2 THE BASICS OF RECURRENCE ANALYSIS—UNIVARIATE RQA

Now that recurrence analysis has been conceptually introduced, we want to expand on this by introducing the different steps that are necessary in order to perform the analysis and to interpret its results. Also, we will introduce readers to the basic, univariate Recurrence Quantification Analysis (RQA), which allows us to quantify the dynamics of the temporal patterns of a single time series. Most of these steps—parameter estimation, parameter selection, and interpretation of the RQA outcome measures—are very much the same for all the different variants of recurrence analysis that are introduced in the following chapters. In case one of these variants demands a somewhat different approach to these steps, you will find a section describing these differences with the respective chapters covering the different recurrence-based analyses.

In the rest of this chapter, we will first describe the parameters that must be estimated before we can conduct RQA and provide guidelines for their estimation. As we will see, the necessity for estimating those parameters directly follows from the dynamics of the time-series data that one is interested in analyzing. In Chapter 7, we will discuss how parameter estimation for individual time series differs from parameter estimation for sample data—that is, multiple instances of time series that one is interested in treating as a group. Finally, we will summarize more exhaustively and in detail all RQA measures and their most common interpretations in the context of applications in psychology and the social sciences. Also, because recurrence-based analysis always yields multiple output measures, we will summarize rationales for selecting specific measures in relation to specific research questions, as well as approaches to combine and reduce the number of individual outcome measures down to a manageable number of variables for inferential statistical analysis. This will be elaborated on in Chapter 7 as well.

PARAMETER ESTIMATION

In the description of a recurrence plot presented in Chapter 1, we used a simplified example of letters taken from a poem. Here the poem is treated as an ordered (time) series, and the letters in this case can be thought of as *values* of a categorical variable. The recurrences in this example are repetitions of identical letters along the verse, that is, perfect matchings, or identities, between such values. The example is simple and intuitive.

However, time series are usually somewhat different objects than a series of letters. The time series we deal with often take the form of real-valued, regularly sampled measurements of a phenomenon of interest. The second-by-second change of stock market indices, daily number of infected people during a pandemic, or interactive behavior of small groups measured by motion coding over time are examples of time series arising in the study of certain phenomena or systems in economics, epidemiology, or psychology. In such cases, the variable measured in time is a real number.

When this is the case, we may find out that time series have only very few values that are identical, which may be for example an obvious consequence of measurement noise, and this has a direct impact on what counts as a recurrence and how a recurrence plot is built. First of all, since the values in such series are continuous, the definition of what is recurrent should not be based on perfect matching (after all, depending on the level of precision in decimal places for those values, perfect matching may actually never occur) but rather on some measure of similarity or dissimilarity, of which the most commonly used is typically the Euclidean distance between them.

For example, when comparing two values at two different time points in a series, let's say 126.12 and 125.65, we may decide that the distance in value between them (here $\Delta = 0.47$) is small enough to be counted as a recurrence, while this may not be the case for two values like 125.12 and 50.31, where $\Delta = 74.81$—here the distance seems to be too big. This is done by setting a *threshold* value, above which two values are considered to be too dissimilar or distant to count as recurrent and, hence, are not represented as points in a recurrence plot. The value of the threshold (or radius[1]) then, is one of the first parameters we will need to consider in recurrence analysis when shifting to real-valued, continuous time series.

Furthermore, another fundamental issue is to establish the dimensionality of the system that generates the observed series and, hence, its phase space reconstruction (in the literature the term *phase space* is also sometimes referred to as state space), a procedure that is at the foundation of nonlinear

time series analysis. This has to do with the number of factors—and their kind of interaction—that drive the system dynamics. In other words, within a dynamical view, such time series are seen as the manifestation and reflect the coupled behavior of many latent influencing factors such as monetary policy and demand for a product, infectiousness of a virus and containment measures, the task that participants must achieve and their skill at task performance, and so on.

In simple terms, when analyzing a single observed time series, we can think of it as a single dimension (variable) of a more complex dynamical system that generates it. Several variables or factors nonlinearly related between them may be at the basis of the phenomenon of study, although we may actually be able to measure or have access to only one of them. However, the oscillations in time, that is, the dynamic behavior in that single dimension, encapsulate the dynamics of all the other dimensions of the system, as they are obviously intrinsically connected and influencing each other.

Let us consider the Lorenz system (Lorenz, 1963). This example system does not originate in the social sciences but has been widely used as a paradigmatic example of complex nonlinear dynamical systems and often appears in other texts illustrating, for example, how recurrence analysis works. In particular, it is a good, transparent example of how this method deals with the problem of dimensionality of systems dynamics. The system has its origin in weather forecasting and is based on the physics of fluid dynamics. Admittedly, such kinds of equations (see eq. 2.1) are hardly found in modeling phenomena from psychology, sociology, or economics. However, the system reflects general properties of dynamical systems and has become a model system that can be used to conceptualize and derive predictions on what to expect from observable phenomena within the social sciences, when we try to understand them within the framework of dynamical complex systems (Favela, 2020). For example, the Lorenz system and similar dynamic systems have been used to generate predictions of phase-transitions when humans gain new insight with regard to the solution of mathematical problems (Stephen et al., 2009), when readers adjust their cognitive processes while switching from one reading task to another (Wallot et al., 2019), or when capturing processes of change in human behavior, such as psychotherapeutic processes (Schiepek & Strunk, 2010).

The Lorenz system is a nonlinear dynamical system of three coupled ordinary differential equations (eq. 2.1), which has been proposed as a simplified model of atmospheric convection processes in a fluid layer uniformly warmed from below and cooled from above, and it models the rate of change

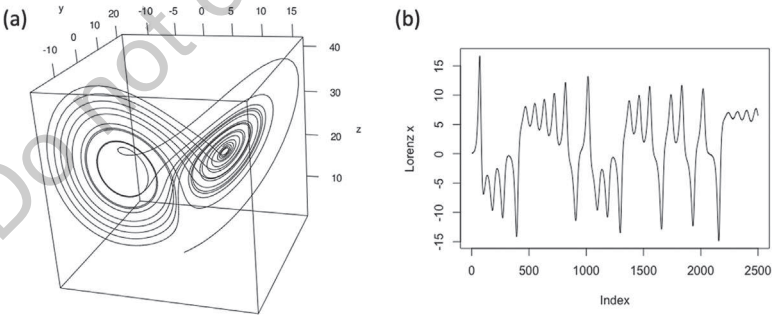
in time of three quantities: x is proportional to the rate of convection, y to the horizontal temperature variation, and z to the vertical temperature variation:

$$\begin{cases} \dot{x} = \sigma y - \sigma x \\ \dot{y} = -xz + rx - y \\ \dot{z} = xy - bz \end{cases} \quad \text{(eq. 2.1)}$$

As can be seen from eq. 2.1, the three differential equations constitute a system since the time derivative functions of each variable depend on values of the other variables. Moreover, the system contains two nonlinear terms (i.e., the dynamics are nonlinear) in xy and $-xz$, while σ (sigma), r, and b are fixed parameters of the system.

The dynamics (i.e., time-dependent trajectories) of the Lorenz system span a three-dimensional state space (Figure 2.1a), since the system is defined by three variables, but as we have seen previously, the time dependent values on each dimension are nonlinearly dependent on the values of the other two dimensions. Now, suppose that we have access to the values of one of these dimensions only—and this is the time series we observe. Obviously, the trajectory or time-dependent representation of this single variable or dimension (Figure 2.1b) looks different than the three-dimensional dynamics of the Lorenz system as a whole (compare Figure 2.1a and Figure 2.1b). However, remember this trajectory is also determined or driven by the other variables, to which in some situations we may not have access.

FIGURE 2.1 ■ Phase Space Portrait of the Lorenz System



Note. Part (a) represents the actual attractor of the Lorenz system as derived from the integration of eq. 2.1 and plotted as a trajectory in the 3-dimensional space defined by the three variables *x*, *y*, and *z*. Part (b) represents the dynamics of the *x*-dimension of the Lorenz system over time, that is, a one-dimensional trajectory.

One fundamental and remarkable result of nonlinear dynamic systems theory (Takens, 1981) is that, in principle, it is possible to reconstruct the full dynamics of a complicated nonlinear system based on one single time series. In other words, we can reconstruct the trajectory of the dynamic behavior of the system in more than one dimension even if we only measured one of them. To extract the "hidden" or unobserved dimensions of the system, from the one we actually observed, we make use of a technique formalized by Takens in 1980: the delay-coordinate embedding. The technique uses past values of a single scalar measurement y from a dynamical system to form a vector that defines a point in a new space of higher dimension. Specifically, after defining a time delay τ, we can construct m-dimensional *reconstruction-space* vectors $\vec{R}_{(t)}$ from m τ-delayed samples of the measurement $y_{(t)}$ (eq. 2.2), such that at every time point t

$$\vec{R}_{(t)} = \left[y_{(t)}, y_{(t-\tau)}, y_{(t-2\tau)}, \ldots, y_{(t-(m-1)\tau)}\right] \qquad \text{(eq. 2.2)}^2$$

The higher dimensional reconstructed phase space is not identical to the real internal dynamics of the system, but even so it has been demonstrated to be isomorphic (i.e., has the same structure) and is guaranteed to be topologically identical to it (i.e., corresponding points in the original and the reconstructed phase space have the same neighbors), which means that many important properties of those dynamics, being invariant under diffeomorphism, can be usefully derived from the reconstructed ones. This concept is called the *embedding* of the (one-dimensional) time series in an m-dimensional reconstructed phase space (Kennel et al., 1992; Sauer et al., 1991).

What we are saying here is that when we observe a time series in our field of study, we should wonder if it is not the expression of more complex, multidimensional systems. In which case the dynamical trajectory of the original system in multidimensional space should be derived first, rather than focusing our analysis on the unidimensional fluctuations we observed (which is, incidentally, the standard approach of linear time series analysis methods).

In other words, when running recurrence analysis, we can and want to analyze the dynamics of the whole system and not just of the single unidimensional time series we observed, which we can do based on the previously mentioned Takens theorem and method. The recurrence plot will then be built not in relation to such single time series but rather in relation to the m-dimensional dynamic trajectory in the reconstructed, embedded phase space. A recurrence will only occur when two points from such trajectory

defined in the m-dimensional reconstructed phase space are close enough, that is, within the threshold value.

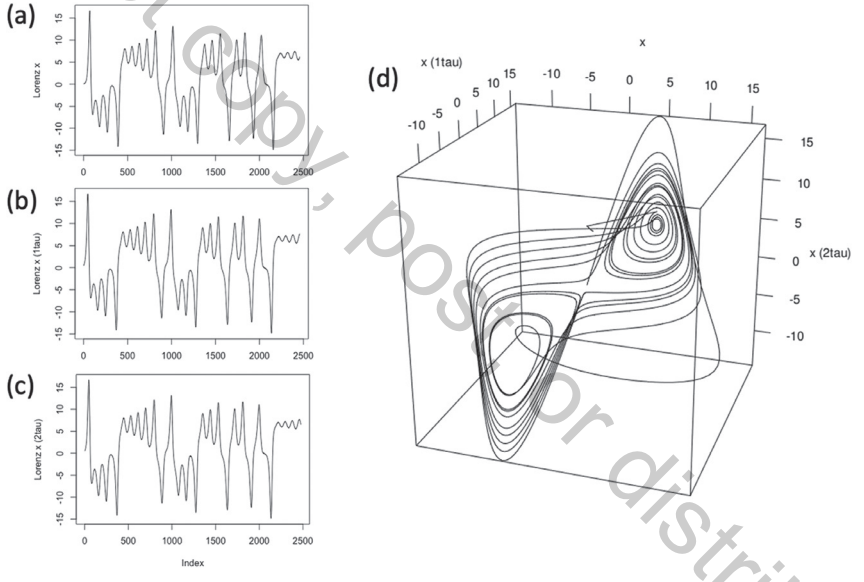
In Figure 2.2 we illustrate how the Takens technique works. Starting from one single dimension of the Lorenz system (Figure 2.2.a, that is, the same as in Figure 2.1.b), we create two additional dimensions (Figure 2.2.b and 2.2.c) that are time-shifted versions by a delay τ (Figure 2.2.b) and $2 * \tau$ (Figure 2.2.c) of the original dimension (Figure 2.2.a). These three dimensions are the ones used as the coordinates of the reconstructed 3-D trajectory of the system (Figure 2.2.d). As we can see, the delay-coordinate embedding serves to reconstruct a dynamic trajectory in three dimensions which is homeomorphic to the original dynamics of the system (Figure 2.1.a).

Given the previous discussion, it becomes clear that to perform any kind of the recurrence-based analyses described in the remainder of the book, three parameters will have to be estimated at first: the **delay parameter** τ, the **embedding (dimension) parameter** m, and the **radius (or threshold)** parameter ε. Under certain circumstances, the values for these parameters can be set without actual estimation (which we will discuss at the end of this section), but nonetheless every statistical software requires values for these parameters to run univariate RQA and its extensions.

The delay parameter τ and the embedding parameter m are jointly necessary to properly unfold the multidimensional dynamics of a time series, which result from (nonlinear) interactions between its constituent components (Takens, 1980). If the value of the embedding parameter $m > 1$, this implies that the latent dynamics from which the unidimensional time series of interest was sampled are more complex than what is apparent in the observed time series, specifically that they have higher dimensionality (Kennel et al., 1992). Figure 2.3 shows a recurrence plot (RP) of the phase spaces presented in Figures 2.1a and 2.2d. As we can see, the RP of the original, three-dimensional data (Figure 2.3a) is quite similar to the RP obtained from the embedded, single dimension of (Figure 2.3b).

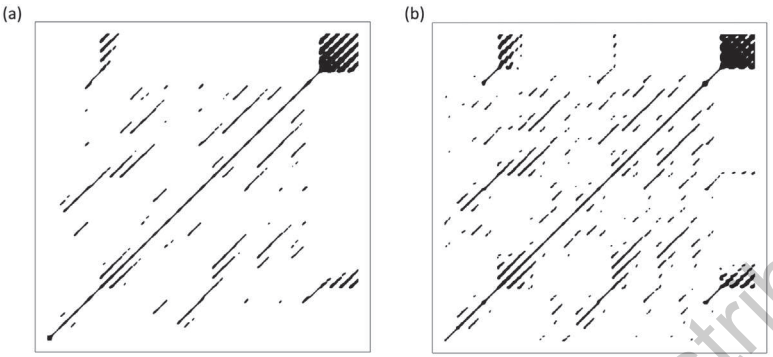
The delay and embedding parameter allow us to reconstruct the higher dimensional dynamics—the phase space of the actual dynamics from which our one-dimensional time series originated. Note, however, that in empirical time series noise can also influence the estimation of the embedding parameters and the RQA outcome measures (Hasson et al., 2008; Thiel et al., 2002). After the embedding in the new m-dimensional phase space of the nonlinear system dynamics, we are interested in computing a recurrence plot in a similar fashion as we saw in the previous chapter. To that end, we now need to set the radius (threshold) parameter. The threshold parameter ε differs from the

FIGURE 2.2 ■ An Illustration Phase Space Reconstruction by Delay Embedding Applied to the Lorenz System



Note. On the left: (a) the *x*-dimension of the original Lorenz system, and (b) and (c) the tau-delayed versions of (a). On the right: (d) a reconstructed three-dimensional version of the Lorenz system, whose coordinates are given by (a), (b), and (c). We can notice the similarity (isomorphism) of the original Lorenz attractor (Figure 2.1.a) and the one reconstructed from the *x*-dimension only.

FIGURE 2.3 ■ Recurrence Plots of the Lorenz System



Note. Depicted here: (a) a recurrence plot of the original phase space of the Lorenz system (Figure 2.1a), and (b) a recurrence plot of the reconstructed phase space (Figure 2.2d).

previous parameters in that it does not affect the phase space but rather how that phase space is translated into recurrence points on the recurrence plot: It provides the size of a tolerance range, within which one point on the trajectory, which is close but not identical to another point on the trajectory, is counted as a recurrence in relation to it. If ε is too low, the recurrence plot will be empty (no recurrence points apart from the LOI) or too sparsely populated. If ε is too high, too many or even all the theoretically possible recurrence points on the RP will be marked, and the dynamics of the time series will not be discernible and quantifiable in terms of recurrence measures (see section on the radius parameter).

Now, let us turn to the process of estimating these parameters. There are different options to do so (Marwan et al., 2007), but we will focus on the most widely used procedure of estimating the delay parameter via the average mutual information function (AMI; Abarbanel, 1996; Fraser & Swinney, 1986; Kantz & Schreiber, 2004) and the embedding parameter via the false-nearest-neighbor function (FNN; Kantz & Schreiber, 2004; Kennel et al., 1992). The mathematical details of these two approaches are described in the associated citations.

THE DELAY PARAMETER τ

Parameter estimation for recurrence-based analyses always starts with the delay parameter τ. Actually, you could think that the embedding parameter (number of dimensions) should be estimated first. After all, if $m = 1$, that would mean that the time series would not be embedded at all, and the delay

parameter would not be applied or needed, so why estimate it? However, as we will see in the next section, in order to obtain a good estimate for m, τ has to be used—hence the need to estimate it first.

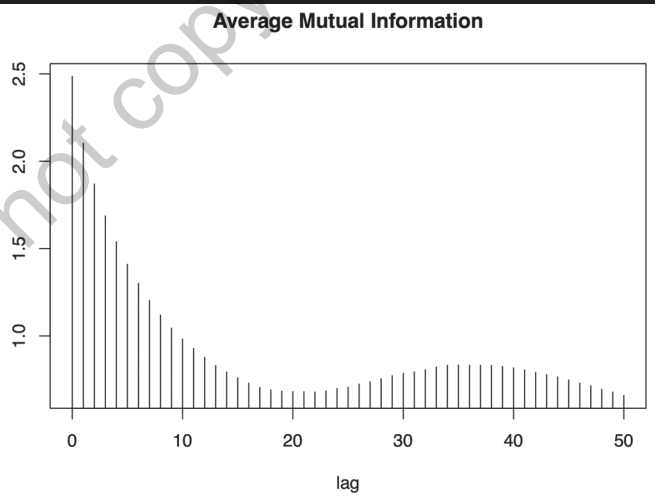
To estimate τ, we use the AMI (Fraser & Swinney, 1986), which is a nonlinear generalization of the autocorrelation function of a time series (ACF). Both mutual information and correlation measure the amount of dependence of two random variables (here the same time series evaluated with itself at different off-set lags), but correlation measures it under the assumption of linearity which we want to avoid. In particular, just as said, we want to calculate values for AMI at multiple lags τ within the time series (auto mutual information) and plot those values as a function of the lag, in the same fashion as in ACF we would calculate and plot the correlation coefficient at multiple lags. The AMI-function informs us about the amount of information shared (dependency or reduction of uncertainty measured in bits) with future values of the time series $x_{(t)}$ as a function of the distance between them, in terms of lags. This means that if the amount of information shared between points at a certain lag L is low, they are somewhat independent from each other, or in other words, by knowing the values of the time series $x_{(t)}$ we can say little about the values at $x_{(t+L)}$, for any t and L.

How does the AMI-function help us to determine the value of the delay parameter τ? When using the method of time delayed embedding, we want to reconstruct the contributions of other, unobserved variables to the potentially multidimensional dynamics of the system from which $x_{(t)}$ was measured. As explained previously, we do this by using time-shifted copies of $x_{(t)}$, which are now surrogates for the other, unobserved dimensions of the original dynamics. In order for these surrogate series to provide new insight about the original, multidimensional dynamics, they should be as independent as possible from the original series. That is, they should contribute as much new, additional information as possible, and our criterion for defining "additional information" is the degree of independence or information shared as given by the AMI value for a particular lag—the lower the AMI, the more independent the shifted time series and the more additional information a surrogate series contributes.

By exploring the AMI function, we should then pick one of the lag-values where AMI is very low or lowest. This implies that time series shifted by that lag-value will be relatively independent from each other and hence each new dimension based on those time-shifted series will contribute new details to the general dynamics of the system or, in other words, this would *unfold* in a higher dimensional space the dynamics of the system under scrutiny.

Figure 2.4 displays one example of the output of the AMI-function. We have arbitrarily chosen to investigate the first 50 lags of the time series; accordingly, we can see the AMI-function over 50 lags. The question arises now, which value to choose? There have been several suggestions: One is to take the first value that drops below a predefined threshold, like $1/e$ (i.e., the first value of the AMI function that is below the standard error of the function) (see Wallot & Mønster, 2018). In our view, this is somewhat problematic, because for empirical data from the social sciences, other threshold values might be just as reasonable—depending on the data properties of a particular time series—but we usually lack a prior theoretical understanding of the data in order to predefine such thresholds. Hence, after Abarbanel (1996), we rather recommend the approach of taking the first local minimum of the AMI-function, which in the previous example leads us to choose $\tau = 21$ as our estimate for the delay parameter. We want the first local minimum also because it saves us data points. Remember, every surrogate series is obtained by shifting the original time series $x_{(t)}$ by τ data points (multiplied by the number of embedding dimensions). That means that we are chopping off that amount of data points from the original time series—data points that are lost for analysis. Hence, choosing the first local minimum, instead of, say the global minimum of the AMI-function, is parsimonious in that sense.

FIGURE 2.4 ■ Average Mutual Information Function of the Time X-Dimension of the Lorenz System



Note. As can be seen, the function has its first local minimum at a value of lag 21, so we chose the value for the delay parameter $\tau = 21$.

However, with empirical data, you will often encounter cases with no clear local minimum. Here you will need to keep several things in mind: First, an odd AMI-function could indicate problems with the data (missing data, outlier, or other problems). If that is not the case, you can ask yourself, for example, how much data would I lose choosing the first as opposed to the second local minimum, and is the additional data worth the drop in AMI? If you find that your AMI-function is bottoming out, you can look for an "elbow" like the scree-plot in factor analysis (Cattell, 1966). That is, you take the value after which no substantial drops in AMI appear anymore.

However, since all these decisions have a subjective component and are uncertain, it is often good practice to use several plausible parameter combinations and check whether and how different parameters influence the results, that is, the actual values of the RQA measures extracted. If those values do not change drastically between the different choices of the parameters' values we are considering, then this means their influence on the results is not critical and will not change the interpretation of the analysis, and we can safely pick up one of them. We will return to this topic in Chapter 7, when talking about exploring the parameter space.

THE EMBEDDING PARAMETER *m*

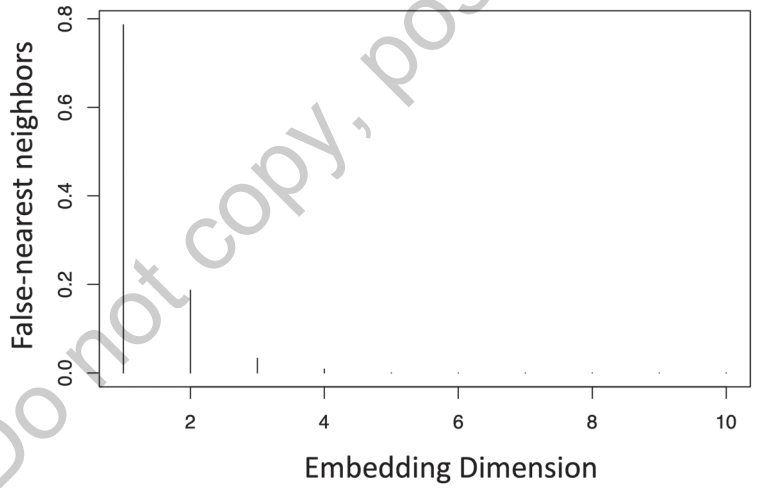
After having obtained a value for the delay parameter, we want to determine the embedding dimension parameter m. We do this by using the FNN. The logic of this procedure is rather simple and intuitive and goes as follows: An unembedded time series can be conceived as a projection in one single dimension of a multidimensional trajectory, which means that some of the points in such time series are close in value to other points therein just as a result of the lack of higher-order dimensionality of the systems from which the time series was observed. They are, in other words, false neighbors of those points, and as soon as we unfold the dynamical trajectory in a higher dimensional embedded phase space they will shift to a different location. If we track the number of false neighbors as we increase in a stepwise manner the number of embedding dimensions of the time series, we will observe that after a given number of embedding dimensions, the number of false neighbors will go to zero.

Practically then, determining m is in many ways similar to determining τ. We will create a plot showing the function of the percentage of FNN for multiple values of dimensions m and want to use the shape of that function to help

us make a decision about which value to choose for the embedding parameter. The false-nearest-neighbor function hence provides an estimate of how distances between neighboring coordinates in phase space change as a function of embedding dimension. When these changes are small, we assume to have found a stable, appropriate dimensionality for our data (Kennel et al., 1992).

It appears now clearly that, in order to explore a number of embedding dimensions and comparing the number of false-nearest neighbors between them, we will need to use, in the FNN recursive computational procedure, the value for the delay τ we already estimated in the previous step. Just as with the maximum number of lags that one wants to investigate using the AMI-function, we will also need to provide a value for the maximum m (i.e., the highest embedding dimension for which we want to compute the function). Figure 2.5 shows the curve of the FNN-function for time series from the Lorenz system, using the input value of 10. As can be seen, the function drops off initially and then bottoms out. Similar to the criteria applied

FIGURE 2.5 ■ False-Nearest-Neighbor Function of the Time X-Dimension of the Lorenz System



Note. As can be seen, the function seems to bottom out at a value of 4. However, an embedding dimension of 3 does not yield an appreciably higher number of false-nearest neighbors, so we choose $m = 3$—also knowing that the true dimensionality of the Lorenz system is 3 (see. eq. 2.1). However, the figure might also be suggestive of an embedding dimension of 4 or 5, if we did not know the true dimensionality of the system. Then, one can try out both parameters and compare the results—a procedure we describe in Chapter 7 under parameter exploration.

to the AMI-function when selecting a value for the delay parameter, we are interested in finding the first value after which the function does not change appreciably anymore.

However, just as with the AMI function, empirical data sometimes give rise to more complicated curves in the FNN-function. Some argue that, if possible, one should select an embedding dimension of the percentage of false-nearest neighbors close to 0. However, there are practical problems with such suggestions. First of all, higher embeddings reduce the number of data points of analysis, and if the number of data points lost to embedding is too high, one needs to make due with lower embedding dimensions. Moreover, empirical data usually have one or more noise components, which can inflate embedding dimension estimates. Accordingly, embedding dimensions would be much too high if one aims at 0% of false-nearest neighbors.

In general, the methodology works best on stable, low-noise systems, and when we deal with real-world data, noise is ubiquitous and can inflate the dimension estimation. It should be noted though, in case of doubts, that it does not hurt to overembed, that is, to select a value for the embedding parameter that is slightly bigger than the first minimum or point of stability (Webber & Zbilut, 2005). Overembedding will unfold the dynamic trajectory in higher dimensions but will not recruit new information. Actually, Takens's theorem (Takens, 1980) guarantees, as a sufficient condition, the preservation of the topological structures of the original trajectory for values of $m > 2d + 1$ where d is the real dimensionality of the system. Clearly, however, increasing the embedding dimension does come at the cost of data points that are used for embedding and, hence, lost for analysis. Moreover, choosing an embedding dimension that is too high can eventually lead even random/stochastic systems to display strong, yet artifactual patterns of recurrence (Marwan et al., 2007, p. 251).

THE RADIUS PARAMETER ε

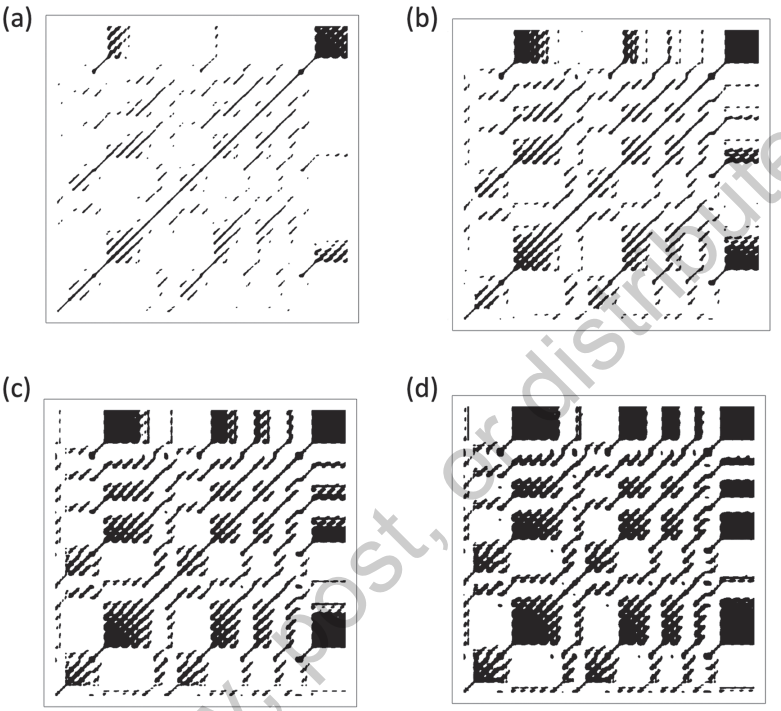
Finally, we need to select a radius parameter (in the literature also often referred to as threshold parameter). The radius parameter (or threshold) provides the width of the tolerance range within which similar data points—or phase space-coordinates—are counted as recurrent. As we increase the value of the radius parameter ε, different data points in the time series with increasingly different numerical values (coordinates in the embedded phase space) are treated as recurrent. The problem is that

the choice of this recurrence threshold may be strongly dependent on subjective judgment from the researcher's side. There are several papers discussing methods for selecting the appropriate value of the radius (e.g., Schinkel et al., 2008), but a systematic study offering universal guidelines in this respect is still to come (Marwan & Webber, 2015). In certain cases, such as when one knows something about the magnitude of the noise component compared to the dynamics of interest of the time series, or when one knows something about the scale on which interesting effects reside, this might provide some limited guidance for setting a value for the radius parameter. Again, however, such knowledge is usually absent in social science applications.

The good news is, however, that the final results of recurrence quantification analysis are relatively scale-invariant and robust across different values of the radius parameter (see the discussion on parameter exploration in Chapter 7). Generally speaking, recurrence analyses are most sensitive when the radius parameter is set to a low, but not too low, value, so that we end up having enough recurrence points to map out the dynamics of a time series but not so many recurrence points that these dynamics drown in mainly meaningless recurrence (Webber & Zbilut, 2005). That is why one first rule of thumb for the choice of the radius is to pick a value based on the maximum distance of the points in the phase space, in terms of a small percentage (less than 10%) of this phase space diameter (Zbilut & Webber, 1992). In other works, some authors recommend selecting a radius value, so that the resulting percentage of recurrence points (REC) lies between 1% and 5% (Webber & Zbilut, 2005), a range that can be even lower for time series with little noise and strong deterministic components. In the case of very noisy or noise-type time series, the resulting REC can also be higher—between 5% and 20% (Wallot, 2017). Figure 2.6 illustrates this, using the x-dimension of the Lorenz system with increasing values for the radius parameter ε. In general, we do not want to count things as recurrent that are increasingly different. Accordingly, the default advice is to aim for a selection of the radius generating lower recurrence rates in the RP, as far as the dynamics of interest are captured in a sufficient clear way on it. Because the Lorenz system is deterministic and here is presented without noise, we would be inclined to choose the sparsest plot, shown in panel 2.6a.

So, in relation to the above suggestions, if one wants to determine which radius to use, one can produce a recurrence rate function by calculating the percentage recurrence (*REC*) for different values of the radius parameter targeting the desired amount or range of REC (see also Chapter 7).

FIGURE 2.6 ■ Recurrence Plots of the Reconstructed Phase Space in Figure 2.2d



Note. As can been seen, with increasing radius parameter ε, more trajectories in phase space are counted as being recurrent, and the number of recurrent structures increases from (a) to (d) (with $m = 3$, $\tau = 21$, of ε being increased from 2, to 4, 6, and 8).

FURTHER PARAMETERS

In the previous discussion, we presented the most important and crucial parameters that have to be determined and set before we run any of the recurrence-based methods. Any of the statistical software solutions[3] computing recurrence analysis will require the user to enter appropriately chosen embedding dimension (m), delay (τ), and radius (ε). But there are additional parameters whose setting, usually, is requested as well. However secondary, they are also relevant for the computation of the analysis and can have an influence on the results obtained. It is useful for a social scientist approaching recurrence quantification analysis for the first time to be at least acquainted with them, and, hence, we will shortly mention those additional parameters and quickly explain their meaning in the present section.

The minimal length of diagonal and vertical lines (sometimes called the parameter LINE) just defines how many adjacent recurrence points in the RP will be needed to count this pattern as a line. We already mentioned that one of the key patterns in RP is diagonal and vertical/horizontal lines, and several recurrence measures are based on some kind of quantification of the points forming such lines. For this reason, it may well be useful to clearly define what counts as a line. The absolute minimum number of adjacent points to be counted as a line is 2, and so this is also the default value typically used in statistical software. Some of them, though, give the user the possibility to customize and make the definition of line more stringent, by increasing the value of this parameter. Notice that such a choice will as a consequence affect recurrence quantification measures like *DET, ADL, MDL*, and others.

Another option users are given in some of the software packages for recurrence analysis is the choice of a distance norm. When computing distances between points (vectors of coordinates) in the phase space of the reconstructed dynamics, the standard choice would be the use of the Euclidean distance norm, that is, the computation of the straight line connecting the two points in the *m*-dimensional Cartesian coordinates of the reconstructed phase space (L2-norm). And actually, this is often the only or the default option given by programs. In theory though, we may conceive the application of other distance norms between points, like minimum or maximum norms (L1-norm and Linf-norm; see Marwan et al., 2007; Webber and Zbilut, 2005), which change the shape of the neighborhood around a point in phase space. Some authors also conceive the application of a so-called order norm, that is, the computation of recurrence plots based on distances between vectors of coordinates in ranked form (Webber, 1996-2021).

Almost all the software packages also give the option of setting a rescale parameter. The rescale parameter acts on the matrix of distances between all the points of the reconstructed dynamical trajectory embedded in the chosen phase space, and may affect the resulting recurrence plot. This parameter typically envisions a few options like rescaling to the mean or maximum distance. Rescaling to mean distance can be taken into consideration for smoothing, in the presence of outlier distances; otherwise, usually, rescaling to maximum distance is applied. This will rescale the matrix to the unit interval (0% to 100%) and will allow the radius parameter to be expressed in a percentage of the maximum distance in the data.

One last option recurrence analysis usually allows users to select is normalization. Normalization involves the choice of whether to standardize the time series before running the analysis. This choice can be crucial and should be

taken into account especially in case we are analyzing two (or more) time series at a time as in cross-recurrence quantification analysis (CRQA) or multidimensional recurrence quantification analysis (MdRQA) (see Chapters 3 and 6).

The question of normalization and rescaling is also related to the choice of the radius parameter. Because recurrence-based analyses are analyses of sequential properties of a time series, and especially with empirical data, time series from different participants or sources can also differ in level or magnitude of variance. These differences in turn also affect how recurrences are defined. However, usually—but not necessarily always—we are interested in disentangling these two aspects, and normalization is a way to do so: If two time series are normalized but differ in terms of recurrence measures, then this is most likely attributable to their dynamics, a difference in the sequential ordering of their values (Shockley, 2005). For further readings on the description of these parameters please see, for example, Webber and Zbilut (2005) and Marwan et al. (2007).

SUMMARIZING RQA OUTPUTS

As anticipated in the previous chapter, recurrence-based analyses provide a wide range of outcome variables that quantify different aspects of the individual or shared dynamics of time series, based on the pattern of recurrence points appearing in the recurrence plot of the series. Table 2.1 describes the most common recurrence measures. The most basic measure is percentage recurrence (*REC*), which we have already introduced and is simply the sum of all recurrence points divided by the size of the recurrence matrix. While *REC* captures how many individual recurrences occur on a recurrence plot, all of the other measures quantify different aspects of clustering of recurrences related to longer lasting temporal structures in a time series.

Further recurrence measures have been proposed to capture different aspects of the dynamics of a time series related to the clustering and connectivity of its values and its complexity (Marwan et al., 2007). Table 2.1 summarizes the most widely used measures.

Just as explained in the rightmost column of Table 2.1, each of the metrics computed based on the RP are meaningful in the context of the dynamical behavior (the trajectory) of the system measured, and within this context they need to be discussed and interpreted. So, for example, DET is a metric that basically gauges periodicity of the behavior/dynamics of the system, as it computes the percentage of recurrences occurring in diagonal structures—strings of data points with similar values and repeating in the same sequence

TABLE 2.1 ■ Summary of the Most Common Recurrence Measures

Variable Name	Definition	Quantifies . . .
Percentage Recurrence (*REC*)	Sum of recurrent points in RP /Size of RP $\frac{1}{N^2}\sum_{i,j=1}^{N} R_{ij}$	repetitiveness of the *values* across the time series. It is a loose indicator of the generalized level of (auto)-correlation of the series.
Percentage Determinism (*DET*)	Sum of diagonally adjacent recurrent points / Sum of recurrence points in RP $\sum_{l=l_{min}}^{N} lP(l) / \sum_{l=1}^{N} lP(l)$	how many of the individual repetitions co-occur in connected trajectories. It is an important index of the fact that dynamic trajectories of the system tend to regularly visit in approximately the same order the same places in a phase space.
Average Diagonal Line Length (*ADL*)	Average length of the diagonal lines in RP $\sum_{l=l_{min}}^{N} lP(l) / \sum_{l=l_{min}}^{N} P(l)$	how long the average repeating trajectory is. It helps distinguish between highly periodic deterministic systems, whose trajectories tend to have long streaks of recurrences, and deterministic but chaotic systems or nondeterministic systems.
Maximum Diagonal Line Length (*MDL*)	Length of longest diagonal line in RP (excluding the main diagonal) $\max\left(\{l_i\}_{i=1}^{N_l}\right)$, $N_l = \sum_{l \geq l_{min}} P(l)$	how long the longest repeating trajectory is. It is an indicator of the chaoticity of the system. The shorter the MDL the less stable'(i.e., more chaotic) the signal is. It was initially connected to one important invariant of dynamical systems, namely the most positive Lyapunov exponent.
Diagonal Line Entropy (ENTR)	Shannon Entropy of the distribution of diagonal line lengths over integer bins. $-\sum_{l=l_{min}}^{N} p(l) \log p(l)$	how complex are the dynamics of the time series. Entropy is a fundamental concept in information theory and can be loosely defined as the amount of uncertainty residing in the signal considered as an information channel. The higher the entropy, the more complex the signal.

Variable Name	Definition	Quantifies . . .
Percentage Laminarity (LAM)	Sum of vertically/horizontally adjacent recurrence points/Sum of recurrence points in RP $\sum_{v=v_{min}}^{N} vP(v)/\sum_{v=1}^{N} vP(v)$	the occurrence/amount of laminar states in the system. It indicates if the system's trajectories (values in phase space) do not change or change very little at subsequent time sampling points, that is, the system *lingers* around certain states.
Trapping Time (TT)	Average length of the vertical/horizontal lines in RP $\sum_{v=v_{min}}^{N} vP(v)/\sum_{v=v_{min}}^{N} P(v)$	how long the lines indicating laminar states are in average. This measure gives an indication of the time the system spends lingering around certain states.
Maximum Vertical Line Length (MVL)	Length of the longest vertical/horizontal line in RP	the maximal length of laminar states in the system. It corresponds to the MDL in the case of vertical/horizontal lines in the RP.
Categorical Area-Based Entropy (catH)	$-\sum_{a>1}^{N_a} p(a) \log p(a)$	complexity of the distribution of size of dynamical states.

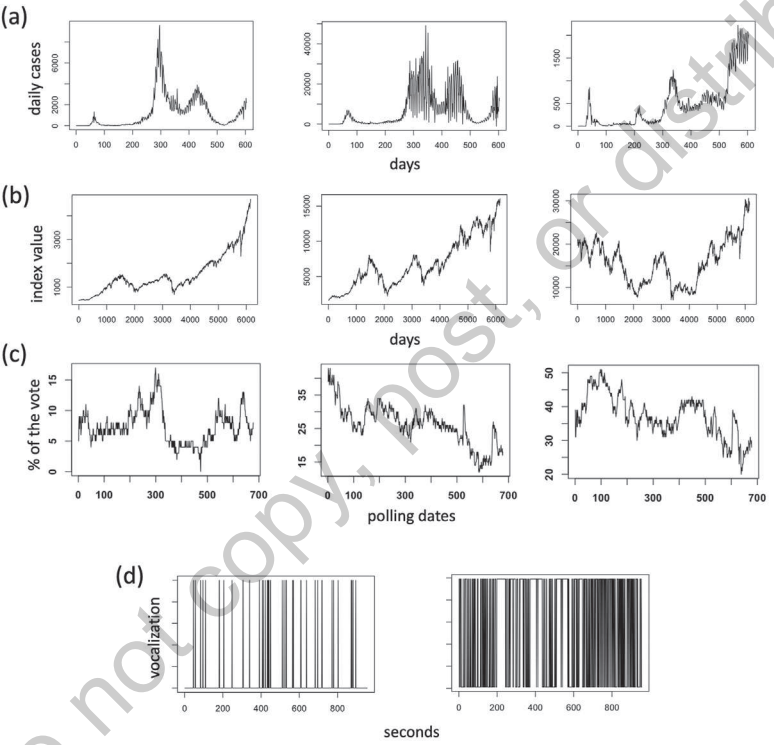
Note. Further RQA measures exist, and others are being developed—for the description of additional measures, see for example Marwan et al. (2007).

after some time. In different experimental or research fields such a periodicity (or its lack) may point to the existence (or not) of a causal process generating (i.e., determining) behavioral cycles whose specific meaning would need to be contextualized according to such field and the research hypotheses put forward. For example, changes in DET may be related to and reflect changes in experimental conditions, as increased cognitive load makes postural sway more regular (deterministic), in order to cope with limited attentional resources (see Pellecchia & Shockley, 2005). In a similar fashion, the LAM metric gauges the general tendency of the system to slow down and dwell around some state in space phase. This metric has allowed and has been used to identify transitions in EEG patterns in two experimental conditions in a behavioral task (probability of appearance of a target sound in an oddball reaction-times paradigm; Marwan & Meinke, 2004).

Finally, let us take a look at some example data that we will use across the book to start to build some intuition about what these measures tell us about

the dynamics of a time series. Figure 2.7 shows example time series that we are going to use across the remaining chapters of the book when discussing further recurrence analysis techniques and illustrating how they can be used to address questions regarding the dynamics of time series data. We will see in the next couple of chapters that different recurrence methods lend themselves to answer different questions, for example, Cross-Recurrence Quantification

FIGURE 2.7 ■ Examples of Empirical Time Series Used in This Book



Note. (a) Time series of daily cases of SarsCov-19 for Austria, Germany, and South Korea (from left to right) from 01/22/2020 to 09/16/2021. Data Source: European Centre for Disease Prevention and Control. (n.d.). *Data on testing for COVID-19 by week and country*. https://www.ecdc.europa.eu/en/publications-data/covid-19-testing; (b) Time series of daily value of the S&P500, DAX, Nikkai225 indices (from left to right) from 04/27/1993 to 11/11/2021. Data Source: Yahoo!. (n.d.). *Yahoo! Finance*. https://finance.yahoo.com; (c) Time series of percentage of the public vote in political polls on federal elections in *G*ermany from 01/07/2000 to 05/12/2023. Data Source: Wahlrect,de e. V. (n.d.). *Sonntagsfrage Bundestagswahl*. https://www.wahlrecht.de/umfragen/index.htm; (d) Second-by-second coded vocalizations of mother (left) and infant (right) during interaction. Data Source: Leonardi et al., 2016. The data sets, together with R-code implementing the analyses, an also be found here: https://osf.io/8ubcj/.

Analysis (Chapter 3) can be used to address questions such as how strongly mother-infant vocalizations couple. The Diagonal Cross-Recurrence Profile (Chapter 4) can be used to address the question of whether mothers or infants lead or follow in terms of their vocalization patterns with regard to each other. Windowed Recurrence Analysis (Chapter 5) can be used to reveal changes in infection dynamics related to the onset of infection waves, or how coupling of mother-infant interactions evolves over time. Finally, Multidimensional Recurrence Quantification Analysis (Chapter 6) can be used to ask questions such as whether the joint dynamics of different stock market indices reveal global patterns of economic trends.

NOTES

1. In the literature, the terms *threshold* and *radius* are used interchangeably. The term *radius* points to the idea that the dissimilarity or distance between values can be measured in a multidimensional space, so that recurrence happens when one point-value is in the vicinities in space of another point-value, that is, it is inside the multidimensional sphere with the chosen radius around the other point-value.
2. Forward delays, where reconstruction-space vectors are indicated as $\left[y_{(t)}, y_{(t+\tau)}, y_{(t+2\tau)}, \ldots, y_{(t+(m-1)\tau)}\right]$, are more commonly used in the vectors notation and practical computation in the literature. However, from a mathematical point of view, the two notations are equivalent while the one used in the text better obeys causality.
3. Please be aware that in alternative software and in the literature those parameters may go under different names, like lag or time lag instead of delay, and threshold or cut-off distance for the radius parameter.