9 Nonlinear Time Series Analysis in a Nutshell

Ralph Gregor Andrzejak

CONTENTS

9.1 INTRODUCTION

Nonlinear time series analysis is a practical spinoff from complex dynamical systems theory and chaos theory. It allows one to characterize dynamical systems in which nonlinearities give rise to a complex temporal evolution. Importantly, this concept allows extracting information that cannot be resolved using classical linear techniques such as the power spectrum or spectral coherence. Applications of nonlinear time series analysis to signals measured from the brain contribute to our understanding of brain functions and malfunctions and thereby help to advance cognitive neuroscience and neurology. In this chapter, we show how a combination of a nonlinear prediction error and the Monte Carlo concept of surrogate time series can be used to attempt to distinguish between purely stochastic, purely deterministic, and deterministic dynamics superimposed with noise.

The framework of nonlinear time series analysis comprises a wide variety of measures that allow one to extract different characteristic features of a dynamical system underlying some measured signal (Kantz and Schreiber 1997). These include the correlation dimension as an estimate of the number of independent degrees of freedom, the Lyapunov exponent as a measure for the divergence of similar system states in time, prediction errors as detectors for characteristic traits of deterministic dynamics, or different information theory measures. The aforementioned nonlinear time series measures are univariate, i.e., they are applied to single signals measured from individual dynamics. In contrast, bivariate measures are used to analyze pairs of signals measured simultaneously from two dynamics. Such bivariate time series analysis measures aim to distinguish whether the two dynamics are independent or interacting through some coupling. Some of these bivariate measures aim to extract not only the strength, but also the direction of these couplings. The Monte Carlo concept of surrogates allows one to test the results of the different nonlinear measures against well-specified null hypotheses.

125

We should briefly define the different types of dynamics. Let $\mathbf{x}(t)$ denote a q-dimensional vector that fully defines the state of the dynamics at time t. For purely deterministic dynamics, the future temporal evolution of the state is unambiguously defined as a function f of the present state:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)). \tag{9.1}$$

Here $\mathbf{x}(t)$ denotes the time derivative of the state $\mathbf{x}(t)$. We further suppose that a univariate time series s_i is measured at integer multiples of a sampling time Δt from the dynamics using some measurement function h:

$$s_i = h(\mathbf{x}(i\Delta t)). \tag{9.2}$$

The combination of Equation 9.1 and Equation 9.2 results in the purely deterministic case. When some noise $\psi(t)$ enters the measurement:

$$s_i = h(\mathbf{x}(i\Delta t), \psi(i\Delta t)), \tag{9.3}$$

we obtain the case of deterministic dynamics superimposed with noise. Here it is important to note that the noise only enters at the measurement. The deterministic evolution of the system state $\mathbf{x}(t)$ via Equation 9.1 is not distorted. In contrast to such measurement noise stands dynamical noise $\xi(t)$ for which we have:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \xi(t)). \tag{9.4}$$

Due to the noise term, the future temporal evolution of $\mathbf{x}(t)$ is not unambiguously determined by the present state. Accordingly, Equation 9.4 represents stochastic dynamics. We again suppose that a measurement function, with or without measurement noise (Equations 9.3 and 9.2, respectively), is used to derive a univariate time series s_i from this stochastic dynamics.

As stated above, we aim to distinguish between purely stochastic, purely deterministic, and deterministic dynamics superimposed with noise. Accordingly, we want to design an algorithm A to be applied to a time series s_i that fulfills the following criteria:

  i. $A(s_i)$ should attain its highest values for time series s_i measured from any purely stochastic process.
  ii. $A(s_i)$ should attain its lowest values for time series s_i measured from a noise-free deterministic dynamics.[*,†]
  iii. $A(s_i)$ should attain some value in the middle of its scale for time series s_i measured from noisy deterministic dynamics.

In particular, the ranges obtained for ii. and iii. should not overlap with the one of i. For iii., $A(s_i)$ should ideally vary with the relative noise amplitude.

In order to study our problem under controlled conditions, we use a time series of mathematical model systems that represent these different types of dynamics. While this chapter is restricted to these model systems, the underlying lecture included results for exemplary intracranial electroencephalographic (EEG) recordings of epilepsy patients (Andrzejak et al. 2001). These EEG times

---

[*] The following text contains a number of footnotes. These are meant to provide background information. The reader might safely ignore them without risk of losing the essentials of the main text.
[†] The polarity of A does not matter. It could likewise attain lowest values for i. and highest values for *ii*.

series as well as the source codes to generate the different time series studied here, the generation of surrogates, and the nonlinear prediction error are freely available.[*]

9.2 DYNAMICS AND TIME SERIES

9.2.1 Nonlinear Deterministic Dynamics—Lorenz Dynamics

As nonlinear deterministic dynamics we use the following Lorenz dynamics.[†]

$$\dot{x}(t) = 10(y(t) - x(t)) \tag{9.5}$$

$$\dot{y}(t) = 39x(t) - y(t) - x(t)z(t) \tag{9.6}$$

$$\dot{z}(t) = x(t)y(t) - (8/3)z(t). \tag{9.7}$$

This set of first-order differential equations can be integrated using, for example, a fourth order Runge–Kutta algorithm.[‡] To obtain time series x_i, y_i, z_i for $i = 1,\ldots,N = 2048$, we sample the output of the numerical integration at an interval of $\Delta t = 0.03$ time units.[§] We denote the resulting temporal sequence of three-dimensional vectors as $L_i = (x_i, y_i, z_i)$. In our analysis, we restrict ourselves to the first component x_i (see Table 9.1 and Figure 9.1). Hence, we have an example of purely deterministic dynamics (Equation 9.1). The discretization in time is obtained from the numerical integration of the dynamics. This integration can be considered as part of the measurement function (Equation 9.2), which furthermore consists of the projection to the first component.

9.2.2 Noisy Deterministic Dynamics—Noisy Lorenz Dynamics

By adding white noise to the first component of our Lorenz time series x_i, we obtain noisy deterministic dynamics:

$$n_i = x_i + 18\psi_i, \tag{9.8}$$

where ψ_i denotes uncorrelated noise with uniform amplitude distribution between 0 and 1 (see Table 9.1 and Figure 9.1). Our time series n_i is an example of a signal generated by deterministic dynamics (Equation 9.1) where noise is superimposed on the measurement function Equation 9.3. To roughly assess the strength of the noise, we have to know that the standard deviation of the noise term ψ_i and the Lorenz time series x_i is approximately 0.29 and 9.6, respectively. With the prefactor of 18, the standard deviation of the noise[¶] amounts to approximately 5.2.

---

[*] www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html; www.cns.upf.edu/ralph.

[†] Equations 9.5 through 9.7 result in chaotic dynamics. However, this is not of our concern here. Rather, we use the Lorenz dynamics as a representative of some aperiodic deterministic dynamics.

[‡] Here we use an algorithm with fixed step size of 0.005 time units. Starting at random initial condition we use some arbitrary but high number (e.g., 10^6) of preiterations to diminish transients.

[§] In the signal analysis literature, one often finds 1024, 2048, 4096, etc. for the number of samples. The reason is that these numbers are integer powers of two: $N* = 2^n$ with n being an integer. For these numbers of samples, one can use the very efficient fast Fourier transform algorithm. For other numbers of samples, one has to use algorithms that divide the number of samples in prime factors and that are slower.

[¶] In assessing the relative strength of noise superimposed to the Lorenz dynamics, it is important to note that a major contribution to the overall standard deviation originates from the switching behavior of this dynamics. The standard deviation of the actual aperiodic oscillations in the two wings is smaller (see Figures 9.1–9.2).

TABLE 9.1
Overview of Different Time Series Studied

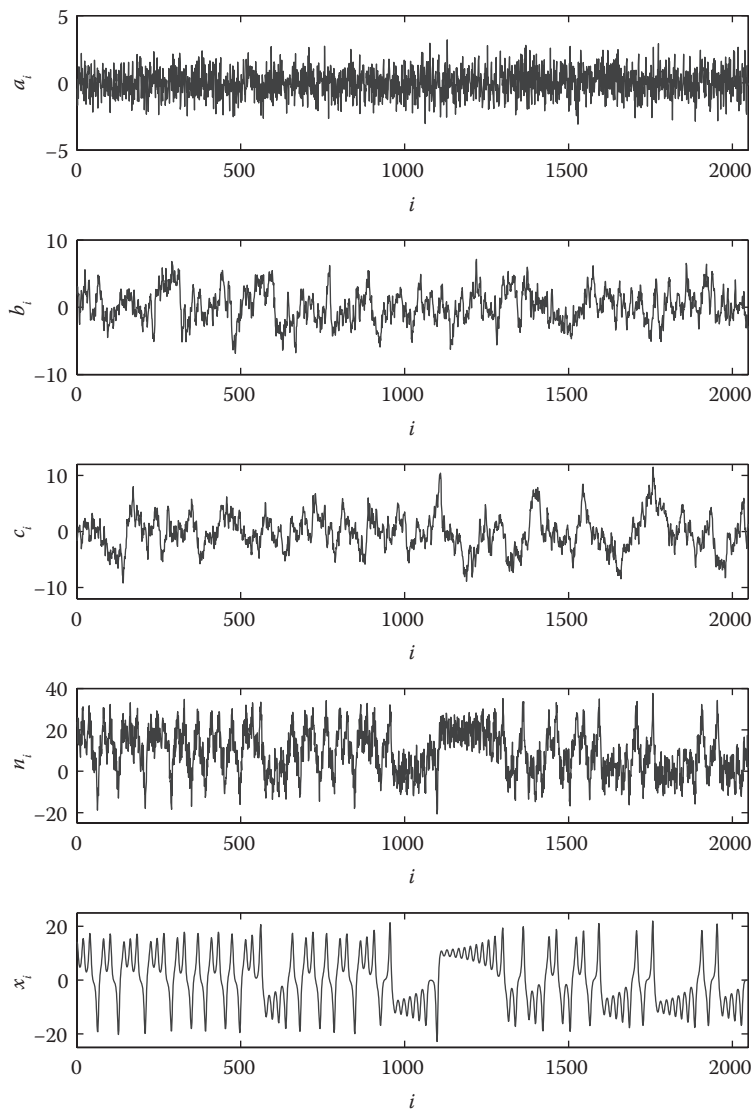| Type		Origin and Parameter
a_i	Stochastic	AR process, Equation 9.9 with $r = 0$
b_i	Stochastic	AR process, Equation 9.9 with $r = 0.95$
c_i	Stochastic	AR process, Equation 9.9 with $r = 0.98$
x_i	Deterministic	x-component of Lorenz dynamics
n_i	Noisy deterministic	x_i superimposed with white noise

FIGURE 9.1 Plots of the different time series studied here.

9.2.3 Stochastic Processes—Linear Autoregressive Models

As a stochastic process, we use a simple but still very instructive example—a first-order linear autoregressive process with varying degrees of autocorrelation:

$$\zeta_{i+1} = r\zeta_i + \xi_i, \tag{9.9}$$

for $i = 1, \ldots, N = 2048$, where ξ_i denotes uncorrelated Gaussian noise with 0 mean and unit variance.[*] We can readily see that for $r = 0$, Equation 9.9 results in white, i.e., uncorrelated noise. Upon the increase of r, within the interval $0 < r < 1$, the strength of the autocorrelation increases. For this chapter, we study three different time series derived from Equation 9.9 using different values of r (see Table 9.1 and Figure 9.1).

In contrast to the time-continuous stochastic dynamics expressed in Equation 9.4, the autoregressive process of Equation 9.9 is defined for discrete time. What is common to both formulations, and what is the important point here, is that the noise is intrinsic to the dynamics; no additional noise enters through the measurement. We should note that the fact that we can write down a rule to generate an autoregressive process (as in Equation 9.9) does not conflict with its stochastic nature in any way. The point is that due to the noise term, ξ_i, an initial condition does not unambiguously determine the future temporal evolution of the dynamics. Furthermore, without this noise term, the autoregressive process would not result in sustained dynamics but rather decay to 0.

9.3 DETERMINISTIC VERSUS STOCHASTIC DYNAMICS

Plotting the three components of the Lorenz dynamics L_i in a three-dimensional state space, we obtain the trajectory of the dynamics that form the famous Lorenz attractor (Figure 9.2, left). Once we specify some state of the Lorenz dynamics $x(0)$, $y(0)$, $z(0)$ at $t = 0$ as an initial condition the future temporal evolution of the dynamics is unambiguously determined by Equations 9.5–9.7. As a result of this deterministic nature of the Lorenz dynamics, its trajectory cannot intersect with itself.[†] A trajectory crossing would imply that the state at the crossing would have two distinct future evolutions, which would contradict with the uniqueness of the solution of Equations 9.5–9.7. Inspecting the trajectory in the left panel of Figure 9.2, we can appreciate that neighboring trajectory segments are aligned. That means that similar present states of the deterministic Lorenz dynamics result in similar states in the near future.

In opposition to deterministic dynamical systems are stochastic dynamics. When we start Equation 9.9 twice for an identical initial condition, $\zeta 0$, and the same r value, we get two completely different solutions. Moreover, similar present states of stochastic dynamics will typically not result in similar future states. Accordingly, trajectories of stochastic dynamics can intersect, and neighboring segments are typically not aligned. This is illustrated in the right panel of Figure 9.2, where we plot the trajectory of simple stochastic dynamics derived from Equation 9.9) in a three dimensional state space.

9.4 DELAY COORDINATES

In the previous section, we saw that the alignment of neighboring trajectory segments can be used as a criterion to try to distinguish deterministic from stochastic dynamics (See the two panels of Figure 9.2). Before we can quantify this criterion we have to realize, however, that in experimental measurements from real-world dynamics we typically cannot access all variables of a dynamical

---

[*] To avoid transients, we again run some arbitrary but high number of preiterations (e.g., 10^4). That is, we start Equation 9.9 at $\zeta_{-10000} = 0$ before recording the samples of the process starting from $\zeta 1$.

[†] The edginess at the outer corners of the wings results from the fixed nonzero step size of the numerical integration.
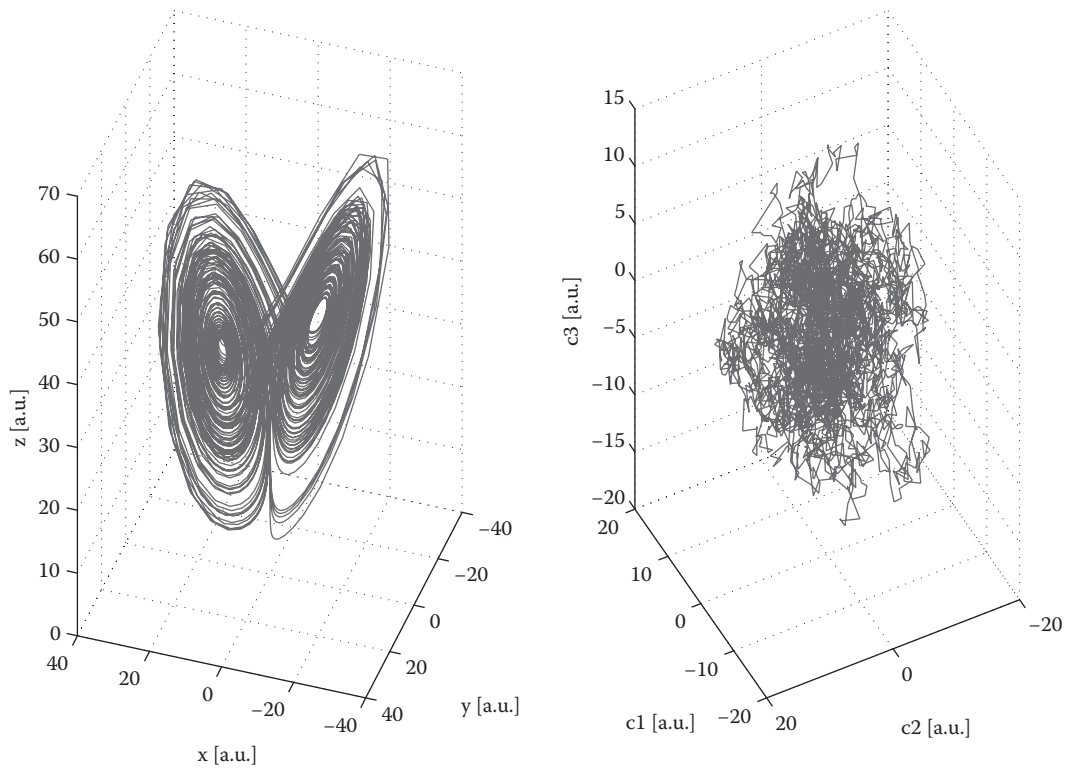
FIGURE 9.2   (left) The three components of the Lorenz dynamics x_i, y_i, and z_i plotted in a three-dimensional state space. (right) Three independent realizations of an autoregressive process (Equation 9.9 for $r = 0:98$) plotted in a three-dimensional state space.

system. Suppose for example that we study a real-world dynamical system that is fully described by the Lorenz differential Equations 9.5–9.7. Suppose, however, that we cannot measure all three components of L_i, but rather only $s_i = x_i$ is assessed by our measurement function. In this case, we can use the method of delay coordinates to obtain an estimate of the underlying dynamics (Takens 1981):

$$v_i = (s_i,\ldots,s_{i-(m-1)\tau}), \tag{9.10}$$

with embedding dimension m and delay τ for $i = 1 + (m - 1)\ \tau\ , \ldots ,N = \eta, \ldots ,N$. Where we define $\eta = 1 + (m - 1)\tau$ to simplify the notation below. An introduction to the embedding theorem underlying the reconstruction of dynamics using such delay coordinates can be found in (Kantz and Schreiber 1997). We restrict ourselves here to a hand-waving argument that Equation 9.10 allows us to obtain a structure that is topologically similar to the original dynamics. In Figure 9.3, we contrast the full Lorenz dynamics L_i with delay coordinate reconstructions obtained from its first component x_i using $m = 3$ and different values of τ. Comparing these structures, we find a striking similarity between the original dynamics and its reconstruction for not too high values of τ. For low values of τ, the dynamics is somewhat stretched out along the main diagonal and is not properly unfolded. At an intermediate value of τ, the two-wing structure of the Lorenz dynamics is clearly discernible from the reconstruction. For too high a value of τ, the dynamics is overfolded.
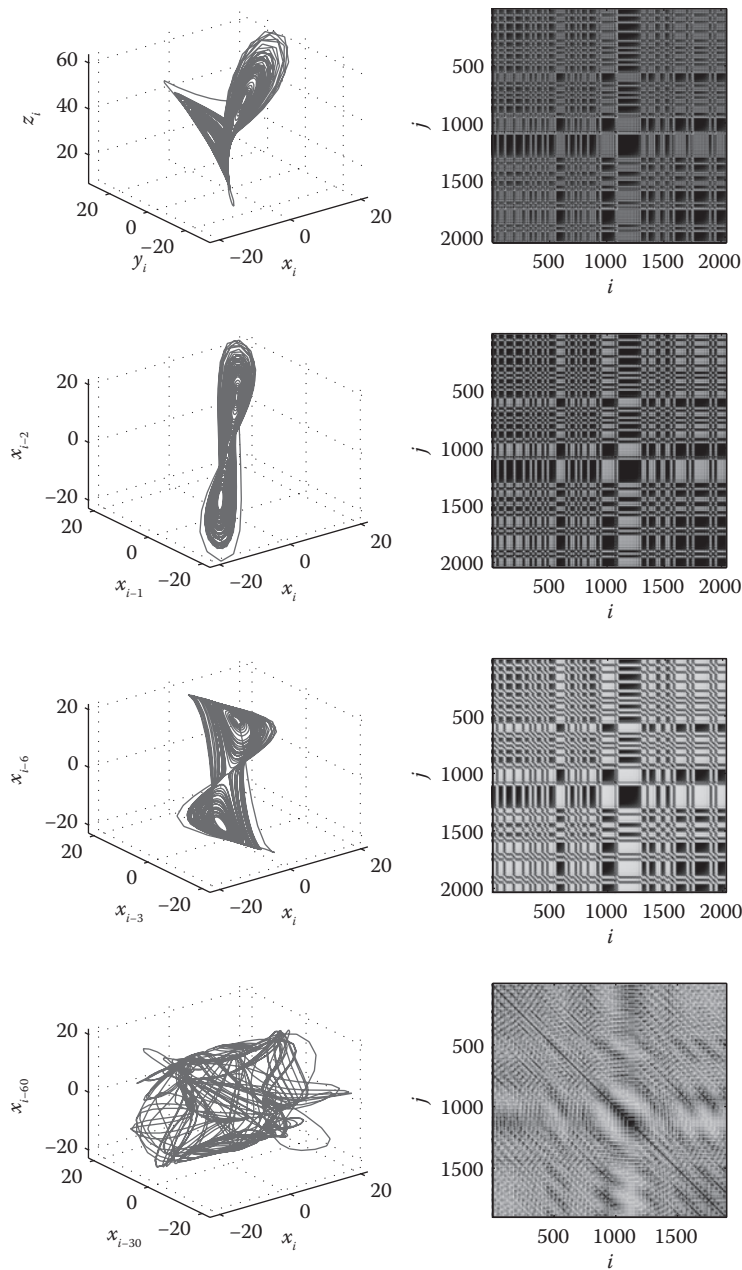
FIGURE 9.3   (first row, left) The three components of the Lorenz dynamics $L_i = (x_i, y_i, z_i)$ plotted in a three-dimensional state space. Hence, this panel is the same as the left panel of Figure 9.2, but viewed from a different angle. (first row, right) Euclidean distance matrix calculated for the Lorenz dynamics L_i. (second row, left) Delay reconstruction of the Lorenz dynamics based on the first component, x_i. Here we use a time delay of $\tau = 1$. For obvious reasons, the maximal embedding dimension that we can use for this illustration is $m = 3$. (second row, right) Euclidean distance matrix calculated from such an embedding, but for $m = 6$. (third and fourth rows) Same as second row, but for $\tau = 6$ and $\tau = 30$, respectively.

We can appreciate this similarity further by inspecting Euclidean distance matrices derived from the full Lorenz dynamics and its delay coordinate reconstruction.[*] For a set of q-component vectors w_i for $i = \eta, \ldots, N$, such a matrix contains in its element Dij the Euclidean distance between the vectors w_i and w_j. The Euclidean distance between two q-component vectors w_1 and w_2 is defined as

$$d(w_1, w_2) = \sqrt{\sum_{l=1}^{q} (w_{1,l} - w_{2,l})^2} .\qquad(9.11)$$

Figure 9.3 shows the distance matrix for the full Lorenz dynamics L_i, which can also be seen as a set of N vectors with $q = 3$ components. This is contrasted to the distance matrix derived from delay coordinate reconstructions v_i, which is a set of N vectors with $q = m$ components. While the overall scales of the different matrices are different, we can observe an evident similarity between the structures found in these matrices for the original and its delay coordinate reconstruction for not too high values of τ. Importantly, this similarity implies that close and distant states in the full Lorenz dynamics are mapped to close and distant states in the reconstructed dynamics, respectively. This topological similarity implies that our criterion of distinction between deterministic and stochastic dynamics, namely the alignment of neighboring trajectory segments, carries over from the original dynamics to the reconstructed dynamics obtained by means of delay coordinates. The different distance matrices derived from the delay reconstructions are, however, not identical to the original distance matrix and the degree of their similarity depends on the embedding parameters m and τ. These aspects will be discussed in more detail below. To get a thorough understanding of the quality of the embedding requires playing with different parameters and viewing the result from different angles. This can be done using the source code referred to in the introduction.

9.5 NONLINEAR PREDICTION ERROR

Above we have identified the alignment of neighboring trajectory segments as a signature of deterministic dynamics. We will now introduce the nonlinear prediction error as a straightforward way to quantify the degree of this alignment (see Kantz and Schreiber (1997) and references therein).

Before we begin, we first normalize our time series s_i to have zero mean and unit standard deviation and variance.[†] Subsequently, we form a delay reconstruction v_i from our time series for $i = \eta, \ldots, N$. To calculate the nonlinear prediction error with horizon h we carry out the following steps for $i_0 = \eta, \ldots, N - h$. We take v_i as reference point and look up the time indices of the k nearest neighbors of the reference point: $\{i_g\}(_{g=1,\ldots,k})$. These k nearest neighbors are simply those points that have the k smallest distances to our reference point in the reconstructed state space. In other words, $\{i_g\}(_{g=1,\ldots,k})$ are the indices of the k smallest entries in row i_0 of the distance matrix derived from v_i.[‡] Now we use the future states of these nearest neighbors to predict the future states of our reference point and quantify the error that we make in doing so by

$$\varepsilon_{i_0} = s_{i_0} + h - \frac{1}{k} \sum_{g=1}^{k} s_{i_g} + h. \qquad(9.12)$$

---

[*] The choice of the Euclidean distance is arbitrary. The same arguments hold when other definitions of distances are used.

[†] In this way we do not have to normalize the prediction error itself and the subsequent notations become simpler.

[‡] Since distance matrices are symmetric, these indices coincide with the k smallest entries in the column i_0 of the matrix. Note however that if v_{i1} is among the k nearest neighbors of v_{i2}, this does not imply that v_{i2} is among the k nearest neighbors of v_{i1}.

Finally, we take the root-mean-square over all reference points:

$$E = \sqrt{\frac{1}{N - h - \eta + 1} \sum_{i_0 = \eta}^{i_0 = N - h} \varepsilon_{i_0}^2}. \tag{9.13}$$

Note that while the nearest neighbors are determined according to their distance in the reconstructed state space, the actual prediction in Equation 9.12 is made using the scalar time series values. Regarding Equation 9.12, it also becomes clear that we must exclude nearest neighbors with indices $i_r > N - h$. Furthermore, we should aim to base our prediction on neighboring trajectory segments rather than on the preceding and subsequent piece of the trajectory on which our reference point v_{i_0} is located on. For this purpose we apply the so-called Theiler correction with window length W (Theiler 1986), by including only points to the set of k nearest neighbors that fulfill $|i_g - i_0| > W$.

For the parameters, we use the following values. We fix the embedding dimension used for the delay coordinates to $m = 6$, the number of nearest neighbors to $k = 5$, the prediction horizon to $h = 5$ samples, and the length of the Theiler correction to $W = 30$. The delay used for the delay coordinates is varied according to $\tau = 1, 2, \ldots, 29$.

Figure 9.4 shows results obtained for our set of exemplary time series. The white noise time series a_i results in E values around 1.1. For this stochastic process, points close to some reference points do not carry any predictive information about the future evolution of this reference point. Neighboring state space trajectory segments of this stochastic process are not aligned. Hence, for white noise the predictions $s_{i_r + h}$ used in Equation 9.12 are no better than guessing a set of k points randomly drawn from s_i.

For the noise-free deterministic Lorenz time series x_i, neighboring trajectories segments are well aligned. Points close to some reference points have a similar future evolution as this reference point. The predictions $s_{i_r + h}$ used in Equation 9.12 are good predictions for $s_{i_0 + h}$. Therefore, we get small values of E. For increasing τ the prediction error increases, reflecting the overfolding of the reconstructed dynamics (see Figure 9.3). Nonetheless, as of now, E seems to fulfil the criteria i. and ii.
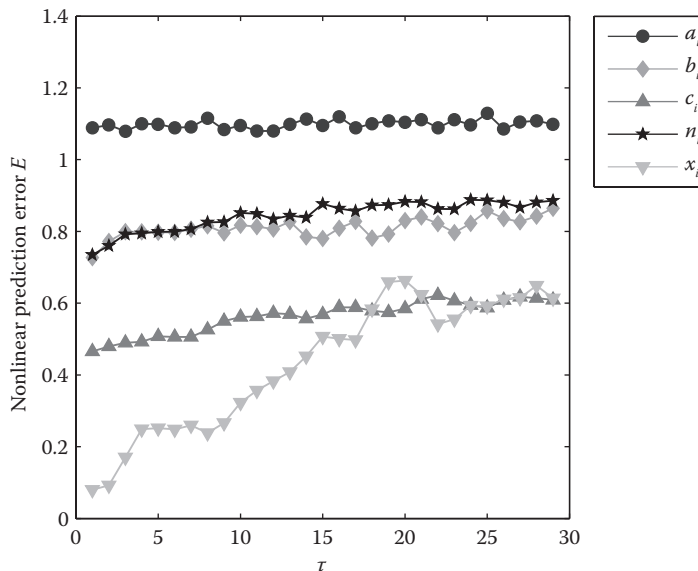

FIGURE 9.4 Dependence of the nonlinear prediction error, E, on the time delay, τ, used for the state space reconstruction. Here, results are shown only for the original time series.

formulated in the Introduction. It allows us to clearly tell apart our deterministic Lorenz time series from white noise. It can be objected that this distinction is not terribly difficult, and could readily be achieved by visual inspection of our data.

We should therefore proceed to our noisy deterministic time series n_i and time series of autoregressive processes with stronger autocorrelation (b_i, c_i). When we look at their results in Figure 9.4, we must realize that E is in fact not a good candidate for Figure 9.4A. For the time series n_i it is clear that the noise must degrade the quality of the predictions of Equation 9.12 as compared to the noise-free case of x_i. This increase in E by itself would not disqualify E, as it is compatible with criterion iii. The problem is that E values for b_i overlap with those for n_i. Results for c_i are even lower than those for n_i and for higher τ, even similar to those of x_i. The reason why these autocorrelated time series exhibit a higher predictability as opposed to white noise is the following: v_{i_1} being a close neighbor of v_{i_2} implies that the scalar s_{i_1} is typically similar to the scalar s_{i_2}. Furthermore, s_{i_1} being similar to s_{i_2} implies, to a certain degree, that s_{i_1+h} is similar to s_{i_2+h}. It is instructive to convince oneself that it is just the strength of the autocorrelation that determines the strength of the "to a certain degree" in the previous statement.

As a consequence, stochastic time series with stronger autocorrelation and noisy deterministic dynamics cannot be told apart by our nonlinear time series algorithm E. We are still missing one very substantial ingredient to reach our aim, which will we introduce in Section 9.6.

9.6 THE CONCEPT OF SURROGATE TIME SERIES

Nonlinear time series analysis was initially developed studying complex low-dimensional deterministic model systems such as the Lorenz dynamics. The brain is certainly not a low-dimensional deterministic dynamics and will typically not behave as one. Rather, it might be regarded as a particularly challenging high-dimensional and complicated dynamical system that might have both deterministic and stochastic features. Problems that arise in applications to signals measured from the nervous system continue to trigger the refinement of existing algorithms and development of novel approaches and, thereby, help to advance nonlinear time series analysis. One important concept in nonlinear time series analysis, which has some of its roots in EEG analysis (Pijn et al. 1991), is the Monte Carlo framework of surrogate time series (Theiler et al. 1992; Schreiber and Schmitz 2000).

Suppose that we have calculated the nonlinear prediction error for some experimental times series and consider the following scenarios. First, we obtain a relatively low E value but have doubts that this is due to some deterministic structure in our time series. Second, we obtain a relatively high E value but we suspect that our time series should have some deterministic structure that is possibly distorted by some measurement noise. In any case, we want to know what a relatively high or a relatively low value of the nonlinear prediction error means. A way to address this problem is to scrutinize our time series further by testing different null hypotheses about it by means of surrogate time series. Such surrogate time series are generated from a constrained randomization of an original time series.[*] The surrogates are designed to share specific properties with the original time series, but are otherwise random. In particular, the properties that the surrogates share with the original time series are selected such that the surrogates represent signals that are consistent with a well-specified null hypothesis.

Here, we use the simple but illustrative example of phase-randomized surrogates (Theiler et al. 1992). To construct these surrogates, first calculate the complex-valued Fourier transform of the original time series. In the next step, randomize all phases of the complex-valued Fourier coefficients between 0 and 2π.[†] Finally, the inverse Fourier transform of these modified Fourier

---

[*] See Schreiber and Schmitz (2000) for the alternative to using a typical randomization to generate surrogates.

[†] In this randomization it is very important to preserve the antisymmetry between Fourier coefficients representing positive and negative frequencies. This antisymmetry holds for real-valued signals. If it is not preserved the inverse Fourier transform will not result in a real-valued time series.

coefficients results in a surrogate time series. An ensemble of surrogates is generated using different randomization of the phases. Randomizing the phase of a complex number does not affect its absolute value, that is, the length of the corresponding vector in the complex plane is preserved. Therefore, the power of the resulting signal and its distribution over all frequencies, as measured via the periodogram, is identical to the one of the original time series. Consequently, the surrogates also have the same autocorrelation function as the original time series. However, regardless of the nature of the original time series, the surrogate time series is a stationary linear stochastic Gaussian process. Whether the original was deterministic or stochastic, stationary or nonstationary, Gaussian or nonGaussian, the surrogate is stochastic, stationary, and Gaussian. In that sense, the surrogate algorithm projects from the space of all possible time series with a certain periodogram to the space of all stationary linear stochastic time Gaussian time series with this same periodogram. To see this point, it is instructive to convince oneself that a phase-randomized surrogate of a phase-randomized surrogate cannot be distinguished from a phase-randomized surrogate of the original time series.

In Figure 9.5, we again show values of the nonlinear prediction error for our exemplary time series (see again Figure 9.4). Here, however, we not only show results for the original time series but also contrast these to results obtained for an ensemble of 19 phase-randomized surrogate time series generated for each original time series. For all time series from autoregressive processes a_i, b_i, and c_i, the results for the surrogates match those of the original time series. The autoregressive processes generated by Equation 9.9 are stationary linear stochastic Gaussian processes. The r value determines the strength of the autocorrelation of this process and thereby the shape of the periodogram. A phase-randomized surrogate of a stationary linear stochastic Gaussian process with any periodogram is a stationary linear stochastic Gaussian process with this very periodogram. Accordingly, phase-randomized surrogates constructed from a_i, b_i, and c_i cannot be distinguished from their corresponding original time series. The null hypothesis is correctly accepted for all three cases.[*]

For both the noise-free Lorenz time series x_i and the noisy Lorenz time series n_i, we find that the original time series results in lower E values as compared to their corresponding ensemble of surrogates. Here, we deal with a purely deterministic time series and a noisy deterministic one. When we construct surrogates from them, their deterministic structures are destroyed. In particular, the randomization destroys the local alignment of neighboring trajectory segments. Regarding the significant difference between the E values for the original time series and the mean values obtained for the surrogates, we can conclude that the null hypothesis is correctly rejected for both x_i and n_i.

From these results, it is only a small step for us to arrive at our aim by defining

$$A = \langle E \rangle \text{surr} - E_{\text{orig}}, \tag{9.14}$$

where the brackets denote the average taken across the ensemble of 19 surrogates. The resulting surrogate-corrected nonlinear prediction error values A are shown in Figure 9.5, and we can see that the definition of Equation 9.14 fulfills criteria i., ii., and iii. that we defined as the aim of this chapter.

9.7 INFLUENCING PARAMETERS

In this chapter, we have plotted the nonlinear prediction error only in dependence on the time delay τ used for the delay coordinates. To fully understand how nonlinear techniques and surrogates work, it is always important to study the influence of all parameters in detail. In our case, these parameters are the delay τ and embedding dimension m used for the delay coordinates, the number of nearest

---

[*] A good check of whether one is still on track in understanding this chapter is to find an explanation of why our different autoregressive processes result in different E values, and why the surrogates match closely with their corresponding original time series in these E values, rather than just undiscriminatingly covering the entire range of E values obtained for a_i, b_i, and c_i.
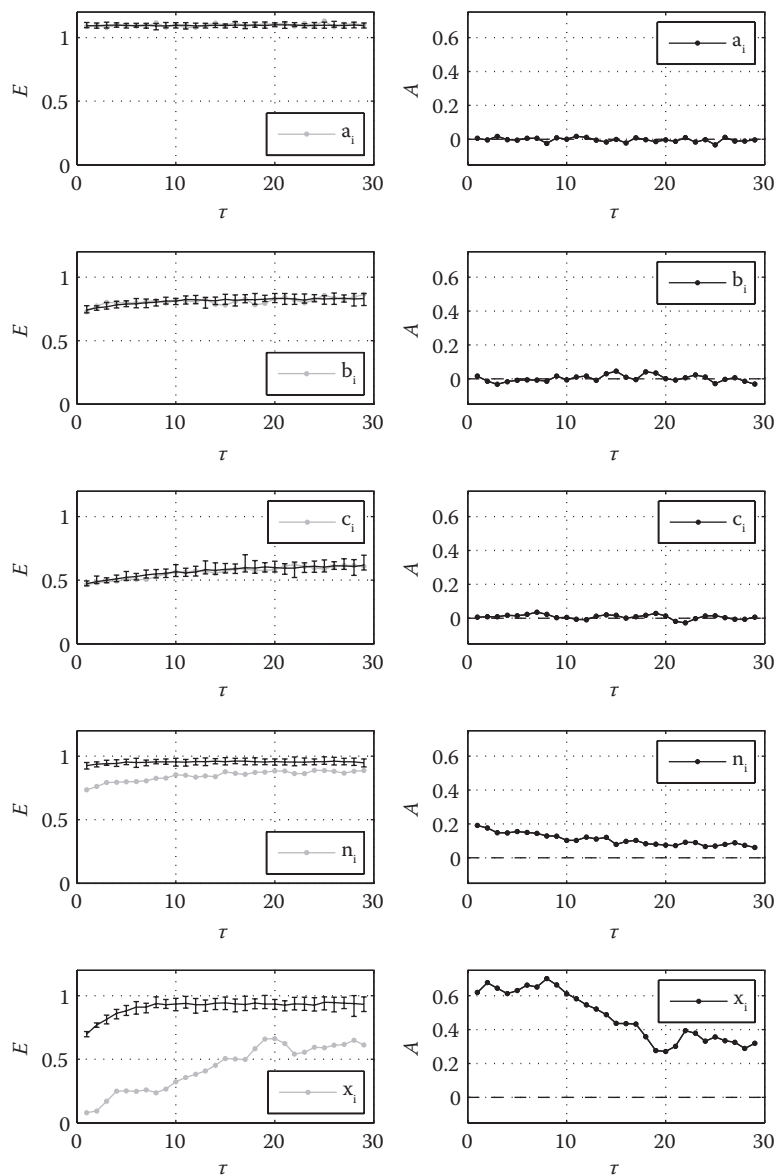
FIGURE 9.5 (left) Dependence of the nonlinear prediction error, E, on the time delay, τ, used for the state space reconstruction. Here, results are shown for the original time series (grey lines with symbols) and 19 surrogate time series (black lines with error bars). The error-bar denotes the range obtained from the surrogates. (right) Dependence of the surrogate-corrected nonlinear prediction error, A, on the time delay, τ. Each row corresponds to results of one specific time series (see legends).

neighbors k and the prediction horizon h, and, finally, the length of the Theiler-correction W. In general, the influence of these different parameters will not be independent from another. For example, it is often the so-called embedding window $\tau(m - 1)$ that matters, rather then the two parameters τ and m independently. There exist very useful guidelines and recipes for an adequate choice of these different parameters (see Kantz and Schreiber (1997) and references therein). However, whenever one is analyzing some type of experimental data, one should play with these parameters and see

how they influence results. Certainly, if one finds some effect in experimental data using nonlinear time series analysis techniques, the significance of this effect should not critically depend on the choice of the different parameters.

We have studied our problem based on only single realizations of our time series. However, whenever we have controlled conditions, we should make use of them and generate many independent realizations of our time series. In this way, we can get more reliable results and also learn more about our measures by regarding their variability across realizations. Another important issue that we have neglected here is that it can be very informative to study the influence of noise for a whole range of noise amplitudes and for different types of noise.

9.8 WHAT CAN AND CANNOT BE CONCLUDED?

First, we can draw a very positive conclusion: We reached the aim that we stated in the Introduction. Using a combination of a nonlinear prediction error and surrogates, we were able to distinguish between purely stochastic, purely deterministic, and deterministic dynamics superimposed with noise. We demonstrated this distinction under controlled conditions using mathematical model systems with well-defined properties. However, how do these results carry over once we leave the secure environment of controlled conditions? Under controlled conditions, we know what dynamics underlie our time series. Here, we can draw conclusions such as the null hypothesis is correctly accepted or rejected, but what conclusions can be drawn when facing real-world experimental time series from some unknown dynamical system?

Let us suppose we want to test the following working hypothesis consisting of two parts. First, the dynamics exhibited by an epileptic focus is nonlinear deterministic. Second, the dynamics exhibited by healthy brain tissue is linear stochastic. Suppose that to test this hypothesis we compare EEG time series measured from within an epileptic focus against EEG time series measured outside the focal area. We calculate the nonlinear prediction error for these EEG time series and phase-randomized surrogates constructed from them. Suppose further that for the focal EEG time series we find significant differences between the results for original EEG time series and those of the surrogates, whereas for the nonfocal EEG we find that the results for the original are not significantly different from the surrogates. (In the lecture underlying this chapter we looked at such results for real EEG time series.) Have we proven our working hypothesis? The answer is no, we have not—neither its first nor its second part.

First, we need to recall that the rejection of a null hypothesis at say $p < 0.05$ in no way proves that the null hypothesis is wrong. It only implies that if the null hypothesis was true, the probability to get our result (or a more extreme result) is less than 5%. Putting that aside, let us assume that the difference between the focal EEG and surrogates is so pronounced that we deem the underlying null hypothesis as indeed very implausible. Still, this does not prove that the dynamics are nonlinear deterministic. Recall that we used phase-randomized surrogates representing the null hypothesis of a linear stochastic Gaussian stationary process. We have to keep in mind that the complement to our null hypothesis is the world of all dynamics that do not comprise all the properties: linear, stochastic, Gaussian, and stationary. Hence, if we deal, for example, with a stationary linear stochastic nonGaussian dynamics, our null hypothesis is wrong. Our test should reject it. Likewise, we might have nonstationary linear stochastic Gaussian dynamics or nonlinear deterministic process. We might also deal with stationary linear stochastic Gaussian dynamics measured by some nonlinear measurement function. We have to realize that the dynamics exhibited by an epileptic focus is nonlinear deterministic and is among these alternatives. However, it is only one of many explanations of our findings. Given only the results of our test, our working hypothesis is not more likely than any of the other explanations of the rejection of our null hypothesis.

We should furthermore note that our test cannot prove the second part of our null hypothesis either. The fact that we could not reject the null hypothesis for our nonfocal EEG does not prove the correctness of this null hypothesis. We cannot rule out that our nonfocal EEG originates from some

nonlinear deterministic dynamics, but our test is not sufficiently sensitive to detect this feature. Our time series might be too short or too noisy for the prediction error to detect the alignment of neighboring trajectories, or we might have used inadequate values of our parameters m, τ, and k.

It is important to keep these limitations in mind when interpreting results derived from nonlinear measures in combination with surrogates.[*] Nonetheless, nonlinear measures in combination with surrogates can still be very useful in characterizing signals measured from the brain. Andrzejak et al. studied the discriminative power of different time series analysis measures to lateralize the seizure-generating hemisphere in patients with medically intractable mesial temporal lobe epilepsy (Andrzejak et al. 2006). The measures that were tested comprised different linear time series analysis measures, different nonlinear time series analysis measures, and a combination of these nonlinear time series analysis measures with surrogates. Subject to the analysis were intracranial electroencephalographic recordings from the seizure-free interval of 29 patients. The performance of both linear and nonlinear measures was weak, if not insignificant. A very high performance in correctly lateralizing the seizure-generating hemisphere was, however, obtained by the combination of nonlinear measures with surrogates. Hence, the very strategy that brought us closest to the aim formulated in the introduction of this chapter—to reliably distinguish between stochastic and deterministic dynamics in mathematical model systems—also seems key to a successful characterization of the spatial distribution of the epileptic process. The degree to which such findings carry over to the study of the predictability of seizures was among the topics discussed at the meeting in Kansas.

REFERENCES

Andrzejak, R. G., K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger. 2001. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E* 64:061907.

Andrzejak, R. G., G. Widman, F. Mormann, T. Kreuz, C. E. Elger, and K. Lehnertz. 2006. Improved spatial characterization of the epileptic brain by focusing on nonlinearity. *Epilepsy Res.* 69:30–44.

Kantz, H., and T. Schreiber. 1997. *Nonlinear time series analysis*. Cambridge: Cambridge Univ. Press.

Pijn, J. P., J. van Neerven, A. Noes, and F. Lopes da Silva. 1991. Chaos or noise in EEG signals; dependence on state and brain site. *Clin. Neurophysiol.* 79:371–381.

Schreiber, T., and A. Schmitz. 2000. Surrogate time series. *Physica D* 142:346–382.

Takens, F. 1981. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence*, Vol. 898 of *Lecture notes in mathematics*, eds. D. A. Rand and L.-S. Young, 366–381. Berlin: Springer-Verlag.

Theiler, J. 1986. Spurious dimensions from correlation algorithms applied to limited time-series data. *Phys. Rev. A* 34:2427–2432.

Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. 1992. Testing for nonlinearity in time series: The method of surrogate data. *Physica D* 58:77–94.

---

[*] Consider as a further principle limitation, a periodic process with a very long period that exhibits no regularities within each period. We would need to observe at least two periods to appreciate the periodicity of the process. In principle, we need an infinite amount of data to rule out that we misinterpret periodic, and hence deterministic, processes as stochastic processes.