

1-26-2016

Chapter 5: Surrogation

Sara A. Myers

Follow this and additional works at: <https://digitalcommons.unomaha.edu/biomechanicsarticles>

 Part of the [Biomechanics Commons](#)

Surrogation

Sara A. Myers

CONTENTS

[Introduction](#)

[Nonlinearity](#)

[General Principles of Surrogation](#)

[Hypothesis Testing \(Discriminating Criterion\)](#)

[Linear Surrogate Methods](#)

[Algorithm 0](#)

[Algorithm 1 \(Fourier Transform\)](#)

[Algorithm 2](#)

[Iterated Amplitude-Adjusted Fourier Transform](#)

[Rejection of Null Hypothesis](#)

[Pseudoperiodic Surrogate Method](#)

[Discriminating Statistics for PPS](#)

[Summary](#)

[Exercises](#)

[Appendix 5.A: Knee Joint Flexion/Extension Angle \(Degrees\)](#)

[Original Series](#)

[PPS Surrogate Series](#)

[References](#)

Progress is impossible without change, and those who cannot change their minds cannot change anything.

George Bernard Shaw (1856–1950)

INTRODUCTION

One of the goals of time series analysis is to understand the underlying mechanisms that generate different dynamics for different time series. If a time series is not a product of random process, then we can assume that some kind of dynamics govern the time series. The question is what kinds of dynamics are controlling the time series. For nonlinear time series analysis, our focus is on nonlinear dynamics, and one of the goals is to characterize those dynamics by applying nonlinear tools. However, it is important to establish evidence of nonlinearity in a time series first in order to avoid obtaining possible spurious results by applying nonlinear tools to the system that does not contain nonlinearity. Second, nonlinearity is considered as one of the key features of time series that exhibit chaos, which has been shown to have a potential link with overall health of the biological system (Amato 1992; Buchman et al. 2001; Cavanaugh et al. 2010; Garfinkel et al. 1992; Goldstein et al. 1998; Orsucci 2006; Slutzky et al. 2001; Toweill and Goldstein 1998; Wagner et al. 1996). Therefore, in terms of detecting chaos in a time series, identifying the presence of nonlinearity in the system is essential.

NONLINEARITY

Nonlinear analysis characterizes the nonlinear properties of time series data. For these analysis tools to provide meaningful results, they must be applied to time series that contain nonlinear structures. When nonlinear tools are used on data without nonlinear structures, false results are obtained due to practical limitations of nonlinear measures. Therefore, it is important to establish the evidence of nonlinearity in a time series prior to applying nonlinear tools.

Before getting into a discussion of methods to identify possible nonlinearity in data, let us first define a nonlinear system. A nonlinear system is defined as a system that does not have a linear origin. This includes a system that may contain nonlinearity, but the underlying dynamics are linear. In this case, the presence of nonlinearity is caused by some measurement distortion, but it is originally generated by a linear stochastic process. We will look at such a case in detail later in this chapter. Furthermore, discussion in this chapter will be limited only to a stationary time series and will not include nonstationary stochastic processes. A time series is considered nonstationary if its distribution changes across time. In other words, the mean and variance of the time series change over different time intervals. Applying surrogate methods to nonstationary time series can lead to problems regarding the proper interpretation of results (Breakspear and Terry 2002; Palus 1996). For example, when the null hypothesis is rejected in a surrogate analysis, there is no way of knowing whether nonlinearity exists in the data or whether the data were generated by a nonstationary stochastic process. This problem was noted by Breakspear and Terry (2002) in their study of electroencephalographic (EEG) data (Breakspear and Terry 2002). The problem of nonstationarity was also highlighted by Peng et al. (1995) in the analysis of heart rate variability. Specifically, nonstationarity makes it difficult to determine whether the structure of the time series is the result of the dynamics of the

system or from changes in the external environment. Therefore, we will restrict our discussion to time series that are stationary.

There are two major approaches to identify the evidence of nonlinearity in a time series in general. The first approach involves the direct application of nonlinear measures (Kaplan and Glass 1995; Mitra et al. 1997), while the second approach involves the application of surrogate methods (Breakspear and Terry 2002; Dingwell and Cusumano 2000; Palus 1996; Stergiou et al. 2004). Methods commonly used for the first approach include the application of the correlation dimension or the largest Lyapunov exponent. The correlation dimension is a measure of self-similarity of a time series, while the largest Lyapunov exponent quantifies the exponential rate of divergence of nearby trajectories in the state space (see Chapters 3, 4, and 8). Both of these measures are applied to attractors reconstructed from an original time series in the state space. Attractors are often associated with nonlinearity and possibly chaotic dynamics. However, the use of these two popular nonlinear measures with experimental data can give spurious results. It has been reported that the correlation dimension of a time series with linear correlations can mimic low-dimensional behavior of the system by giving finite noninteger values (Osborne and Provencale 1989). Noise in a time series can cause the largest Lyapunov exponent to be positive, indicating the presence of chaos where there is none (Rapp et al. 1993). The use of other nonlinear measures besides correlation dimension and the largest Lyapunov exponent are also limited in terms of detecting nonlinearity in a time series since the probability distributions of those measures on time series with finite data length are unknown (Palus 1995; Pompe 1993; Prichard and Theiler 1995). Thus, applications of these nonlinear measures alone in detecting nonlinearity, possibly chaotic behavior in the system have been shown to be difficult (Miller et al. 2006; Schreiber and Schmitz 2000; Theiler and Rapp 1996). Moreover, applications of these nonlinear measures often involve subjective judgment of a researcher such as finding an appropriate scaling region or threshold value and lack in certainty. We particularly observe such procedures with methods like detrended fluctuation analyses and recurrence quantification analyses. To compensate for these weaknesses of the first approach of direct applications of nonlinear measures, the second approach of applying surrogate methods is often used. The second approach can be considered as an indirect approach in a sense that attempts to identify the evidence of nonlinearity by excluding that a time series has a linear origin through statistical hypothesis testing.

GENERAL PRINCIPLES OF SURROGATION

Surrogate methods were originally developed to prevent misdiagnoses of random stochastic processes from being characterized as chaotic dynamical processes or vice versa (Stergiou et al. 2004; Theiler et al. 1992; Theiler and Rapp 1996). They take a form of hypothesis testing to determine whether a given time series is consistent with a specific null hypothesis. The general procedure of a surrogate method is as follows (Figure 5.1). First, a null hypothesis is specified, and from the original time series an

ensemble of surrogate time series is generated that are consistent with this null hypothesis. The null hypothesis is typically what researchers want to show that is not true. An example of null hypothesis would be that the time series was generated by a linear stochastic process. If the data are nonlinear, the test statistic results will be the difference between the original and surrogate time series and the null hypothesis will be proven false. If the results are the same, the null hypothesis fails to be proven false, and the original time series is a linear stochastic process. These surrogate time series must preserve some properties (mean, variance, and/or power spectra), which correspond to the underlying null hypothesis. Then, discriminating statistics such as the correlation dimension are computed for both the original and the ensemble of surrogate time series. The values of the discriminating statistics between the original time series and the distribution of values of discriminating statistics obtained from the surrogate time series are compared. If the value of the discriminating statistics from the original time series does not fall within the distribution of the discriminating statistics of the surrogates, the null hypothesis should be rejected. As it is stated differently, if the results between original and surrogate are different, the null is rejected, and if the discriminating statistic is the same between original and surrogate, the null is accepted. The following tools are examples of discriminating statistics that have been used before for this purpose: the correlation dimension (Diks 1996; Grassberger and Procaccia 1983; Small 2005; Small and Judd 1998; Small and Tse 2002; Yu et al. 2000), the largest Lyapunov exponent (Kantz and Schreiber 1997; Wolf et al. 1985), approximate entropy (Miller et al. 2006; Pincus 1991), sample entropy (Lamothe et al. 2010, 2011; Rathleff et al. 2011), higher and cross moments (Keenan 1985; Tsay 1986), a simple skewed difference statistic (Theiler et al. 1992), Volterra polynomials (POL) (Barahona and Poon 1996; Kugiumtzis 1999), and the local average mapping (LAM) (Schreiber and Schmitz 1997).

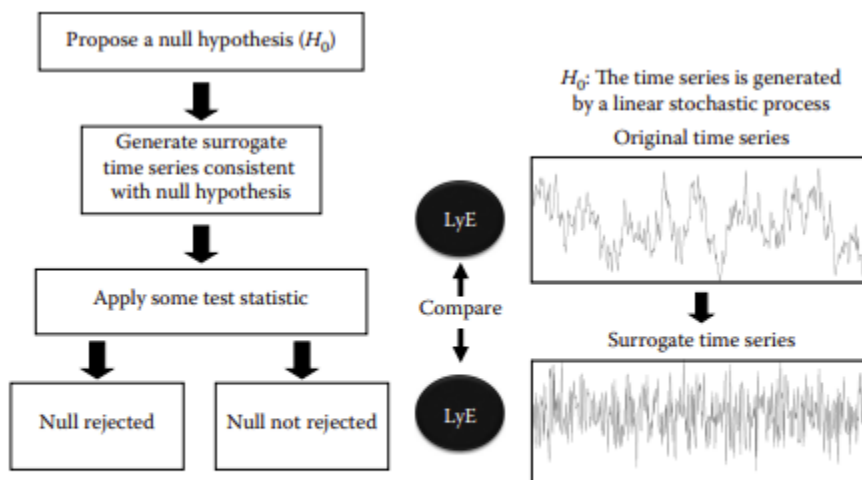


FIGURE 5.1 The general procedure of surrogate methods. These steps can be applied with one of the surrogate algorithms presented in this chapter. If the discriminant statistics are significantly different between the original time series and the surrogate time series, the null is rejected.

This approach to use surrogate methods for determining whether a given time series is consistent with a specific null hypothesis has been applied to identify the evidence of nonlinearity in many biological systems such as postural control, ECG, EEG, gait mechanics, and so forth (Acharya et al. 2005; Breakspear and Terry 2002; Buzzi et al. 2003; Chang et al. 1994; Cignetti et al. 2009; Collins and De Luca 1995; Costa et al. 2014; Ehlers et al. 1998; Govindan et al. 1998; Hausdorff et al. 1995; Ivanov et al. 1996; Janjarasjitt et al. 2008; Kugiumtzis 2001; Kunhimangalam et al. 2008; Kurz et al. 2008; Ladislao and Fioretti 2007; Little et al. 2006; Martinerie et al. 1998; Miller et al. 2006; Myers et al. 2013; Nurujjaman et al. 2009; Palus 1996; Porta et al. 2007; Preatoni et al. 2010; Rieke et al. 2003; Rombouts et al. 1995; Stam et al. 1997; Stergiou et al. 2004; Zhang et al. 2007; Zhao et al. 2008; Wurdeman et al. 2014). Table 5.1 details the surrogate methodology and whether determinism was found in these studies. Throughout the various methods and biological time series utilized in these studies, a common theme was the presence of nonlinear patterns and consistence of rejecting the null hypothesis in biological time series.

Since surrogate methods take the form of a null hypothesis testing, considerations must be given to the selection of discriminating statistics and discriminating criteria. Hypothetically, it should not matter which discriminating statistic is used. However, there are two different views on the selection of discriminating statistics. One view is that all nonlinear statistics should be able to detect the presence of nonlinearity by rejecting the null hypothesis at different significance levels. The alternative view is that the mismatch between a surrogate algorithm and discriminating statistics can lead to a spurious result. There are many different surrogate algorithms, which will determine which discriminating statistic is appropriate, based on the origination of the time series data. In general, a discriminating statistic must give consistent results for both surrogates and original time series if the null hypothesis is true. If the null hypothesis is not true, the discriminating statistic of the original time series should be different from the distribution of the discriminating statistics for its surrogates. The use of multiple discriminating statistics is encouraged in order to establish the evidence of nonlinearity in a time series since there may be cases where one discriminating statistic is not sufficient (Kugiumtzis 2001). For example, using parameters such as the largest Lyapunov exponent, correlation dimension, global false nearest neighbors, average mutual information and others on both simulated and actual EEG data, Kugiumtzis (2001) argued that different nonlinear methods characterize different aspects of data. However, if there is nonlinearity in the data, those different nonlinear methods should be able to detect the presence of nonlinearity by rejecting the null hypothesis at different significance levels respectively. On the other hand, it has been pointed out that the mismatch between surrogate algorithm and discriminating statistics can occur and can lead to a spurious result (Small et al. 2001, 188101). Therefore, considerations should be given to the right match between a specific surrogate algorithm and discriminating statistics (Zhang et al. 2007). We will show some examples of mismatch between a discriminating statistics and surrogate algorithm later in this chapter.

TABLE 5.1

A Selection of Studies in the Literature That Have Implemented Surrogate Methods in Biological Signals

Article	Measured Behavior	Surrogate Procedure	Discriminating Criterion	Rejecting the Null, Results
Collins and De Luca (1995)	Posture	Algorithm 0	Scaling exponent	Yes , correlations in the time series are due to underlying dynamic processes.
Stergiou et al. (2004)	Gait	Algorithm 1	Lyapunov exponent (LyE)	Yes , gait variability is deterministic in patients with Parkinson's even without medication.
Stam et al. (1997)	EEG	Algorithm 1	Correlation coefficients	Yes , it may be deterministic; however, the test is not definitive.
Ehlers et al. (1998)	EEG	Algorithm 1	Slope asymmetry, correlation dimension, one percent radius, redundancy, Kaplan's δ/ϵ .	Yes , EEG is not modeled effectively as linearly filtered Gaussian noise.
Ivanov et al. (1996)	EEG	Algorithm 1	Correlation dimension, correlation parameter (P_c)	Yes , data differed significantly from linear stochastic time series.
Martinerie et al. (1998)	Intracranial recordings	Algorithm 1	Correlation density	Yes , data differed significantly from stochastic fluctuations.
Breakspear and Terry (2002)	EEG	Algorithm 1	Phase synchrony	Yes , phase coherence cannot be described by purely linear methods for a small number of EEG epochs.
Govindan et al. (1998)	ECG	Algorithm 1	Correlation dimension	Yes , the results indicate the presence of nonlinear structure in the time series.
Rombouts et al. (1995)	EEG	Algorithm 2	Correlation dimension	Yes , evidence for nonlinearity is found yet EEG dynamics are low dimensional.
Chang et al. (1994)	Spinal cord monosynaptic reflex	Algorithm 1	Local flow, local dispersion, and nonlinear prediction	Yes , deterministic structure was found in 2/4 of the time series for the decerebrate state.
Palus (1996)	EEG	Algorithm 1	Slopes of redundancy	Yes , the hypothesis of a linear stochastic process is rejected.

(Continued)

TABLE 5.1 (Continued)
A Selection of Studies in the Literature That Have Implemented Surrogate Methods in Biological Signals

Article	Measured Behavior	Surrogate Procedure	Discriminating Criterion	Rejecting the Null, Results
Miller et al. (2006)	Gait	Algorithm 1, PPS	LyE, approximate entropy (ApEn)	Yes , significant differences were found for both LyE and ApEn values.
Hausdorff et al. (1995)	Gait	Algorithm 0	Scaling exponent α from DFA	Yes , fluctuations in the stride intervals were unlike those obtained from an uncorrelated random walk.
Buzzi et al. (2003)	Gait	Algorithm 1	LyE	Yes , LyE were found to be significantly different indicating that the time series may reflect nonlinear processes.
Zhang et al. (2007)	ECG, speech	PPS	Correlation coefficient, average cycle divergence rate (ACDR, analogous to LyE), and clustering coefficient	Yes , differences were observed in all measures.
Rieke et al. (2003)	EEG from epilepsy patients	IAAFT	Coarse-grained flow average (Kaplan and Glass 1995)	Yes , nonlinearity is evident in the EEG.
Kugiumtzis (2001)	Simulations, EEG, exchange rate data	Algorithm 2, IAAFT	Volterra polynomials, local average mapping, mutual information, LyE, correlation dimension, simple nonlinear averages	Yes , but results depend on the parameter examined suggesting surrogate testing should be performed using multiple methods and parameters.
Kurz et al. (2008)	Robot gait	PPS	LyE	Yes , support the notion that the deterministic variations present in gait may be partly governed by the intrinsic mechanical dynamics of the locomotive system.
Kunhimangalam et al. (2008)	EEG	Algorithm 1	Correlation dimension, LyE	Yes , in case of LyE of EEG data, the change due to surrogating is small suggesting that it is not representing the system complexity properly but there is a marked change in the case of correlation dimension value due to surrogating.

(Continued)

TABLE 5.1 (Continued)

A Selection of Studies in the Literature That Have Implemented Surrogate Methods in Biological Signals

Article	Measured Behavior	Surrogate Procedure	Discriminating Criterion	Rejecting the Null, Results
Janjarsajitt et al. (2008)	Neonatal EEG	Algorithm 1	Correlation dimension	Yes , dimensional complexity cannot be rejected for active sleep but can be for other sleep states.
Acharya et al. (2005)	EEG during sleep	Algorithm 1	Correlation dimension, fractal dimension, LyE, ApEn, Hurst exponent, phase space plot, recurrence plots	Yes , correlation dimension and ApEn show that the original data contain nonlinear features.
Nurujiaman et al. (2009)	EEG	Algorithm 0	Probability distribution function, Hurst exponent	Yes , the activity of an epileptic brain is non-Gaussian in nature, but not normal brain activity.
Little et al. (2006)	Speech	IAAFT	Time-delayed mutual information	Yes , the assumption of dynamical nonlinearity and/or non-Gaussianity is supported but linear Gaussian theory is also a weak approximation.
Porta et al. (2007)	Short-term heart beat intervals	Algorithm 0, Algorithm 1, Algorithm 2, IAAFT	Porta's (P) method, modified Kantz and Schreiber's (MKS) method, Sugihara and May's (SM) method	No , short-term heart period variability at rest was mostly linear.
Ladislao and Fioretti (2007)	Posture	IAAFT	LyE	Yes , postural sway derives from a process exhibiting weakly chaotic dynamics.
Cignetti et al. (2009)	Cross country skiing joint angles	Algorithm 1	LyE, correlation dimension	Yes , temporal variations in the data had a deterministic origin.
Zhao et al. (2008)	Human cardiac data	PPS	Complexity, correlation dimension	Yes , bounded aperiodic determinism exists in ECG and pulse time series but dynamic noise means that it is not possible to conclude the underlying system is chaotic.

(Continued)

TABLE 5.1 (Continued)
A Selection of Studies in the Literature That Have Implemented Surrogate Methods in Biological Signals

Article	Measured Behavior	Surrogate Procedure	Discriminating Criterion	Rejecting the Null, Results
Preatoni et al. (2010)	Race walking joint angles and ground reaction forces	PPS	Sample entropy (SampEn)	Yes , SampEn was significantly lower for all measures than the surrogates.
Myers et al. (2013)	Gait (lower limb joint angles)	PPS	LyE	Yes , significant differences were found for the ankle and knee joint angle time series at baseline and ankle postvascular occlusion.
Costa et al. (2014)	Glucose levels	Algorithm 0	SampEn	Yes , the 100 shuffled surrogate time series were consistently more entropic than the original first difference time series.
Wurdeman et al. (2014)	Gait (ankle motion)	PPS	LyE	Yes , results revealed the presence of underlying deterministic structure.

HYPOTHESIS TESTING (DISCRIMINATING CRITERION)

In this section, hypothesis testing will be discussed. The first question that needs to be answered after computing discriminating statistics for both original time series and its surrogates is what criteria should be used to determine whether the null hypothesis should be rejected or not. One way to conduct a hypothesis test is by using a parametric criterion. The mean (μ_H) and standard deviation (σ_H) of discriminating statistics for surrogate data are used to calculate the significance S with a unit of “sigma,” which in turn is used to construct a confidence level of inference (Theiler et al. 1992):

$$S = \frac{|Q_D - \mu_H|}{\sigma_H} \quad (5.1)$$

For example, the rejection of the null hypothesis at the 95% level of confidence is indicated by significance of about two “sigmas.” The assumption for this criterion is that the distribution of discriminating statistics is Gaussian. However, it has been shown that the distributions of many nonlinear measures do not follow a Gaussian distribution (Schreiber and Schmitz 2000; Theiler et al. 1992).

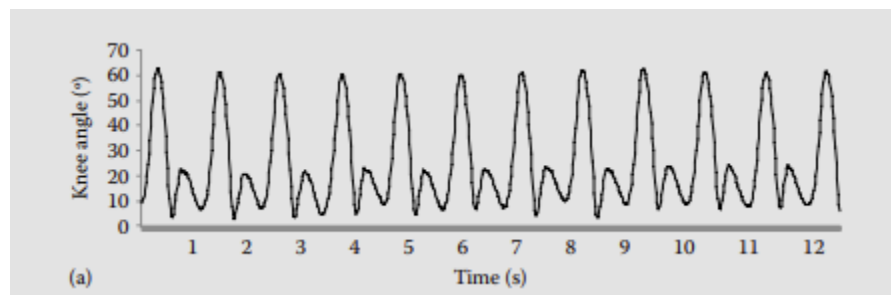
Therefore, another criterion, the rank-order criterion is often used in the literature since it is more robust in terms of defining significance. This criterion namely examines the ranks of discriminating statistics of an original time series and surrogates. Suppose N surrogate time series were generated from the original time series and the discriminating statistics Q were computed for each surrogate time series and the original time series. Then, there are $(N + 1)$ Q 's in total. Let Q_D be the discriminating statistics value for the original time series, and Q_1, Q_2, \dots, Q_N be the discriminating statistic's values for the surrogates. Now, all these $(N + 1)$ discriminating statistic's values are ranked in an increasing order. If the original time series is generated by a process which is consistent with the null hypothesis, the probability of Q_D to be the smallest or the largest will be $1/(N + 1)$. According to the rank-order criterion, the null hypothesis is rejected when Q_D is the smallest or the largest values among $(N + 1)$ Q 's. For a one-sided test, $1/(N + 1)$ is regarded as a false rejection rate while $2/(N + 1)$ for a two-sided test. Therefore, in order to conduct hypothesis testing at 95% significance level, 19 surrogates must be generated for a one-sided test and 39 surrogates for a two-sided test.

Let us look at an example of a surrogate test using sample entropy (SampEn), which is a measure of regularity as a discriminating statistics (Example Box 5.1). Nineteen surrogate time series were generated from an original time series. Twenty sample entropy (SampEn) values were computed for the original time series and the nineteen surrogates. After ranking those 20 SampEn values, we may reject the null hypothesis at 95% confidence level if the SampEn value of the original time series was the smallest value (Figure 5.2; Appendix 5.A). In this case, the probability of a false rejection is 5%.

EXAMPLE BOX 5.1 EXAMPLE OF SURROGATE TESTING

This example uses the data from Appendix 5.A to go through the steps of surrogate testing. For space purposes, only the original and one surrogate time series are included. For this example, sample entropy (SampEn) will be used as the discriminant, which is a measure of regularity. The data are a continuous knee flexion/extension time series from walking.

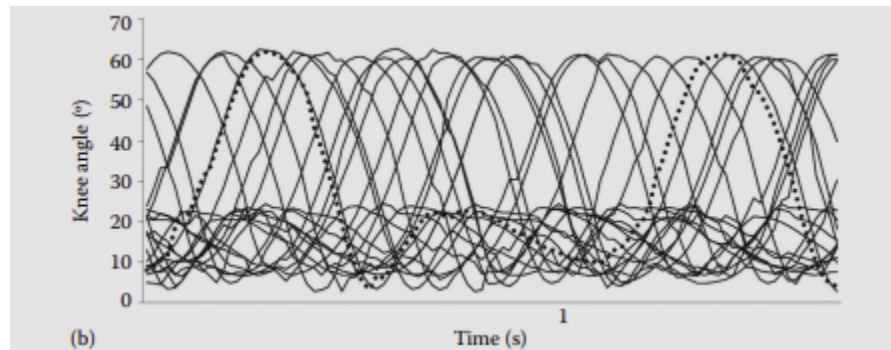
1. *Plot the data and identify the null hypothesis.* It is important to plot the data to quickly inspect that the data are as expected (no missing points, proper length, etc.). Identifying the null hypothesis will determine which surrogate algorithm should be implemented. In our example, a knee flexion/extension time series is used that has inherent periodicity due to the repetition of gait cycles. The repeating cycles of the time series are evident in the graph in the following. Therefore, an appropriate null hypothesis would be that our time series is consistent with a periodic orbit perturbed by uncorrelated noise. To test this hypothesis, we will need to use the pseudoperiodic surrogate (PPS) method.



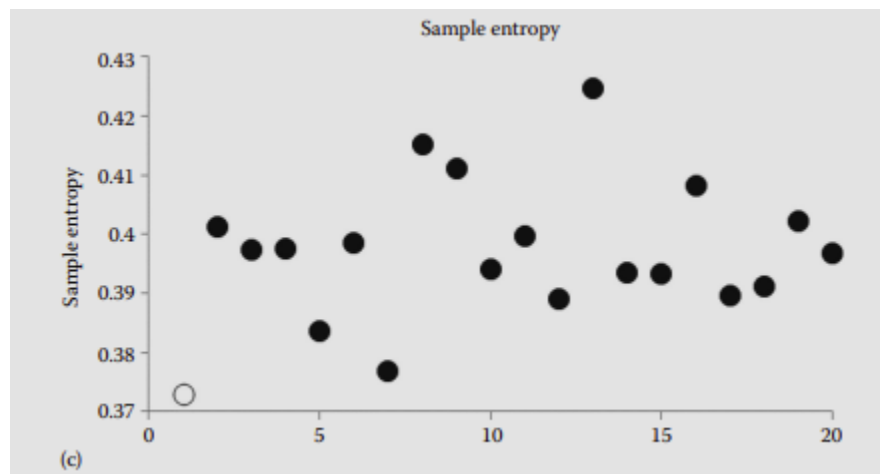
2. *Calculate and choose parameters for surrogate method.* As mentioned earlier, the PPS algorithm is appropriate for this data. To implement the PPS algorithm, the state space must be reconstructed. Thus, the embedding dimension and the time lag from the original series must be calculated to generate the surrogate time series. Another parameter that must be determined is the noise radius, which defines the amount of noise in a surrogate. The proper p is selected to maximize the number of short segments that are the same for the original time series and the surrogate. For this particular time series, the embedding dimension is 6, time lag is 10 and the noise radius is 0.5.

3. . Using the same parameters, a series of 19 surrogates should be calculated from the original time series. Each of these surrogates should be plotted to ensure that it was generated correctly. In the example, the dotted line is the original time series and the 19 surrogates are all plotted

in black. The surrogates resemble the original time series, which is correct. Only 1.5 s of data is shown for clarity purposes.



4. *Calculate the discriminating statistic.* SampEn, a measure of regularity is our chosen discriminating statistic. To calculate SampEn, some parameters need to be chosen first. Briefly, SampEn calculates the logarithmic probability that run patterns that are close (within tolerance r) for m observations remain close (with the same tolerance r) on the next incremental comparison. For this example, $m = 2$ and $r = 0.2$ will be used based on previous studies using entropy calculations on kinematic data (Georgoulis et al. 2006; Myers et al. 2012; Yentes et al. 2013). Next, all SampEn values from the original and surrogate should be plotted. In our example, the open circle is the original time series and the solid circles are the surrogates. It is clear in the graph that the original time series has the lowest entropy value and therefore is not within the distribution of the surrogates. The SampEn values for the original and surrogate time series from Appendix 5.A are 0.373 and 0.401, respectively. Keep in mind, however, that each time a surrogate series is generated, it adds dynamics noise. So, a surrogate you generate from the example data will not match the provided surrogate series. However, you can use the example surrogate series to test whether your SampEn calculations are correct using the same parameters used in this example.



5. *Test for significance between the discriminating statistic values of the original time series and the series of surrogates.* There are multiple ways that determine whether the discriminating statistic values of the original time series are different from the surrogates. One of the most common ways as used in this example is a simple rank test. Based on the plot from step 4, it is clear that we can reject the null hypothesis at 95% confidence level because the SampEn value of the original time series is the smallest value. This means that the regularity of the original time series is significantly different than that of the surrogates. That would leave a 5% probability that our rejection of the null is false. Remember the null hypothesis was that our original knee joint flexion/extension time series is consistent with a periodic orbit perturbed by uncorrelated noise. Since we rejected the null hypothesis, we have concluded that our time series is not an orbit with noise, but contains identifiable dynamics.

LINEAR SURROGATE METHODS

As we have mentioned earlier in the chapter, there are many different surrogate algorithms. In this section, the most commonly applied surrogate algorithms, which are called linear surrogate methods, will be discussed. Linear surrogate methods were originally developed by Theiler and colleagues (Theiler et al. 1992; Theiler and Rapp 1996), and later improvements were made by Schreiber et al. (Schreiber and Schmitz 2000). These linear surrogate methods are applied to a stationary irregular time series without any long-term trend or periodicity. If trends, especially periodic trends, are present (i.e., gait kinematic data), other algorithms are more appropriate as we will see later in the chapter. They are known as Algorithm 0, Algorithm 1 or Fourier transform surrogate, and Algorithm 2 or amplitude adjusted Fourier transform (AAFT), and each algorithm deals with a different null hypothesis.

Algorithm 0

First, let us examine Algorithm 0, which is used to test whether there is any evidence that a time series has any dynamics at all. Therefore, surrogate time series generated by Algorithm 0 should be consistent with the null hypothesis of an independent and identically distributed (IID) noise with unknown (random) mean and variance. Such a surrogate time series is generated by randomly shuffling an original time series, destroying temporal correlations (Example Box 5.2). Since this method permutes temporal order of a time series without replacement, a surrogate time series preserves the same probability distribution as the original time series (Figure 5.3).

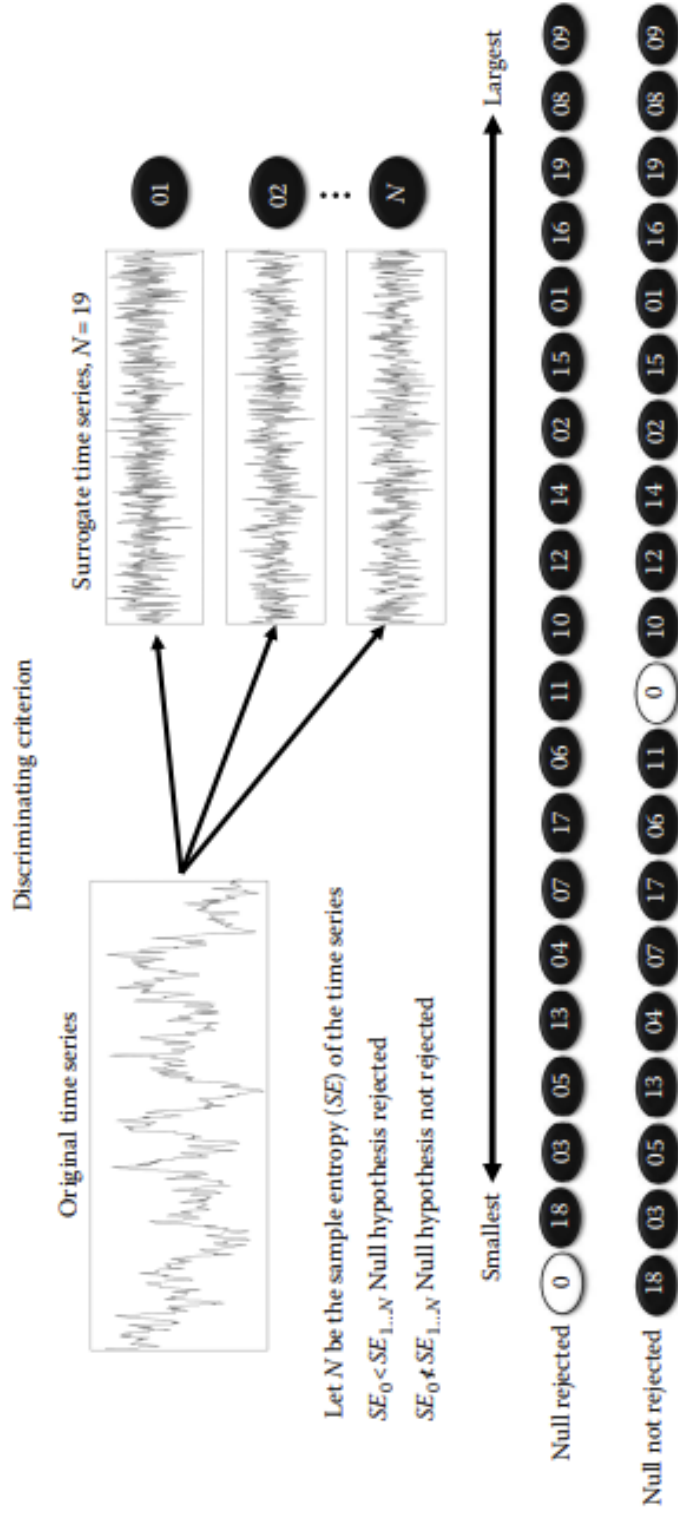


FIGURE 5.2 An example of surrogate testing with the sample entropy (SampEn) as the discriminant statistic. A total of 19 surrogates are generated and SampEn is calculated for the original and each surrogate. The process of generating 19 surrogate time series is the same, regardless of which surrogation algorithm is determined to be appropriate for the data. A rank order test is used to verify that the original time series has the lowest SampEn. The null hypothesis would be rejected if SampEn of the original series has the lowest value of all of the surrogate series values. The null would be accepted if the SampEn value for the original is not the lowest value. (Adapted from Pincus, S.M. et al., *J. Clin. Monit.*, 7(4), 335, 1991.)

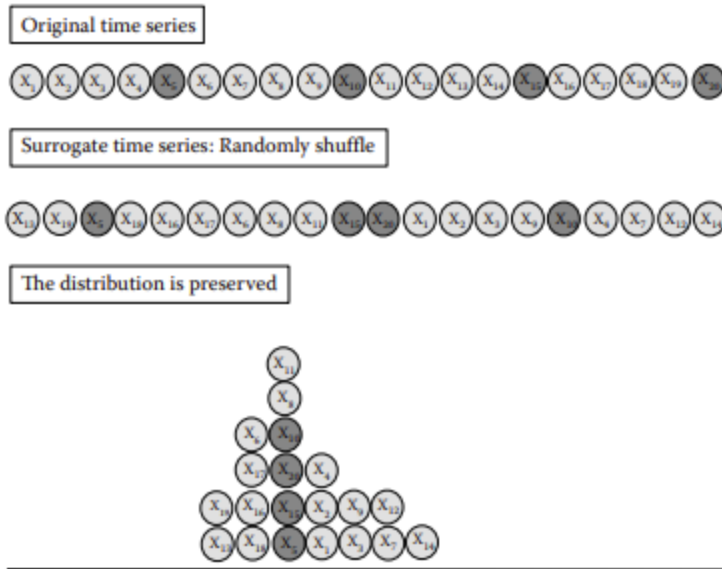
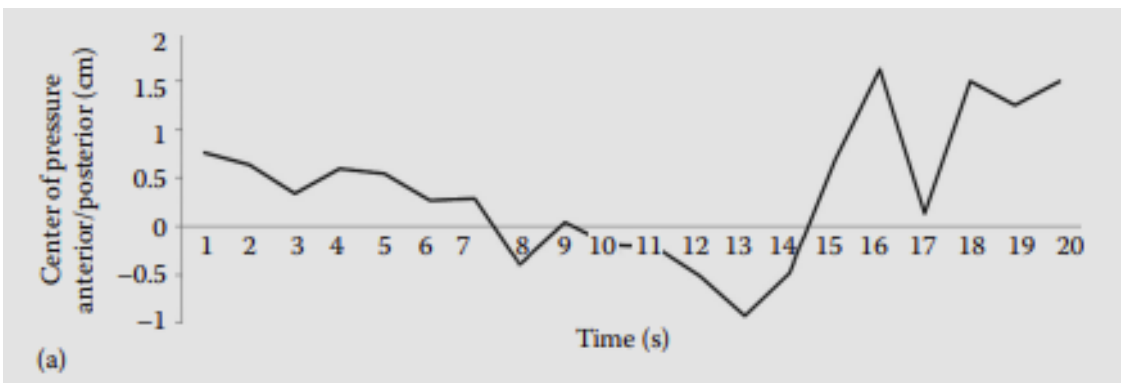


FIGURE 5.3 A graphical representation of the Theiler et al. Algorithm 0 surrogate. The original and surrogate (created using the Algorithm 0) time series have the same frequency distribution, even though they now have completely different time structure.

EXAMPLE BOX 5.2 NUMERICAL EXAMPLE OF THEILER ET AL. ALGORITHM 0 SURROGATE METHOD

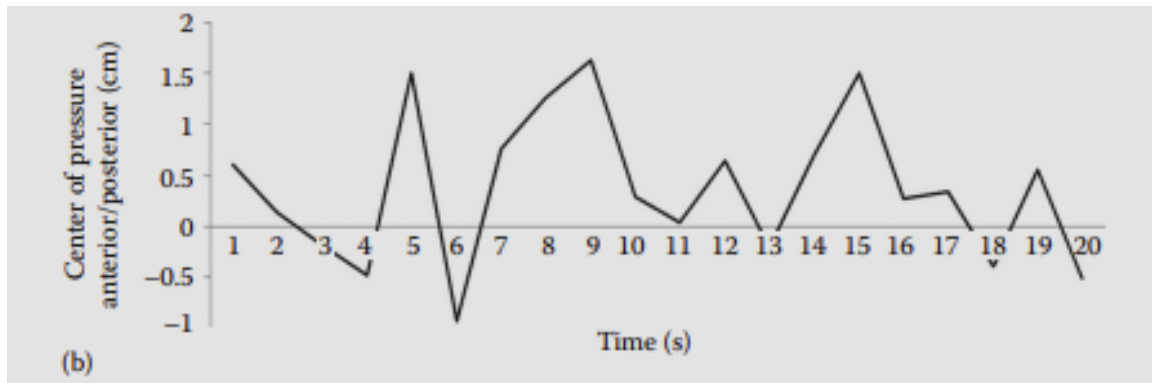
This example uses the following 20 data points from 20 s of standing posture sampled at 1 Hz. {0.76, 0.64, 0.34, 0.60, 0.55, 0.27, 0.29, -0.39, 0.04, -0.19, -0.21, -0.51, -0.93, -0.48, 0.69, 1.63, 0.13, 1.5, 1.26, 1.5}.

1. Plot the data and identify the null hypothesis. The null hypothesis for A0 is that the data are consistent with independent and identically distributed noise with unknown mean and variance.



2. Randomly shuffle the original time series. This can be done with the randn function in MATLAB®. After shuffling, our time series is as follows:

{0.60, 0.13, -0.19, -0.48, 1.5, -0.93, 0.76, 1.26, 1.63, 0.29, 0.04, 0.64, -0.21, 0.69, 1.5, 0.27, 0.34, -0.39, 0.55, -0.51}



This step would be repeated so that a total of at least 19 surrogates are generated to be able to reject the null with 95% confidence that the original time series is not within the distribution of the surrogates.

3. Calculate the discriminant statistic. For this example, the discriminant is not calculated because the 20 points data series would be too short to provide meaningful results, but this would be the next step.

4. Test for significance between the discriminating statistics values of the original time series and the series of A0 surrogates.

Suppose, surrogate time series are generated by Algorithm 0 and some discriminating statistic is applied, and the null hypothesis of the test is rejected. What does this indicate? The rejection of the null hypothesis of IID noise implies the evidence of some structure in a time series. If the null is accepted, it means that the original time series has no determinable dynamics, that is, there are no correlations among the data points of the time series. Thus, the next step is to test whether this structure has a linear origin or not by using Algorithm 1 or Algorithm 2.

Algorithm 1 (Fourier Transform)

The null hypothesis in Algorithm 1 addresses that a time series is generated from a linear Gaussian stochastic process, which is equivalent to a linearly filtered noise. Algorithm 0 involves random shuffling of a time series. This process destroys both linear and nonlinear correlations among data points. In order to preserve linear correlations, shuffling must be done in the frequency domain. Therefore, the first step of generating surrogate time series by Algorithm 1 is to perform a Fourier transform, that is, taking data from the time domain to the frequency domain, where Fourier transformed data have corresponding amplitudes and phases. The next step is to shuffle the phases while preserving the amplitudes in the frequency domain. The final step is to take the inverse Fourier transform in order to transform back the data to the

time domain from the frequency domain (Figure 5.4; Example Box 5.3). A surrogate time series generated by Algorithm 1 preserves the linear correlations, which are represented by the discrete Fourier power spectrum as the original time series while any additional structure should be destroyed. However, surrogate time series does not preserve the probability distribution of the original data, which can lead to a false rejection of the null hypothesis. There is an especially obvious discrepancy in the probability distribution between a coarsely grained time series and its surrogate generated by Algorithm 1.

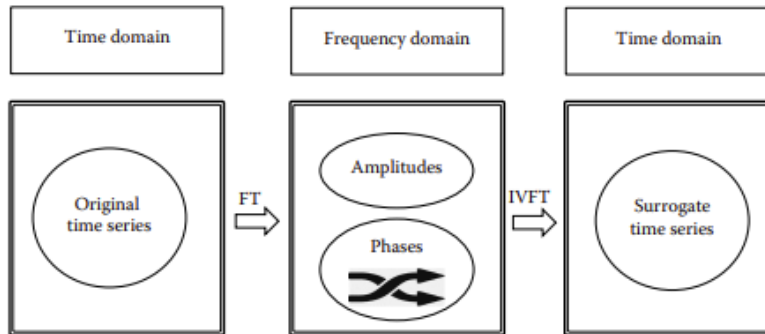
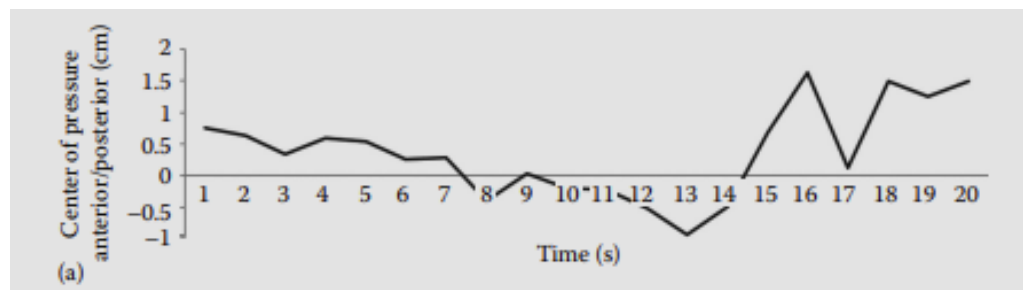


FIGURE 5.4 The original time series undergoes fast Fourier transformation to move the series from the time domain to the frequency domain. Then phase randomization is performed. Inverse Fourier transformation provides the resulting surrogate time series.

EXAMPLE BOX 5.3 NUMERICAL EXAMPLE OF THEILER ET AL. ALGORITHM 1 SURROGATE METHOD

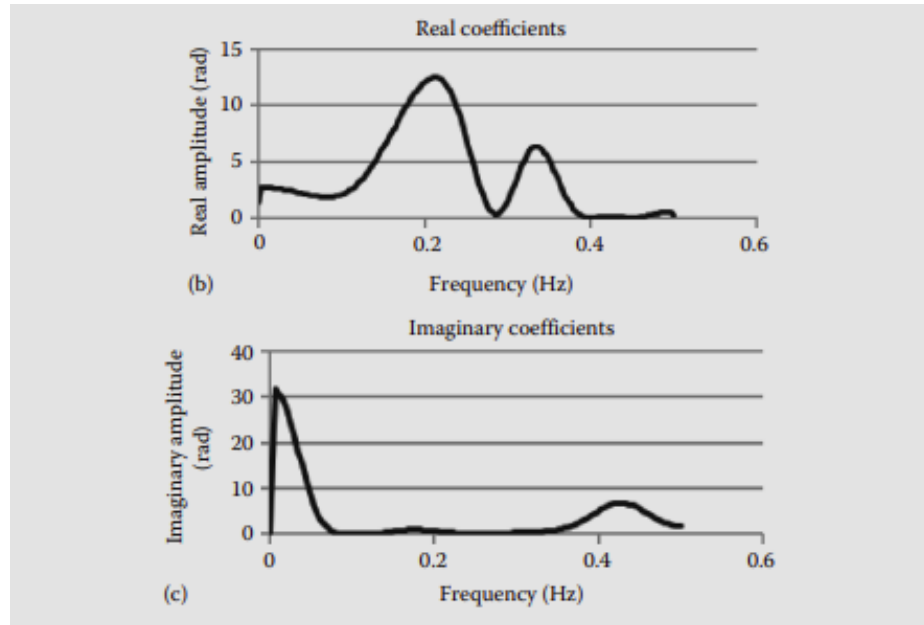
This example will use the same 20 data point time series as Example Box 5.2, but continue through the steps for the A1 algorithm.

1. Plot the data and identify the null hypothesis. The appropriate null hypothesis for the A1 algorithm is that the time series is generated from a linear Gaussian stochastic process equivalent to a linearly filtered noise. The time series is: {0.76, 0.64, 0.34, 0.60, 0.55, 0.27, 0.29, -0.39, 0.04, -0.19, -0.21, -0.51, -0.93, -0.48, 0.69, 1.63, 0.13, 1.5, 1.26, 1.50}.

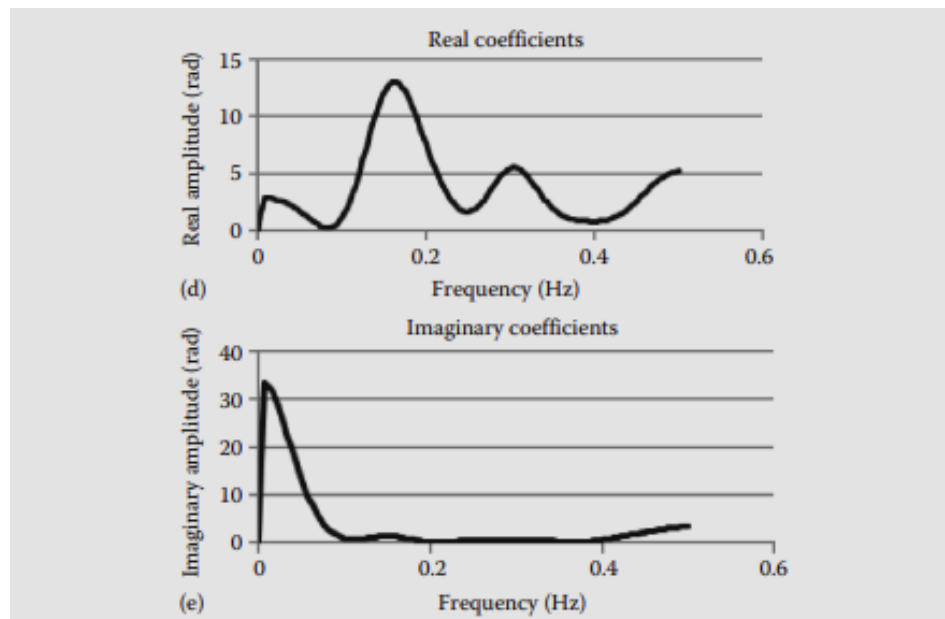


2. Perform a Fourier transform. Since randomly shuffling the time series removes linear and nonlinear correlations among data points, shuffling must be done in the frequency domain. This can be done with the fft

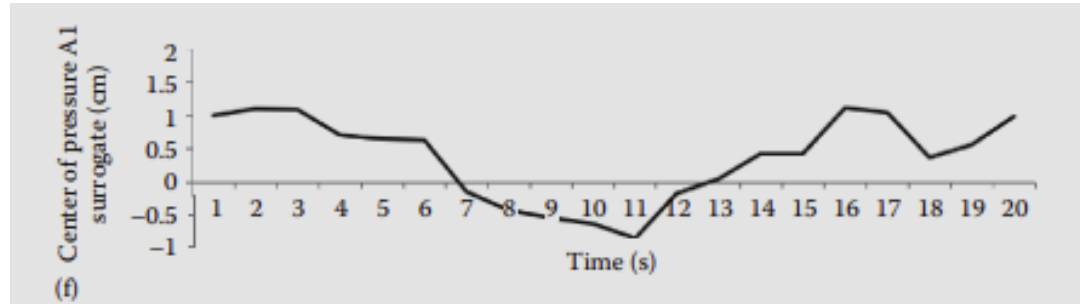
function in MATLAB®. After the transform, the series is “moved” from the time domain to the frequency domain. The result is a list of coefficients representing real and imaginary parts. The coefficients represent the amount that each frequency contributes to the overall signal. Plots of the real and imaginary parts of each Fourier coefficient are as follows:



3. Shuffle the phases while preserving the amplitudes. The graphs in the following are the real and imaginary amplitudes of the Fourier coefficients after the phases have been shuffled. It is clear that the frequency contributions are different from the original.



4. Take the inverse Fourier transform. This step takes the data back into the time domain. The surrogate time series is {1.00, 1.10, 1.09, 0.72, 0.65, 0.64, -0.14, -0.42, -0.52, -0.63, -0.83, -0.16, 0.06, 0.44, 0.44, 1.11, 1.05, 0.37, 0.57, 0.98}. The new time series is plotted:



5. Calculate the discriminant statistic. For this example, we will not calculate the discriminant because the 20 data point time series would be too short to provide meaningful results, but this would be the next step.
6. Test for significance between the discriminating statistics values of the original time series and the series of A1 surrogates.

Let us look at an example. An autoregressive time series with length $N = 2000$ called AR is generated by the following Equation 5.2:

$$X_{t+1} = 0.9X_t + \delta \quad (5.2)$$

where δ is an error term.

AR is generated by the linear Gaussian stochastic process. Thus, the null hypothesis of Algorithm 1 should not be rejected. Another time series called AR2 is created by making AR coarse (removing the precisions). Even though little difference is visually observed between the plots of the two time series, the number of unique values each time series has is quite different (Figure 5.5a). The number of unique values of AR is the same as its data length (2000) while AR2 only has 71 unique values, even though it is still 2000 data points long. Nineteen surrogate time series are generated from AR and AR2, respectively, but again it is hard to differentiate between the surrogate time series of each time series by a visual inspection (Figure 5.5b). However, the results of the null hypothesis testing using the SampEn as discriminating statistics are different between AR and AR2 (Figure 5.5c). For AR, the SampEn value of the original time series falls within the distribution of the SampEn values of the surrogates, and this is what we have expected, that is, not rejecting the null hypothesis. However, the SampEn value of AR2 is the smallest SampEn values among the SampEn value of its surrogate time series. As a result, we may reject the null hypothesis at 95% confidence level. The discrepancy in the results between AR and AR2 hypothesis testing seems to be due to

an increase in the number of values in surrogate time series generated from AR2. The SampEn value of AR2 is much lower than the SampEn value of AR since AR2 has fewer numbers of unique values than that of the AR2. However, the process of generating surrogate time series using the Fourier transform preserves the frequency distribution (Figure 5.6), but not the probability distribution (Figure 5.7). Thus, Algorithm 1 increases the number of unique values in surrogate time series generated from AR2: they all have 2000 unique values. This jump in the increase in the number of values in the surrogate time series is reflected in the value of SampEn of surrogates generated from AR2. It created discrepancies in the SampEn values between AR2 and its surrogates, causing the false rejection of the null hypothesis. In order to avoid statistical bias, it is better that the probability distribution of an original time series and surrogate are the same. Therefore, for a coarsely grained time series, Algorithm 1 may not be an appropriate surrogate method to apply.

Again, let us suppose surrogate time series are generated by Algorithm 1 and some discriminating statistics are applied, and the null hypothesis is rejected. What does the rejection of the null hypothesis of a linearly filtered noise imply? It may be indicating evidence of a more complex structure in a time series such as nonlinearity. However, even though a time series contains nonlinearity, it may be due to some distortion caused by a measurement procedure, and the origin of underlying dynamics may be linear stochastic. To test such a case, Algorithm 2 can be used.

Algorithm 2

Algorithm 2 deals with the null hypothesis of a static and monotonic nonlinear filter. Here, a measurement procedure is regarded as a nonlinear filter and enhances fluctuations in a system with linear dynamics as this linear stochastic system passes through the filter. Therefore, the time series contains nonlinearity, which enhances the system with linear dynamics as it goes through the filter, but this nonlinearity is not in the dynamics. This process is known as nonlinear distortion and is described by the following equation where $h(\cdot)$ is an invertible, static measurement function, $\{x_n\}$ is the underlying linear stochastic process, and $\{y_n\}$ is the observed data:

$$y_n = h(x_n) \tag{5.3}$$

A time series in test is assumed to have failed a test for originating from a Gaussian process by Algorithm 1 and therefore shown that it is not generated by a linear Gaussian stochastic process. However, if the invertible, static measurement function may be able to explain this deviation from the Gaussian distribution, then it may be shown that the underlying dynamics are a linear Gaussian stochastic process. Algorithm 2 deals with such a scenario. Since the underlying dynamics are considered as a linear Gaussian stochastic process, part of the procedure to generate surrogate time series by Algorithm 2 also involves Algorithm 1. It is necessary to rescale the time series first so that the distribution will be Gaussian, and we can apply Algorithm 1. Example Box 5.4

goes through the procedure of generating a surrogate time series by Algorithm 2 by using a simple 10 data point time series.

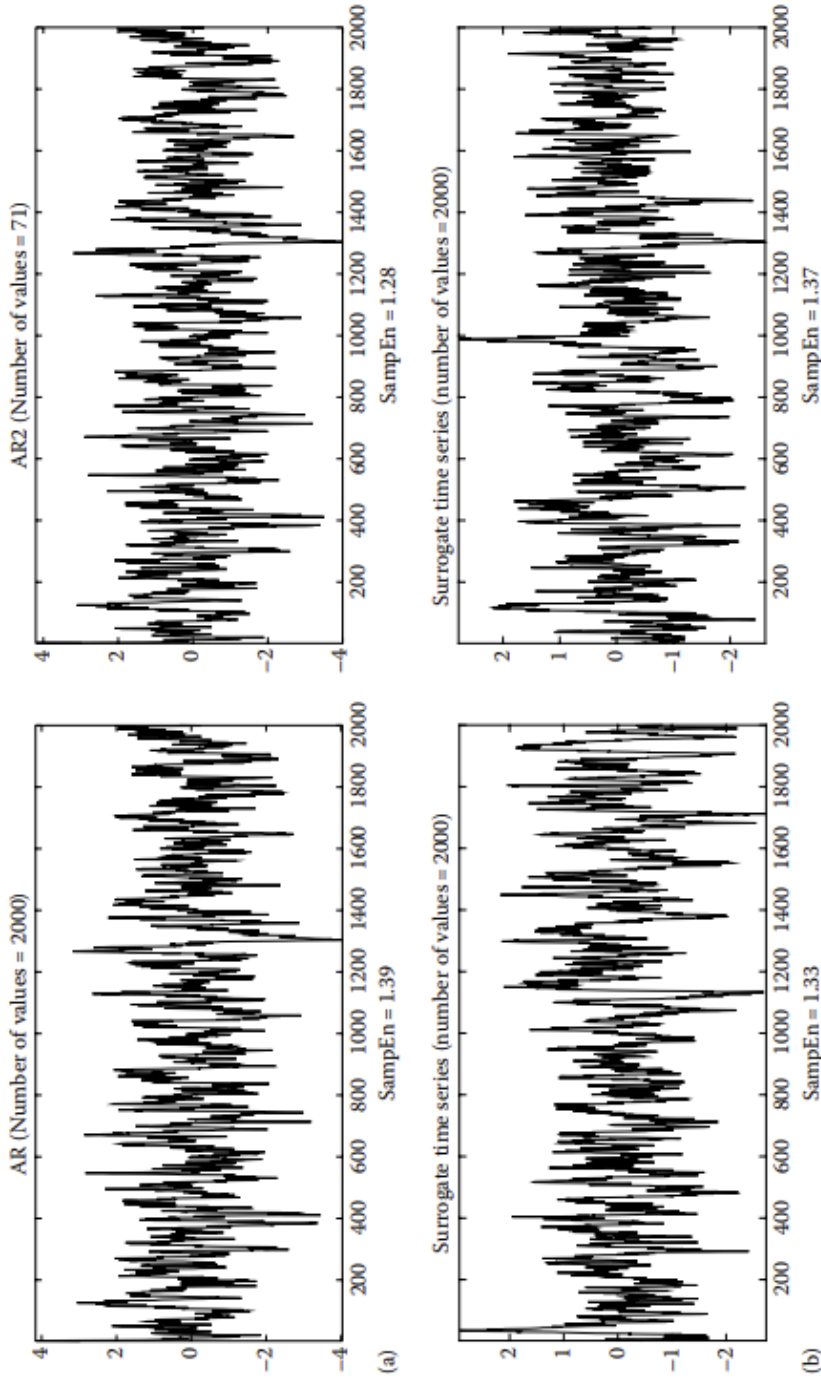


FIGURE 5.5 (a) Plots of a linear Gaussian stochastic process (AR) and a course AR (AR2), in which the precisions were removed. There is little visual difference, but the number of unique values in AR is 2000, while AR2 only has 71. (b) The surrogate time series of AR and AR2 are also visually similar. However, Algorithm 1 increases the number of unique values in AR2 to 2000. (Continued)

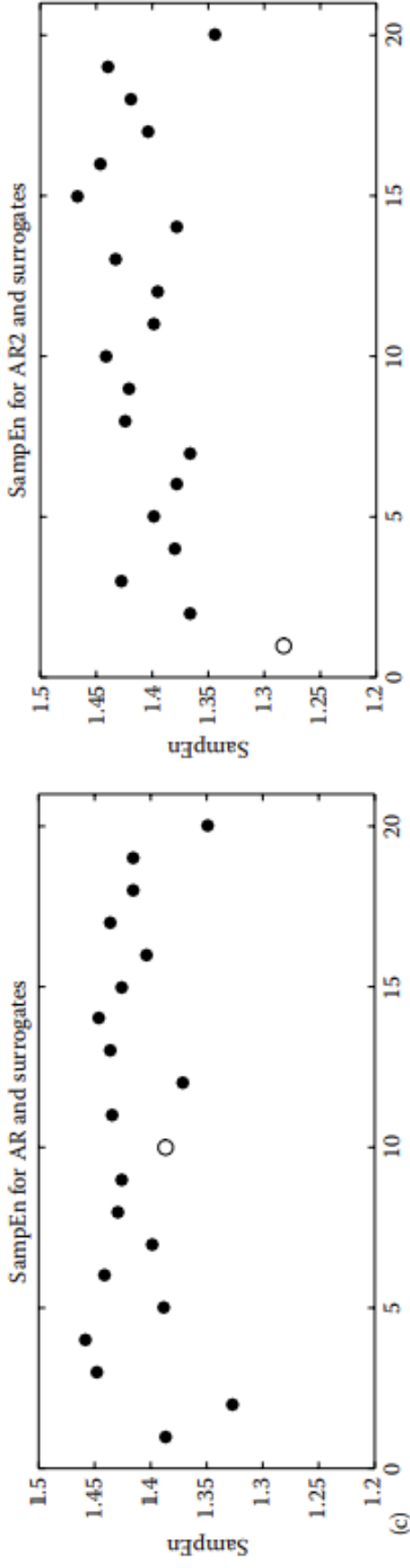


FIGURE 5.5 (Continued) (c) The open circle represents the SampEn of the original time series and the solid circles are the surrogate SampEn values on each graph. Results of testing the null hypothesis using SampEn as a discriminant statistic are different when using AR and AR2. This difference occurs because the Fourier transform increases the number of unique values in the surrogate time series generated from AR2. This increase resulted in discrepancies in the SampEn values between AR2 and its surrogates, leading to a false rejection of the null hypothesis.

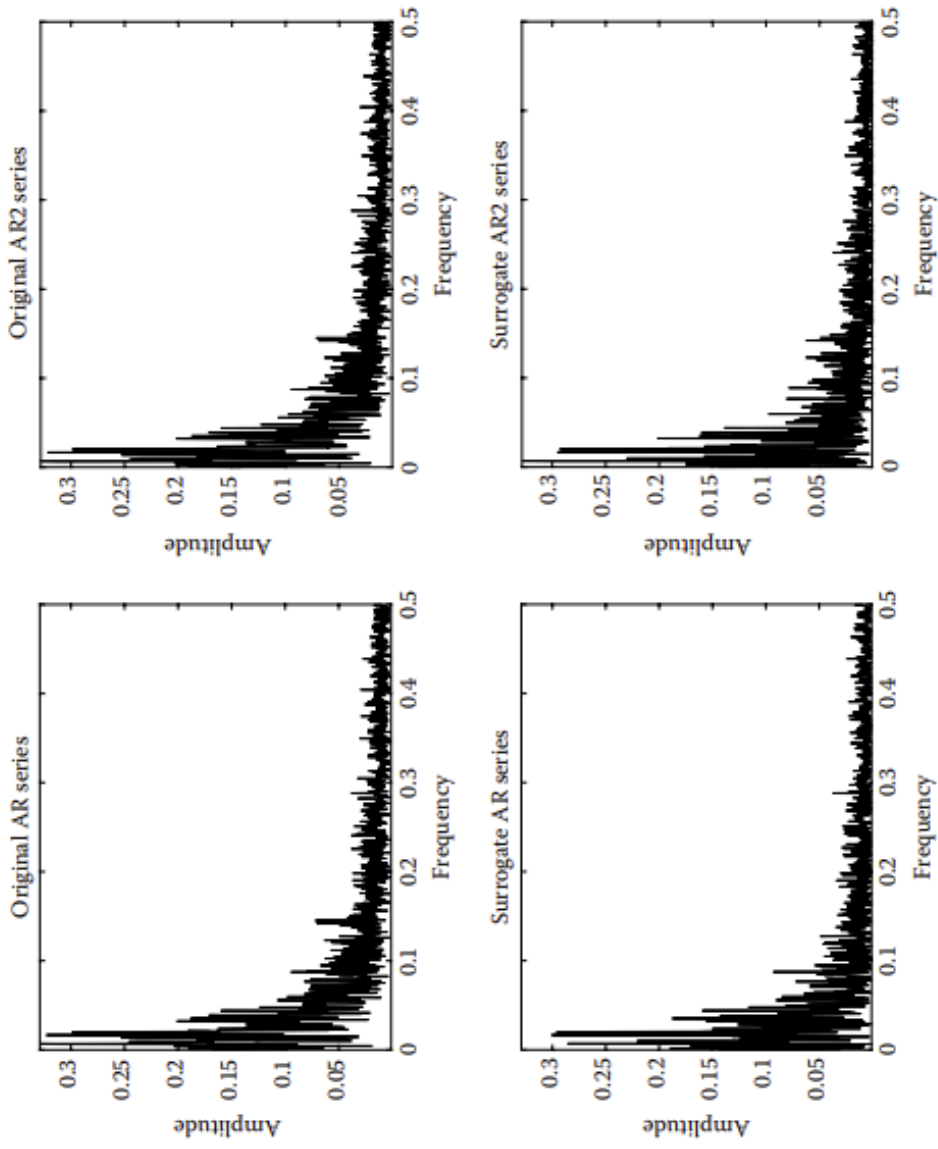


FIGURE 5.6 The power spectra of the series in Figure 5.5 are shown. Algorithm 1 preserves the frequency distribution in both the AR and AR2 time series.

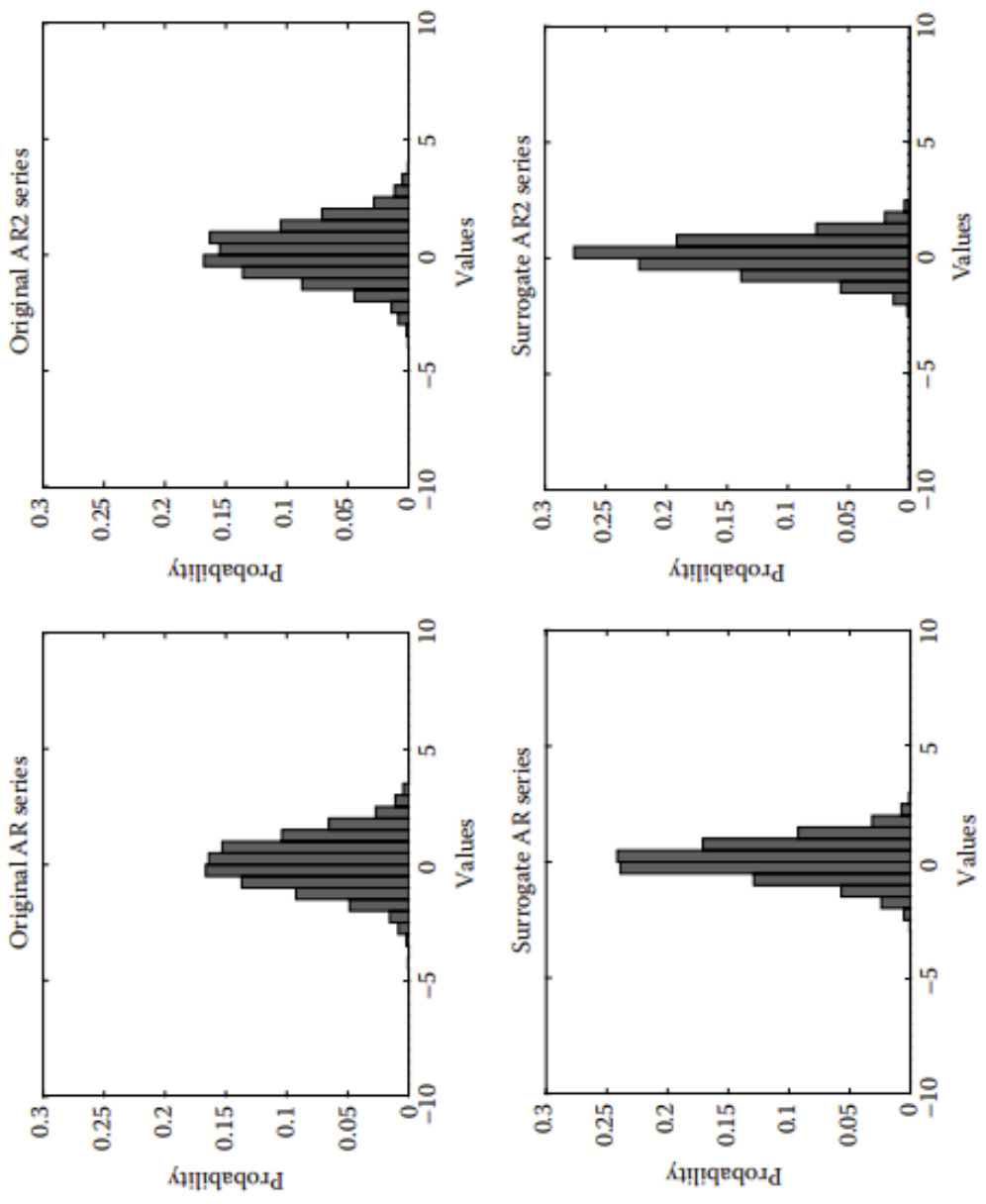
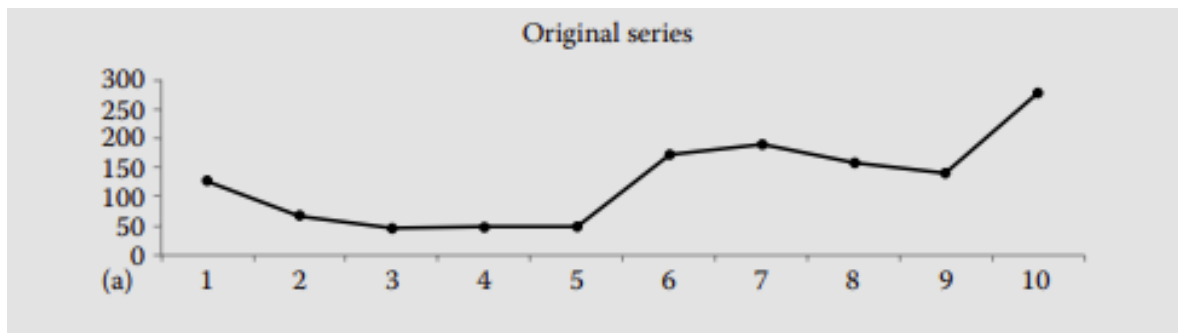


FIGURE 5.7 The probability distribution of AR and AR2 from Figure 5.5. Algorithm 1 does not preserve this distribution, which can lead to a false rejection of the null hypothesis.

EXAMPLE BOX 5.4 NUMERICAL EXAMPLE OF THEILER ET AL. ALGORITHM 2 OR THE Aaft SURROGATE METHOD

This example will illustrate the process of generating a surrogate time series by Algorithm 2.

1. Plot the data and identify the null hypothesis. The appropriate null hypothesis for the A2 algorithm is that the time series is generated from a static and monotonic nonlinear filter. The original time series for this example is: $\{x_n\} = \{125.00, 67.37, 46.33, 49.64, 49.83, 172.36, 188.22, 157.40, 138.77, 276.97\}$.



2. Rank order $\{x_n\}$ and make a time series of rank. If we rank order the time series, it would be: $\{46.33, 49.64, 49.83, 67.37, 125.00, 138.77, 157.40, 172.36, 188.22, 276.97\}$. So, the rank order time series would take the position of these numbers as 1–10, in the order of the original series. The time series of the ranks would be: Rank_original = $\{5, 4, 1, 2, 3, 8, 9, 7, 6, 10\}$.

3. Generate a random time series which has the same length as $\{x_n\}$ and has a Gaussian distribution. We call this random time series $\{y_n\} = \{1.50, -0.28, 0.50, 1.034, -1.51, -2.34, 0.26, -0.52, 0.23, 0.32\}$.

4. Sort $\{y_n\}$ in an ascending order, which is called Sorted_ $\{y_n\}$.
Sorted_ $\{y_n\} = \{-2.34, -1.50, -0.52, -0.28, 0.23, 0.26, 0.32, 0.50, 1.03, 1.50\}$

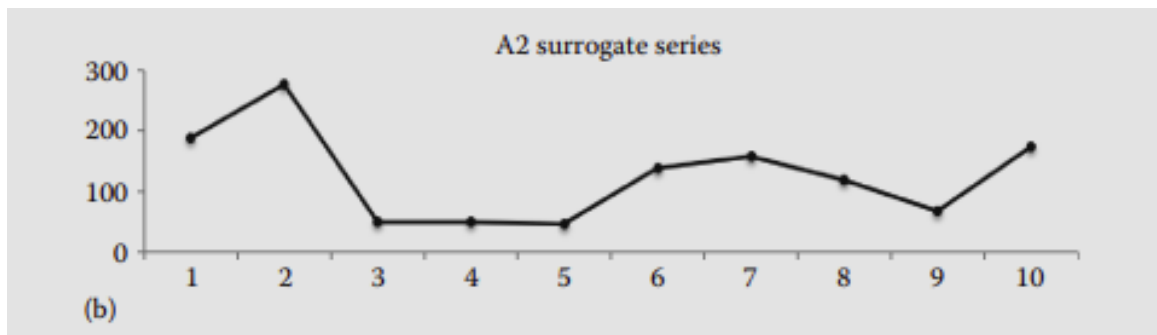
5. Reorder Sorted_ $\{y_n\}$ according to Rank_original and let it be called Reordered_ $\{y_n\}$. Reordered_ $\{y_n\} = \{0.23, -0.28, -2.34, -1.51, -0.52, 0.50, 1.03, 0.32, 0.26, 1.50\}$. Reordered_ $\{y_n\}$ has the same time evolution as the $\{x_n\}$, but the distribution is Gaussian.

6. Apply Algorithm 1 on Reordered_ $\{y_n\}$. That is taking the Fourier transform and shuffling phases while keeping the amplitudes and then applying the inverse Fourier transform. This series is called Surrogate_ $\{y_n\} = \{1.07, 1.19, -0.55, -1.06, -1.48, -0.06, 0.17, -0.12, -0.28, 0.30\}$.

7. Rank order Surrogate $\{y_n\}$. This results in Rank_Surrogate $\{y_n\} = \{5, 4, 3, 9, 8, 6, 7, 10, 1, 2\}$.

8. Reorder Rank_Surrogate $\{y_n\}$. This sequence is called Final_reorder = $\{9, 10, 3, 2, 1, 6, 7, 5, 4, 8\}$.

9. Arrange $\{x_n\}$ according to the Final_reorder. Now, a surrogate has been generated as Surrogate = $\{188.22, 276.97, 49.83, 49.64, 46.33, 138.77, 157.40, 125.00, 67.37, 172.36\}$. By plotting the series in the following, it is clear that the surrogate contains the same values as the original, but in a different order.



The chart following further illustrates how the surrogate series has changed from the original series. Two highlighted values show 46.33 moving from position 3 in the original series to position 5 in the surrogate series and 188.22 moving from position 7 to position 1.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
125.00	5	1.50	-2.34	0.23	1.07	5	9	188.22
67.37	4	-0.28	-1.50	-0.28	1.19	4	10	276.97
46.33	1	0.50	-0.52	-2.34	-0.55	3	3	49.83
49.64	2	1.03	-0.28	-1.50	-1.06	9	2	49.64
49.83	3	-1.50	0.23	-0.52	-1.48	8	1	46.33
172.36	8	-2.34	0.26	0.50	-0.06	6	6	138.77
188.22	9	0.26	0.32	1.03	0.17	7	7	157.40
157.40	7	-0.52	0.50	0.32	-0.12	10	5	125.00
138.77	6	0.23	1.03	0.26	-0.28	1	4	67.37
276.97	10	0.32	1.50	1.50	0.30	2	8	172.36

A surrogate time series generated by Algorithm 2 preserves the amplitude distribution of the original time series and linear correlations among data points or power spectrum of the original time series. Algorithm 2 deals with more realistic situations than Algorithm 1. However, limitations with Algorithm 2 have also been

reported (Kugiumtzis 2001; Schreiber and Schmitz 1997). When a time series is short and strongly correlated, rescaling of the inverse Fourier transformed data can change the linear correlations of the time series, causing a discrepancy between the power spectrum of the original time series and the surrogate time series (Figure 5.8). This discrepancy can lead to a false rejection of the null hypothesis, especially with a discriminating statistic which is sensitive to correlations among data points such as polynomial and local average fit, mutual information, and largest Lyapunov exponent (Kugiumtzis 2001). To deal with this problem, Schreiber et al. (1996) proposed an improved method known as the iterated amplitude-adjusted Fourier transform (IAAFT), and this has been widely used (Lehnertz et al. 2001; Poggi et al. 2004; Rieke et al. 2003).

Iterated Amplitude-Adjusted Fourier Transform

The steps to generate surrogate time series by the iterated amplitude-adjusted Fourier transform (IAAFT) are as follows (Figure 5.9). Let $\{x_n\}$ be the original time series where $n = 1, 2, \dots, N$:

1. Sort $\{x_n\}$ in an ascending order and store this sorted time series Sorted_ $\{x_n\}$.
2. Take the Fourier transform of $\{x_n\}$ and store the squared, the amplitudes of the Fourier transform of $\{x_n\}$, $X_k^2 = |\sum x^n e^{2\pi kn/N}|^2$
3. Shuffle $\{x_n\}$ and take the Fourier transform and call it FTRandomized $\{x_n\}$.
4. To adjust the power spectrum, replace the squared amplitudes of FTRandomized $\{x_n\}$ by $\{X_k^2\}$. The phases are kept unchanged. Then transform back by taking the inverse Fourier transform.
5. The procedure at step 4 will change the amplitude distribution. Therefore, adjust the amplitudes by ranking the values of this time series and replacing them by the values of Sorted_ $\{x_n\}$.
6. However, again the procedure at step 5 may alter the power spectrum, so step 4 and step 5 are repeated until some convergence is achieved.

A surrogate time series generated by IAAFT preserves the power spectrum of an original time series much better than the other surrogate algorithms (Figure 5.10). For a numerical example, please see Example Box 5.5. This algorithm is trying to maintain the underlying linear correlations. However, there is no guarantee that iterations will eventually converge. Furthermore, there is also a concern that surrogate time series generated by IAAFT for a data with short length may not have

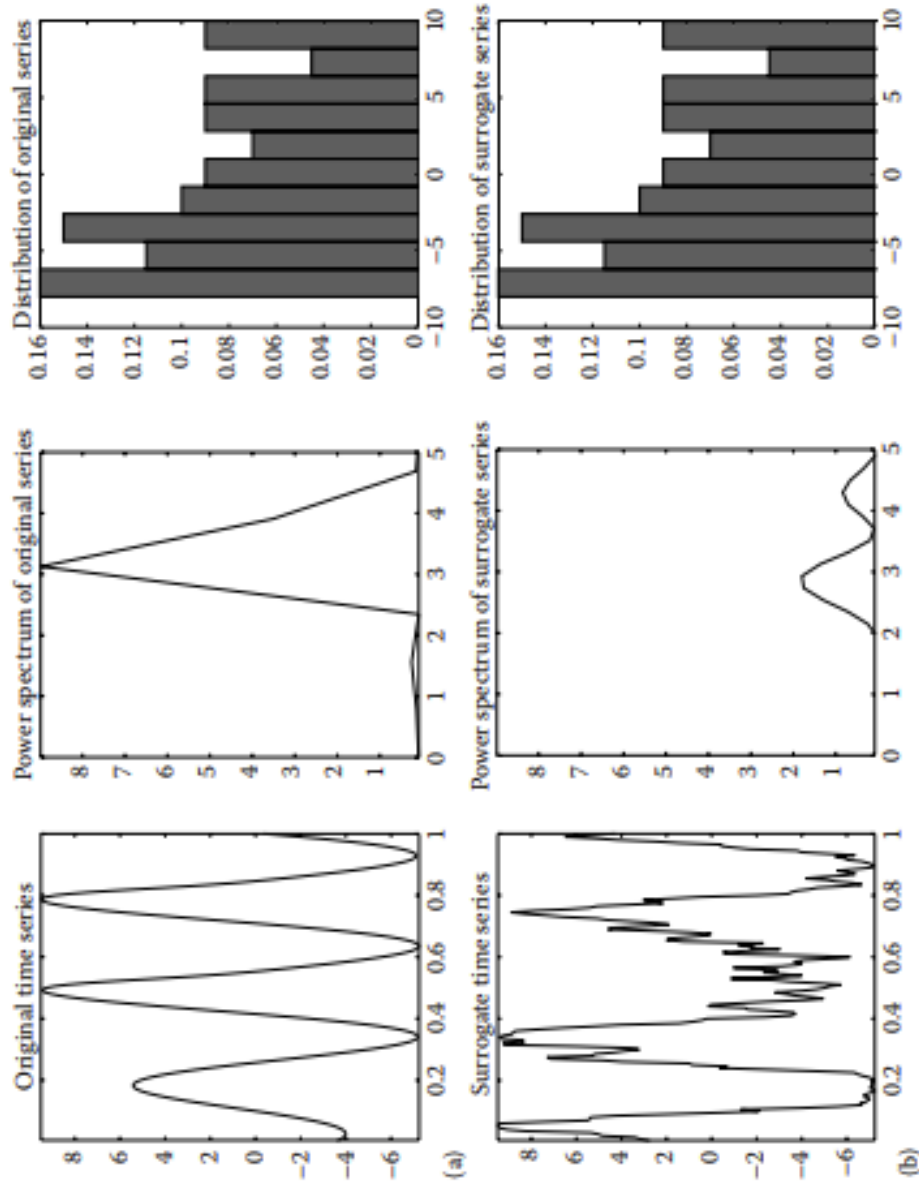


FIGURE 5.8 Theiler et al. Algorithm 2 preserves the amplitude distribution and the power spectrum of the original time series. However, when the time series is short and strongly correlated, the linear correlations of the data can change during the inverse Fourier process. When comparing the (a) original time series, power spectrum, and amplitude spectrum to the (b) surrogate time series, power spectrum, and amplitude, it is clear that the amplitude is preserved with algorithm 2, but not the power spectrum.

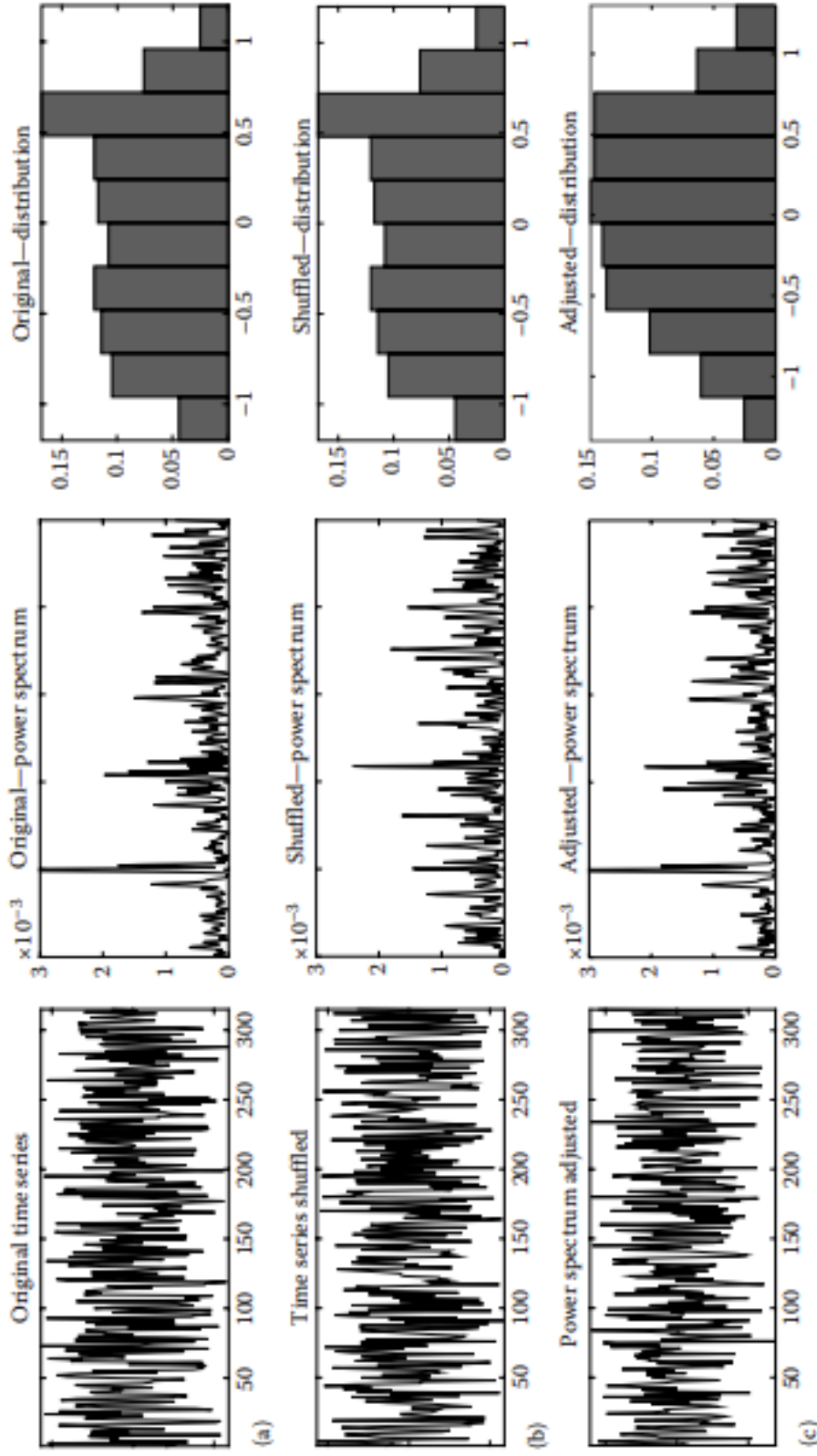


FIGURE 5.9 The IA-AFT algorithm is the best to preserve the power spectrum of the original time series. (a) The original time series, power spectra, and distribution are shown. (b) The shuffled time series results in a different power spectrum, but the distribution is the same. (c) The power spectrum is adjusted by replacing the squared amplitudes of the Fourier transform of the shuffled time series with the squared amplitudes of the Fourier transform of the original time series. This brings the power spectrum closer to the original time series but results in changes to the amplitude distribution. (*Continued*)

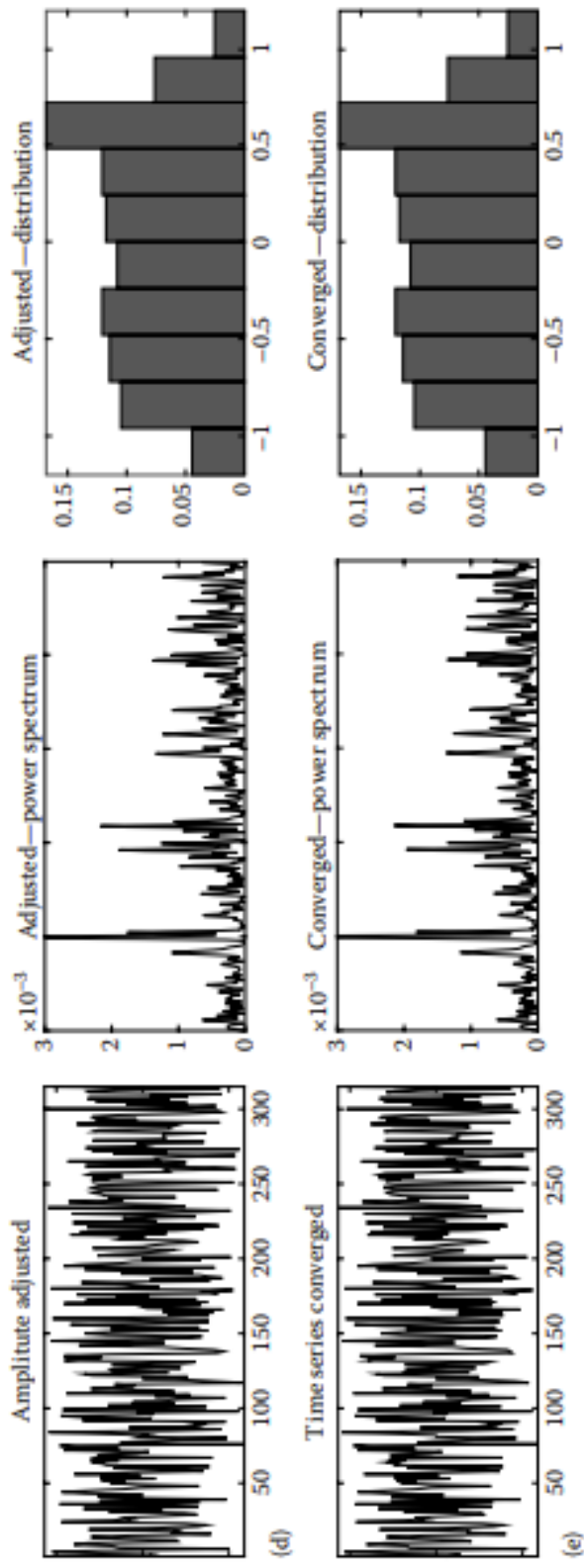


FIGURE 5.9 (Continued) (d) The amplitudes are then adjusted by ranking the values of the time series in (c) and replacing them with the ranked values of the time series in (b). (e). This procedure can alter the power spectrum again, so the process of adjusting the power spectrum and the amplitude distribution are repeated until convergence is achieved with both measures.

A surrogate time series generated by IAAFT preserves the power spectrum of an original time series much better than the other surrogate algorithms (Figure 5.10). For a numerical example, please see Example Box 5.5. This algorithm is trying to maintain the underlying linear correlations. However, there is no guarantee that iterations will eventually converge. Furthermore, there is also a concern that surrogate time series generated by IAAFT for a data with short length may not have enough randomization, which makes hypothesis testing against a specific system rather than a general class of the system (Small and Judd 1998; Small and Tse 2002). Or, put another way, the surrogate time series may not be different enough from the original.

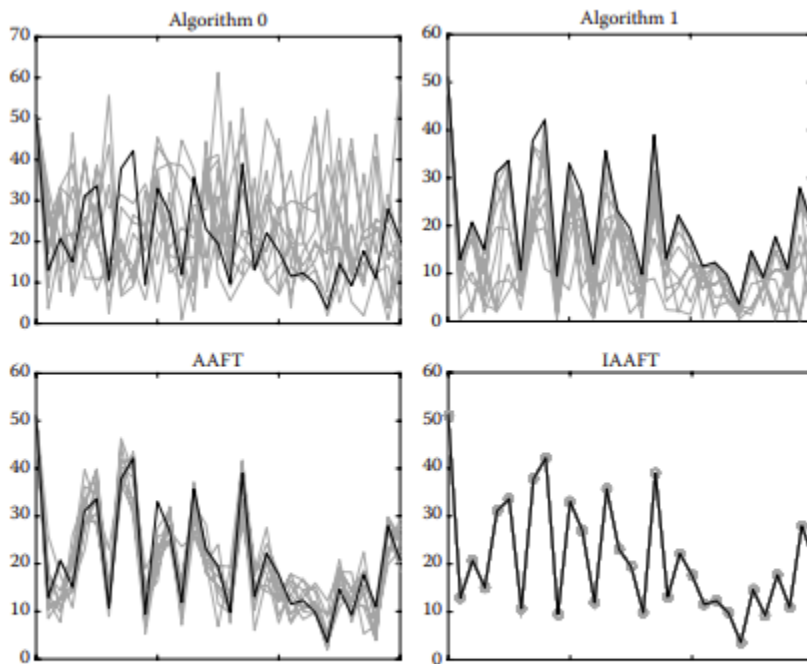
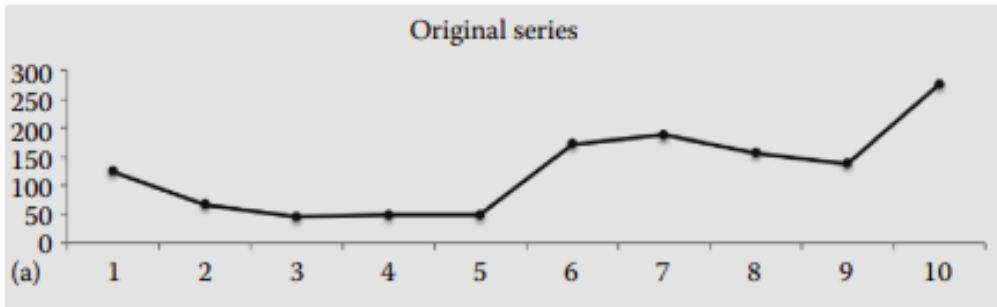


FIGURE 5.10 These plots of a portion of the power spectra of the original time series (gray) and 19 surrogate time series generated by Algorithm 0 (black), Algorithm 1 (black), AAFT (black), and IAAFT (circles). It is visually clear that the spectra are altered the least in IAAFT.

EXAMPLE BOX 5.5 NUMERICAL EXAMPLE OF THE IAAFT

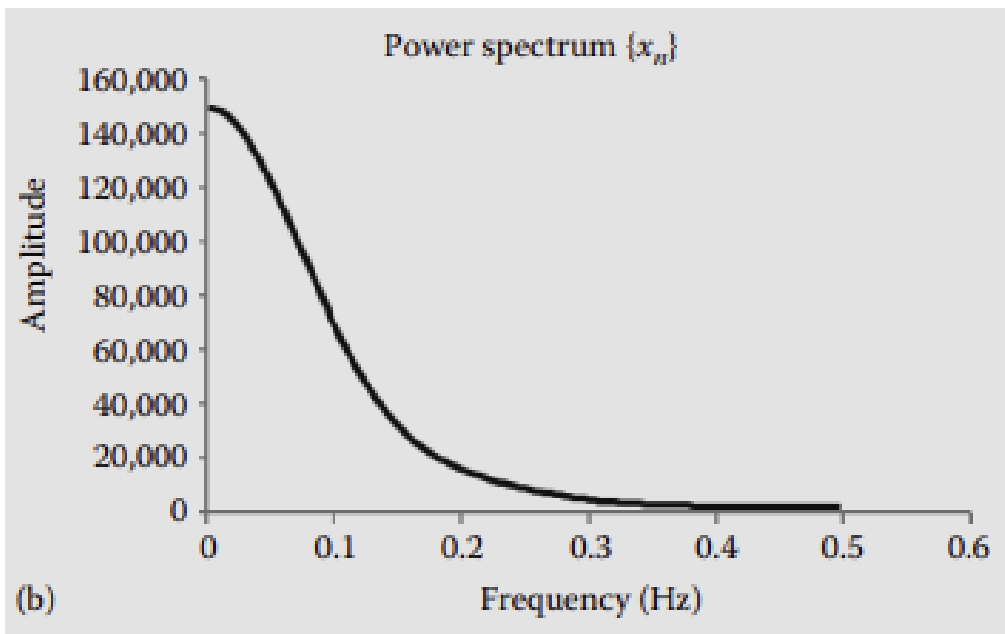
This example will use the same 10 data point time series as Example Box 5.4, which is: $\{x_n\} = \{125.00, 67.37, 46.33, 49.64, 49.83, 172.36, 188.22, 157.40, 138.77, 276.97\}$.

1. Plot the data, identify the null hypothesis, and Sort $\{x_n\}$ in an ascending order. The appropriate null hypothesis for this algorithm is that the time series is generated from a linear Gaussian stochastic process.

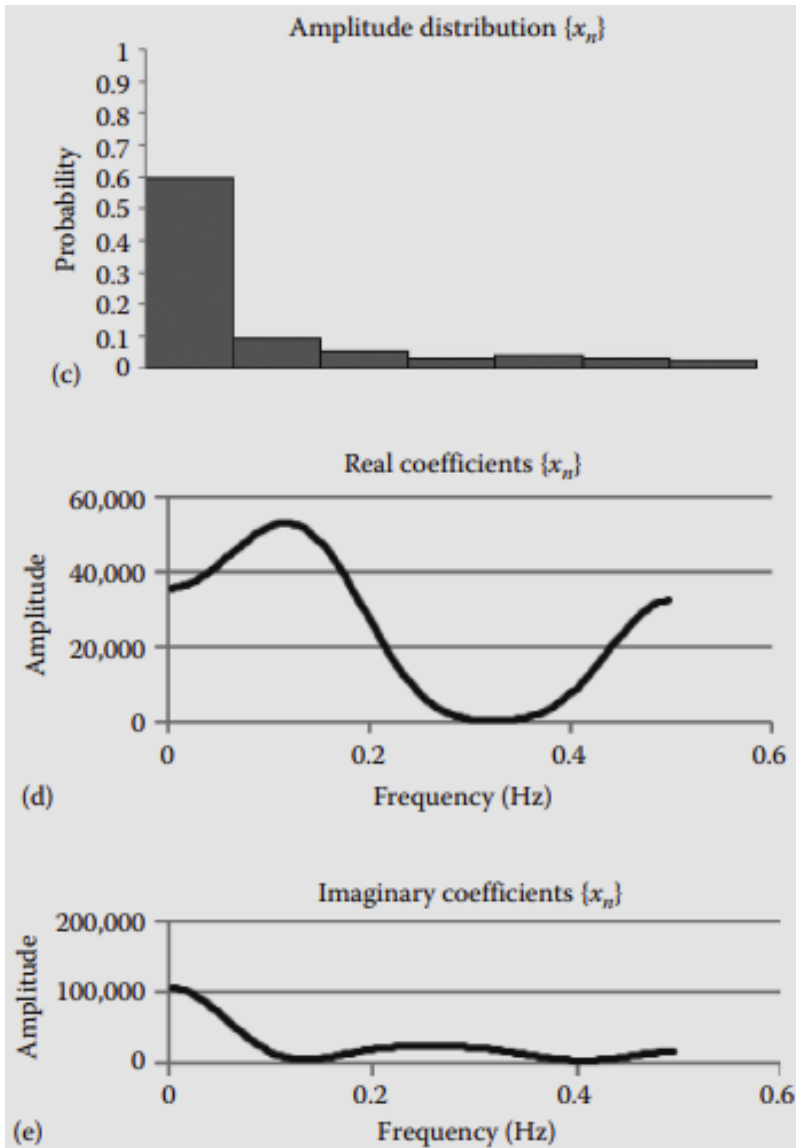


The new time series is Sorted_{x_n} = {46.33, 49.64, 49.83, 67.37, 120.00, 138.77, 157.40, 172.36, 188.22, 276.97}.

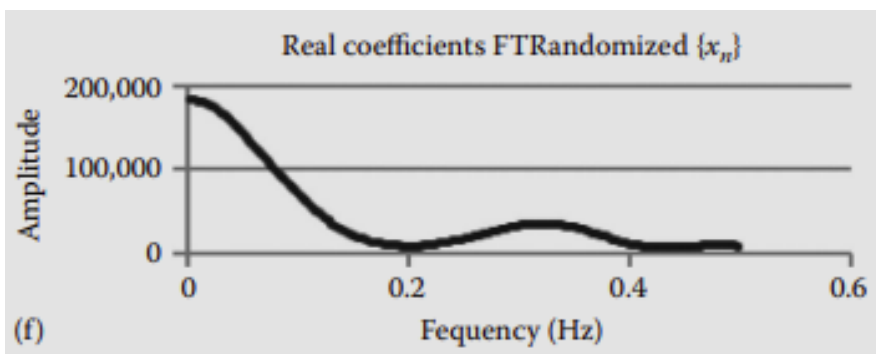
2. Take the Fourier transform of {x_n} and store the squared amplitudes of the Fourier transform of {x_n}, $X_k^2 = |\sum x^n e^{2\pi kn/N}|^2$.

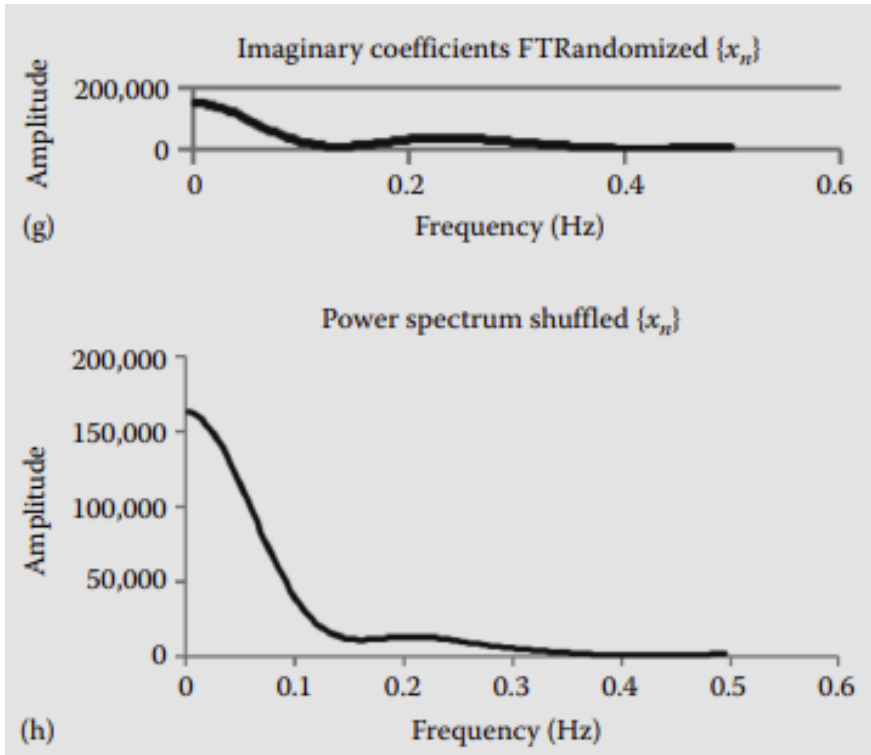


The transform moves the series from the time domain to the frequency domain. The power spectrum of {x_n} is shown. The Fourier transform results in a list of real and imaginary coefficients in the different frequencies. The plots of the frequency components from the Fourier transform are shown in the following. The stored squared amplitudes of the Fourier transform of {x_n} that are stored as follows: {X_k^2} = {1271.89, 396.23, 170.51, 222.24, 77.16, 175.59, 77.16, 222.24, 170.51, 396.23}. The amplitude distribution of {x_n} is shown on the next page.

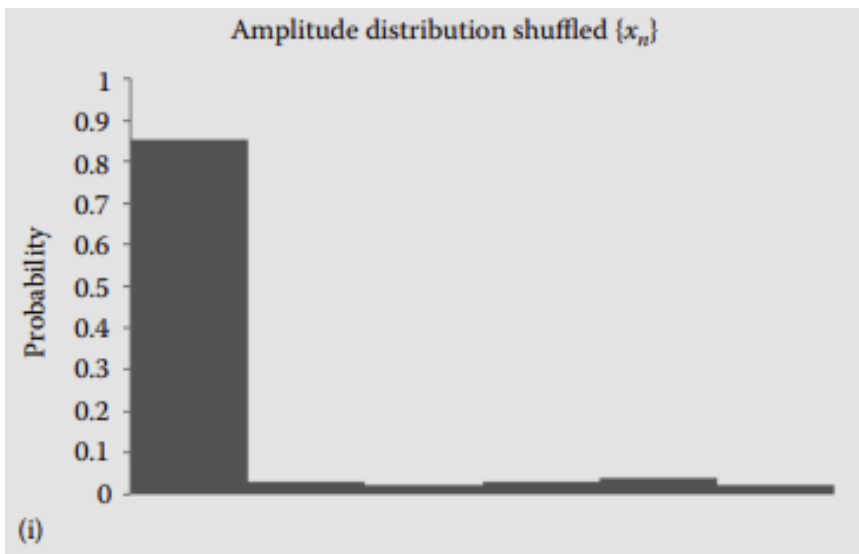


3. Shuffle $\{x_n\}$ and take the Fourier transform of the shuffled series. This series is called FTRandomized $\{x_n\}$. Shuffled $\{x_n\} = \{49.64, 276.97, 172.36, 188.22, 120.00, 46.33, 49.83, 157.40, 138.77, 67.37\}$. The FTRandomized $\{x_n\}$ results in the following real and imaginary coefficients:



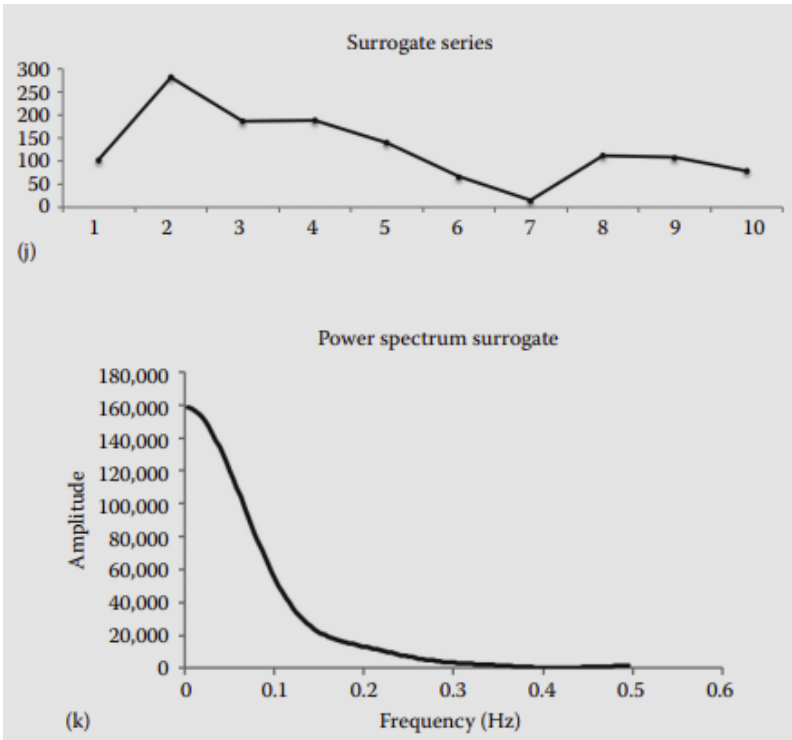


The power spectrum of Shuffled $\{x_n\}$ is different from the spectrum of $\{x_n\}$, so needs to be adjusted. However, the amplitude distribution Shuffled $\{x_n\}$ is similar with that of the $\{x_n\}$.

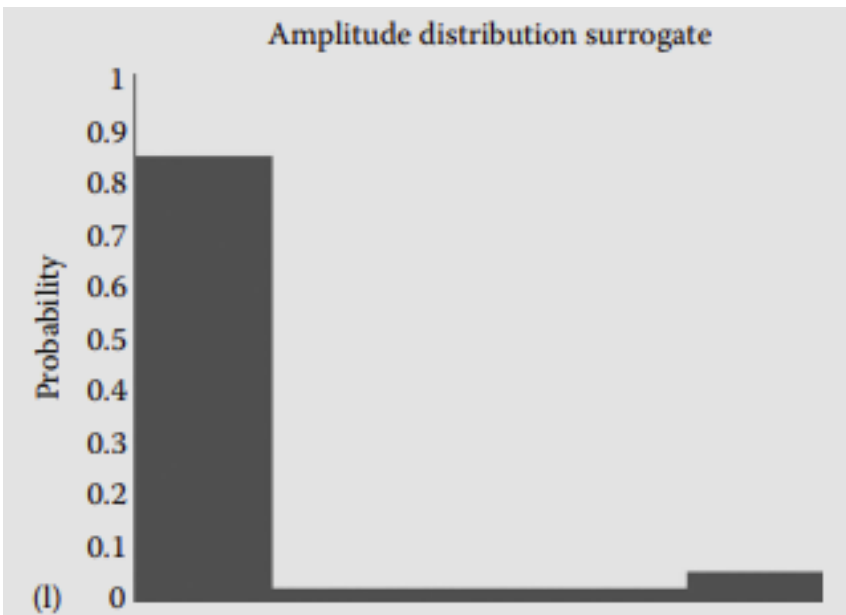


4. The power spectrum is adjusted by replacing the squared amplitudes of FTRandomized $\{x_n\}$ with $\{X_k^2\}$. This randomizes the series without changing the phases. Then, the series is transformed back by using the inverse Fourier transform. That results in a surrogate $\{102.19, 281.05,$

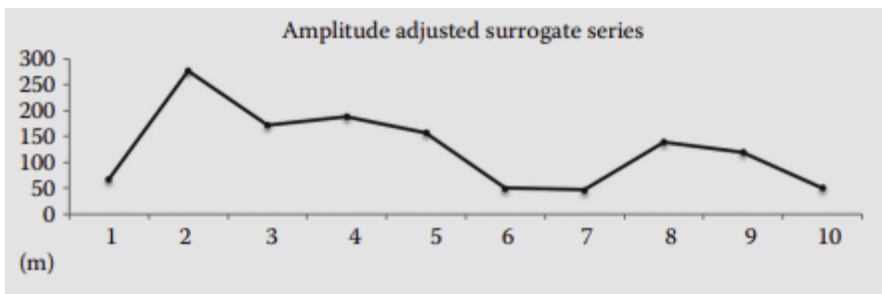
185.90, 187.87, 139.30, 65.28, 13.55, 111.26, 107.20, 78.28}. This time series now looks like:



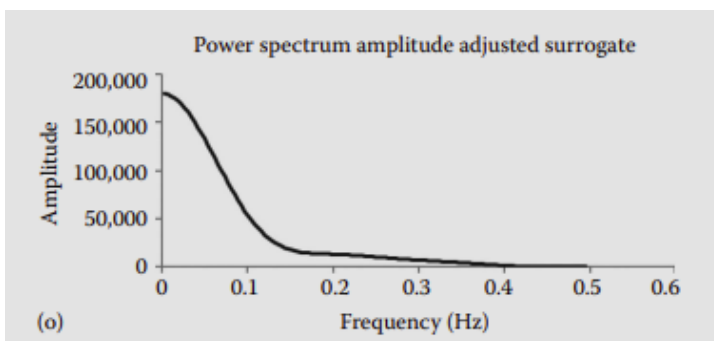
Now, the power spectrum of the adjusted surrogate series is similar with the power spectrum as $\{x_n\}$. However, the amplitude distribution of the surrogate is different from that of $\{x_n\}$. Keep in mind that although the differences are small, our example time series is only 10 data points. Thus, these differences would be magnified for longer series.



5. *Adjust the amplitude distribution.* The procedure from Step 4 will change the amplitude distribution. The amplitudes are adjusted by ranking the values of the surrogate time series and replacing them by the values of Sorted_{x_n}. Rank order of the surrogate is {13.55, 65.28, 78.28, 102.19, 107.20, 111.26, 139.30, 185.90, 187.87, 281.05} which results in the Rank_surrogate = {7, 6, 10, 1, 9, 8, 5, 3, 4, 2}. Now, the rank is sorted and ranked according to its position. For example, you can see that rank number 1 of Rank_surrogate was in the 4th position, rank number 2 in the 10th position, etc., which results in Sorted_rank_surrogate = {4, 10, 8, 9, 7, 2, 1, 6, 5, 3}. The final part of this step is to replace them with the values of Sorted_{x_n}. From step 1, Sorted_{x_n} = {46.33, 49.64, 49.83, 67.37, 120.00, 138.77, 157.40, 172.36, 188.22, 276.97}. When the Sorted_rank_surrogate is replaced with the values from Sorted_{x_n}, the 4th number from Sorted_{x_n} is placed 1st in our new series, the 10th number from Sorted_{x_n} 2nd, and so on, resulting in the Amplitude_adjusted_surrogate = {67.37, 276.97, 172.36, 188.22, 157.40, 49.64, 46.33, 138.77, 120.00, 49.83}.



6. *The procedure in step 5 may again alter the power spectrum. Thus, step 4 and step 5 are repeated until some convergence is achieved and both the power spectrum and the amplitude distribution are close to being preserved.* By examining the amplitude distribution shown in the following, it is clear that the amplitude distribution has changed and is (again) similar with {x_n}, but the power spectrum has changed (again). However, the differences between this power spectrum and {x_n}, are less than between the spectrum of Shuffled {x_n} and {x_n}.



7. *Calculate the discriminant statistic.*

8. *Test for significance between discriminating statistic values of the original time series and the series of IAAFT surrogates.*

REJECTION OF NULL HYPOTHESIS

So far different linear surrogate algorithms have been discussed as well as how surrogate time series are generated by preserving some linear properties of an original time series while destroying other dynamics (Figures 5.11 and 5.12). The rejection of null hypothesis has also been discussed as being an indication of the presence of more complex dynamics than linear dynamics in the original time series. However, results of surrogate methods do not provide any definite answers regarding the exact nature of the underlying dynamics of the original data. The rejection of the null hypothesis only indicates that the underlying dynamics of the original time series are not consistent with the null hypothesis. In addition, even if there is no significant difference between the original and the surrogate time series, it cannot be concluded that they are from the same population. It may be simply due to inadequate statistics. Therefore, as is the case for other nonlinear methods, multiple nonlinear tools should be applied instead of solely depending on the results obtained by a surrogate method for data analysis.

Another question one may raise is whether the surrogate methods that were discussed so far can be applied to biological time series that exhibit inherent trends. In fact, many experimental data exhibit strong periodicity such as gait, human speech, ECG and so forth (Buzzi et al. 2003; Miller et al. 2006; Schreiber and Schmitz 2000; Small and Tse 2002; Stergiou et al. 2004; Zhang et al. 2007). What happens if the above mentioned surrogate methods are used with such time series? The hypothesis tests for linear surrogate methods are not suitable for a time series with periodicity: such time series are inconsistent with the null hypothesis of a linearly filtered noise. As a result, surrogate time series generated from a time series with periodicity by using linear surrogate methods have geometric structures different from that of the original time series (Figure 5.13). This would lead to a higher rejection rate of the null hypothesis than what should actually occur simply because the structure of the time series in the surrogate is changed (Algorithm 0 and 1) or because the surrogate has changed geometric structure (AAFT and IAAFT). Therefore, we need another testing hypothesis and surrogate algorithm for a time series with regular persistent fluctuations or underlying periodicity.

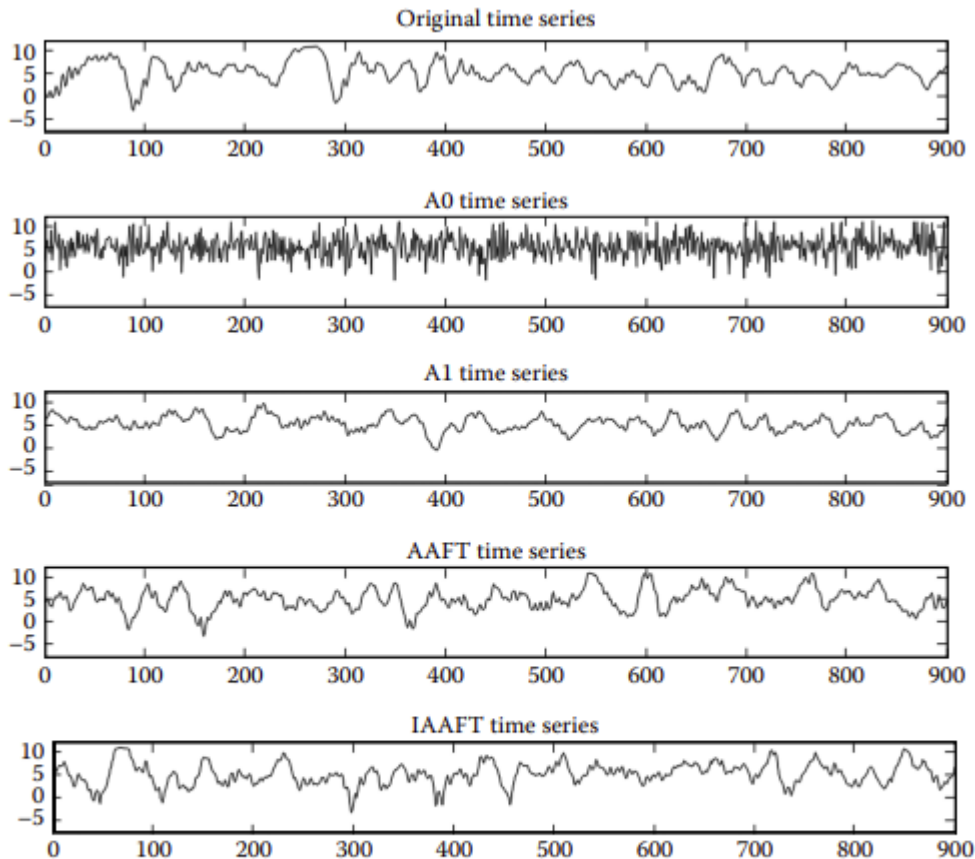


FIGURE 5.11 Original time series of center of pressure (COP) data from standing posture in the anterior–posterior (AP) direction and surrogate time series generated by different algorithms.

PSEUDOPERIODIC SURROGATE METHOD

Small et al. (2001) introduced a method called the pseudoperiodic surrogate (PPS) algorithm (Small et al. 2001). Pseudoperiodic time series are defined as time series that have a noisy periodic orbit perturbed by either dynamical noise or observational noise or have an oscillatory chaotic flow. The power spectrum of these time series display clear spikes (Small 2005) (Figure 5.14) due to the dominant inherent frequencies. Since such a time series already exhibits a deterministic behavior, what is actually sought in this case is if there is any type of nonlinear structure on the fluctuations that are on top of these inherent dominant frequencies. A human electrocardiogram is an example of a signal with periodic orbits that result from successive heartbeats. However, there may be additional order within the dynamics of the sinus rhythm, or it could be consistent with uncorrelated noise (Small et al. 2001). Therefore, the null hypothesis of the PPS states that a time series is consistent with a periodic orbit perturbed by uncorrelated noise. To test this null hypothesis, the PPS generates a surrogate time series that keeps the large-scale behavior of the original time series but does not preserve any additional small-scale dynamics that can be regarded as chaotic, linear, or nonlinear deterministic structure. Therefore, intracycle

dynamics, which are dynamic patterns within one period of a cyclic pattern, are preserved, while intercycle dynamics, which are dynamic patterns between different periods across a cyclic pattern of the time series, are altered. Two alternative hypotheses are suggested: (1) deterministic nonperiodic intercycle dynamics and (2) a periodic orbit with correlated noise. In the case of the electrocardiogram, the application of the PPS algorithm showed that the sinus rhythm has deterministic nonperiodic intercycle dynamics (Small et al. 2001).

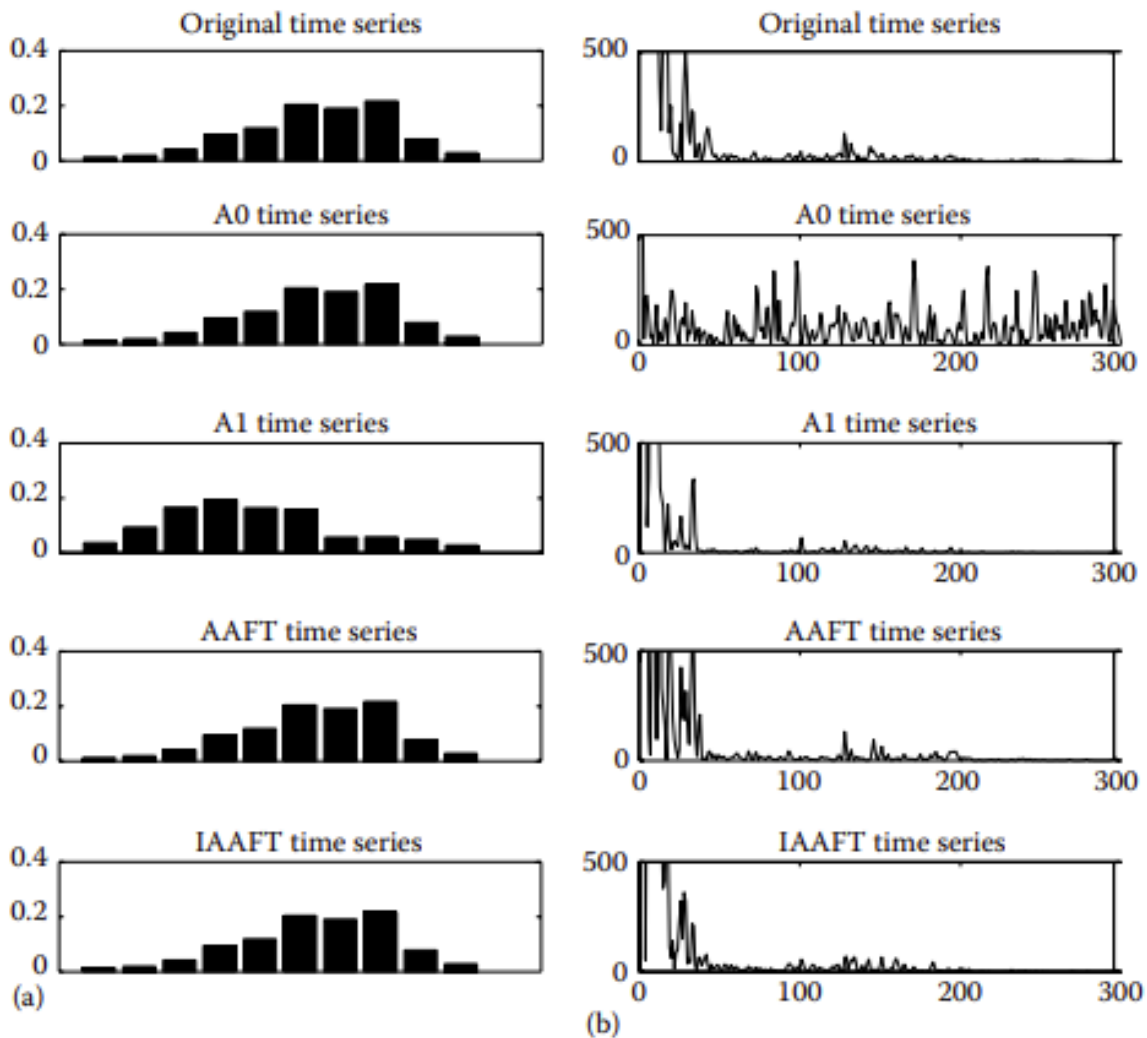


FIGURE 5.12 The probability distribution (a) and power spectra (b) of the original time series and all surrogate time series generated by the different algorithms from Figure 5.11.

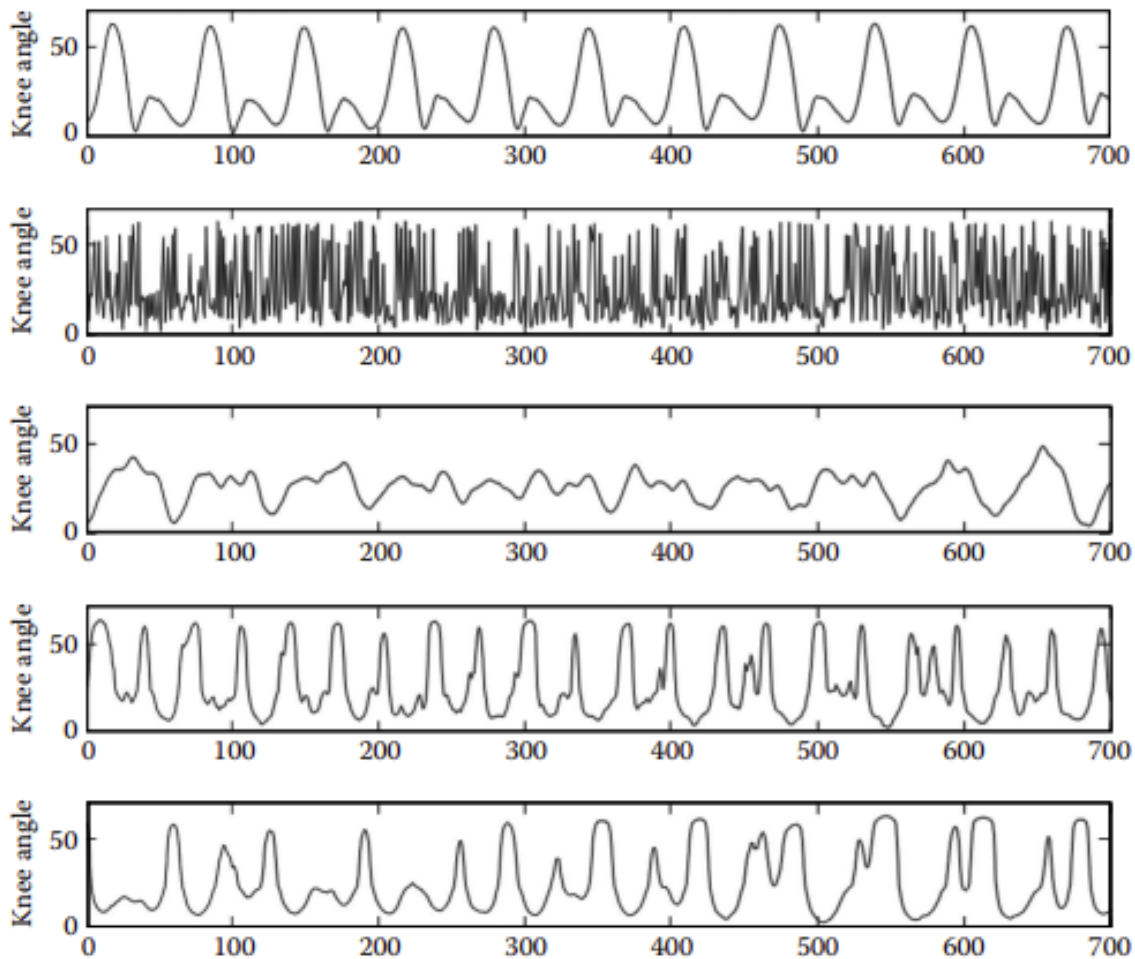


FIGURE 5.13 Applying linear surrogate methods is not suitable for time series with inherent periodicity. Knee flexion and extension angle data from human gait produces cycles that correspond to steps. Using Algorithms, A0 and A1 would result in false rejection of the null hypothesis because these methods change the structure of the time series. Similarly, AAFT and IAAFT algorithms change the geometric structure of the time series and would also lead to false rejection of the null.

The first step of generating a surrogate time series by the PPS algorithm involves the reconstruction of the state space. Therefore, the embedding dimension (m) and the time lag (τ) need to be defined. The embedding parameters to compute discriminating statistics, which require the reconstruction of the state space, are usually set to be the same for the entire data sets. However, for generating a surrogate time series, the embedding parameters should be specific to each data set (Small and Tse 2002). The other parameter the PPS requires is the noise radius (ρ), which defines the amount of noise in a surrogate (Small et al. 2001). If noise radii are too large, then the surrogate time series will be too distinct from the original while if noise radii are too small, the original and surrogate time series will be too similar, which may result in a false positive

result for the hypothesis testing (Figure 5.15). Noise radii should be chosen such that the fine intercycle dynamics are removed, but the intracycle dynamics are preserved. Small et al. (2001) suggested selecting a p that maximizes the number of short segments (length ≥ 2) that are the same for the original time series and the surrogate. These segments represent the amount of correlation between the surrogate and the original data sets (Small et al. 2001). If p is too large, surrogate time series will be too different from the original time series because the dynamics were poorly approximated. If p is too small, surrogate time series will be too similar to the original time series.

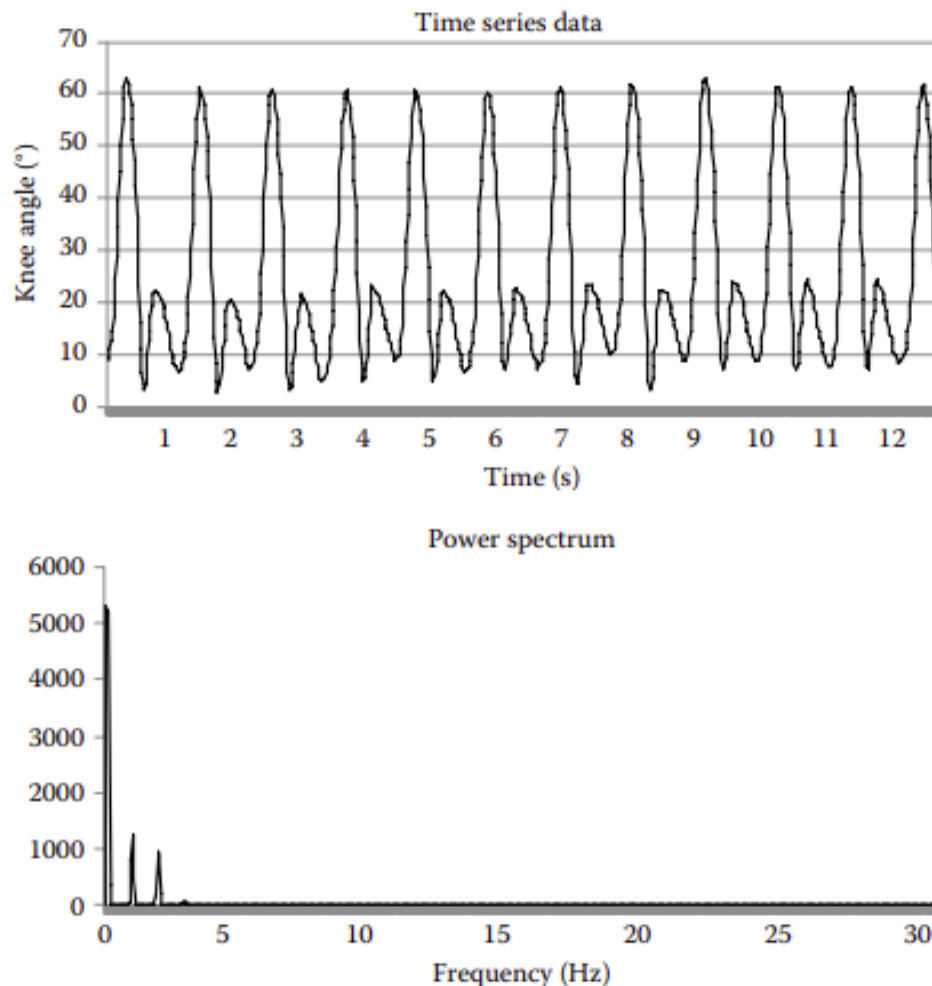


FIGURE 5.14 Knee flexion and extension angle time series have a periodic orbit that cycles with every step. The power spectrum displays a clear spike, which demonstrates the frequency of the dynamics in the time series. Such time series already exhibit deterministic behavior, so a different algorithm, the pseudoperiodic surrogate method, is required to identify whether additional determinism exists in the system.

Let us look at the procedure of generating a surrogate time series with the PPS:

1. Select the embedding dimension (m) and time lag (τ) for time delay embedding reconstruction.
2. Randomly select an embedded point as an initial condition (Figure 5.16a). This embedded point is a delay vector that has m elements, and we call it v_1 .
3. Randomly select a neighboring vector to v_1 and call it v_2 (Figure 5.16b). The neighbors are chosen with a certain probability equation.
4. Randomly select a neighboring vector to v_2 and call it v_3 (Figure 5.16c).
5. Repeat this procedure until the number of vectors that we select reaches the length of the original time series (Figure 5.16d).
6. A surrogate time series is generated by taking the first element of the selected delay vectors.

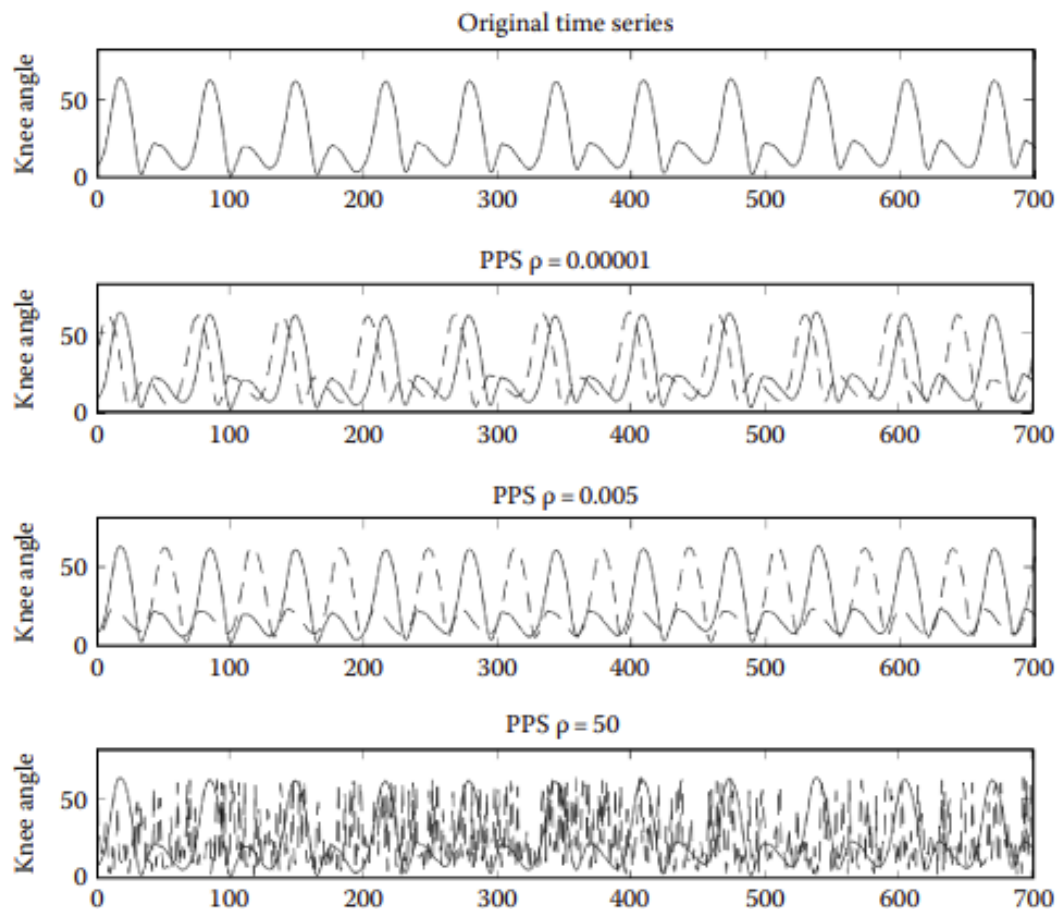


FIGURE 5.15 The noise radius is an important parameter for implementing the pseudoperiodic surrogate algorithm. If the noise radius is too small, the original and surrogate time series will be too similar. If the noise radius is too large, the surrogate time series will be too distinct from the original. Noise radii should be chosen such that the fine intercycle dynamics are removed, but the intracycle dynamics are preserved.

The surrogate time series generated by the PPS can be considered as a random walk on the attractor; therefore, it follows the same vector field as the original time series but is contaminated with dynamic noise (Figures 5.17 and 5.18). This addition of dynamic noise destroys subtle deterministic intercycle dynamics, including periodic dynamics with correlated noise, pseudoperiodic chaos or any deterministic nonperiodic intercycle dynamic behavior (Small and Tse 2002; Zhao et al. 2008). The PPS algorithm has been applied to gait kinematics data (Miller et al. 2006).

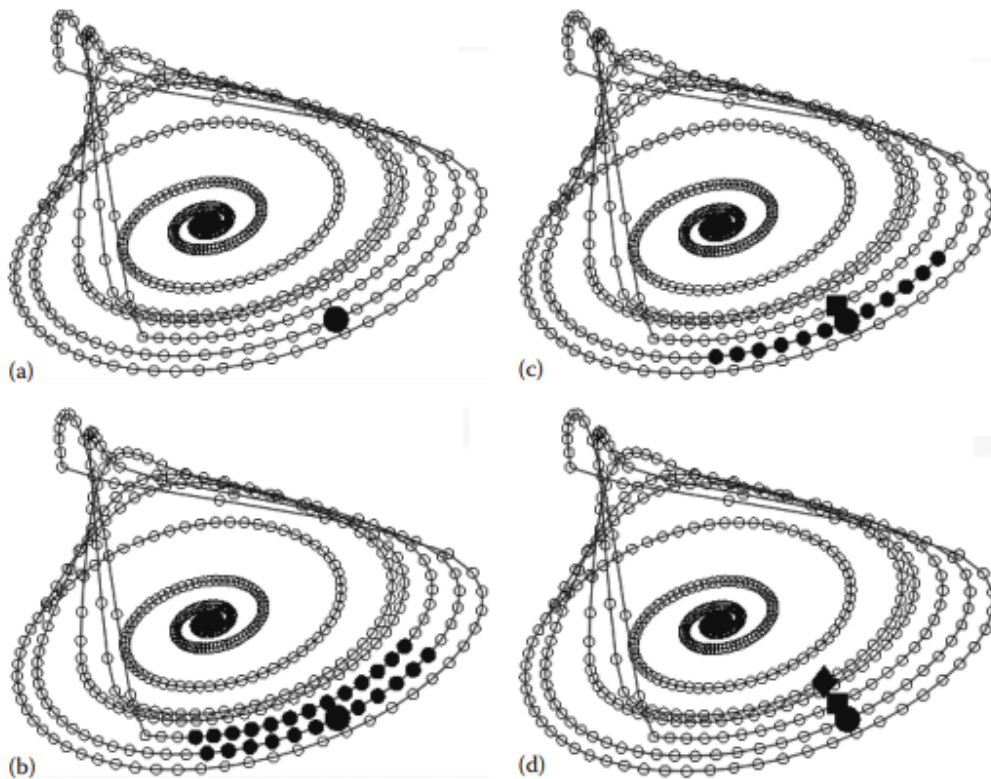


FIGURE 5.16 The process of the pseudoperiodic surrogate algorithm includes (a) picking an embedded point on the attractor as an initial condition, (b) identifying the neighbors within a certain probability equations, (c) randomly picking one neighboring vector and identifying its neighbors, and (d) randomly picking one neighbor's neighboring vector. This process is repeated until the number of vectors selected equals the length of the original time series. In (a), the initial condition is shown as a circle and its neighbors are diamonds in (b). The neighbors of a randomly selected neighboring vector are shown in (c) and one neighbor's neighboring vector is shown as a circle in (d).

Specifically, knee angle kinematic time series from healthy subjects were evaluated using the PPS algorithm, and Theiler et al. algorithm 0. The average time lag for the series was 9.833 and the average embedding dimension was 6.333. The noise radii that maximized the number of short segments that are the same for the original time series and the surrogate was 3.351. The paper demonstrated that Theiler et al. algorithm 0 destroyed the intracycle dynamics of the gait time series by changing the overall shape, which resulted in a false rejection of the null hypothesis. The PPS

algorithm did not alter the intracycle dynamics of the original time series, which made it more appropriate to explore the presence of underlying processes within these dynamics. Example Box 5.1 shows the general surrogation procedure using the PPS algorithm. The data for a knee flexion and extension angle, along with one surrogate generated using the PPS algorithm are included in Appendix 5.A. The SampEn values of the original and surrogate series, along with the parameters used are included in Example Box 5.1.

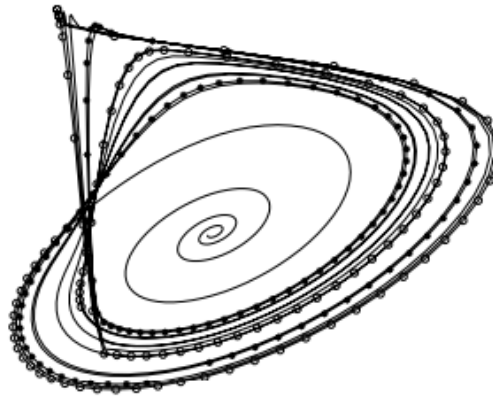


FIGURE 5.17 The phase portrait of the original Rossler attractor is plotted as a solid black line. The trajectories of the original attractor are indicated with open circles, while the trajectories of the pseudoperiodic surrogate attractor are indicated with asterisks. Clearly, the surrogate attractor follows the same vector field as the original, but since it has been contaminated with dynamic noise, the trajectory is not identical to that of the original attractor.

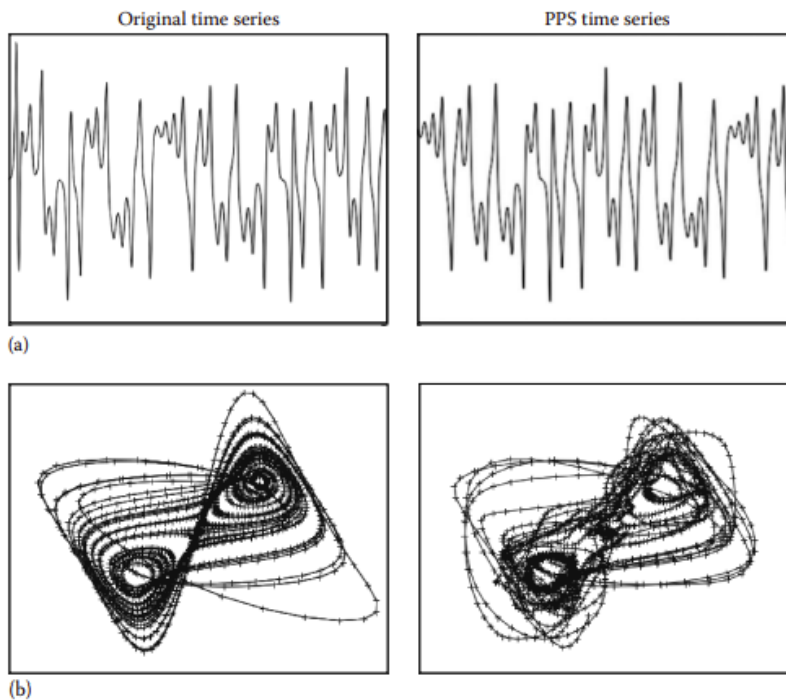


FIGURE 5.18 The plot of the original and pseudoperiodic surrogate time series (a) and two dimensional vector phase portraits (b) from the Lorenz attractor. The surrogate is a random walk in the same vector field as the original time series but is contaminated with dynamic noise.

DISCRIMINATING STATISTICS FOR PPS

As for discriminating statistics to be used with the PPS, it was reported that for both theoretical and experimental time series, the correlation dimension performed the best compared to other nonlinear tools such as the Shannon's entropy, prediction error, mutual information, kurtosis and skewness (Small and Tse 2002). However, Zhang et al. (2007) reported that the correlation dimension and the PPS algorithm failed to differentiate between a chaotic system and a noisy periodic orbit. In this study, a theoretical time series of the chaotic Rossler system (Equation 5.4) was used as the original time series. If the correlation dimension was an appropriate discriminating statistic, then the null hypothesis should have been rejected. However, the correlation dimension of the original time series lied within the dimension distribution of the surrogates and could not differentiate between the original and the surrogate time series. This is due to the fact that the correlation dimension characterizes the distribution of the points in the state space; however for the PPS algorithm, the distribution of the original and surrogate time series should follow the same pattern. Thus, an alternative discriminating statistic must be used when implementing PPS surrogation:

$$\begin{aligned}\dot{x} &= -(y+z) \\ \dot{y} &= x+ay \\ \dot{z} &= 2+z(x-4)\end{aligned}\tag{5.4}$$

Zhang et al. (2007) introduced alternative new methods to be used as discriminating statistics with the PPS. Those methods do not require the reconstruction of the state space but take the cycle in the time series as the basic processing unit. The discriminating statistics depend on the correlation coefficient between cycles and are supposed to be more robust to nonstationarity in data and different kinds of noise. There are separate methods for detecting the temporal, or time index, correlation and the spatial, or configuration of cycles, correlation. The basic ideas of the methods proposed by Zhang et al. (2007) are the following:

1. Divide a pseudoperiodic time series into consecutive cycles C_i ($i = 1, 2, \dots, m$).
2. Find correlation coefficients (ρ) between cycles as a measure of their distance in phase space.
3. Characterize the similarity of wave form between a pair of cycles. A large ρ means there is a higher level of similarity. Two cycles will also be close in the phase space with a higher ρ .

For chaotic systems, the distance between two nearby cycles will increase exponentially over time due to sensitivity to initial conditions, whereas the correlation will drop exponentially. More specifically, Zhang et al. presents two specific alternative surrogate methods. The first is average cycle divergence rate, which detects a correlation

between the temporal cycles. For chaotic systems, the distance between two nearby cycles will increase exponentially over time, due to the sensitive dependence on initial conditions. The correlation between two cycles is expected to drop exponentially as the number of cycles increase. The second alternative method investigates the fluctuation of the degree of distribution of cycles in the phase space and is quantified through the variance of the normalized derivative. In this method, the degree distribution curve and the variance of the normalized derivative distribution are calculated. The degree distribution curve provides the distribution of cycles in the phase space. A chaotic system will show multiple distribution peaks, whereas a noisy periodic system will show a Poisson distribution of peaks. The peaks or smoothness of the distribution are quantified by the variance of the normalized derivative. A chaotic system will have a high value, and a noisy periodic system, which has more homogeneous smoothness, will have a low value.

A more recent surrogate method called the small shuffle surrogate was introduced by Nakamura and Small (2005) to investigate whether there are dynamics in irregular fluctuations (short term variability), even if the fluctuations are modulated by trends or periodicities. This algorithm generates surrogates that preserve long-term behaviors but destroys local structures. The null hypothesis is that irregular fluctuations are independently distributed (temporally uncorrelated) random variables. This algorithm changes the flow of information in the data and can be used to detect whether dynamics are present or not, regardless of whether the time series are linear or nonlinear. The authors propose that autocorrelation or average mutual information are appropriate to use as the discriminating techniques. The autocorrelation function and average mutual information answer the question regarding how much future data points are determined by past data points. To test the hypothesis, 39 surrogate time series are developed (two-sided test) and discriminating statistics calculated for the original and surrogate time series. If the values of the discriminating statistics of the original fall within the distribution of the surrogates, the null is not rejected. Time series with no dynamics (random process) have autocorrelation function, and average mutual information values fall within the distribution of the small surrogate shuffle distribution. Time series that contain dynamics will result in autocorrelation function and average mutual information values that are separate from the surrogate distribution. This method was robust in systems that contain long-term trends and those contaminated by stochastic noise.

One limitation of the small shuffle surrogate method is that it cannot distinguish between linear or nonlinear phenomena, as both types of systems exhibit some type of dynamics, which would lead to rejecting the null hypothesis. Nakamura et al. (2006) proposed a modification to the small shuffle surrogate that would test for the presence of nonlinearity in time series containing long-term trends and short-term fluctuations called truncated Fourier transform surrogate. This method assumes that the frequencies of irregular fluctuations are higher than the long-term trends and that when the data are linear, all phases can be treated as linear data when the power spectrum is preserved.

The null hypothesis for this algorithm is that irregular fluctuations are generated by a stationary linear system. The algorithm destroys nonlinearity in the irregular fluctuations and preserves the trends or periodicities. This is done by randomizing those phases from the power spectrum in the higher-frequency domains while maintaining the low-frequency phases. Thus, the major difference from the previous small shuffle surrogate is that not all phases are randomized, but only those in the higher-frequency domain. This algorithm requires the selection of a parameter to determine which frequencies will be randomized. This limitation of the truncated Fourier transform surrogate led Rios et al. (2015) to develop two new methods using decomposition to improve the surrogation techniques. The new techniques are the empirical mode decomposition–Fourier transform and empirical mode decomposition–amplitude-adjusted Fourier transform. These techniques rely on decomposing the data into a set of monocomponents plus residuals, with the residuals demonstrating the time series trend. Next, traditional surrogate methods are applied on each monocomponent, which results in a set of monocomponent surrogates. Finally, the set of surrogates is combined and retrended by adding the residuals back into the time series from the first step. This algorithm allows testing for the presence of linear and nonlinear behaviors in both stationary and nonstationary time series.

SUMMARY

In this chapter, we have discussed surrogate methods, which take the form of hypothesis testing. We examined several different surrogate algorithms, four linear surrogate methods, the pseudoperiodic surrogate method, and the small shuffle surrogate. The linear surrogate methods are designed to be applied to a stationary irregular time series without any long-term trend or periodicity, while the pseudoperiodic is applied to time series with periodicity. The small shuffle surrogate provides information regarding dynamics in irregular fluctuations, even if the fluctuations are modulated by trends or periodicities. Each algorithm generates a surrogate time series, which is consistent with a specific null hypothesis. Surrogate methods are used as an indirect approach to identify the nature of a time series. They try to narrow down the possibility of what a time series is by eliminating the possibilities of what a time series is not. Surrogate methods can identify whether a “hidden” structure exists within the data, but not necessarily tell if chaos exists. A surrogate method alone cannot decide what the time series is but is an extremely helpful tool when used with other nonlinear tools.

EXERCISES

1. State the general procedure of conducting a surrogate test.
2. State the null hypothesis of Algorithm 0.
3. State the null hypothesis of Algorithm 1.
4. State the null hypothesis of Algorithm 2.

5. What is the difference between Algorithm 2 and IAAFT?
6. Can we use linear surrogate methods when a time series exhibits strong periodicity? Why or why not?
7. State the null hypothesis of the PPS algorithm.
8. What three parameters are necessary for the PPS algorithm?
9. Can you determine what a time series is by using surrogate methods?
10. Explain two kinds of discriminating criteria.

APPENDIX 5.A: KNEE JOINT FLEXION/EXTENSION ANGLE (DEGREES)

Original Series

8.9612	12.578	25.092
9.7076	14.736	30.052
10.999	17.075	34.983
12.622	19.491	40.42
14.419	21.488	45.34
17.115	22.29	50.399
20.207	22.265	54.797
24.619	22.061	57.619
29.044	21.629	59.679
34.142	21.241	60.858
39.573	20.817	61.133
44.759	21.04	60.848
49.739	20.637	59.408
54.843	19.316	57.758
58.542	18.678	55.071
60.774	17.495	51.788
62.389	16.124	48.584
62.196	14.93	43.881
62.171	13.69	39.086
61.575	12.439	33.177
59.821	11.267	26.737
57.643	10.207	20.074
54.715	9.302	13.561
51.169	8.5962	8.1246
47.151	7.7525	4.782
42.075	7.2274	2.8538

PPS Surrogate Series

16.558	6.6854	45.34
17.772	7.0044	49.475
16.352	7.5202	54.155
15.246	8.3092	56.516
12.269	9.5507	58.711
11.255	10.634	59.818
10.341	12.254	60.459
8.8039	14.471	60.116
7.7423	17.542	58.864
9.302	21.09	56.819
8.5962	25.092	54.73
7.7525	30.052	51.788
7.2274	34.983	48.584
6.8028	40.42	43.881
39.086	13.089	16.833
33.177	15.602	15.748
28.6	18.385	14.811
23.003	25.092	13.767
17.55	26.077	12.133
12.775	34.983	10.818
8.3107	40.42	10.144
4.782	45.34	9.6492
4.7991	50.399	9.1698
5.4521	54.797	8.8224
7.252	57.619	9.3043
10.041	59.679	8.6863

35.762	6.8028	2.8926	12.566	61.933	8.6412
29.722	6.6854	4.4985	14.798	61.578	8.9573
23.035	7.0044	7.0089	17.933	59.408	9.8134
16.215	7.5202	9.2506	20.797	57.758	8.9071
11.144	8.3092	10.876	23.074	55.071	9.8471
6.6787	9.5507	12.778	23.312	51.788	10.48
4.3748	10.634	15.692	22.75	48.584	12.879
3.4397	12.254	18.496	21.821	43.881	18.64
4.6151	14.471	20.266	21.809	39.086	22.644
7.3564	17.542	20.526	21.635	33.177	26.84
10.249	21.09	20.45	21.346	26.737	31.358
20.485	34.294	18.385	20.976	20.331	36.497
20.501	28.575	22.163	20.004	14.562	41.339
20.069	22.217	26.077	19.105	10.1	46.514
19.438	15.664	30.343	17.911	6.2972	51.32
19.219	10.865	34.79	16.509	7.4885	55.239
18.51	6.7063	39.653	15.515	8.1968	58.468
17.474	4.36	44.762	15.515	10.589	59.867
16.718	3.2073	49.475	14.619	13.886	59.928
15.219	3.9469	54.155	14.001	16.381	61.25
14.25	6.542	56.516	11.015	18.542	61.654
13.251	8.452	58.711	8.8039	20.648	59.089
12.093	11.194	59.818	7.7423	20.571	56.91
11.015	13.136	60.459	6.3726	22.131	54.625
9.7966	14.735	60.116	5.5298	22.771	51.881
8.6375	16.063	58.864	4.9179	22.457	48.264
8.0029	18.232	56.819	4.9681	22.075	45.083
7.4718	19.856	54.73	4.9942	21.818	40.315
7.0539	20.942	51.357	5.4474	21.668	34.958
7.0923	21.685	47.767	5.7399	21.053	28.6
7.624	21.156	43.319	6.7546	20.976	22.251
7.991	20.515	38.25	7.8583	20.004	15.247
8.9692	20.305	31.784	9.234	18.694	10.459
10.382	19.543	25.41	11.085	17.514	8.0987
12.074	18.632	18.241	7.4885	57.409	8.2302
14.513	17.751	12.775	8.1968	59.772	8.2302
17.673	17.33	8.3107	5.4572	60.752	7.3732
21.398	15.568	5.7746	7.9289	61.196	8.0175
25.6	14.439	4.7991	13.743	59.528	9.3349
30.065	12.788	5.4521	16.208	60.719	10.395
34.683	11.664	7.252	18.112	59.582	11.618
39.781	10.341	10.041	19.88	57.577	13.484
45.339	8.8039	12.566	22.265	55.071	15.74

49.923	7.7423	14.798	22.589	51.788	12.254
54.053	6.3726	17.933	23.685	48.584	14.471
57.503	5.5298	20.797	23.955	40.315	17.542
59.414	4.9179	23.074	23.039	34.958	21.09
60.413	4.9681	23.312	22.423	27.034	25.092
60.424	4.9942	22.75	21.961	20.331	30.052
60.287	5.4474	21.821	21.158	13.561	34.983
59.089	5.7399	21.809	20.332	10.865	40.42
56.91	6.7546	21.635	19.148	6.7063	45.34
54.625	7.8583	21.346	17.707	4.36	50.399
51.825	9.234	20.976	16.558	3.2073	54.797
47.954	11.085	20.004	15.607	3.9469	57.619
44.343	13.089	18.694	14.216	6.542	59.679
39.828	15.602	17.514	13.369	9.2506	60.858
16.833	6.172	59.904	12.272	10.876	60.459
15.748	8.7978	60.064	11.467	12.778	60.116
14.811	11.818	59.735	10.876	14.735	59.582
13.767	14.175	59.006	9.6281	16.063	57.577
13.046	16.626	57.236	8.833	18.232	54.643
12.233	19.409	55.149	8.2302	23.398	48.584
11.395	21.455	51.881	7.1681	23.498	43.881
11.043	22.41	48.264	8.6673	23.151	39.086
10.098	21.86	45.083	8.6682	23.046	31.784
9.3043	21.326	40.315	5.7399	21.959	25.41
8.6863	21.17	34.958	9.4577	21.603	18.241
8.6412	20.953	29.086	10.94	21.213	12.775
8.9573	20.711	23.003	12.879	20.607	8.3107
9.8134	19.943	17.55	15.348	19.525	5.7746
11.062	19.012	12.304	18.331	18.704	4.7991
13.032	17.888	8.9619	21.913	17.439	5.4521
15.424	16.889	7.1428	26.323	16.742	7.252
18.64	15.826	6.5955	30.906	15.858	10.041
22.644	14.575	7.6152	34.676	15.029	12.566
26.84	13.611	9.3094	38.509	13.786	14.798
31.358	12.269	11.563	43.778	13.241	17.933
36.497	11.255	13.711	48.795	9.302	14.735
41.339	9.9009	15.32	53.255	8.5962	16.063
46.514	9.164	18.008	18.232	54.73	18.64
51.32	7.9652	20.571	19.856	51.357	22.644
55.239	7.6236	22.131	20.45	47.767	26.84
58.468	6.8844	22.771	20.45	43.319	31.358
59.867	6.7978	22.457	20.485	40.315	36.497
60.618	6.8674	22.075	20.501	34.958	41.339

60.685	7.3732	21.818	20.515	29.086	46.514
59.528	8.0175	21.668	20.305	23.003	51.32
58.682	9.0707	21.053	19.543	17.55	55.239
56.665	10.739	20.497	18.632	12.304	58.468
54.062	12.226	20.088	17.751	8.9619	59.867
51.362	14.413	19.105	17.33	7.1428	60.618
47.681	17.15	17.911	15.219	6.5955	60.685
43.842	20.42	16.509	14.25	10.589	61.22
38.084	24.323	15.515	13.251	13.886	59.521
32.691	28.578	14.619	12.093	16.381	57.577
26.653	33.036	14.001	11.015	18.542	54.643
20.583	37.901	12.752	9.7966	20.648	51.513
14.562	43.144	11.41	7.7423	23.207	45.418
10.1	47.896	10.876	9.6281	24.454	40.387
6.719	52.56	9.6281	8.833	23.568	37.955
4.9821	55.894	8.833	8.2302	23.362	32.211
4.8589	58.446	8.2302	7.1681	22.61	25.922
7.1681	23.498	47.993	7.5316	22.823	20.583
7.5316	23.151	42.995	7.6462	21.783	14.671
7.6462	23.046	37.955	8.155	20.661	10.459
8.155	22.685	32.211	9.0405	19.748	8.0987
9.0405	22.416	25.922	10.48	18.595	7.3227
10.48	22.112	19.329	12.019	17.153	6.8289
12.019	21.613	13.066	14.143	16.45	7.6683
14.143	20.718	8.1002	16.536	15.826	10.584
16.536	19.749	4.9102	20.392	16.124	12.135
19.937	18.704	3.4886	24.59	15.246	14.628
23.942	17.439	3.7459	28.798	14.534	16.402
28.48	16.742	5.4572	33.379	13.217	18.799
33.381	15.858	7.9289	38.638	12.099	20.716
38.509	15.029	10.387	43.54	11.107	23.362
43.778	13.786	12.289	48.942	10.065	24.356
48.795	13.241	14.95	53.917	9.5547	23.884
53.255	12.462	18.168	57.409	9.081	23.039
57.514	11.624	20.786	59.772	7.6539	22.423
59.803	11.14	22.191	60.413	7.8662	21.961
60.379	10.688	22.535	60.424	7.0044	21.158
60.745	10.094	22.263	60.116	11.062	21.814
61.129	10.217	22.175	58.864	13.032	21.437
59.998	10.508	21.973	56.819	15.424	20.085
58.18	11.127	21.959	19.294	16.208	60.848
56.05	12.127	21.603	17.772	18.112	59.408
52.888	13.695	21.213	14.216	19.88	56.819

49.539	15.538	20.607	12.272	22.265	54.73
45.418	17.848	19.525	11.467	23.548	51.357
40.387	20.735	18.82	10.207	23.685	47.767
35.095	24.59	17.654	9.302	22.263	43.319
28.6	28.798	16.639	8.5962	22.175	38.25
22.251	33.379	15.502	9.0675	21.973	31.784
15.247	38.638	14.507	8.6673	21.959	25.41
10.352	43.54	13.862	8.6682	21.603	18.241
6.2817	48.942	12.873	7.0044	21.213	12.775
4.618	53.522	12.018	6.7546	20.607	8.1246
4.2304	57.697	10.793	12.226	16.889	6.2817
5.4652	59.845	10.054	18.64	15.826	4.618
8.1121	61.454	9.0675	22.644	14.575	3.2073
10.589	61.933	8.6673	26.84	13.611	6.172
12.855	61.578	8.6682	31.358	13.767	8.7978
14.994	61.012	8.7991	36.497	13.046	9.2506
17.341	59.332	9.3349	41.339	12.233	10.876
19.835	57.317	10.757	46.514	11.395	14.994
21.98	54.946	12.119	51.32	11.043	18.008
23.398	51.479	14.471	55.239	10.098	20.571
16.985	20.085	10.92	58.468	9.3043	22.131
20.392	19.294	8.6063	59.867	8.6863	22.771
24.368	17.772	7.0982	60.752	8.6412	22.457
28.529	16.352	6.5939	61.196	8.9573	22.075
33.403	15.246	8.1424	59.528	9.8134	21.818
38.097	14.534	10.584	58.682	11.127	21.668
43.153	13.217	13.743	59.332	10.48	21.053
48.927	12.099	16.208	57.317	14.368	20.497
53.825	11.107	18.112	54.946	12.254	18.632
57.878	10.065	19.88	51.479	14.471	19.105
60.481	9.5547	22.265	47.993	17.542	16.889
61.898	9.081	23.548	42.995	21.09	15.826
62.526	8.6737	24.356	37.955	25.092	13.894
62.566	8.5694	23.884	33.049	30.052	13.174
61.809	8.8998	23.039	27.394	34.983	12.269
60.491	9.3981	22.423	21.394	40.42	13.046
58.281	10.395	21.961	13.066	45.34	11.467
55.879	11.618	21.158	9.3651	50.399	10.7
52.87	13.484	20.332	7.3227	54.797	9.7143
49.361	15.74	19.148	6.8289	57.619	8.8553
45.07	18.586	17.707	7.6683	59.679	9.3043
40.907	21.849	16.558	10.584	60.858	8.6863
35.457	25.907	15.607	13.743	61.133	8.6412

29.854	30.262	14.216	7.5316	21.158	8.9619
23.848	34.676	13.369	8.8998	20.332	4.618
17.678	39.309	12.272	11.062	19.148	4.2304
13.003	44.134	11.467	13.032	19.525	5.4652
9.3651	49.565	10.7	15.424	18.82	6.542
7.3227	53.917	9.7143	15.538	17.654	8.452
6.8289	57.409	8.8553	17.848	16.639	11.194
7.6683	59.772	8.6787	20.735	13.369	13.136
9.6401	60.752	7.9992	24.368	12.233	14.735
12.135	61.196	7.6539	28.529	11.395	16.063
14.628	61.22	7.8662	36.497	11.043	18.232
16.402	60.719	7.9757	41.339	7.6236	19.856
18.799	59.112	8.4738	46.514	8.2302	22.457
20.716	56.994	9.4577	51.32	7.1681	21.821
22.589	54.797	10.94	55.239	8.6412	21.809
23.685	51.513	12.879	58.468	8.9573	21.635
23.955	47.928	15.348	59.867	9.8134	21.346
23.613	43.74	18.331	60.618	8.4738	20.976
23.391	38.666	21.913	59.528	9.4577	20.004
22.926	33.049	26.323	58.281	10.94	18.694
22.43	27.394	30.906	55.879	12.019	17.514
21.814	21.394	35.439	52.87	14.143	16.833
21.437	15.337	40.819	49.361	16.536	15.748
46.206	24.454	16.796	45.07	19.937	14.811
50.497	23.568	19.678	40.907	23.942	13.767
54.896	23.362	23.314	35.457	28.48	12.752
58.094	22.61	27.119	29.854	33.381	13.786
59.928	22.823	32.058	23.848	38.509	13.241
60.787	21.783	37.112	17.678	44.134	12.462
61.003	20.661	42.176	13.003	49.565	11.624
60.438	19.748	47.556	9.3651	53.522	11.14
59.521	18.595	52.439	7.3227	57.697	10.688
57.651	17.153	56.779	6.8289	59.845	10.094
55.088	16.45	59.074	7.6683	61.454	10.217
51.606	14.902	60.953	7.9289	61.933	11.127
48.204	13.894	61.25	10.387	61.578	12.127
44.285	13.174	61.654	12.289	61.012	13.695
39.07	12.133	60.786	14.95	58.281	15.538
33.371	10.818	59.582	18.168	52.888	17.848
27.034	10.144	57.577	20.786	49.539	20.735
20.331	9.6492	54.643	22.191	45.418	24.368
14.671	9.1698	51.619	23.498	40.387	28.529
10.459	8.8224	47.6	23.884	35.095	33.403

8.0987	8.5052	43.06	23.039	28.6	38.097
7.4885	8.4115	37.446	22.423	23.003	43.153
8.1968	8.2302	31.552	21.961	14.562	48.927
10.589	8.8595	24.704	53.825	58.682	38.084
13.886	8.9071	18.513	57.878	56.665	32.691
16.381	9.8471	12.488	60.481	54.062	26.653
18.542	10.919	8.4084	61.898	51.825	20.583
20.648	12.492	6.2972	61.196	47.954	14.562
23.207	14.368		61.22	43.842	

REFERENCES

- Acharya, U. R., O. Faust, N. Kannathal, T. Chua, and S. Laxminarayan. (2005). Non-linear analysis of EEG signals at various sleep stages. *Computer Methods and Programs in Biomedicine* 80 (1): 37–45.
- Amato, I. (June 26, 1992). Chaos breaks out at NIH, but order may come of it. *Science* 256 (5065): 1763–1764.
- Barahona, M. and C. Poon. (May 16, 1996). Detection of nonlinear dynamics in short, noisy time series. *Nature* 381: 215–217.
- Breakspear, M. and J. R. Terry. (2002). Nonlinear interdependence in neural systems: Motivation, theory, and relevance. *The International Journal of Neuroscience* 112 (10): 1263–1284.
- Buchman, T. G., J. P. Cobb, A. S. Lapedes, and T. B. Kepler. (2001). Complex systems analysis: A tool for shock research. *Shock* 16 (4): 248–251.
- Buzzi, U. H., N. Stergiou, M. J. Kurz, P. A. Hageman, and J. Heidel. (2003). Nonlinear dynamics indicates aging affects variability during gait. *Clinical Biomechanics* 18 (5): 435–443.
- Cavanaugh, J. T., N. Kochi, and N. Stergiou. (2010). Nonlinear analysis of ambulatory activity patterns in community-dwelling older adults. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 65 (2): 197–203.
- Chang, T., S. J. Schiff, T. Sauer, J. P. Gossard, and R. E. Burke. (1994). Stochastic versus deterministic variability in simple neuronal circuits: I. monosynaptic spinal cord reflexes. *Biophysical Journal* 67 (2): 671–683.
- Cignetti, F., F. Schena, and A. Rouard. (2009). Effects of fatigue on inter-cycle variability in cross-country skiing. *Journal of Biomechanics* 42 (10): 1452–1459.

- Collins, J. J. and C. J. De Luca. (1995). The effects of visual input on open-loop and closed-loop postural control mechanisms. *Experimental Brain Research* 103 (1): 151–163.
- Costa, M. D., T. Henriques, M. N. Munshi, A. R. Segal, and A. L. Goldberger. (2014). Dynamical glucometry: Use of multiscale entropy analysis in diabetes. *Chaos* 24 (3): 033139.
- Diks, C. (1996). Estimating invariants of noisy attractors. *Physical Review E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 53 (5): R4263–R4266.
- Dingwell, J. B. and J. P. Cusumano. (2000). Nonlinear time series analysis of normal and pathological human walking. *Chaos* 10 (4): 848–863.
- Ehlers, C. L., J. Havstad, D. Prichard, and J. Theiler. (1998). Low doses of ethanol reduce evidence for nonlinear structure in brain activity. *The Journal of Neuroscience* 18 (18): 7474–7486.
- Garfinkel, A., M. L. Spano, W. L. Ditto, and J. N. Weiss. (Aug 28, 1992). Controlling cardiac chaos. *Science* 257 (5074): 1230–1235.
- Georgoulis, A. D., C. Moraiti, S. Ristanis, and N. Stergiou. (2006). A novel approach to measure variability in the anterior cruciate ligament deficient knee during walking: The use of the approximate entropy in orthopaedics. *Journal of Clinical Monitoring and Computing* 20 (1): 11–18.
- Goldstein, B., D. Towell, S. Lai, K. Sonnenthal, and B. Kimberly. (1998). Uncoupling of the autonomic and cardiovascular systems in acute brain injury. *The American Journal of Physiology* 275 (4 Pt 2): R1287–R1292.
- Govindan, R. B., K. Narayanan, and M. S. Gopinathan. (1998). On the evidence of deterministic chaos in ECG: Surrogate and predictability analysis. *Chaos* 8 (2): 495–502.
- Grassberger, P. and I. Procaccia. (1983). Measuring the strangeness of strange attractors. *Physica D* 9: 189–208.
- Hausdorff, J. M., C. K. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger. (1995). Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. *Journal of Applied Physiology* 78 (1): 349–358.
- Ivanov, D. K., H. A. Posch, and C. Stumpf. (1996). Statistical measures derived from the correlation integrals of physiological time series. *Chaos* 6 (2): 243–253.
- Janjarsjitt, S., M. S. Scher, and K. A. Loparo. (2008). Nonlinear dynamical analysis of the neonatal EEG time series: The relationship between sleep state and complexity. *Clinical Neurophysiology* 119 (8): 1812–1823.

- Kantz, H. and T. Schreiber. (1997). *Nonlinear Time-Series Analysis*. Cambridge, U.K.: Cambridge University Press.
- Kaplan, D. and L. Glass. (1995). *Understanding Nonlinear Dynamics*. New York: Springer-Verlag.
- Keenan, D. (1985). A Tukey non-additivity type test of time series nonlinearity. *Biometrika* 72 (1): 39–44.
- Kugiumtzis, D. (1999). Test your surrogate data before you test for nonlinearity. *Physical Review E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 60 (3): 2808–2816.
- Kugiumtzis, D. (2001). On the reliability of the surrogate data test for nonlinearity in the analysis of noisy time series. *International Journal of Bifurcation and Chaos* 11 (7): 1881–1896.
- Kunhimangalam, R., P. K. Joseph, and O. K. Sujith. (2008). Nonlinear analysis of EEG signals: Surrogate data analysis. *IRBM* 29 (4): 239–244.
- Kurz, M. J., T. N. Judkins, C. Arellano, and M. Scott-Pandorf. (2008). A passive dynamic walking robot that has a deterministic nonlinear gait. *Journal of Biomechanics* 41 (6): 1310–1316.
- Ladislao, L. and S. Fioretti. (2007). Nonlinear analysis of posturographic data. *Medical & Biological Engineering & Computing* 45 (7): 679–688.
- Lamoth, C. J., E. Ainsworth, W. Polomski, and H. Houdijk. (2010). Variability and stability analysis of walking of transfemoral amputees. *Medical Engineering & Physics* 32 (9): 1009–1014.
- Lamoth, C. J., F. J. van Deudekom, J. P. van Campen, B. A. Appels, O. J. de Vries, and M. Pijnappels. (2011). Gait stability and variability measures show effects of impaired cognition and dual tasking in frail people. *Journal of Neuroengineering and Rehabilitation* 17 (8): 2.
- Lehnertz, K., R. G. Andrzejak, J. Arnhold, T. Kreuz, F. Mormann, C. Rieke, G. Widman, and C. E. Elger. (2001). Nonlinear EEG analysis in epilepsy: Its possible use for interictal focus localization, seizure anticipation, and prevention. *Journal of Clinical Neurophysiology* 18 (3): 209–222.
- Little, M. A., P. E. McSharry, I. M. Moroz, and S. J. Roberts (2006). Testing the assumptions of linear prediction analysis in normal vowels. *The Journal of the Acoustical Society of America* 119 (1): 549–558.
- Martinerie, J., C. Adam, M. Le Van Quyen, M. Baulac, S. Clemenceau, B. Renault, and F. J. Varela. (1998). Epileptic seizures can be anticipated by non-linear analysis. *Nature Medicine* 4 (10): 1173–1176.

- Miller, D. J., N. Stergiou, and M. J. Kurz. (2006). An improved surrogate method for detecting the presence of chaos in gait. *Journal of Biomechanics* 39 (15): 2873–2876.
- Mitra, S., M. A. Riley, and M. T. Turvey. (1997). Chaos in human rhythmic movement. *Journal of Motor Behavior* 29 (3): 195–198.
- Myers, S. A., J. M. Johanning, I. I. Pipinos, K. K. Schmid, and N. Stergiou. (2012). Vascular occlusion affects gait variability patterns of healthy younger and older individuals. *Annals of Biomedical Engineering*.
- Myers, S. A., J. M. Johanning, I. I. Pipinos, K. K. Schmid, and N. Stergiou. (Aug, 2013). Vascular occlusion affects gait variability patterns of healthy younger and older individuals. *Annals of Biomedical Engineering* 41 (8): 1692–1702.
- Nakamura, T. and M. Small. (2005). Small-shuffle surrogate data: Testing for dynamics in fluctuating data with trends. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics* 72 (5 Pt 2): 056216.
- Nakamura, T., M. Small, and Y. Hirata. (2006). Testing for nonlinearity in irregular fluctuations with long-term trends. *Physical Review E* 74: 026205.
- Nurujjaman, M. D., R. Narayanan, and A. N. S. Iyengar. (2009). Comparative study of nonlinear properties of EEG signals of normal persons and epileptic patients. *Nonlinear Biomedical Physics* 3: 6.
- Orsucci, F. F. (2006). The paradigm of complexity in clinical neurocognitive science. *The Neuroscientist* 12 (5): 390–397.
- Osborne, A. and A. Provencale. (1989). Finite correlation dimension for stochastic systems with power-law spectra. *Physica D* 35: 357–381.
- Palus, M. (1995). Testing for nonlinearity using redundancies: Quantitative and qualitative aspects. *Physica D* 80: 186–205.
- Palus, M. (1996). Nonlinearity in normal human EEG: Cycles, temporal asymmetry, nonstationarity and randomness, not chaos. *Biological Cybernetics* 75 (5): 389–396.
- Peng, C. K., S. Havlin, H. E. Stanley, and A. L. Goldberger. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* 5 (1): 82–87.
- Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America* 88 (6): 2297–2301.
- Pincus, S. M., I. M. Gladstone, and R. A. Ehrenkranz (1991). A regularity statistic for medical data analysis. *Journal of Clinical Monitoring* 7 (4): 335–345.

- Poggi, D., A. Porporato, L. Ridolfi, J. Albertson, and G. Katul. (2004). Interaction between large and small scales in the canopy sublayer. *Geophysical Research Letters* 31 (5): L05102.
- Pompe, B. (1993). Measuring statistical dependences in a time series. *Journal of Statistical Physics* 73 (3/4): 587–610.
- Porta, A., S. Guzzetti, R. Furlan, T. Gnecci-Ruscione, N. Montano, and A. Malliani. (2007). Complexity and nonlinearity in short-term heart period variability: Comparison of methods based on local nonlinear prediction. *IEEE Transactions on Bio-Medical Engineering* 54 (1): 94–106.
- Preatoni, E., M. Ferrario, G. Donà, J. Hamill, and R. Rodano. (2010). Motor variability in sports: A non-linear analysis of race walking. *Journal of Sports Sciences* 28 (12): 1327–1336.
- Prichard, D. and J. Theiler. (1995). Generalized redundancies for time series analysis. *Physica D* 84 (476): 493.
- Rapp, P. E., A. M. Albano, T. I. Schmah, and L. A. Farwell. (1993). Filtered noise can mimic low-dimensional chaotic attractors. *Physical Review E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 47 (4): 2289–2297.
- Rathleff, M. S., A. Samani, C. G. Olesen, U. G. Kersting, and P. Madeleine. (2011). Inverse relationship between the complexity of midfoot kinematics and muscle activation in patients with medial tibial stress syndrome. *Journal of Electromyography and Kinesiology* 21 (4): 638–644.
- Rieke, C., F. Mormann, R. G. Andrzejak, T. Kreuz, P. David, C. E. Elger, and K. Lehnertz. (2003). Discerning nonstationarity from nonlinearity in seizure-free and pre-seizure EEG recordings from epilepsy patients. *IEEE Transactions on Bio-Medical Engineering* 50 (5): 634–639.
- Rios, R., M. Small, and R. de Mello. (2015). Testing for linear and nonlinear Gaussian processes in nonstationary time series. *International Journal of Bifurcation and Chaos* 25 (1): 155013.
- Rombouts, S., R. Keunen, and C. Stam. (1995). Investigation of nonlinear structure in multichannel EEG. *Physics Letters A* 202: 352–358.
- Schreiber, T. and A. Schmitz. (1997). Classification of time series data with nonlinear similarity measures. *Physical Review Letters* 79: 1475.
- Schreiber, T. and A. Schmitz. (2000). Surrogate time series. *Physica D* 142: 346–382.
- Slutzky, M. W., P. Cvitanovic, and D. J. Mogul. (2001). Deterministic chaos and noise in three in vitro hippocampal models of epilepsy. *Annals of Biomedical Engineering* 29 (7): 607–618.

- Small, M. (2005). *Applied Nonlinear Time Series Analysis: Application in Physics, Physiology and Finance*. Singapore: World Scientific Publishing.
- Small, M. and K. Judd. (1998). Comparison of new nonlinear modelling techniques with applications to infant respiration. *Physica D* 117: 283–298.
- Small, M. and C. K. Tse. (2002). Minimum description length neural networks for time series prediction. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics* 66 (6 Pt 2): 066701.
- Small, M., D. Yu, and R. G. Harrison. (2001). Surrogate test for pseudoperiodic time series data. *Physical Review Letters* 87 (18): 188101.
- Stam, C. J., T. C. van Woerkom, and R. W. Keunen. (1997). Non-linear analysis of the electroencephalogram in Creutzfeldt-Jakob disease. *Biological Cybernetics* 77 (4): 247–256.
- Stergiou, N., U. H. Buzzi, M. J. Kurz, and J. Heidel. (2004). Nonlinear tools in human movement. In *Innovative Analyses of Human Movement*, ed. N. Stergiou, pp. 63–90. Champaign, IL: Human Kinetics Publ.
- 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.
- Theiler, J. and P. E. Rapp. (1996). Re-examination of the evidence for low-dimensional, nonlinear structure in the human electroencephalogram. *Electroencephalography and Clinical Neurophysiology* 98 (3): 213–222.
- Toweill, D. L. and B. Goldstein. (1998). Linear and nonlinear dynamics and the pathophysiology of shock. *New Horizons* 6 (2): 155–168.
- Tsay, R. (1986). Nonlinearity tests for time series. *Biometrika* 73: 461–466.
- Wagner, C. D., B. Nafz, and P. B. Persson. (1996). Chaos in blood pressure control. *Cardiovascular Research* 31 (3): 380–387.
- Wolf, A., J. B. Swift, H. L. Swinney, and J. A. Vastano. (1985). Determining Lyapunov exponents from a time series. *Physica* 16D: 285–317.
- Wurdeman, S. R., S. A. Myers, and N. Stergiou. (2014). Amputation effects on the underlying complexity within transtibial amputee ankle motion. *Chaos* 24 (1): 013140.
- Yentes, J. M., N. Hunt, K. K. Schmid, J. P. Kaipust, D. McGrath, and N. Stergiou. (Feb, 2013). The appropriate use of approximate entropy and sample entropy with short data sets. *Annals of Biomedical Engineering* 41 (2): 349–365.
- Yu, D., M. Small, R. G. Harrison, and C. Diks. (2000). Efficient implementation of the Gaussian kernel algorithm in estimating invariants and noise level from noisy

time series data. *Physical Review E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 61 (4 Pt A): 3750–3756.

Zhang, J., X. Luo, T. Nakamura, J. Sun, and M. Small. (2007). Detecting temporal and spatial correlations in pseudoperiodic time series. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics* 75 (1 Pt 2): 016218.

Zhao, Y., J. Sun, and M. Small. (2008). Evidence consistent with deterministic chaos in human cardiac data: Surrogate and nonlinear dynamical modeling. *International Journal of Bifurcation and Chaos* 18 (1): 141.