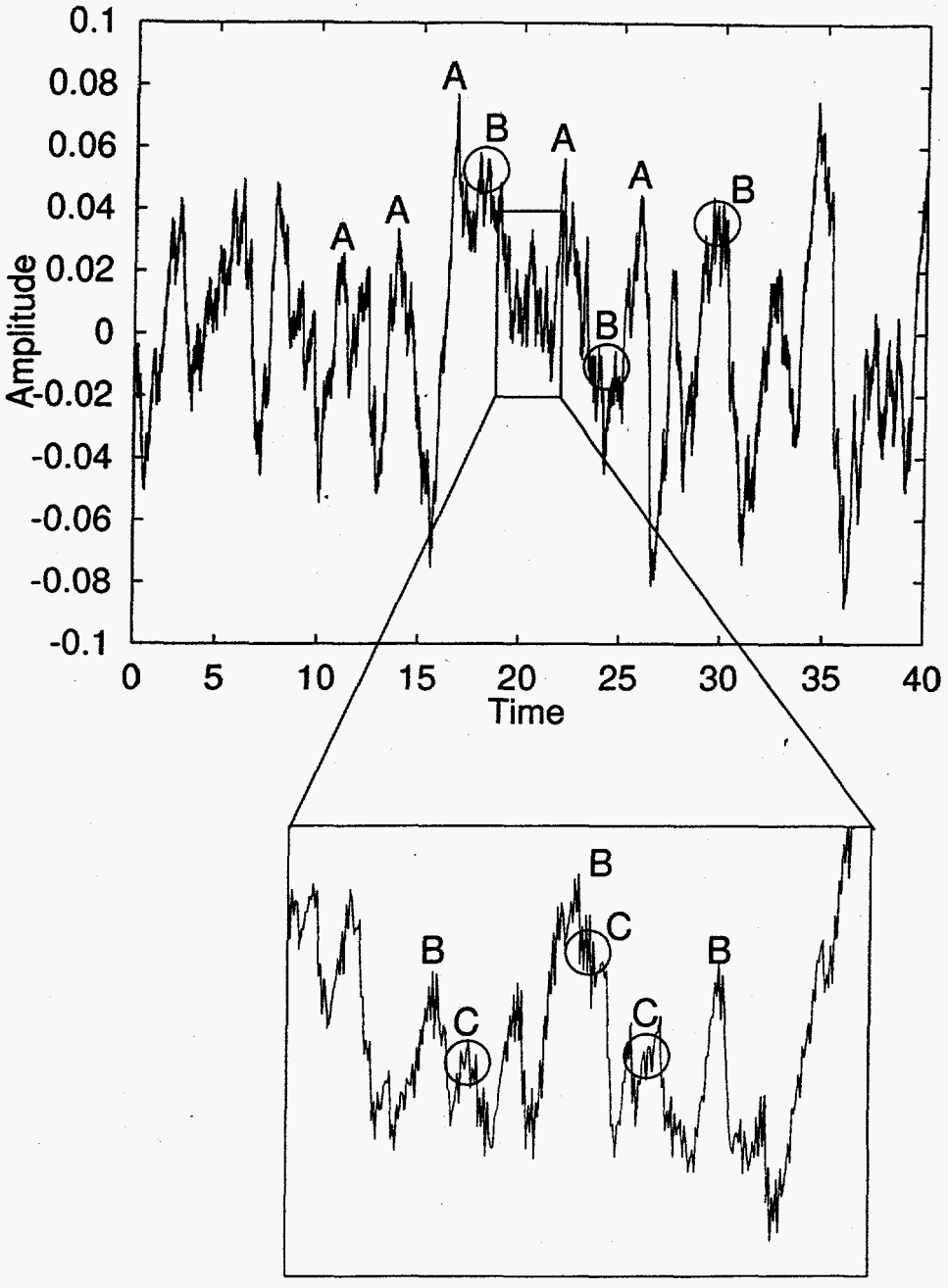
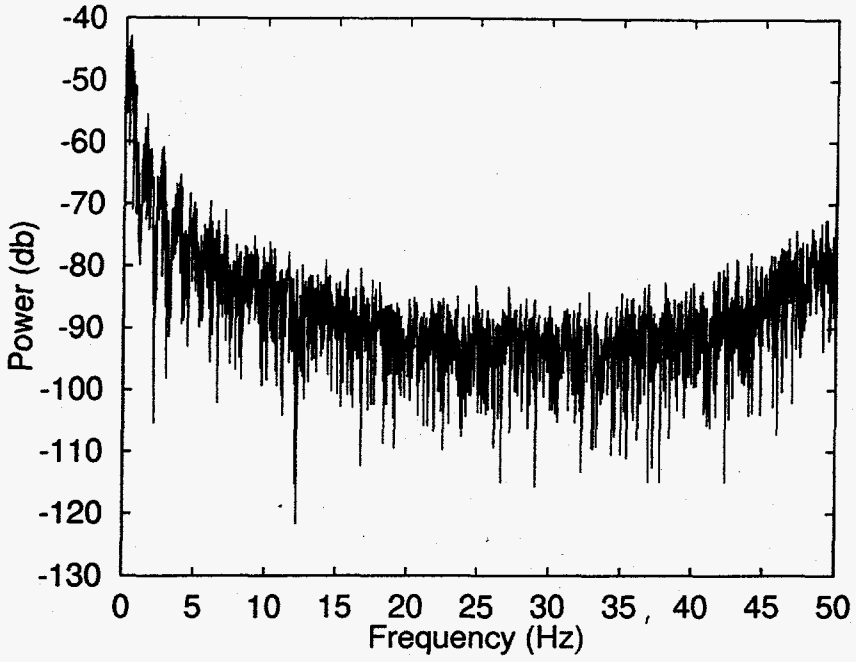
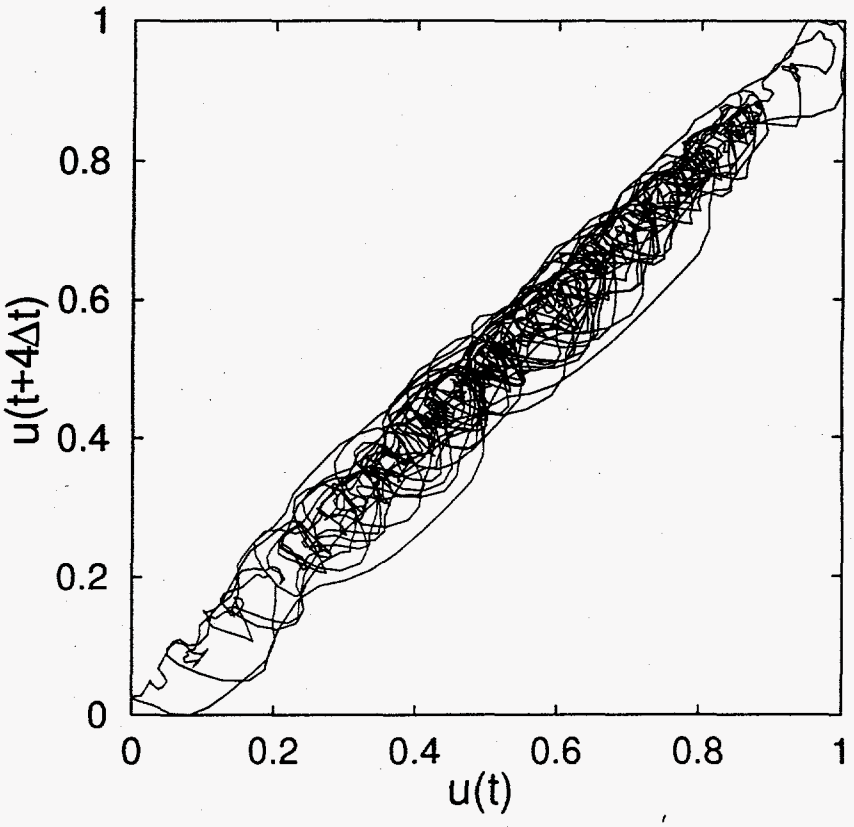


FIG. 2. Sample time series from Burgers' equation driven by the modified logistic map. Large-scale features are indicated by A; superposed on these are smaller fluctuations labeled B. Inset shows details of these fluctuations and evidence of yet finer structures denoted by C.



(a)



(b)

FIG. 3. Sample results from Burgers' equation driven by the modified logistic map. (a) Power spectrum. (b) Delay map (delay=4).

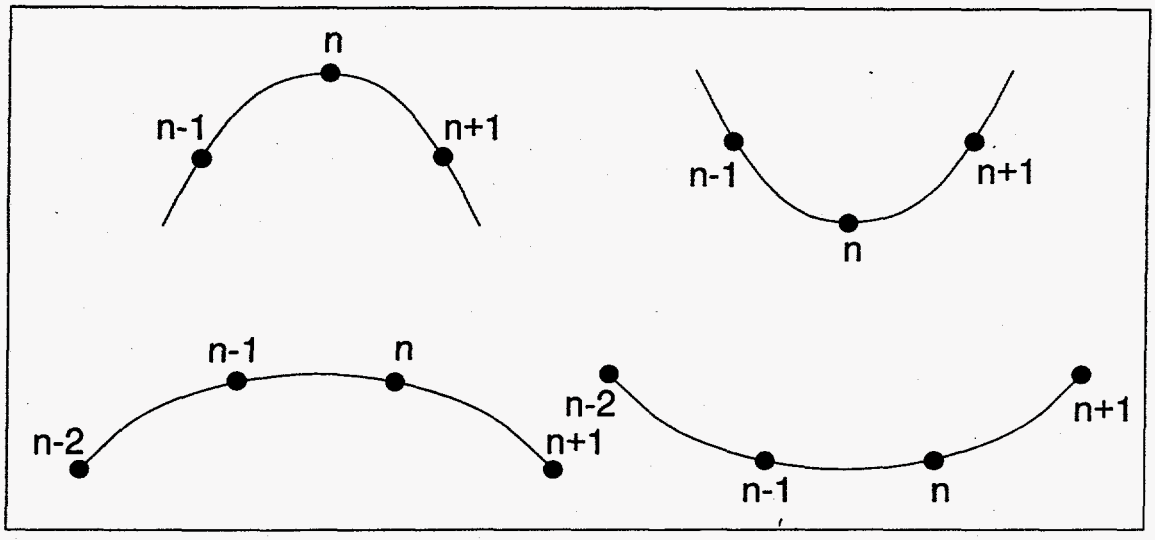


FIG. 4. Oscillation types contributing to the intermittency distribution.

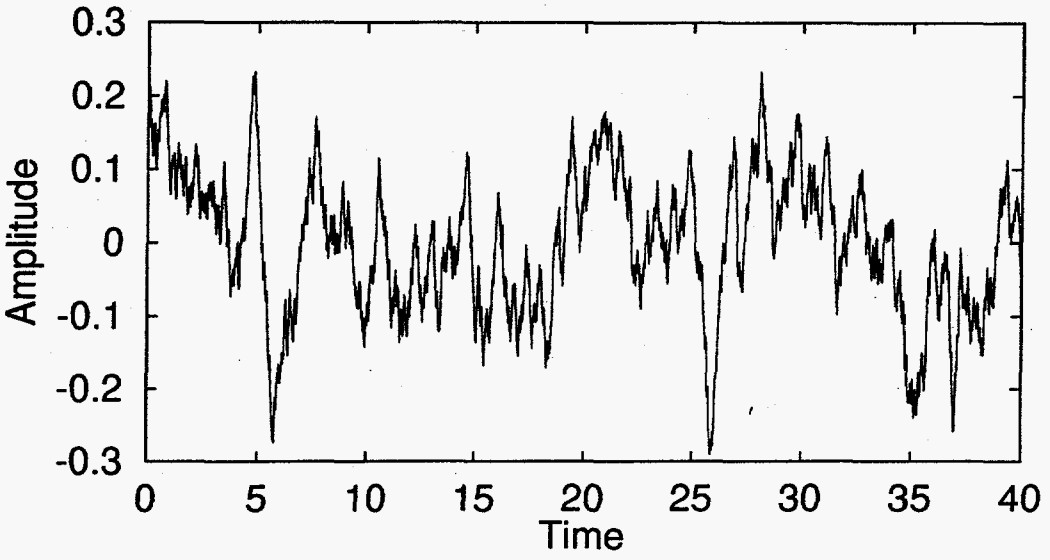


FIG. 5. Time series from Burgers' equation driven by the modified logistic map. Map data are fit to this time series.

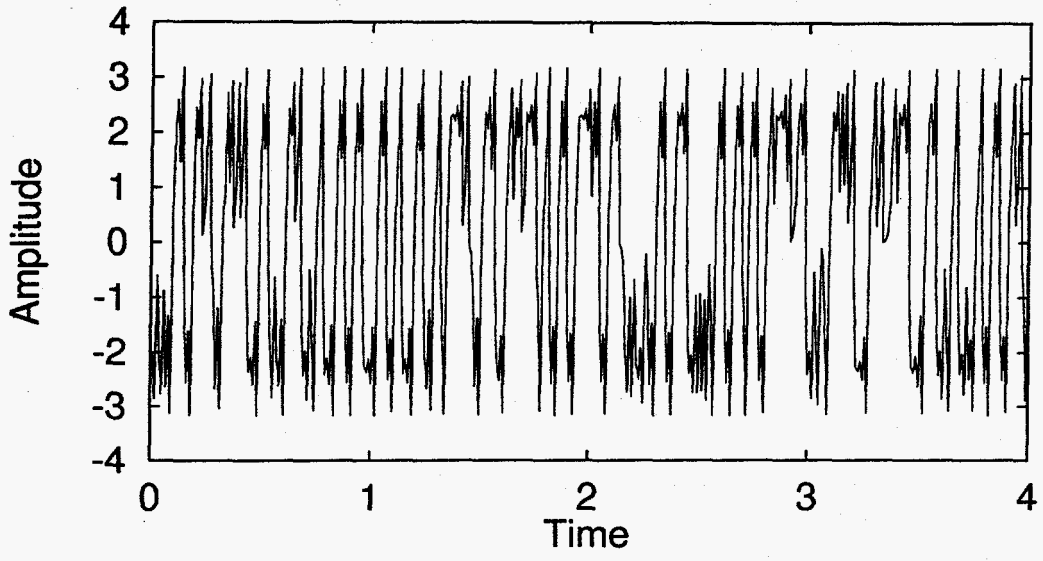


FIG. 6. Time series from the linear combination of logistic maps, first attempt.

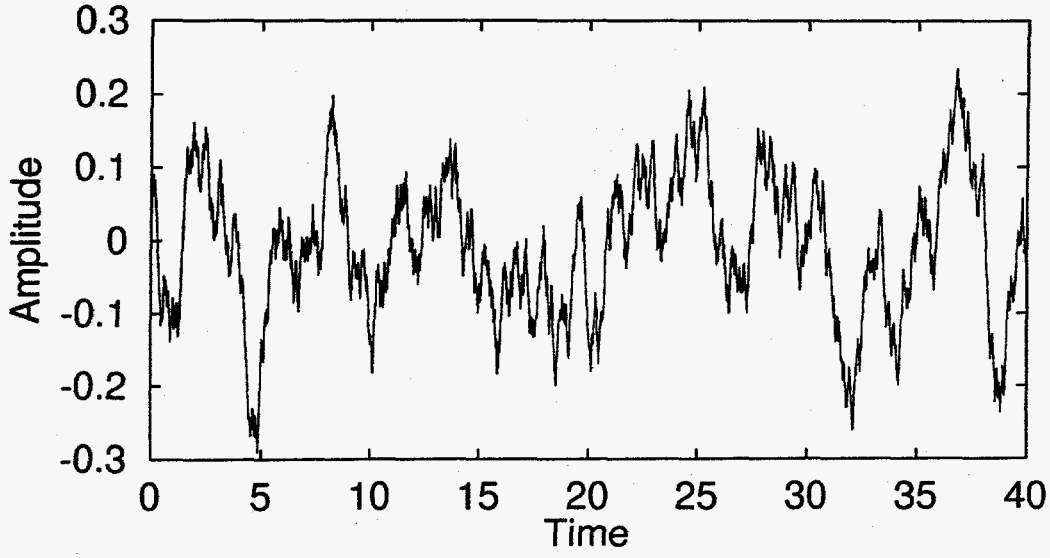
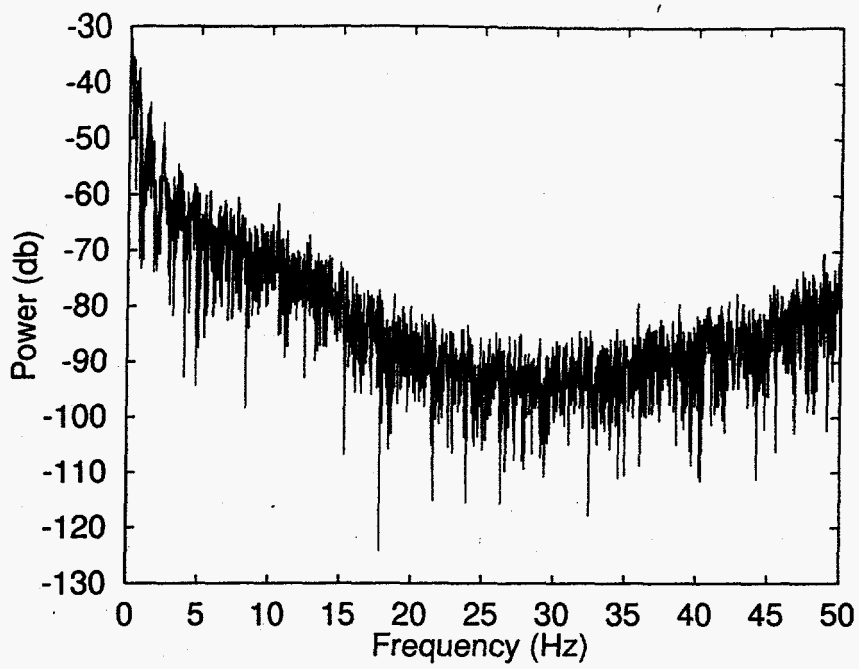
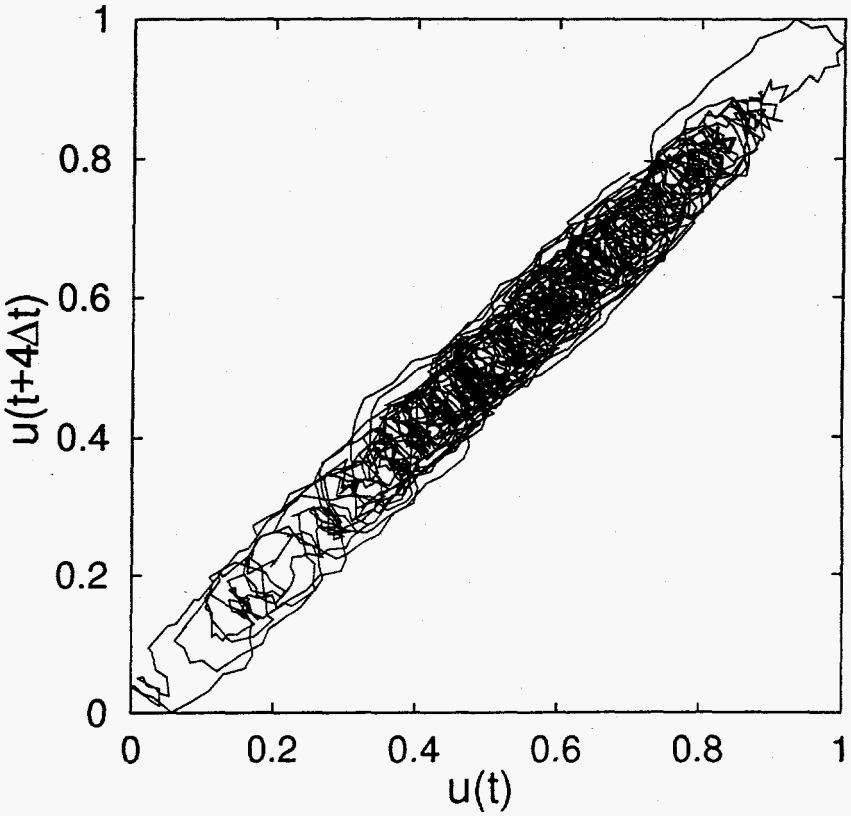


FIG. 7. Time series from the combination of chaotic maps, best case.

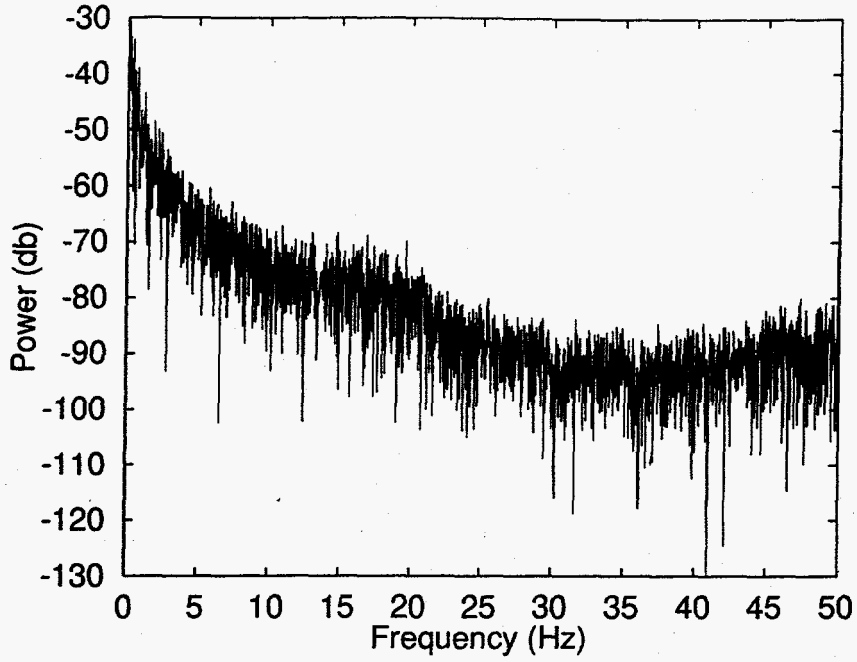


(a)

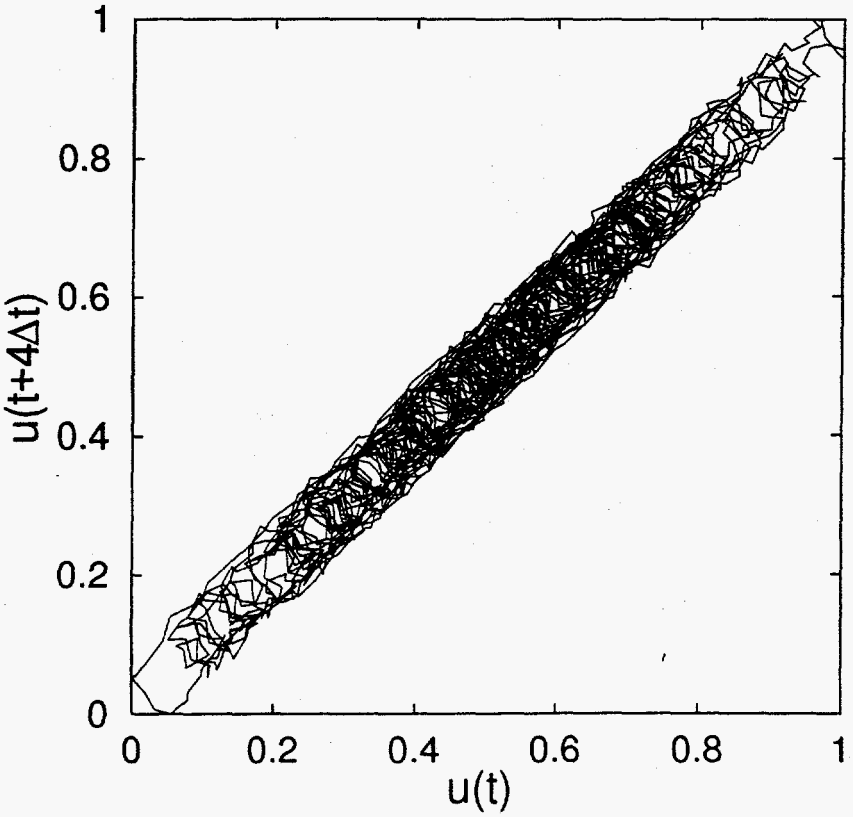


(b)

FIG. 8. Results from Burgers' equation driven by the modified logistic map. (a) Power spectrum. (b) Delay map (delay=4).

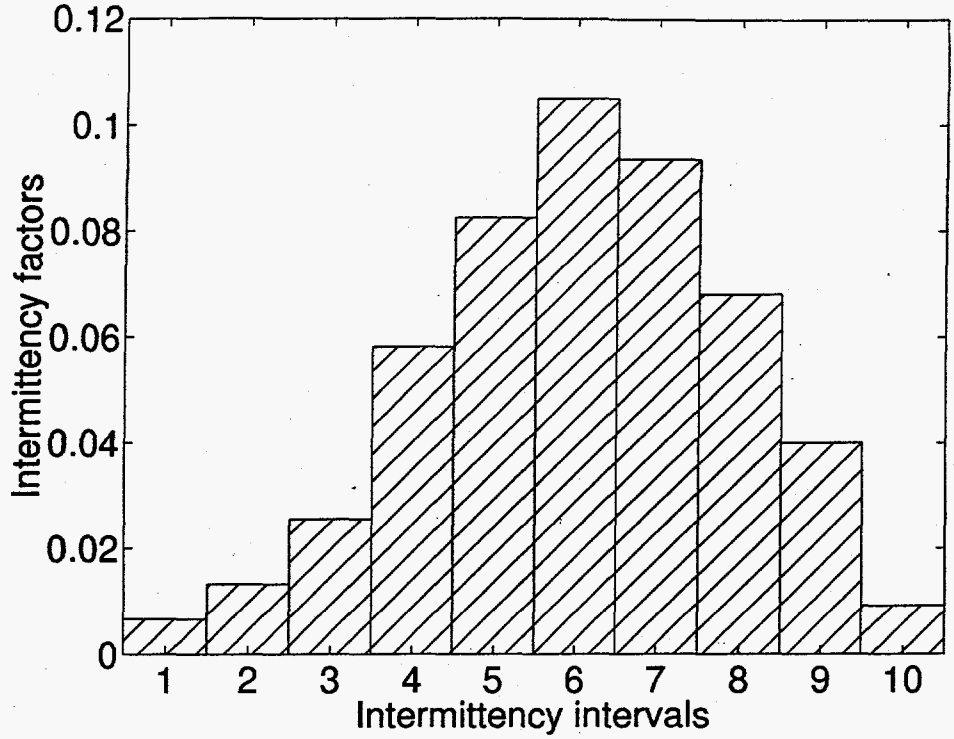


(a)

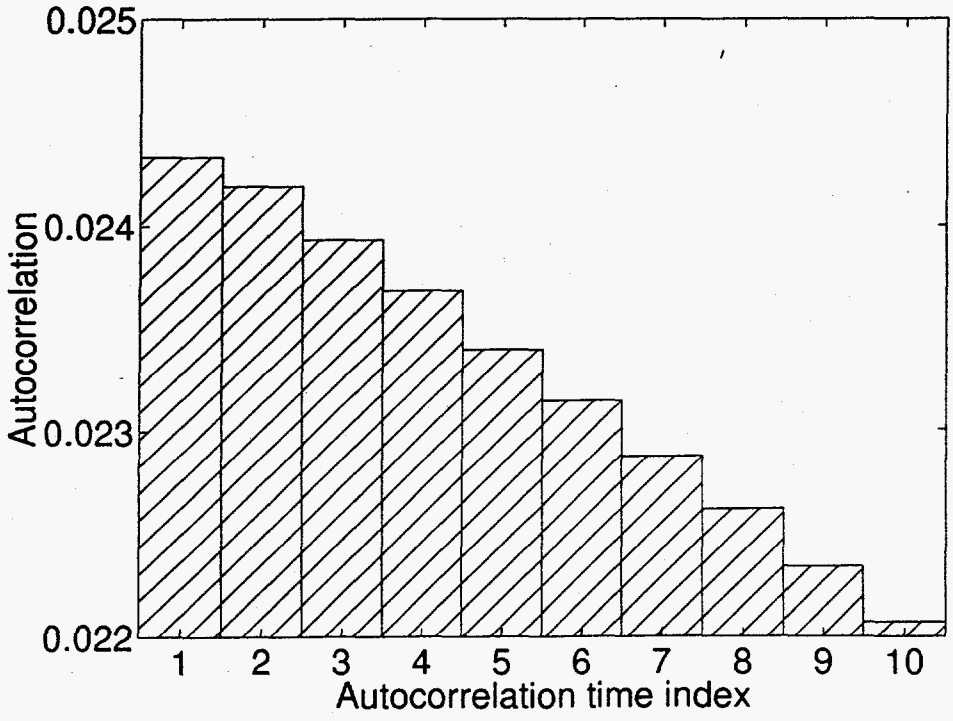


(b)

FIG. 9. Results from the linear combination of chaotic maps, best case. (a) Power spectrum. (b) Delay map (delay=4).

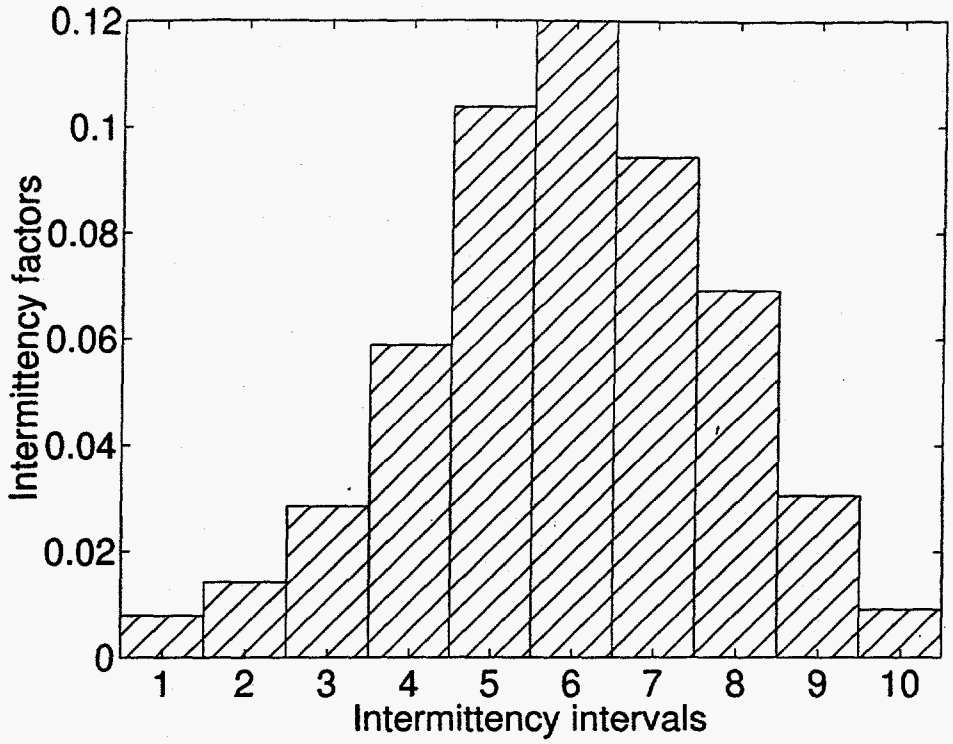


(a)

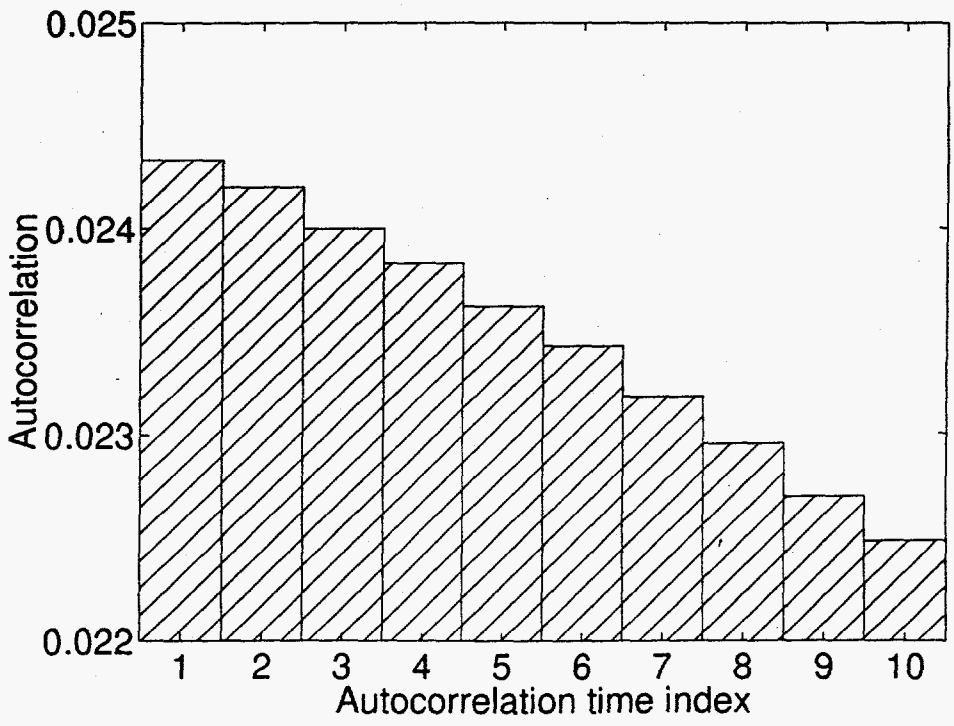


(b)

FIG. 10. Results from Burgers' equation driven by the modified logistic map. (a) Intermittency distribution. (b) Autocorrelations.



(a)



(b)

FIG. 11. Results from the linear combination of chaotic maps, best case. (a) Intermittency distribution. (b) Autocorrelations.

Map, k	α_k	b_k	ω_k	d_k
1	7.2735 [-1]	-3.86410	1	1
2	3.3730 [-4]	-3.94434	4	2
3	1.3650 [0]	4.16815	$1'$	1

$\theta = 0.00801$

Table 1. Map parameters for the best case.

Property No. i	Description	Experimental Value p_i^{meas}	Modeled Value p_i^{model}	Weights ϕ_i	Contribution To Q $\phi_i(\delta p_i)^2$
1	Average ($\overline{u}$)	-2.841 [-4]	-8.553 [-3]	1.547 [-3]	1.058 [-7]
2	Positive average ($\overline{u^+}$)	7.510 [-2]	7.099 [-2]	1.547 [-3]	2.602 [-8]
3	Negative average ($\overline{u^-}$)	-7.998 [-2]	-7.888 [-2]	1.547 [-3]	1.901 [-9]
4	Average variation ($\overline{V}(u)$)	7.076 [-3]	6.680 [-3]	1.547 [-3]	2.420 [-10]
5	Positive average variation ($\overline{V^+}(u)$)	6.916 [-3]	6.607 [-3]	1.547 [-3]	1.482 [-10]
6	Negative average variation ($\overline{V^-}(u)$)	7.174 [-3]	6.666 [-3]	1.547 [-3]	3.991 [-10]
7	Slope sign changes, total ($f_{s\pm}$)	3.556 [-1]	3.527 [-1]	1.547 [-3]	1.329 [-8]
8	Slope sign changes, u positive ($f^+_{s\pm}$)	3.660 [-1]	3.571 [-1]	1.547 [-3]	1.215 [-7]
9	Slope sign changes, u negative ($f^-_{s\pm}$)	3.447 [-1]	3.488 [-1]	1.547 [-3]	2.603 [-8]
10	Time average L^1 norm	7.745 [-2]	7.516 [-2]	1.547 [-3]	8.137 [-9]
11	Time average L^2 norm	1.500 [-2]	1.474 [-2]	1.547 [-3]	9.752 [-11]
12	Overall minimum (u_{min})	-2.909 [-1]	-2.908 [-1]	1.547 [-3]	3.078 [-11]
13	Overall maximum (u_{max})	2.334 [-1]	2.337 [-1]	1.547 [-3]	1.533 [-10]
14	Positive minimum (u^+_{min})	4.067 [-5]	5.451 [-5]	1.547 [-3]	2.961 [-13]
15	Positive maximum (u^+_{max})	2.334 [-1]	2.337 [-1]	1.547 [-3]	1.533 [-10]
16	Negative minimum (u^-_{min})	-2.909 [-1]	-2.908 [-1]	1.547 [-3]	3.078 [-11]
17	Negative maximum (u^-_{max})	0.000 [0]	0.000 [0]	1.547 [-3]	0.000 [0]
18	Number of zero crossings/time (N_0/N)	2.833 [-2]	2.809 [-2]	1.547 [-3]	9.227 [-11]
19	Number of $\overline{u}$ crossings/time (N_0/N)	2.833 [-2]	2.809 [-2]	1.547 [-3]	9.227 [-11]
20	Number of $\overline{u^+}$ crossings/time (N_{mean+}/N)	2.247 [-2]	2.149 [-2]	1.547 [-3]	1.476 [-9]
21	Number of $\overline{u^-}$ crossings/time (N_{mean-}/N)	2.394 [-2]	1.954 [-2]	1.547 [-3]	2.990 [-8]
22	Autocorrelation ($C(u, \Delta t)$)	2.434 [-2]	2.433 [-2]	2.612 [-2]	3.025 [-13]
23	Autocorrelation ($C(u, 2\Delta t)$)	2.419 [-2]	2.420 [-2]	2.642 [-2]	2.202 [-12]
24	Autocorrelation ($C(u, 3\Delta t)$)	2.393 [-2]	2.400 [-2]	2.700 [-2]	1.185 [-10]
25	Autocorrelation ($C(u, 4\Delta t)$)	2.369 [-2]	2.383 [-2]	2.757 [-2]	5.938 [-10]
26	Autocorrelation ($C(u, 5\Delta t)$)	2.340 [-2]	2.362 [-2]	2.824 [-2]	1.405 [-9]
27	Autocorrelation ($C(u, 6\Delta t)$)	2.315 [-2]	2.343 [-2]	2.887 [-2]	2.302 [-9]
28	Autocorrelation ($C(u, 7\Delta t)$)	2.287 [-2]	2.318 [-2]	2.956 [-2]	2.810 [-9]
29	Autocorrelation ($C(u, 8\Delta t)$)	2.262 [-2]	2.296 [-2]	3.023 [-2]	3.455 [-9]
30	Autocorrelation ($C(u, 9\Delta t)$)	2.234 [-2]	2.270 [-2]	3.100 [-2]	4.072 [-9]
31	Autocorrelation ($C(u, 10\Delta t)$)	2.207 [-2]	2.248 [-2]	3.176 [-2]	5.415 [-9]
32	Intermittency factor (f_{I_1})	6.595 [-3]	7.816 [-3]	3.556 [-1]	5.304 [-7]
33	Intermittency factor (f_{I_2})	1.319 [-2]	1.417 [-2]	8.889 [-2]	8.486 [-8]
34	Intermittency factor (f_{I_3})	2.540 [-2]	2.858 [-2]	2.397 [-2]	2.416 [-7]
35	Intermittency factor (f_{I_4})	5.813 [-2]	5.887 [-2]	4.576 [-3]	2.457 [-9]
36	Intermittency factor (f_{I_5})	8.256 [-2]	1.038 [-1]	2.269 [-3]	1.025 [-6]
37	Intermittency factor (f_{I_6})	1.050 [-1]	1.265 [-1]	1.402 [-3]	6.477 [-7]
38	Intermittency factor (f_{I_7})	9.355 [-2]	9.428 [-2]	1.767 [-3]	9.489 [-10]
39	Intermittency factor (f_{I_8})	6.815 [-2]	6.913 [-2]	3.330 [-3]	3.179 [-9]
40	Intermittency factor (f_{I_9})	4.006 [-2]	3.053 [-2]	9.638 [-3]	8.746 [-7]
41	Intermittency factor ($f_{I_{10}}$)	9.038 [-3]	9.038 [-3]	1.893 [-1]	0.000 [0]

Least-squares functional, $Q = 3.766$ [-6].
Maximum contribution to Q is from property 36.

Table 2. Statistical properties of data from Burgers' equation and chaotic map model.