

Recurrence plots of experimental data: To embed or not to embed?

Joseph S. Iwanski*

Department of Applied Mathematics
University of Colorado, Boulder, CO 80309-0526

Elizabeth Bradley†

Department of Computer Science
University of Colorado
Boulder, CO 80309-0430

Abstract

A recurrence plot is a visualization tool for analyzing experimental data. These plots often reveal correlations in the data that are not easily detected in the original time series. Existing recurrence plot analysis techniques, which are primarily application-oriented and completely quantitative, require that the time-series data first be embedded in a high-dimensional space, where the embedding dimension d_E is dictated by the dimension d of the data set, with $d_E \geq 2d + 1$. One such set of recurrence plot analysis tools, Recurrence Quantification Analysis, is particularly useful in finding locations in the data where the underlying dynamics change. We have found that the same results can be obtained with no embedding. The work presented in this paper represents the beginning of an attempt to improve upon recurrence plot analysis in a way that incorporates and exploits their rich structural characteristics.

*Supported by a NSF Mathematical Sciences Graduate Research Traineeship #DMS-925-6335

†Supported by NSF NYI #CCR-9357740, NSF #MIP-9403223, ONR #N00014-96-1-0720, and a Packard Fellowship in Science and Engineering from the David and Lucile Packard Foundation.

Lead Paragraph In general, time-series analysis methods begin with — or at least include — delay-coordinate embedding, a well-established means of reconstructing the hidden dynamics of the system that generated the time series. If the embedding is done correctly, the theorems involved guarantee that certain properties of the original system, known as dynamical invariants, are preserved in the embedded, or reconstruction space. This is an extremely powerful correspondence, implying that many conclusions drawn from the reconstruction-space dynamics are also true of the real, underlying dynamics. There are, of course, some important caveats; one of the most limiting is that correct embeddings are not easy to construct. Methods in the time-series analysis literature use a variety of heuristics to solve the (significant) problems that are inherent in this process, and the resulting algorithms are often computationally expensive. Time-series analysis methods that do not require embedding are, therefore, extremely desirable. We present evidence suggesting that recurrence plots are such a method. Our conclusion is twofold. First, when using current methods of recurrence plot analysis one need not embed the data. Second, we note that better methods of recurrence plot analysis are needed, methods that take into account the structural and qualitative aspects of these fascinating plots. This work represents a first step towards this goal.

1 Introduction

First introduced in a 1987 paper by Eckmann, Kamphorst, and Ruelle[3], the recurrence plot (RP) is an analysis tool for experimental time-series data. An RP is a two-dimensional representation of a single trajectory. The time series spans both ordinate and abscissa and each point (i, j) on the plane is shaded according to the distance between the two corresponding trajectory points y_i and y_j . In an unthresholded RP (UTRP) the pixel lying at (i, j) is color-coded according to the distance, while in a thresholded RP (TRP) the pixel lying at (i, j) is black if the distance falls within a specified threshold corridor and white otherwise. For instance, if the 117th point on the trajectory is 14 distance units away from the 9435th point, the pixel lying at $(117, 9435)$ on the RP¹ will be shaded with the color that corresponds to a spacing of 14. Figure 1 shows UTRPs generated from two very different data sets: a time series derived by sampling the function $\sin t$ and a time series from the well-known Lorenz system. The colors on these plots range from dark blue for very small spacings to red for large inter-point distances, as shown on the calibration bars in the figure. With this in mind, the sine-wave RPs are relatively easy to understand; each of the “blocks” of color simply represents half a period of the signal². The lower RPs in the figure, generated from a chaotic data set, are far more complicated, although they too have block-like structures resembling what might be expected from a periodic signal. This signal, though, is not periodic, so the repeated structural elements in the plot beg an explanation.

Recurrence plots are intricate and visually appealing. They are also useful for finding hidden correlations in highly complicated data. Moreover, because they make no demands on the stationarity of a data set, RPs are particularly useful in the analysis of systems whose dynamics may be changing. Although the literature in this area is not extensive, the use of recurrence plots in time-series analysis has become more common in recent years, particularly in the area of physiology. Webber and Zbilut[14], for instance, used recurrence plot analysis to discern between “quiet” and “active” breathing in laboratory rats, and Kaluzny and Tarnecki[4] used RPs to study neuronal spike trains in cats. RPs have been used in mathematical problems primarily to identify transition points in non-stationary data sets. Trulla, Giuliani, Zbilut, and Webber[13], for instance, analyzed the dynamics of the logistic equation, varying the driving parameter smoothly and leading the time series between chaotic and periodic regimes. They concluded that RP analysis compares favorably to classical statistical approaches as a means for analyzing chaotic data, particularly in its ability to detect bifurcations. Very recently, Casdagli[2] used RPs to characterize time series generated by dynamical systems driven by slowly varying external forces.

Our study of recurrence plots has been motivated by the desire to give meaning to the fascinating structures that they exhibit. Previous work in this area is primarily

¹Point $(9435, 117)$ will be shaded similarly. As originally conceived, however, RPs are not necessarily symmetric[3].

²Recall that the shading of the point (w, v) on this plot reflects the distance between $\sin w$ and $\sin v$; if $w - v = \pi/2$, for example, that distance is large.

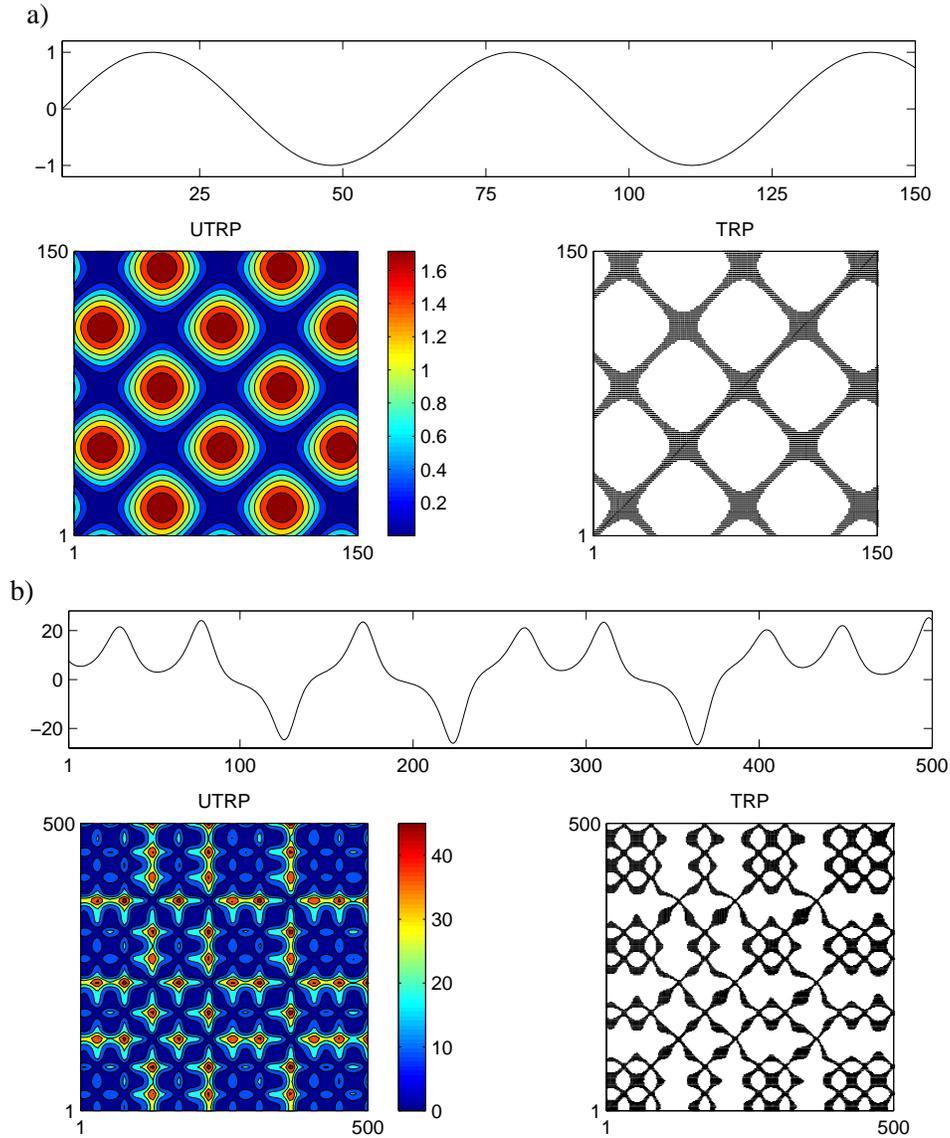


Figure 1: Recurrence plots: (a) a sine-wave signal and the corresponding recurrence plots (b) a more-complicated signal from the Lorenz system and the corresponding recurrence plots. The colorful plots to the left, known as unthresholded recurrence plots (UTRPs), show the distance relationships between all points in the signal via a color map. The black and white plots to the right are called thresholded recurrence plots (TRPs) because they highlight only those points that fall within a prescribed distance range. In (a), for example, this range of distances, called a threshold corridor, is $[0, 0.25]$; in (b) the threshold corridor is $[0, 5.0]$. One can see the TRP patterns in the UTRPs by careful examination of the UTRP colorbars for the given threshold corridor values.

application-oriented and completely quantitative. We wish, rather, to extend, formalize, and systematize recurrence plot analysis in a meaningful way that is based both in theory and experiment and that targets both quantitative and qualitative properties. An important consequence of such a formalization is the power that it would lend to the RP as an analysis tool. For example, knowledge of how periodicity and chaos manifest on RPs and how bifurcations affect the geometry and topology of their structure would allow us to use these plots as a means of determining when a system has, for example, moved from a limit cycle to a chaotic regime. The work described in this paper is a first step towards this type of formalization.

While examining several RPs of a particular data set, we noticed that their appearances seemed to remain qualitatively unchanged with changing embedding dimension. Figure 2 illustrates this for data from an angle sensor on a parametrically driven pendulum. When RPs of other, unrelated data sets also exhibited this type of surprising behavior, the question naturally arose as to whether the quantitative aspects of recurrence plots were independent of the embedding dimension as well. In this paper, we present a suite of numerical calculations on simulated and experimental data sets that explore this question. Specifically, we demonstrate that the Recurrence Quantification Analysis (RQA) of [13] appears not to depend on the embedding dimension. Analytic justification of this result is a current focus of our effort and the topic of a forthcoming paper.

This result suggests that recurrence plots are more powerful than was previously believed. In all four of the studies mentioned in the related work paragraph above, for example, the data were first embedded in R^{d_E} for some $d_E > 1$, using the familiar method of delay coordinates[8]. Contrary to current time-series analysis arcana, however, our evidence suggests that d_E need only be equal to one if the data are to be analyzed with RQA. This means that no embedding need be done.

2 RPs and Recurrence Quantification Analysis

In this section, we briefly outline some of the basic features of RPs and describe how one generates an RP of an experimental data set. The standard first step in this procedure is to reconstruct the dynamics by embedding the one-dimensional time series in d_E -dimensional reconstruction space using the method of delay coordinates. Given a system whose topological dimension is d , the sampling of a single state variable is equivalent to projecting the d -dimensional phase-space dynamics down onto one axis. Loosely speaking, embedding is akin to “unfolding” those dynamics, albeit on different axes. The Takens theorem guarantees that the reconstructed dynamics, if properly embedded, are equivalent to the dynamics of the true, underlying system in the sense that dynamical invariants such as generalized dimensions and the Lyapunov spectrum, for example, are identical[7, 12]. The process of constructing a correct embedding is the subject of a large body of literature and numerous heuristic algorithms and arguments; Abarbanel’s recent text[1] gives a good summary of this extremely active field. The basic problem is

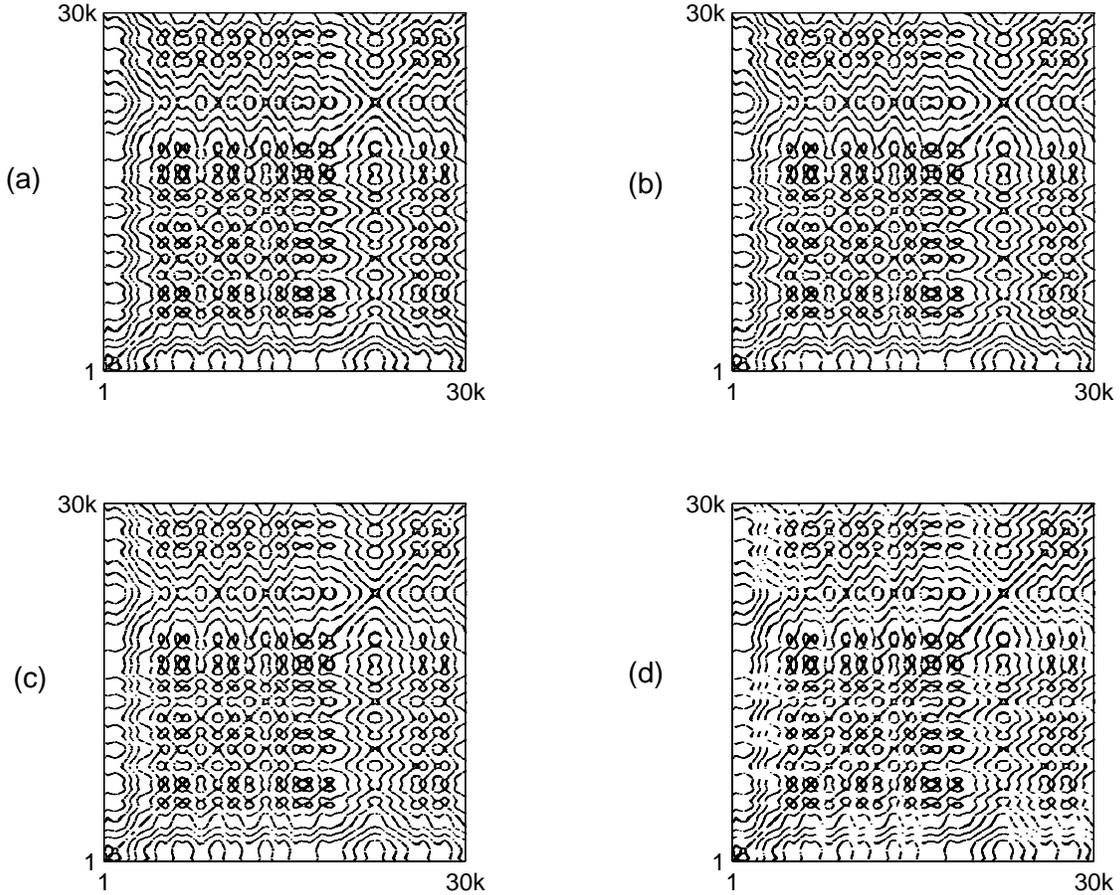


Figure 2: These four thresholded recurrence plots were generated from a data set gathered from an angle sensor on a parametrically driven pendulum. TRPs (a)–(d) represent embedding dimensions 1–4, respectively, and all four plots have identical threshold corridors and time delay values. Note the striking structural similarity: the only apparent variation is a lightening of the TRP with increasing embedding dimension.

to choose two parameters — the delay and the reconstruction-space dimension — that guarantee an *embedding*³ of the data. This process is difficult because in general one does not have prior knowledge of d 's correct value; all one has is a one-dimensional time series and from this one would like to learn as much as possible about the system that generated the signal. Given a trajectory in the embedded space, finally, one constructs an RP by computing the distance between every pair of points (y_i, y_j) using an appropriate norm and then shading each pixel (i, j) according to that distance.

2.1 Delay Coordinate Embedding

To reconstruct the dynamics, we begin with experimental data consisting of a time series:

$$\{x_1, x_2, \dots, x_N\}$$

Delay-coordinate reconstruction of the unobserved and possibly multi-dimensional phase-space dynamics from this single observable x is governed by two parameters, embedding dimension d_E and time delay τ . The resultant trajectory in R^{d_E} is:

$$\{y_1, y_2, \dots, y_m\}$$

where $m = N - (d_E - 1)\tau$ and

$$y_k = (x_k, x_{k+\tau}, x_{k+2\tau}, \dots, x_{k+(d_E-1)\tau})$$

for $k = 1, 2, \dots, m$. Note that using $d_E = 1$ merely returns the original time series; one-dimensional embedding is equivalent to not embedding at all. Proper choice of d_E and τ is critical to this type of phase-space reconstruction and must therefore be done wisely; only “correct” values of these two parameters yield embeddings that are guaranteed — by the Takens Theorem[12] and subsequent work by Packard, Crutchfield, Farmer and Shaw[7] and Sauer, Yorke and Casdagli[8] — to be topologically equivalent to the original (unobserved) phase-space dynamics.

Assuming that the delay-coordinate embedding has been correctly carried out, it is natural to assume that the RP of a reconstructed trajectory bears great similarity to an RP of the true dynamics. Furthermore, we expect any properties of the reconstructed trajectory inferred from this RP to be true of the underlying system as well. This is, in fact, the rationale behind the standard procedure of embedding the data before constructing a recurrence plot.

2.2 Constructing the Recurrence Plot

Recurrence plots are based upon the mutual distances between points on a trajectory, so the first step in their construction is to choose a norm D . For the work presented

³Used precisely, the term embedding refers to a one-to-one map that also preserves tangent directions.

here, we use the maximum norm, although in one dimension the maximum norm is, of course, equivalent to the Euclidean p-norm. We chose the maximum norm for two reasons: for ease of implementation and because the maximum distance arising in the recurrence calculations (the difference between the largest and smallest measurements in the time series) is independent of embedding dimension d_E for this particular norm. This means that we can make direct comparisons between RPs generated using different values of d_E without first having to re-scale the plots. Using the Euclidean 2-norm, on the other hand, inter-point distances increase with embedding dimension simply because length along a new dimension can only contribute to the total distance. It is trivial to show that distance $D(y_i, y_j)$, as measured by Euclidean p-norm or the maximum norm, is non-decreasing with respect to d_E . The point of using the maximum norm was to mitigate this effect.

Next, we define the *recurrence matrix* A as follows:

$$A(i, j) = D(y_i, y_j), 1 \leq i, j \leq m$$

$$D(y_i, y_j) = \max_{1 \leq k \leq d_E} |x_{i+(k-1)\tau} - x_{j+(k-1)\tau}|$$

We then generate an unthresholded recurrence plot (UTRP) of the time series by plotting matrix A as a contour plot (see figure 1). Since current recurrence-plot analysis methods, including Trulla *et al.*'s RQA, focus on thresholded recurrence plots (TRPs), the next step is to use the UTRP to choose a *threshold corridor*, $[\delta_l, \delta_h]$. This is done by first visually examining the UTRP in order to find interesting structures, and then using the corresponding values from the UTRP colorbar as values for δ_l and δ_h , thus isolating these structures. *Interesting* structures are those whose appearance is best described as being somewhat continuous – resembling an underlying skeleton for the rest of the UTRP – and that persist for several different threshold corridors. For data from the Lorenz and Rössler systems, this procedure is relatively easy, but for other systems in which such structural organization may not be present we are relegated to the somewhat ad hoc process of choosing a threshold corridor that represents some percentage of the total range of recurrence distances present in the UTRP. The latter method is what is generally used in the literature. Once the threshold corridor has been chosen, it is used in order to generate a *thresholded recurrence matrix* B :

$$B(i, j) = \begin{cases} 1 & \text{if } \delta_l \leq D(y_i, y_j) \leq \delta_h \\ 0 & \text{otherwise} \end{cases}$$

Finally, the TRP is generated by darkening all pixels (i, j) that correspond to nonzero entries in matrix B .

The choice of threshold corridor $[\delta_l, \delta_h]$ is critical; too large a corridor results in saturation of the entire TRP — where every pixel is black — while a corridor that is too narrow will not be adequately populated with points to support the analyses that follow. Besides being critically important, the selection of threshold corridor is also difficult to

systematize in any sensible way. Solutions in the literature are unsatisfying; Webber and Zbilut, without comment, prescribe a threshold corridor corresponding to the lower ten percent of the entire distance range present in the corresponding UTRP. Our procedure, which uses corridor boundary values that isolate “interesting” structures in the UTRP, is also somewhat unsatisfying, and we are working on developing a better formalization. In the meantime, it is an adequate preliminary approach; for instance, it allowed us to reproduce the results in [13], even though that paper did not specify a threshold corridor. Moreover, our procedure allows us to isolate and examine interesting structures *across the range of recurrence distances* — unlike most existing TRP approaches, which specify threshold corridors of the form $[0, r]$. This is an important advantage, as it allows us to examine recurrence structures comprised of points that are not false near neighbors (FNNs) in reconstruction space⁴.

2.3 Recurrence Quantification Analysis

Perhaps the key mathematical issue in any attempt to use RPs to analyze experimental data is that of quantifying the structure that appears in the plots. Trulla, Guilianni, Zbilut, and Webber[13] have devised a set of quantifying analyses, collectively called Recurrence Quantification Analysis (RQA), to address this problem. The remainder of this section covers the RQA procedure in detail. We view these techniques as the best-formulated and most-general approach to RP analysis that has been developed to date. However, its lumped statistical nature means that RQA cannot capture many of the spatiotemporal details of the dynamics. Moreover, it appears that the standard first step in this procedure — that of embedding the data — may be unnecessary.

In order to perform RQA on a data set, we first construct a TRP, choosing a threshold corridor $[\delta_l, \delta_h]$ as described in the previous section, and then use that TRP to compute five statistical values. The first of these statistics, termed *% recurrence* (REC), is simply the percentage of points on the TRP that are darkened (i.e., those pairs of points whose spacing falls within the corridor). This percentage is precisely what is used to compute the correlation dimension of a data set — Kaplan and Glass[5], for instance, define correlation dimension as the slope of the linear region in the S-shaped *% recurrence vs. corridor width* plot. RQA, however, stops short of extending the analysis beyond the simple calculation of the percentage of dark points on the plot. The second RQA statistic is called *% determinism* (DET); it measures the percentage of recurrent points in a TRP that are contained in lines parallel to the main diagonal. The main diagonal itself is excluded from these calculations because points there are trivially recurrent. Diagonal lines are included in the analysis if and only if they meet or exceed some prescribed minimum length threshold. Intuitively, DET measures how “organized” a TRP is. The third RQA statistic, called *entropy*, is closely related to *% determinism*. Entropy (ENT) is calculated by binning the diagonal lines defined in the previous paragraph according to their lengths and using the following formula:

⁴The process of unfolding the attractor — increasing d_E — effectively eliminates FNNs, points that are neighbors in low-dimensional space due only to projection and not to the underlying dynamics[6].

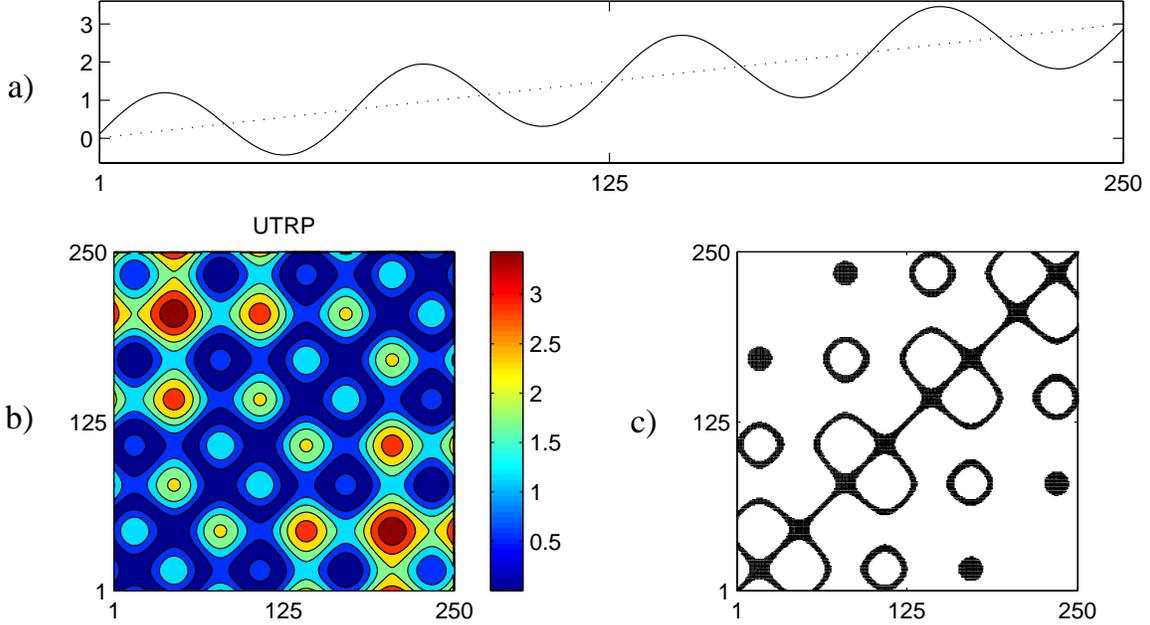


Figure 3: (a) a simple sine-wave signal is combined with an upward linear signal (dashed line) to yield an upward-trending sine wave. (b) The corresponding UTRP “pales” away from the main diagonal; the color shifts from blue to red as distances increase. (c) The TRP for a threshold corridor of $[0, 0.25]$. Note that recurrences are less frequent away from the main diagonal. This generally means that points are recurrent only if they are close in time.

$$ENT = - \sum_{k=1}^N P_k \log P_k$$

where N is the number of bins and P_k is the percentage of all lines that fall into bin k . According to Shannon’s information theory[9], predictability decreases with increasing entropy, so one would expect low values of ENT for TRPs of chaotic data sets, for example. The fourth RQA statistic, termed TREND, measures how quickly a TRP “pales” away from the main diagonal. As the name suggests, TREND is intended to detect non-stationarity in the data. Figure 3 shows a time series of a sine-wave signal with an upward linear trend, together with RPs — thresholded and unthresholded — of that signal. Although both plots retain the periodic block structure induced by the periodicity of the signal, the shading of the UTRP “fades” away from the diagonal because the linear trend adds to the inter-point distances in spite of the periodicity. The fifth and final RQA statistic is $1/line_{max}$, where $line_{max}$ is the longest line found during the computation of DET. Eckmann *et al.* claim that line lengths on RPs are directly related to the inverse of the largest positive Lyapunov exponent[3]. Short $line_{max}$ values are therefore indicative of chaotic behavior. In a purely periodic signal — the opposite extreme — lines tend to be very long, so $1/line_{max}$ is very small.

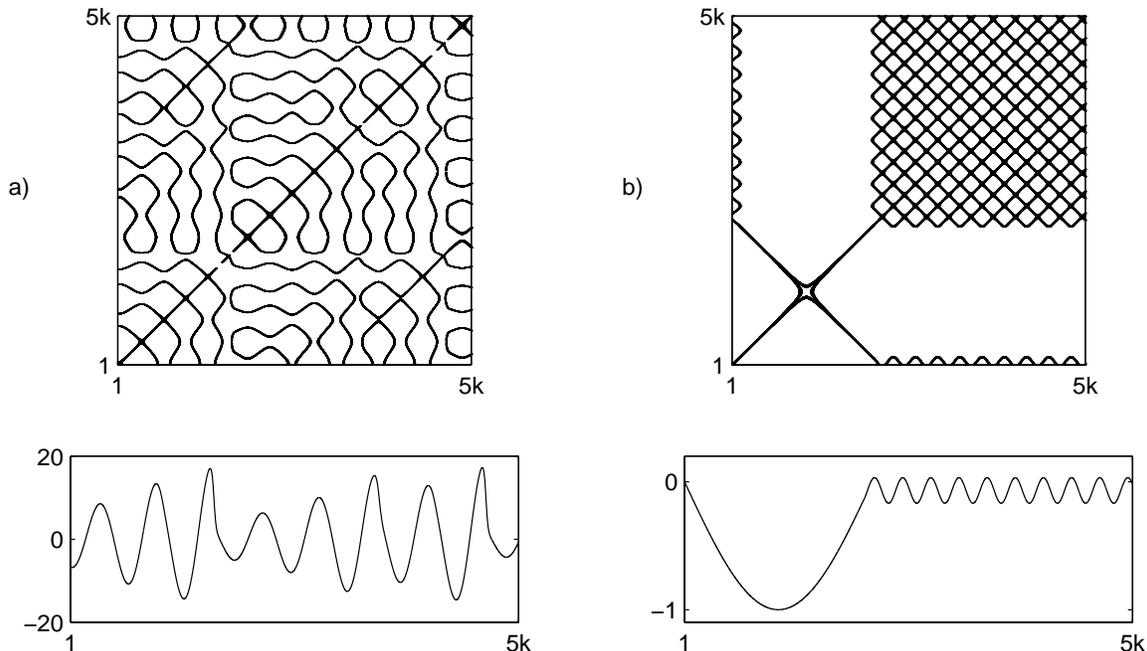


Figure 4: RQA results on structurally dissimilar TRPs can be almost identical. These two very different TRPs, one (a) from the Rössler system and one (b) a sine-wave signal of varying period, have equal or near-equal values of REC (2.1%) and DET (42.9% for the Rössler data and 45.8% for the varying-period sine wave). Both TRPs were generated from 5000-measurement time series using an embedding dimension of 1 and a minimum line length of 5.

While the ultimate goal of our work is to improve upon existing methods of RP analysis, we recognize that RQA is probably the best current set of RP analysis tools. However, the lumped statistics of RQA do not measure much of the qualitative structure of recurrence plots; in figure 4, for example, we show two structurally dissimilar RPs that are almost identical from the standpoint of RQA. It is our goal to devise RP analysis methods in which these types of structural, qualitative differences are clearly evident and easy to analyze. The ultimate intent of this line of research is to refine RP analysis techniques to the point where they can be used to gain greater insight into the underlying signals from which the RPs were generated.

3 Experiments and Results

One of the more intriguing — and puzzling — characteristics of recurrence plots is the structural stability that they exhibit with increasing embedding dimension. That is, qualitative features that are visible in RPs generated using $d_E = 1$ often persist in RPs that are based on higher embedding dimensions — even in complex, high-dimensional data sets. This may appear counterintuitive, as the delay-coordinate embedding process

is designed to “unfold” the underlying dynamics from the one-dimensional time series. Following this line of reasoning, it might be natural to expect that the RP structure of a formally correct reconstruction ($d_E \geq 2d + 1$) would be different from RPs of partially unfolded dynamics ($d_E = 1, 2, \dots, 2d$): that the RP structure would change with embedding dimension until the “correct” d_E was reached. However, this is not generally the case, as is clearly visible in figure 2.

The remainder of this section explores this issue in the context of both quantitative and qualitative measures of RP structure. Specifically, we demonstrate that RP structure remains qualitatively unchanged with d_E for two simulated systems (Lorenz and Rössler), one experimental system (the driven pendulum), and a simple sine-wave time series. We then investigate how changes in embedding dimension affect the quantitative RP features, as captured by three of the five RQA statistics. Finally, we duplicate the analysis of [13] — wherein RQA was used to detect dynamical variations in a nonstationary time series derived from the logistic map — and we extend that analysis to the Lorenz system. The results of these analyses were uniformly independent of embedding dimension.

3.1 Qualitative Appearance of TRPs for Various Systems

To demonstrate their structural stability with increasing embedding dimension, we present TRPs of four different time series. The first consists of the x -component of a fourth-order Runge-Kutta integration of the familiar Lorenz system:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} \sigma(y - x) \\ rx - y - xz \\ xy - bz \end{bmatrix} \quad (1)$$

with $\sigma = 16$, $b = 4$, $r = 45.92$ — values that yield the classic two-lobed chaotic Lorenz attractor — and an integrator timestep of 0.01. The second time series was derived from a similar integration and sampling of the Rössler system:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} -(y + z) \\ x + cy \\ a + xz - bz \end{bmatrix} \quad (2)$$

with $a = 0.2$, $b = 10.0$, $c = 0.15$ and a timestep of 0.01. The third data set was gathered experimentally from an angle sensor on a parametrically driven pendulum, and the fourth signal was a uniform sampling of a simple, noise-free sine wave covering several periods. The first three data sets were chosen because they exhibit chaotic dynamics of known dimension; the sine-wave signal was chosen because it makes the basic elements of RP structure easy to see and understand. Each of these signals covered a different range of values and thus required a different threshold corridor⁵.

⁵For each time series, the same threshold corridor value was used for each of the four embedding dimensions.

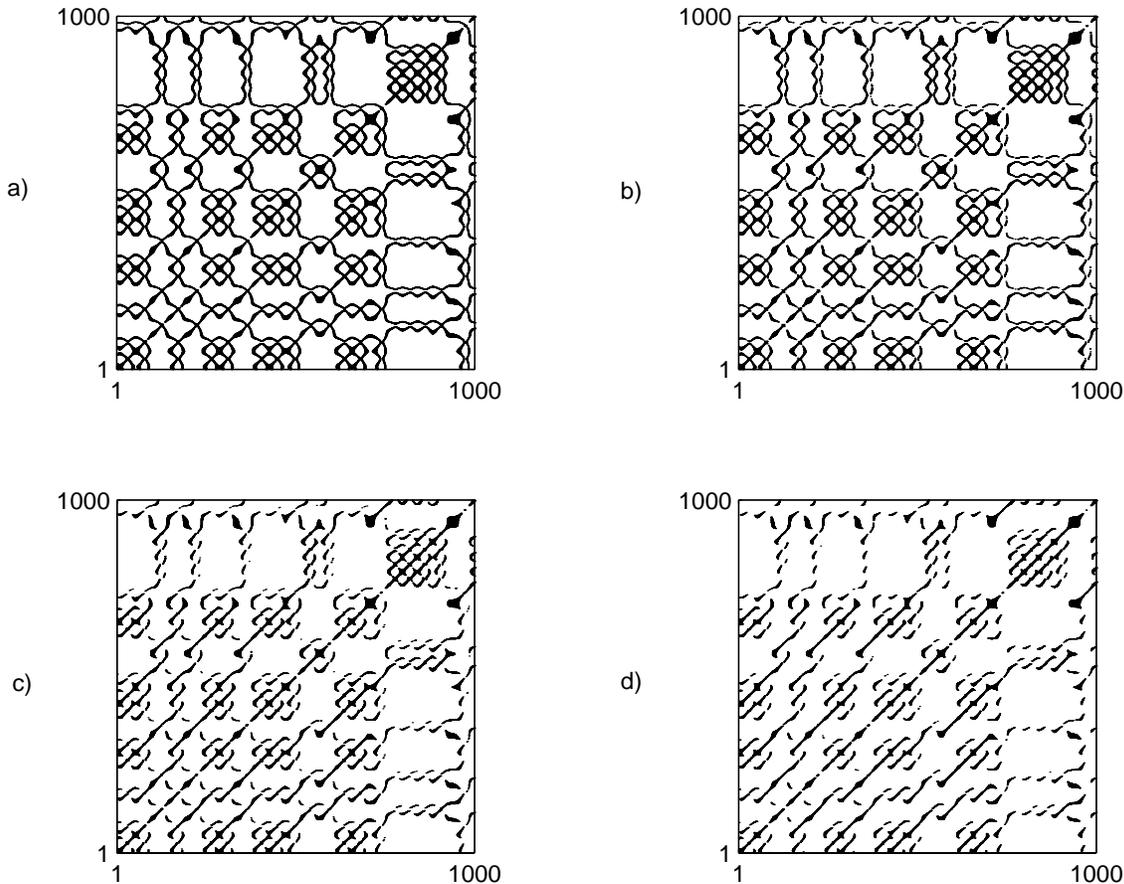


Figure 5: These four TRPs were generated from data from a numerical integration of the Lorenz system. TRPs (a)–(d) represent embedding dimensions 1–4, respectively; identical threshold corridors and time delay values were used for all four plots. Note the striking structural similarity: the only apparent variation is a lightening of the TRP with increasing embedding dimension.

Figure 2 on page 6 shows recurrence plots generated from the chaotic pendulum data set with four different embedding dimensions. Note that the qualitative features of the four TRPs are essentially the same. The major difference is in the gradual fading of the TRPs as d_E increases. This is largely due to the fact that an identical threshold corridor was used to generate each plot. As noted in section 2.2, the distance between any two points on a delay-coordinate reconstructed trajectory is *non-decreasing* with increasing d_E . The threshold corridors in the TRPs in figure 2 are all of the form $[\epsilon, r]$, where ϵ is some small number⁶. The net effect of a fixed threshold corridor in the face of increasing embedding dimension is that fewer and fewer pairs of points are recurrent, causing the TRP to fade. RP structure also appears to be independent of d_E for the Lorenz and Rössler data sets, as shown in figures 5 and 6. The same is also true of the sine-wave time series (figure 7).

⁶We do not use 0 as the lower bound in order to avoid trivial recurrences.

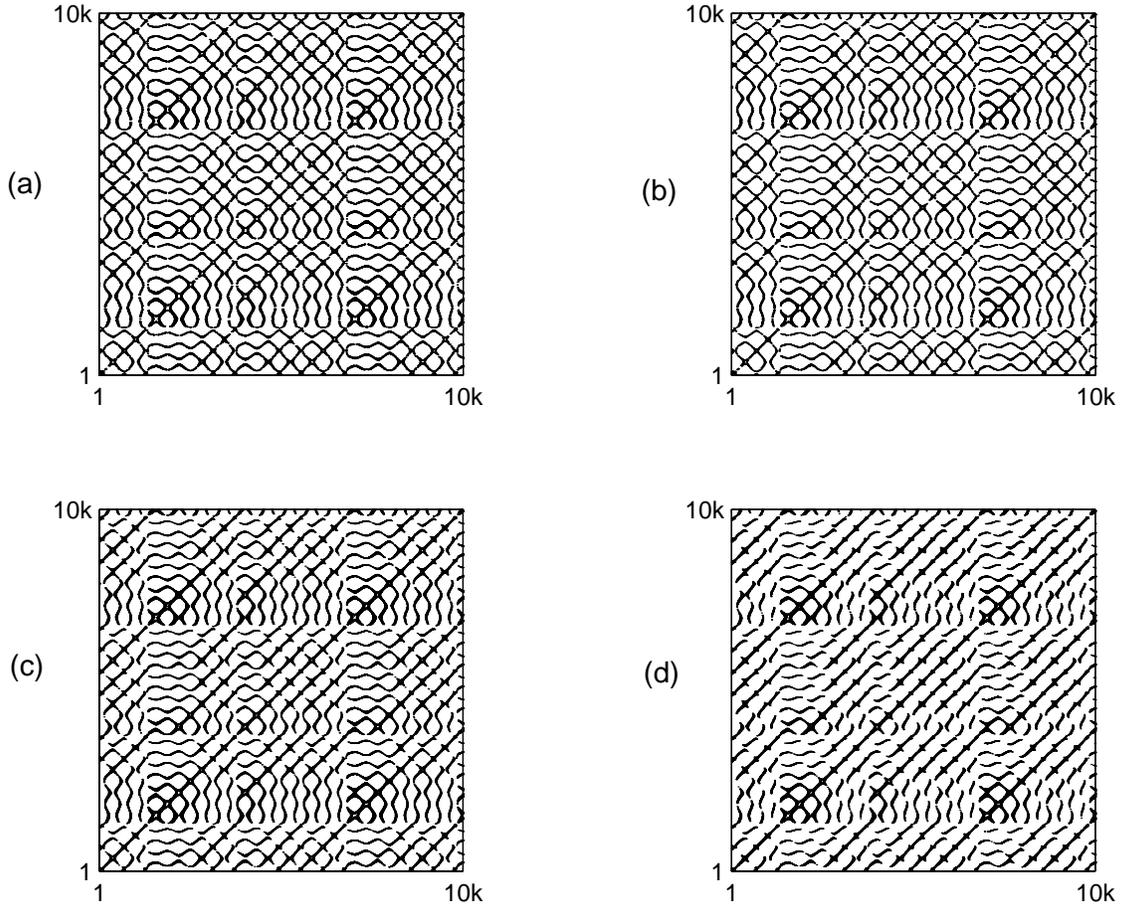


Figure 6: These four TRPs were generated from data from a numerical integration of the Rössler system. TRPs (a)–(d) represent embedding dimensions 1–4, respectively, and all four plots used identical threshold corridors and time delay values. Again, note the striking structural similarity: the only apparent variation is a lightening of the TRP with increasing embedding dimension.

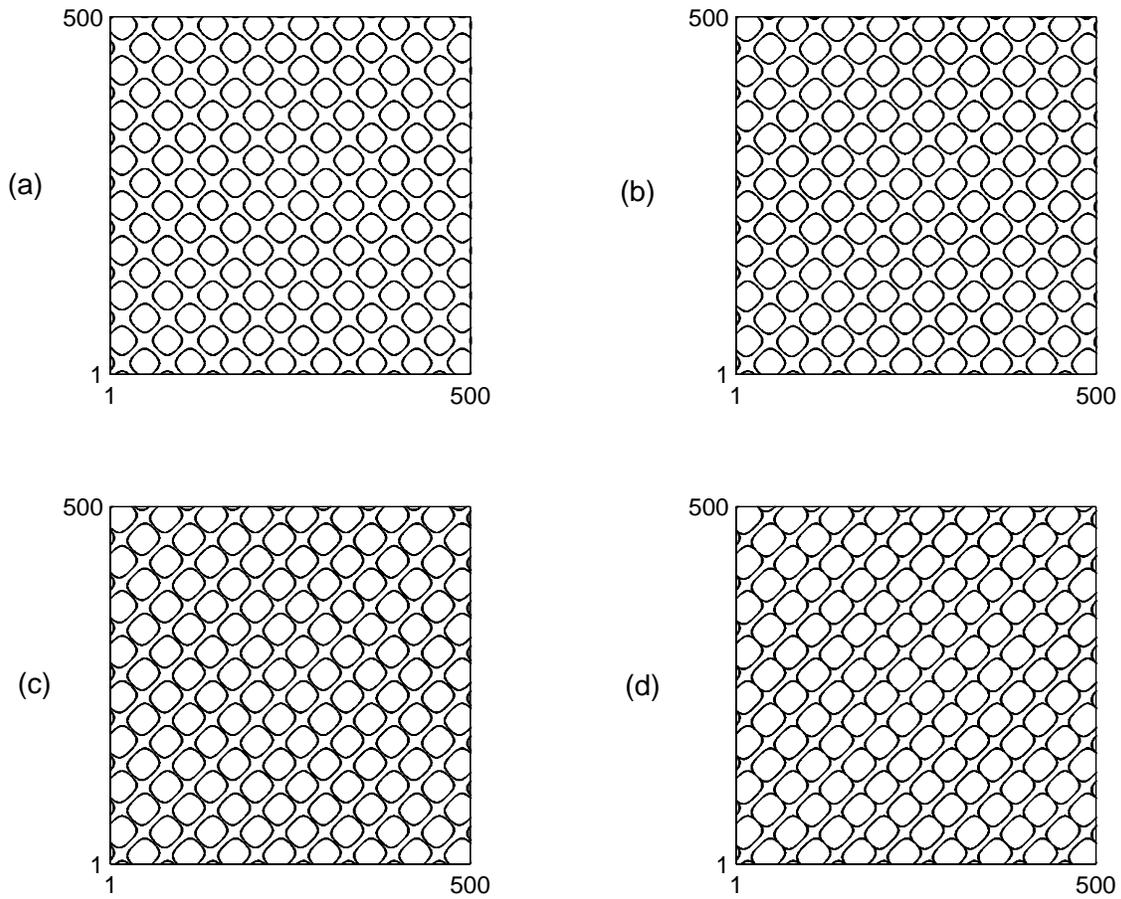


Figure 7: These four TRPs were generated from noise-free sine-wave data. TRPs (a)–(d) represent embedding dimensions 1–4, respectively; all four plots used identical threshold corridors and time delay values. As in the three other data sets, the embedding dimension does not appear to affect the general nature of the RP structure.

The fact that the qualitative features of RPs are independent of embedding dimension in all of the (very different) data sets that we have examined is highly suggestive. Analytical justification of this strong pattern of similarity is one of the current focuses of our research effort. Of particular interest are two aspects of this behavior. First, why does it happen? Recurrence plots expose *distance* relationships in the data — relationships that, should (perhaps) change as the dynamics are unfolded by the embedding process. An obvious ansatz is that since the time series really is only one dimensional, perhaps it makes sense that recurrence patterns present in one-embeddings capture its essential properties. However, such an explanation seems to violate the whole point of the delay-coordinate embedding process, wherein the *multi-dimensional* characteristics of the dynamics are recovered from one-dimensional signals. The second goal of our current research is to work out a formally justifiable and meaningful way to codify the structural characteristics of recurrence plots. Lumped statistical measures, such as those of RQA, are a good starting point, but these methods cannot capture the spatiotemporal details of the dynamics. We are currently investigating various pattern recognition and topological analysis techniques — perhaps focusing on the unstable periodic orbits embedded within the chaotic attractors in the data sets — in order to develop methods that allow sensible and useful structural classification and comparison of RPs.

3.2 Applying RQA to the Dynamic Logistic Map Time Series

As mentioned in section 1, RPs can be used to detect changes in the dynamics of a system. Trulla, Giuliani, Zbilut, and Webber[13], for instance, used RQA to recognize bifurcations in the dynamics of the logistic equation. In this section, we duplicate these results *without embedding the data*⁷.

The well-known logistic map is given by the following construction:

$$x_{n+1} = \alpha x_n(1 - x_n)$$

for some choice of α , which we will call the dynamic parameter, and some initial condition x_0 . When studying this well-known system, one usually fixes the parameter α and iterates the map from x_0 , discarding the first few hundred (or thousand) iterates in order to allow any transient behavior to die out. This map exhibits a variety of periodic and chaotic behaviors for dynamic parameter values between $\alpha = 2.8$ to 4.0 , and the parameter values at which the bifurcations occur are also well known.

In order to show that RQA is an effective method for detecting bifurcations, Trulla, Giuliani, Zbilut, and Webber[13] applied their analyses to a somewhat different logistic map time series — one in which the *transient behavior was not allowed to die out*. They began with dynamic parameter $\alpha = 2.8$ and initial condition $x_0 = 0.6$. After each iteration, they incremented α by 0.00001 until $\alpha = 4.0$, yielding a nonstationary time series of 120,001 measurements of x . They then embedded the data in three-dimensional space, divided the trajectory into 11,920 epochs of 800 points, in which epoch $(k + 1)$

⁷In the original experiment, Trulla, Giuliani, Zbilut, and Webber used $d_E = 3$.

began ten points farther along⁸ the original time series than epoch k , and calculated REC and DET for each epoch in the sequence. Their main conclusion was that these two RQA statistics were better than classical statistical measurements, such as mean and standard deviation, at distinguishing between periodicity and chaos.

We repeated this analysis and obtained identical results. The results for $d_E = 3$ appear in figure 8: part (a) shows the signal itself, while (b), (c), and (d) give the REC, DET, and $1/line_{max}$ results, respectively, for $d_E = 3$ and time delay $\tau = 1$. The vertical lines on the figure indicate bifurcation points, as determined by careful examination of the signal. On the whole, the RQA statistics pick these bifurcation points up very nicely. Where the bifurcation involves a transition between periodic and chaotic behavior, there are abrupt changes in all three statistics; in the direction from periodicity to chaos, REC and DET increase and $1/line_{max}$ decreases (and vice versa). Changes in all three statistics also accompany bifurcations between different types of periodic orbit — e.g., the two-cycle to four-cycle bifurcation near $\alpha = 3.48$. In $1/line_{max}$, these appear as small spikes, which are somewhat obscured by the scale and the vertical lines. At the period-doubling bifurcations, REC drops sharply because points are suddenly not recurrent with respect to points on the other branch of the orbit. DET also drops, although not as sharply; the bifurcation leads the dynamics into *another* periodic regime, where the recurrent points are still organized into coherent structures.

In order to explore the role that embedding dimension plays in RQA, we then repeated the same analysis with *different* embedding dimensions. The results were virtually identical. Figures 9, 10, and 11 illustrate this, showing the same three RQA statistics for the cases $d_E = 1, 2, 3$, and 4. The vertical lines in these figures are the same as those in figure 8; in all cases the time delay $\tau = 1$. The correspondence is striking: all three RQA statistics pick up the bifurcation points just as well when $d_E = 1$ as when $d_E = 3$, as is apparent from a vertical comparison of figures 9–11. There are some minor shape differences in the plots, but the bifurcation points are equally distinguishable in all four embeddings. This implies that the results and claims in [13] may not really require embedding the data.

The logistic map is, of course, fundamentally one-dimensional, so one might naturally expect *any* data analysis results to be independent of d_E ⁹. The next logical step in exploring the effects of embedding dimension on the qualitative and quantitative structure of recurrence plots is to run a similar experiment on data from a higher-dimensional system.

3.3 Applying RQA to the Dynamic Lorenz Time Series

The results in the previous section suggest that RQA performed on 1-embeddings of time-series data is a useful method for detecting bifurcations. In this section, we test

⁸In this formulation, epoch k overlaps with the last 790 points of epoch $(k - 1)$ and with the first 790 points of epoch $(k + 1)$

⁹Trulla *et al.* do not give any justification for having embedded this time series in R^3 .

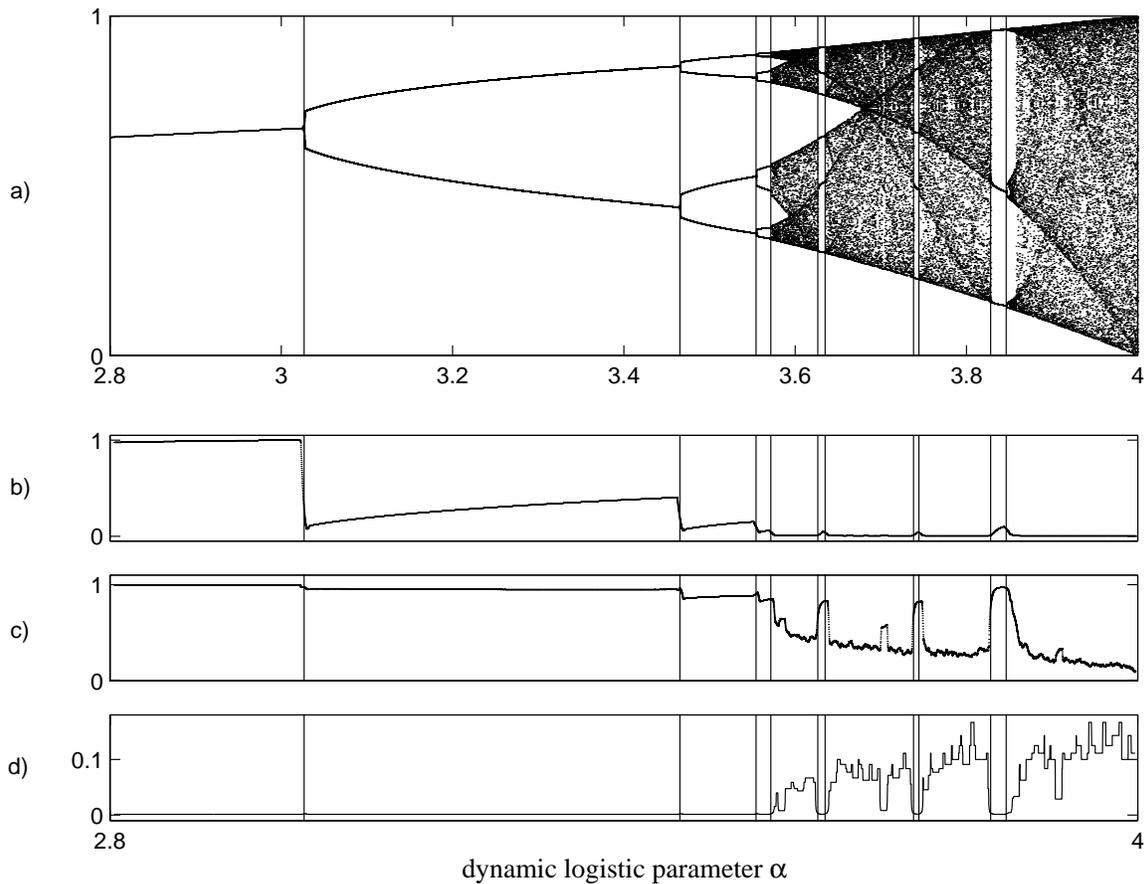


Figure 8: (a) the dynamic logistic map time series, plotted vs. parameter α . The vertical lines are drawn at bifurcation points of the signal. This is not the standard bifurcation diagram of the logistic map; here, the transients were not allowed to die out. (b), (c) and (d) are REC, DET, and $1/line_{max}$, respectively, plotted vs. parameter α , using embedding dimension $d_E = 3$ and time delay $\tau = 1$. In all three RQA calculations, threshold corridor was $[0, 0.0149]$ and minimum line length was 2. Note that the RQA statistics are effective indicators of the dynamical bifurcations.

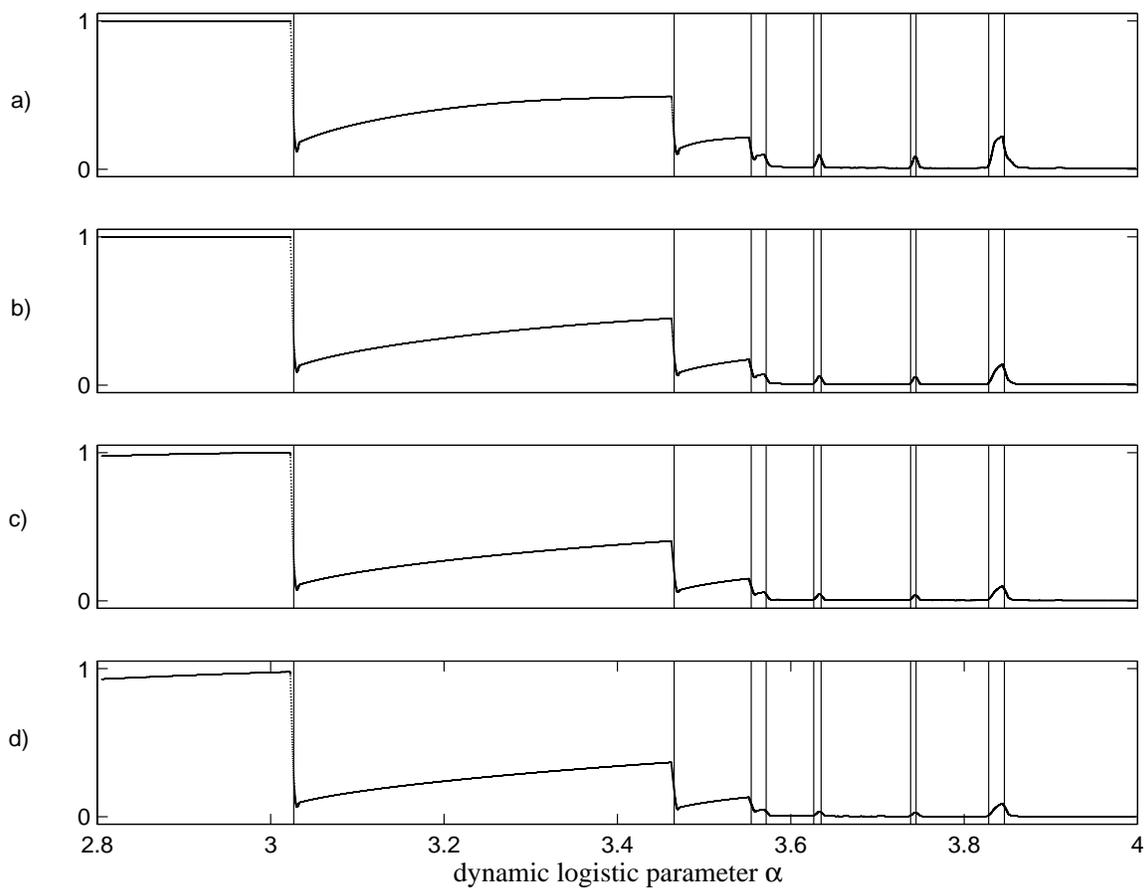


Figure 9: The RQA statistic REC computed from the dynamic logistic map data set shown in part (a) of the previous figure. (a)–(d) show REC results for embedding dimension $d_E = 1$ –4, respectively, with time delay $\tau = 1$. Part (c) is the same plot as part (b) of the previous figure. Note how the REC values shown in all four plots pick up the bifurcation points of the signal equally well — even the one with $d_E = 1$. Threshold corridor was $[0, 0.0149]$.

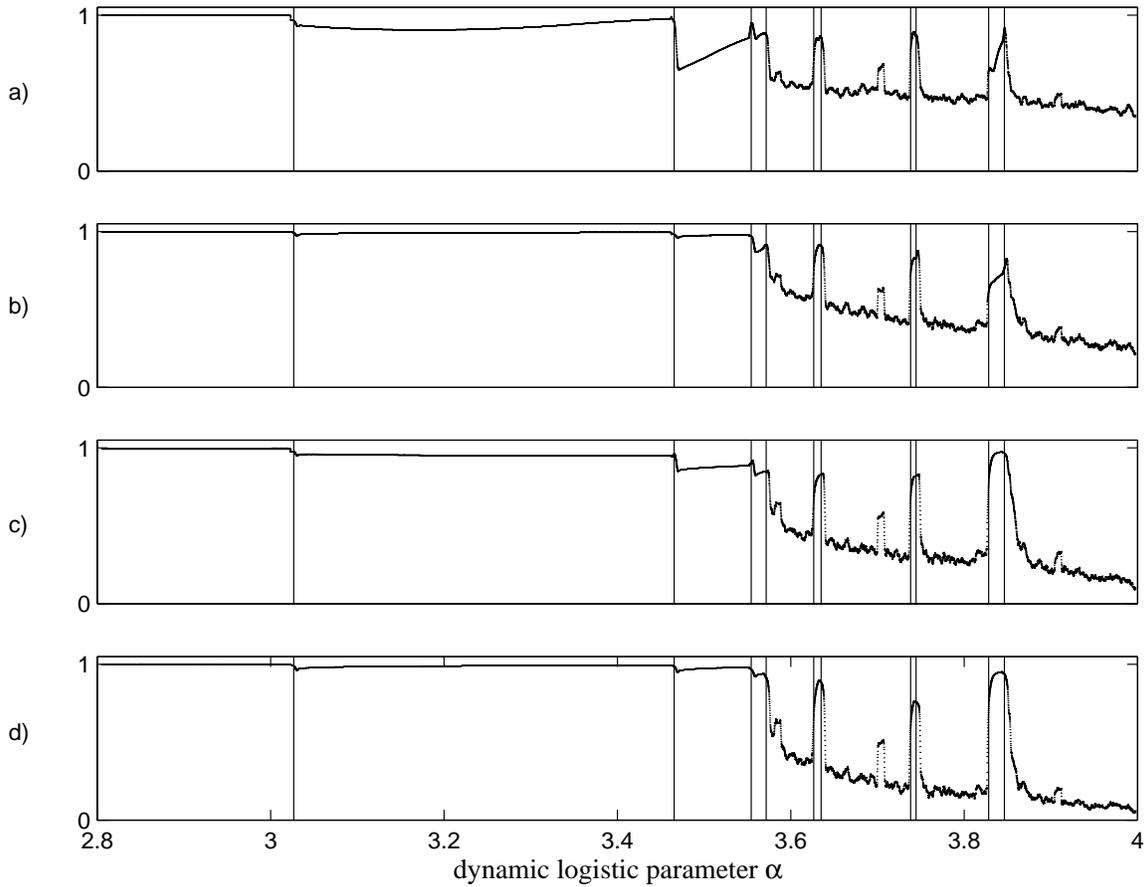


Figure 10: The RQA statistic DET computed from the dynamic logistic map data set shown in part (a) of figure 8. (a)–(d) show DET results for embedding dimension $d_E = 1$ –4, respectively, with time delay $\tau = 1$. As before, all four DET computations pick up the bifurcation points of the signal, including the $d_E = 1$ calculation. Minimum line length was 2. Part (c) is the same plot as part (c) of figure 8.

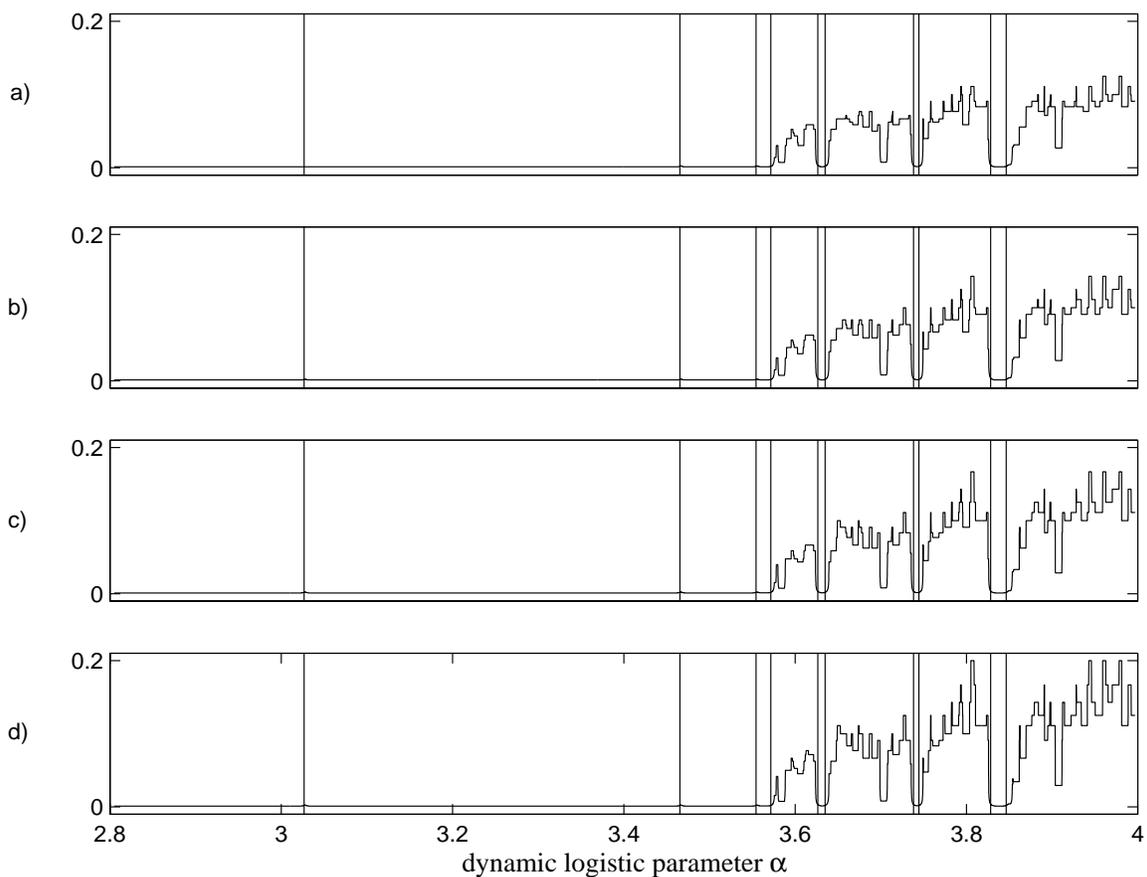


Figure 11: The RQA statistic $1/line_{max}$ computed from the dynamic logistic map data set shown in part (a) of figure 8. (a)–(d) show $1/line_{max}$ results for embedding dimension $d_E = 1-4$, respectively, with time delay $\tau = 1$. As in the previous two figures, this statistic clearly picks up the bifurcation points of the signal, even if the data were not embedded. The initial flat area of this plot corresponds to the long initial periodic regime of the signal, for which we would expect a zero value of the leading positive Lyapunov exponent. Part (c) is the same plot as part (d) of figure 8.

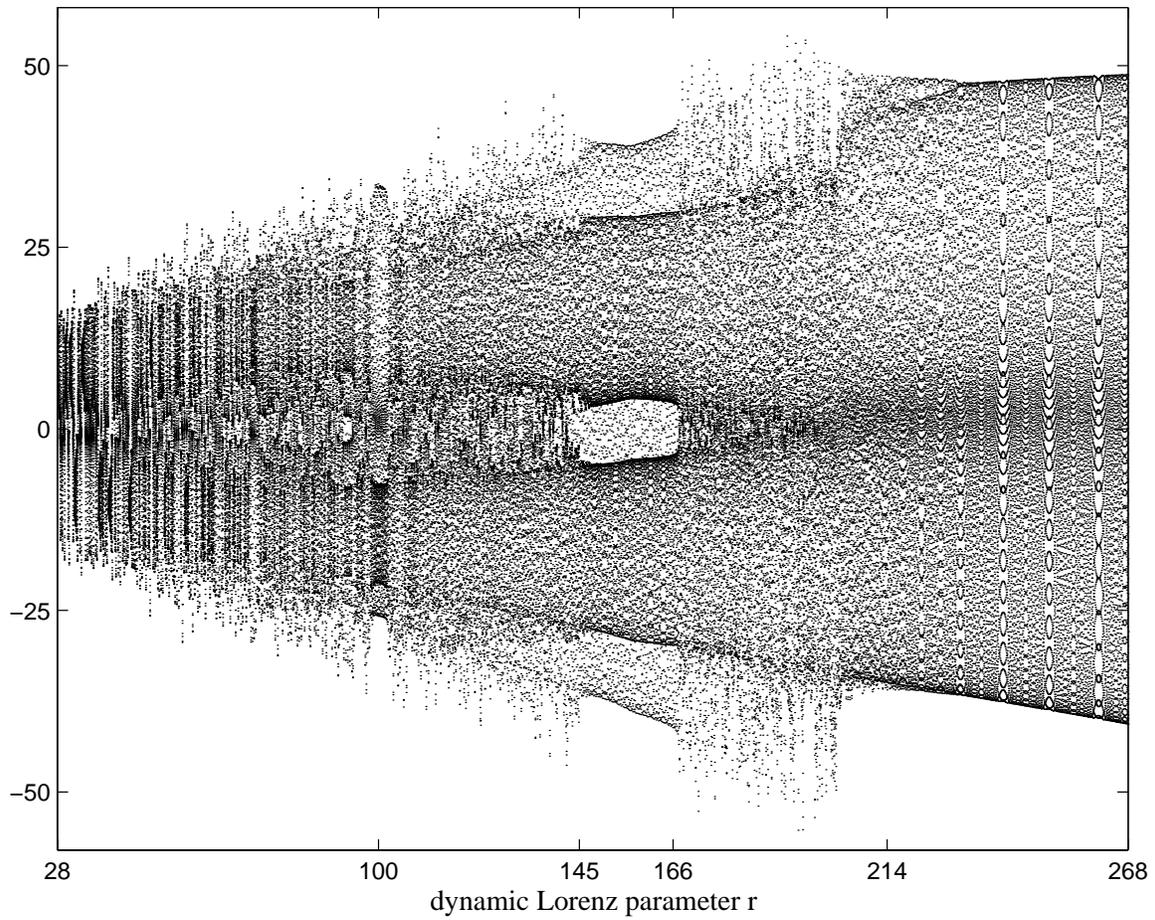


Figure 12: Dynamic Lorenz time series. This signal was generated from the Lorenz system of equations (1) by fixing $\sigma = 10.0$, $b = 8/3$, and incrementing the dynamic parameter r from 28.0 to 268.0 by 0.002 at each integration step.

this hypothesis further by applying RQA to various embeddings of time-series data from the Lorenz system. Specifically, we integrated the Lorenz equations numerically with fourth-order Runge-Kutta and a timestep of 0.01, holding σ and b fixed (10 and $8/3$, respectively) and varying r from 28.0 to 268.0, and used the x value as the time series. The r -increment was 0.002 per integration step. The signal is shown in figure 12. For this particular range of parameter values, the behavior of this system is extremely well studied; see, for example, [10, 11]. As in the case of the logistic map data of section 3.2, this time series is somewhat unusual: it includes some amount of transient behavior at each step. One could, of course, allow the transient to die out each time r was incremented before starting to gather time series samples. However, part of the point of this technique is to be able to detect bifurcations as they occur, so an experiment with dynamically changing behavior is an appropriate test case.

The results of RQA on the dynamic Lorenz time series are shown in Figures 13, 14, and 15. The vertical lines on these plots identify three of the known periodic windows: $99.524 < r < 100.795$, $145 < r < 166$, and $r > 214.4$. It is clear from figure 13 that $1/line_{max}$ is an effective periodic window indicator: all three of the windows correspond to markedly lower values for this RQA statistic. The fluctuation of the values at the beginning of the periodic windows probably stems from the transient nature of the signal; if the transient is slow, the dynamics may require some time to reach the periodic orbit. The flat areas at the very left of the $1/line_{max}$ plots — which grow shorter as d_E increases — might also lead one to conclude (mistakenly) that the signal is periodic for r -values just greater than 28. We believe that this anomaly is due in part to the fact that the attractor size changes with r , while we used a single, fixed threshold corridor for the entire signal¹⁰. A better experiment might be to adapt the threshold corridor in accordance with attractor size; we are currently working out how to do so in a sensible manner. At any rate, it is clear that all four $1/line_{max}$ plots are qualitatively quite similar. Figure 14 shows REC results for the same signal. The first periodic window (around $r = 100$) is markedly obvious and the second one is indicated by a leveling out of REC, but the third one does not leave a strong, well-delineated signature on these plots. Again, this is probably due in large part to the effects of the change in attractor size with parameter r . The similarity of the four plots is the main point we wish to stress here. The DET results for the dynamic Lorenz signal are shown in figure 15. The obvious similarity between the four plots further supports our hypothesis that the $d_E = 1$ RQA calculation is a sufficient indicator of the changes in the dynamics of this signal. All four curves show a marked increase for the first known periodic window (around $r = 100$) and an increase and leveling out across the second periodic window (between $r = 145$ and $r = 166$), followed by an abrupt decrease above $r = 166$. The third periodic window ($r > 214.4$) is reflected in all four plots by a significant increase in DET.

It is worth noting that the DET results are highly sensitive to the choice of minimum line length. For example, when we ran the same experiment using a minimum line length of 2 (instead of 10), the $d_E = 1$ DET graph was significantly different from the other

¹⁰We did not see this type of anomaly in the dynamic logistic map experiment because the domain of that map is $[0,1]$, independent of α .

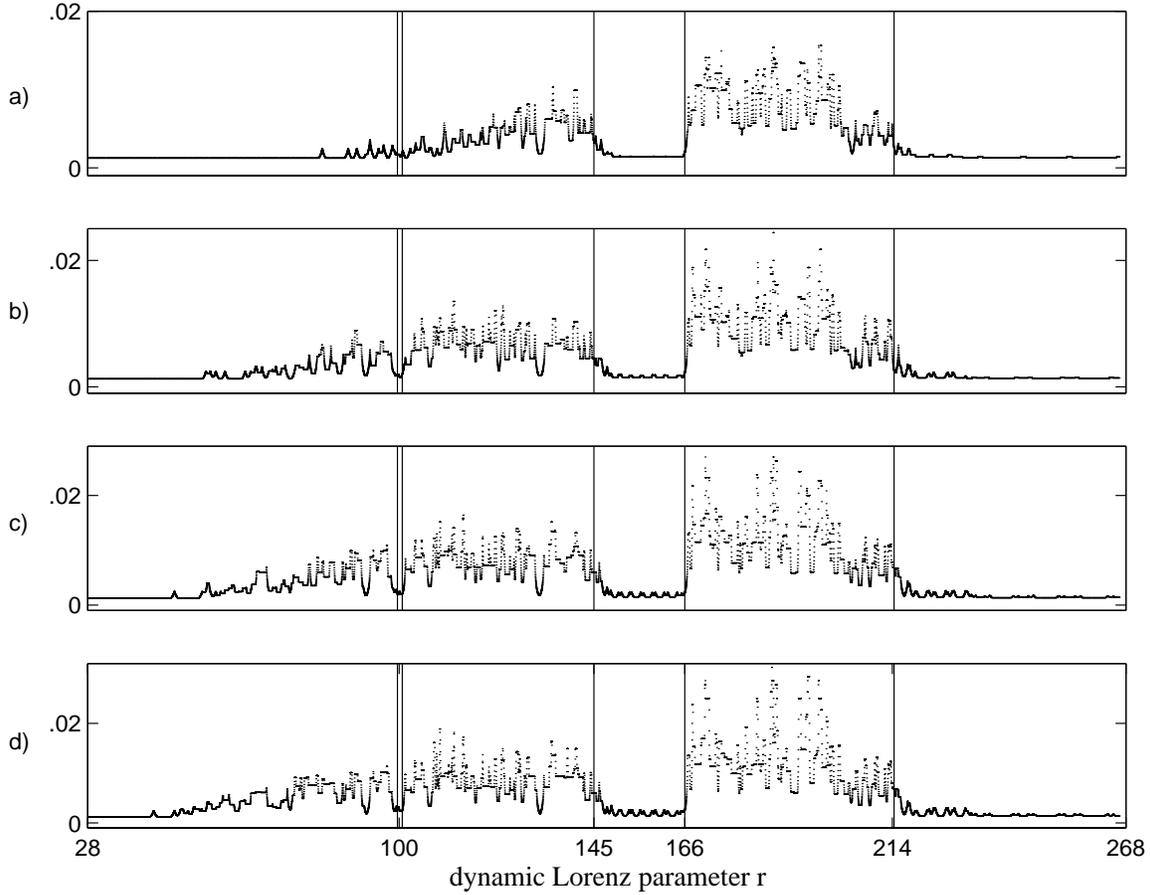


Figure 13: RQA statistic $1/line_{max}$ for the dynamic Lorenz time series. (a)–(d) show $1/line_{max}$ results for $d_E = 1-4$, respectively, with time delay $\tau = 1$. The vertical lines indicate several parameter ranges for which the Lorenz system is known to be periodic. This statistic was the best one overall at picking up the known periodic windows, which correspond to low (near zero) values of $1/line_{max}$. Since $1/line_{max}$ is related to the leading positive Lyapunov exponent, this makes sense; one would expect the leading Lyapunov exponent to be positive in chaotic regions and zero in periodic regions. Additionally, there is little significant variation between the four plots. (Note different vertical scales.)

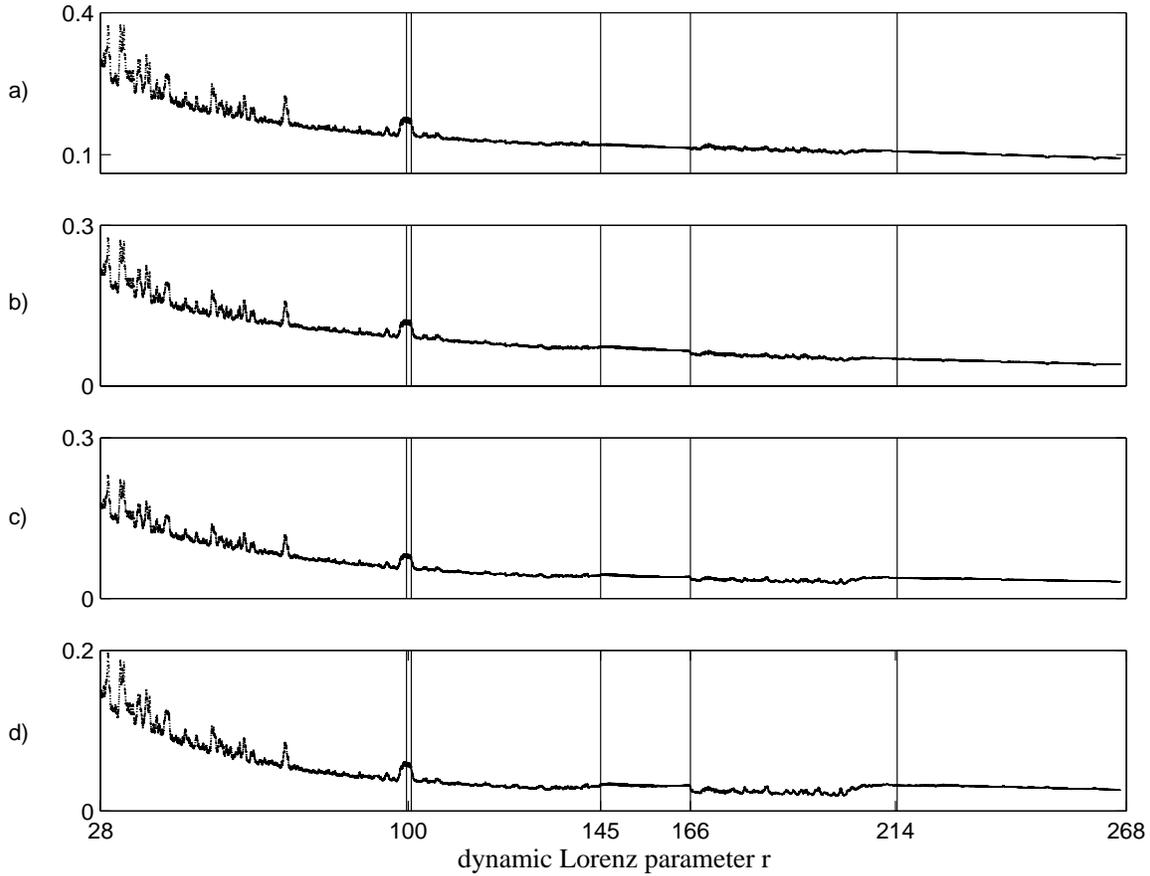


Figure 14: RQA statistic REC for the dynamic Lorenz time series. (a)–(d) give REC results for $d_E = 1-4$, respectively, with time delay $\tau = 1$. The vertical lines indicate several parameter ranges for which the Lorenz system is known to be periodic. This statistic was the best indicator of the first periodic window (the bulge at $r \approx 100$). As before, there is little significant variation between the four plots. (Note different vertical scales.)

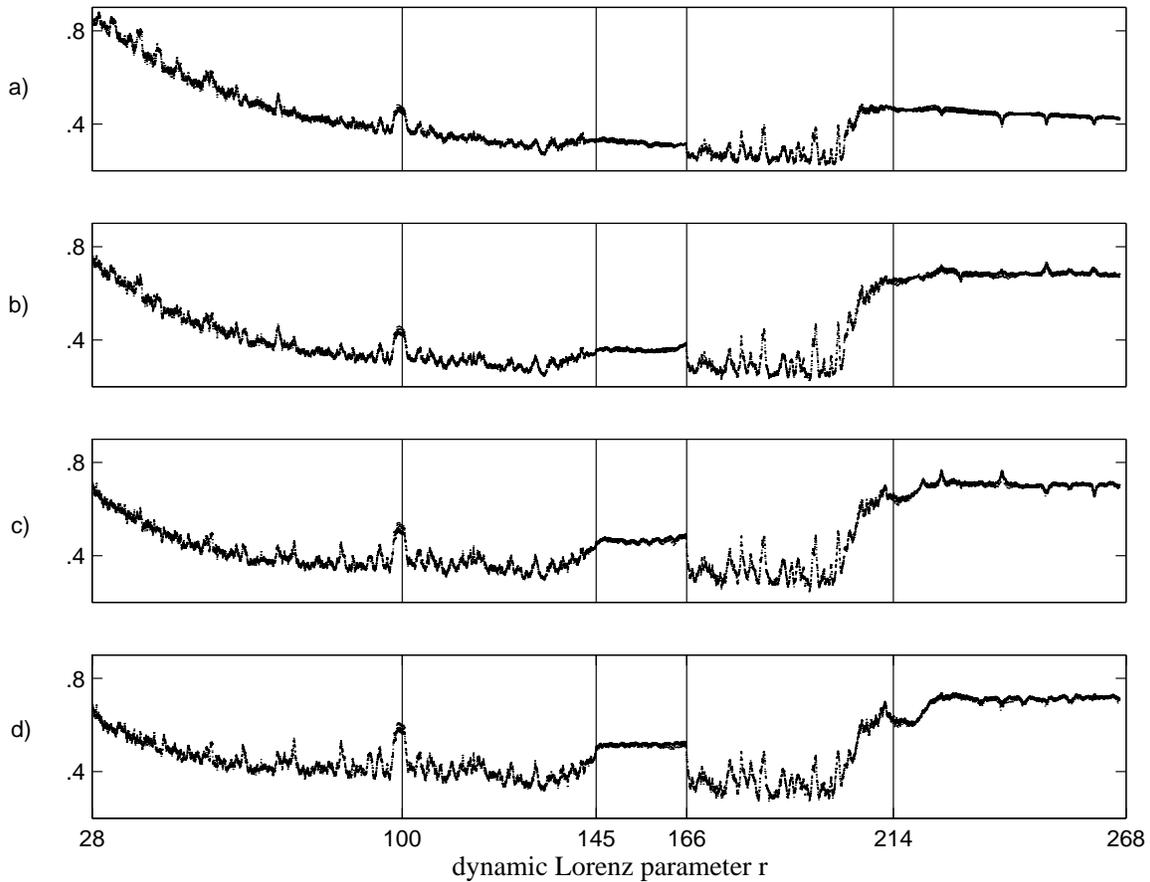


Figure 15: RQA statistic DET for the dynamic Lorenz time series. (a)–(d) give DET results for $d_E = 1-4$, respectively, with time delay $\tau = 1$ and a minimum line length of 10. Note the similarity of the plots, which all give indication of the changing dynamics that this signal is known to exhibit.

three. It is important to be careful when choosing parameters for RQA; we recommend running experiments using several different parameter choices. One of the shortcomings of RQA is revealed here – that of how to intelligently choose the analysis parameters.

The obvious next step, on which we are currently working, is to perform RQA analyses on dynamic data sets from systems whose known dimension is greater than three. Similar results on such a data set would further strengthen our hypothesis that in order to perform RQA on a data set one need *not* first perform any delay-coordinate embedding.

4 Conclusions and Future Work

Recurrence plots of experimental data appear to be quantitatively and qualitatively independent of embedding dimension. The results presented here support this hypothesis through diverse examples in several dynamical systems. In particular, we have demonstrated that the Recurrence Quantification Analysis of Trulla, Giuliani, Zbilut, and Webber[13], a statistical, quantitative RP analysis method, does not require time-series embedding. That is, for the purpose of finding bifurcation points in data sets, $d_E = 1$ is adequate for computation of the RQA statistics. RQA analysis involves *relative* comparisons; we do not know, say, whether a REC value of 2.1% or 1.7% is the *correct* value for a given data set (or even what it might mean for one value to be “correct”). For the purpose of finding bifurcation points in a data set, however, the $d_E = 1$ calculation of RQA statistics works just as well as the calculations that involve a higher embedding dimension.

RQA is the best extant set of tools for RP analysis, but its lumped statistics cannot capture the details of the spatiotemporal dynamics of a time series. The ultimate goal of this line of research is to develop structural analysis tools, based in theory and verified in experiment, that will allow us to systematically and reliably classify the qualitative structure of these intriguing plots. We are focusing on methods that exploit pattern recognition techniques to identify and classify the topological features of RPs.

If $d_E = 1$ recurrence plots do indeed capture the same information as do those generated using higher embedding dimensions, then recurrence plot analysis can proceed in the absence of the considerable difficulties and hazy heuristics that inform the delay-coordinate embedding process. Of course, the difficulties that are inherent to recurrence plot analysis itself will remain, but the overall process will have been simplified significantly.

References

- [1] H. D. I. Abarbanel. *Analysis of Observed Chaotic Data*. Springer, 1995.
- [2] M. Casdagli. Recurrence plots revisited. *Physica D*, 108:12–44, 1997.
- [3] J.-P. Eckmann, S. Kamphorst, and D. Ruelle. Recurrence plots of dynamical systems. *Europhysics Letters*, 4:973–977, 1987.
- [4] P. Kaluzny and R. Tarnecki. Recurrence plots of neuronal spike trains. *Biological Cybernetics*, 68:527–534, 1993.
- [5] D. Kaplan and L. Glass. *Understanding Nonlinear Dynamics*. Springer-Verlag, New York NY, 1995.
- [6] M. B. Kennel, R. Brown, and H. D. I. Abarbanel. Determining minimum embedding dimension using a geometrical construction. *Physical Review A*, 45:3403–3411, 1992.
- [7] N. Packard, J. Crutchfield, J. Farmer, and R. Shaw. Geometry from a time series. *Physical Review Letters*, 45:712, 1980.
- [8] T. Sauer, J. A. Yorke, and M. Casdagli. Embedology. *Journal of Statistical Physics*, 65:579–616, 1991.
- [9] C. E. Shannon. *The Mathematical Theory of Communication*. University of Illinois Press, 1964.
- [10] C. Sparrow. The Lorenz equations. In H. Haken, editor, *Synergetics: A Workshop*, pages 111–134. Springer Verlag, 1977.
- [11] S. H. Strogatz. *Nonlinear Dynamics and Chaos*. Addison-Wesley, Reading, MA, 1994.
- [12] F. Takens. Detecting strange attractors in fluid turbulence. In D. Rand and L.-S. Young, editors, *Dynamical Systems and Turbulence*, pages 366–381. Springer, Berlin, 1981.
- [13] L. Trulla, A. Giuliani, J. Zbilut, and C. Webber. Recurrence quantification analysis of the logistic equation with transients. *Physics Letters A*, 223:255–260, 1996.
- [14] C. Webber and J. Zbilut. Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, 76:965–973, 1994.