A unified approach to attractor reconstruction

Louis M. PecoraLinda MonizJonathan NicholsThomas L. Carroll

Citation: Chaos **17**, 013110 (2007); doi: 10.1063/1.2430294
View online: http://dx.doi.org/10.1063/1.2430294
View Table of Contents: http://aip.scitation.org/toc/cha/17/1
Published by the American Institute of Physics

A unified approach to attractor reconstruction

Louis M. Pecora
Code 6362, Naval Research Laboratory, Washington, D.C. 20375

Linda Moniz
Department of Mathematics, Trinity College, Washington, D.C. 20017

Jonathan Nichols
Code 5673, Naval Research Laboratory, Washington, D.C. 20375

Thomas L. Carroll
Code 6362, Naval Research Laboratory, Washington, D.C. 20375

(Received 30 June 2006; accepted 2 December 2006; published online 13 February 2007)

In the analysis of complex, nonlinear time series, scientists in a variety of disciplines have relied on a time delayed embedding of their data, i.e., attractor reconstruction. The process has focused primarily on intuitive, heuristic, and empirical arguments for selection of the key embedding parameters, delay and embedding dimension. This approach has left several longstanding, but common problems unresolved in which the standard approaches produce inferior results or give no guidance at all. We view the current reconstruction process as unnecessarily broken into separate problems. We propose an alternative approach that views the problem of choosing all embedding parameters as being one and the same problem addressable using a single statistical test formulated directly from the reconstruction theorems. This allows for varying time delays appropriate to the data and simultaneously helps decide on embedding dimension. A second new statistic, undersampling, acts as a check against overly long time delays and overly large embedding dimension. Our approach is more flexible than those currently used, but is more directly connected with the mathematical requirements of embedding. In addition, the statistics developed guide the user by allowing optimization and warning when embedding parameters are chosen beyond what the data can support. We demonstrate our approach on uni- and multivariate data, data possessing multiple time scales, and chaotic data. This unified approach resolves all the main issues in attractor reconstruction. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2430294]

What is the requirement for the determination of vector components for the reconstruction of an attractor from time series? This is a problem that has been studied for many years and everyone agrees that the problem consists of finding time delay, embedding dimension, and in the multivariate case which time series to use for each coordinate (although the latter is a much neglected problem). Most work has arbitrarily divided the problem into finding the delay and embedding dimension, separately. This division is the source of many problems. In addition, almost all approaches, and there are many, rely on weak heuristic methods or a choice of arbitrary scales (e.g., what constitutes a false neighbor) which are usually unknown. We view the construction of vectors for attractor reconstruction as a problem in finding a coordinate system to represent the dynamical state. What is mathematically necessary for any good coordinate system is that the coordinates be independent; this requirement is highlighted in Taken's theorem. To this end we develop a statistic to test for general, nonlinear functional dependence called the continuity statistic. This allows us to simultaneously test for delays and the necessity of adding more embedding dimensions with the same statistic. This helps us determine the best coordinates and gives us feedback on how well our reconstruction is progressing. For chaotic systems we also need to know when we are taking delays that are too long. This latter situation is always mentioned, but never dealt with. We establish an undersampling statistic based on the geometric description of the attractor manifold when there are long delays. This latter statistic alerts us when we are using embedding parameters beyond what can be supported by the data. Together the continuity and undersampling statistic tell the practitioner when he/she is doing the reconstruction right—something other statistics do not do.

I. INTRODUCTION

One of the most powerful analysis tools for investigating experimentally observed nonlinear systems is attractor reconstruction from time series which has been applied in many fields.[1,2] The problem of how to connect the phase space or state space vector $\mathbf{x}(t)$ of dynamical variables of the physical system to the time series $s(t)$ measured in experiments was first addressed in 1980 by Packard *et al.*[3] who showed that it was possible to reconstruct a multidimensional state-space vector by using time delays (or advances which we write as positive delays) with the measured, scalar time series $s(t)$. Thus, a surrogate vector $\mathbf{v}(t)=(s(t),s(t+t),s(t+2t),\ldots)$ for $\mathbf{x}(t)$ could be formed from scalar measurements. This is essentially a particular choice of coordinates

in which each component is a time-shifted version of the others with the time shift between adjacent coordinates the same.

Takens[4] and later Sauer *et al.*[5] independently put this idea on a mathematically sound footing by showing that given any time delay τ and a dimension $\Delta \geqslant 2$ box-counting dimension $(\mathbf{x})+1$ then, nearly all delay reconstructions are one to one and faithful (appropriately diffeomorphic) to the original state space vector $\mathbf{x}(t)$. These important theorems allow determination of system dynamical and geometric invariants from time series in principle.

The above theorems are existence proofs. They do not directly show how to get a suitable time delay τ or embedding dimension Δ from a finite time series. From the very start[3] emphasis has been put on heuristic reasoning rather than mathematically rigorous criteria for selecting statistics to determine τ and Δ and that remains true up to the present.[2] But many issues are unsettled. For example, there are no clearcut statistical approaches for dealing with multiple time scales, multivariate data, and avoiding overly long time delays in chaotic data even though these problems are common in many systems and are often acknowledged as important issues in attractor reconstruction.

Various approaches to the τ problem have been autocorrelation, mutual information,[6] attractor shape,[7] and predictive statistics based on various models.[8–10] All these approaches have serious short comings (see Kantz and Schreiber,[2] Grassberger *et al.*,[11] and Abarbanel[1] for critiques).

An early approach to the determination of τ was presented by Liebert *et al.*[12] which also combined a search for the proper embedding dimension that presaged the subsequent false nearest neighbor method (see below). The main idea is that near neighbors may become distant points when the embedding dimension is increased. This is an observation of the topology of embeddings and their projections onto lower dimensions. Leibert *et al.* devised a nearest neighbor statistic using 10 nearest neighbors (an empirically derived number) and examined its change as an embedding dimension is changed. There is some need to mitigate against oscillators in the statistic which is done empirically. However, the statistic is among the earliest to try to determine τ and embedding dimension together in some fashion.

About the same time Grassberger *et al.*[11] approached the problem of reconstruction by allowing the delays to not be fixed, but suggest that a higher-dimensional version of mutual information should be used to capture the longest delay and thereafter the shorter ones. They show how the mutual information will change as dimensions are added to the reconstruction (cf. Ref. 13).

Empirical and heuristic approaches to τ determination continue to the present. For example, the recent work by Garcia and Almeida[14] who after appropriately criticizing mutual information approaches go on to use a nearest neighbor measure to allow a determination of τ. They allow τ to vary when a different embedding dimension is considered in an attempt to optimize their result although τ is kept the same for all coordinates.

To determine Δ Kennel *et al.*[15,16] developed false nearest neighbor (FNN) statistics (cf. Ref. 12). These statistics require one to choose an arbitrary threshold which ignores all structures under that scale which may be considerable for many attractors. Furthermore, in chaotic systems the statistic can be skewed by the divergence of nearby points on the attractor or the existence of two time scales so that the procedure may not truly terminate. That is, given a chaotic time series additional delays will eventually cause the FNN statistic to increase from trajectory divergences alone, leading to the incorrect addition of embedding dimensions. In addition, in time series with vastly different time scales the FNN statistic may achieve a minimum when the initial delays are equal to the fastest time scale incorrectly suggesting an embedding dimension lower than the true dimension.

Cao[17] has suggested a scale-free FNN approach, but this struggles with the fact that time-delay embeddings of any signal (even noise) will cause the statistic to reach an asymptote to values similar to those at the same level as deterministic systems. One should also compare Fraser's paper[13] which attempts to go beyond a two-dimensional statistic by using redundancies. The latter paper by its own admission has difficulties with computation with probability distributions when the embedding dimension increases.

The problem of a long time window in chaotic systems is often mentioned in passing, but only with simple admonitions that somehow the product $\tau\Delta$ should not get "too large." This is an acknowledgment of the above-mentioned FNN problem with diverging chaotic trajectories. This problem is pointed out in many early papers and one should consult[2] references therein for more information. Such problems point to the fact that until now there have not been statistics to check on overly long embedding times or to guide the user on the quality of the attractor reconstruction.

For a good overview of the geometry of attractor reconstruction and the effects of various choices of τ and Δ along with the effects of noise on the reconstruction one should consult the papers of Casdagli *et al.*[18]

Finally, very little has been done with multivariate time series. Usually data analysis simply extends what is done for univariate time series[8,19] thus retaining the shortcomings. There are very few tests for optimal choices of time series to use from a multivariate set (outside of eliminating linear dependence using singular value decomposition[20]) or the Gamma test of Jones.[21] Time series with multiple time scales pose another difficult and unaddressed problem (an exception is the more recent work in Ref. 9, although those approaches rely on specific models).

We claim that the above approaches to attractor reconstruction often artificially divide the problem into two problems thereby causing more difficulties than necessary while failing to address typical problems of multiple time series, overembedding[22,23] and multivariate data. In this paper we show that only one criterion is necessary for determining embedding parameters. This single criterion is the determination of functional relationships among the components of $\mathbf{v}(t)$ (the reconstruction vector). This allows us to find τ and Δ simultaneously and deal with the unsolved problems of multivariate data, excessively large $\tau\Delta$ and multiple time scales. For this reason we refer to our approach as a unified approach to attractor reconstruction. Our method does not

rely on the choice of a threshold or some number of nearest neighbors, but rather requires the practitioner to choose a level of confidence for his or her application of the attractor reconstruction. Each level of confidence will be dependent on the application and context and is independent of the method presented here.

II. CONTINUITY AND UNDERSAMPLING STATISTICS

A. Continuity or function statistic

Takens[4] showed that, generically, $s(t+\tau)$ is functionally independent of $s(t)$. The requirement of any embedding or attractor reconstruction is that one must construct vectors which have independent coordinates. We take this to be the central criterion since without it there is no embedding or reconstruction. This criterion is explicitly stated in Takens theorem[4] (p. 369) and echoed again in the embedding theorems of Sauer *et al.*[5] (p. 582) which extend the embedding theorems to arbitrary compact invariant sets with mild conditions. Thus, the previous observations about delays and functional independence enables the embedding.

With this in mind we can provide a general requirement which quantifies the functional independence of reconstruction coordinates. Consider a multivariate time series data set $\{s_i(t)\,|\,i=1,\ldots,M\}$ sampled simultaneously at equally-spaced intervals, $t=1,\ldots,N$. Then for each $\mathbf{v}(t)$ component we can sequentially choose among M time series and various delays (not necessarily equal). Thus, suppose we have a $\mathbf{v}(t)$ of d dimensions $\mathbf{v}(t)=(s_{j_1}(t+\tau_1),s_{j_2}(t+\tau_2),\ldots,s_{j_d}(t+\tau_d))$, where the $j_k\in\{1,\ldots,M\}$ are various choices from the multivariate data and, in general, each τ_k is different for each component; usually $\tau_1=0$. To decide if we need to add another component to $\mathbf{v}(t)$, i.e. increase the dimension Δ of the embedding space, we must test whether the new candidate component, say $s_{j_{d+1}}(t+\tau_{d+1})$ is functionally independent of the previous d components. Mathematically we want to test the equality,

$$s_{j_{d+1}}(t+\tau_{d+1})\underset{?}{=}f(s_{j_1}(t+\tau_1),s_{j_2}(t+\tau_2),\ldots,s_{j_d}(t+\tau_d)), \qquad (1)$$

for any function $f:\mathbf{R}^d\to\mathbf{R}^1$. Equation (1) is the rigorous criterion for generating new $\mathbf{v}(t)$ components and is a general requirement for independent coordinates.[24] Using this approach we continue to add components to $\mathbf{v}(t)$ until all possible candidates for new components from all time series and for all τ values are functions of the previous components. In this way we have found Δ and all delays *simultaneously* and are done. Note, we have not separated the τ and Δ problems—they are found together. Below we develop a statistic to select $s_{j_{d+1}}$ and τ_{d+1} that fulfills the independence criterion.

In the case of embedding of finite data, a distinction must be made between the dimension of the dynamical system that produces the data (and an appropriate embedding dimension for it) and of the reconstructed object that the *data* describes. The aim of the procedure described here is both to make that distinction and to embed the data so that the reconstruction is in the maximum dimension *guaranteed by the data to exhibit independent coordinates*, but no higher or lower. The procedure described herein also excludes the delays that are explicitly forbidden by the embedding theorems, that is, delays equal to the period or twice the period of a periodic orbit.

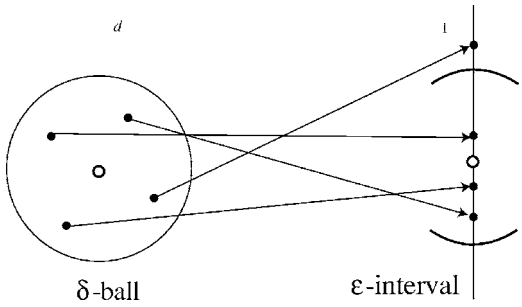

FIG. 1. The δ and ε sets and their data points in R^d and R^1, respectively. Of the k points ($k=4$, here) in the δ ball l ($l=3$, here) are mapped into the ε interval and 1 is mapped outside. The probability of interest by which we accept or reject the null hypothesis is the cumulative binomial for getting three or more points in the ε interval with probability p (see text).

B. Continuity statistic

The expression of a mathematical property [Eq. (1)] as a statistic is a necessary step if we want to analyze data for such a property. This inevitably involves a heuristic step and the introduction of extra parameters or thresholds which are unavoidable. Our objective is to make the heuristic a strong one in that the statistic should capture the mathematics and provide a confidence level. Thus we aim for a test for functional independence at some maximal or optimal level. We do this in analogy with the linear case (for example, singular value decomposition) in which we can find many linearly independent coordinates, but we choose those whose independence is maximum in some sense.

We build our test statistic for functional relationships [Eq. (1)] on the simple property of continuity. It captures the idea of a function mapping nearby points in the domain to nearby points in the range yet assumes nothing more about the function. We use a version of a continuity statistic[25,26] with a new null hypothesis[27] that is more flexible in the presence of noise.

For example, suppose $\mathbf{v}(t)$ has d coordinates and we want to test if we need to add another $d+1$st component. Equation (1) is a test for evidence of a continuous mapping $\mathbf{R}^d\to\mathbf{R}^1$ [Eq. (1)]. Following the definition of continuity we choose a positive ε around a fiducial point $S_{j_{d+1}}(t_0+\tau)$ in $\mathbf{R}^1$ (τ is fixed for now). We pick a δ around the fiducial point $\mathbf{v}(t_0)=(s_{j_1}(t_0+\tau_1),\ldots,s_{j_d}(t_0+\tau_d))$ in $\mathbf{R}^d$ corresponding to $s_{j_{d+1}}(t_0+\tau)$. Suppose there are k points in the domain δ set. Of these suppose l points land in the range ε set. We invoke the continuity null hypothesis that those l points landed in the ε set by chance with probability p. This is shown schematically in Fig. 1. A good choice for data which we use here is $p=0.5$, i.e., a coin flip on whether the points are mapped from δ into ε. Other choices are possible,[25,27] but $p=0.5$ is actually a standard null, harder to reject and very robust under additive noise. Now we pick the confidence level α at which we reject the null hypothesis. If the probability of getting l or more points in the ε set (a binomial distribution)

TABLE I. The number of points from the δ set that must map into the ε set to reject the null hypothesis at the 95% confidence level (α=0.05). The number of δ points and ε points are the same from 5 to 7, after which the number of δ points can be less than the number of ε points, i.e., one or more δ points can be mapped out of the ε set and the null hypothesis can still be rejected. Below 5 points the significance is less than 95% and the null hypothesis cannot be rejected.

Number of δ points	Number of points in the ε set	Number of points not in the ε set
5	5	0
6	6	0
…	…	…
8	7	1
9	8	1
10	9	1
11	9	2
12	9	3
13	10	3

is less than α, we reject the null. For example, if we choose α=0.05, this insures that there are at least 5 points in both δ and ε balls to reject the null hypothesis. Table I shows the number of points in the δ set that must be in the ε set to reject the null hypothesis at the α=0.05 level. Note that not requiring all δ points to map into the ε set makes the statistic less susceptible to noise or other measurement problems. We comment more on choosing confidence levels below. We note here that an essential part of the reconstruction process is to report not only delays and embedding dimension, but the null hypotheses and the confidence levels used. We do this throughout.

We repeat the above process by decreasing ε and varying δ until we cannot reject the null. Call the smallest scale at which we can reject the null our continuity statistic ε^*. We sample the data sets at many fiducial points and calculate the average $\langle\varepsilon^*\rangle$. The data determines ε^*, the smallest scale for which we can claim a function relationship. If the time series is too short to support an embedding, this will be reflected in large and uniform ε^* for any delay.

We suggest that a way to view this statistic is analogous to how we look upon tests for linear independence. If we examine normalized covariances of separate time series, we rarely see either 1's or 0's, but rather numbers between 0 and 1. The closer to 0 a covariance is the less linearly dependent are the two associated time series. In this more general case, the larger $\langle\varepsilon^*\rangle$ the more functionally independent are the two coordinates.

If we succeed in reducing $\langle\varepsilon^*\rangle$ by adding components at proper delays we will know we are doing well in reconstructing the attractor. When we can no longer reduce $\langle\varepsilon^*\rangle$, we are done—this is the best we can do with the given data set. Specifically, we apply the continuity test sequentially, and start with one component and build up $\mathbf{v}(t)$ one dimension at a time by examining the continuity statistic $\langle\varepsilon^*\rangle$ as a function of delay and/or data set. If possible we choose τ at a local maximum of $\langle\varepsilon^*\rangle$ to assure the most independent coordinates [as in Eq. (1)]. If $\langle\varepsilon^*\rangle$ remains small out to large τ, we do not need to add more components; we are done and $\Delta=d$. In the multivariate case we can extend this to generating an $\langle\varepsilon^*\rangle$ for each potential new component $s_j(t+\tau), j=1,\ldots,M$ of $\mathbf{v}(t)$ and choose that component which has the least functional relationship [a maximum of $\langle\varepsilon^*\rangle$] to the previous d components.

C. Comments on choosing a confidence level

All statistics and tests including those for finding reconstruction parameters will require some threshold or numerical parameters to be chosen. These can be deciding on the number of points to examine locally, setting a FNN threshold, or choosing a method of averaging quantities that mitigates against spurious behavior. However, the amount of ambiguity over these choices can be greatly diminished by choosing a null hypothesis that directly captures the central mathematical property to be tested and setting a confidence level for rejecting or accepting that null hypothesis. This latter approach is better simply because it can be evaluated in terms of probabilities rather than absolute quantities which give no inkling of confidence level. That is, one has a gauge into how likely the results are to be valid.

In our case the null hypothesis and confidence level α determine the number of local points necessary to reject the null hypothesis and accept that there is a mapping at some ε^* level (see Table I). It is important to report the null hypothesis and the confidence level along with the ε^* values. We do this throughout.

Choosing an α value *a priori* is generally not possible. The usual choices are 95% or 99% confidence levels, α =0.05 and α=0.01, respectively, since these are standards in statistics research as they relate to standard deviations in normal probability distributions. However, in some cases one can do a risk analysis or use decision theory to determine α. That is, one knows the rewards and penalties for choosing correctly or incorrectly regarding the null hypothesis. By choosing a desired reward lower bound or penalty upper bound we can calculate what value of α we should use. Such considerations are common in medical, engineering, and communications decisions and analysis.[28] For more information on decision theory one should consult Jaynes.[29]

Our purpose here is not to enter into decision theory since that takes us beyond attractor reconstruction and will be specific to each application and its context. Instead we note that if a risk analysis is available, then it will provide a confidence level that we can immediately use in the determination of the best (in the sense of the risk) reconstruction. To put this another way, we push the choice of the "cutoff" parameter α further up to the user of the method since that is where a confidence level can be sensibly chosen. We contend that this is an advantage to our approach along with the use of a statistic close to the original mathematical requirement.

D. A self-consistent test, the undersampling statistic

In principle the continuity statistic should be the only test we need since it not only determines a good set of $\mathbf{v}(t)$ components, but also a "stopping point" when there are no more independent components, hence automatically giving Δ

and τ together. But real data are finite in number and resolution. Because one cannot get nearest neighbors arbitrarily close for finite data, eventually, with large enough delays any two reconstruction vectors from a chaotic time series will necessarily have some components appear randomly scattered on the attractor. This phenomenon is endemic to all chaotic systems. Mathematically the problem is that in looking at points that are distant in time we are essentially looking at points distributed on a high iteration of the flow or map. This indicates the manifold of the system is very folded and contorted and we are undersampling it. Thus, one cannot discern a loss in functional relation from an undersampled manifold both of which give a large $\langle \varepsilon^* \rangle$. It is easy to see this effect in a simple one-dimensional map like the logistic map. Kantz and Olbrich[22] showed a similar overembedding phenomenon for simple maps in high dimensions.

We developed an undersampling statistic to detect when $\mathbf{v}(t)$ components enter this undersampling regime. We use the null hypothesis that at least one of the $\mathbf{v}(t)$ components is randomly distributed on the attractor. We test to see if component-wise difference between a point $\mathbf{v}(t_0)$ and its nearest neighbor $\mathbf{v}_{\mathrm{NN}}$ in $\mathbf{R}^{d+1}$ (the combined δ and ε spaces) is on the order of typical distances between randomly distributed points on the attractor. To do this test we generate the baseline probability distribution of differences between randomly chosen points from each time series. It is easy to derive the distribution $\sigma_j(\xi)$ of differences ξ between randomly chosen points from the probability distribution $\rho_j(x)$ of values in the jth time series. We calculate $\rho_j(x)$ for each time series $j=1,\ldots,M$ (e.g., by binning the data). Then it is easy to show under the assumptions of randomly chosen points that $\sigma(\xi)=\int \rho(x)\rho(\xi-x)dx$. Then for each component i of the difference between a point $\mathbf{v}(t_0)$ and its nearest neighbor $\mathbf{v}_{\mathrm{NN}}-\mathbf{v}(t_0)$ the probability of getting a value less than or equal to $\xi_i=|(\mathbf{v}_{\mathrm{NN}}-\mathbf{v}(t_0))_i|$ (the NN distance for the ith component) at random from the j_ith time series is $\Gamma_i = \int_0^{\xi_i} \sigma_{j_i}(\zeta)d\zeta$. Let $\Gamma=\max\{\Gamma_i\}$. As with the continuity statistic we average Γ over fiducial points on the reconstruction.

We choose the confidence level β for rejection or acceptance of the undersampling null hypothesis with the comments of the previous section in mind. We often choose the standard level for rejection at $\beta=0.05$, so that if $\Gamma \leqslant 0.05$ we reject the null hypothesis and accept that the time delays are not too large so that the average distances do *not* appear to be distributed in a random fashion. We do this at times in the following examples, but we also display another approach that is an advantage of our unified method. This is to monitor Γ versus τ as we add reconstruction components (dimensions) and when Γ increases precipitously we stop. The main point is that we report not only on the reconstruction parameters (Δ, τ_i, and for multivariate time series which components were used for the reconstruction vector), but also what the null hypothesis rejection levels are. Thus, we might say we have a two-dimensional reconstruction that uses time series #1 at a delay of 0 and time series #2 at a delay of 15 with an $\alpha=0.05$ and a $\beta=0.1$. We remark more on this in the conclusions.

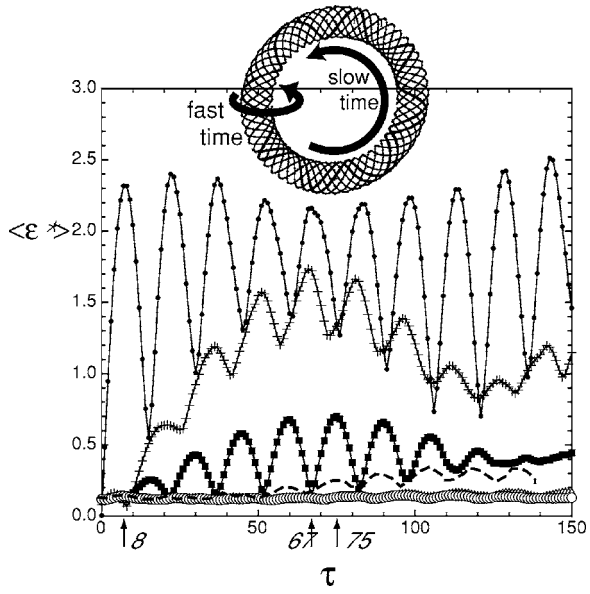

FIG. 2. Continuity statistic for a quasiperiodic system. Inset: The torus for the quasiperiodic system with two time scales. The continuity statistic for the quasiperiodic system for the embeddings 0 (–●–); 0, 8 (–|–); 0, 8, 67 (–■–); 0, 8, 67, 75 (–▽–); and 0, 8, 67, 75, 112 (–○–); and the constant $\tau=8$ embedding (---) using eight dimensions. The first four delays are noted in the figure.

III. APPLICATIONS

We now present the continuity statistic $\langle \varepsilon^* \rangle$ and the undersampling statistic Γ for some typical systems of increasing complexity using univariate and multivariate data. Our notation for labeling different $\langle \varepsilon^* \rangle$ results for Eq. (1) is to simply label the current delays in $\mathbf{v}(t)$. So 0,10 labels the $\langle \varepsilon^*(\tau) \rangle$ which is testing for a functional relation from the two-dimensional $\langle \varepsilon^* \rangle$ (the time series and a ten-step advanced shifted version) to the next possible third-dimensional component time shifted by τ. All time series are normalized to zero mean and standard deviation of 1. For all our statistics our data points are gathered using a temporal exclusion window[30] or using strands in place of points[15] to avoid temporal correlations.

A. Quasiperiodic, multiple time scale system

A nontrivial case is a quasiperiodic system with different time scales. Figure 2 inset shows the torus of this system. The slow and fast times are in the ratio of $2.5\pi:1$ (approximately 8:1) and the time series is sampled at 32 points per fast cycle. The continuity statistic shows only four dimensions are needed since $\langle \varepsilon^* \rangle$ falls to and remains at a low level after adding the $\tau_4=75$ component. This is correct since any 2-torus can be embedded in four dimensions or less. The two time scales are correctly captured in the ratio of $\tau_2:\tau_3$. In comparison, using the standard constant delay from the first minimum of the mutual information takes more than eight dimensions to embed the torus in the sense that both continuity statistics (indicating that independent coordinates can still be added) or false near neighbor statistics (indicating

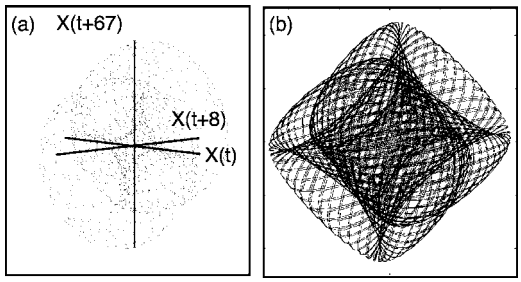

FIG. 3. The attractor reconstruction using delays 0, 8, and 67. Part (a) is just points and part (b) are connecting trajectories through those points.

that the torus has been completely unfolded) do not reach their respective terminal values in fewer than eight dimensions using the standard constant delay.

In Fig. 3, we plot three of the components of the torus attractor using delays of 0, 8, and 67. The attractor fills the space equally in all directions and shows only a slight self-intersection (hence the need for a fourth dimension with $\tau_4 = 75$). In Fig. 4 the reconstruction uses the standard constant delay of the first minimum of the mutual information resulting in delays of 0, 8, and 16. Note, that in some views [Fig. 3(b)] the attractor is rather flat and more self-intersection appears. This is consistent with the fact that $\langle \varepsilon^* \rangle$ does not decrease until we have added many more components.

B. Lorenz attractor reconstruction

We tested the unified approach on a chaotic three-dimensional Lorenz system[31] with $\sigma = 10$, $b = 8/3$, $\rho = 60$. The x time series was generated using a fourth-order Runge-Kutta solver with a time step of 0.02 and 64000 points. We calculated $\langle \varepsilon^* \rangle$ by averaging ε^* over 500 random points on the reconstructions. The results are shown in Fig. 5. This system is chaotic with $\langle \varepsilon^* \rangle$ eventually increasing with τ because of undersampling so we add the undersampling statistic, Γ, at the bottom of the figure.

A three-dimensional $\mathbf{v}(t)$ significantly lowers the $\langle \varepsilon^* \rangle$ value out to near 300 time steps where it begins to rise. If we try to add another $\mathbf{v}(t)$ component, for example, with a $\tau = 350$, the undersampling statistic rises drastically and goes above the 5% confidence level indicating that the increases in $\langle \varepsilon^* \rangle$ result from a folded manifold that is undersampled. Nonetheless, the time window over which we have a good embedding is rather wide, about 300 time steps. Using the

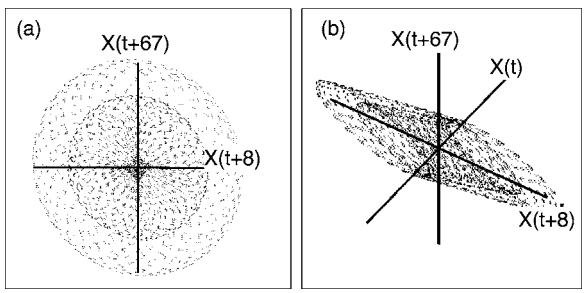

FIG. 4. The attractor reconstruction using delays 0, 8, and 16 as would be done in a constant delay reconstruction.

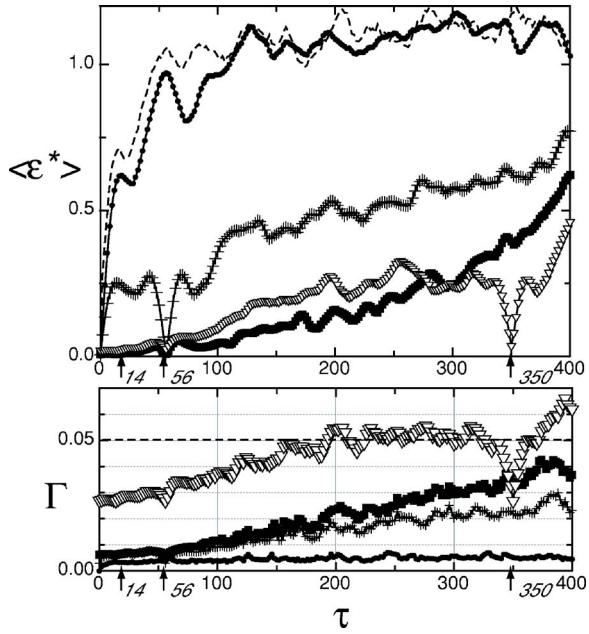

FIG. 5. Continuity $\langle \varepsilon^* \rangle$ and undersampling Γ statistics for Lobster time series plus $\langle \varepsilon^* \rangle$ when 16% Gaussian white noise is added to time series (---). Advances: 0 (–●–); 0, 56 (–|–); 0, 56, 14 (–■–); 0, 56, 14, 350 (–▽–). The delays selected are shown along the τ axes.

first minimum in the mutual information ($\tau \approx 16$) would require a $\mathbf{v}(t)$ of about 16 components to accomplish the same reconstruction with constant τ embeddings in the sense that independent coordinates are not achieved with fewer than 16 time delays. In other words, the constant τ embedding does not unfold the attractor as efficiently. The continuity statistic provides confidence that details in the attractor down to a scale of 0.2 attractor standard deviation are real. This detail can be lost using arbitrary thresholds of other tests (e.g., mutual information and FNN). The dashed line shows the effect of Gaussian white noise of 16% of the time-series standard deviation added to the data. Despite this high noise level much of the $\langle \varepsilon^* \rangle$ structure remains. This noise robustness comes from our choice of the $p = 0.5$ probability for the $\langle \varepsilon^* \rangle$ null hypothesis. Similar results occur from shorter time series.

The dimension suggested by the unified embedding procedure is less than that deemed to be sufficient by the embedding theorems, that is, more than twice the box dimension of the attractor. This fact would seem to indicate that the unified embedding is somehow amiss, but it proves to illustrate a point. The embedding theorems give a *sufficient* dimension for a generic mapping to be an embedding. Indeed, Sauer *et al.*[5] (p. 593) indicate that a smaller dimension (than twice the box dimension) may be entirely sufficient for genericness on most of the attractor, and give the *maximum* dimension of the self-intersection set. The procedure described here minimizes self-intersections by finding independent coordinates and gives a clear indication that independent coordinates are not possible. In the Lorenz example, this was not the case. The data determine the dimension of the embedding in this procedure by determining the number of possible independent coordinates.

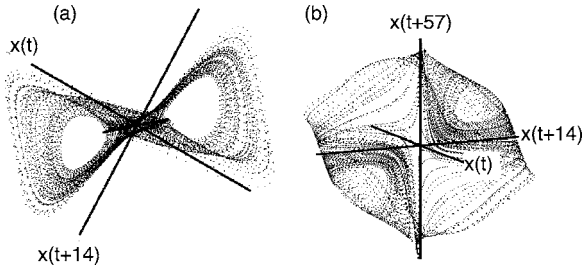

FIG. 6. The Lorenz attractor using delays 14 and 56.

Figure 6 shows the Lorenz attractor from two different views. In Fig. 6(a) we see the "standard" double-winged attractor shape. In Fig. 6(b) we see that the third dimension is not flat, but more extended than plots of the three Lorenz components (x,y,z) because of the use of a longer delay. Figure 6(b) displays a third component that is more independent of the first two components.

C. A multivariate test case

For the case of multivariate time series using all three of the Lorenz system's components, x, y, and z we calculated $\langle\varepsilon^*\rangle$ for all possible combinations of components allowing for time delays or advances (±100 time steps). The results imply a surprising conclusion. The best set of components for the three-dimensional $\mathbf{v}(t)$ is $[x(0),x(17),z(0)]$, i.e., no y component. We emphasize that the "best" here means (in the context of Sec. II) the most mutually independent which is our statistical criterion. The $\langle\varepsilon^*\rangle$ values for the y time series are below those of the x time series at almost all delay values. On reflection we realize that there is no reason why the variables of the physical equations of motion are best for attractor reconstruction.

The numerical tests were done as follows. We started with three cases (x, y, or z) as the first component of the reconstruction vector. We then examined all possible mappings from each of those cases to (x, y, or z) at advanced and delayed times from −100 to +100 times steps at 0.02 (in Lorenz system time units) for each step averaging ε^* at 400 random points on the attractor for each delay value. After choosing a second component $[x(t+17)]$ we examined all mappings from these two to (x, y, or z) at advanced and delayed times as above. From these we chose $z(t)$. Computationally such a detailed statistical study takes on the order of 10 h on a 1.25 GHz computer. This is certainly not an exhaustive test, since combinatorically more possibilities exist. It is in fact a version of a greedy algorithm,[32] but we expect it to be almost optimal because of the low dimension of the Lorenz system.

D. An optimization approach to the reconstruction

Thus far we have taken the simple direct route of determining the delays sequentially by choosing relative maxima. This can be characterized as a form of greedy algorithm.[32] A more general and effective way to calculate the proper delays for a time series is to embed the signal in d dimensions, and choose all d delays simultaneously by numerical optimization. In this optimization procedure, the time series is embedded in d dimensions, and the statistic $\langle\varepsilon^*\rangle$ is calculated for the next $d+1$st component at each delay from 1 to 100. Since the goal of the unified approach is to minimize the continuity statistic we average the $\langle\varepsilon^*\rangle$ values for the 100 delays. This delay-averaged quantity $\langle\langle\varepsilon^*\rangle\rangle$ can be thought of as the average continuity of the d-dimensional embedding to the next $(d+1$st) component. This calculation is then done using various d values while monitoring the undersampling statistic.

We chose a downhill simplex method, with simulated annealing[33] (there were many local minima) for the numerical minimization. In order to speed the calculation, rather than vary the radius δ of the d-dimensional embedding in $\mathbf{R}^d$, a fixed number of neighboring points on the d-dimensional embedding was chosen (40 in this case). From the binomial distribution, if there is a 50%, probability of any particular point landing in the ε set, there is a 5% chance of having 25 points all land in the ε set, so in this routine, ε is the radius on the $d+1$st component which contains 25 of the 40 points from the δ neighborhood on the d-dimensional embedding.

The minimum value of $\langle\langle\varepsilon^*\rangle\rangle$ was calculated from a 30 000 point time series from the Lorenz system, with a time step of 0.03. For the calculation, 10 000 fiducial points were randomly chosen from the 30 000 point time series. For a three-dimensional embedding, the minimum $\langle\varepsilon^*\rangle$ was 0.561, at delays of (0, 54, 34). For a four-dimensional embedding, the minimum $\langle\varepsilon^*\rangle$ was 0.510, at delays of (0, 52, 35, 68). The optimized approach does show that adding a fourth component is probably warranted although the improvement in the reconstruction is not big. We can compare these with our greedy algorithm given above by multiplying the latter's delays by 2/3 since the Runge-Kutta time step for the latter was 0.02 while here we used 0.03. This gives (integer) delays of (0, 37, 9) in step sizes of 0.03. One pair of delays (37 and 35) agree, but the greedy algorithm resulted in the addition of a short delay whereas the optimized version suggests adding components to the reconstruction vector using larger delays. There were actually many local minima seen using the optimization method, so it is not surprising that different algorithms give different results. There is more than one "good" set of delays for reconstruction, just as there is more than one "bad" set of delays. The optimization approach is more rigorous, but simply choosing local maxima on the $\langle\varepsilon^*\rangle$ plot works almost as well if there are obvious maxima.

E. Neuronal data

We applied these statistics to neuronal data taken from a lobster's stomatogastric ganglia[34] in Fig. 7. These neurons are part of a central pattern generator and exhibit synchronized bursting similar to Hindmarsh and Rose models.[34,35] We find that the time delays $(\tau=0,278,110,214,15)$ range from 15 to 278, a factor of nearly 20. Five dimensions are required for a reconstruction at an undersampling statistic confidence level of $\beta=0.1$. This agrees qualitatively with a detailed model of the ganglia by Falcke *et al.*[36] Using a constant delay established by the 1st minimum of the mutual information $(\tau=32)$ five dimensions still does not capture the

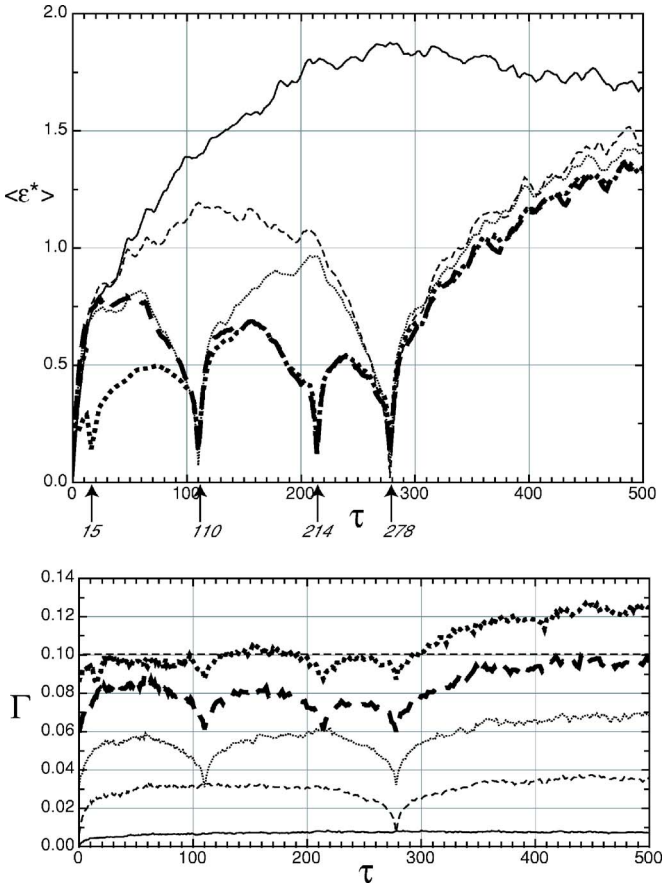

FIG. 7. Continuity $\langle \varepsilon^* \rangle$ and undersampling Γ statistics for Lobster time series. Advances: 0 (—); 0, 278 (---); 0, 278, 110 (···); 0, 278, 110, 214 (—·—); 0, 278, 110, 214, 15 (-·--). The location of delays 15, 110, 214, and 278 are shown.

long time scales beyond $\tau=160$, whereas we have structure out to near $\tau=300$. Adding larger delays causes the undersampling statistic to go much higher than 0.1 and we decided to stop with the five delays we have, but we always note what confidence level we used. Figure 8 shows a plot of three of the neuronal reconstruction vector components using 0, 15, and 278. In Fig. 8(a) the spikes are well-resolved into open trajectories that captures their dynamics. From the angle of Fig. 8(a) the long-time undulations appear as a tail because at the time scales of this view they are highly correlated. However, Fig. 8(b) includes a better view capturing the delay of 278. This long delay opens up the tails and exposes the variations of the slow time part of the dynamics. One can see how the spikes "ride" on top of the undulations.

IV. CONCLUSIONS

Our unified approach uses statistics which faithfully adhere to the rigorous mathematical conditions of the Takens' embedding theorem $\langle \varepsilon^* \rangle$ and the accurate geometric picture of excessive time delays (Γ). Both statistics give good indications when we are have a good reconstruction $\langle \varepsilon^* \rangle$ and when we have gone beyond what the data will allow (Γ), e.g., undersampling. Hence, unlike other reconstruction approaches we have an indication of the quality of the embedding at a given level of confidence.

As we mentioned the $\langle \varepsilon^* \rangle$ statistic is a generalization of the concept of linear correlation, but note that it is asymmetric in general. For example, given two time series, say $x(t)$ and $y(t)$, we might get very small $\langle \varepsilon^* \rangle$ for the test of functionality $x(t) \rightarrow y(t)$ indicating $y(t)$ is a function of $x(t)$, but get a large $\langle \varepsilon^* \rangle$ for the reverse relationship $y(t) \rightarrow x(t)$ indicating that $x(t)$ is not a function of $y(t)$. A simple example is $y(t)=x^2(t)$.

The statistics $\langle \varepsilon^* \rangle$ and Γ are data dependent as all statistics should be. Adding more time series points will lower the continuity statistic and improve the undersampling statistic so that the acceptable time window for an embedding will enlarge and more $\mathbf{v}(t)$ components will not be needed. The statistics replace arbitrary time, length, and point-count thresholds with probabilities so that physical scales (which are often unknown) are derived and data-dependent, but not assumed at the outset.

The data-dependent method has the advantage (over current methods) of yielding unambiguous information about the ability to embed a data set at all. If it is not possible to choose successive delays so that the continuity statistics fall, the data set does not include enough information to produce independent embedding coordinates.

We have chosen "standard" statistical confidence levels here of $\alpha=0.05$ and $\beta=0.05$ (two sigmas) in some cases. In others we stop the reconstruction when the undersampling statistic falls precipitously and give the reconstruction parameters along with the level for null hypothesis rejection. In all problems with statistics choosing the level for the null

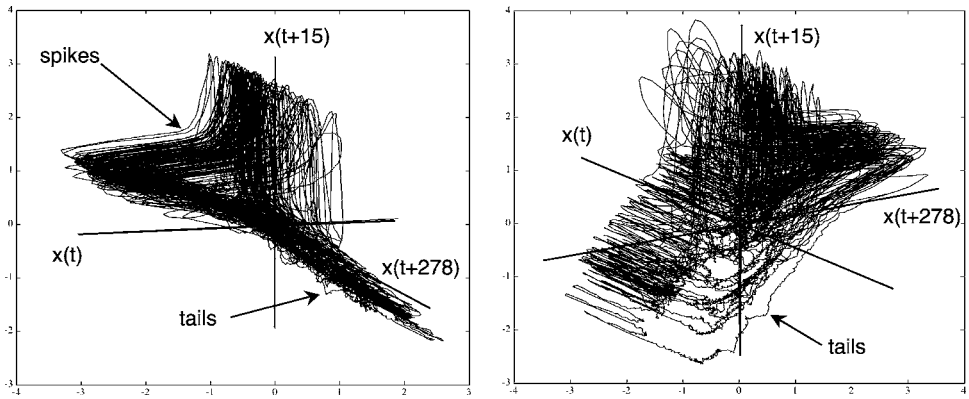

FIG. 8. Two three-dimensional views of the neuronal attractor using delays of 0, 15, and 278. The fast spikes are well resolved and the slow undulations (tails) are open.

hypothesis rejection is a separate problem and usually depends on each particular situation, hence the confidence level cannot be given *a priori*, but should be chosen by the end user. Finding the best confidence level would involve doing something like a risk or decision analysis for which payoffs and penalties for correctly or incorrectly rejecting the null hypothesis are known. Whether such an analysis is or is not possible (and often it is not), we believe that these levels should always be given so others can judge the quality of the reconstruction. Because our statistics do not involve arbitrary scales, but rather null hypothesis we can always give the probability that the reconstruction is right.

Finally, we note that the problem of observability for delay embeddings might be addressed with our statistic since in this case there is no closed form solution for chaotic systems.[37]

ACKNOWLEDGMENTS

We would like to acknowledge helpful conversations with M. Kennel, L. Tsimring, H. D. I. Abarbanel, and M. A. Harrison.

[1]H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer, New York, 1996).
[2]H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University. Press, Cambridge, 1997).
[3]N. H. Packard, J. P. Crutchfield, J. D. Farmer *et al.*, Phys. Rev. Lett. **45**, 712 (1980).
[4]F. Takens, in *Dynamical Systems and Turbulence, Warwick 1980*, edited by D. Rand and L.-S. Young (Springer, Berlin, 1981), p. 366.
[5]T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **64**, 579 (1991).
[6]A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).
[7]G. Kember and A. C. Fowler, Physica D **58**, 127 (1993); M. T. Rosenstein, J. J. Collins, and C. J. De Luca, *ibid.* **73**, 82 (1994).
[8]L. Cao, A. Mees, and J. K. Judd, Physica D **121**, 75 (1998).
[9]K. Judd and A. Mees, Physica D **120**, 273 (1998).
[10]D. Kugiumtzis, Physica D **95**, 13 (1996).
[11]P. Grassberger, T. Schreiber, and C. Schaffrath, Int. J. Bifurcation Chaos Appl. Sci. Eng. **1**, 521 (1991).
[12]W. Leibert, K. Pawlezik, and H. G. Schuster, Europhys. Lett. **14**, 521 (1991).
[13]A. Fraser, IEEE Trans. Inf. Theory **35**, 245 (1989).
[14]S. P. Garcia and J. S. Almeida, Phys. Rev. E **71**, 037204 (2005).
[15]M. Kennel and H. D. I. Abarbanel, Phys. Rev. E **66**, 026209 (2002).
[16]M. B. Kennel, R. Brown, and H. D. I. Abarbanel, Phys. Rev. A **45**, 3403 (1992).
[17]Lingyue Cao, Physica D **110**, 43 (1997).
[18]M. Casdagli, S. Eubank, J. D. Farmer *et al.*, Physica D **51**, 52 (1991); J. F. Gibson, J. D. Farmer, M. Casdagli *et al.*, *ibid.* **57**, 1 (1992).
[19]S. Boccaletti, L. M. Pecora, and A. Pelaez, Phys. Rev. E **65**, 066219 (2001).
[20]D. Broomhead and G. P. King, Physica D **20**, 217 (1986).
[21]A. P. M. Tsui, A. J. Jones, and A. G. de Oliveira, Neural Comput. **10**, 318 (2002); D. Evans and A. J. Jones, Proc. - R. Soc. Edinburgh, Sect. A: Math. **458**, 2759 (2002).
[22]H. Kantz and E. Olbrich, Chaos **7**, 423 (1997).
[23]E. Olbrich and H. Kantz, Phys. Lett. A **232**, 63 (1997).
[24]Wendell Fleming, *Functions of Several Variables* (Springer-Verlag, New York, 1977).
[25]L. Pecora, T. Carroll, and J. Heagy, Phys. Rev. E **52**, 3420 (1995).
[26]Louis M. Pecora, Thomas L. Carroll, and James F. Heagy, in *Nonlinear Dynamics and Time Series: Building a Bridge Between the Natural and Statistical Sciences, Fields Institute Communications*, edited by C. D. Cutler and D. T. Kaplan (American Mathematical Society, Providence, 1996), Vol. 11, p. 49.
[27]L. Moniz, L. Pecora, J. Nichols *et al.*, Struct. Health Monit. **3**, 199 (2003).
[28]David Middleton, *An Introduction to Statistical Communication Theory* (IEEE, Piscataway, 1996).
[29]E. T. Jaynes, *Probability Theory, The Logic of Science* (Cambridge University Press, Cambridge, 2003).
[30]J. Theiler, Phys. Rev. A **34**, 2427 (1986).
[31]E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).
[32]G. Strang, *Introduction to Applied Mathematics* (Wellesley-Cambridge, Wellesley, 1986), p. 611.
[33]W. H. Press, B. P. Flannery, S. A. Teukolsky *et al.*, *Numerical Recipes* (Cambridge University Press, New York, 1990).
[34]R. C. Elson, A. I. Selverston, R. Huerta *et al.*, Phys. Rev. Lett. **81**, 5692 (1998).
[35]J. L. Hindmarsh and R. M. Rose, Proc. R. Soc. London, Ser. B **221**, 87 (1984).
[36]Martin Falcke, Ramón Huerta, Mikhail I. Rabinovich *et al.*, Biol. Cybern. **82**, 517 (2000).
[37]L. Aguirre and C. Letellier, J. Phys. A **38**, 6311 (2005).