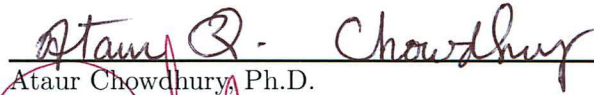


RECURRENCE ANALYSIS METHODS  
FOR THE CLASSIFICATION OF NONLINEAR SYSTEMS


By


Mark Graybill

RECOMMENDED:

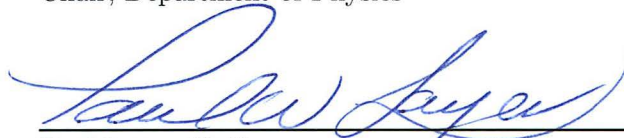
  
Ataur Chowdhury, Ph.D.

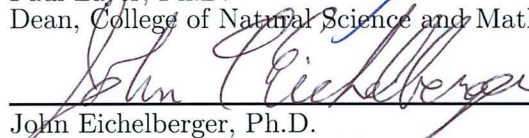
  
David Newman, Ph.D.

  
Renate Wackerbauer, Ph.D.  
Advisory Committee Chair

  
Curt Szuberla, Ph.D.  
Chair, Department of Physics

APPROVED:

  
Paul Layer, Ph.D.  
Dean, College of Natural Science and Mathematics

  
John Eichelberger, Ph.D.  
Dean of the Graduate School

4/25/14  
Date



RECURRENCE ANALYSIS METHODS  
FOR THE CLASSIFICATION OF NONLINEAR SYSTEMS

A  
THESIS

Presented to the Faculty  
of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

By

Mark Graybill, B.A.

Fairbanks, Alaska

May 2014

## Abstract

Recurrence is a common phenomenon in natural systems: A system enters and leaves a state, but after a given period of time, passes near that same state again. Many complex signals, such as weather cycles, heartbeats, or neuron firing patterns, all show recurrence. The recurrence plot (RP) displays all times  $j$  where a system returns near a state it has occupied at time  $i$ , giving rise to upward-sloping diagonal lines where a system follows a recurrent path, orthogonal lines when the system changes very slowly, or many disconnected points where a system's behavior is unpredictable. Investigation of the RP can then proceed through recurrence quantification analysis (RQA).

Three new measures for RQA were developed: *diagonality*, quantifying diagonal lines, *verticality*, quantifying vertical lines, and *periodicity* quantifying the arrangement of recurrence points in periodic structures. These new measures were applied alongside classical recurrence measures to explore trends in random data, identify periodicity and chaotic behavior in the logistic map, estimate the dimensionality of the Lorenz attractor, and discriminate between persistent data signals.

In collaboration with biologist Dr. Michael Harris, RQA methods were applied to the discrimination of two neuron types: serotonergic cells are believed to stimulate respiration, while nonserotonergic cells are implicated in respiratory inhibition. Typical discrimination methods compare mean and standard deviation of firing rates to a reference line, which correctly classifies serotonergic cells but incorrectly classifies many nonserotonergic cells. Voltage signals from such cells were converted into inter-spike intervals. Convergence required trials containing over 300 spikes for biological methods, and over 1000 for full investigation using RQA. Whether such cells can be discriminated from baseline firing patterns remains an open question.



## Table of Contents

Signature Page . . . . .	i
Title Page . . . . .	iii
Abstract . . . . .	v
Table of Contents . . . . .	vii
List of Figures . . . . .	ix
Acknowledgements . . . . .	xi
<b>Chapter 1: Introduction . . . . .</b>	<b>1</b>
1.1 The Recurrence Plot . . . . .	2
1.2 Tolerance Threshold and Embedding Dimension . . . . .	4
1.3 Features of Recurrence Plots . . . . .	8
<b>Chapter 2: Methods . . . . .</b>	<b>11</b>
2.1 Classic Recurrence Measures . . . . .	11
2.2 Dependence of Classic RQA Measures on Changing RR . . . . .	15
2.3 New Measures For Recurrence Quantification Analysis . . . . .	20
2.4 Convergence Behavior of RQA Measures . . . . .	22
<b>Chapter 3: Application to Synthetic Data . . . . .</b>	<b>25</b>
3.1 The Logistic Map . . . . .	25
3.1.1 RQA of the Logistic Map . . . . .	28
3.2 The Lorenz System . . . . .	31
3.2.1 RQA for the Lorenz System . . . . .	33
3.3 Persistence Data . . . . .	35
3.3.2 RQA for Persistence Data . . . . .	37
3.4 Summary of Synthetic Data Investigation . . . . .	39
<b>Chapter 4: Application to Natural Data . . . . .</b>	<b>41</b>
4.1 Preliminary Analysis . . . . .	43
4.2 Minimum Trial Length . . . . .	45
4.3 Conclusions . . . . .	48
<b>Chapter 5: Conclusion . . . . .</b>	<b>49</b>
5.1 Outview . . . . .	51
<b>References . . . . .</b>	<b>53</b>



## List of Figures

1.1	RP of “Surfin’ Bird” . . . . .	3
1.2	RPs of continuous systems . . . . .	5
1.3	$\varepsilon$ -tube . . . . .	5
1.4	Variation of $\varepsilon$ . . . . .	6
1.5	Embedding dimension . . . . .	8
1.6	The Logistic map . . . . .	9
2.1	Trend for nonstationary signals . . . . .	15
2.2	$\varepsilon$ in random data . . . . .	16
2.3	RQA measures on random data . . . . .	17
2.4	$\varepsilon$ in chaotic data . . . . .	18
2.5	RQA measures on chaotic data . . . . .	19
2.6	New measures on random data . . . . .	21
2.7	New measures on chaotic data . . . . .	21
2.8	Convergence in random data . . . . .	23
2.9	Convergence in chaotic data . . . . .	24
3.1	Cobweb diagrams, logistic map . . . . .	26
3.2	Orbit diagram, logistic map . . . . .	27
3.3	RQA, logistic map . . . . .	29
3.4	Lorenz attractor . . . . .	31
3.5	Recurrence, Lorenz attractor . . . . .	32
3.6	RQA, Lorenz attractor . . . . .	34
3.7	Persistence data . . . . .	35
3.8	Return plots, persistence data . . . . .	36
3.9	Discrimination, persistence data . . . . .	37
3.10	RQA, persistence data . . . . .	38
4.1	Voltage signal . . . . .	42
4.2	Neuron classification . . . . .	42
4.3	Return plots, neuron data . . . . .	43
4.4	RPs, neuron data . . . . .	44
4.5	RQA, neuron data . . . . .	45



4.6	Classification by series length . . . . .	46
4.7	High frequency signal . . . . .	46
4.8	Convergence in statistics . . . . .	47
4.9	Convergence, RQA measures . . . . .	47

### List of Tables

3.1	Classification of Periodic Windows . . . . .	3
-----	--	---

## Acknowledgements

Thanks go to my advisor, Dr. Renate Wackerbauer, for her dedication to teaching and her interest in the recurrence plot, to Dr. Ataur Chowdhury and Dr. David Newman for their help and advice, and to Dr. Michael Harris and Kimberly Iceman for making their data available for analysis. Thanks also to Jacopo Lafranceschina and Douglas Ogata for their insights in problem-solving, and lastly, thanks to my wife, Jeanette Graybill, for her patience and support throughout my graduate career.



## Chapter 1

### Introduction

*“And this slow spider which creepeth in the moonlight, and this moonlight itself, and thou and I in this gateway whispering together, whispering of eternal things—must we not all have already existed? And must we not return and run in that other lane out before us, that long weird lane—must we not eternally return?”*

—Friedrich Nietzsche, Also Sprach Zarathustra

Recurrence is a common phenomenon in natural systems: A system enters and leaves a state, but, after a given period of time, comes near that same state once again. Such recurrence is not always regular, and the time interval between recurrences may be infinite in some systems, meaning that a given state will never recur. Nevertheless, many common systems, such as weather cycles, heartbeats, or neuron firing patterns, may be highly complex, and yet highly repetitive.

Not all deterministic systems are capable of producing complex recurrence. One-dimensional systems always either diverge, or converge to a steady state of perpetual recurrence. And two-dimensional systems at most give rise to simple periodicity. But in systems with greater than two dimensions, chaos may arise, characterized by topological *folding* in phase space, combined with *stretching* through sensitivity to initial conditions. The Lyapunov exponent is a measure of the strength of this sensitivity (see, *e.g.*, Schmitz, 2001 or Rosenstein *et al.*, 1993). In systems exhibiting a chaotic attractor, such sensitivity causes even adjacent trajectories to diverge. But such trajectories are ultimately recurrent because they are bounded; with nowhere else to go, they must eventually return near previously visited states. Yet in a chaotic system, the time between such recurrences can vary wildly, as the system shifts between chaotic regions, or in transient systems even escapes the chaotic saddle into a steady state where there is no time between recurrences, or to a divergent path that never returns.

## 1.1 The Recurrence Plot

When faced with a system exhibiting complex behavior, the scientific approach is to analyze, quantify, and particularly distinguish it from other systems; the *recurrence plot* (RP) is excellently suited to this purpose.

First introduced by Eckmann, the RP is a two-dimensional plot of a time series, with black dots at every  $i, j$  coordinate where the value of the series  $x$  at time  $i$  is approximately equal to the value of  $x$  at time  $j$  (Eckmann, Kamphorst, & Ruelle, 1987). Such a plot shows, at a glance, all places where a state in a time series loosely returns to a previously occupied state.

Consider for instance the lyrics to “Surfin’ Bird” (Frazier, White, Harris, & Wilson, 1963, side A, The Trashmen), with letters converted to numbers (Fig. 1.1 a). The recurrence plot (Fig. 1.1 b) is rich with recurrent points, shown here as black dots. The lower left corner corresponds to the start of the series, where the singer tells us that “b-bird’s the word.” Midway through the song, the singer goes into a seizure, singing “Bbbbbbbbbbbbbb...” and then many repeats of “papapapa...;” the cross at approximately (800, 800) represents this interruption before the song moves into its last half, where the singer repeats “papa ooma mow mow,” corresponding to the upper-right portion of the RP. Note a momentary return to “bird is the word” near (1400, 1400), which echoes into the upper-left and lower-right areas of the RP. The entire upper left and lower right triangle of any RP must be symmetric, since a recurrence between  $(i, j)$  entails recurrence at  $(j, i)$ . Note also the diagonal line running from the lower left to upper right corner. This main diagonal is referred to as the *line of interest* (LOI), and exists in all RPs because  $x(i) = x(j)$  must always be satisfied when  $i = j$ .

To formalize the concept of the RP, we begin with a dynamical system  $\dot{x} = f(x)$  with  $x \in \mathbb{R}$  being a state of the system.<sup>1</sup> Then the recurrence plot  $R(i, j)$  is described by

$$R(i, j) = \Theta(\varepsilon - |x(i) - x(j)|). \quad (1)$$

---

<sup>1</sup>In principle, one may consider a dynamical system in multiple dimensions,  $\dot{\vec{x}} = \vec{f}(\vec{x})$ , with  $\vec{x} \in \mathbb{R}^n$ . This is not a problem for generating the recurrence plot, but throughout this document the one-dimensional time series is used as a basis for analysis.

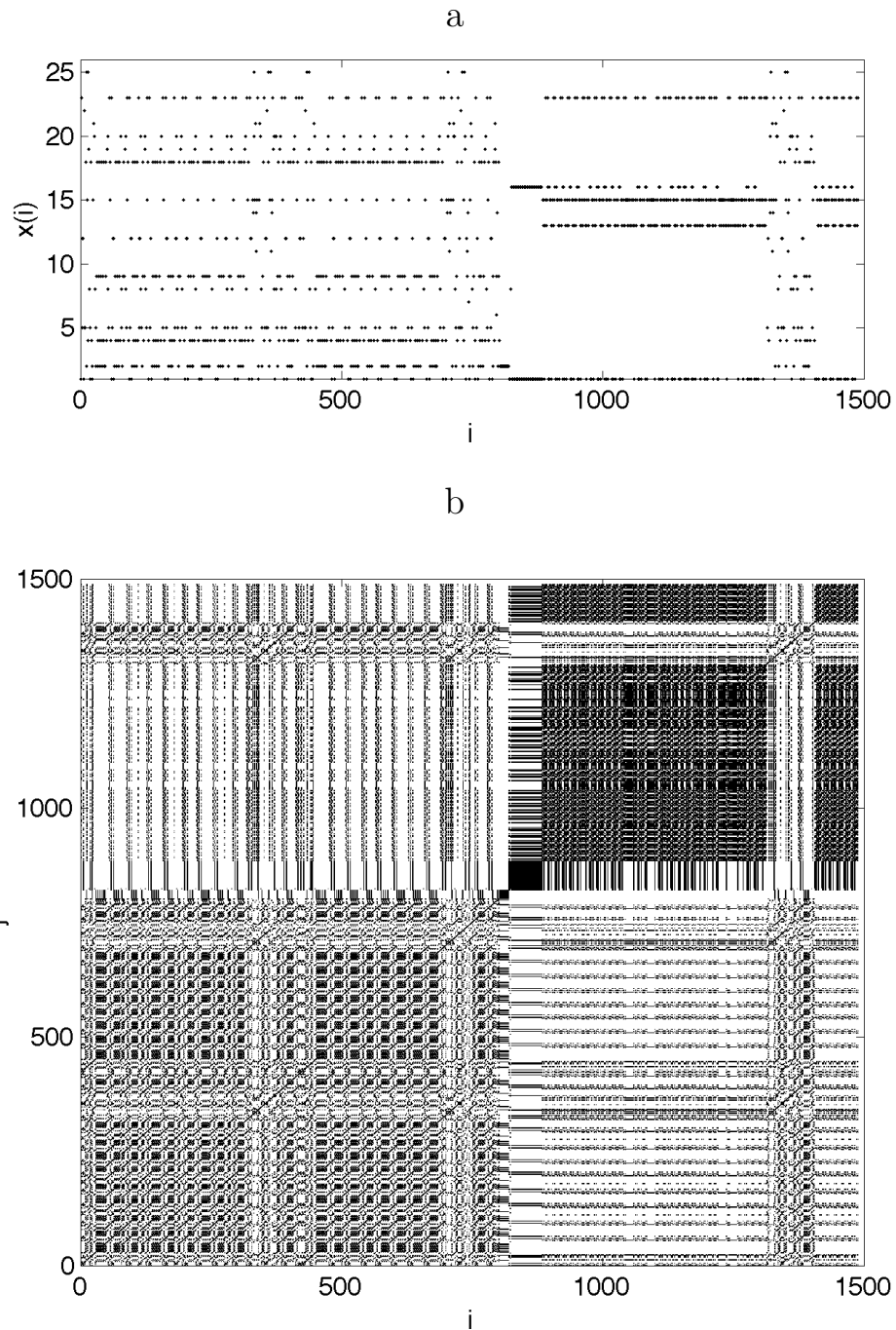


Figure 1.1: Lyrics to the Trashmen's "Surfin' Bird," with letters converted into numbers where  $a = 1$ ,  $b = 2$ ,  $c = 3$ , etc. (a), and recurrence plot (b).

This process compares two states at time  $i$  and  $j$ , plotting a black dot ( $R(i, j) = 1$ ) wherever the states come arbitrarily close to one another, and a white dot ( $R(i, j) = 0$ ) otherwise. Here, “arbitrarily close” is defined by the *tolerance threshold*,  $\varepsilon$ . This user-selected threshold describes the radius of the neighborhood in phase space, within which a state is said to recur.

To see how the RP looks when applied to continuous time series, Fig. 1.2 shows periodic (a) and chaotic (b) signals, with RPs below (c, d). Crosses in the RPs arise in the main diagonal where the signal has a peak or valley; here,  $x(i + \Delta i) \sim x(i - \Delta i)$ . Further crosses also appear above and below the main diagonal. In the sinusoidal plot (c), they appear regularly because, for period  $T$ ,  $x(i + T) \sim x(i)$ ; in the chaotic signal (d) their appearance is more erratic.

Notice also in the chaotic signal, Fig 1.2 (d), that the positive values past  $i = 300$  create a dark region in the upper right corner similar to the RP of the sinusoidal signal (Fig 1.2 a). Recurrences in the lower right and upper left corners of Fig. 1.2 (d) result from the positive spike at the start of the series echoing these positive values later on.

## 1.2 Tolerance Threshold and Embedding Dimension

Critical to the generation of an RP is the tolerance parameter,  $\varepsilon$ . It is this value that allows the RP to interpret whether recurrence has occurred anywhere in a time series or not. Recurrent trajectories in phase space can be described by an  $\varepsilon$ -tube, enclosing similar paths as shown in Fig. 1.3. For larger values of  $\varepsilon$ , these tubes will be wider and more recurrences may be counted, though even for small  $\varepsilon$ , the restricted range of possible states occupied by the system ensures that recurrence in a chaotic system is inevitable.

One must carefully consider what value for  $\varepsilon$  to choose when analyzing a time series. If the tolerance threshold is too large, then almost every point will be interpreted as neighboring almost every other point without reflecting genuine recurrence. This leads to a plot which is very dark, or even flat black. Figure 1.4 shows the RPs for the same sinusoidal and chaotic signals with much higher values for  $\varepsilon$ . The lines in (a) thicken, obscuring any possible fluctuations in the signal, while areas of spurious recurrence appear in (b) near  $i = 200, j = 400$ . Conversely, if the tolerance threshold is chosen too strictly, the RP may miss genuine periods of recurrence, as shown in Fig 1.4 (c, d).

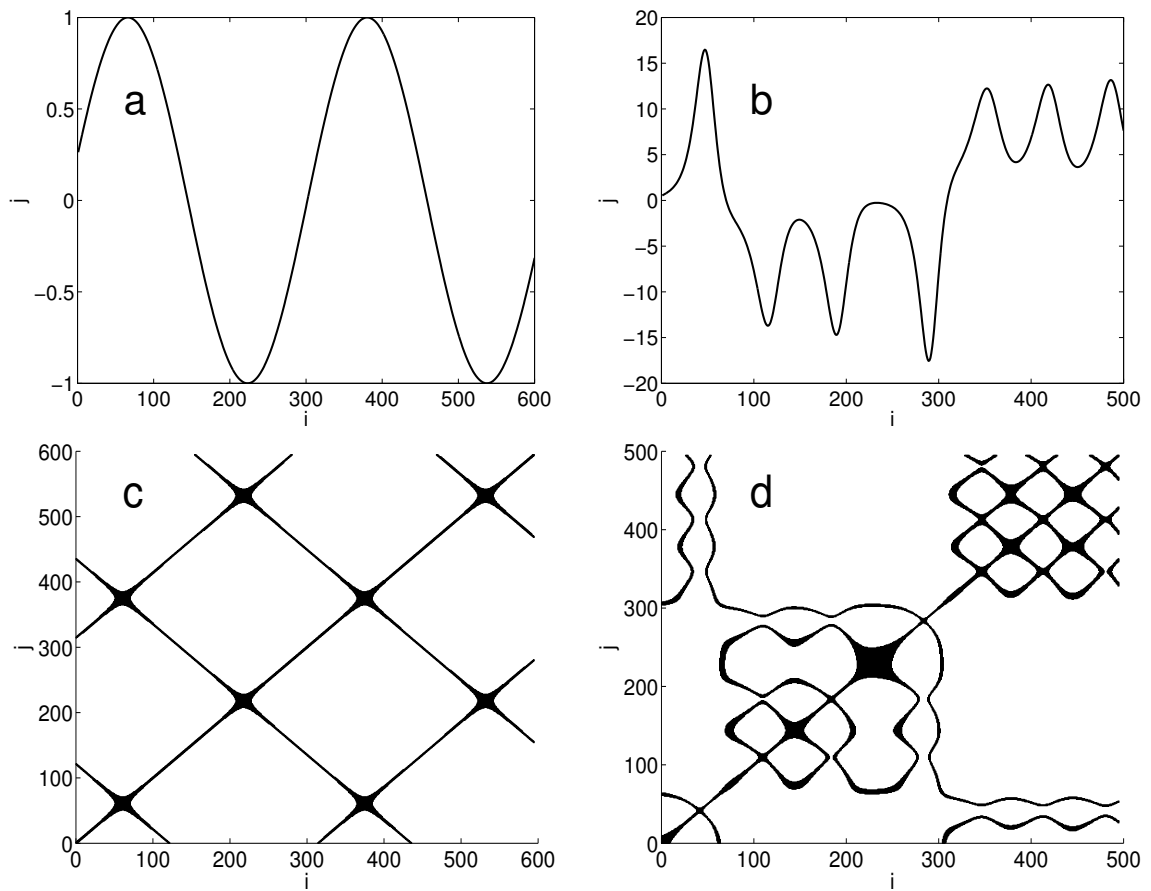


Figure 1.2: Position ( $x(i)$ ) vs. time ( $i$ ) for (a) Sinusoid ( $x(i) = \sin(0.02i + 0.3)$ ) and (b) Lorenz attractor,  $\rho = 28$ ,  $\sigma = 10$ , and  $\beta = 8/3$ . Associated RPs are shown below, for (c) sinusoid, with tolerance  $\varepsilon = 0.02$  and (d) Lorenz attractor with tolerance  $\varepsilon = 1$ .

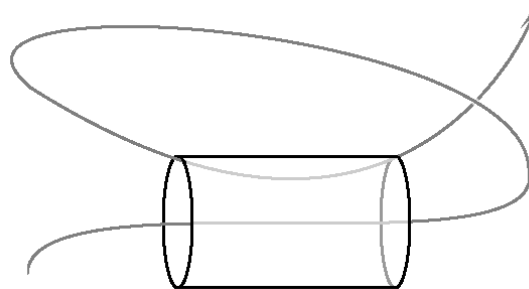


Figure 1.3: Graphical representation of an  $\varepsilon$ -tube in phase space with radius  $\varepsilon$ , enclosing the trajectory during an initial and later pass through the same region.



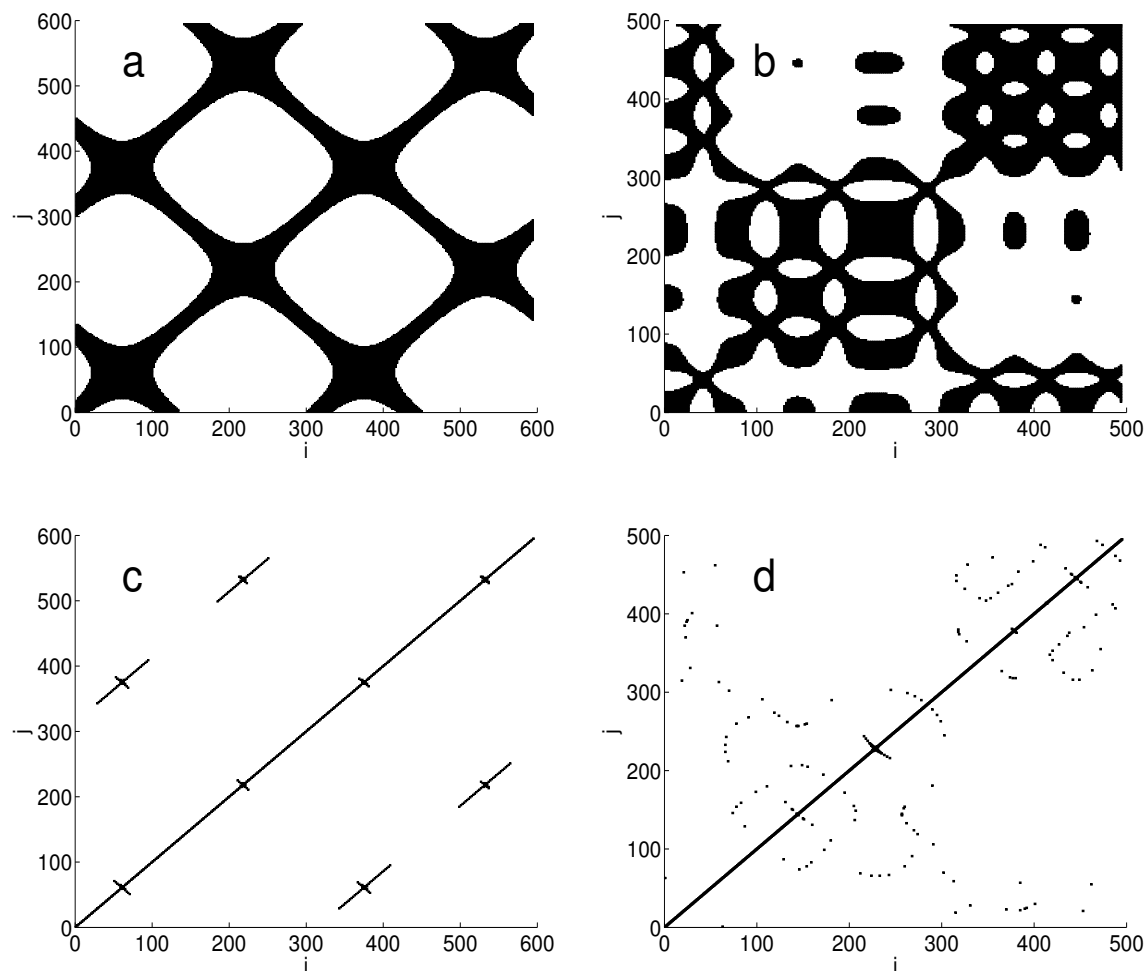


Figure 1.4: Recurrence plots for sinusoidal signals from Fig. 1.1, generated with tolerance  $\varepsilon = 2$  (a) and tolerance  $\varepsilon = 0.002$  (c). To the right, RPs are shown for the Lorenz attractor, with tolerance  $\varepsilon = 6$  (b) and tolerance  $\varepsilon = 0.01$  (d).

Although there are no hard-and-fast rules for the choice of  $\varepsilon$ , it is generally selected so that more than 1% (Mindlin & Gilmore, 1992), but no more than 10% (Koebe & Mayer-Kress, 1992; Zbilut & Webber, 1992) of the points in the RP are darkened. Beyond this, the choice of  $\varepsilon$  depends on the kind of signal under analysis. In a data series formed from a signal with observational noise of standard deviation  $\sigma$ , it is desirable to pick out the signal underlying the noise, so  $\varepsilon$  should be chosen as significantly larger than the standard deviation of the noise; since 99% of the points under a Gaussian distribution fall within a window 5-standard deviations wide,  $\varepsilon > 5\sigma$  is a common choice. (Marwan, Romano, Thiel,

& Kurths, 2007). This does not work well for highly noisy data, but for such cases trying to derive more information than the average value or mean frequency may not be possible.

A further choice in generating a recurrence plot is represented by the *embedding dimension*. If a system progresses from  $x(i)$  to a new value  $x(i + 1)$ , and at a later time  $j$ , has the same value as at  $x(i)$ , but progresses to a state  $x(j + 1)$  having the same value seen at  $x(i - 1)$ , the RP will show downward-sloping diagonal lines. The rising and falling of the signals in Fig. 1.1 produce these downward-sloped diagonals in the RPs. Although these runs can easily arise in a random signal, such behavior is not possible in a deterministic system where every state originates directly from the preceding state. Here, such downward-sloping diagonal lines indicate that the raw signal must come from a system of higher dimensionality than 1.

The determinism of the underlying system can be recovered through a higher dimensional embedding of the scalar signal, by selecting a characteristic time constant,  $\tau$ , and then generating a recurrence plot in  $n$ -dimensions by

$$R(i, j) = \Theta \left( \varepsilon - \sqrt{\sum_{k=0}^{n-1} [x(i + k\tau) - x(j + k\tau)]^2} \right). \quad (2)$$

Here  $k = 0$  refers to the first coordinate,  $k = 1$  to the second coordinate, and so on up to  $k = n - 1$  corresponding to the  $n^{\text{th}}$  coordinate. So for example, the RP for a signal with two-dimensional embedding would examine the Euclidian distance between the state at  $x(i), x(i + \tau)$  and  $x(j), x(j + \tau)$ , returning a 1 if the distance were no greater than  $\varepsilon$ , and a 0 otherwise.

Figure 1.5 shows RPs for the signals from Fig. 1.1 with a two-dimensional embedding which eliminates the downward-sloped diagonals from both the sinusoidal (Fig. 1.5 a) and chaotic (Fig. 1.5 b) signals, to resolve the underlying determinism in each system with upward-sloping diagonals.

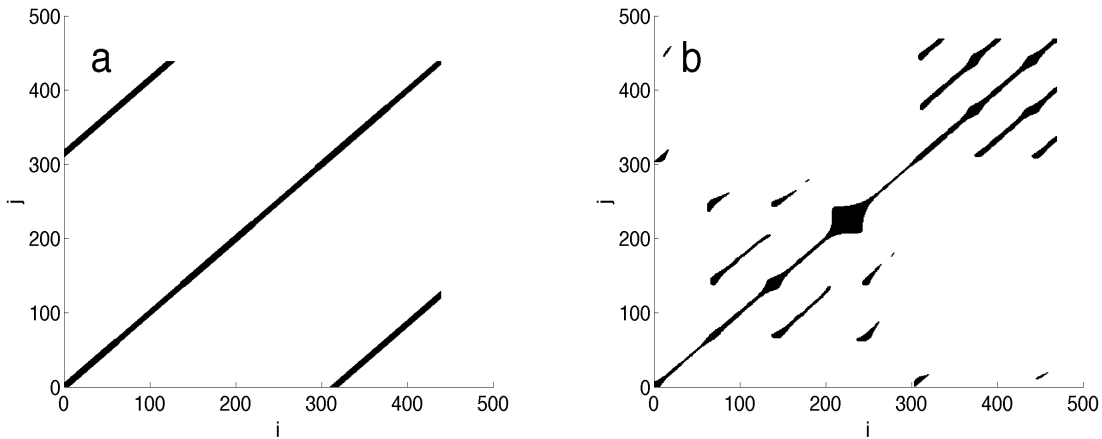


Figure 1.5: Recurrence plots for signals from Fig. 1.1 in a two dimensional embedding. a) Sinusoid, with tolerance  $\varepsilon = 0.01$  and time constant  $\tau = 78$ . b) Lorenz attractor, with tolerance  $\varepsilon = 1.5$  and time constant  $\tau = 13$ .

### 1.3 Features of Recurrence Plots

The most basic feature of an RP is the upward-sloping diagonal line parallel to the LOI. Any time a diagonal line is observed parallel to the LOI, this indicates recurrence, not only of an isolated point but of a *path* through an  $\varepsilon$ -tube in phase space. The longer the diagonal line, the longer the two paths are similar. Diagonal lines may be repeatedly spaced parallel to one another, indicating periodic runs along the same path, or they may be staggered, indicating complex returns to many different paths. Like the different words in an infant’s babbling, or the different routes a fox takes in the evenings as it leaves its lair, every repeated sequence will appear in the RP as a diagonal line. The more diagonals in the RP, and the longer these diagonals are, the more regularity is contained in the signal.

Compare the chaotic signals in Fig. 1.6; (a) is more regular, and shows longer diagonals in its RP than (b). The LOI itself, however, never contains any information; all states are equal to themselves at time  $i = j$ . Long diagonal lines can occur directly below and above the LOI, particularly when dealing with high-resolution data. Since  $\varepsilon$  is nonzero, these false “points of recurrence” may be darkened in the RP, but actually represent the system slowly changing from one state to another within the same  $\varepsilon$  tube. So it is customary to exclude not only the LOI, but this entire corridor of points falling a distance  $W$  or less from the LOI, known as the *Theiler window* (Marwan *et. al*, 2007). Theiler recommends choosing for

exclusion all points within  $W = \tau$  steps from the LOI, where  $\tau$  is the autocorrelation time of the raw data series. (Theiler, 1986) Diagonal lines within this region may be excluded from analysis.

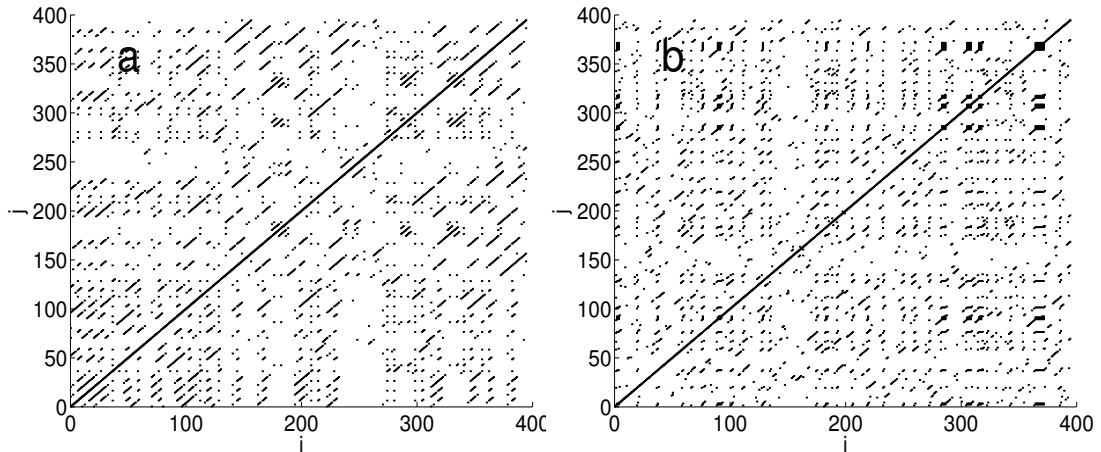


Figure 1.6: Recurrence plots for the logistic map,  $x(i + 1) = r x(i) [1 - x(i)]$ , with (a)  $r = 3.65$ , tolerance  $\varepsilon = 0.003$ , and (b)  $r = 4.0$ , tolerance  $\varepsilon = 0.006$ .

Recurrence plots contain other features besides upward-sloping diagonal lines, however. Vertical and horizontal lines may appear where a state changes only very slowly, so that each successive state is within  $\varepsilon$  of the preceding states. This occurs when a time series becomes “trapped” in a state for some time before moving on. Downward sloping diagonals occur where the time series shows the same pattern running forwards and backwards, as in the sinusoidal signals seen previously (Fig. 1.1), suggesting the need for a higher embedding dimension. White bands appear when there is an anomalous point in the RP. In a heavily recurrent RP, such “odd men out” indicate either that some states are exceptional, or else that some kind of transition has occurred. Squarish mats of diagonals appear when a system enters an area of temporary periodicity; this is often seen in chaotic systems such as the Rössler system (Rössler, 1976) or Lorenz system (Lorenz, 1963). Lastly, the RP may show a fading to white in the upper left and upper right corners; this indicates a long term trend for the data to consistently rise, or consistently fall, over time.

As a glance at the figures above will show, the recurrence plot is often worth having simply for what it reveals to us visually. However, it is best if we have some means of *quantifying* point distributions in an RP. Therefore, after generating the recurrence plot for a data series, the next step is to analyze its behavior.



## Chapter 2

### Methods

*Recurrence quantification analysis* (RQA) is the process of investigating the behavior of a data series with length  $N$  by numerically analyzing the features of its recurrence plot. Physicists and mathematicians have developed several common measures for analyzing these features (Marwan *et. al.*, 2007).

#### 2.1 Classic Recurrence Measures

These measures, as defined here, consider the entire recurrence plot of area  $N^2$ , even though the RP is symmetric above and below the main diagonal (that is, the “line of interest,” or LOI). But in practice, RPs may be generated using different tolerance thresholds ( $\varepsilon$ ) above and below the LOI (see for example Eckmann *et al.*, 1987), or the analysis can focus only on the upper left half of the RP for better computational speed.

The *recurrence rate* (RR) represents the recurrence point density,

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}. \quad (3)$$

$RR$  is calculated by summing up all black points and dividing by the plot area, to give a proportion ranging from 0 to 1. In practice, the summation usually excludes the LOI and all points within the Theiler Window.  $RR$  gives the chance that any point chosen at random from a recurrence plot will be black.

The *diagonal histogram*,

$$P(l) = \sum_{i,j=1}^N (1 - R_{i-1,j-1})(1 - R_{i+l,j+l}) \prod_{k=0}^{l-1} R_{i+k,j+k}, \quad (4)$$

returns the number of diagonal lines of precisely a given length,  $l$ . Here, diagonal lines are defined as a run of upward-sloping dark points bracketed by white points on either end.

The *vertical histogram* is defined analogously as

$$P(v) = \sum_{i,j=1}^N (1 - R_{i,j-1})(1 - R_{i,j+v}) \prod_{k=0}^{v-1} R_{i,j+k}, \quad (5)$$

representing the number of vertical lines of a given length,  $v$ .

The *average diagonal line length* is defined as

$$L = \frac{\sum_{l=L_{min}}^N lP(l)}{\sum_{l=L_{min}}^N P(l)}. \quad (6)$$

$L$  sums up every point in each diagonal line of length  $L_{min}$  or greater, and then divides by the total number of such diagonal lines. Although  $L$  may be calculated for any  $L_{min}$ , choosing  $L_{min} = 1$  gives the overall average diagonal line length, and can thus be interpreted as the mean prediction time (Marwan *et al.*, 2007).

*Trapping time* (TT) represents the average vertical line length,

$$TT = \frac{\sum_{v=V_{min}}^N vP(v)}{\sum_{v=V_{min}}^N P(v)}. \quad (7)$$

It is calculated analogously to  $L$ , by summing every point in every vertical line, and then averaging over all vertical lines. This measure indicates the amount of time that the system remains in the same state, or the tendency for adjacent points in the original time series to take the same value.

The *determinism* (DET) reflects how often a sequence recurs in a dataset, and for how long. (Marwan, 2011). It is defined by

$$DET = \frac{\sum_{l=L_{min}}^N lP(l)}{\sum_{i,j=1}^N R_{i,j}}, \quad (8)$$

which gives the proportion of recurrence points forming diagonal structures of a given length  $L_{min}$  or longer. It sums over all points contained in these lines and then divides by the total number of recurrence points. Despite its name, determinism does not provide proof of physical determinism in the sense of each point leading to one unique subsequent point; data signals may be generated with nondeterministic features and still return high DET values.

The *laminarity* is a measure of the overall tendency of the system to remain in the same state,

$$LAM = \frac{\sum_{v=V_{min}}^N vP(v)}{\sum_{i,j=1}^N R_{i,j}}. \quad (9)$$

Laminarity gives the proportion of recurrence points forming *vertical* structures of a given length  $V_{min}$  or longer.

The *Shannon entropy* (ENTR) is a way to quantify the probability distribution of diagonal lines. It is calculated as

$$ENTR = - \sum_{l=L_{min}}^N \frac{P(l)}{N_l} \ln \left( \frac{P(l)}{N_l} \right), \quad (10)$$

which evaluates the frequency distribution of the diagonal line lengths. It reflects the complexity of the RP with respect to diagonal lines. For uncorrelated noise the value of ENTR is rather small, whereas a regular signal such as a sine wave returns a much higher ENTR (Marwan, 2007). In contrast, Shannon entropy as originally introduced (Shannon, 1948) describes the information encoded in a data signal (Ihara, 1993) which is highest when the values in the signal are evenly distributed. But here, a high value for ENTR indicates regularity in a signal, returning values near zero for random signals.

The *tau recurrence rate* ( $RR_\tau$ ), describes the dynamics on a fixed timescale  $\tau$ , with  $j = i + \tau$ ,

$$RR_\tau = \frac{1}{N - \tau} \sum_{i=1}^{N-\tau} R_{i,i+\tau}. \quad (11)$$

The tau recurrence rate determines the recurrence rate along each diagonal separated from the LOI by a distance  $\tau$ . So  $\tau = 0$  corresponds to the main diagonal itself, where  $RR_0 = 1$ .  $\tau > 0$  corresponds to a positive time delay, for diagonals a distance  $\tau$  above the main diagonal, and  $\tau < 0$  to a negative time delay, below the main diagonal.  $RR_\tau$  is a type of autocorrelation function, giving information on the time scales separating recurrent paths. It can be interpreted as the probability that a state recurs after  $\tau$  time steps (Marwan *et. al*, 2007).



*Trend* quantifies long-term changes by the degree to which the recurrence density drops off with distance from the main diagonal. Trend discriminates between broadly homogeneous RPs and those with an overall “fading out” away from the LOI; such RPs result from nonstationary time series with data values that, on average, increase or decrease with time. Trend is defined by

$$TREND = \frac{\sum_{\tau=1}^{\tilde{N}} (\tau - \tilde{N}/2)(RR_{\tau} - \langle RR_{\tau} \rangle)}{\sum_{\tau=1}^{\tilde{N}} (\tau - \tilde{N}/2)^2}, \quad (12)$$

which sums over all  $RR_{\tau}$  for changing  $\tau$ , multiplying by the distance each diagonal is from the LOI, and then averaging by dividing by the square of these distances. Note that  $\tilde{N} < N$  is used instead of just  $N$ , in order to exclude the edges of the recurrence plot where few recurrences will be found. The choice of  $\tilde{N}$  is up to the user;  $\tilde{N} = N - 10$  is claimed to be sufficient for noise, but in a more regular signal, it is suggested to subtract as much as 10 times the order of magnitude of the autocorrelation time (Marwan *et. al*, 2007). Trend is closely related to the Pearson product-moment correlation coefficient, (Stigler, 1989) except that the denominator only references the variance in  $\tau$  rather than normalizing with the standard deviation in each variable. Trend commonly returns small values on the order of  $-10^{-6}$ , since  $\tau$  ranges from 1 to  $\tilde{N}$  while  $RR_{\tau}$  ranges only from 0 to 1.

Figure 2.1 shows TREND for three typical signals. In purely random data (a) the TREND is  $-1.41 \times 10^{-8}$ , increasing in magnitude to  $-3.79 \times 10^{-6}$  for data with a subtle upward tendency (b). Even in data with a strong linear trend (c), the TREND measure is still only  $7 \times 10^{-6}$ .

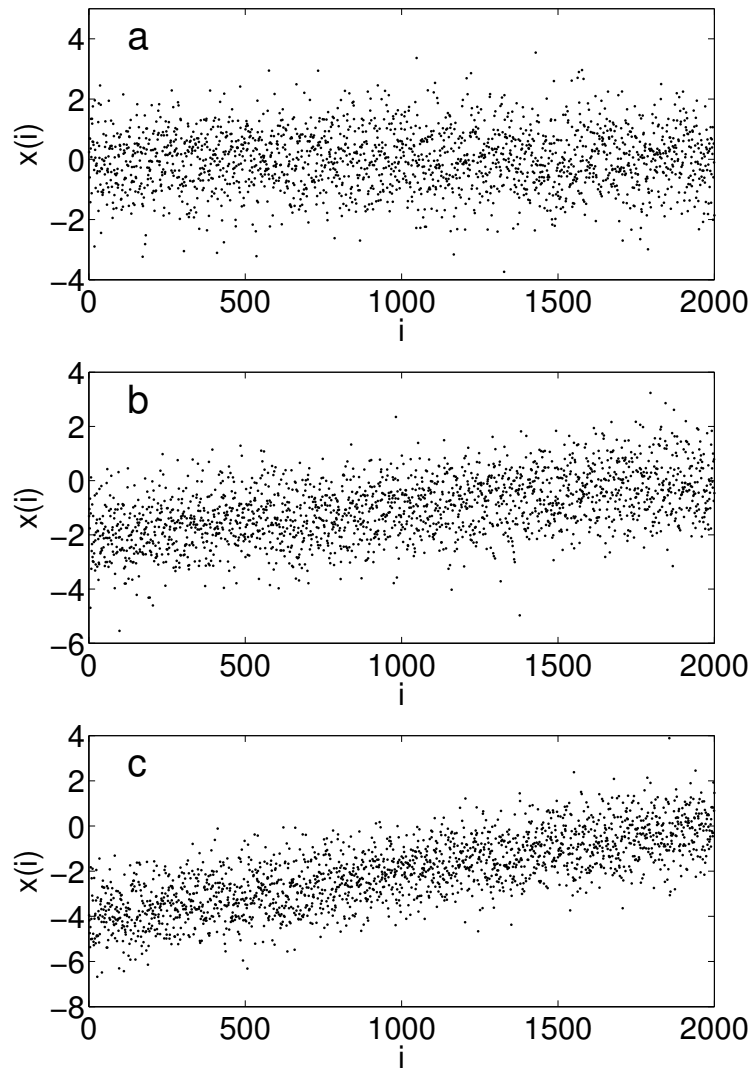


Figure 2.1:  $x(i)$  and TREND values for three Gaussian noise signals with mean  $\mu$  and standard deviation  $\sigma$ . a)  $\mu = 0.000, \sigma = 1$ ; TREND =  $-1.41 \times 10^{-8}$ . b)  $\mu = 0.001i, \sigma = 1$ ; TREND =  $-3.79 \times 10^{-6}$ . c)  $\mu = 0.002i, \sigma = 1$ ; TREND =  $-6.93 \times 10^{-6}$ .

## 2.2 Dependence of Classic RQA Measures on Changing RR

That the choice of tolerance ( $\varepsilon$ ) remains open to the analyst sometimes creates difficulties interpreting RQA results, because the above measures are strongly influenced by the basic recurrence rate (RR). A higher  $\varepsilon$  will give rise to a more densely populated RP, which will

inevitably contain more and longer diagonal and vertical lines, thereby increasing the values for  $P(l)$  and  $P(v)$ , and all measures based on them (DET, LAM, ENTR, L, TT).

Using a random signal, the dependence of RQA measures on  $\varepsilon$  and RR were investigated. As Fig. 2.2 shows, RR has a virtually linear relationship with  $\varepsilon$ . Figure 2.3 shows the dependence of other RQA outcomes on changing RR; all measures increase with RR. Determinism (Fig. 2.3 a) shows some fluctuations at low RR values; as  $\varepsilon$  increases, it creates more points in the RP, but in an RP with few existing recurrences, these new points may be disconnected from others. This fluctuation is minimal past  $RR = 0.02$ . Laminarity and Shannon entropy (Fig. 2.3 b and c) show a more direct dependence on RR, rising by over an order of magnitude. Note that DET (a) and LAM (b) return lower values for higher  $L_{min}$  and  $V_{min}$ , since shorter structures will be discounted as  $L_{min}$  and  $V_{min}$  are increased.

The maximum lengths of diagonal and vertical lines ( $L_{max}$  and  $V_{max}$ ; Fig. 2.3 d) likewise increase with RR, by a factor of approximately 3. The reciprocal of the longest diagonal line length is claimed to relate directly to the Lyapunov exponent in a dynamical series (Trulla *et al.*, 1996; Webber and Zbilut, 1994). However, the Lyapunov exponent is a dynamical invariant, but  $L_{max}$  is not, and their relationship has been described as “not as simple as it was mostly stated in the literature” (Marwan *et al.*, 2007).

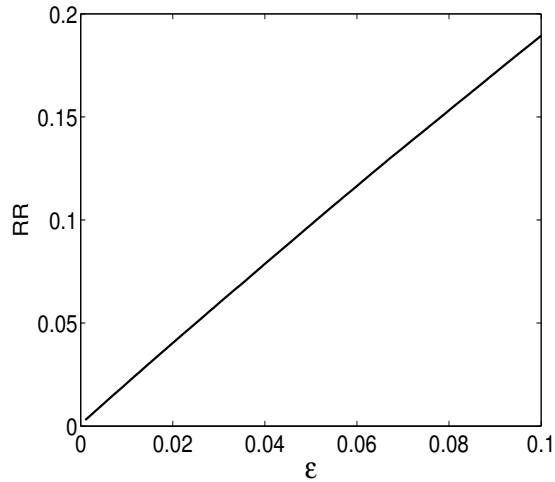


Figure 2.2: Recurrence rate (RR) vs. tolerance threshold ( $\varepsilon$ ) for a time series of uniform random data,  $x \in [0, 1]$ ,  $N = 2000$ .  $\varepsilon$  varies from 0.001 to 0.1 in steps of 0.001.

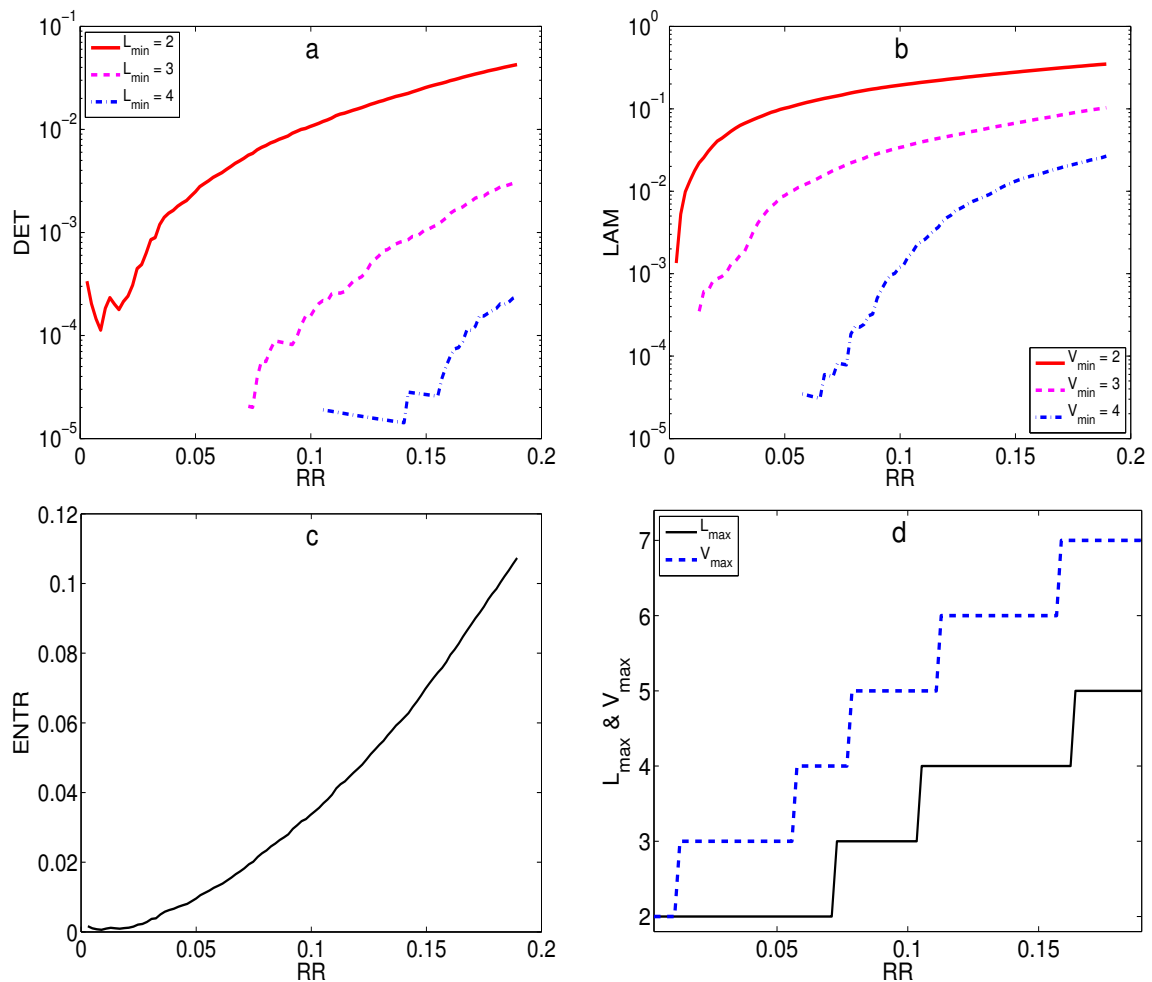


Figure 2.3: RQA measures vs. recurrence rate (RR) for the random time series from Fig. 2.2, showing determinism (DET) for three minimum diagonal lengths ( $L_{min}$ , a), Laminarity (LAM) for three minimum vertical lengths ( $V_{min}$ , b), Shannon entropy (ENTR, c), and Longest Diagonal ( $L_{max}$ , black) and Vertical ( $V_{max}$ , blue) Lines (d).

This behavior is not restricted to random signals; a similar relationship between  $\varepsilon$ , RR, and other RQA outcomes can be seen in other series. Figure 2.4 shows RR vs.  $\varepsilon$  in a nonlinear signal. Although the relationship shows slightly more curvature, RR still rises monotonically with rising  $\varepsilon$ .

Figure 2.5 gives a similar picture for other measures. In the chaotic signal, DET (Fig 2.5 a), LAM (Fig 2.5 b), and ENTR (Fig 2.5 c) each rise almost linearly with RR. These higher values are expected given the higher regularity in the RP. We also observe higher

values for  $L_{max}$  than in the random series. But both  $L_{max}$  and  $V_{max}$  increase monotonically with RR, just as observed in the random signal.

In both random and chaotic data, higher  $\varepsilon$  gives rise to a higher RR, in turn increasing other RQA measures. Since the finite-time recurrence rate can fluctuate within a single signal, or within different signals generated from different conditions (*e.g.* stimulated and unstimulated neurons), interpretation of RQA values can be difficult.

One solution is to carry out analyses using a fixed RR. (This is not always an option, however; with signals taking only  $m$  discrete values, RR can never fall below  $1/m$ .) Another solution is to develop RQA measures which are more stable with respect to RR. The following section details three new RQA measures created in an attempt to minimize the effects of variability in RP density.

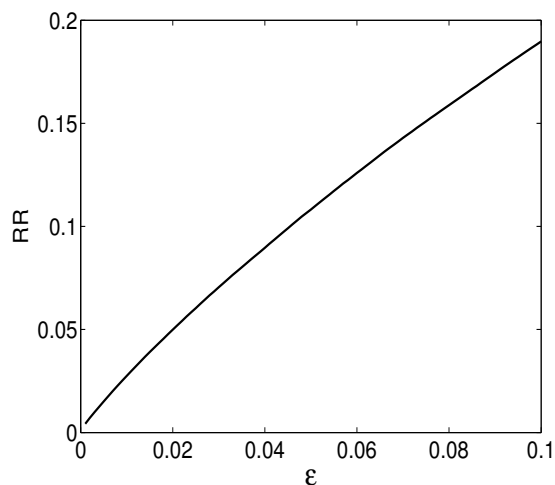


Figure 2.4: Recurrence rate (RR) vs. tolerance threshold ( $\varepsilon$ ) for a chaotic signal (logistic map,  $x_{n+1} = r x_n(1 - x_n)$ , with  $r = 4$ ).  $\varepsilon$  varies in steps of 0.001.

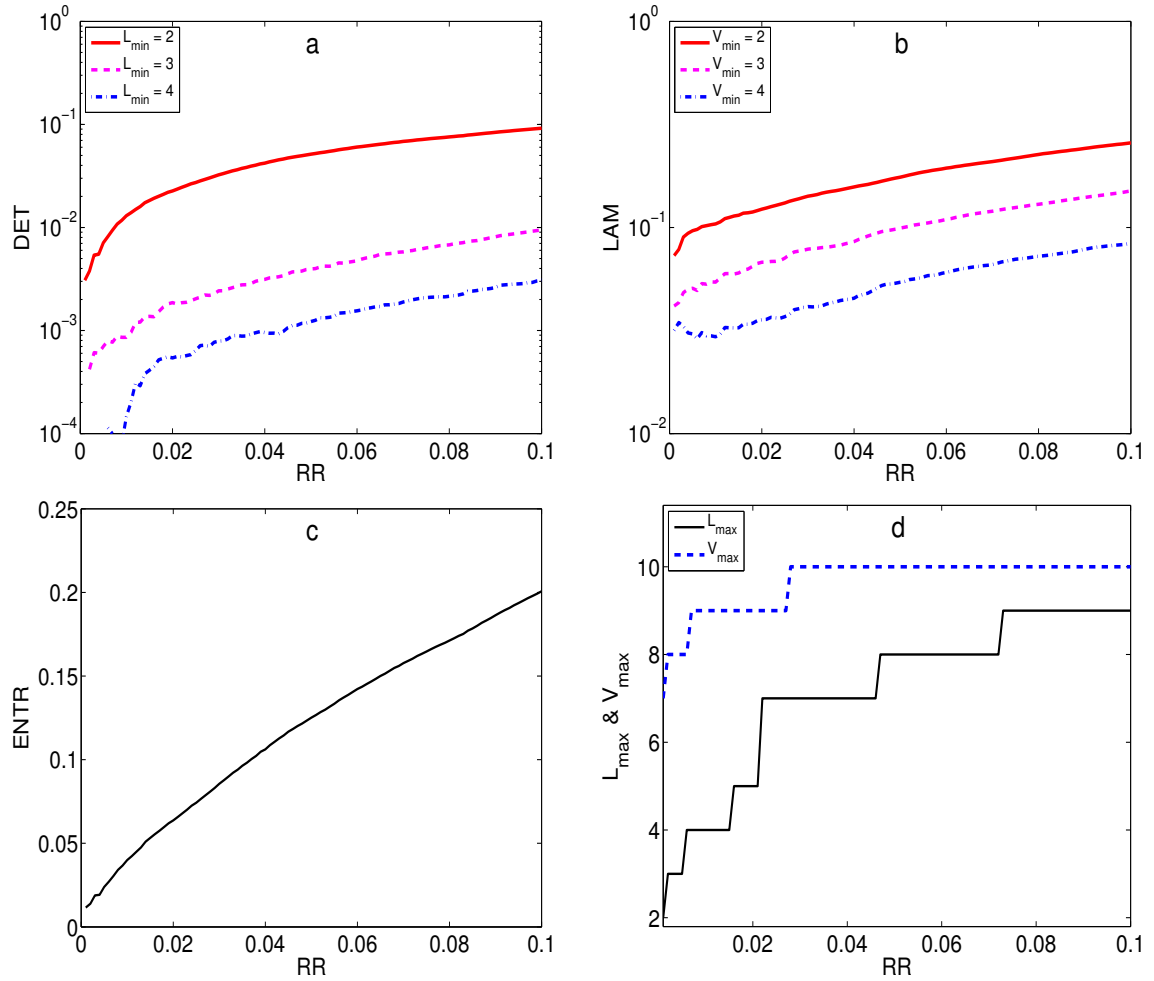


Figure 2.5: RQA measures vs. recurrence rate (RR) for the chaotic time series from Fig. 2.4, showing determinism (DET) for three minimum diagonal lengths ( $L_{min}$ , a), Laminarity (LAM) for three minimum vertical lengths ( $V_{min}$ , b), Shannon entropy (ENTR, c), and longest diagonal ( $L_{max}$ , black) and vertical ( $V_{max}$ , blue) lines (d).

### 2.3 New Measures For Recurrence Quantification Analysis

We now define new measures, similar to the classical tools used in RQA, but more robust with regards to changing RR. The first of these is *diagonality*,

$$D = \frac{\sum_{i,j=2}^N R_{i,j} R_{i-1,j-1}}{\sum_{i,j=2}^N R_{i,j}} - RR, \quad (13)$$

which represents the overall degree to which points in an RP are arranged in upward-sloping diagonal lines. The first term in D describes the number of dark points in the RP with dark lower-left neighbors, normalized by the total number of black points. The recurrence rate is subtracted away from this, because RR is the convergent value of the first term in  $D$  for a large RP of uniform random noise. Simulations verify that D does indeed go to zero with increasing  $N$  for recurrence plots of uniform random noise. Inspection of the fraction in the first term in D reveals why; the chance for the lower-left neighbor of any given point to be black in a random RP is RR.

D returns a positive result *only* if there are more points present in upward-sloping diagonals in an RP than in a reference signal of infinitely long uniform random noise. The result is a measure of the overall tendency of a system to follow recurrent paths, which is more robust to fluctuations in the recurrence rate than classical measures.

*Verticality* is defined as

$$V = \frac{\sum_{i=1}^N \sum_{j=2}^N R_{i,j} R_{i,j-1}}{\sum_{i=1}^N \sum_{j=2}^N R_{i,j}} - RR. \quad (14)$$

V represents the number of dark points having dark neighbors directly underneath. Like LAM, verticality measures the overall tendency of the system to remain in the same state from time  $i$  to  $i + 1$ . V (and also D) can return the same value if an RP had many short lines or a few long lines.

*Period- $n$*  measures the periodicity in a time series. It is defined as

$$P_n = \frac{\sum_{i=1}^N \sum_{j=n+1}^N R_{i,j} R_{i,j-n}}{\sum_{i=1}^N \sum_{j=n+1}^N R_{i,j}} - RR. \quad (15)$$

Period- $n$  returns the black points with a black neighbor exactly  $n$  spaces below them, normalized by the total number of black points, with  $RR$  subtracted away.  $P_n$  yields information similar to  $RR_\tau$  for a fixed  $\tau$ ; for instance, a recurrence plot with a period 5 orbit should be expected to return a high value for  $P_5$ , but lower values for  $P_4$  or  $P_6$ . If the periodic behavior extends over a long enough time, high values will be found for many  $P_{kn}$  where  $k = 1, 2, 3, \dots$  and  $n$  is the period length. Note that  $P_1$  is identical to  $V$ .

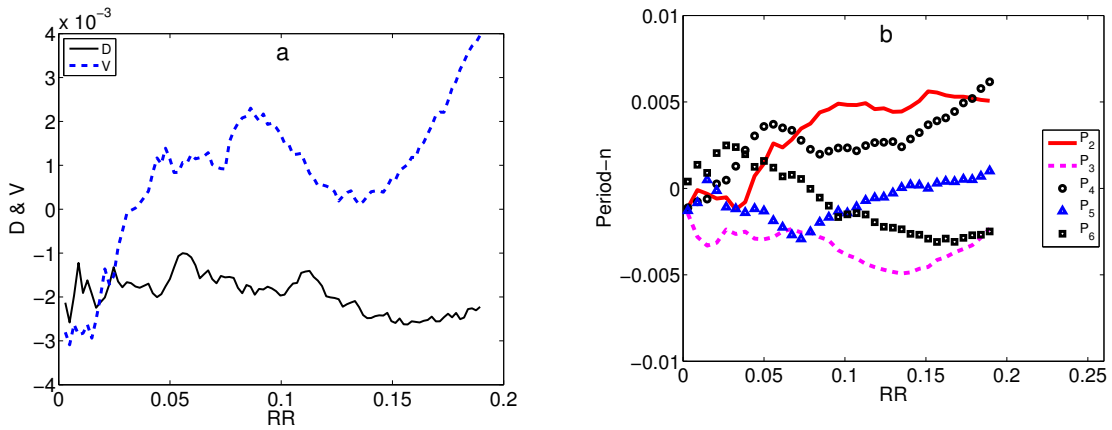


Figure 2.6: RQA measures vs. recurrence rate (RR) for the random time series from Fig. 2.2, showing (a) diagonality (D) and verticality (V), and (b) period- $n$  ( $P_n$ ).

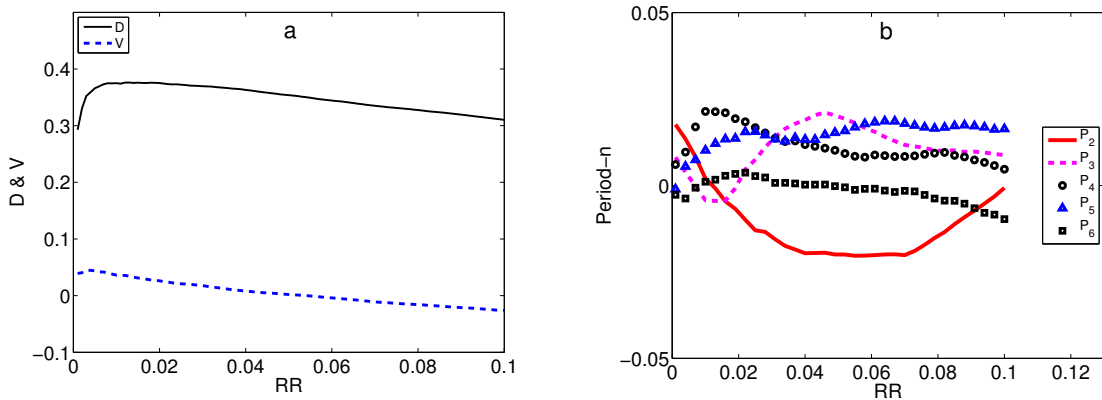


Figure 2.7: RQA measures vs. recurrence rate (RR) for the chaotic time series from Fig. 2.4, showing (a) diagonality (D) and verticality (V), and (b) period- $n$  ( $P_n$ ).



To investigate the properties of these new RQA measures, each was applied to the random signal from section 2.2. Figure 2.6 shows the behavior of these measures with changing RR. All measures are confined to  $\pm 0.006$  of the average value regardless of the recurrence rate in the random signal.

A greater dependence on RR was found in the chaotic signal. Although the relationship between RR and the new RQA measures is not perfectly attenuated, the measures do not increase monotonically with RR. Instead, Fig. 3.4 shows they are confined to fairly narrow envelopes, particularly in the typical range where  $RR \in (0.01, 0.10)$ .

V and  $P_n$  vary no more than  $\pm .025$ . For diagonality, where non-zero values are expected due to the organization in the signal, the total variation is no more than  $\pm .05$ . The diagonality curve in Fig 2.7 (a) is promising in being virtually flat between  $RR = 0.01$  and  $0.10$ , which is the typically suggested range for RQA. Thus, while these new measures do not eliminate the dependency on RR, they do allow comparisons to be made even across signals with slightly different recurrence rates.

## 2.4 Convergence Behavior of RQA Measures

The length of the available time series is an important consideration throughout time series analysis. Sometimes it is not possible to obtain long signals, *e.g.* when checking the short-term effects of a sudden stimulus on a neural network, and in such cases one must work with the data as it exists.

As the following figures show, most RQA measures do not converge for short lengths. RQA methods were repeatedly applied to a signal of uniform random noise with increasing data length, starting with a length  $N = 1000$  up to 20,000 points.

For random data, the recurrence rate, determinism, diagonality, verticality, trend, and Shannon entropy are shown in Fig. 2.8. While there is some fluctuation even as far out as  $N = 15,000$ , these measures generally fall near their final values by  $N = 2000$ . Even as early as  $N = 1000$ , all measures were confined to within 0.01 of their final value. However, verticality shows more fluctuations than the other measures, and does not clearly converge by  $N = 15,000$ .

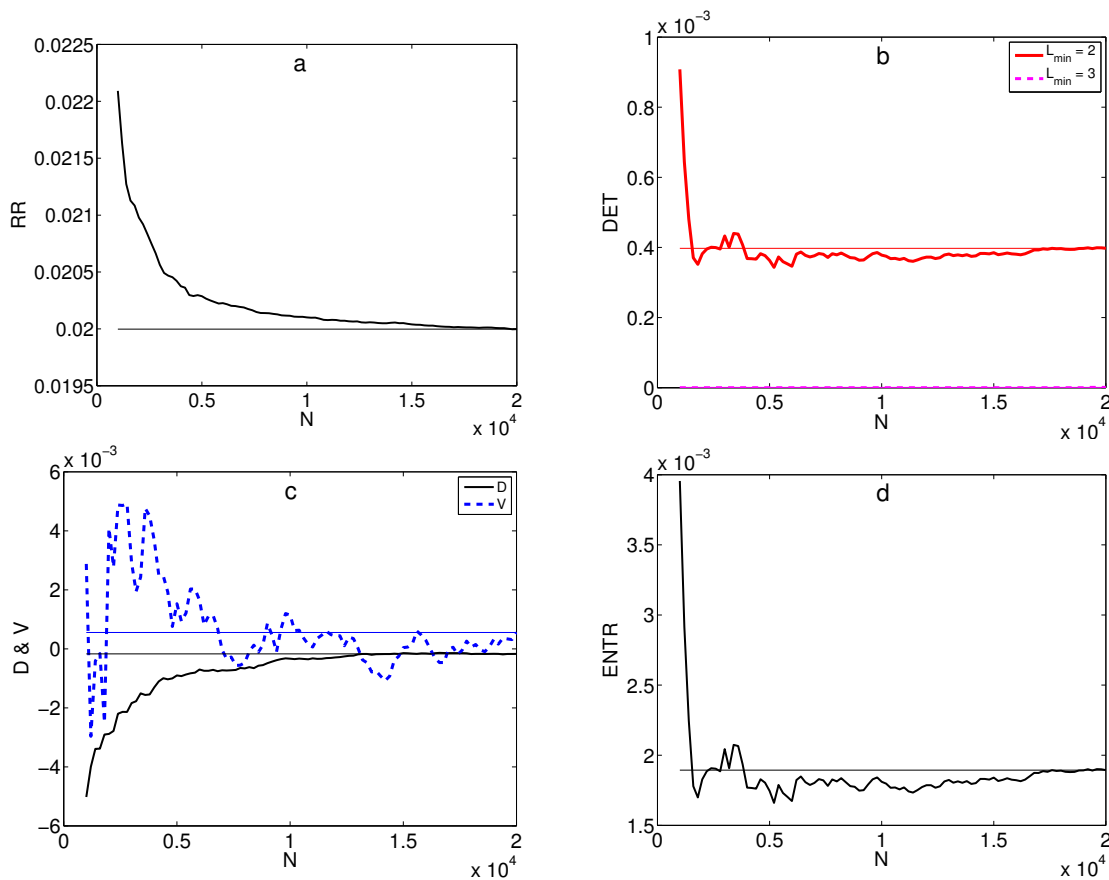


Figure 2.8: Convergence behavior for selected RQA measures applied to a random signal (uniform over the interval  $[0,1]$ , with  $\varepsilon = 0.01$ ), showing recurrence rate (RR, a), determinism (DET, b), diagonality and verticality (D and V, c), and Shannon entropy (ENTR, d) all vs. time (N). Horizontal lines show final values to guide the eyes.

RQA will typically be applied to signals containing more structure than found in random noise. Figure 2.9 shows RQA measures applied to a chaotic signal. Convergence is less clear for this signal, with V fluctuating as much as 0.03 past  $N = 2000$ . The other measures were more stable, with variation below 0.01 by  $N = 1000$ , similar to that seen in the random signal. Therefore,  $N = 1000$  appears to be generally sufficient for accuracy of two decimal places, though verticality may still fluctuate by 0.03 or higher.

For shorter time series, recurrence quantification analysis can still be used, particularly if the circumstances surrounding a non-stationary phenomenon are well understood (*e.g.* a heartbeat after stimulus). In such cases, RQA measures only represent the behavior of a specific response *over a specific time period*, which may change at other times. This is most

relevant when applying RQA to short time series, where  $N$  may not be sufficient at the chosen sampling rate to fully represent the dynamics of the system; inferences regarding the overall characteristics of a less well-known signal typically cannot be drawn without sampling for a longer time. For the signals investigated here,  $N = 1000$  was usually sufficient, although extremely long data series with  $N > 20,000$  would remain beneficial for measuring long-run behavior to high precision.

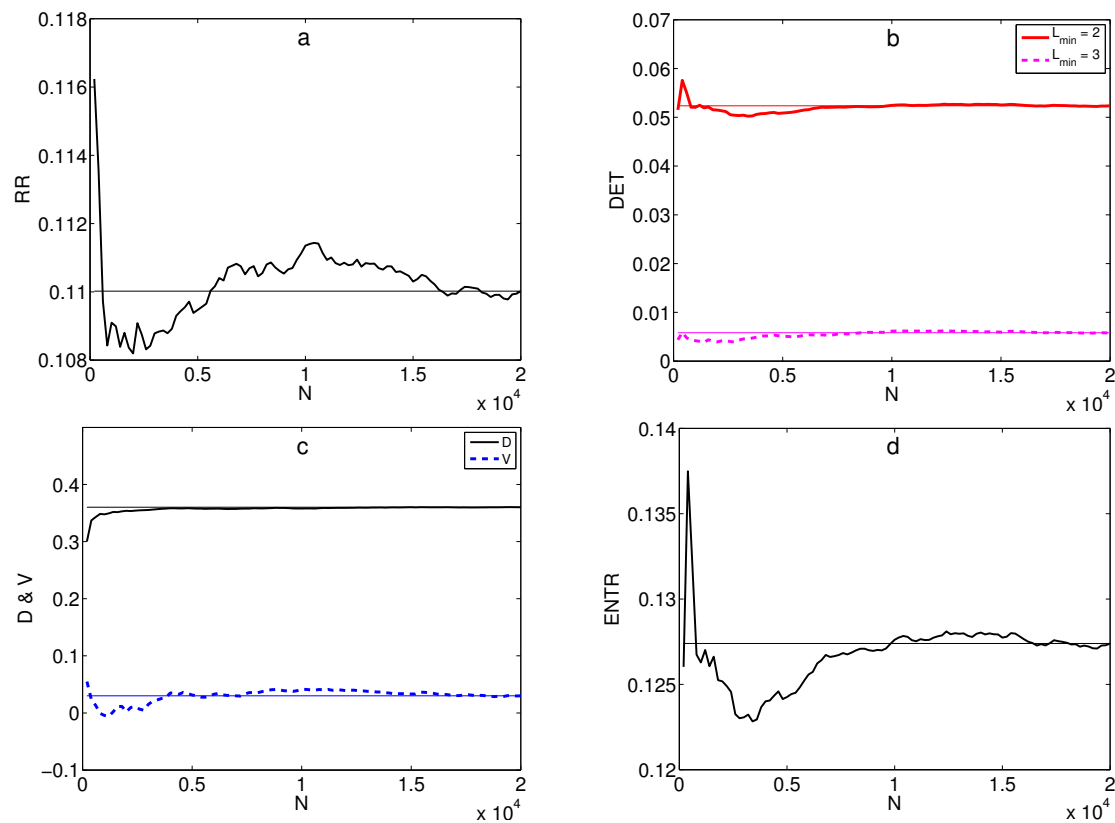


Figure 2.9: Convergence behavior for selected RQA measures applied to a chaotic signal (Logistic map,  $r = 4$ ,  $x_{n+1} = r x_n(1 - x_n)$ ,  $\varepsilon = 0.05$ ). Shown are the recurrence rate (RR, a), determinism (DET, b), diagonality and verticality (D and V, c), and Shannon entropy (ENTR, d).

## Chapter 3

### Application to Synthetic Data

To develop an intuition for RQA, we can analyze synthetic data with known dynamical properties. Three signals are investigated, generated from the logistic map (May, 1976), the Lorenz system (Lorenz, 1963), and persistence data (Newman, 2013).

#### 3.1 The Logistic Map

Though popularized as a chaotic system by Robert May in the 1970s, the Logistic Map was introduced as a model for population fluctuations before the development of chaos theory (Ausloos and Dirickx, 2006). It is today one of the best studied systems in nonlinear dynamics, and remains a highly attractive equation for its complex behavior, arising paradoxically from extreme simplicity (May, 1976). In the logistic map, each state  $x_{n+1}$  at discrete time  $n + 1$  depends on the state at time  $n$  by

$$x_{n+1} = rx_n(1 - x_n). \quad (16)$$

This equation maps  $x_n$  on the interval  $[0,1]$  to  $x_{n+1}$  on the same interval. Figure 3.1 a) shows a typical orbit, or sequence of successive states  $\{x_n\}_0^\infty$ , for  $r \in [0, 3]$ . In this range,  $x_n$  reaches a stable fixed point. As  $r$  increases, a period-doubling bifurcation occurs at  $r = 3$ , giving rise to a period-2 orbit for  $x \in [3, 1 + \sqrt{6}]$ , in Fig. 3.1 (b). As  $r$  increases, the period doubles to four, eight, and any number  $2^k$ . Figure 3.1 (c) shows a typical period-8 orbit. This period doubling continues to the accumulation point,  $r_\infty \sim 3.570$ , where the period is infinite; beyond this point the dynamics become chaotic, as seen in Fig. 3.1 (d).

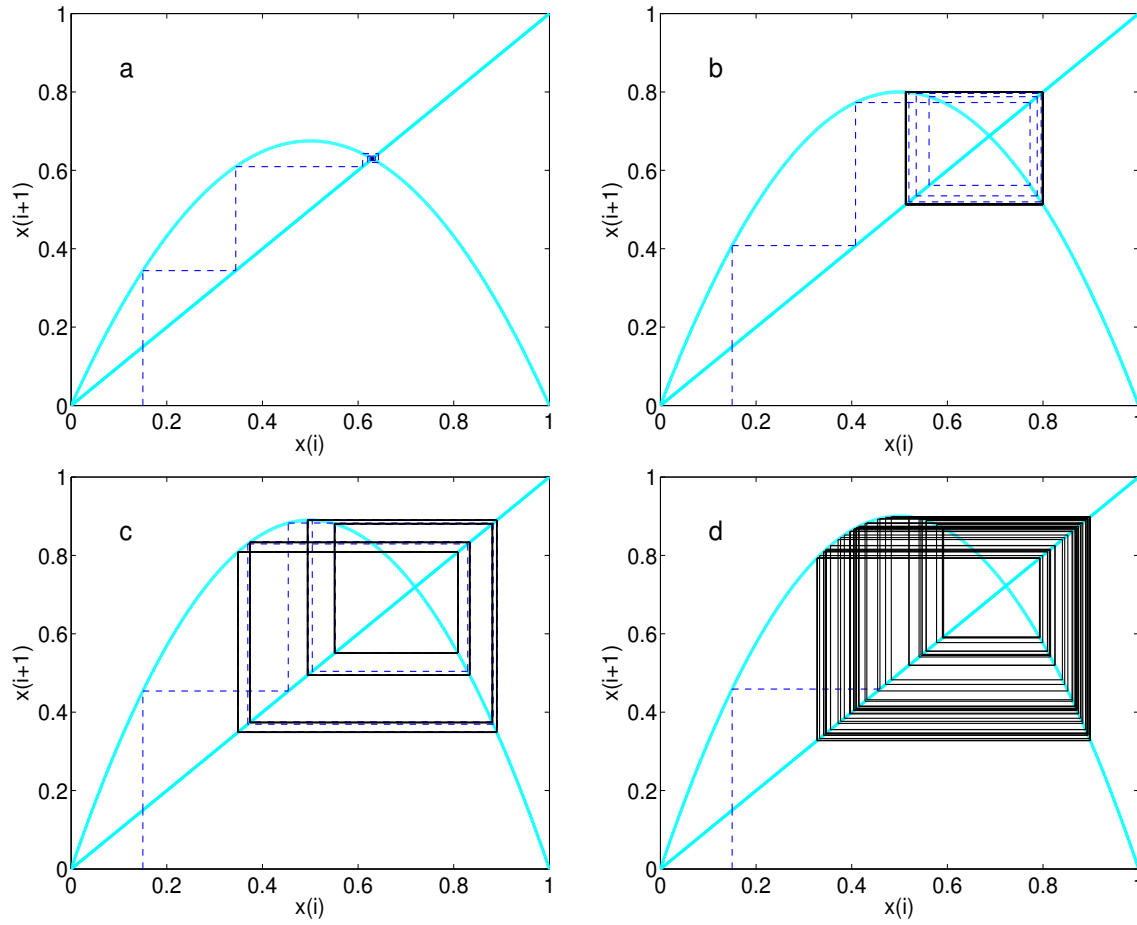


Figure 3.1: Orbits of the logistic map for typical initial values, with  $r = 2.7$  (a),  $r = 3.56$  (b),  $r = 3.2$  (c), and  $r = 3.6$  (d). Transient behavior is in dashed lines.

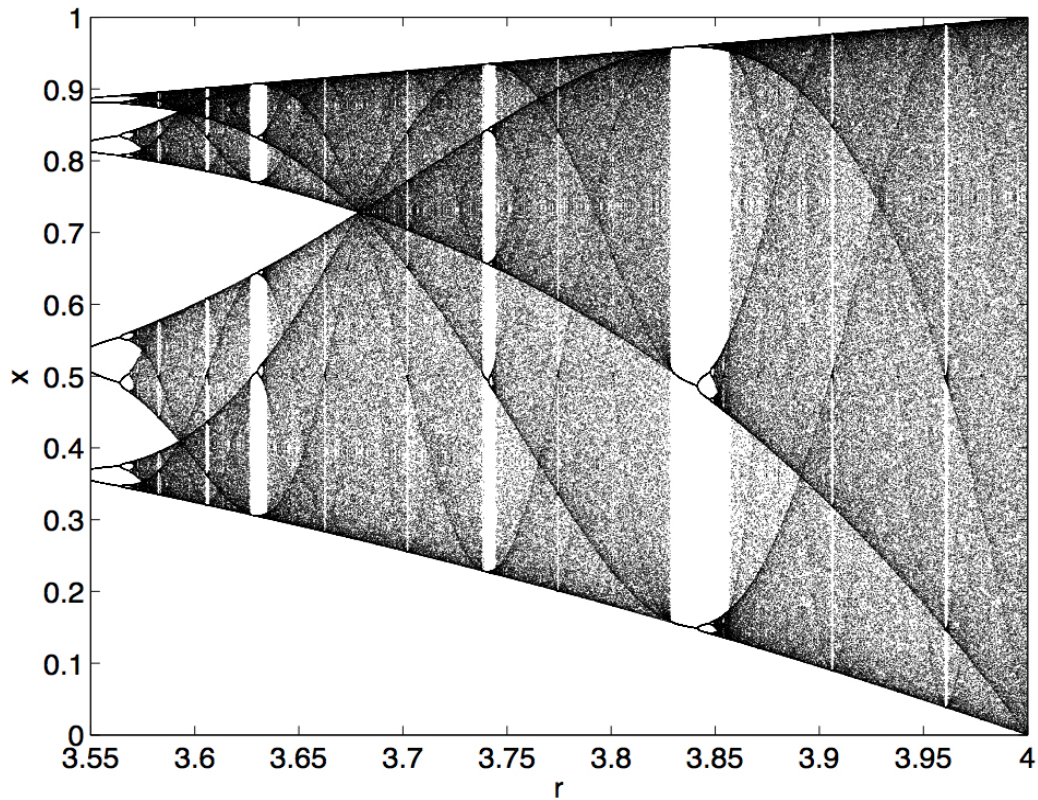


Figure 3.2: Orbit diagram for the logistic map, where  $r$  ranges from 3.55 to 4, showing all possible long-term values for the system in black. Darker areas indicate values visited more frequently by the system.

Plotting the long term behavior for varying  $r$  results in the orbit diagram, Fig. 3.2; orbits are visible in the vertical direction. When  $r > r_\infty$  the dynamics become chaotic, interspersed with periodic windows. This chaotic regime is of particular interest, exhibiting islands of periodicity. The logistic map returns values at superstable points with much greater probability than others. These superstable points are visible as oscillating dark lines in Fig. 3.1.4, which merge near  $r \sim 3.68$ . To identify the periodic windows, transitions, band merging points, and other features of the logistic map is an important challenge for RQA.

### 3.1.1 RQA of the Logistic Map

Figure 3.3 shows the RQA for the chaotic regime of the logistic map. The  $RR$  (Fig. 3.3 a) varies considerably across  $r$ -values. The tolerance parameter,  $\varepsilon$ , was chosen to give a modest recurrence rate based on analysis of the logistic map at  $r = 4$ , but  $RR$  rose almost to 0.35 in the periodic regions. This is not entirely unexpected, as  $x_n$  is constrained to  $m$  specific values in windows with orbit of period  $m$ ; regardless of the value chosen for  $\varepsilon$ ,  $RR$  can never be lower than  $1/m$  in a long signal with only  $m$  possible values. We may infer the number of possible values for  $x$  in the region near  $r = 3.83$ ;  $RR$  is 0.33 there, so this may be a period 3 orbit. But it is not possible to know that the orbit is periodic from  $RR$  alone.

The changing values for  $RR$  also make it difficult to use either  $DET$  or  $ENTR$  as direct measures of the regularity in the system, since both are highly influenced by  $RR$ . The spikes in each measure do correspond with areas of regularity in the logistic map, but all three plots look very much the same, with the exception of  $DET$  for  $L_{min} = 3$  which returns very low values. Knowing the characteristics of the logistic map, the  $x$ -values are quantized only where the system's behavior is periodic; thus the close relationships between  $DET$ ,  $ENTR$ , and  $RR$  are entirely expected, reflecting the signal's high regularity in low-period orbits.

Diagonality (Fig. 3.6 d) confirms the predictability of the signal yielding diagonal lines in the RP. Excluding the periodic windows,  $D$  declines throughout the chaotic regime, reflecting the increasing chaos as the Lyapunov exponent increases. A dip in diagonality immediately after  $r = 3.85$  reveals the irregularity of the signal during the chaos-chaos transition just past the period-3 window's bifurcations, when the chaotic attractor spontaneously increases in size.

Each visible periodic window has some representation in  $D$ ,  $DET$ , and  $ENTR$ . Measures based on vertical lines are low in these areas because the values always change from time  $n$  to  $n + 1$ ; verticality (Fig. 3.6 d) remains near zero, but dips into negative values in periodic regimes.<sup>2</sup> The curious positive spikes in verticality at  $r = 3.68$  and  $r = 4.0$  are seen also in the  $DET$  plot for  $L_{min} = 3$ . The orbit diagram at Fig. 3.2 reveals that the first such spike corresponds to the band-merging point where the superstable points merge; when  $r = 3.68$ , it the system lingers by  $x = 0.73$ . The second spike occurs when the superstable points merge at  $r = 4.0$ . Smaller spikes in  $V$  occur at  $r = 3.76$ ,  $r = 3.79$ ,  $r = 3.88$ ,  $r = 3.93$ , and  $r = 4.0$ , corresponding to the crossing of superstable points in Fig. 3.2.

---

<sup>2</sup>Note that  $D$ ,  $V$ , or  $P_n$  may be negative due to the subtraction of  $RR$ ; a negative value in one of these measures may indicate that points in the RP are organized according to another pattern. For instance,  $V$  and  $P_4$  are negative at  $r = 3.83$ , because this is the highly predictable period-3 window;  $D$  and  $P_3$  return strong positive values there.

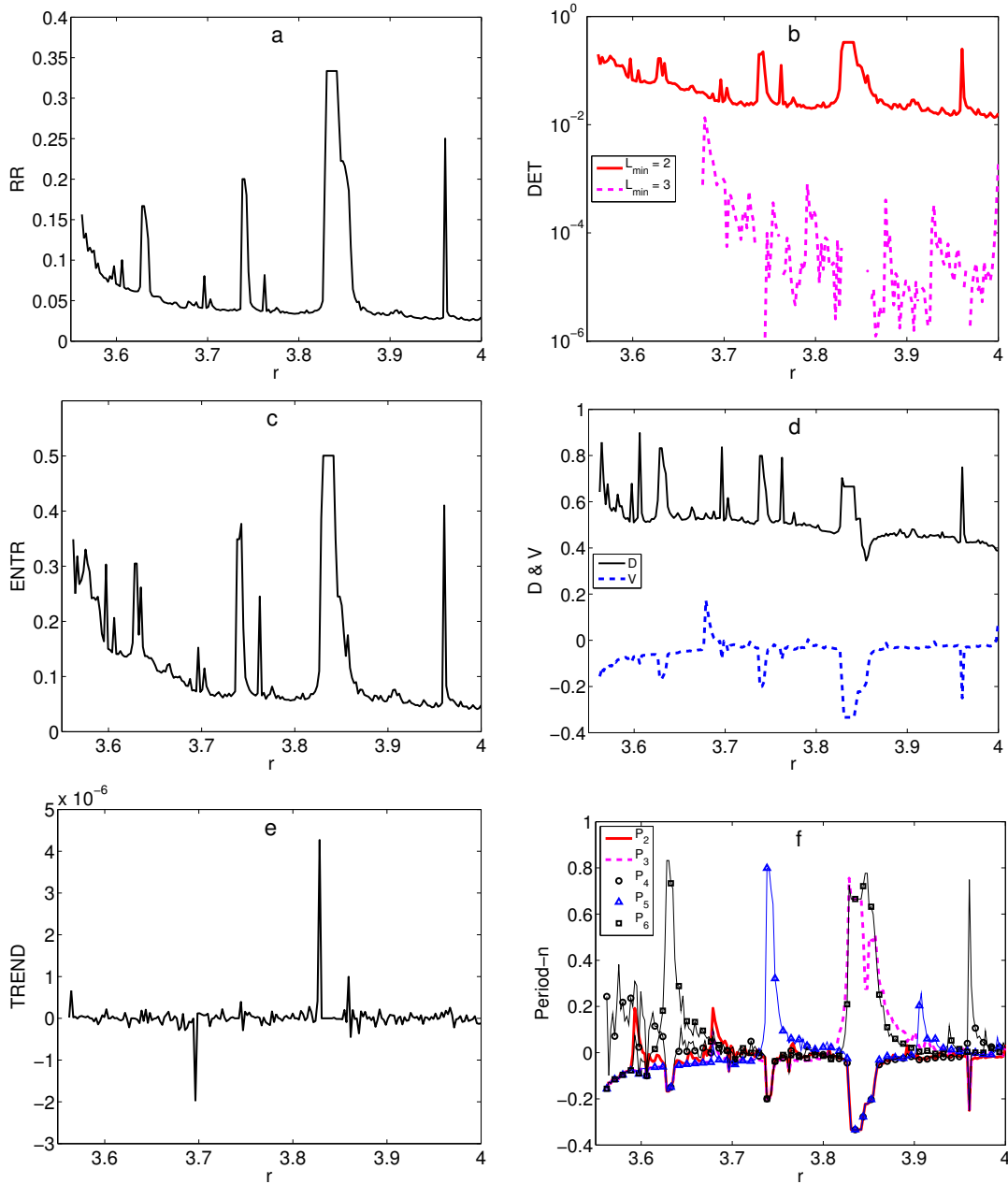


Figure 3.3: Selected RQA measures applied to the chaotic regime of the logistic map, showing the recurrence rate (RR, a), determinism (DET, b), Shannon entropy (ENTR, c), diagonality and verticality (D and V, d), TREND (e), and Period-n ( $P_n$ , f) all vs.  $r$  varying from 3.55 to 4 in 200 steps. Length  $N = 10,000$  and tolerance  $\varepsilon = 0.001$ .



TREND appears to be largely independent of  $r$ , probably because it is insensitive to the changing Lyapunov exponent in the system; TREND returns significant values when the average values in a signal rise or fall with time, but the logistic map returns high and low values without any long-term dependence on time. TREND returns unusual values near  $r = 3.70$  and  $r = 3.83$ . The first of these appears to correspond with a narrow period-7 window, and the next with the start of the large period-3 window, although the significance of these areas is unclear. They may have appeared randomly due to the length of the time series used, but could also be genuine features not fully resolved because of the resolution.

The  $P_n$  measures pick out and classify periodic windows. The windows identified are given in Table 3.1; checking these against the orbit diagram of the logistic map (Fig. 3.2) verifies that these classifications are correct.

Table 3.1: Classification of Periodic Windows

$r$	<i>Period</i>
3.63	6
3.74	5
3.83	3
3.96	4

There are many more periodic windows that went unclassified by  $P_n$ . For example, a narrow period-10 window is visible in the bifurcation diagram close to  $r = 3.60$ , and a period-7 window can be seen near  $r = 3.70$ . These were not found by  $P_n$  because  $n$  was only plotted for  $n \in [1, 6]$ . However, these windows do show up as spikes in the plots of DET, ENTR, and DET (Figs. 3.3 b-d); they were not undiscovered, simply unclassified.

In conclusion, an analyst lacking any knowledge of the logistic map could use RQA to identify the features of the system at different parameter values. One cannot classify the signal as chaotic, or determine system invariants, but values for D and DET in these intermediate ranges do give a sense of aperiodicity. The band merging, chaos-chaos transitions, periodic windows, and increasing chaos throughout the typical regions of the map can also be identified through analysis of the recurrence plot. This suggests RQA measures will also be effective for investigating new and unknown time series.

### 3.2 The Lorenz System

Edward Lorenz was one of the founding thinkers in the field of Nonlinear Dynamics. Building a twelve-dimensional computer model for atmospheric motion, he gradually noticed the patterns did not always change as predicted. When he tried to re-run the system from a previous time, the model gave different behavior—tiny rounding errors in his re-entered data compounded to produce completely new behavior (Gleick, 2011). This sensitivity to initial conditions shown by all chaotic systems is known today as the “butterfly effect.” (Hilborn, 2004)

As Lorenz went on to investigate chaos, he discovered much simpler systems that can give rise to chaotic behavior. One of these chaotic systems, published in his 1963 article, *Deterministic Nonperiodic Flow* (Lorenz, 1963), consisted of only three coupled differential equations, which are known today as the Lorenz System,

$$\begin{aligned}\frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z.\end{aligned}\quad (17)$$

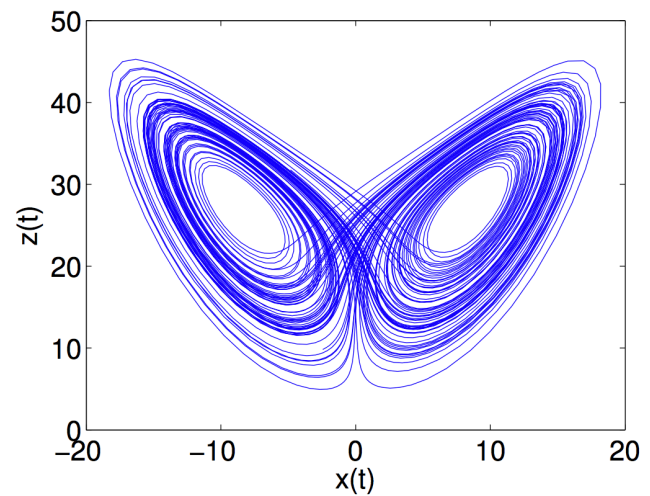


Figure 3.4: Lorenz Attractor for  $\rho = 28$ ,  $\sigma = 10$ , and  $\beta = 8/3$ , pictured in  $x$  and  $z$ .

For specific parameter values, the system exhibits a strange attractor, generated with time step  $\Delta t \sim 0.0164$  and shown in Fig. 3.4 in two dimensions.

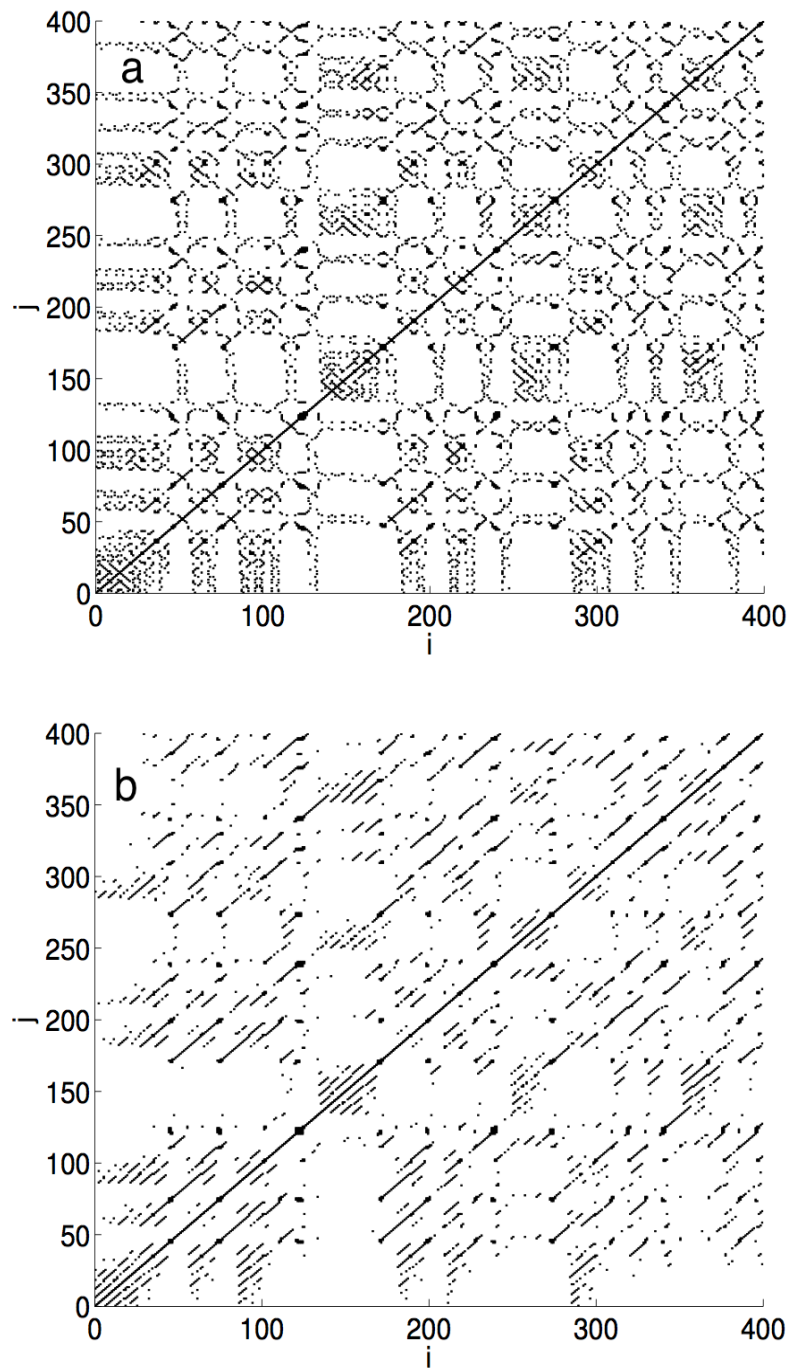


Figure 3.5: Recurrence plots for the Lorenz system, generated from a timestep  $\Delta t \sim 0.0164$ , with (a) no embedding,  $\varepsilon = 0.8$  and (b) two-dimensional embedding,  $\tau = 2 \Delta t$ ,  $\varepsilon = 2.0$  (b). Different values of  $\varepsilon$  were chosen to match recurrence rates across RPs.

### 3.2.1 RQA for the Lorenz System

The RP for the  $x$ -component of the Lorenz System in Fig. 3.5 (a) shows downward-sloping diagonals, which might seem to imply the system runs backwards and forwards in time. But recall from the RPs of the sine curve seen in Fig. 1.1 (c) and 1.5 (a) that, in a deterministic system, such downward sloping diagonals indicate an improper embedding dimension.

The Lorenz attractor has a fractal dimension of 2.06 (*i.e.*, very slightly greater than two). Two dimensions alone are insufficient to generate chaos in a deterministic system, but a two-dimensional embedding yields an acceptable approximation for RQA. Embedding the signal in two dimensions can be accomplished by plotting the signal as  $x(i + \tau)$  vs.  $x(i)$ , where  $\tau$  is the characteristic time-scale of the system.<sup>3</sup> Then, the RP is generated by measuring the Euclidian distance between  $((x(i), x(i + \tau)))$  and  $((x(j), x(j + \tau)))$  for all  $i, j = 1$  to  $N$ . Using the characteristic time separation  $\tau = 2\Delta t, \varepsilon = 2$  resolves the determinism of the system, as shown in Fig. 3.5 (b). The squarish mats from the first RP (Fig. 3.5 a) are echoed in the RP for the embedded signal as only upward-sloping diagonals, verifying that the dimensionality of the system is approximately two.

The 2D-embedded Lorenz system shows similar relationships between RQA measures and tolerance to those seen on other signals. RR and DET increase steadily with increasing  $\varepsilon$ , as shown in Fig. 3.6 (a, b). ENTR (not shown) also rises virtually linearly with  $\varepsilon$ . Diagonality, plotted in Fig. 3.6 (c) reaches a plateau for  $\varepsilon \in [0.5, 1.5]$ ; again, this is the region where  $RR \in [0.01, 0.10]$ , where RQA is normally applied.  $P_n$  returns the highest values for  $P_2$ , as expected given the characteristic time-scale  $\tau = 2\Delta t$ .

---

<sup>3</sup> $\tau$  is determined from the first zero-crossing of the autocorrelation function (Kim, Eykholt, & Salas, 1999); this is also the time separation for which a two-dimensional plot from one-dimensional data,  $x(i + \tau)$  vs.  $x(i)$ , approximately recovers the shape of the strange attractor.

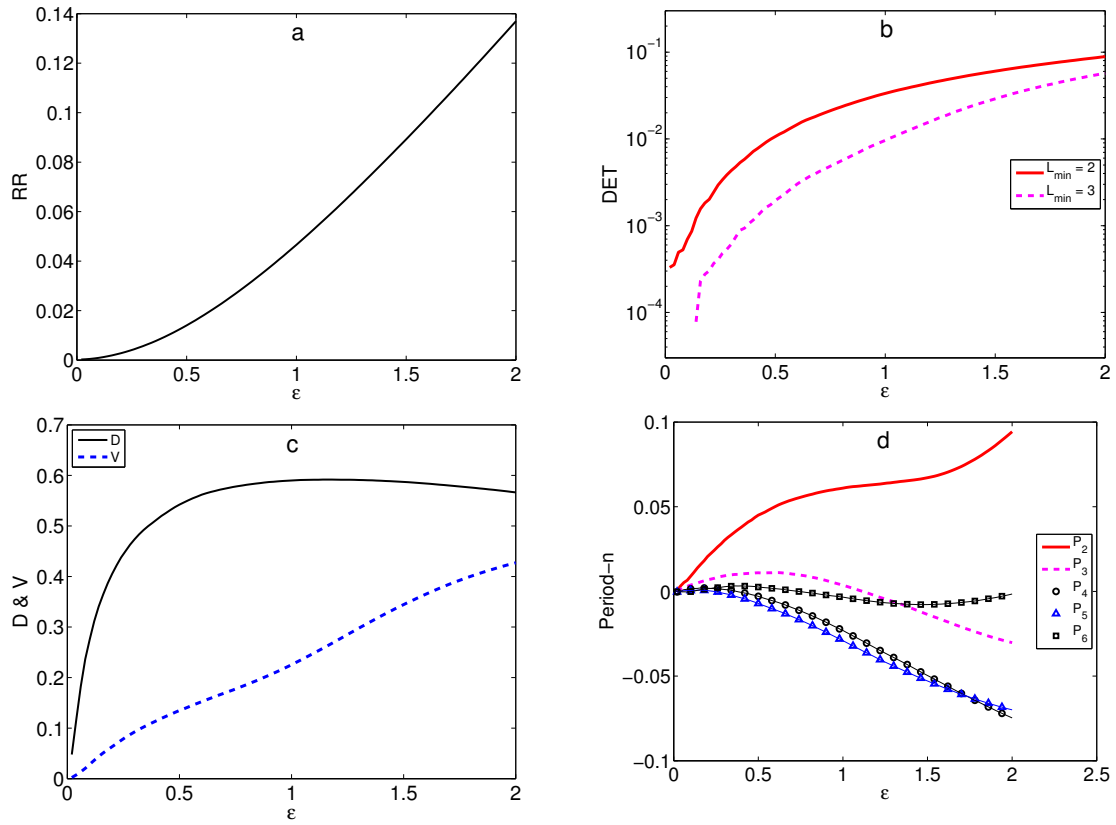


Figure 3.6: RQA measures applied to the two-dimensionally embedded signal from Fig. 3.5, showing recurrence rate (RR, a); determinism (DET, b), diagonality and verticality (D and V, c) and Period- $n$  ( $P_n$ , d) as a function of tolerance ( $\epsilon$ ). RR values range from 0.0 to 0.14.

### 3.3 Persistence Data

Rearranging random data can create an information-rich signal. Beginning with a time series of Gaussian noise, the sequence of data points may be shuffled to create a *persistent* signal where similar  $x$ -values follow one another. An *antipersistent* signal was generated analogously, such that  $x$ -values tend to oscillate across the mean (Newman, 2013).

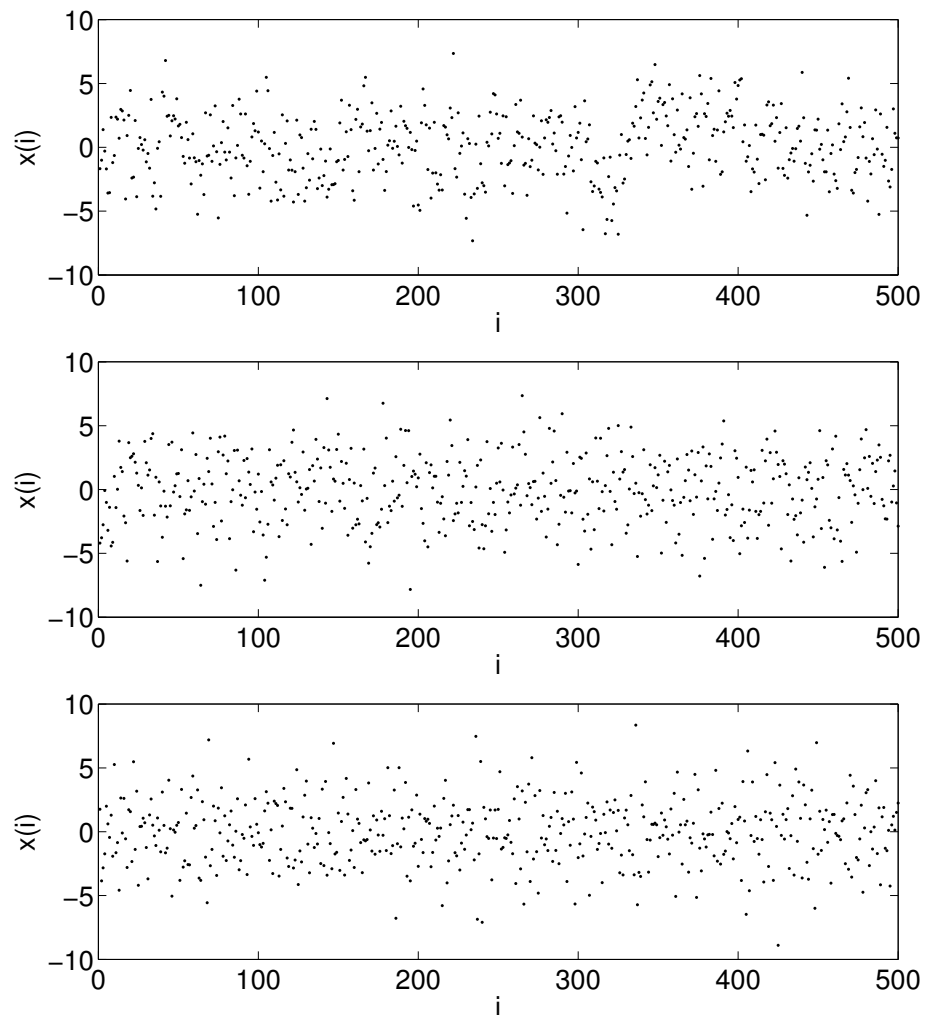


Figure 3.7: First 500 dataspoints in a) persistent, b) random, and c) antipersistent signals.

These three signal types (Fig. 3.7) have identical probability distributions, but can be readily discriminated by plotting  $x(i+1)$  vs.  $x(i)$  (Figs. 3.8 a, b, c). Persistent data form an upward slope, random data form a centered scatterplot, and antipersistent data form a downward slope. This arises high  $x(i)$  in a persistent series tends to be followed by another high  $x(i+1)$ , while high  $x(i)$  in an antipersistent series tends to be followed by a low  $x(i+1)$ .

Recurrence plots for each dataset are shown in Fig. 3.8 (d, e, f). Small squarish mats of diagonals appear in the RPs for persistent (d) and antipersistent (f) data, with a more disorganized RP for random data (e). This visual difference suggests the possibility of signal discrimination through RQA.

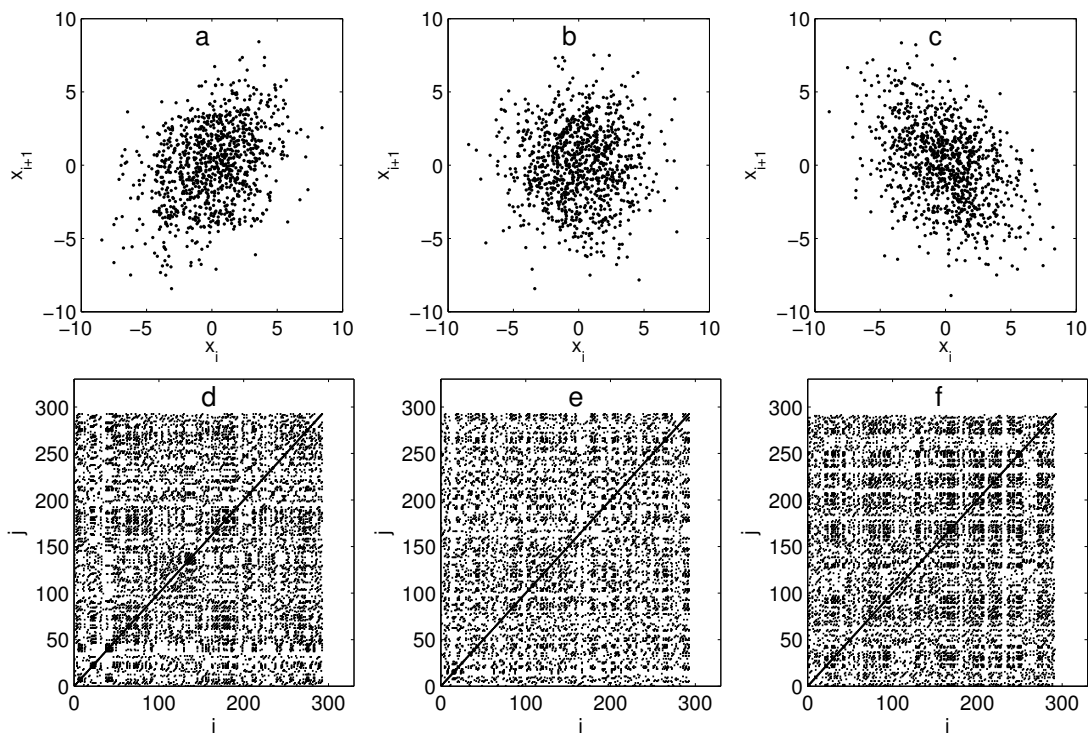


Figure 3.8:  $x(i+1)$  plotted against  $x(i)$  for (a) persistent, (b) random (b), and (c) antipersistent series. RPs are shown below for the first 300 data points of the same (d) persistent, (e) random, and (f) antipersistent series, with tolerance  $\varepsilon = 1.7$ . Each signal originated from the same Gaussian noise function with mean  $\mu = 0$ , standard deviation  $\sigma = 1$ , and length  $N = 10,000$ .

### 3.3.1 RQA for Persistence Data

The clear visual differentiation of signal types in two-dimensional plots of  $x(i+1)$  vs.

$x(i)$  (Fig. 3.8) suggests that two dimensions may be sufficient for embedding, with a time separation of  $\tau = 1$ . After attempting different embeddings, it was found that the signals could be indeed discriminated well under a two-dimensional embedding with  $\tau = 1$ .

Figure 3.9 a) shows the results for diagonality only, though all diagonal-based measures (DET, ENTR, L, and D) discriminated between the random and rearranged signal types similarly, with both persistent and antipersistent signals showing higher values than the random signal. Because the antipersistent signal tends to switch from high to low values, it is more likely to follow recurrent paths from low to high and high to low, producing more diagonal lines in the RP.

Figure 3.10 gives the results for verticality. All vertical-based measures (LAM, TT, and V) discriminated well between the three signal types. The tendency in persistent data for each data value to remain near the preceding value results in longer vertical lines, while antipersistent data tends to change values even more often than random data.

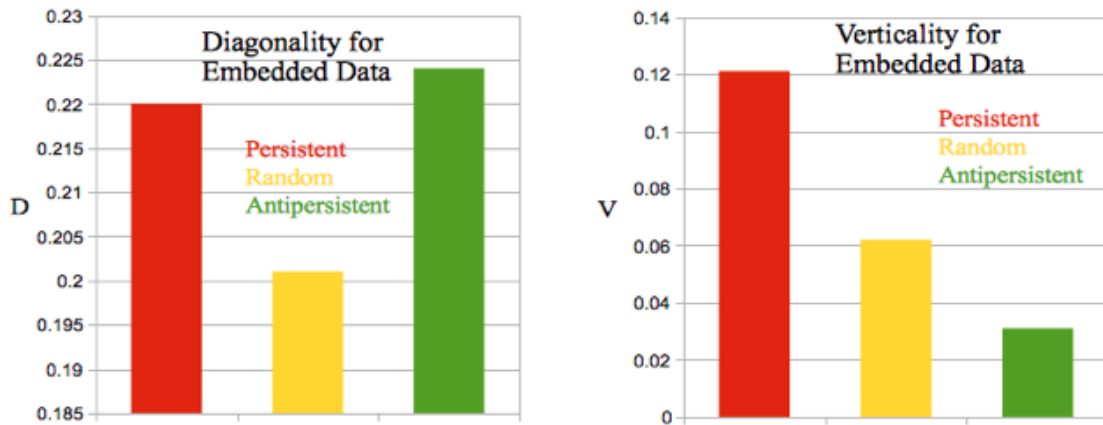


Figure 3.9. Comparison of diagonality (D, a) and verticality (V, b) from the RP of the three signals in Fig. 3.8, embedded in two dimensions with  $\tau = 1$ ,  $\varepsilon = 1.7$ .

All three signals returned values above zero for D and V; this is likely a consequence of the two-dimensional embedding, as the values for D and V were near zero for RPs of the one-dimensional signals. Since all three embedded signals had very similar recurrence rates (persistent,  $RR = 0.105$ ; random,  $RR = 0.097$ ; and antipersistent,  $RR = 0.105$ ) the values for D and V are comparable across time series.

The RQA measures for these time series showed no unusual dependence on  $\varepsilon$ . To provide a sample, Fig. 3.10 shows selected RQA plots for persistent data, with all measures rising



monotonically as  $\varepsilon$  increases from 0.1 to 2.0. No significant changes in the discrimination were observed for any values of  $\varepsilon$  within this range.

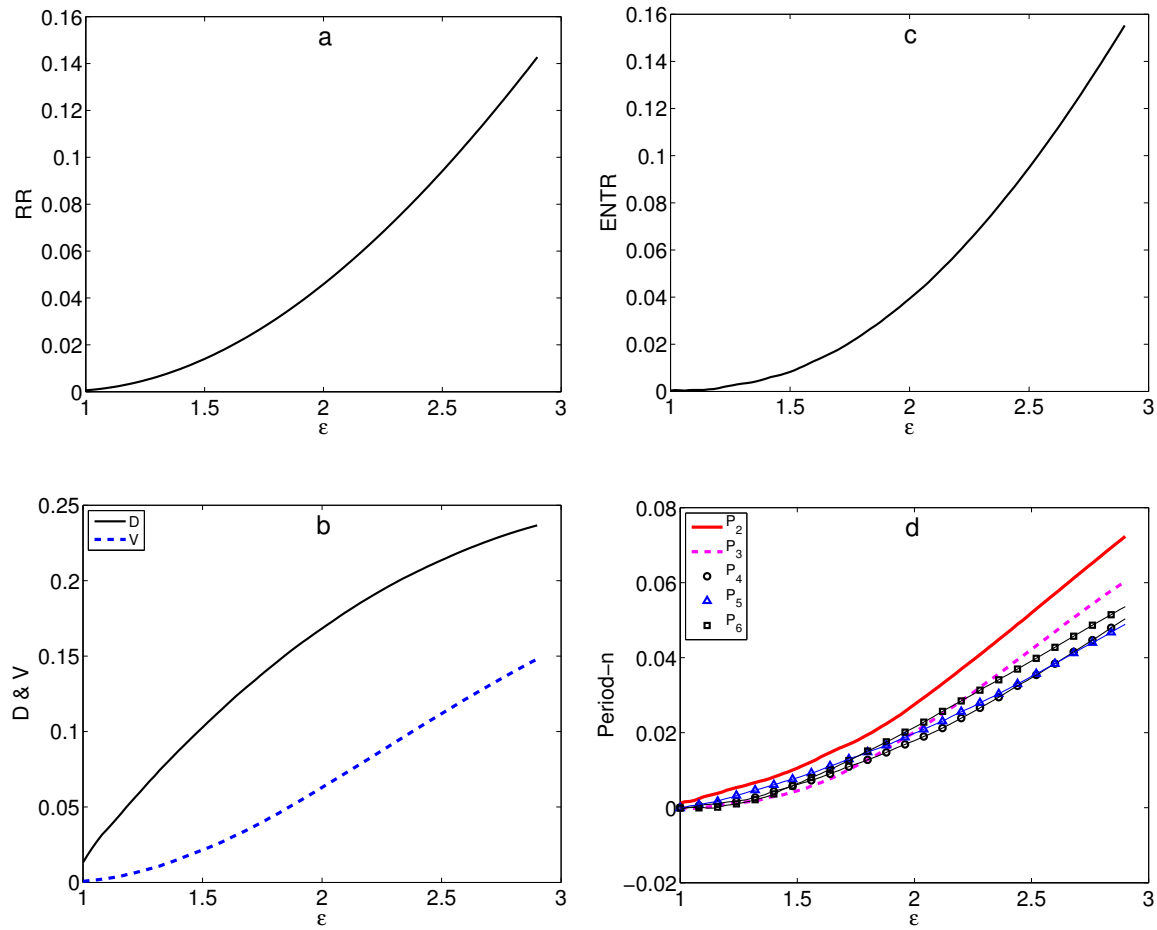


Figure 3.10. Selected RQA measures vs. tolerance ( $\varepsilon$ ) for persistent data, showing the recurrence rate (RR, a), diagonality and verticality (D and V, b), Shannon entropy (ENTR, c), and Period-n ( $P_n$ , d).

### 3.4 Summary of Synthetic Data Investigation

When interpreting RQA values, the classic measures of Shannon entropy, determinism, and laminarity return values in the range  $[0, 1]$ , with high values indicating more organization of the RP in the form of diagonal or vertical lines. But user selection of shorter or longer  $L_{min}$  and  $V_{min}$  can push the results closer to 0 or 1. Similarly, these measures are highly dependent on RR, so that choosing very small or large  $\varepsilon$  can also push these values close to 0 or 1. Because of this, ENTR, DET, and LAM should be interpreted in a relative sense to reveal differences between signal types or parameter values in the system under study.

Diagonality, verticality, and period- $n$  return values in a similar range, though shifted downward by the subtraction of the recurrence rate. Higher values again indicate more regularity in the form of more time spent following recurrent paths, remaining near a previous value, or following periodic trajectories. These new measures are more robust with respect to the recurrence rate in the typical range of  $RR \in [0.01, 0.1]$ , allowing for somewhat more objective interpretation; for the signals studied here, values from approximately 0.2 to 0.6 corresponded to chaos, with higher values seen only in periodic signals. Values near 0 are found in uniform random noise, although small negative values can also occur when the arrangement of the RP contains fewer lines or periodic values than uniform random noise.

Overall, we have seen that RQA provides a tool for investigating and discriminating between time series. RQA may not be ideal for all applications; persistence datatypes were very easily discriminated by plotting  $x(i+1)$  vs.  $x(i)$ . There are many alternative methods for investigating time series, such as statistical analysis (Anderson, 2011) or spectral analysis (Marple, 1987).

Each step in the RQA process results in a loss of information. Converting a raw signal into a recurrence plot, and further reducing this signal to a few values, represents a massive reduction in the information available—two very different signals can have highly similar values for DET, LAM, and other measures. If discrimination is the goal, however, removing extraneous information can be highly useful to produce a clear picture. Ultimately RQA represents one tool among many, and there is no reason to believe that data analysis begins and ends with the recurrence plot.



## Chapter 4

### Application to Natural Data

The neurological structure underlying respiration, though complex, is gradually yielding to investigation (DePuy, Kanbar, Coates, Stornetta, & Guyenet, 2011; Messier, Li, & Nattie, 2008). An emerging model has serotonergic neurons firing with increased frequency in response to elevated blood serum levels of CO<sub>2</sub> (Richerson, 2004), with nonserotonergic neurons firing with decreased frequency in response to the same stimulus (Messier *et al.*, 2004). Although these neuron types can be discriminated by their response to CO<sub>2</sub> challenges, it is also useful to discriminate between them according to baseline behavior. The standard means for discrimination used in biology involves calculating the mean and standard deviation of inter-spike intervals and comparing these to a reference line (Mason, 1997). Although serotonergic cells are typically classified using this method, many non-serotonergic cells are misclassified as serotonergic. RQA may allow classification of these cells.

We used raw data that were collected by Dr. Harris' research group (Institute of Arctic Biology, University of Alaska Fairbanks). Voltage spikes were recorded from the brains of anesthetized rats, as shown in Fig. 4.1 (Harris & Iceman, 2012). Raw data were then converted into inter-spike intervals (ISI) by measuring the time delay between spikes; spikes were counted as any voltage increase past a given threshold, set above the noise band. This threshold was not identical across signals because of sampling differences changing the size of the noise band. Although noise presented a problem in some datasets, genuine voltage spikes had a characteristic pattern (Fig. 4.1 b): a rise of increasing steepness, followed by a plunge below baseline, and final relaxation to baseline. In difficult cases, true voltage spikes could be distinguished from noise by this characteristic shape.

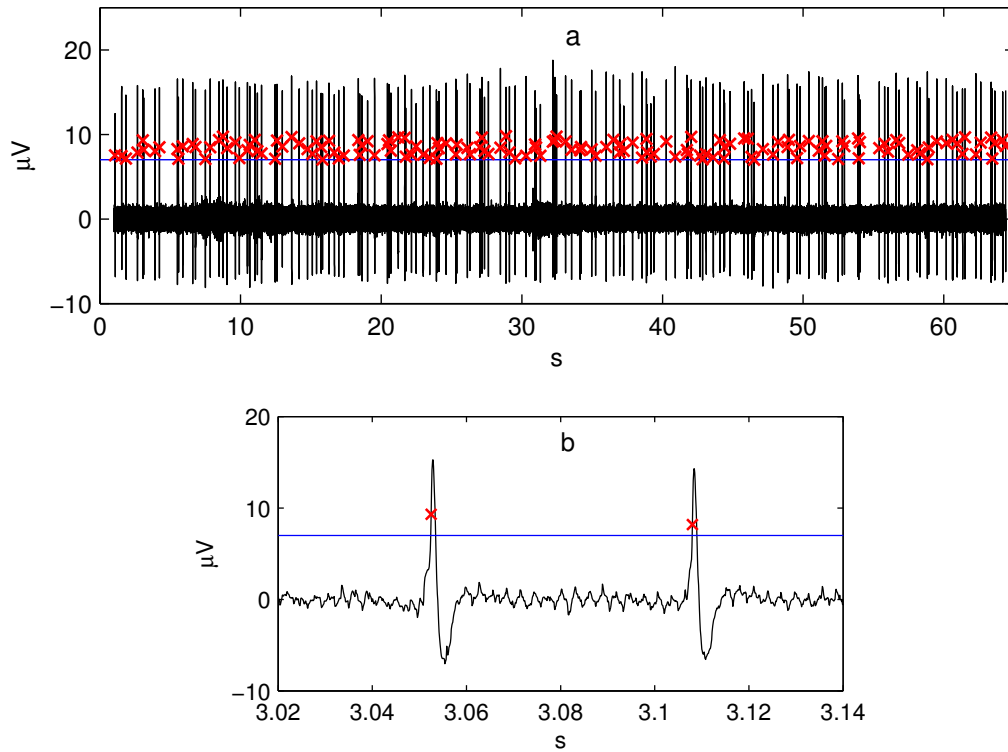


Figure 4.1. Voltage signal for a typical neuron (a) and close-up (b). A horizontal line marks the  $7\mu\text{V}$  threshold for determining spike positions, marked with red crosses.

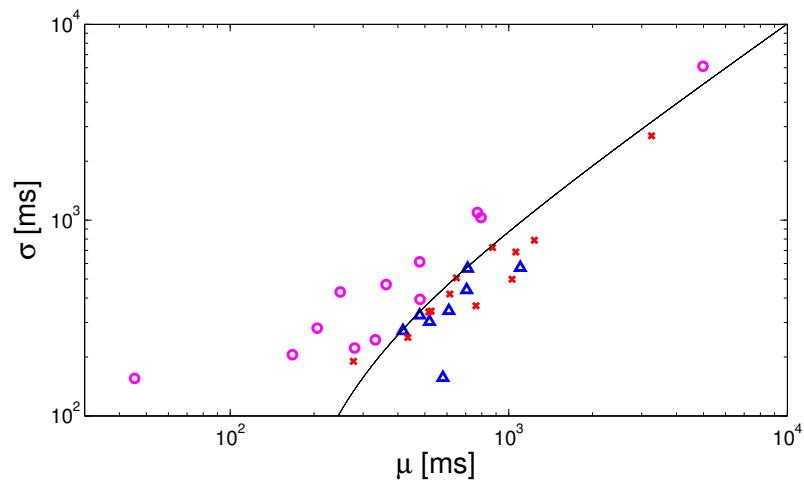


Figure 4.2. Standard deviation ( $\sigma$ ) vs. mean ( $\mu$ ) for serotonergic neurons (blue triangles), nonserotonergic neurons (pink circles) and misclassified nonserotonergic neurons (red crosses). The discriminant line  $\sigma = (\mu - 146)/0.98$  is plotted in black.

## 4.1 Preliminary Analysis

Thirty-nine data runs were received and separated into three categories: serotonergic, nonserotonergic, and misclassified nonserotonergic. Most runs were short, taken at lengths below 5 minutes and containing fewer than 200 spikes, although one run contained 1365 spikes due to very high firing frequency.

The discrimination method typically used in biology was applied, comparing the standard deviation ( $\sigma$ ) and mean ( $\mu$ ) of a cell's ISI to a threshold as seen in Fig. 4.2; the line  $\sigma = (\mu - 146)/0.98$  is the classical discriminant criterion for classifying serotonergic and nonserotonergic cells (Mason, 1997). All signals from serotonergic cells are confined to the region on or below the discriminant line, although signals from nonserotonergic cells are scattered both above and below the line. One cell, originally misclassified in the biologists' dataset (red cross at  $(10^{2.4}$  ms,  $10^{2.3}$  ms)), was correctly classified here, possibly due to differences in threshold selection, although most of the initially misclassified nonserotonergic cells remained on the lower (serotonergic) side of the line.

The Pearson correlation of  $x_n$  against  $x_{n+1}$ , can be a quick method for discrimination in many time series (Fig. 3.8). As a biological argument for this method of analysis, some cells might require longer relaxation times after firing rapidly twice in a row. However, as shown in Fig. 4.3 for two typical signals, the plots were relatively diffuse, and the available time series were generally too short to reveal significant correlations between successive inter-spike-intervals.

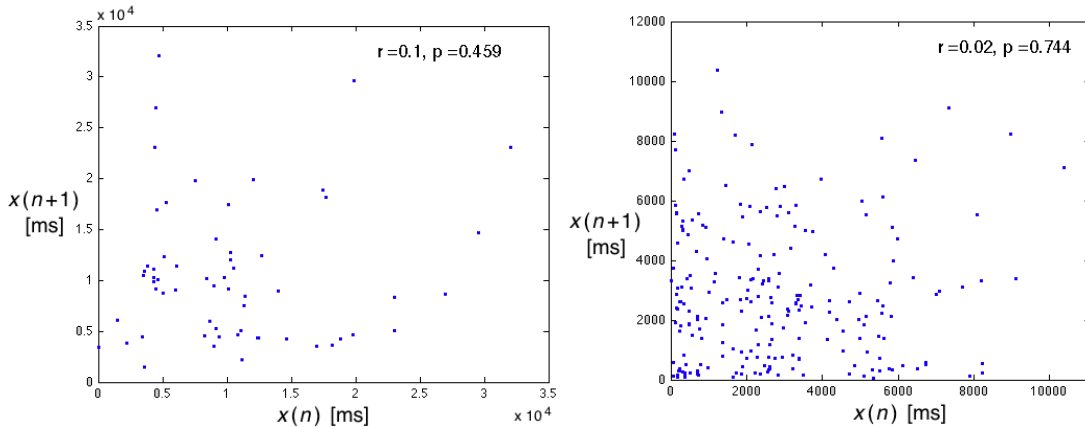


Figure 4.3. Typical return plots of inter-spike interval ( $x_{n+1}$ ) vs. ( $x_n$ ) for a misclassified serotonergic cell (a) and nonserotonergic cell (b). Pearson correlation coefficients ( $r$ ) with significance levels ( $p$ ) are shown in each figure.

Turning to RQA, Fig. 4.4 shows RPs for three typical signals. The plots appear random or nearly so, with  $D$  and  $V$  returning values near or below 0. It might be argued that the nonserotonergic cell (Fig. 3.3 b) lingers somewhat longer in the same state with  $V = 0.03$  vs.  $V = 0.01$  for the serotonergic cell (Fig. 3.3 a) and  $V = -0.01$  for the misclassified nonserotonergic cell (Fig. 3.3 c). But such small differences are not necessarily meaningful, given that none of the plots were greater than length  $N = 250$ .

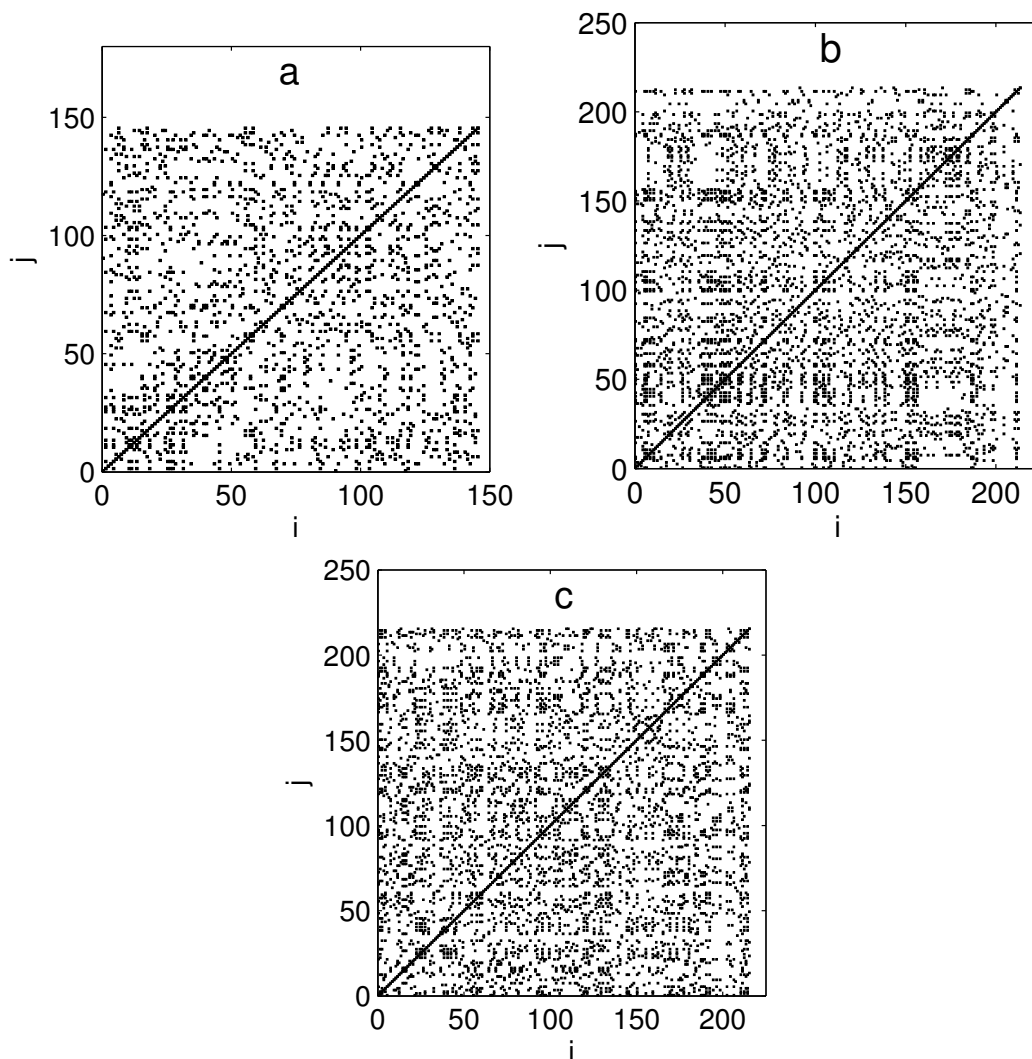


Figure 4.4. Typical recurrence plots for a serotonergic cell (a), nonserotonergic cell (b), and misclassified nonserotonergic cell (c), with tolerance ( $\varepsilon$ ) chosen to give a recurrence rate of  $RR = 0.1$  across all cases.

Overall findings are summarized in Fig. 4.5. RQA values for all serotonergic cells are confined to low values, in a way that suggests the possibility for discrimination. But most nonserotonergic cells also return values in a similar range. The signal with the highest recurrence rate originated from the nonserotonergic cell previously mentioned with a very high firing rate, leading to a reduced ISI range.

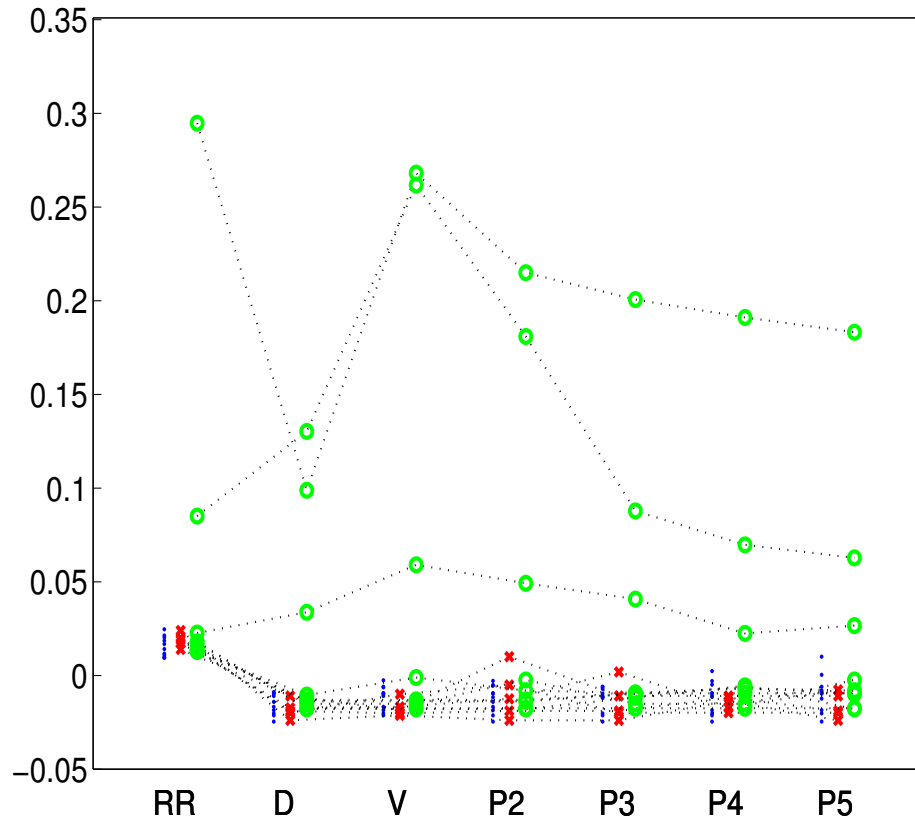


Figure 4.5. Recurrence rate (RR), diagonality (D), verticality (V) and Period-n ( $P_2$ - $P_5$ ) for serotonergic (blue dots), misclassified serotonergic (red crosses), and nonserotonergic cells (green). Dotted lines connect values for same data series. All measures were calculated from recurrence plots using a tolerance threshold ( $\varepsilon$ ) of 1 ms.

## 4.2 Minimum Trial Length

Throughout the analysis, the difference in the lengths of the data series were striking. Some time series had fewer than 50 data values, raising the question of whether they were long enough for analysis, and what the minimum usable data length should be.



To investigate the robustness of classification, the time series were split into parts and re-plotted according to each part's mean and standard deviation. Figure 4.6 shows the process for a longer dataset; the discriminant line was crossed when the entire series of length  $N = 353$  was split into lengths below  $N = 200$ . Given that only 12 of the 29 available time series were above this length, the number of ISI counted for any given neuron was clearly a problem for discrimination.

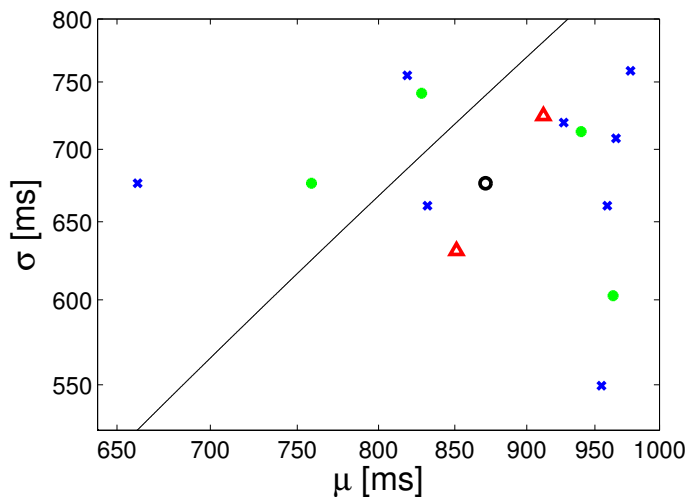


Figure 4.6. Log-Log plot of ISI standard deviation ( $\sigma$ ) vs. mean ( $\mu$ ) for a single time series (black circle) split into segments of length 50 (blue crosses), 100 (green dots), and 200 (red triangles). The discriminant line  $\sigma = (\mu - 146)/0.98$  is plotted in black.

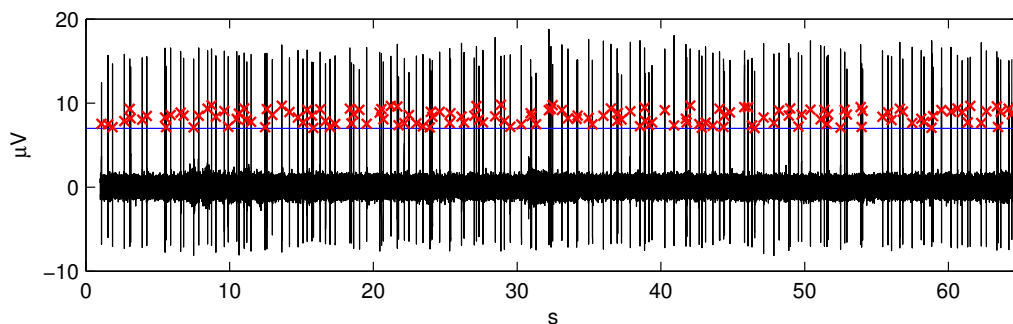


Figure 4.7. Voltage signal for high frequency neuron. A horizontal line marks the  $7\mu V$  threshold used to determine spike positions, marked with red crosses.

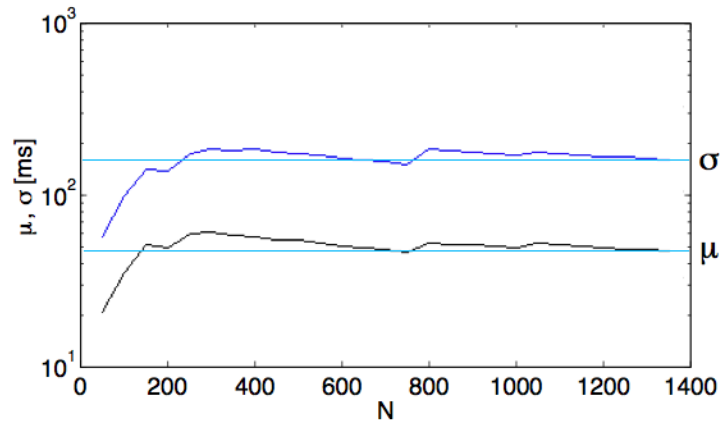


Figure 4.8. ISI standard deviation ( $\sigma$ , blue) and mean ( $\mu$ , black) vs. data length ( $N$ ) for the high-frequency cell. Horizontal lines are added to guide the eyes.

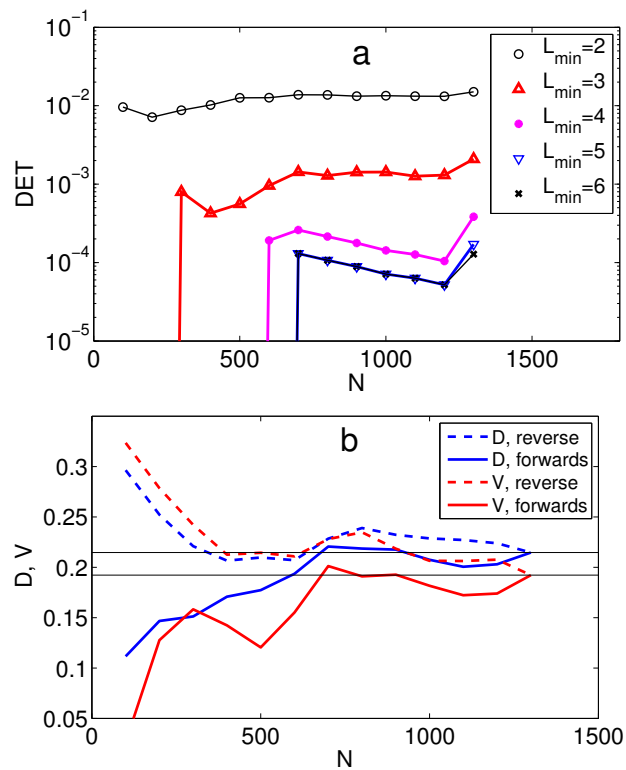


Figure 4.9. Convergence behavior of RQA measures for the signal from Fig. 4.7. a) determinism (DET) for varying  $L_{min}$ . b) Diagonality (D, blue) and verticality (V, red). Convergence behavior was explored by starting from the beginning of the series (solid lines) and from the end (dashed lines); horizontal lines are added to guide the eyes.

In order to determine the minimum length necessary for convergence, the high-frequency nonserotonergic series with 1365 spikes was used (Fig. 4.7). Figure 4.8 shows convergence of mean and standard deviation with increasing series length. These measures, typically used in biology, did not converge until 300 ISI, suggesting that classification of other cells with more typical firing rates by mean and standard deviation also requires longer runs than were available. This, together with the analysis of segments from Fig. 4.6, suggests that longer time series are necessary to accurately classify neuron types.

The convergence of diagonality and determinism with series length were investigated for the same dataset. Determinism (Fig 4.9 a) increases with increasing data length; the longest diagonal line also increased steadily towards 600 and then jumped from  $L_{max} = 6$  to  $L_{max} = 11$  between  $N = 600$  and  $N = 700$ . Convergence behavior of diagonality and verticality are shown in Fig. 4.9 (b), starting from the beginning and the end of the signal; as is typical for these measures, D and V do not converge until  $N > 600$ .

### 4.3 Conclusions

The minimum data length for typical classification methods is at least 200 ISI, and longer series are necessary for evaluation via RQA. Although longer runs were not available, this result remains informative for research methods in biology. The problem of misclassified serotonergic cells by plotting mean and standard deviation might only be a consequence of insufficient sampling time; if so, this issue could be resolved by sampling out to ISIs of 300 or more.

The deeper question of whether the cell types can be distinguished by their firing *patterns* remains unresolved. RQA measures generally returned low values, indicating that the signals could be so complex as to be indistinguishable from random noise. Yet the sample size, and particularly the trial lengths, were too small to draw any conclusions. Further research in this direction will require ISIs of 1000 or greater; given typical firing frequencies, trial runs of ten minutes are recommended. Although such data are required for the analysis, they might not be feasibly achieved experimentally if the “brain condition” is non-stationary over such lengths.

## Chapter 5

### Conclusion

The recurrence plot provides a method for time series analysis by displaying all points in a time series that have similar values at time  $i$  and  $j$ . Diagonal lines and other features in the recurrence plot reveal underlying organization in the time series, and can be used in recurrence quantification analysis (RQA) to identify characteristics of a single signal, or discriminate between different signal types.

Three new RQA measures were developed: *Diagonality* quantifies recurrent points in diagonal lines, *verticality* quantifies recurrent points in vertical lines, and *period- $n$*  quantifies recurrent points in periodic trajectories. These and classical RQA measures were applied to the logistic map (May, 1976), identifying band merging points, periodic windows, increasing chaos, and other transitions with changing parameter values. RQA quantified the dimensionality of the Lorenz attractor (Lorenz, 1963) as roughly two, and discriminated between persistent, random, and antipersistent data series of differing Hurst exponents (Hurst, 1951). After examining convergence behavior of RQA measures on these well-studied systems, the same methods were applied to neuron data with the goal of classifying different cell types on the basis of their firing patterns. The available data, however, were of insufficient length for classification, with a minimum series length of 600 spikes being required. Whether such neurons can be discriminated on the basis of RQA remains an open question.

When first approaching a new signal, it helps to select a reasonable tolerance threshold ( $\varepsilon$ ) and visually inspect the recurrence plot of the signal without embedding before deciding how to move forward. Many upward-sloping diagonal lines indicate a highly regular signal, and the characteristic time-scale can be inferred from the spacing between such lines. Downward sloping diagonals suggest the need for a higher embedding dimension. And areas without recurrence indicate the presence of exceptional points taking different values from most of the data series.

Technical finesse is required in selecting the appropriate tolerance parameter, since the choice of tolerance threshold determines the number of recurrence points in the recurrence plot. In comparing signals, the option exists to standardize plots for different signals on the recurrence rate, tolerance threshold, or neither. The analyses discussed here generally controlled for the recurrence rate by selecting the tolerance to give recurrence rates near 0.03. However, it is not necessary to standardize either the recurrence rate or the tolerance;

if datasets vary greatly in range, one option is to standardize the tolerance as a varying fraction of the signal range or multiple of the standard deviation of the signal's noise.

The length of the series must also be considered. In the signals investigated here, gross features were detected on the recurrence plot of even very short series of about 200 points, but convergence behavior of RQA measures were insufficient for fine discrimination in time series of length below 1000, as was concluded for the neuron signals analyzed in this work. This may not be the case with other signals; more datapoints would be needed to resolve the dynamics from a signal with high resolution. Long data series are not always available, however. In signals with intermittent behavior, or in experiments with short-term response, RQA measures can only be taken as a representation of the system during that transitory state.

Even given a series of appropriate length, the choice of the embedding dimension remains. Given a scalar signal from an  $n$ -dimensional system, the series may be embedded in higher dimensions to recover the dynamics. Deterministic systems should not show downward sloping diagonals in their recurrence plots. Finding the characteristic time-scale of the signal and embedding on that scale is required.

Lastly, interpretation of the results may not be straightforward. For instance, chaotic signals seem to give values for diagonality in the range from 0.2 to 0.6, with random noise returning lower values, and periodic signals returning higher values, but these are not fixed limits. Classical measures, such as determinism, can give any value from 0, for random signals with a low tolerance threshold, to 1, for regular signals or recurrence plots generated with very high tolerance. There are also conceptual pitfalls in the terminology of Recurrence Quantification Analysis. The RQA measures for “entropy” and “determinism” have unusual meanings. The Shannon entropy as introduced by Shannon (1948) describes the information encoded in a data signal (Ihara, 1993), whereas here the Shannon entropy is a measure of structural regularity; and a high value on determinism indicates regularity in the signal and recurrence plot, but not necessarily determinism in the classical sense of cause and effect.

## 5.1 Outview

Discrimination between serotonergic and nonserotonergic neuron types remains a subject for further investigation, since these two cells play different roles in respiration, specifically, excitation and inhibition of breathing rate. Current datasets were not long enough to achieve convergence in RQA measures. Longer time series of baseline behavior would allow

for the nature of these cells and their differences to be fully explored.

Some outstanding methodological issues remain in the field of Recurrence Quantification Analysis. The measure TREND, as defined in RQA, reflects nonstationarity in signals with rising or falling values over time, measuring this through a formula very similar to the Pearson product-moment correlation coefficient, (Stigler, 1989) except that it is normalized differently to return very small results, below  $10^{-5}$  in absolute magnitude. Further work may be worthwhile to produce a measure returning values on a similar scale as the others, and corresponding more clearly to the Pearson product-moment correlation coefficient. Regarding the new measures of diagonality, verticality, and period-n, there is currently no clear way of determining whether a small positive reading indicates a significant deviation from random noise. Determining statistical  $p$ -values for these measures in data series of different lengths would also be useful in order to apply rigorous hypothesis tests to the results of RQA.

Compared to other quantities, such as the Lyapunov exponent or fractal dimension, which determine dynamical invariants of a system (Strogatz, 2001). RQA measures do not classify time series in such a mathematical sense. Ultimately it would be desirable to ground RQA measures on a rigorous mathematical foundation. The Lyapunov exponent identifies the existence of, and quantifies the degree of, chaos, but it is not clear whether RQA measures can do this. They remain highly useful from a purely empirical standpoint, but much could be gained from theoretical work establishing the relationship such measures have to existing dynamical invariants.

Finally, the values within time series investigated here have generally fallen within a continuous range. But text, musical scores, dice-rolls, codes, and other quantized signals offer rich areas for exploration. Particularly, differences in musical genres could be investigated to create algorithms for discriminating between types of music. Or, questions regarding the difference in composer eminence could be explored by using RQA. *e.g.*, would Mozart or Beethoven show more irregularity in their music scores than less well-known composers such as Schulhoff? Ultimately, any phenomenon displaying recurrent behavior is a candidate for analysis via the recurrence plot.



## References

- Anderson, T. (2011) *The statistical analysis of time series. Vol. 19.* New York: John Wiley & Sons.
- Ausloos, M., and Dirickx, M. (Eds.) (2006) *The logistic map and the route to chaos: From the beginnings to modern applications.* New York: Springer.
- DePuy, S. D., Kanbar, R., Coates, M. B., Stornetta, R. L., and Guyenet, P. G. (2011) "Control of breathing by raphe obscurus serotonergic neurons in mice." *The Journal of Neuroscience* 31: 1981.
- Eckmann, J. P., Kamphorst, S. O., & Ruelle, D. (1987) "Recurrence plots of dynamical systems." *Europhys. Lett*, 4: 973.
- Frazier A., White C., Harris S., Wilson T. Jr. (1963) "Surfin' Bird," recorded by The Trashmen. On *Surfin' Bird, 7*" record. Garret: Minneapolis.
- Gleick, J. (2011) *Chaos: Making a new science.* New York: Open Road Media.
- Harris, M. and Iceman, K. (2012) Raw data courtesy of Dr. Michael Harris and RA Kimberly Iceman, UAF.
- Hilborn, R. (2004) "Sea gulls, butterflies, and grasshoppers: A brief history of the butterfly effect in nonlinear dynamics." *American Journal of Physics* 72: 425.
- Hurst, H. E. (1951) "Long term storage capacity of reservoirs." *Trans. Am. Soc. Civ. Eng.* 116: 770.
- Ihara, S. (1993) *Information theory for continuous systems.* Singapore: World Scientific.
- Kim, H., Eykholt, R., and Salas, J. D. (1999) "Nonlinear dynamics, delay times, and embedding windows." *Physica D: Nonlinear Phenomena*, 127: 48.
- Koebbe, M. and Mayer-Kress, G. (1992) "Use of recurrence plots in the analysis of time-series data." *Proc. SFI Studies in the Science of Complexity, eds. Casdagli, M. & Eubank, S. (Addison-Wesley, Redwood City):* 361.
- Lorenz, E. N. (1963) "Deterministic nonperiodic flow." *Journal of the Atmospheric Sciences*, 20: 130.
- Marple Jr., S. L. (1987) *Digital spectral analysis with applications.* New Jersey: Prentice-Hall, Inc.
- Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007) "Recurrence plots for the analysis of complex systems." *Physics Reports*, 438: 237.
- Marwan, N. (2011) "How to avoid potential pitfalls in recurrence plot based data analysis." *International Journal of Bifurcation and Chaos*, 21: 1003.



Mason, P. (1997) “Physiological identification of pontomedullary serotonergic neurons in the rat.” *Journal of Neurophysiology* 77: 1087.

May, R. (1976) “Simple mathematical models with very complicated dynamics.” *Nature*, 261: 459. Messier, M. L., Li, A., and Nattie, E. E. (2004) “Inhibition of medullary raphe serotonergic neurons has age-dependent effects on the CO<sub>2</sub> response in newborn piglets.” *Journal of Applied Physiology* 96: 1909.

Mindlin, G. M. and Gilmore, R. (1992) “Topological analysis and synthesis of chaotic time series,” *Physica D* 58: 229.

Newman, D. (2013) Raw data courtesy of Dr. David Newman, UAF.

Richerson, G. (2004) “Serotonergic neurons as carbon dioxide sensors that maintain pH homeostasis.” *Nature Reviews Neuroscience* 5: 449.

Rosenstein, M. T., Collins, J. J., and De Luca, C. J. (1993) “A practical method for calculating largest Lyapunov exponents from small data sets.” *Physica D: Nonlinear Phenomena*, 65: 117.

Rössler, O. E. (1976) “An Equation for Continuous Chaos.” *Physics Letters* 57A (5): 397.

Schmitz, R. (2001) “Use of chaotic dynamical systems in cryptography.” *Journal of the Franklin Institute*, 338: 429.

Shannon, C. E. (1948) “A Mathematical Theory of Communication.” *Bell System Technical Journal* 27: 379.

Stigler, S. M. (1989) “Francis Galton’s Account of the Invention of Correlation.” *Statistical Science* 4: 73.

Strogatz, S. (2001) *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering*. New York: Perseus books.

Theiler, J. (1986) “Spurious dimension from correlation algorithms applied to limited time-series data.” *Phys. Rev. A* 34: 2427.

Trulla, L. L., Giuliani, A., Zbilut, J. P., and Webber Jr., C. L. (1996) “Recurrence quantification analysis of the logistic equation with transients.” *Physics Letters A* 223: 255.

Webber, C. L., and Zbilut, J. P. (1994) “Dynamical assessment of physiological systems and states using recurrence plot strategies.” *Journal of Applied Physiology* 76: 965.

Zbilut, J. P. & Webber Jr., C. L. (1992) “Embeddings and delays as derived from quantification of recurrence plots.” *Phys. Lett. A* 171: 199.