1 WHAT IS RECURRENCE ANALYSIS?

What is recurrence analysis? In brief, the term recurrence quantification analysis (RQA) covers a range of time series analysis techniques that detect recurring temporal patterns in a single or across multiple time series. Such patterns can be used to reliably quantify time series that are nonstationary and have complex dynamics, such as the relationship between mother-infant vocalization, the evolution of COVID-Infection dynamics, or shared patterns of fluctuations in different stock market indices—all of which we will discuss in this book. Recurrence analysis stands aside other popular and well-known time series analysis techniques because while the latter ones basically assume that values in a time series can be predicted, usually using linear relationships, based on past values of the time series under scrutiny, RQA posits the existence of an underlying dynamical system, no matter how complex or nonlinear, which often (additionally) lives in a latent multidimensional space and can be reconstructed and addressed therein. The analyzed time series are then manifestations of trajectories of the dynamical system in such a phase space—this concept will be better clarified in the next chapter. The goal of the analysis, hence, is not just one of mathematically modeling the time series, for example, in order to predict future values like in standard time series analysis techniques, but to assess the characteristics and quality of the underlying dynamical system, for example, its complexity, its stability and instability, under which conditions it transitions from one stable state to another, and so on etc.

In the remainder of this chapter, we will characterize RQA in greater depth by providing readers with a conceptual introduction of what recurrence is, how it can be used to analyze data—particularly time series—and what its merits and limitations are. In subsequent chapters, we will build on this foundation by introducing specific analysis techniques that solve particular data analysis problems and illustrate them using contemporary research that has employed these analyses. In this respect, due to our personal interests and inclinations, there may be a natural skew toward the behavioral and psychological sciences, which is also the field where RQA has been applied with some success in the last 20 years, after arising in physics and physiology. However, we will try to point to the few relevant applications in the general social sciences field and the promises that it brings therein.

Our goal is that after reading through the book and following the examples, readers will become competent users of recurrence-based analysis techniques and can apply these analyses to their own data. This book also contains the most central equations underlying recurrence analysis; however, all the analysis will be described conceptually in the first place, and a working understanding of the usage of recurrence analysis should be achieved also by those less mathematically inclined. Moreover, readers can access an online repository accompanying this book, which contains the data that we use throughout the book, as well as the *R*-code that was used to analyze and display the data. The repository can be found here: https://osf.io/8ubcj/.

We have used *R*, because *R* is freely available and has up-to-date functions implementing different recurrence analysis techniques—particularly the "crqa" package (Coco & Dale, 2014; Coco et al., 2020). At the end of this book, we will also provide an overview of other platforms and resources that can be used to conduct such analyses, particularly Matlab and Python. At any rate, even if any reader might be more inclined toward one or the other of these software and programming languages, the knowledge gained from the present book is not specific to any of these platforms. Our guidance on parameter estimation, the analysis of sample data, and best practices summarizes the current state of the field and how they apply to recurrence-based analysis and is independent from any particular platform or implementation.

In the three sections of this chapter, we will first introduce the concept of recurrence and the recurrence plot (RP), which lie at the heart of all recurrence-based analyses. Then, we will show how recurrence measures, statistical descriptions of time series data, can be derived from RPs. Finally, we will close the chapter by outlining the main advantages, but also some limitations, of using recurrence analysis.

THE RECURRENCE PLOT

As the name implies, "recurrence" involves repetition of some kind. The advent of RQA began with the seminal papers of Eckmann et al. (1987), who were the first to introduce recurrence plots as a visualization for the dynamics of time series, and (a few years later) of Zbilut and Webber (1992), who were the first to propose definitions for statistical measures that could be derived from such recurrence plots. In the years since, recurrence plots have been developed to analyze multivariate data, and further measures have been proposed to capture other properties of time series (see Marwan et al., 2007).

So, what is a recurrence plot? Basically, a recurrence plot is a binary matrix that results from the cross comparison of the values of a time series. The values of this matrix are coded as "1" when the cross-comparison is counted as a recurrence (i.e., when a value repeats) and is hence represented by a graphical mark in the plot (a point or a filled cell in the matrix), or as "0" if the cross comparison is not counted as a recurrence (i.e., the values compared are different. This gives rise to a blank space in the plot (an empty cell in the matrix). This is easiest to understand when looking at categorical data, such as sequences of letters. Take, for example, the following first few lines of the poem for children, "Popcorn" by Helen H. Moore (from *A POEM A DAY* by Helen H. Moore, copyright © 1997 by Helen H. Moore, used by permission of Scholastic Inc.):

Pop, pop, popcorn,
Popping in the pot!
Pop, pop, popcorn,
Eat it while it's hot!
Pop, pop, popcorn,
Butter on the top!
When I eat popcorn,
I can't stop!

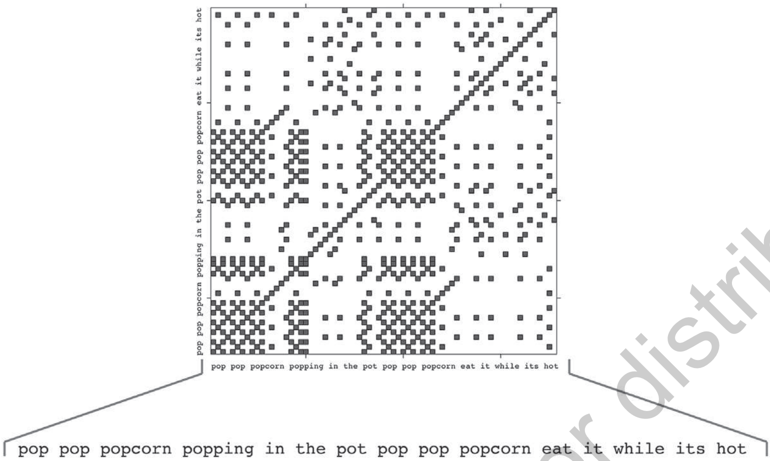
Clearly, poetry typically entails repeating features, such as rhymes and meters, and it is easy to see that several such repetitions are implemented in the above poem, such as the repetition of the word "pop," or of similar syllables at the end of the verses, that is, "corn," "pot," "corn," and "hot' in the first stanza. Let's try to make these repeating patterns visible in a recurrence plot. To keep it simple, Let's consider only the first stanza, ignore punctuation, and treat all letters as lower case:

pop pop popcorn popping in the pot pop pop popcorn eat it while its hot

This sequence of letters is now displayed as a recurrence plot in Figure 1.1.

As described previously, a recurrence plot is a binary matrix, indicating whether a particular value in a sequence is a *recurrence* of another value in the same sequence or not. Please note, that in this case but also more generally, an ordered sequence of values can be loosely seen and interpreted as a

FIGURE 1.1 ■ Recurrence Plot of the Letter Sequence of the First Stanza of the Poem "Popcorn"



Note. Black squares are recurrence points and indicate where letters in the sequence repeat themselves. Note how individual recurrence points appear within patches or diagonal lines of recurrence, indicating the repetition of whole subsequences of letters in the first stanza.

Text from "Popcorn" in *A POEM A DAY* by Helen H. Moore. Copyright © 1997 by Helen H. Moore. Used by permission of Scholastic Inc.

"time-series." In the case of nominal sequences, such as letter strings, recurrence means simple "repetition" of the *values* (i.e., the letters), and this is very easy: A repetition occurs only if an identical element in this sequence is encountered again. So, we indicate the first letter, "p," as being recurrent with the third letter, also "p," but not with any other letter that is not "p." And every time we re-encounter an identical element, we mark this as a recurrence point on the plot—in our case, as black dots filling the cells of the matrix.

Note that this simple univariate or auto-recurrence plot (which is so named because we look at repetitions within the same, single time series, analogously to autocorrelation) has several graphical features. First of all, note the uninterrupted line of recurrence running along the main diagonal, from the lower-left to the upper-right of the plot. This is the so-called line of identity (LOI). The LOI is always present in an auto-recurrence plot, even if the sequence was made of randomly drawn values. Why? Because the line of identity charts all the recurrences at lag0, that is when we compare every value in the sequence with itself, and by definition every value is identical with itself, and hence recurrent. This also means that in the case of a univariate or auto-recurrence plot the line of identity is not informative about the dynamics of a time series or the structure of a sequence, and it should be discarded.

Note next that the recurrence plot is symmetrical around the LOI. That is, the pattern of recurrences in the lower-right triangle of the RP is a mirror image of the pattern of recurrences in the upper-left triangle of the plot, both capturing lagged recurrences in the sequence. They are lagged patterns, because when we consider diagonals to the right and left of the LOI, we are really examining how the shifting of our sequence by a certain number of letters (i.e., by certain lags) changes the pattern of recurrences. For example, the first diagonal to the left or right of the line of identity charts auto-recurrences at lag one, that is, when we move the sequence of the poem's letters by one letter forward or backward. We will later see that there are other kinds of recurrence plots where no line of identity exists and where the patterns of recurrences are not the same across the lower-left and upper-right parts of the plot, for example, in a cross-recurrence plot (CRP), which we will discuss in Chapter 3.

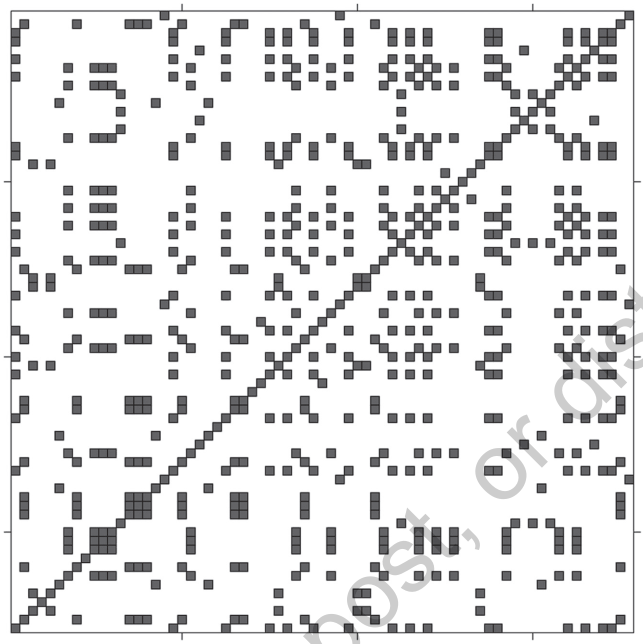
Finally, notice that recurrences are not evenly distributed across the plot. Rather, we see that certain areas of the plot are, in general, more densely populated than others, and that recurrences cluster differently across the plot. This clustering is how the structure of letter sequences of the poem becomes apparent. In other words, the point we want to make here is that the clustering of recurrences in the RP will reflect some qualities of the *dynamics*, that is, how values change and repeat in time of the specific sequence we are analyzing.

To better show this point, compare Figure 1.1 to Figure 1.2, where in the latter you can see a recurrence plot of the same letters of the first stanza when their sequence is shuffled. In general, shuffling sequences will destroy the intrinsic orderliness of the dynamics of the sequence, and this will lead to a statistically homogeneous distribution of recurrence points on the plot and no particular clustering of sorts.

The function of shuffling time-series and sequences of values is hence exactly that of creating a baseline, a random version of the same set of *data* where the original dynamics is lost. The visual differences between Figure 1.1 and Figure 1.2 are easily apparent.

However, we can go a step further. Instead of providing merely a qualitative visualization tool of the dynamics of a time series, recurrence plots can be used to quantify the structure of repetitions of time series and sequences. We will introduce how this quantification works and what features it is based on in the next section.

FIGURE 1.2 ■ Recurrence Plot of the Shuffled Letter Sequence of the First Stanza of the Poem "Popcorn"



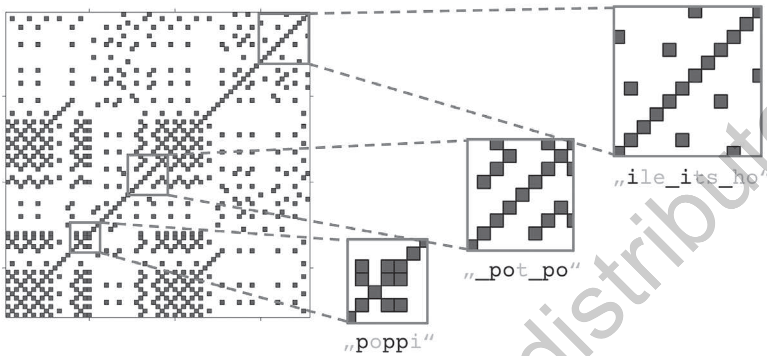
Note. When compared to Figure 1.1, note how individual recurrence points are now much more evenly dispersed across the plot and how the prominence of diagonal lines of recurrences has been reduced by shuffling.

DERIVING RECURRENCE MEASURES

In order to understand how recurrence plots can be used to quantify dynamics of a time series or the structure of a sequence, let's take a closer look at the kinds of patterns of a recurrence plot. Three kinds of patterns are of particular importance for the majority of measures that can be defined for a recurrence plot: isolated recurrence points, diagonally adjacent recurrence points, and horizontally/vertically adjacent recurrence points. Figure 1.3 shows the recurrence plot of the first stanza of "Popcorn," highlighting these three different kinds of recurrence patterns.

Isolated recurrent points indicate that there is a single match between values in a sequence and that this match is not embedded into a larger systematic pattern of the sequence.

FIGURE 1.3 ■ Illustration of Three Important Kinds of Recurrence Patterns, Using the First Stanza of the Poem "Popcorn"



Note. This figure, depicting the recurrence plot for the first stanza of the poem "Popcorn," focuses on isolated recurrence points, diagonally adjacent recurrence points, and vertically/horizontally adjacent recurrence points. While isolated recurrence points occur when only an individual value is repeated (i.e., "i" or "_"), diagonally adjacent recurrence points or diagonal lines occur when a whole subsequence of different values is repeated (i.e., "_po"). Vertically/horizontally adjacent recurrence points or vertical/horizontal lines occur when the same value repeats itself several times one after another (i.e., "pp").

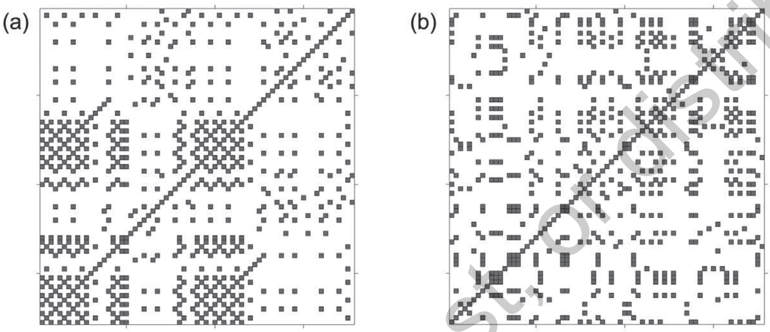
Diagonally adjacent recurrence points show that there is a whole subsequence of values in the sequence that is repeated at later times. For example, the "pop" triplet, which is particularly prominent in the first stanza of the poem, will generate several diagonally adjacent points on the recurrence plot, or smaller diagonal lines.

Finally, horizontally/vertically adjacent recurrence points indicate that the time series or sequence has settled into a stable (or only slowly changing) state. Such a pattern is not very prominent in our example, and it's apparent only in the direct repetition of the double "p" in the fourth word, "popping."

Now these types of clusters can be quantified by several measures that capture different aspects of the dynamics or sequential structure. We would like to introduce only some of these measures here to illustrate how quantification of a recurrence plot works and will return to this matter in the next chapter with fuller and more formal treatment.

The simplest measure is the percentage of recurrence points in a plot or *recurrence rate* (*REC*). It is, simply, the sum of all recurrence points (excluding recurrences that are part of the LOI) divided by the sum of all possible recurrence points, given by the size of the recurrence plot. For instance, in the toy-dataset considered of the popcorn poem, out of 71^2 cells in the recurrence matrix (the poem is 71 characters long) only 610 (681 if we also count recurrences along the LOI) are recurrence points, and hence REC=610/71^2, which gives a proportion of about 0.121 or 12.1% (see also Figure 1.4a)[1, 2]. This measure mainly captures the general level of recurrence in a sequence.

FIGURE 1.4 ■ Comparing the Recurrence Plots of the Original (a) and Shuffled (b) First Stanza of the Poem "Popcorn" on Selected Recurrence Measures



REC = (610 / 71^2) * 100 = 12.1
DET = (253 / 610) * 100 = 41.3
ADL = 253 / 66 = 3.8
LAM = (28 / 610) * 100 = 4.6

REC = (610 / 71^2) * 100 = 12.1
DET = (128 / 610) * 100 = 21.0
ADL = 128 / 62 = 2.1
LAM = (147 / 610) * 100 = 24.1

Note. REC = recurrence rate, *DET* = percent determinism, *ADL* = average diagonal line length, *LAM* = percent laminarity. While *REC* remains constant, shuffling reduces both *DET* and *ADL*, as the sequential ordering of the text is compromised. At the same time, *LAM* increases for the shuffled data. This is because sequences of letters in language rarely exhibit more than two adjacently repeating letters, if any, and the chance of finding such repetitions of letters in a shuffled sequence is actually higher compared to what would be expected from linguistic structure.

A measure that captures the structure of the time series in terms of larger patterns is called *percent determinism*, or simply *determinism* (*DET*), which is the sum of all recurrence points with diagonally adjacent neighbors divided by the sum of all recurrence points. It captures how strongly a time series or sequence is structured in terms of larger diagonally oriented lines, and for the example of our poem, it shows how many of the individual letters recur as substrings—syllables, words, or multiword groups. In our popcorn example, 253 out of the 610 recurrence points in the matrix have diagonally adjacent neighbors and hence DET=253/610, that is, a proportion of 0.415 or 41.5%

(Figure 1.4). A related measure, the *average diagonal line length* (*ADL*), quantifies the average length of such substrings. Since there are 66 of such diagonal line structures in our example, the *average diagonal line length* would be *ADL*=253/66=3.8.

There are also measures based on vertically oriented lines, such as *percent laminarity* (*LAM*), which is defined as the sum of all recurrence points with vertically adjacent neighbors (as well as horizontal ones, given the symmetrical nature of the plot around the LOI) divided by the sum of all recurrence points in the plot. In our example there are only 28 recurrence points with vertically/horizontally adjacent neighbors, and so *LAM*=28/610=0.046 or 4.6%. This measure captures how many of the individual recurrence points cluster in unchanging (or slowly changing) states. As indicated previously, there are only a few of such patterns in our example data, in fact only the repetition of the two adjacent "p"s in "popping," as we would expect from written language.

There are further measures, but, as mentioned at the beginning of this section, we will provide a broader overview of recurrence measures in the next chapter. As this is a book for social scientists, some of these measures may not be readily interpretable from their mathematical definition alone. However, as we progress through this book, introducing different recurrence-based analyses and working our way through different applications, you will become increasingly familiar with them and will learn how to interpret them. Likewise, some of the names of these measures may sound unfamiliar, like *determinism* or *laminarity*. This is because of their origin in physics and dynamic systems theory. *Determinism* originally referred to the deterministic part of an equation relative to the importance of its stochastic component or noise (although it turned out that the value of *determinism* alone is not always a sufficient indicator to judge whether the underlying mechanisms that generated a particular time series are purely deterministic, there have been applications in econometrics using this measure to evaluate the dynamics of asset prices in this regard (see e.g., Aparicio et al., 2010). Similarly, *laminarity* refers to laminar flow, for example in aerodynamics, and loosely refers to the fact that flow—value over time—is relatively stable and unperturbed. However, it is not crucial to know the origins of these terms, although it can be useful, in order to interpret the measures that they refer to in the specific field of application we work in, and you will be fluent in interpreting them by the time you are done reading this book.

Finally, let us close by illustrating the four measures described previously. Take a look at Figure 1.4. Here you will see the original recurrence plot of the first stanza of "Popcorn," as well as the recurrence plot for the shuffled sequence of letters of this stanza. In the following, you'll find the values for *REC, DET, ADL*, and *LAM*[2].

As you can see, shuffling the letter sequence does not change *REC*. Please note that this is a special property of our example data—a sequence of individual letters. In this case, *REC* is dependent only on the distribution of discrete values (letters) in a sequence, and since this distribution is not changed by shuffling (we have the same set of letters), *REC* is not impacted by shuffling. This is different for data at the interval measurement level and data that were *embedded* before performing recurrence quantification, a topic that will be clarified in the next chapter.

In contrast, the shuffled sequence is noticeably lower in terms of *DET* and *ADL*, indicating that the systematic subsequences of which a text, particularly a poetic text, is built, are destroyed by the shuffling procedure. The value of ADL for the original poem shows us the average length of repeating subsequences—based on syllables, words, and word-groups—and that this grouping is largely destroyed by shuffling.

As we have discussed previously, direct repetitions of letters do not occur often in language, so the original poem is higher on *DET* compared to *LAM*, which captures such direct repetitions—in a way representing a kind of stand-still of the letter sequence. Curiously, the shuffling procedure tends to increase *LAM* as compared to the original. This is because the base rate of letters makes it more probable to find—by chance—adjacent identical letters in a random sequence than what we would expect from language, as indicated by the slightly heightened value of *LAM* for the shuffled letter sequence.

We have so far illustrated how the basic core of recurrence-based analyses—recurrence and the recurrence plot—works and how measures can be derived. You now understand the basics of recurrence analysis, upon which the following chapters of this book will build to extend your knowledge toward different applications of RQA. We will close this chapter by summarizing briefly, in a final section, the main advantages and disadvantages of recurrence analysis and provide a brief outlook of what to expect in the remaining chapters of this book.

ADVANTAGES AND LIMITATIONS OF RECURRENCE ANALYSIS

Recurrence-based analyses for time series data have some clear advantages compared to other traditional methods but also a few limitations.

In the first place, we should notice that the single most important characteristic of recurrence analysis when compared to traditional time-series analysis methods at large, is that the method arises in the context of chaos and nonlinear dynamical systems theory. In this sense, the goal and scope of the analysis diverges considerably from the standard linear time series analysis methods. The purpose of time series analysis in its standard or traditional form can be summarized into two general aspects: *understanding* and *modeling* the mechanism which gives rise to an observed series and *predicting* or *forecasting* a future value of a series based on the history of the values in the series. In fact, they can be seen as two faces of the same process. Once we manage to understand and model the system generating the time series, we then have a tool (i.e., the model) that we can use to predict future values of a series. Moreover, this approach relies on the so-called linear paradigm.

This endeavor—which is after all common to all sciences—takes a different accent as we approach nonlinear methods in time series analysis, like recurrence-based ones. The modeling effort is not always very actively pursued and is much reduced here, because if a linear model can only be linear in one way (what needs to be estimated is a set of parameters) nonlinear models constitute an immensely vast class of models, which is often hard to identify when starting from real (observed) data. The goal, hence, lies rather in probing and characterizing the dynamics of the system generating the time series under different conditions, in assessing its stability or instability, in detecting phase transitions in its behavior, and nonlinear time series analyses like the recurrence-based methods treated here are based on phase space reconstruction, not alike the better-known standard time series analysis methods.

Recurrence-based analyses are a class of analysis techniques that are model-free. As such, they make very few assumptions. Particularly, they make no assumptions about the distribution of the underlying data to be analyzed, which is necessary in the case we assume linearity as in standard time-series analysis. They are extremely robust in the face of outliers and extreme values, and they can be applied to time series that are nonstationary or contain strong nonlinearities (Marwan et al., 2007). Moreover, they are very versatile: They can be applied to interval scale and nominal data alike (Dale et al., 2011), which is impossible to do within a standard time series analysis framework. They can be used to analyze or classify the dynamics of individual time series (Webber & Zbilut, 1994), compute correlational and coupling measures between two time series (Shockley et al., 2002), and approach the analysis of multidimensional time series (Wallot et al., 2016b) and leader-follower relationships (Richardson & Dale, 2005)—all within the same coherent framework. All these applications will be presented in detail in this book. Further applications spinning off from the same framework, also include Chromatic/Anisotropic RQA (Cox et al., 2016), Conceptual Recurrence Plots (Angus et al., 2011), Multiplex Recurrence Networks (Eroglu et al., 2018), or Recurrence Network Analysis (Gao et al., 2013), among others, which we will briefly summarize at the end of the book.

The main disadvantages partly originate from the model-free nature of recurrence-based analysis techniques as well. Without a model, these techniques are not very suitable in themselves for predicting future behavior of a time series. To achieve this, one needs either to introduce model assumptions or to use recurrence statistics in conjunction with other inferential statistics models or machine learning techniques. Moreover, the analysis of samples of time series data often requires an exploration of the parameters space for a particular analysis (Wallot & Leonardi, 2018a), which generates additional steps in the analysis and introduces a certain amount of subjectiveness. To the discussion of these topics, we have devoted a whole chapter at the end of this book (Chapter 7). Finally, for most applications, a set of parameters must be estimated prior to analysis, particularly in the treatment of continuously sampled data, a topic to which we turn in the next chapter, and as of now, there is no universal, automated procedure for doing this.

NOTES

1. For simplicity we are not dropping the recurrence points of the LOI in these examples.
2. It should be noted that in the literature different authors have their preferred, slightly different names and symbols for these measures. And so *recurrence rate* is also often named *percent recurrence* and may be indicated by the symbol %REC or simply RR. *Determinism* is somewhere else named *percent determinism* and indicated by %DET, just like *laminarity* and *percent laminarity* (%LAM). In the same fashion, ADL is also indicated by *MeanL*, *MeanLine* or simply *L* in other works.