Biological algorithm for data reconstruction

Daniel J. Cross, Ryan Michaluk, and Robert Gilmore
Physics Department, Drexel University, Philadelphia, Pennsylvania 19104, USA
(Received 5 June 2009; revised manuscript received 8 January 2010; published 24 March 2010)

An algorithm inspired by Genome sequencing is proposed which "reconstructs" a single long trajectory of a dynamical system from many short trajectories. This procedure is useful in situations when many data sets are available but each is insufficiently long to apply a meaningful analysis directly. The algorithm is applied to the Rössler and Lorenz dynamical systems as well as to experimental data taken from the Belousov-Zhabotinskii chemical reaction. Topological information was reliably extracted from each system and geometrical and dynamical measures were computed.

DOI: 10.1103/PhysRevE.81.036217 PACS number(s): 05.45.Tp, 05.45.Pq

I. INTRODUCTION

A recent article [1] explored the possibility of reconstructing a long time-series from multiple short time-series in order to carry out fractal dimension calculations. The motivation for such a procedure is the desire to study dynamics in situations when long time-series simply are not available, either from practical (cost) or inherent (short-lived phenomenon) limits in an experiment.

An important instance of this can occur in electroencephalographic recordings using multipin sensors [2,3]. Recordings are made from many (~100) pins over long times during which behavior modality can change often, rapidly, and significantly. To analyze brain activity during a *single* modality, long signals from each pin can be decomposed into short segments representing each of the modalities experienced. The short signals from all electrodes during a single modality can then be synthesized into a longer signal using the methods described below. The resulting long signal describing a single behavior modality can then be used in standard ways for geometric, dynamical, and topological analyses.

The method of reconstruction in [1] is based on recurrence plot methods. Start with many short trajectories of vector data. At a given point x in the reconstruction determine the successor as follows. Calculate the nearest neighbors of x in phase-space within some distance ξ and then jump with probability p to the successor (in a short data set) of a randomly chosen neighbor, otherwise to the successor of x. If x is at the end of a short segment, jump to the future of a neighbor with probability one.

We offer a different method inspired by biological DNA methods. Since we start with many actual short trajectories of the system under study, we wish to take full advantage of this data and glue these trajectories together in a way that preserves as much of the given dynamics as possible. We then use the reconstructed time-series to compute the box-counting dimension and largest Lyapunov exponent as well as extract periodic orbits and compute topological indices. This method is illustrated for the Rössler and Lorenz attractors. We also reconstruct and analyze an attractor occurring in experimental data taken from the Belousov-Zhabotinskii chemical reaction.

The organization of the paper is as follows. In Sec. II we describe the reconstruction algorithm. We use this algorithm to reconstruct the Rössler attractor from many short data sets in Sec. III. In Sec. IV we calculate geometric and dynamical measures as well as perform a topological analysis on the attractor reconstructed in Sec. III. In Secs. V and VI we do the same for the Lorenz system. We then apply the reconstruction algorithm to experimental data taken from the Belousov-Zhabotinskii reaction in Sec. VII and analyze the resulting reconstruction in Sec. VIII. We consider variations in the algorithm in Sec. IX and summarize our results in Sec. X.

II. RECONSTRUCTION

In order to synthesize a number of smaller time-series fragments into a longer time series, we used a variant of the shotgun approach to genome sequencing [4]. The biological procedure involves taking a long DNA sequence, breaking it up into overlapping shorter segments, exponentially amplifying the number of each of these segments, and then finding overlapping regions on segments that allow the reconstruction of the entire original DNA sequence. We have adopted a variant of this matching method that does not employ segments overlapping in time. Our method, using nonoverlapping segments, works because each segment exists in a part of a strange attractor. As a consequence, segments that are nonoverlapping in the time domain are effectively overlapping in the phase-space domain, where the overlap tests are actually carried out.

Suppose we have N data sets of length L, where each data point is a vector in some n-dimensional phase space. (The algorithm may be used on data sets of unequal length with obvious modifications.) Our data sets will then be labeled as $D^j=\{x_1^j,x_2^j,\ldots,x_L^j\}$, where $1\leq j\leq N$.

Form a new trajectory as follows. First, predefine a threshold $\epsilon>0$ and a minimum and maximum length such that $0<L_{min}<L_{max}<L$. These are the overlap search parameters. For each l satisfying $L_{min}\leq l\leq L_{max}$ define a distance $\langle i,j\rangle_l$ between segments D^i and D^j by

$$\langle i,j\rangle_l=\frac{1}{l}\sum_{k=1}^{l}\|x_{L+1-k}^i-x_k^j\|. \tag{1}$$

These functions measure the average Euclidean distance between the last l points in D^i and the first l points in D^j (see Fig. 1).

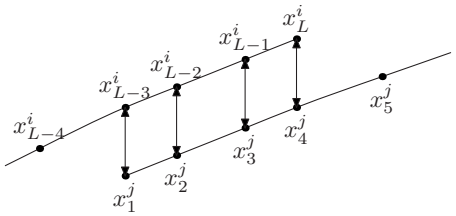

FIG. 1. A comparison between the last four iterates of sequence i with the first four of sequence j.

Begin with an arbitrary initial segment D^{j_0} and minimize the distance $\langle j_0, j\rangle_l$ over every l and every segment D^j with $j \neq j_0$. If $j=j_1$ minimizes the distance to j_0 with overlap $l=l_1$, then the next sequence in the reconstruction will be D^{j_1}. Next minimize the distance $\langle j_1, j\rangle_l$ over every l and every $j \neq j_0, j_1$, since we wish to avoid complete circuits (this "use and do not replace" condition is eliminated in Sec. IX D). Continue in this fashion until the minimum distance exceeds the threshold ϵ (likely) or all data sets have been exhausted (not likely). This produces a sequence of data sets and overlaps $\{j_0, l_1, j_1, l_2, j_2, \ldots\}$ from which we will reconstruct the trajectory.

Build a long trajectory by concatenating the short trajectories, but with an appropriate average of the points where data sets overlap. More specifically, begin by writing down the first $L-l_1$ data points of the first data set D^{j_0}. Next, linearly interpolate between the last l_1 points of D^{j_0} with the first l_1 points of D^{j_1} and then write down the next $L-l_1-l_2$ points of D^{j_1}. Then interpolate between the remaining l_2 points that overlap with the first l_2 points of D^{j_2}, etc.

The linear interpolation is a simple weighted average of the points. Specifically, if the points in the overlap region are $x^{j_1}_{L-l+1}, \ldots, x^{j_1}_L$ and $x^{j_2}_1, \ldots, x^{j_2}_l$, the interpolated points will be

$$\hat{x}_k = \frac{l+1-j}{l+1} x^{j_1}_{L-l+k} + \frac{k}{l+1} x^{j_2}_k, \tag{2}$$

where $1 \leq k \leq l$. We note that when $k=0$ or $k=l+1$ this expression gives (respectively) the last or first point on a short sequence not in the overlap region (see Fig. 2).

This scheme can be repeated for different values of ϵ or by starting from different initial data sets until a sufficiently long data set has been reconstructed. We emphasize that, even by changing these parameters, it is unlikely that all the original data will be used in the reconstructed trajectory. The advantage of this method is that most of the reconstructed data set consists of actual trajectories and the overlapping regions are smoothly interpolated. The search conditions can also be relaxed so that segments can be matched away from their ends. This and other variants are discussed further in Sec. IX

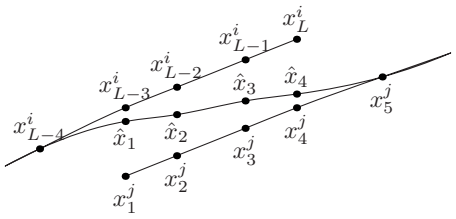

FIG. 2. Linear interpolation of the reconstruction in an overlapping region.

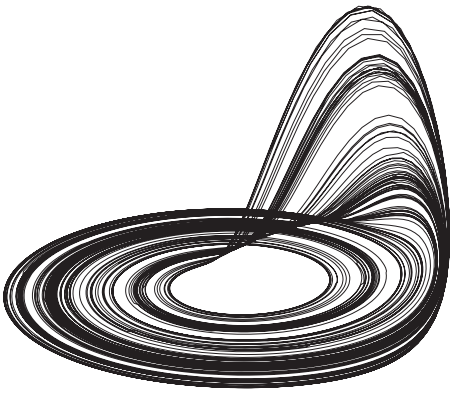

FIG. 3. Rössler attractor reconstructed from the short data sets. For clarity, only the first half of the reconstructed data set was plotted.

III. RECONSTRUCTION OF THE RÖSSLER ATTRACTOR

Once a time-series has been reconstructed by the method of Sec. II, it can be used as a surrogate for a single long data set in order to learn about the dynamics. In order to test this method we integrated the Rössler equations,

$$\dot{x} = -y - z,$$

$$\dot{y} = x + ay,$$

$$\dot{z} = b + z(x-c), \tag{3}$$

at the parameter values $(a=b=0.2,\ c=5.7)$ with an integration step of 0.01. After letting transients die out we recording every fifth point (x_i, y_i, z_i), which corresponds to approximately 130 samples per cycle [5]. From a single long data set we extracted $N=500$ data sets, each of length $L=200$. This was accomplished by removing, after every 200 points, a certain number of points, each number chosen at random between 50 and 150. The resulting short data sets each represent approximately a period and a half of the dynamics and may be regarded as independent.

We applied the reconstruction algorithm with a threshold of $\epsilon=0.5$ (or approximately 2.5% of the attractor diameter) and overlap parameters $L_{min}=5$ and $L_{max}=30$. The results are not sensitive to the choice of L_{min} and L_{max}.

The reconstruction algorithm produced a chain of 159 of the data sets, resulting in a data set 29 032 points long, or roughly one-third of the total data. The mean overlap $\langle\langle i, j\rangle\rangle$ between the data sets was 0.0943, which corresponds to only 20% of the threshold (0.5% of the attractor diameter). The reconstructed attractor is shown in Fig. 3.

We chose the $(-y, z)$-half-plane as a Poincaré section. Every short trajectory intersected this plane either once or twice. A return map was constructed from the subset of short segments that intersected the Poincaré section twice. This return map is shown in Fig. 4 (+ signs). A return map for the

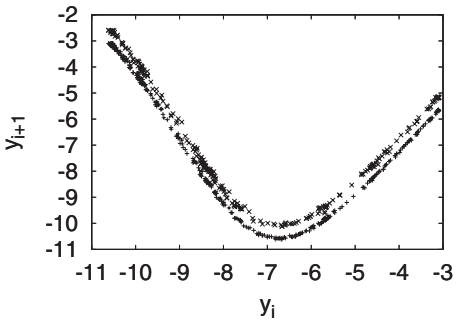

FIG. 4. Comparison of the first return maps from the original and reconstructed data: + return map from the original 500 short data sets; × return map from the reconstructed data. The latter is shifted upward by 0.5.

reconstructed data (×, rigidly displaced upward) is also shown in this figure.

IV. ANALYSIS OF RÖSSLER DATA

A. Geometrical measures

We performed a standard box-counting dimension calculation on the reconstructed attractor, yielding a dimension of 1.92 (see Fig. 5). This agrees quite well with the value 1.94 calculated on the original full data set. Both values are reasonably close to the accepted value of 2.01. We note that in [1] the correlation dimension was similarly underestimated for an ensemble of reconstructions of the Rössler attractor, yielding values of the form $2-\epsilon$ with $0 \leq \epsilon \leq 0.6$ (cf. Fig. 10(c) in [1]). In fact, underestimation of this quantity is characteristic [6].

For an interesting comparison we also made a dimension calculation on a data set consisting of all the short trajectories simply lumped together without any ordering or attempt at reconstruction. In this case we found the dimension to be 1.93, which is consistent with the previously calculated values. We believe this is explainable by appealing to ergodicity. In this case the reconstruction of long trajectories is unnecessary for geometrical measures, at least assuming the short trajectories well-sample the attractor.

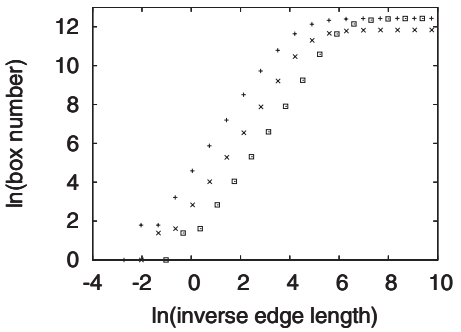

FIG. 5. Log-log plot (log base e) of box number versus inverse edge length for Rössler attractor: + the original attractor; × reconstructed attractor (offset right by 1); □ the union of all short trajectories (offset right by 2). The slope in the stable range is the box-counting dimension.

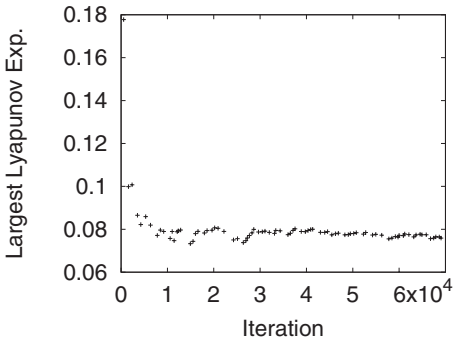

FIG. 6. Plot of the cumulative average of the largest Lyapunov exponent for the reconstructed Rössler attractor using the method of Wolf *et al.* [7].

B. Dynamical measures

Next we calculated the largest Lyapunov exponent using the methods of Wolf *et al.* [7] and Rosenstein *et al.* [8]. Using Wolf's method we first calculated the value 0.070 from the equations of motion. The values for the full experimental and reconstructed attractors were 0.068 and 0.075, respectively (Fig. 6). The respective errors are 2.9% and 7.1%.

Following Rosenstein's method, the value obtained for the original attractor (see Fig. 7) was 0.062 (11.4% error) and the value obtained for the reconstructed attractor was 0.069 (1.4% error). In opposition to the box-counting dimension estimate, no attempt was made to estimate the largest Lyapunov exponent from the union of all short trajectories since this dynamical measure requires a trajectory to follow.

C. Topological measures

We applied the method of close returns to the reconstructed Rössler time series to search for unstable periodic orbits in the strange attractor. We searched for orbits up to period ten, finding at least one orbit of each period. Each orbit was identified by name (its symbol sequence) by using the first return map. The linking numbers for the orbits through period five were computed and are summarized in

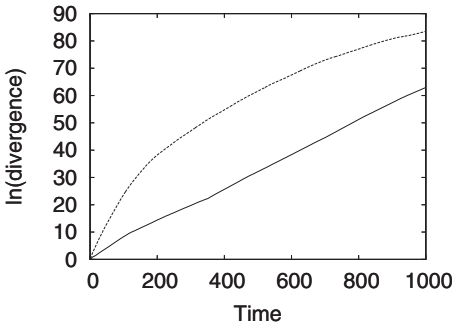

FIG. 7. Plot of ⟨ln(divergence)⟩ versus time for Rössler attractor using the method of Rosenstein [8]: (solid) original attractor; (dashed) reconstructed attractor. The slope in the stable range is largest Lyapunov exponent.

TABLE I. Table of linking numbers of periodic orbits through period 5 for the reconstructed Rössler attractor. Orbits are labeled by their itinerary on the Poincaré section.

	1	01	001	011	0111	01111
1	0	1	1	1	2	2
01	1	1	2	2	3	4
001	1	2	2	3	4	5
011	1	2	3	2	4	5
0111	2	3	4	4	5	8
01111	2	4	5	5	8	8

Table I (all linking numbers are negative, which has been suppressed). This table agrees with the known linking numbers of these orbits in the Rössler system [9].

Two of the periodic orbits we extracted from the reconstructed time series possess positive topological entropy. One was the period eight orbit 00101101 with a topological entropy of 0.346 034. The other was the period seven orbit 0010101, which has a topological entropy of 0.476 818 and is shown in Fig. 8. This figure shows in alternating symbols the trajectories from the short data sets used to reconstruct the periodic orbit. The interpolation scheme illustrated in Fig. 2 was not used in producing this orbit. In spite of this the discontinuities between overlapping segments are barely discernable.

The self-relative rotation rates of the period seven orbit, from which the entropy may be calculated, were found to be $(0)^1(\frac{2}{7})^4(\frac{3}{7})^2$, as explained in [5,10]. The existence of periodic orbits with positive topological entropy proves that the underlying dynamics is chaotic [9]. Moreover, the topological entropy of the system is bounded below by the entropy of any of its orbits. In the present case the attractor has an entropy no less than 0.476 818.

V. RECONSTRUCTION OF THE LORENZ ATTRACTOR

Next we applied our reconstruction algorithm to the attractor generated by the Lorenz equations,

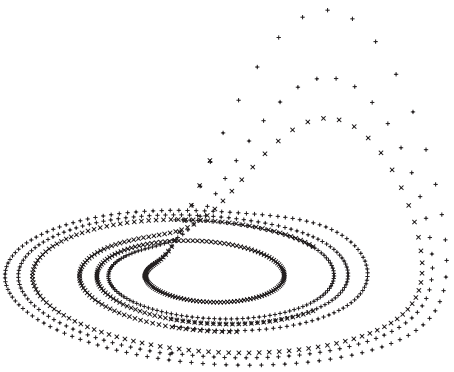

FIG. 8. The period seven orbit 0010101 extracted from the reconstructed Rössler data set. The original data sequences are plotted alternating between + and × without interpolation.

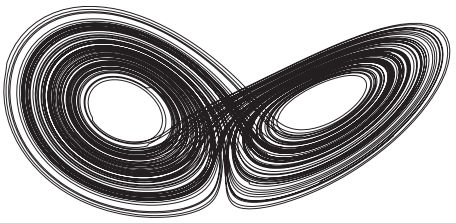

FIG. 9. Lorenz attractor reconstructed from the short data sets. For clarity, only the first third of the reconstructed data set was plotted.

$$\dot{x} = \sigma(y - x),$$

$$\dot{y} = x(R - z) - y,$$

$$\dot{z} = xy - bz, \tag{4}$$

at the parameter values ($\sigma=10$, $R=28$, $b=8/3$). In this case we used an integration step of 0.0058/5 and recorded every fifth point (after letting transients die out). This corresponds to about 130 samples per period with respect to the two-component Poincaré section [11,12]. We extracted $N=500$ data sets of length $L=200$ by removing random intervals of length between 50 and 150 points from a single long time-series. These parameters are comparable to those used for the Rössler system.

We applied the reconstruction algorithm with a tolerance of $\epsilon=2.1$ (or approximately 2.5% of the attractor diameter) and overlap parameters $L_{min}=5$ and $L_{max}=30$. The results are not sensitive to the choice of L_{min} and L_{max}. Again, these values are comparable to those used in the reconstruction of the Rössler attractor.

The reconstruction algorithm produced a chain of 335 of the data sets, resulting in a data set 61 027 points long, or roughly one third of the total data. The mean overlap between the data sets was 0.07, which corresponds to only 3.3% of the tolerance. The reconstructed attractor is shown in Fig. 9.

For the Poincaré section we chose two planes, one at $y=-9$ and the other at $y=+8$, and recorded the successive x values. The values were reoriented to measure distance away from fixed points. Figure 10 shows the sections for the origi-

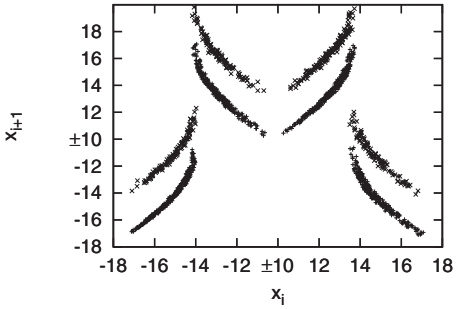

FIG. 10. Comparison of the first return maps from the original and reconstructed data: + return map from the original data; × return map from the reconstructed data. The latter is shifted upward by 3.0 for ease of comparison.

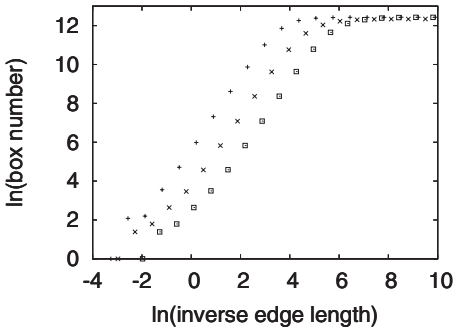

FIG. 11. Log-log plot (log base e) of box number versus inverse edge length for Lorenz: + the original attractor; × reconstructed attractor (offset right by 1); □ the union of all short trajectories (offset right by 2). The slope in the stable range is the box-counting dimension.

nal (+ signs) and reconstructed (×, rigidly displaced upward) data.

VI. ANALYSIS OF LORENZ DATA

A. Geometrical measures

For both the original and reconstructed attractor we calculated a box-counting dimension of 1.83 (see Fig. 11). These values are somewhat lower than the accepted value of 2.07, corresponding to nearly a 12% error. As for the Rössler attractor, we also calculated the value for the union of all data sets and found a value of 1.87, which closely agrees with the other values and in fact is closer to the accepted value. Again, we believe this is explainable by appealing to ergodicity.

B. Dynamical measures

For the differential equations, Wolf's method [7] gave a value of 0.90 for the largest Lyapunov exponent. Using Wolf's method (see Fig. 12), the value determined from the original attractor was 0.85 (5.6% error) and the value determined from the reconstructed attractor was 0.96 (6.7% error). For the original attractor the Rosenstein method [8] (see Fig. 13) gave a value of 0.83 (7.8% error), while for the reconstructed attractor it gave a value of 0.84 (6.7% error).

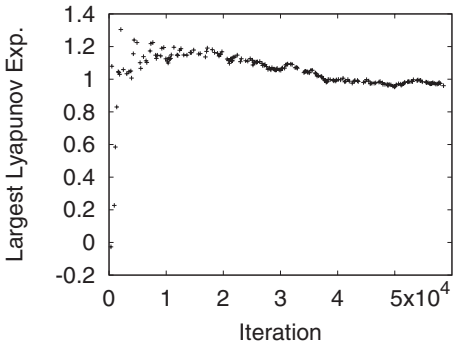

FIG. 12. Plot of the cumulative average of the largest Lyapunov exponent for the reconstructed Lorenz attractor in Wolf's method.

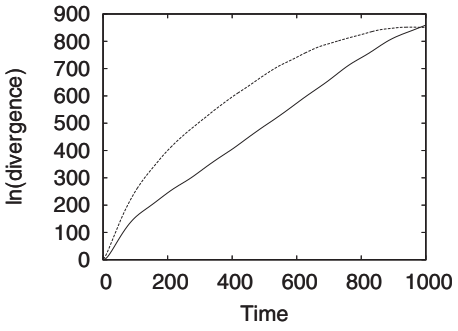

FIG. 13. Plot of ⟨ln(divergence)⟩ versus time for Lorenz attractor: (solid) original attractor; (dashed) reconstructed. The slope in the stable range is largest Lyapunov exponent.

C. Topological measures

As for the Rössler system we applied the method of close returns on the reconstructed Lorentz time series to search for unstable periodic orbits in the attractor. We searched for orbits from period two up to period ten, finding at least one orbit of each period and all 14 orbits up to period five. Each orbit was identified by name (its symbol sequence) by using the first return map. The linking numbers for all 14 orbits through period five were computed and found to agree with their known values [9]. For brevity, only the linking numbers of the six orbits with periods two through four are summarized in Table II.

The period five orbits LLLRR and RRRLL are the covering orbits of the Rössler orbit 00101 [13]. Since the image Rössler orbit has positive topological entropy (0.543 535), so do the covering orbits. Since the reconstructed attractor includes these positive entropy orbits we conclude that the system is chaotic. As a result, the topological entropy of the attractor is bounded below by 0.543 535.

VII. RECONSTRUCTION OF BELOUSOV-ZHABOTINSKII CHEMICAL REACTION DATA

In order to test the reconstruction algorithm on experimental data we used a time series from the Belousov-Zhabotinskii chemical reaction [14,15]. The data set is a scalar time series that measures the concentration of a bromine ion, which oscillates chaotically in the experiment. The data cover approximately 523 periods of oscillation, with an av-

TABLE II. Table of linking numbers of periodic orbits through period 5 found in the reconstructed Lorentz attractor. Orbits are labeled by their itinerary on the Poincaré section.

	LR	LLR	LRR	LRRR	LLLR	LLRR
LR	1	1	1	1	1	2
LLR	1	2	1	1	2	2
LRR	1	1	2	2	1	2
LRRR	1	1	2	3	1	2
LLLR	1	2	1	1	3	2
LLRR	2	2	2	2	2	3

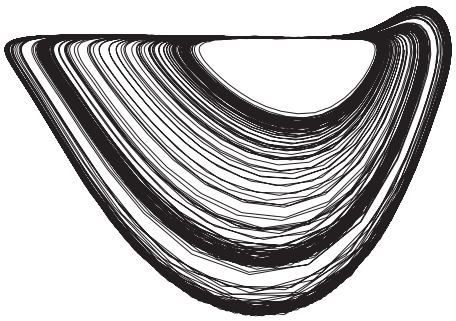

FIG. 14. Plot of d_i vs $\dot{d}_i$ for Belousov-Zhabotinskii data. The flat region of the data is apparent.

erage sample rate of 125 points per period. We applied the reconstruction algorithm directly to the scalar series and then constructed an embedding in three dimensions from the reconstructed time series.

Before discussing the details of the reconstruction we first review the subtleties of embedding Belousov-Zhabotinskii reaction data [9,16]. The scalar time series shows a slow, almost uniform rise from each minimum to the following maximum. The slope of this rise is, for all practical purposes, the same for all 523 oscillations. As a result, direct differential embeddings of the data are problematic. The projection onto the two components $(d_i, \dot{d}_i \simeq d_i - d_{i-1})$ has long, very flat regions. This projection is shown in Fig. 14. The second derivative, $\ddot{d}_i \simeq d_{i+1} - 2d_i + d_{i-1}$, is approximately zero in this region. This flat region in the three dimensional differential embedding is nearly one-dimensional and contains all the orbit crossings, causing them to be indistinguishable. The embedding problem is mitigated by using an integral rather than the second derivative. The embedding using these three coordinates, integral, original time series, and differential, resolves this flat region and distinguishes orbit crossings, as is seen in Fig. 15. For more details see [16].

We adopted the following embedding [16]

$$x_i = (d_i - \langle d \rangle) + \lambda x_{i-1},$$

$$y_i = d_i,$$

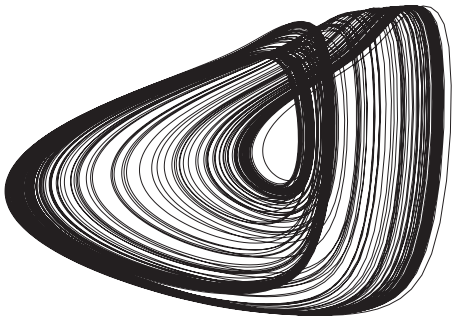

FIG. 15. Integral-differential embedding of Belousov-Zhabotinskii data. Projection onto the first two coordinates (x_i, y_i) is shown.

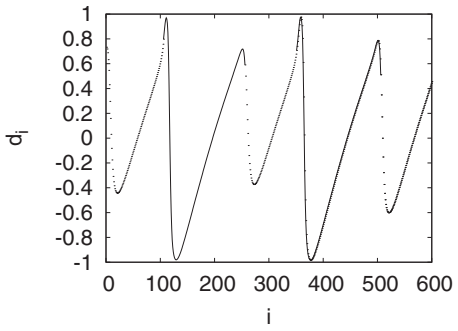

FIG. 16. Belousov-Zhabotinskii time series. The solid lines show how the short segments were constructed from consecutive maxima. The flat region is indicated by the nearly constant upward slope preceding each maximum.

$$z_i = d_i - d_{i-1}, \tag{5}$$

where $\langle d \rangle$ is the long time average over the data set reconstructed from scalar data sets and $\lambda = 1 - 10^{-2}$ is related to the memory of the integral. The resulting embedding is shown in Fig. 15.

The existence of this flat region apparent in the yz projection shown in Fig. 14 is a problem not only for constructing embeddings, but for the data reconstruction algorithm. In this region all data sets are highly correlated and the reconstruction algorithm yields a very close match for any two data segments that overlap there. A naïve matching of segments in this region results in reconstructions that do not truly represent the original dynamics.

To avoid this problem, short data segments that have end points in the flat region were not considered during the reconstruction. In the present case we created the short segments from a single long data set. Since creating these segments by making random cuts in the original data set would result in short segments ending in the flat region which must be discarded, we chose instead to maximize the number of short segments by making cuts before and after successive maxima in the original time series (see Fig. 16). The maxima fall outside the flat region. To prevent perfect overlap between adjacent segments, every other segment was discarded. For this time series it is always possible to do this "segment tailoring" on all segments if $L \geq 2(T + L_{max}) \simeq 272$ for $T \simeq 125$ (samples/cycle) and $L_{max} = 11$. The procedure created 261 short data sets, each covering just over one period of oscillation.

The reconstruction algorithm was applied to these data sets with the overlap fixed at eleven points and the tolerance set at 2.5% of the range of the maxima. The reconstruction utilized 227 of the 261 data sets. We embedded the reconstructed data set using the integral-differential embedding given in Eq. (5). The result is visually indistinguishable from the embedding of the original full data set shown in Fig. 15.

VIII. ANALYSIS OF THE BELOUSOV-ZHABOTINSKII CHEMICAL REACTION DATA

Using the method of close returns we isolated periodic orbits through period five from the embedded reconstructed

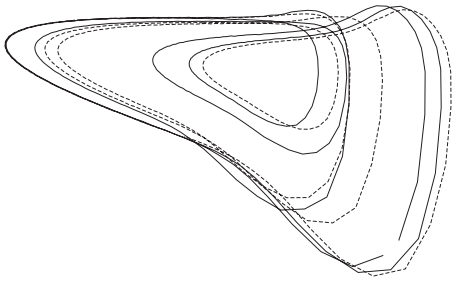

FIG. 17. Period four (dashed) and five (solid) orbits from the reconstructed attractor. The curves lie very close to each other in the furthermost left region resulting in unreliable linking numbers.

data. The calculated linking numbers of the extracted orbits agree with known values with the single exception of the period four and five pair. The linking number of these orbits was difficult to compute because some of the segments comprising the orbits were too close to resolve in the reconstruction. This difficulty can be seen in Fig. 17, which shows the period four and period five orbits. The table of linking number is given in Table III. Unfortunately, no periodic orbits with positive topological entropy were extracted from the reconstructed time series.

IX. VARIATIONS IN THE ALGORITHM

A. Undersampled data

It may occur that data from an experiment are undersampled, so that the dynamics is not represented with sufficient resolution. For example, the construction of a Poincaré map can be problematic when iterates are generally far away from the chosen section. Especially in "fast" regions of the flow, straight lines interpolated between successive data points can intersect each other, obviously violating the uniqueness theorem and misrepresenting the dynamics.

In such cases a Fourier interpolation scheme can be utilized to effectively increase the sample rate to a sufficient level [5]. One finds a trigonometric representation of the data in terms of sine and cosine functions and evaluates them at intermediate values of angle to interpolate between the known points. This is a natural choice for interpolation of recurrent systems as the trigonometric functions are themselves periodic.

TABLE III. Table of linking numbers of periodic orbits through period 5 found in the reconstructed Belousov-Zhabotinskii data. Orbits are labeled by their itinerary on the Poincaré section. Parentheses indicates the correct value that could not be reliably calculated. All values are negative, which has been suppressed.

	1	01	011	0111	01011
1	0	1	1	2	2
01	1	1	2	3	4
011	1	2	2	4	5
0111	2	3	4	5	(8)
01011	2	4	5	(8)	8

B. Scalar time series

It is most common experimentally to record scalar rather than vector time-series data. When one has multiple short scalar data sets, one could first embed each of the individual data sets and then apply the reconstruction algorithm. However, as we demonstrated in Sec. VII it is also possible to apply the algorithm directly on short scalar data sets and then proceed to embed the reconstructed scalar time series. Care must be taken so avoid matching segments with end points in flat regions of the data.

C. Utilization of unused data

As has been pointed out, the reconstruction algorithm will not generally use all of the original data points. Our reconstruction of the Rössler attractor in Sec. III used about one-third of the total data. Starting with a particular data set and search parameters, the algorithm finds the best next data set repeatedly until it can no longer stay within tolerance. However, having found this initial long data set, we can repeat the procedure on the remaining unused data sets to obtain another reconstructed data set, and so-on. We can treat each of these reconstructions as independent long data sets and analyze each separately.

Another way to utilize more data is to relax the restriction that the overlap comparison is just between heads and tails of data sets. One can search for the best match of a certain data set's tail with any sequence of points in any other data set. If the best match is found mid-data set, we can cut the data set at that point, and save the unused portion for another trial later. This variant of the algorithm would also tend to have smaller average distances between matched sequences since the search space is larger.

D. Direct determination of periodic orbits

It is possible, by slightly altering the algorithm, to run a search for periodic orbits directly on the individual data segments without having to construct an intermediate long data set. This can be done by stitching together the short times series into approximately closed orbits by allowing the algorithm to match any intermediate part of the last added segment with the beginning of the initial segment (implementing a "use and replace" condition). The advantage here is to use all the data equally when searching for the periodic orbits, rather than a certain subset. The reconstruction algorithm will always prefer the closest segment it finds, but if several segments are within the close-return threshold they could all be considered when searching for the orbits.

E. Comparison with the Shotgun Approach

To make a direct comparison with the actual shotgun procedure in biology, we integrated the Rössler equations as before, but instead of removing random intervals to obtain the short data sets, we chopped them up into overlapping segments, just as in the original procedure in biology. The amount of overlap was chosen at random between 5 and 30 points. This resulted in a total of 500 data sets.

After randomizing the order of the data sets, the reconstruction algorithm found three chains in the data, the first using 496 data sets, the second three data sets, and the third one data set. Running the algorithm on these three resulting data sets returned the full original data set.

X. CONCLUSIONS

This paper has explored an algorithm for reconstructing long time series from several short time series inspired by the shotgun approach to genome sequencing. The Rössler and Lorenz systems were studied in detail. We showed that geometrical and dynamical measures could be extracted from the reconstructed dynamics with a degree of confidence comparable to calculation with long clean data sets or with the differential equations in the case of dynamical measures. More importantly, we successfully performed a topological analysis of the reconstruction. Many periodic orbits of both systems were found, including several with positive topological entropy. The presence of positive entropy orbits demonstrates that both systems were chaotic. We successfully applied the reconstruction algorithm to experimental scalar data from the Belousov-Zhabotinskii reaction. Next, we looked at variations in the algorithm for dealing with undersampled data, with scalar data, for utilizing unused data, and for making direct searches for periodic orbits. We discussed the search for periodic orbits without explicit data reconstruction, and finally compared our algorithm directly with the biological procedure. The variations in the algorithm are currently being implemented and will appear in a forthcoming paper. The algorithm we have presented offers a robust procedure for extracting topological information about a dynamical system from several short data sets.

[1] C. Komalapriya, M. Thiel, M. C. Romano, N. Marwan, U. Schwarz, and J. Kurths, Phys. Rev. E **78**, 066217 (2008).
[2] E. M. Maynard, C. T. Mordhausen, and R. A. Normann, Electroencephalogr. Clin. Neurophysiol. **102**, 228 (1997).
[3] P. R. Kennedy and R. A. E. Bakay, NeuroReport **9**, 1707 (1998).
[4] J. C. Venter *et al.*, Science **291**, 1304 (2001).
[5] R. Gilmore, Rev. Mod. Phys. **70**, 1455 (1998).
[6] W.-L. Lee, Y.-C. Chen, and K.-S. Hsieh, *IEEE International Symposium on Circuits and Systems*, ISCAS 2002 (IEEE, Bellingham, WA, 2002), Vol. 2, pp. 656–659.
[7] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, Physica D **16**, 285 (1985).
[8] M. T. Rosenstein, J. J. Collins, and C. J. D. Luca, Physica D **65**, 117 (1993).
[9] R. Gilmore and M. Lefranc, *The Topology of Chaos* (Wiley, New York, 2002).
[10] H. G. Solari and R. Gilmore, Phys. Rev. A **37**, 3096 (1988).
[11] T. D. Tsankov and R. Gilmore, Phys. Rev. Lett. **91**, 134104 (2003).
[12] T. D. Tsankov and R. Gilmore, Phys. Rev. E **69**, 056206 (2004).
[13] C. Letellier and R. Gilmore, Phys. Rev. E **63**, 016206 (2000).
[14] K. Coffman, W. D. McCormack, Z. Noszticzius, R. H. Simoyi, and H. L. Swinney, J. Chem. Phys. **86**, 119 (1987).
[15] D. P. Lathrop and E. J. Kostelich, Phys. Rev. A **40**, 4028 (1989).
[16] G. B. Mindlin, H. G. Solari, M. A. Natiello, R. Gilmore, and X. J. Hou, J. Nonlinear Sci. **1**, 147 (1991).