False neighbors and false strands: A reliable minimum embedding dimension algorithm

Matthew B. Kennel
Institute for Nonlinear Science, University of California, San Diego, Mail Code 0402, La Jolla, California 92093-0402

Henry D. I. Abarbanel
Institute for Nonlinear Science, Department of Physics and Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, Mail Code 0402, La Jolla, California 92093-0402

(Received 7 February 2002; published 23 August 2002; publisher error corrected 22 November 2002)

The time-delay reconstruction of the state space of a system from observed scalar data requires a time lag and an integer embedding dimension. We demonstrate a reliable method to estimate the minimum necessary embedding dimension that improves upon previous methods by correcting for systematic effects due to temporal oversampling, autocorrelation, and changing time lag. The method gives a sharp and reliable indication of the proper dimension. With little computational cost, the method also distinguish easily between infinite-dimensional colored noise—including noisy periodicity—and low-dimensional dynamics.

DOI: 10.1103/PhysRevE.66.026209 PACS number(s): 05.45.−a, 05.10.−a

I. INTRODUCTION

One of the foundations of the analysis of chaotic time series is the Takens embedding theorem [1,2] that allows us to unfold the attractor of a system from observations of a single dynamical variable. Suppose that the dynamical system which is the source of some observations evolves in a multidimensional phase space $\mathbf{x}(n)=[x_1(n),x_2(n),\ldots,x_d(n)]$ as $\mathbf{x}(n+1)=\mathbf{F}(\mathbf{x}(n))$, and that we observe a scalar component of $\mathbf{x}(n)$ or a scalar projection of the $\mathbf{x}$ space: $s(n)=S(\mathbf{x}(n))$. The theorem guarantees that if we perform a correct time-delay embedding of this scalar signal, we may construct a d_E-dimensional space that inherits many of the properties of the original d-dimensional $\mathbf{x}$ space even though that $\mathbf{x}$ space is unknown to us. The scalar data are replaced by the vectors

$$\mathbf{y}(n)=[s(n),s(n+T),\ldots,s(n+(d_E-1)T)] \qquad (1)$$

as the description of the dynamical system. This d_E-dimensional space is the reconstructed state space. In $\mathbf{y}(n)$ T is the "reconstruction time delay," an integer multiple of the sampling time. The integer d_E is the "embedding dimension." When d_E is large enough, the points in this reconstructed phase space are related to the original space through a diffeomorphism, namely, a smooth, differentiable, and invertible transformation. Many important properties of the original dynamics are preserved, including that we have an implied deterministic time evolution in the $\mathbf{y}$ space corresponding to the unknown evolution in the $\mathbf{x}$ space. That is, if $\mathbf{x}(n)$ is followed by $\mathbf{x}(n+1)$ in the evolution of the underlying system, then its counterpart $\mathbf{y}(n)$ is followed by $\mathbf{y}(n+1)$ in the reconstructed state space. Also since the original orbits $\mathbf{x}(n)$ cannot cross themselves in the original phase space, when d_E is large enough, the $\mathbf{y}(n)$ orbits cannot cross themselves in the reconstructed phase space. The theorem states that when $d_E>2d_A$, where d_A is the box counting dimension of the orbit in the original space, this gives a *sufficient* condition for a correct embedding [2]. In the unrealistic limit of an infinite amount of infinitely precise data, *any* time delay T will work. In practice, of course, there are approximate upper and lower limits to the values of T which are acceptable.

We address the question "Given actual observed data, what d_E do we require to ensure a deterministic mapping?" From a purely topological point of view, there is no penalty as long as a sufficiently large d_E is employed, but there is no benefit using a larger than necessary dimension. Practically, however, there may be a penalty associated with using too large a value for d_E. When dealing with finite amounts of finite precision data, one would usually like to use as few coordinates as possible. First, as one increases d_E, the noise that inevitably accompanies any real signal contaminates all components of the vector $\mathbf{y}$, and will populate every dimension whether or not there is a real signal there. This is obviously undesirable. Second, if one uses the reconstructed phase-space vectors to make empirical models for prediction or control, then using unnecessarily high dimension has a significant penalty, especially as the computational cost and number of free parameters of many estimation and prediction methods scale exponentially with d_E. Finally, an indication of the minimum embedding dimension puts an upper bound on the dimensionality of the system. A strong empirical indication of a topologically satisfactory small embedding dimension provides evidence that the observed signal may have, in fact, arisen from a low-dimensional dynamical system. These are the core issues discussed in this paper.

Our goal is to find the minimum necessary embedding dimension denoted d_E. Ding *et al* [3] demonstrated that the commonly used correlation dimension reaches its correct value, in principle, as soon as the trial embedding dimension d_E is an integer just greater than the correlation dimension of the attractor. For purposes of computing the correlation dimension, then, $d_E>d_A$ will do. This may still be insufficient to unfold the attractor for purposes of revealing the dynamics. As a simple example, consider a limit cycle in a flow. This attractor has a geometric fractal dimension of one. In some two-dimensional projections this limit cycle could be

topologically equivalent to a circle, but in others, it could appear as Arabic numeral ''8.'' At the intersection point the implied vector field in the reconstructed state space is not well defined and the embedding is inadequate for dynamical reconstruction though it is sufficient for determining the geometric fractal dimension (i.e., one). This is not an isolated pathology; in general, an embedding dimension adequate for evaluating the correlation dimension may not be sufficient when dealing with the actual underlying dynamics which is at the source of the observations.

The now popular false nearest neighbor method [4,5] yields the necessary global dimension d_E for unfolding an attractor using only observed data. It assures that points have state-space neighbors that are a result of the dynamics rather than being projected from near one another as an artifact of using too low an embedding dimension. Choosing an inappropriately low d_E will cause parts of the attractor, which are widely separated in the original, but unknown, state space to overlap spuriously in the reconstructed space. To detect this, the false nearest neighbors test constructs vectors from the data in dimension $d=1$, then $d=2$, and so forth, asking at each stage what fraction of nearest neighbors in the data set, as seen in dimension d, fail to remain close in dimension $d+1$. When all nearest neighbors are true, that is, do not significantly move apart when we go to the next dimension, we have found the necessary minimum embedding dimension $d_E=d$.

This paper demonstrates a number of improvements to the earlier method that address some of the issues we have encountered in dealing with various data sets from fluid dynamics, climatic variations, nonlinear circuits, laser dynamics, and neuronal systems. The outcome of the suggestions we make here is a more reliable algorithm that eliminates various systematic effects with embedding dimension so as to give quite reliable answers. Specifically, we provide corrections (1) to account for the neighborhood properties of oversampled continuous data, (2) to account for autocorrelation when a small time delay is used, and (3) and to account for sparsely populated regions of the attractor.

A principal goal is to provide a *sharp* indication of the minimum embedding dimension by eliminating systematic effects with dimension, sampling rate, and correlation that might otherwise make the determination of d_E less certain. The improved statistic also provides a fast and powerful test against colored noise with linear correlation. If such a signal has strong spectral peaks, so we might call it ''noisy periodicity,'' have smooth and often ''deterministic-appearing'' structure in phase space. It is important to be able to reliably distinguish such signals from low-dimensional deterministic, chaotic dynamics.

II. NEIGHBORS: TRUE AND FALSE

The implementation of our ideas boils down to choosing criteria for determining an ''illegal projection.'' For each observed point on the attractor $\mathbf{y}(i)$ in d dimensions we find the *nearest neighbor*, called $\mathbf{y}^{\mathrm{nn}}(i)$. We then examine the same vectors $\mathbf{y}^{\mathrm{nn}}(i)$ and $\mathbf{y}(i)$ in dimension $d+1$. This simply means adding the $(d+1)^{th}$ component $s(i+d_E T)$ to $\mathbf{y}(i)$, and similarly to $\mathbf{y}^{\mathrm{nn}}(i)$. If the distance between $\mathbf{y}^{\mathrm{nn}}(i)$ and $\mathbf{y}(i)$ in $d+1$ dimensions is very large, then we have a *false nearest neighbor*. This is the hallmark of an illegal projection caused by improper embedding. Efficient computer algorithms [6,7] can find near neighbors of a point in an N point data set in $O(\ln N)$ time, rendering the overall algorithm $O(N \ln N)$ rather than $O(N^2)$. The latter would make the algorithm computationally infeasible for substantial datasets.

One part of the our earlier algorithm [5] compared the ratio of the distance incurred by considering the $(d+1)$th component to the nearest neighbor distance in d dimensions. A sufficiently high value of this ratio was deemed to indicate a ''false neighbor,'' namely, points close in dimension d were far apart in dimension $d+1$. The specific test for false neighbors was to evaluate if

$$\frac{|y_{d+1}(i)-y_{d+1}^{\mathrm{nn}}(i)|}{R_d(i)}>\rho_r, \tag{2}$$

with

$$R_d(i)^2=\frac{1}{d}\sum_{k=1}^{d}[s(i+(k-1)T)-s^{\mathrm{nn}}(i+(k-1)T)]^2 \tag{3}$$

the distance in dimension d. If this ratio is greater than some threshold ρ_r, we deem these points $\mathbf{y}(i)$ and $\mathbf{y}^{\mathrm{nn}}(i)$ false nearest neighbors. Typically we take $10<\rho_r<20$.

This criterion alone had a flaw: the typical distance even to nearest neighbors can be substantial (even a large fraction of the global attractor size) in higher-dimensional spaces, contrary to our low-dimensional intuition. This results in a downward bias of this ratio statistic with increasing embedding dimension. This occurs since the typical denominator, namely, the distance in dimension d, got so large that an authentically false neighbor would not result in a sufficiently high ratio to trip the false neighbor alarm. An infinite-dimensional white noise data set would indicate a small proportion of false neighbors at a sufficiently high embedding dimension because of this effect. We still needed this test to deal with significantly autocorrelated data, but we present here an alternate approach that appears to work better.

In our earlier paper [5] we dealt with the small noisy data set with a second ratio test comparing the augmented distance in going from dimension $d\rightarrow d+1$ to the overall size of the attractor. If we are using time-delay embedding, then each $\mathbf{y}(n)$ embedded in $d+1$ dimensions consists of the same vector as embedded in d dimensions with one new element in the last position. In the case of a false nearest neighbor, the additional large distance in $(d+1)$ dimensions will occur solely in the last component, because the first d components are constructed to be small by virtue of being nearest neighbors. Imagine that the data are, in fact, white noise of standard deviation σ. The absolute difference between the $(d+1)$th components of the vectors, $|\mathbf{y}_{d+1}^{\mathrm{nn}}(i)-\mathbf{y}_{d+1}(i)|$ will in probability be as large as $\sqrt{2}\sigma$ and many nearest neighbors will register as false, as desirable for this class of signals. This suggested another test for a false neighbor: $\mathbf{y}^{\mathrm{nn}}(i)$ is a false neighbor if

$$\frac{||\mathbf{y}^{\rm nn}_{d+1}(i)-\mathbf{y}_{d+1}(i)||}{R_A}>\rho_a, \tag{4}$$

where R_A is a scalar measuring the overall size of the attractor, and ρ_a a fixed parameter chosen externally, with typical values between 1 and 2. We call this the ''absolute'' test, because it compares the size of the nearest neighbor deviation to an absolute quantity independent of the nearest neighbor distance. A reasonable choice for R_A is

$$R_A=\left(\frac{1}{N}\sum_{i=1}^{N}[s(i)-\overline{s}]^2\right)^{1/2} \tag{5}$$

with $\overline{s}$ the average value of the observed scalar data. With this test there is less intrinsic bias in its variation with d.

As our first improvement for the current work, we consider systematic effects with dimension; namely that distances increase with the dimension attempted. In Eq. (4) the numerator will scale as this but the denominator, being derived from scalar statistics, will not. Our criterion of a false neighbor uses the numerator from Eq. (2) over the denominator of Eq. (4),

$$\frac{|y^{\rm nn}_{d+1}(i)-y_{d+1}(i)|}{R_A}>\rho_a. \tag{6}$$

In the actual algorithm presented below, we employ this criterion to judge groups of neighboring points, not merely single neighbors. Furthermore, with the decorrelation transformation also mentioned below, a relative ratio test (2) is superfluous.

III. THE ALGORITHM AND STATISTIC

A. Removing spatial neighbors nearby in time

Often one is presented with *oversampled continuous data*. The underlying physics may be governed by ordinary differential equations, leading to smoothly observables smoothly varying in time. One must sample this data at a finite time resolution, and one ought to sample at a rate fast enough to capture all the important dynamical behavior without aliasing. Within reason, from this point of view, the more oversampling, the better it is, as it results in a smoother reconstructed flow.

Oversampling introduces complications into the false nearest neighbor algorithm. The data analyst might suspect oversampling when statistics of the time series demonstrate significant autocorrelation. For instance, if the first local minimum (if extant) of the average mutual information is a large integer times the sampling time τ_s, then the data is most likely oversampled. One can check this by removing every other data point, and the average mutual information should show a minimum half as large as in the original data.

This section discusses methods that discount systematic geometric phenomena that could have a strong effect on the measured false neighbor statistic yet are irrelevant to the minimum embedding dimension question.

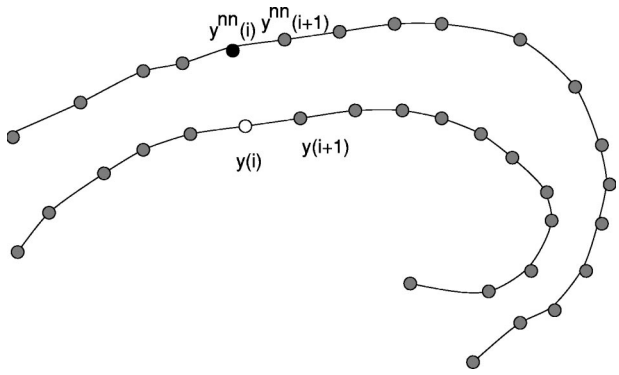

FIG. 1. An illustration of false neighbors and false strands.

If the sampling rate is fast, and successive scalar observables differ only slightly, then temporally successive points in the reconstructed state space will also be close in spatial distance. With such an oversampled trajectory, however, the ''nearest neighbor'' to a reference point $\mathbf{y}(i)$ will very often be the next point $\mathbf{y}(i+1)$ or, perhaps, the previous point $\mathbf{y}(i-1)$. The distance between the $(d+1)$th coordinates of these two points, the quantity going into the false neighbors test, will then *never* be large, simply because of the high oversampling rate. Under these circumstances, until the number of data becomes enormous, one will never register a false nearest neighbor, regardless of whether there is an actual illegal projection. The average distance between points when we have N data is $\approx 1/N^{1/d}$ in dimension d. This eventually becomes less than the distance between spatially successive points in oversampled data, but the number of data may need to be very large to accomplish this.

To illustrate this matter consider Fig. 1. Clearly the point designated $\mathbf{y}^{\rm nn}(i)$, and not $\mathbf{y}(i+1)$, is the nearest neighbor we had in mind in developing the false nearest neighbor method. So we must tell the algorithm some way to distinguish the correct, but potentially false, neighbor from the nearest point along the same section of orbit lying close by dint of having oversampled the data. The fix for this problem is now well known. When searching for nearest neighbors, one ignores all points within a certain temporal *decorrelation interval W* around $\mathbf{y}(i)$. We want to exclude the entire ''trajectory segment'' temporally associated with $\mathbf{y}(i)$, and not necessarily just $\mathbf{y}(i\pm 1)$. If we did only that, then $\mathbf{y}(i\pm 2)$ could end up being a nearest neighbor for a highly oversampled trajectory.

This ''decorrelation correction'' was applied by Theiler [8–10] in the context of computing the correlation dimension [11] from observed data. Osborne and Provenzale [12] found that high-dimensional *colored* noise could result in apparently finite values of the correlation dimension, thereby casting doubt on some findings of low dimensionality in experimental data by using this method. Theiler demonstrated that this finite-dimensional correlation dimension came about by counting neighbors on the same trajectory segment. The decorrelation correction eliminated this effect.

The results ought not depend very much on the trajectory segment interval chosen, as long as it is long enough to completely exclude the local trajectory segment, and not so long

that it does not exclude too many valid parts of the dataset. This sounds like smoke and mirrors, but we have found that a good rule of thumb is to choose an interval about two or three times the first minimum of the average mutual information. We have found good results when we choose the decorrelation interval to be a time scale on which the average mutual information has reached its asymptotic zero value. If the results change substantially with varying decorrelation interval, that is a warning that the data may be nonstationary.

B. False strands

The standard decorrelation window is not entirely a satisfactory correction, however. Another important issue is illustrated in Fig. 1, showing a typical situation for legal embeddings. If, by contrast, the two orbit segments were illegally projected then generically they tend to make a close pass at some strong angle instead of remaining close with nearly parallel flow vectors. The point $\mathbf{y}^{\mathrm{nn}}(i)$ is a true nearest neighbor to $\mathbf{y}(i)$, and $\mathbf{y}^{\mathrm{nn}}(i+1)$ is a true nearest neighbor to $\mathbf{y}(i+1)$. *There is no new information in this fact.* It is a simple consequence of the oversampled trajectory and autocorrelated data. As the sampling rate of the data is increased, trajectory segments in a ''true neighborhood'' will acquire many more points that register as true neighbors, as these are just iterates of previously found pairs of true neighbors. If we increase the sampling rate in a false neighborhood, the number of false neighbors would increase by a small amount at most, because false trajectory segments cross at oblique angles and cannot remain parallel to the reference trajectory for very long.

The overall systematic effect is to diminish the final statistic: the *proportion of false neighbors*, as the sampling rate is increased. By choosing a high enough sampling rate the proportion of false neighbors could move down to a small number, even with an unacceptably low embedding dimension. This has nothing to do with the geometric issue we are addressing in determining a good value for d_E.

There is yet another problem: as the dimension increases, close true trajectory segments may stay together for longer and longer lengths of time due to the geometry of high-dimensional spaces. This imparts another downward tendency to the false nearest neighbors statistic, which becomes stronger as dimension increases.

Our approach to addressing this matter is to examine and count *nearest strands* rather than focus on nearest neighbors alone. Given pairs of real nearest neighbor points we collect sets of these pairs which are direct temporal iterates of one another. We collect a whole strand of trajectory nearby to the trajectory we are presently examining. These sets comprise *nearest strand pairs*. The central statistic is then to find the number of false nearest strands compared to the total number of strands. Thus we eliminate the previously mentioned systematic error due to the changing sampling rate. Accordingly we name the proposed method ''false nearest strands.''

How do we identify strand pairs? A strand pair is a list of index pairs $[i,J(i)]$ which in turn, designate the indices of nearest neighbors in phase space. For each $\mathbf{y}(i)$ we find the nearest nontrivial neighbor $\mathbf{y}^{\mathrm{nn}}(i)$. This nearest spatial neighbor possesses time index $J(i)$; namely, $J(i) = \mathrm{argmin}_J ||\mathbf{y}_d(J(i)) - \mathbf{y}_d(i)||$, subject to $|J(i)-i|>W$. Next we examine temporal predecessors of the reference pair $[i,J(i)]$, $[k,J(k)]$ starting at $k=i-1$ and counting down to $k=i-W$. We identify that k^* closest to i for which $[i,J(i)]$ is a direct temporal iterate of some point pair $[k^*,J(k^*)]$, that is $J(i)-i=J(k^*)-k^*$. The pair $[i,J(i)]$ is declared to be on the same *strand pair* as $[k^*,J(k^*)]$, and we append it to the appropriate list. Each point pair $[i,J(i)]$ is associated with at most one previously extant strand pair, the one containing $[k^*,J(k^*)]$. If $[i,J(i)]$ is not a direct iterate of any $[k,J(k)]$ for $k\in[i-W,i-1]$ then we create a new strand pair and make $[i,J(i)]$ its first member. The pair $[1,J(1)]$ is initialized as the first point in the first strand.

After examining all data, we will have accumulated a collection of strand pairs, each of whose elements are pairs of points that have identical temporal offsets from each other and pointwise are nontrivial nearest neighbors. If two trajectory segments travel parallel to one another and stay as nearest neighbors for a significant amount of time, as in Fig. 1, all these points will end up on the same single strand pair, and thus they get counted only once. We denote N_s as the cardinality of the completed set of strand pairs.

Designating strand pairs in this way automatically accounts for the situation when more than one trajectory segment stays close to the reference segment, and the identity of the *nearest* one jumps back and forth between these trajectory segments. If there were two segments for example, then some point pairs will be assigned to one strand pair and some to the other, but there will be just two separate strand pairs constructed for this region. This becomes important if there are a number of parallel trajectories in some region of phase space but with data contaminated with a slight amount of noise. Our definition of strand pairs deals with this situation well, in contrast to the slightly more intuitive notion of requiring strand pairs to encompass only consecutive points. That alternative works well on very clean data, but would generate many new separate strands with a small amount of noise, which would nonetheless not represent ''new'' information.

Once we have defined strand pairs, we define true and false nearest strands. There are many different reasonable choices, such as, (1) a strand pair is false if any of its points are false or (2) a strand pair is false if the closest point pair is false. We found best results if we choose to designate a strand pair as false when the averaged ''extra distance'' is too large. Suppose we have two strands of a strand pair S, the set of reference points $\mathcal{R}_S$ and the set of neighbors $\mathcal{N}_S$, in total $m=S$ point pairs. We evaluate the average distance of the $d+1$th component of the vector $\mathbf{y}(k)$ ($k\in\mathcal{R}_S$ being a time index of the points on the reference strand) from the $d+1$th component of its nearest neighbor $\mathbf{y}(J(k))$ on the $\mathcal{N}_S$ In particular, we compute

$$\Delta(S)=\frac{1}{m}\sum_{k\in\mathcal{R}_S}|y_{d+1}(J(k))-y_{d+1}(k)|. \tag{7}$$

Each element of the summation is a distance exactly the numerator in Eq. (6), with arithmetic average $\Delta(\alpha)$. If

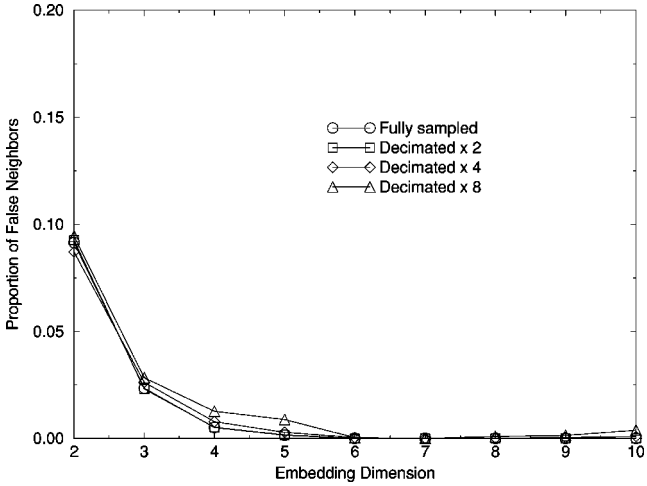

FIG. 2. False neighbors as a function of decimation interval, without sampling corrections. The data comes from the Lorenz model of the atmosphere. The initial data are 16 384 points sampled with $\tau_s=0.5$. The data are then halved, then quartered, then cut by a factor of 8. This reduces the oversampling by hand.

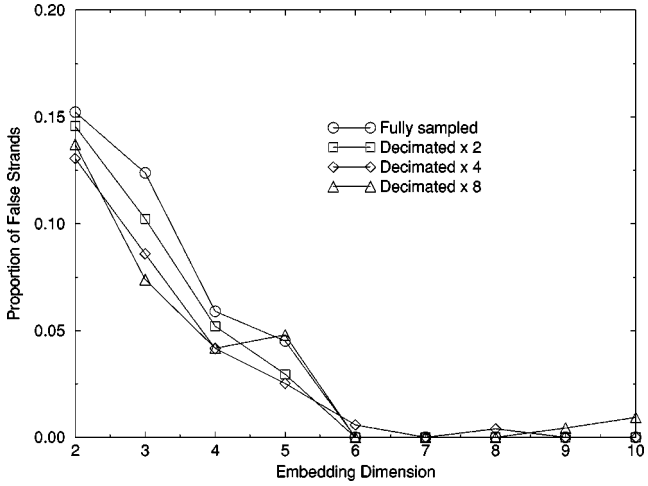

FIG. 3. False strands as a function of decimation interval, including both sampling corrections. The data come from the Lorenz model of the atmosphere. Initial data are 16 384 points sampled with $\tau_s=0.5$. The data are then halved, then quartered, then cut by a factor of 8. This reduces the oversampling by hand.

$\Delta(S)/R_A>\rho$ then this strand pair is declared false. The final statistic is the proportion of false strand pairs to the total number of strand pairs.

Our goal is that each different strand pair should be approximately independent information. One can then make an admittedly very rough estimation of the uncertainty by further assuming Poisson statistics. If the proportion of false strand pairs is N_f/N_s, one could guess a standard deviation of $\pm\sqrt{N_f}/N_s$, or employ a binomial model. With substantial oversampling, using false points would greatly overestimate the precision.

With these oversampling corrections, one can use large amounts of highly oversampled data with greater confidence. Figure 2 shows the proportion of false neighbors without oversampling corrections on the 16 384 points of the x coordinate of a three-variable system of equations due to Lorenz [13],

$$\frac{dx(t)}{dt}=-y^2-z^2-a(x-F),$$

$$\frac{dy(t)}{dt}=xy-bxz-y+G,$$

$$\frac{dz(t)}{dt}=bxy+xz-z, \tag{8}$$

with parameters $a=0.25$, $b=4.0$, $F=8.0$, and $G=1.0$, sampled at intervals $\tau_s=0.05$. The attractor has a dimension of about 2.5, which means it is rather more complicated than the traditional Lorenz attractor.

As an initial time delay we chose a reasonable $T=16$—the average mutual information criterion indicates $T=17$. The figure demonstrates what happens to the false neighbors statistic as this dataset is decimated by successive factors of 2 (scalar data excised) with the time delay in units of τ_s successively halved to maintain the same geometry, thereby effectively lowering the sampling rate (increasing τ_s). There is a convergence of the standard false neighbors criterion to zero starting at $d=4$.

We often use as a rule of thumb a false nearest neighbor reading of less than 1% to indicate the necessary embedding dimension given clean data, but it would be rather difficult to make the distinction among dimensions 4 or 5 or 6 here. If one increases the sampling rate the false nearest neighbor statistic decreases. The same attractor that may have appeared to require six dimensions would perhaps seem embeddable in four or five dimensions simply by virtue of being more oversampled, which cannot have any substantial effect on the topology. By contrast, Fig. 3 shows the results when the two oversampling corrections are included. For $d<6$ there are always a substantial fraction of false strands, well above 4%, with a clean drop to *zero* at $d=6$, regardless of the sampling rate. This time, the proportion of false strands for $d<6$ *increases* with increasing sampling rate: the attractor is more fully defined in all places, thereby providing more opportunities to precisely observe the close crossings from illegal projections. We hypothesize that for $d=4$ or $d=5$ there are a modest number of illegal projections but when looking only pointwise they may be overwhelmed by the increasing number of true neighbors in a highly oversampled systems. With the strand correction the drop-off, and therefore the confidence we have in choosing $d_E=6$, becomes sharper with more data. With the oversampling corrections, more data mean better results: a higher fraction of false strands for d_E too small, and a lower fraction of incorrect false neighbors for correct d_E.

C. Removing linear correlation

Autocorrelation among observations $s(n)$ has another effect on the observed minimum embedding dimension statistic. When the embedding time delay is short, then autocorre-

lated behavior will make the proportion of false neighbors registered appear small regardless of the actual dynamics. A too short time delay folds the attractor onto a thin tube along the $\mathbf{y}_1=\mathbf{y}_2=\cdots=\mathbf{y}_d$ line. Consider the ''absolute'' false neighbors test

$$\frac{|\mathbf{y}^{\mathrm{nn}}_{d+1(i)}-\mathbf{y}_{d+1}(i)|}{R_A}>\rho_a\,. \tag{9}$$

If T is short, then $\mathbf{y}_{d+1}(i)$ is usually close to $\mathbf{y}_d(i)$ by autocorrelation, likewise for $\mathbf{y}^{\mathrm{nn}}(i)$ also. By construction, $\mathbf{y}^{\mathrm{nn}}_d(i)$ will be close to $\mathbf{y}_d(i)$, and therefore $\mathbf{y}^{\mathrm{nn}}_{d+1}(i)$ will automatically end up being close to $\mathbf{y}_{d+1}(i)$ and thus will not register as a false neighbor, by the absolute criterion. If we ask that false neighbors evolve to macroscopic distances (that is, with ρ on order of unity), we must permit this to happen. Autocorrelated data and a short time delay will suppress this. The net effect is to again suppress observed false neighbors with smaller T. Thus we might erroneously infer a too-small embedding dimension as a result of choosing a small time delay. With large enough autocorrelation, one will not register *any false neighbors*, despite the presence of an inappropriate embedding. It reduces confidence in the results when the apparent minimum embedding dimension changes drastically with the time delay—much more than permitted by even a loose interpretation of the embedding theorem. The standard approach is to fix a time-delay presumed to be good by another means, such as the average mutual information criterion [14]. However, we have found some examples where that estimate of time delay may be excessively large, as when the reciprocal of the local Lyapunov exponent is large relative to the autocorrelation or mutual information time. That is, in regions of phase space where the system is especially unstable, average mutual information may yield a misleadingly large time delay for that section of phase space.

To overcome this problem, we apply a global linear transformation to the state space which removes all linear cross correlation among the components and renders each auto correlation unity. With $\mathbf{z}$ denoting the transformed space $\mathbf{y}\rightarrow\mathbf{z}$, this means that the cross correlation among components

$$C(\mathbf{z}_i,\mathbf{z}_j)=\frac{1}{N}\sum_{m=1}^{N}[\mathbf{z}_i(m)\mathbf{z}_j(m)]=\delta_{ij}\,, \tag{10}$$

$i,j=1,2,\ldots,d$, and we have N total d-dimensional data points. The mean is subtracted from the $\mathbf{y}(m)$ vectors prior to the transformation. If $\mathbf{z}_{d+1}(m)$ has no linear cross correlation with $\mathbf{z}_d(m)$, then we will not suffer the problem mentioned in above. One well-known and standard transformation satisfying this requirement is the rotation and stretching to normalized principal components; for instance one transforms to $d+1$ dimensional principal coordinates $\mathbf{z}$, finds neighbors in the d-dimensional subspace, and examines the $d+1$ coordinate for true and false strands.

This linear correlation between components is successfully removed but it presents a problem for our use, however: the necessary global embedding dimension found in this new space may *not* be the same as that of the original time-delay embedded space. For example consider a one-dimensional quadratic map: with $s(n+1)=f(s(n))$, the minimum embedding dimension is one, and we see a parabola if we plot the two-dimensional phase space $\mathbf{y}(n)=[s(n+1),s(n)]$. The linear transformation to principal coordinates $\mathbf{z}=\mathbf{T}\mathbf{y}$ in two dimensions might rotate this parabola so that the mapping is no longer one-to-one from the first coordinate of $\mathbf{z}$ to the second, meaning that the minimum embedding dimension is greater than 1 in this space. We require that the minimum embedding dimension of the transformed space be the same as the original time-delay embedded space. In the Appendix we derive the rotation that preserves this property: a linear transformation in $d+1$ dimensions from $\mathbf{y}\rightarrow\mathbf{z}$ that not only normalizes each component and eliminates linear cross correlation, but also guarantees that the first d coordinates of $\mathbf{z}$ are functions only of the first d coordinates of $\mathbf{y}$. In $\mathbf{z}$ space we find nearest neighbors in d dimensions, and examine the difference between the $(d+1)$th coordinates of neighbors—which are now guaranteed to be linearly decorrelated from all the other d coordinates, and have unit standard deviation, thus implying $R_A=1$. We may now hope to observe false neighbors or strands even when the embedding time delay is short.

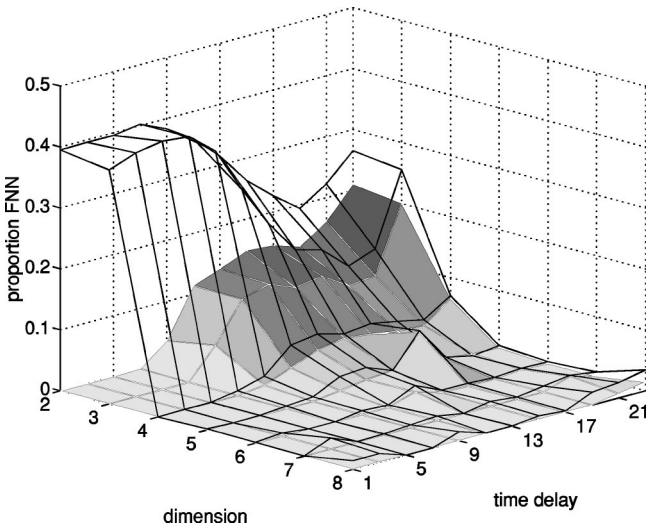

FIG. 4. Proportion of false nearest strands as a function of time delay (multiples of τ_s) and trial embedding dimension. 25 000 data points were taken from the atmospheric model of Lorenz. The parameters for the numerical simulation were as noted earlier in the text; $\tau_s=0.05$ which leads to a suggested $T=17$ from considerations of average mutual information. The opaque surface includes neighbor decorrelation (strand) corrections but not the normalized principal component transformation. The clear mesh adds the global linear decorrelation.

IV. NUMERICAL RESULTS AND CONTRAST WITH NOISE PROCESSES

A. Some numerical simulations

Our first example uses 25 000 data points generated from the small atmospheric model of Lorenz [13]. Figures 4 and 5 show the global false neighbors statistic with (clear mesh) and without (solid surface) the decorrelation transformation.

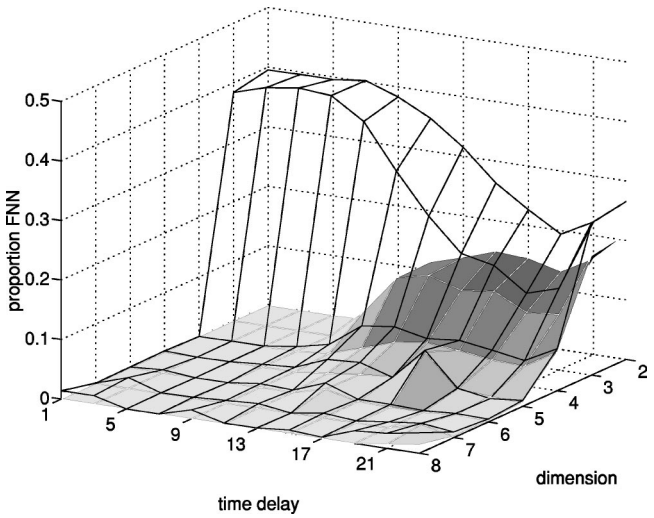

FIG. 5. This is the same as Fig. 4 viewed from a different angle.

The average mutual information criterion suggests $T=17$ for this data set. This falls into the range of our rule of thumb of being oversampled data, though not by much. Near $T=17$, the augmented global false neighbors criterion suggests that the best embedding dimension is $d_E=6$. This is certainly consistent with a box counting dimension $d_A\approx 2.5$. With smaller time delays, starting below 11, the false neighbor statistic calculated on untransformed time-delay embedded data decreases rapidly for all dimensions due to the previously mentioned systematic effect. The transformed data, however, show a very clear drop-off at $d=4$ of the false neighbors statistic from almost 40% to less than 1%. Thus the method clearly demonstrates that $T=7$ and $d=4$ is a good embedding. The untransformed data also show a drop-off to near zero at $d=4$ for values of T between 7 and 11, but the prudence suggests this may be a result of the strong downward bias in embedding dimension for small time delays. In this example, we find that a smaller dimension $d_E=4$ gives a good set of embedding coordinates $\mathbf{y}(n)$ at $T=7$. The time delay of $T=17$ suggested by average mutual information requires $d_E=6$. It is important to note that, even theoretically, that *the minimum necessary global embedding dimension need not be invariant with changing time delay* because such a change constitutes a nonlinear change of coordinates. What should be invariant is the local dimension required to capture the dynamics and the correct number of Lyapunov exponents of the dynamics [18]. The decorrelation transformation will allow one to safely employ smaller time delays than those necessary to ''fill out the attractor'' (remove componentwise correlation) as chosen by methods such as average mutual information and still be confident that a signal of few false neighbors is *not* the result of autocorrelation that masks detection of topological self-crossings.

The message here is that there are many perfectly acceptable embedding spaces with $d_E>d_A$ and d_E less than the sufficient condition indicated by the embedding theorem: $d_E>2d_A$. The different embeddings all meet the criterion of zero (or very small) proportion of false neighbors or false strands, if oversampling is an issue. The different embeddings in the viewpoint we take here are all time-delay reconstructions of phase space and differ in the choice of time delay. Different choices of time delay are different nonlinear choices of coordinates for the d_E-dimensional space, and there is no reason why the *global* dimension required should not change as we vary the time delay used. As noted the *local* dimension must be the same as that is the dimension of the active dynamics. Fractal dimensions and Lyapunov exponents will be unaltered under these coordinate changes [11]. The usual point of view has been to choose the time delay T based on mutual information, and then to choose d_E based on false neighbors. It is entirely logical to choose both T and d_E based on false strands, since each choice constitutes a different choice of coordinate systems for the reconstruction of the dynamics. Presuming each choice has a sufficiently low proportion of false strands, deterministic evolution is assured and model making can proceed [11].

B. Noise: White and colored

With the decorrelation transformation, the improved algorithm provides a good test distinguishing real low-dimensional dynamics from correlated, colored noise signals. This kind of noise (very high-dimensional dynamics) can confuse standard false neighborlike tests. For example, if the spectrum is very red, falling rapidly as frequency increases, it implies strong autocorrelation, and thus with the previous arguments, there could be few false neighbors or lack of determinism observed even though the dynamics are formally infinite dimensional—even with a decorrelation correction in choosing neighbors.

Removing the linear correlations between components is prewhitening of the data and serves to give somewhat equal emphasis to all frequency components of the data. As our transformation of the Appendix is a linear transform of the original data, it does no harm from the theoretical point of view of the embedding theorem. All local properties will be preserved, and when there really are many degrees of freedom in the signal, this whitening transformation will reveal them. Of course, whitening does tend to amplify noise, which may have the practical consequence of degrading statistical quantities computed on low-dimensional attractors reconstructed from somewhat noisy observations. We comment that we are not performing the kind of bleaching that is disrecommended by Ref. [19], where successive scalar residuals that occur after a linear fit is performed to the original data are themselves reembedded into a new state space. That procedure amplifies noise unacceptably, and does not necessarily preserve topological invariants.

Consider vectors drawn from the samples of independent white noise. The $(d+1)$th component will have no relation to the previous d components, and thus choosing nearest neighbors in that space provides no help in predicting the $(d+1)$th component. Thus the criterion for false neighbors

$$|\mathbf{z}_{d+1}^{\mathrm{nn}}-\mathbf{z}_{d+1}|>\rho_a \tag{11}$$

will be met very frequently giving a high proportion of false neighbors: with $\rho_a=1$ around 50% being typically seen as $\mathbf{z}_{d+1}^{\mathrm{nn}}$ and $\mathbf{z}_{d+1}$ would be essentially uncorrelated scalars drawn randomly from the distribution. One would observe a

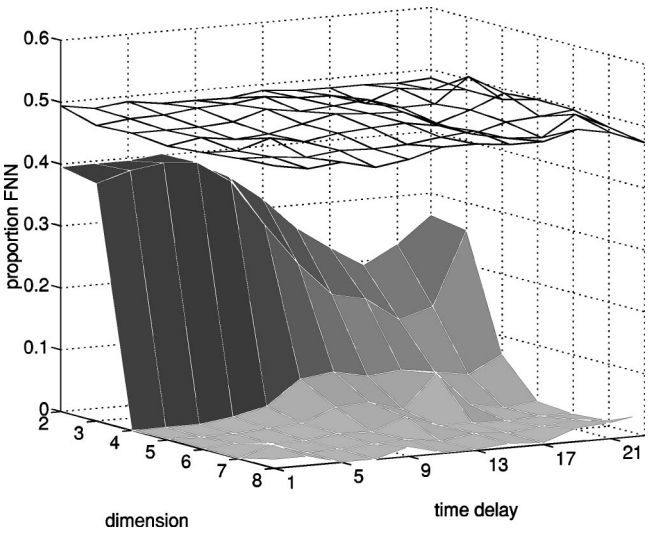

FIG. 6. False nearest strands as a function of time delay and trial embedding dimension. The opaque surface shows the original low-dimensional signal from the model of Lorenz. The clear mesh shows surrogate data: noise with same power spectrum as the true signal. The distinction is quite clear, and supports our foreknowledge that the original signal came from a low-dimensional dynamical system.

nearly constant and high level of false neighbors on the absolute test—independent of embedding dimension and time delay. This is immediately distinguishable from those seen with low-dimensional dynamics.

The key observation is that this situation persists *even with correlated noise*. The reason is that our transformation removes all linear correlation among components, but there is nothing else to the noise signal besides linear correlation. Therefore, when examining the evolution of neighbors in the transformed phase space, the noise appears effectively white. The present algorithm provides a very powerful and easy-to-interpret statistic to distinguish colored noise from authentic low-dimensional dynamics. Figure 6 shows the false neighbors statistic for the low-dimensional dataset previously employed compared with a surrogate noise dataset, constructed with a Fourier transform method, which has the same power spectrum, and therefore the same autocorrelation structure [20,21]. (Furthermore, the two signals compared had the same approximate one-dimensional probability distribution as a result of the ''histogram transformation'' on scalar observables as employed in Ref. [20].)

The contrast is striking. Compared to the authentic signal, the noise results show a much higher level of false neighbors as well none of the characteristic changes seen on a dynamical signal with varying time delay and embedding dimension. The presence or absence of this variation with embedding parameters provides another immediately apparent test as to whether an unknown signal comes from low-dimensional dynamics or an essentially stochastic source.

We suggest to users familiar with the original algorithm [5] that when employing all the new corrections in this work, the standard for an ''acceptable'' percentage of false strands to deem a dataset sufficiently low dimensional be not as strict. The corrections preclude overcounting of close neighbors and may amplify noise, including quantization error. Noise will result in a noise floor reasonably flat with embedding dimension. We point out that the analysis of Ref. [22] should carry over for making a correction for uncorrelated noise by loosening the criterion (6). If one has knowledge of the noise amplitude bound δ on top of a low-dimensional signal one may include in the criterion a loosening to eliminate ''noise-induced'' false neighbors.

$$\frac{|y_{d+1}(i)-y_{d+1}^{\mathrm{nn}}(i)|}{R_A} > \rho_a + 2\delta. \tag{12}$$

In Ref. [22] the authors also consider the deviation due to high local expansion, i.e., a large Jacobian. We comment that a small region of very high expansion may be indistinguishable from a false neighbor. Unfortunately given data alone it may be difficult to estimate the noise level and reliably identifying regions of local expansion is even more difficult. One would first need a good global and local embedding of the data that are required with the results of present algorithm. Thus we do not consider such corrections in the subsequent results.

C. Last component shuffling, an alternative to surrogate data

Owing to the nature of the decorrelation transformation, we can employ a slightly different ''null test'' that is faster and easier than generating surrogate data sets with the same Fourier power spectrum and then recomputing the false neighbor statistic. With linearly correlated noise, the $(d+1)$th coordinate is rendered a random variable independent of the previous parts of the vector. Thus we can compute the level of false neighbors that one would expect from noise alone by assuming that the $(d+1)$th coordinate has nothing to do with the previous ones. This is accomplished by generating new vectors z' whose last component has been shuffled randomly among all the vectors in the database and by computing the proportion of false neighbors on this set. The choice of nearest neighbors depends only on the first d components, which remain unchanged, thus one may use exactly the same indices of nearest neighbors as for the actual data and no additional nearest neighbor search (which comprises the bulk of the computation time) is needed. The additional cost for the shuffled statistic is hence very low. Figure 7 compares the results from a Fourier transform based surrogate dataset to those found by shuffling the last component using the original dataset, after the decorrelation transformation. The shuffling technique predicts the approximate level of false neighbors that one would see with surrogate data of colored noise. Though this may seem a very minor technical detail, it is true that the shuffling and equivalent Fourier spectrum tests are not completely equivalent: the shuffling method will preserve the one-dimensional probability distribution of the original signal whereas surrogate data methods typically generate Gaussian distributions in every projection.

In some ways, the shuffling test is even more direct to the point than surrogate data. Assuming all linear correlation has been removed it asks ''is there any predictable low-

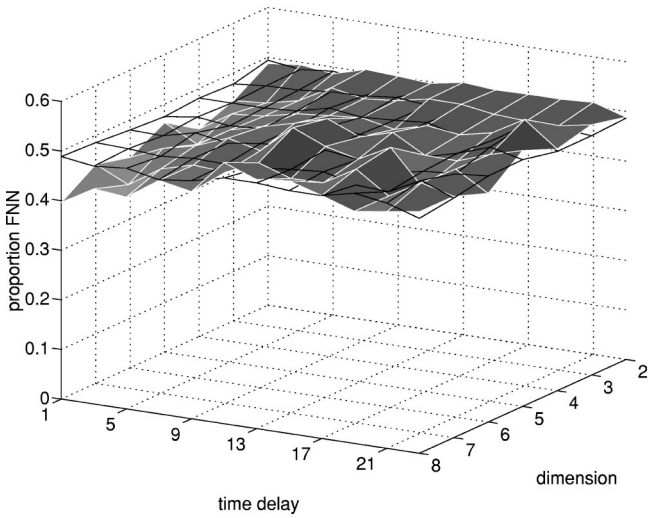

FIG. 7. Proportion of false nearest strands coming from (1) surrogate data (clear surface), noise with the same power spectrum, and (2) data coming from shuffling the last component of the data vector (opaque surface). The data were from the small Lorenz model.

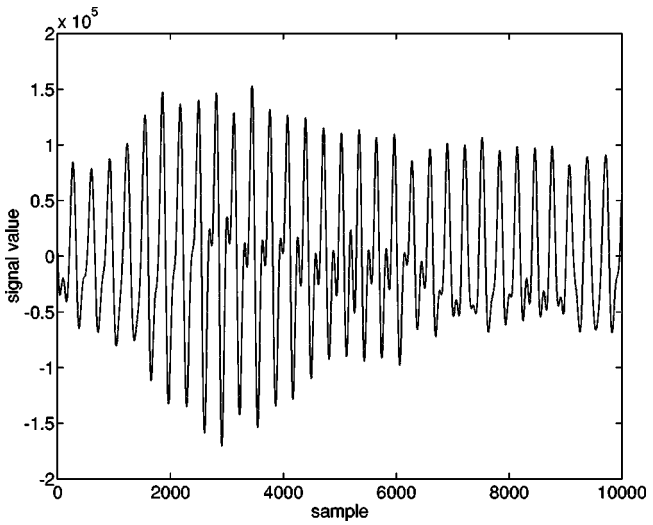

FIG. 8. A sample from a tricky synthetic time series which is noisy periodicity and only resembles chaos. The system was a linear filter of white noise, whose transfer function had poles very close to the unit circle in the *z* plane. More physically this is like noise driving a number of extremely resonant linear oscillators.

dimensional structure in the original data'' by generating what one would see if the answer were ''no, one might have chosen that last coordinate at random.'' At minimum, it is certainly far less computationally intensive than surrogate data methods, requiring only negligibly larger effort than performing the false strands test to begin with.

The shuffling method provides a sharper test for low-dimensional dynamics. One could easily perform the obvious Monte Carlo simulation involving multiple shufflings and derive the appropriate statistical hypothesis test comparing the false neighbors statistic of the original data to the distribution from an ensemble of shuffled sets. This appears to us to be overkill in many practical situations: given reasonably low-dimensional signals, the false neighbor statistic is powerful enough and gives results easily distinguishable by eye from those arising from colored noise.

This technique also gives correct results where Fourier surrogate data methods as in Refs. [20,21] may fail. We synthesized an intentionally tricky highly oversampled data set which came from a few highly resonant oscillators driven by white noise. In signal processing language, there were poles in the *z*-plane transfer function very close to the unit circle, and thus some sharp peaks in the power spectrum. The data were, in fact, generated as a linear filter of Gaussian white noise. Some of the characteristic frequencies were integral multiples of each other, and some were not. The resulting time series was very smooth and resembled, even to the experienced eye, low-dimensional chaotic data despite truly arising from a purely linear stochastic process. Figure 8 shows a time series segment of this dataset, and Figure 9 shows a three-dimensional state-space embedding. By visual inspection, this data looks low dimensional with interesting phase-space structure, in contrast to the ''tangled-spaghetti'' phase-space appearance of broadband colored noise. Nearly all other tests we applied to this data set gave results indicative of low-dimensional chaos, including the Fourier spectrum surrogate data method [20]. In that method there is an extra Kolmogorov-Smirnov test performed on prospective surrogate data before they are accepted into the ensemble of valid surrogates. This check had to be disabled to get any results at all, otherwise all surrogates were rejected, correctly indicating the inapplicability of that test to this dataset.

Figure 10 shows the proportion of false neighbors for the noisy periodic data as well as its last component shuffled counterpart. Here the level of false neighbors is very high in the original data and remains near the 50% seen with the shuffled set. It shows no interesting variation with embedding dimension or time delay, which we have seen is indicative of low-dimensional chaos. The present shuffling algorithm successfully and unequivocally distinguishes this signal from authentic low-dimensional chaotic data. All the computation necessary for Fig. 10 took less than 30 min on a standard SPARC 10 computer.

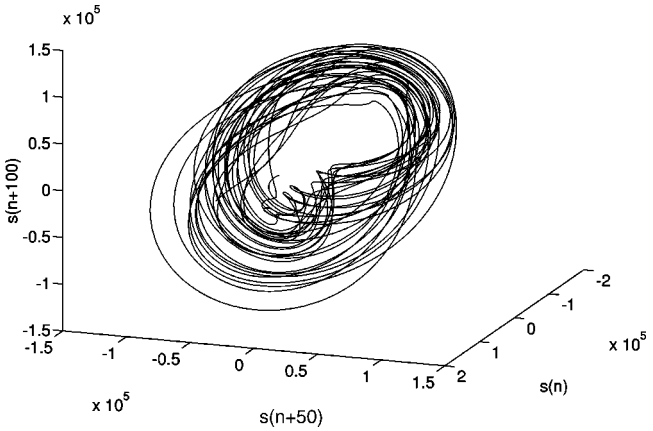

FIG. 9. Time-delay embedding in $d_E=3$ of the ''noisy periodicity'' data set.

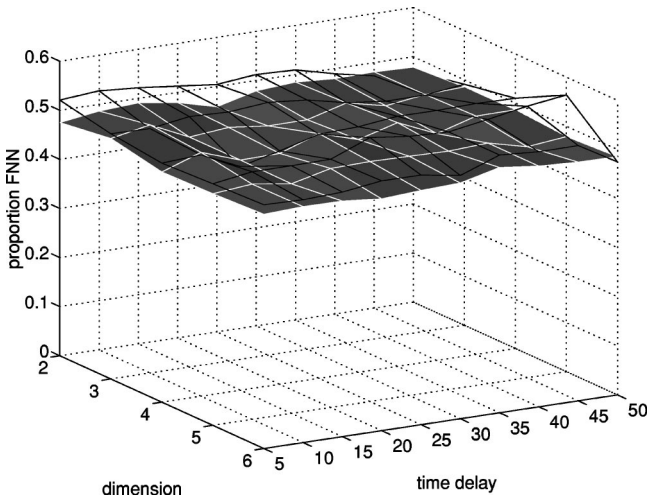

FIG. 10. Opaque surface: Proportion of false nearest strands for the ''noisy periodicity'' dataset. Clear surface: The same with shuffled last component on each vector.

Figure 11 shows the results applied to data which are a static scalar nonlinear transformation ($y=x^3$) of the previous noisy periodicity dataset. In this case, the proportion of false neighbors is smaller than that seen with the shuffled data, but the absolute level is not small and does not show the proper structure with time delay and embedding dimension to indicate chaos. The algorithm diagnoses these data as a nonlinear transformation of noisy periodicity.

V. EXPERIMENTAL EXAMPLES

We now illustrate the above developments with examples from numerical simulations and from laboratory data taken from three sources: (1) Data from a chaotic nonlinear circuit; (2) data from the chaotic fluctuations of a Nd:YAG (yttrium aluminum garnet) laser with an intracavity doubling crystal; and (3) data from the pressure fluctuations in a ''fluidized bed'' of small particles.

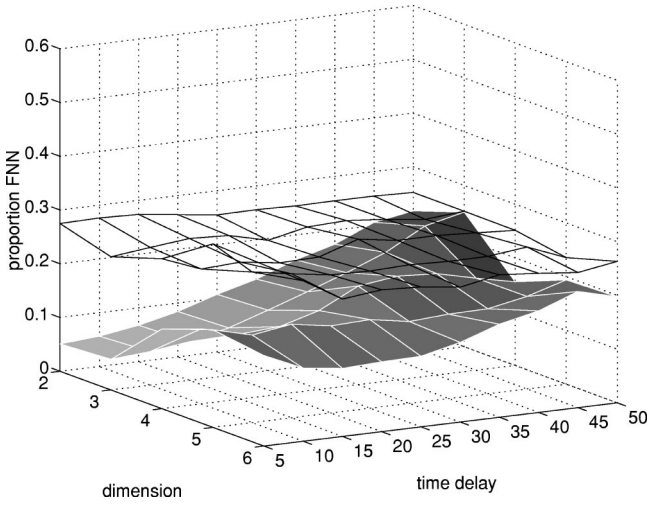

FIG. 11. Opaque surface: The false strands for the noisy periodicity dataset with a time-independent nonlinear transformation. Clear surface: Same with shuffled last component on each vector.

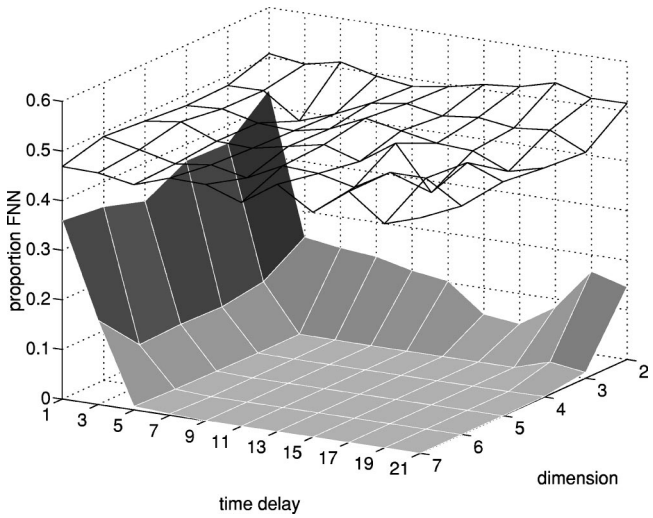

FIG. 12. Opaque surface: Proportion of false neighbors for the chaotic circuit described in the text. Clear surface: Same data with the last component of the vectors shuffled at random.

A. Electronic circuit

The first system is a clean nonlinear circuit described in greater detail in Refs. [23,24]. It consists of a nonlinear amplifier $[V\rightarrow\alpha f(V)]$ whose linear feedback loop has an RC_1 low pass filter and an LC_2 resonance. This is part of an experimental setup intended for studying the synchronization of physical chaotic systems. Figure 12 shows the statistic on a dataset of voltages measured across one of the capacitors. This circuit is part of an experimental setup to investigate chaotic synchronization [23,24]. This system is clearly low dimensional, with $d=3$ indicated as a good embedding dimension over a range of time delays. Figure 13 shows a different measurement of the attractor of the same circuit at the same experimental parameters, but measuring the voltage across the other capacitor. Again, the test shows the data are emphatically more predictable than colored noise, indicating a good embedding now at $d=4$ across a wide plateau of time

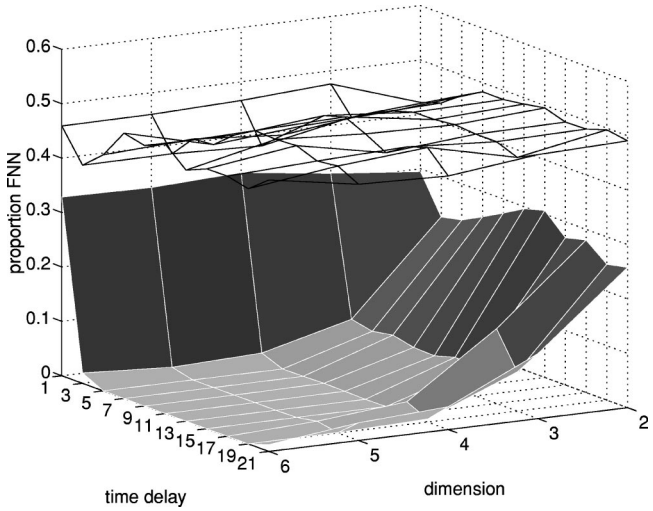

FIG. 13. Same computation for the nonlinear circuit as in Fig. 12, but with a different measured output voltage.

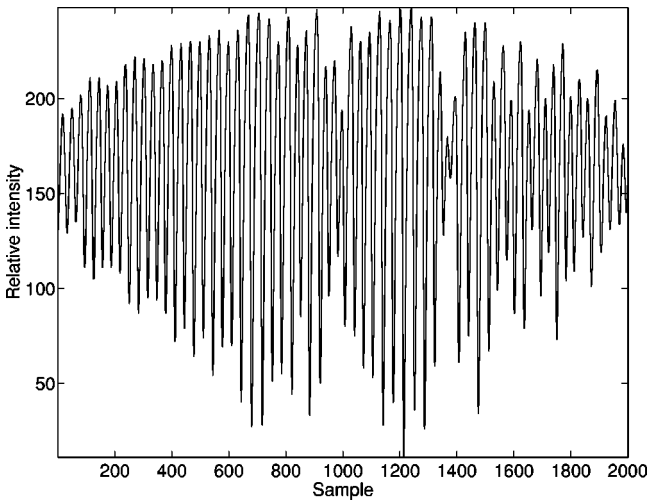

FIG. 14. Sample of time series from controllable Nd:YAG laser dataset. There is a strong natural periodicity.

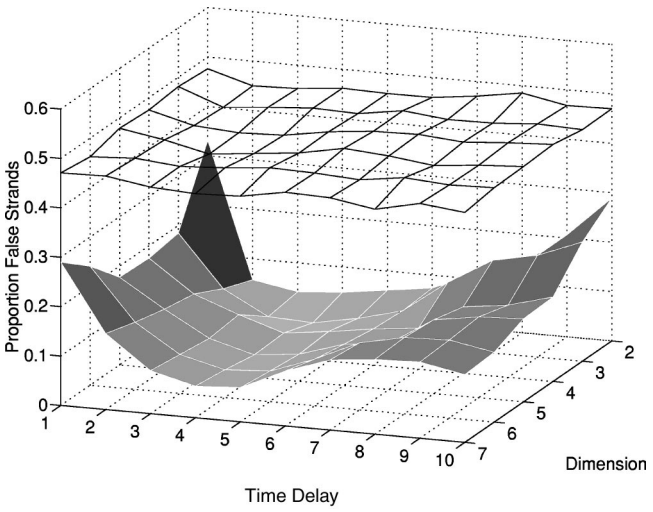

FIG. 15. Opaque surface: Proportion of false neighbors for intensity fluctuations of the laser, for controllable dataset. The minimum occurs at $d_E=3$, with time delay $T=4$ in units of the sampling interval, which is 0.5μ s. Clear surface: statistic with last vector component shuffled.

delays. The behavior in the d_E-T plane clearly indicates a clean low-dimensional dynamical system (similar to results seen with simulations of ODE's), as expected, given our knowledge of the physical system.

B. Laser

We now turn to the experimental data of beam intensities of a Nd:YAG laser with nonlinear frequency doubling crystal in the cavity [25]. The first dataset examined came from a regime where a dynamical control scheme was able to operate—however, the control was not activated here, as these data demonstrate the autonomous chaotic dynamics intrinsic to the laser. There were three electromagnetic modes active in the laser cavity, all polarized in the same direction. The dataset was 100 000 points long, digitized with 8 bit precision recorded at a sampling rate of 2 MHz. Figure 14 shows a segment of the time series. For all calculations on these laser data, we set the decorrelation time interval $W=25$ and $\rho_a=1.0$.

Figure 15 shows the proportion of false strands evaluated on this data set. In contrast to the circuit data, the statistic does not converge to near zero for a sufficiently high embedding dimension. This is because, in contrast to the previous data, this data set is quite noisy, both from experimental quantization noise (about 0.4%) plus what are believed to be fluctuations from spontaneous quantum emission (possibly dynamically amplified) in the lasing medium itself. Nevertheless there is a good minimum in the embedding statistic for $d_E=3$ and $T=3$ or $T=4$. As the embedding dimension increases past its best value, the statistic tends either to plateaus at a nonzero level or even starts to increase again. This is a result of using the normalized principal component transformation (A5) that, as previously mentioned, tends to amplify noise in higher coordinates. (The local maximum at $d_E=4$ for some time-delays is apparently a spurious artifact of the particular state-space structure, as it vanishes upon mild filtering as seen below). With chaotic signal with moderate noise, as with these data, one typically sees a *minimum* or at least a significant plateau in our statistic at a good embedding dimension, in contrast to most other algorithms that usually demonstrate only a subtle "break" in the change of some quantity that may be very ambiguous when used on realistically nontrivial noisy data.

We attempted to separate noise from chaos with a simple low-pass digital linear filter, and computed the false strand statistic on the filtered data. Figure 16 shows the power spectral density of the original and the filtered signal. Despite the appearance of this figure, the filtering was not particularly radical in reconstructed state space, eliminating only 1.4% of the power in the original signal. At frequencies below 150 kHz there was hardly any alteration of phase or amplitude. Figure 17 shows the false strands statistic. Now, there is a very clear convergence at $d_E=3$ for short time delays, re-

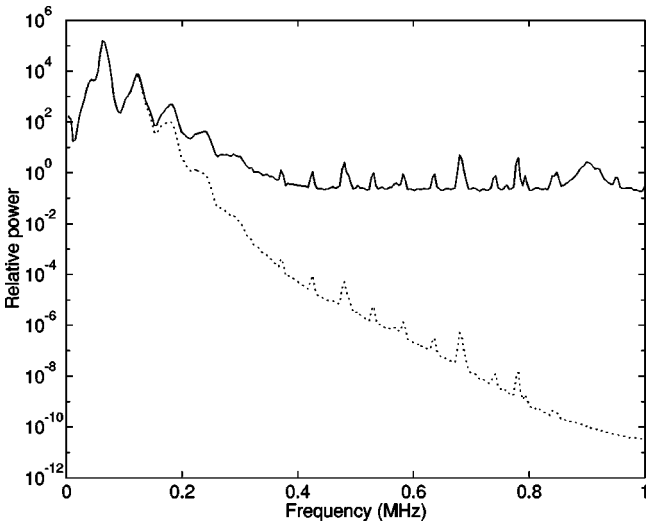

FIG. 16. Power spectral density of original and low-pass filtered laser intensity signal. Filter eliminates 1.4% of power in original signal.

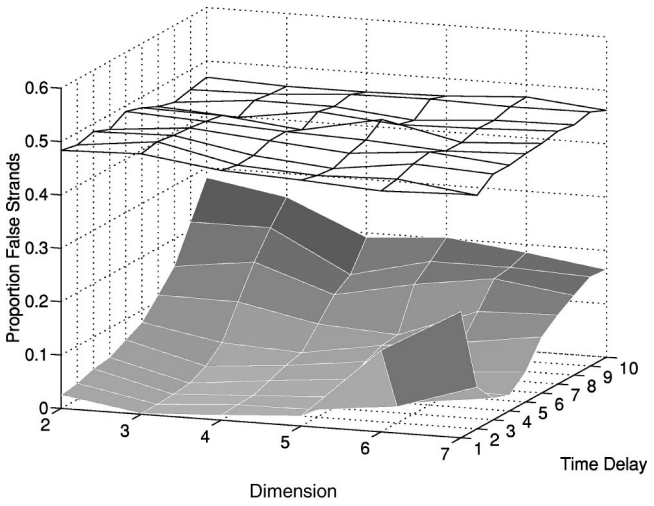

FIG. 17. Opaque surface: Proportion of false neighbors for intensity fluctuations of the controllable laser dataset passed through a low-pass linear digital filter. A clear minimum at a value near zero is seen at $d_E=3$ with time delays between $T=1$ and $T=4$ in units of the sampling interval. Clear surface: same data with last component shuffled.

doubling our confidence in the low-dimensional nature of the signal. The fact that the statistic at $d_E=2$ is not all that large is because the signal is, in fact, approximately periodic in the short term, and so can ''almost'' be embedded in two dimensions.

Figure 18 compares the statistics for different time delays. The mutual information criterion [14] suggests using $T=9$ in order to maximally decorrelate components of the state space. In this case, the results suggest that $d_E=4$ would be necessary at this time delay. However, by using the special principal component transformation, we see that we can, in

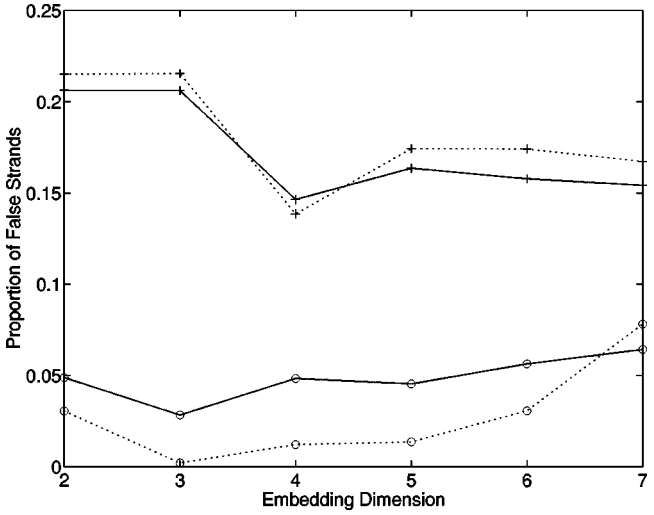

FIG. 18. Lower solid curve: statistic for unfiltered dataset at best (lowest minimum) time delay, $T=4$. Lower dotted curve: statistic for filtered dataset at best time delay, $T=2$. Both these indicate $d_E=3$. Upper solid curve: statistic for unfiltered dataset at time delay suggested by mutual information, $T=9$. Upper dotted curve: statistic for filtered dataset at same time delay, $T=9$.

fact, employ a smaller embedding dimension by going to smaller time delays, which also results in a smaller statistic, implying a ''more deterministic'' mapping. (Noise, a more complicated attractor, and finally high local Lyapunov exponents can mimic the effect of topologically false neighbors.) The transformation ensures that the improvement in the statistic is not illusory, reflecting only correlation between $d+1$ and the previous d vector components. It must be admitted that the normalization in the transformation amplifies noise by effectively taking differences, but we emphasize that one need *not* continue to employ this transformed space in further analysis, prediction, or control of the system. By our special construction that distinguishes our transformation from the standard principal components transformation, the topological properties of this space are the same as those of the untransformed time-delay embedded space with the same time delay—including the presence or absence of dynamically illegal self-intersections. Therefore, if one finds that $d_E=3$ is good at $T=2$ or $T=4$ viewed in the transformed space, one can then use the ordinary time-delay embedded space at $T=2$ or $T=4$, knowing that it has been topologically unfolded. Incidentally, the false strands statistic at $T=9$ computed on the untransformed space also shows a minimum/plateau at $d_E=4$ but at a much smaller level of the statistic, as a result of the lack of noise amplification. The statistic computed below $T=5$ or so on the untransformed space is zero everywhere. The fact that we see $d_E=4$ both with and without the transformation at $T=9$ reinforces the conclusion that $d_E=4$ is genuine and we need not worry that $T=9$ does not still show effects of linear correlation. The message is that examining the statistic on data subjected to the global linear transformation increases confidence in the validity of the result even if the original vector space is employed for later purposes. Finding an embedding that works with three dimensions instead of four may be quite significant, as there are sophisticated topological analyses of attractors based on linking numbers of embedded periodic orbits whose mathematics only works properly in three dimensions.

We turn to another dataset from the same laser operated at different physical parameters, such that the experimental control scheme fails to operate. An off-line Fabry-Perot interferometer shows that there now seem to be three electromagnetic modes in one polarization and two in the other one active in this dynamical system. Again there were 100 000 data points at a sampling rate of 2 MHz. The dynamics appears more complicated—the strong approximate periodicity seen in the previous dataset is no longer obvious. (Fig. 19). Figure 20 shows the false strand results. In comparison to the shuffled data results, we identify this dataset as chaotic, but higher dimensional and noisier than the previous one, as the proportion of false strands never reaches a very small value. The best minimum seems to occur at $d_E=4$ at $T=2$, with the statistic remaining relatively constant or slightly increasing for higher dimensions, as expected with noise. For $T=3$ and $T=4$ the statistic also appears to plateau at $d_E=4$, lending some confidence that the value found at $T=2$ is correct. Minimum embedding dimension is not invariant with time delay, but often does not change drastically. At

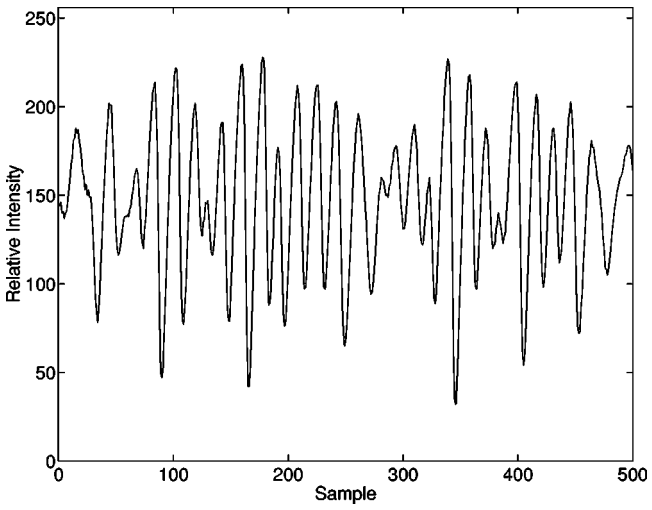

FIG. 19. Sample of time series from uncontrollable Nd:YAG laser dataset.

$T=1$ the amplification of noise in the global decorrelation appears to make results worse.

As with the controllable dataset seen before, these data have a substantial amount of noise, and so we run the algorithm on a low-pass filtered version of the data. Figure 21 shows power spectra for original and filtered signal. Figure 22 demonstrates the improvement that the low-pass filtering has made, as seen by the lower false strand statistic, a minimum statistic of 3.5% false strands, less than half the minimum value seen without filtering. The results now demonstrate evidence of low dimensionality. For further analysis,

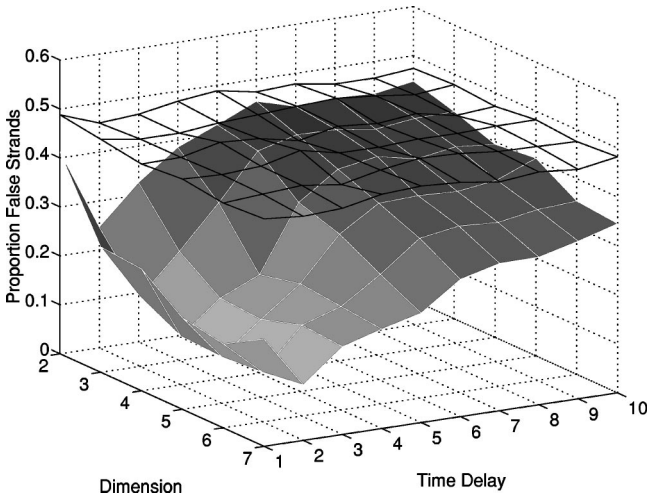

FIG. 20. Opaque surface: Proportion of false strands for intensity fluctuations of the laser, for uncontrollable dataset. The minimum occurs at $d_E=4$, with time delay $T=2$ in units of the sampling interval, which is 0.5 μs. Compared with the controllable dataset, these data require a higher embedding dimension and have more noise. Mutual information time is $T=5$, which does *not* provide a good embedding, as can be seen by the high absolute level of false strands at that time delay, as well as the absence of any clear indication of d_E. Clear surface: statistic with the last vector component shuffled.

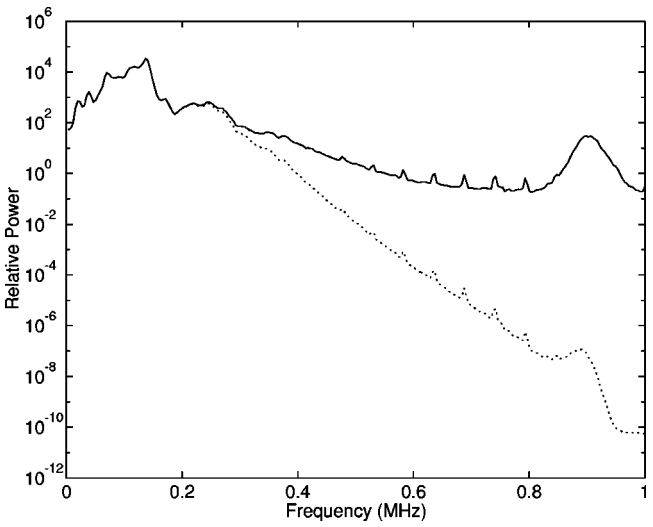

FIG. 21. Power spectral density of original and low-pass filtered laser intensity signal, uncontrollable dataset. Filter eliminates 0.86% of power in original signal.

we would recommend using an embedding dimension $d_E=4$ and time delay $T=2$. The decorrelation global linear transformation is not strictly necessary, and a pure time-delay embedding does not introduce the transformation's noise amplification but has significant autocorrelation. The mutual information criterion [14] recommended the use of $T=5$ which unfortunately did not give a clear indication of the embedding dimension. The freedom in T allowed by our method permitted us to search for an embedding that does show honest evidence of low-dimensionality. Without the principal component transformation we would have been required to stay at $T=5$, which did not show convincing low dimensionality on this realistically difficult experimental dataset. The filtering was not particularly severe, eliminating

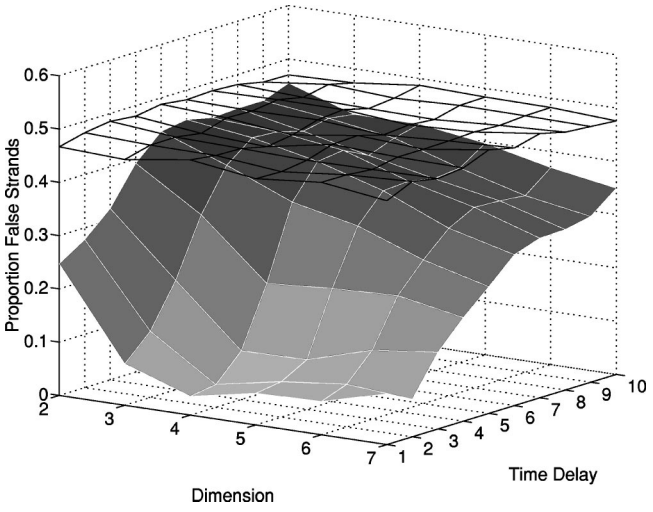

FIG. 22. Opaque surface: Proportion of false strands for intensity fluctuations of the laser, for low-pass filterd version of uncontrollable dataset. Good minima occur at $d_E=4$, with time delay $T=1$ or $T=2$. Clear surface: statistic with last vector component shuffled.

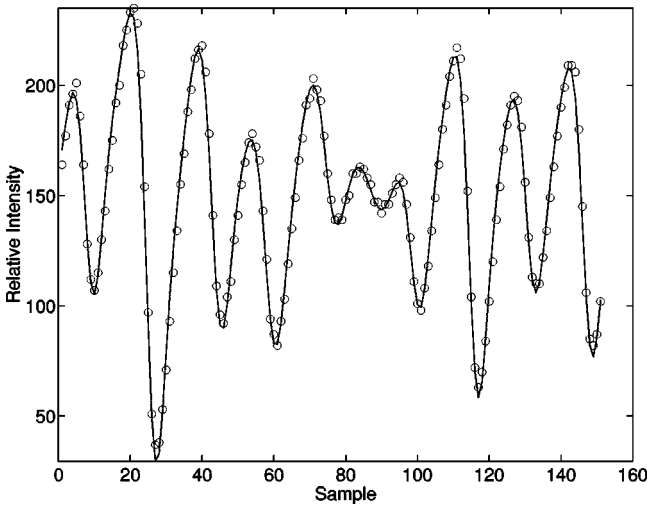

FIG. 23. Closeup of filtered (solid line) and original (circles) uncontrollable laser dataset. The filtering only smooths and does not otherwise change gross characteristics of the time series.

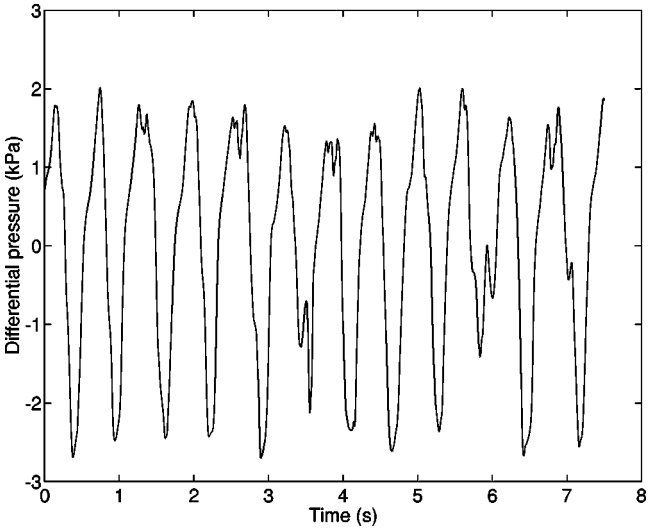

FIG. 24. Fluidized bed pressure time series. There is a strong natural periodicity.

only 0.86% of the power in the original signal. Figure 23 superimposes the filtered and original signals. Because of the properties of the decorrelation transformation we can be assured the improvement in the results was not spurious—an artifact of the filter low dimensionality out of a high-dimensional noise signal The phase-space structures of original and filtered signals appear reasonably similar to the eye.

Filters with either lower or higher cutoff frequencies did not appear to give as good results as that shown previously, considering the best result to be that which gives the deepest minimum in the statistic. Less filtering left too much noise in the signal, and more filtering apparently harmed the reconstructed phase-space structure, resulting in less clear evidence of chaos, and thus a higher statistic. The false strand statistic now provides a nontrivial criterion to choose the type and amount of filtration that gives the most cleanly low-dimensional signal. We emphasize this is ''blind'' noise reduction—we have no specific model for the system—and so it is quite significant to have an independent means to evaluate the success of the filtering. In this regard a simple ''empirical prediction error'' criterion could be fooled; a very severe low-pass filter can result in extremely smooth data that are very predictable in the short term, yet do not have a low-dimensional deterministic embedding. We anticipate that noise reduction methods designed specifically for chaotic data [26] would work better than the simple filter employed here, and that our statistic would be useful as an arbiter of how much cleaning by these algorithms one should employ.

C. Fluidized bed

Our final experimental example is a pressure time series from a fluidized bed experiment. The apparatus was a 2.5 m long hollow tube of 10 cm inner diameter. A quantity of 575 μm micron mean diameter stainless steel bearings was subjected to a regulated upward flow of air from a distributor plate on the bottom of the tube. With sufficient air flow, the mass of particles become suspended and possess ''fluidlike'' behavior, hence the name. With slightly higher air flow, the system demonstrates a number of complicated dynamical regimes, one of which we briefly examined. The experiment itself is an idealized model of particulate transport devices used in industrial applications. The quantity measured was from a differential pressure transducer taken between 23 cm and 35 cm above the air input grate, which defines the bottom. The pressure output at each cross section was the average of four individual sensors distributed around the circumference of the device. The data, 60 000 points in total, was digitized at 12 bits at 200 Hz.

In the regime examined, the data seem to exhibit approximately periodic or chaotic behavior with a clear natural frequency viewed by the eye (Fig. 24). Chaos might be a likely hypothesis, but the results of the false strand statistic surprisingly suggest otherwise. Figure 25 demonstrates a high level of false strands that is nearly completely independent of time delay and embedding dimension. There is no compelling evidence for low dimensionality in these data, with a noisy linear system with a sharp resonance being a reasonable alternative. The overall level is lower than that seen with shuffled data, indicating the presence of some sort of nonlinearity. We note that the Fourier transform surrogate data method [20] rejected its null hypotheses of linearly correlated noise and the same warped with a static nonlinearity on this dataset, implying that the dataset was more predictable to a statistically significant degree. It is not clear whether this is because the data authentically do not fit either null hypothesis, or whether this is because of the finite-frequency resolution flaw inherent in the Fourier transform method when applied to a signal with a significant dominant frequency. The false strand method appears to be more useful in positively identifying low dimensionality when one looks for proper structure with T and d_E to indicate chaos. The older false neighbors method [5] showed ambiguous results that could potentially have been construed as indicating a $d_E=6$ embedding; the strand correction of this paper eliminated that spurious decrease with increasing d_E, and the decorrelation transformation eliminated systematic changes with T.

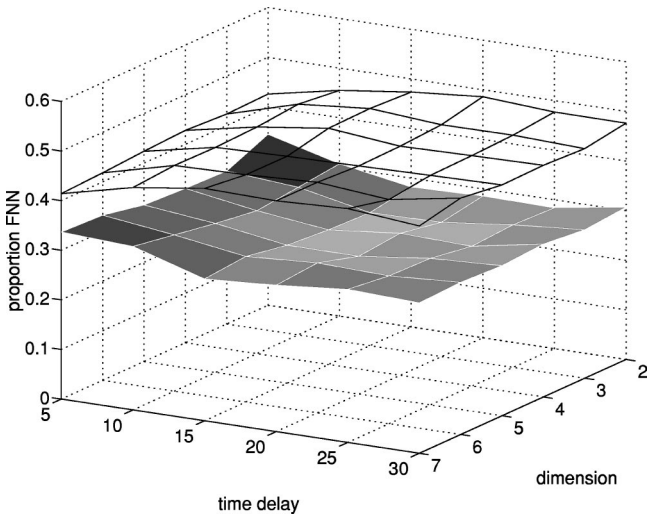

FIG. 25. Opaque surface: False strand statistic for fluidized bed pressure signal. Clear surface: Statistic for shuffled last component. The fact that the opaque surface is below the level of the clear surface indicates the presence of nonlinearity, but the absence of structure in the *T-d* plane and the relative high value of the statistic does not support the hypothesis of low-dimensional chaos. Noisy periodicity is the more likely diagnosis.

VI. DISCUSSION AND CONCLUSION

The essential idea behind our method and nearly all other methods to determine the minimum necessary embedding dimension from data is to identify or quantify the prevalence of local phase-space regions where the crossing of orbits indicates inability to predict where orbits should go—that is, it is an indication that in the embedding dimension chosen, the system is not deterministic.

For example, Casdagli [27] recommends constructing nonlinear empirical predictors that approximate the unknown implied dynamical system using the observed data, and then seeing where the measured average prediction error drops off sharply with increasing embedding dimension. This is a step in the right direction, as it is sensitive to the accidental lack of determinism, but we wish to have a specific and easy method that directly attacks the minimum embedding dimension problem. With complicated prediction functions—designed to give good predictions, but not calculate embedding dimension—one has to consider systematic variations in performance with dimensionality that are independent of the topological question of determining the embedding dimension.

Good empirical predictors are not necessarily the best tests of minimum embedding dimension. Most statistics of this nature, and indeed most alternative methods proposed for determining embedding dimension [28–31], employ various kinds of averaging that seem to obscure the desired result. The first question is whether to compute an average quantity over some neighborhood. If one is interested in good *predictions* then this is beneficial, as it averages out the effect of noise and aids statistical precision in the fitting processes. The consequence of an incorrect embedding is not low-level noise but occasional self-crossings that result in *macroscopic* divergences at future iterations. Including a false neighboring trajectory along with good neighbors will certainly result in a poorer fit overall and hence higher error, but this is a roundabout method of finding incorrect embeddings, as one cannot distinguish this situation from that where the general error level is high due to other reasons, including bad choices of model. There is no chance that a self-crossing would be missed when counting single nearest neighbors, as one examines the entire reference trajectory. The point is that one is interested only in the *extreme events* likely to result from misembedding, not a typical prediction.

There is also an advantage to having an easily interpreted absolute scale for the statistic in contrast to the one where one examines relative changes with trial embedding dimension. That is, convergence to a small proportion of false strands, for example 1%–5%, suggests a good embedding of low-dimensional chaos in contrast to convergence at 20%.

One also would like to eliminate systematic effects that change with dimension. One of the most frequently encountered problems is that the typical distance to nearest neighbors increases substantially as dimension is increased without regard to self-crossing. This means that statistics based on ratios of vector distances will often have a strong bias with dimension, usually resulting in more apparent ''predictability'' as dimension is increased. This fact also makes examining a fixed-sized neighborhood problematic as the relevant distances change with trial dimension. A fixed number of neighbors is thus easier, and if one takes that to its logical conclusion, one ends up examining single nearest neighbors.

Finally, we believe that the problems we examined with oversampled data would show up in any local phase-space statistic. It is not at all clear to us how to correctly account for a high sampling rate with statistics substantially more complicated than false nearest neighbors.

Most published techniques developed to distinguish chaos from colored linear noise have relied on the surrogate data technique (with the exception of time-reversibility methods). An exhaustive list would be excessively long, but we point out Ref. [32] which advocated the use of surrogate data as a backstop against spurious observation of low false neighbors in correlated time series seen in the original false neighbors method of Ref. [5]. There are some undesirable technicalities with surrogate data methods and our direct alternative may be an attractive alternative. Typically surrogates are constructed by inverting a phase-randomized version of the discrete Fourier transform of the signal. In generating surrogate data, the Fourier transform assumes that the signal has a period equal to the length of the dataset, i.e., the signal wraps back around and repeats itself. If the first and last points of the scalar signal do not ''match'' (and usually they do not) the effect is to introduce what looks like a sharp discontinuity which translates into more high frequency power than there should be. If the original data is reasonably noisy then this end effect has little consequence. If, however, the original dataset is smooth (little power in high frequencies), then the surrogate datasets made from the Fourier transform method will have noticeably higher frequency noise because the Fourier transform was sensitive to a high-frequency-containing discontinuity in the signal. If the original data were just very smooth colored noise then this test will spu-

riously reject the null hypothesis because the surrogates will be less predictable than the original due to this excess of higher frequency noise. There is the minor issue that the simplest efficient subroutines for discrete Fourier transform only operate on datasets sized in powers of two, whereas the technique presented in this work has no such restriction.

In many ways the surrogate data technique is often *too* powerful—''power'' in the purely statistical sense—because it can easily sense a deviation from the particular null process embodied in the surrogates, but without giving further insight into the nature of the rejection. The data may be from some sort of noise process with a nonlinearity somewhat different from the Gaussianizing histogram transformations often used in simple null hypotheses [20,21]. Local prediction error statistics used frequently in surrogate data methods are designed to show a distinct difference with low-dimensional chaos, and so a very clear rejection of its null hypothesis over many time delays and a plateau with increasing d often suggests low dimensionality, but in our opinion, not as convincingly as low proportion of false strands in an absolute scale. In surrogate data methods one looks for strong predictability *relative* to appropriately constructed colored noise. There is no way to know just how much more predictable one should be with deterministic chaos. With false strands, we do know that, at least, in principle, the statistic should go to *zero* when we have a deterministic process unfolded in the correct embedding dimension.

In sum, we have demonstrated an algorithm to find the minimum embedding dimension from observed data. It offers a number of advantages over previous approaches, such as follows.

(1) Consideration of effects due to high sampling rate data. Large amounts of well sampled data are often desirable for state-space algorithms and statistics, as they trace out the attractor well. Changes in sampling rate irrelevant to fundamental dynamics can cause systematic changes with many statistical quantities, often giving spurious tendencies toward indicating low dimensionality. The strand concept is a intuitively attractive and experimentally successful countermeasure.

(2) Consideration of effects due to high autocorrelation, which often comes with well sampled data of chaotic flows. The normalized principal component transformation eliminates all linear correlation between components of the state space and thus many systematic convergences of empirical statistics as the time delay gets short. This permits one to explore a wider range of time delays than if one were required to keep T sufficiently large to ensure component wise decorrelation with an ordinary embedding. We have seen examples where time delays smaller than those normally employed appear to give cleaner embeddings at a lower dimension and a more secure indication of low dimensionality. We constructively demonstrate an orthogonal rotation of the standard principal coordinate space that ensures the topological answer found in the transformed space remains valid for the untransformed time-delay embedding space.

(3) The elimination of the systematic tendencies pushing down the old false neighbor statistic for small T allows one to plot our statistic over the entire d_E-T plane. Authentic low-dimensional data show a characteristic behavior, with false strand proportion bottoming out or plateauing once the correct d_E has been achieved, and with the overall level of false strands increasing past the optimal T (because of chaotic decorrelation, and a global increase in the complexity) of the reconstructed attractor, and with the statistic sometimes increasing below the best T as noise is further amplified.

(4) By comparison to results seen when shuffling the last component, the statistic gives a good practical of easily distinguishing chaos from colored noise. It requires much less computation than the standard surrogate data approach. In addition, it successfully distinguishes noisy periodicity and other intentionally tricky highly resonant linear systems from low-dimensional chaos; these may frequently fool surrogate data methods. We do not present results using a more rigorous null hypothesis test, though this test is obvious in construction given the sets of extra distances on original and shuffled data. We feel that as a way of positively identifying low-dimensional predictability instead of a mere rejection of a null, the presence of characteristic behavior in d_E-T as well as a reasonably small absolute value of the statistic is more compelling than just a rejection of a particular null test.

Given the comparative simplicity and low computational resources necessary for this algorithm, we believe it to be a useful tool as a ''first look'' into the nonlinear dynamical characteristics of many kinds of experimental signals, in the way that the Fourier transform is a window into gross frequency domain behavior and other linear features in a system.[1]

ACKNOWLEDGMENTS

We thank members of INLS/UCSD for advice and discussion, mentioning Reggie Brown and Sid Sidorowich among others. M.K. is indebted to Lou Pecora of the Naval Research Laboratory for valuable communication and especially thanks Dave Pritchard of the Center for Nonlinear Studies, Los Alamos National Laboratory for comments on the manuscript. We are grateful to N. Rulkov of INLS for providing useful data. We also thank C. Stuart Daw of Oak Ridge National Laboratory and Nick van Goor of the University of Tennessee for providing the dataset.

APPENDIX: DECORRELATION TRANSFORMATION

With a large $N\cdot(d+1)$ matrix $\mathbf{A}$ of the $\mathbf{y}$ vectors of the dataset, $\mathbf{A}_{ij}=\mathbf{y}(i)_j$, we compute the singular value decomposition with conventional algorithms,

$$\mathbf{A}=\mathbf{U}\mathbf{D}\mathbf{V}^T \tag{A1}$$

with $\mathbf{U}$ and $\mathbf{V}$ orthogonal and $\mathbf{D}$ diagonal. With these, the transformation to normalized principal coordinates is

[1]See Ref. [33].

$$\mathbf{B}=\mathbf{A}\mathbf{V}\frac{1}{N^{1/2}}\mathbf{D}^{-1}, \tag{A2}$$

with $\mathbf{B}$ the $N\cdot(d+1)$ matrix of transformed vectors $\mathbf{z}$. In this framework, the cross correlation requirement is satisfied,

$$\frac{1}{N}\mathbf{B}^T\mathbf{B}=I. \tag{A3}$$

We apply an additional orthogonal rotation designed such that the first d coordinates in the new state space are a function of *only* the first d coordinates of the original delay space. If the transformation to principal coordinates is $\mathbf{B}=\mathbf{A}\mathbf{T}$ then it is easy to show that any additional orthogonal rotation $\mathbf{C}=\mathbf{B}\mathbf{Q}=\mathbf{A}\mathbf{T}\mathbf{Q}$ preserves the identity cross-correlation matrix. For our purposes we need not preserve the other property of the conventional principal coordinates transformation—that the projection of the data along each axis has the maximum variance/power not accounted for by previous axes. It thus suffices to find an orthogonal matrix $\mathbf{Q}$ such that $\mathbf{T}\mathbf{Q}$ has zeros all along the bottom row except for the last column. This ensures that the first d columns of $\mathbf{C}$, the transformed state space, only depend on the first d columns of $\mathbf{A}$, the original time-delayed space. We constructively find such a matrix as follows. Take the $(d+1)\times(d+1)$ matrix $\mathbf{T}=N^{-1/2}\mathbf{V}\mathbf{D}^{-1}$ and extract the last row $\mathbf{x}_j=\mathbf{T}_{d+1,j}$. Then the necessary orthogonal transformation is

$$\mathbf{Q}=H(\mathbf{x}-|\mathbf{x}|\mathbf{e}_{d+1}), \tag{A4}$$

with $\mathbf{e}_{d+1}$ the unit vector in the $(d+1)$th direction, and the Householder matrix [15] $\mathbf{H}(\mathbf{z})=I-2\mathbf{z}\mathbf{z}^T/|\mathbf{z}|^2$. In this coordinate system, it is a fortuitous coincidence that the space formed by the first d components is an orthogonal rotation of the space we would get if we had performed an ordinary normalized principal component transformation on the d-dimensional time-delayed embedded vectors. To sum up, the N by $d+1$ matrix of vectors $\mathbf{B}$ that we actually embed to find false strands is

$$\mathbf{B}=\mathbf{A}\left(\mathbf{V}\frac{\mathbf{D}^{-1}}{\sqrt{N}}\right)\mathbf{Q}=\mathbf{A}\begin{bmatrix}\star & \star & \star & \star\\ \star & \star & \star & \star\\ \vdots & & \ddots & \vdots\\ 0 & 0 & \text{zeros} & \star\end{bmatrix}. \tag{A5}$$

The standard and modified rotation to principal coordinates share a common feature: at least for small time delays, successively higher coordinates correspond to successively higher orders of derivatives [16]. This means that higher-dimensional embeddings processed by a normalized principal component transformation will tend to successively amplify noise and thus increase the proportion of observed false neighbors as the test examines the last, and therefore noisiest component. The $(d+1)$th component of the transformed space is similar to a finite difference approximation of the dth derivative of the signal. We mention that any prerotation one might attempt on the standard principal components space (to distribute the noise across all components) is undone by the rotation in Eq. (A4). This is the price that one pays for ensuring no linear correlation among components and requiring the only last component of the transformed space depends on the last component of the initial space. The transformations in this section cure the problems found in Ref. [17].

[1] J.P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[2] T. Sauer, J.A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).
[3] M. Ding *et al.*, Phys. Rev. Lett. **70**, 3872 (1993).
[4] D.T. Kaplan and L. Glass, Phys. Rev. Lett. **68**, 427 (1992); D.T. Kaplan and L. Glass, Physica D **64**, 431 (1993).
[5] Matthew B. Kennel, R. Brown, and H.D.I. Abarbanel, Phys. Rev. A **45**, 3403 (1992).
[6] J.H. Friedman, J.L. Bentley, and R.A. Finkel ACM Trans. Math. Softw. **3**, 209 (1977).
[7] R.F. Sproull, Algorithmica **6**, 579 (1991).
[8] J. Theiler, Phys. Rev. A **34**, 2427 (1986).
[9] J. Theiler, Phys. Rev. A **41**, 3038 (1990).
[10] J. Theiler, Phys. Lett. A **155**, 480 (1991).
[11] H.D.I. Abarbanel, R. Brown, J. Sidorowich, and Lev Sh. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993).
[12] A.R. Osborne and A. Provenzale, Physica D **35**, 357 (1989).
[13] E.N. Lorenz, Tellus, Ser. A **36A**, 98 (1984).
[14] A.M. Fraser and H.L. Swinney, Phys. Rev. **33A**, 1134 (1986).
[15] G.H. Golub and C.F. Van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, 1989).
[16] J.F. Gibson, J.D. Farmer, M. Casdagli, and S. Eubank, Physica D **57**, 1 (1992).
[17] D.R. Fredkin and J.A. Rice, Phys. Rev. E **51**, 2950 (1995).
[18] H.D.I. Abarbanel and M.B. Kennel, Phys. Rev. E **47**, 3057 (1993).
[19] J. Theiler and S. Eubank, Chaos **3**, 771 (1993).
[20] M.B. Kennel and S. Isabelle, Phys. Rev. A **46**, 3111 (1992).
[21] J. Theiler *et al.*, Physica D **58**, 77 (1992); D. Prichard and J. Theiler, Phys. Rev. Lett. **73**, 951 (1994).
[22] C. Rhodes and M. Morari, Phys. Rev. E **55**, 6162 (1997).
[23] R. Brown, N. Rulkov, and E. Tracy, Phys. Rev. E **49**, 3784 (1994).
[24] N. Rulkov, A. Volkovskii, A. Rodriguez-Lozano, E. Del Rio, and M. Velarde, Int. J. Bifurcation Chaos Appl. Sci. Eng. **2**, 669 (1992).
[25] H.D. I Abarbanel, Z. Gills, C. Liu, and R. Roy, Phys. Rev. A **53**, 440 (1996)
[26] E.J. Kostelich and T. Schreiber, Phys. Rev. E **48**, 1752 (1993).

[27] M. Casdagli, Physica D **35**, 335 (1989).
[28] W. Liebert, K. Pawelzik, and H.G. Schuster, Europhys. Lett. **14**, 521 (1991).
[29] K. Pyragas and A. Cenys, Litov. Fiz. Sb. **27**, 437 (1987).
[30] J. Gao and Z. Zheng, Phys. Lett. A **181**, 153 (1993).
[31] R. Huerta, C. Santa Cruz, J.R. Dorronsoro, and V. Lopez, Phys. Rev. E **49**, 1962 (1994).
[32] R. Hegger and H. Kantz, Phys. Rev. E **60**, 4970 (1999).
[33] FORTRAN90 source code and X86 Linux binary software to perform the false strand calculation is available at ftp://lyapunov.ucsd.edu/pub/nonlinear/fns.tgz and at the AIP Electronic Physics Auxiliary Publication Service, http://www.aip.org/pubservs/epaps.html. See EPAPS Document No. E-PLEEE8-66-021208, which includes EPAPS files Readme.TXT, false_Strands_f90.f90, fns_linux_binary, lor84x.dat, and installation.txt