International Journal of Bifurcation and Chaos, Vol. 11, No. 4 (2001) 983–997

SURROGATE DATA PATHOLOGIES AND THE FALSE-POSITIVE REJECTION OF THE NULL HYPOTHESIS

P. E. RAPP*,†, C. J. CELLUCCI‡ and T. A. A. WATANABE†
†*Department of Pharmacology and Physiology,*
The Medical College of Pennsylvania and Hahnemann University,
The Arthur P. Noyes Clinical Research Center,
Norristown State Hospital, Norristown, Pennsylvania, USA
‡*Department of Physics, Ursinus College,*
Collegeville, Pennsylvania, USA

A. M. ALBANO
Physics Department, Bryn Mawr College,
Bryn Mawr, Pennsylvania, USA

T. I. SCHMAH
Département de Mathématiques, École Polytechnique Fédérale de Lausanne,
Lausanne, Switzerland

Received May 23, 2000; Revised June 29, 2000

It is shown that inappropriately constructed random phase surrogates can give false-positive rejections of the surrogate null hypothesis. Specifically, the procedure erroneously indicated the presence of deterministic, nonlinear structure in a time series that was constructed by linearly filtering normally distributed random numbers. It is shown that the erroneous identification was due to numerical errors in the estimation of the signal's Fourier transform. In the example examined here, the introduction of data windowing into the algorithm eliminated the false-positive rejection of the null hypothesis. Additional guidelines for the use of surrogates are considered, and the results of a comparison test of random phase surrogates, Gaussian scaled surrogates and iterative surrogates are presented.

1. Introduction: The Method of Surrogate Data

The method of surrogate data [Theiler *et al.*, 1992; Schreiber & Schmitz, 1999]; for citations of the earlier surrogate literature see [Rapp *et al.*, 1993] has become a central tool for validating the results of dynamical analysis. Generically stated, the procedure can be reduced to four steps. First, a dynamical measure, for example, the Lyapunov exponent or correlation dimension is applied to the original time series obtaining the result M_{Orig}. Second, surrogate data sets, which will be described presently, are constructed using the original data set. Third, the dynamical measure that was applied to the original time series is applied to the surrogates. The average value of the measure obtained with surrogates

---

*Author for the correspondence. Clinical Research Center, Norristown State Hospital, 1001 Sterigere Street, Norristown, PA 19401, USA.
E-mail: Paul.E.Rapp@Drexel.edu

is denoted $\langle M_{\mathrm{Surr}}\rangle$. Fourth, a statistical criterion is used to determine if M_{Orig} and $\langle M_{\mathrm{Surr}}\rangle$ are sufficiently different. If they are, the surrogate null hypothesis (the original and the surrogate data come from the same population) is rejected.

The procedure used to construct surrogate data sets from the original time series depends on the null hypothesis being examined. If the null hypothesis is: under measure M, the original time series is indistinguishable from identically distributed random numbers, then the surrogate time series can be constructed by a random shuffle of the time series. A random shuffle is a very unexacting test of deterministic structure. Filtered random numbers, for example, have a sequence-sensitive structure that is destroyed by a random shuffle. If a sequence-sensitive measure is used, filtered noise appropriately rejects the random shuffle null hypothesis. Random phase surrogates address the null hypothesis: under measure M the original time series is indistinguishable from linearly filtered random numbers. Random phase surrogates are constructed by calculating the original time series' Fourier transform, randomizing the phase and calculating the inverse transform. The resulting time series is the surrogate data set.

It is essential to test the software used to generate surrogate data against well standardized time series of both types, that is, against deterministic signals where the null hypothesis should be rejected, and against time series based on random numbers and filtered random numbers where the null hypothesis should not be rejected. The importance of validation calculations with the second class of example is inadequately appreciated. Using a statistical criterion the null hypothesis is rejected if M_{Orig} and are $\langle M_{\mathrm{Surr}}\rangle$ are sufficiently different. Errors in surrogate software can result in surrogate data sets that are pathologically different from the original time series. If this occurs, $\langle M_{\mathrm{Surr}}\rangle$ is always very different from $\langle M_{\mathrm{Orig}}\rangle$, and a false-positive rejection of the surrogate null hypothesis results. Tests where a failure to reject the null hypothesis is the correct response are essential.

An example of this validation procedure was constructed with three data sets. The first data set was a time series of 8,192 iterates of the x-coordinate of the Lorenz equations.

$$\frac{dx}{dt} = cy - cx$$

$$\frac{dy}{dt} = rx - y - xz$$

$$\frac{dz}{dt} = -bz + xy$$

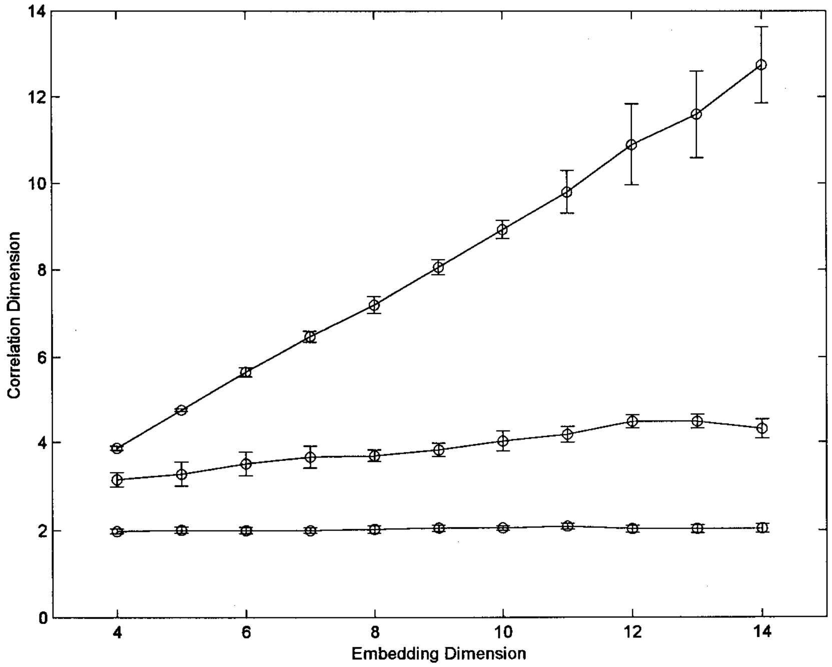


Fig. 1. Estimated correlation dimension for three time series as a function of embedding dimension. The three time series were: uniformly distributed random numbers (upper function), the time series produced by linearly filtering that random data (middle function), and a time series generated by integrating the Lorenz equations (lower function).

where $c = 10$, $b = 8/3$, $r = 28$, $\delta t = 0.006$ and $x_0 = y_0 = z_0 = 1$. The second data set, is composed of random numbers uniformly distributed on $[0, 1]$. The third data set was constructed by linearly filtering the set of random numbers. The Fourier transform of 8,192 points was calculated. The jth harmonic of the Fourier series was multiplied by F_j

$$F_j = \max[0,\, 1 - kj^2] \quad k = 0.37 \times 10^{-6}$$

The filtered data set is the inverse transform of this modified Fourier series. In these calculations, the applied measure was the correlation dimension. The procedure implemented here for estimating the correlation dimension is outlined in the first appendix. In all calculations, 8,192 points were used. The results are presented in Fig. 1.

The estimated correlation dimension of the random numbers is seen to increase with the embedding dimension. (The embedding dimension is defined in Appendix A.) This behavior is consistent with random systems. In the case of the Lorenz data set the correlation dimension, approximately 2.03, is stable with increasing embedding dimension. This is often taken as being a definitive identifying characteristic of determinist systems. However, in the case of the filtered random numbers the function of correlation dimension versus embedding dimension is also nearly constant once the embedding dimension is sufficiently large. Historically, it was this type of observation that motivated the construction of random phase surrogates.

The results obtained with fifty random phase surrogates for these three systems are shown in Fig. 2. Examined visually, it seems that in the case of the Lorenz data, the values obtained with the original signal and with its surrogates are clearly separated. This indicates that the surrogate null hypothesis can be rejected. That is, under this measure, the signal can be distinguished from linearly filtered random numbers. Figure 2 also shows the results obtained with random data. As expected, the values of the correlation dimension for the original time series and its surrogates are essentially indistinguishable. This is as it should be; random numbers should fail to reject the surrogate null hypothesis. The results obtained with filtered random numbers and its random phase surrogates are also similar to each other, suggesting that the null hypothesis cannot be rejected. Thus, the procedure correctly identified the underlying random nature of the dynamical process that generated the filtered time series.

Rather than relying on a visual examination, a systematic statistical procedure should be used to justify rejection of the surrogate null hypothesis. Three candidate criteria for doing this have been outlined in a previous publication [Rapp *et al.*, 1994]. In summary they are as follows. Under the nonparametric criterion [Barnard, 1963; Hope, 1968], the null hypothesis is rejected only if $M_{\text{Orig}} > M_{\text{Surr}}$ or $M_{\text{Orig}} < M_{\text{Surr}}$ for all surrogates. This is a very demanding criterion. A single aberrant value of M_{Surr} can result in the failure to reject the null hypothesis. As an alternative, one can calculate P_M, the Monte Carlo probability of the surrogate null hypothesis. In the case where the discriminating measure is dimension, the value of dimension obtained with the original time series will, in the case of a deterministic system, be less than that obtained with random phase surrogates. In that case

$$P_M = \frac{(\text{Number of Cases } D \le D_{\text{Orig}})}{(\text{Number of Cases})}\,,$$

because the value in the numerator includes the contribution of the original time series, $P_M > 0$. The number in the denominator is the number of surrogates computed plus one. P_M is particularly appropriate if the number of surrogates is small. (The question of an appropriate choice of N_{Surr}, the number of surrogates, is addressed presently.)

If the values of the discriminating measure obtained with surrogates form a normal distribution, then a more sophisticated analysis is possible. The Z value is defined as

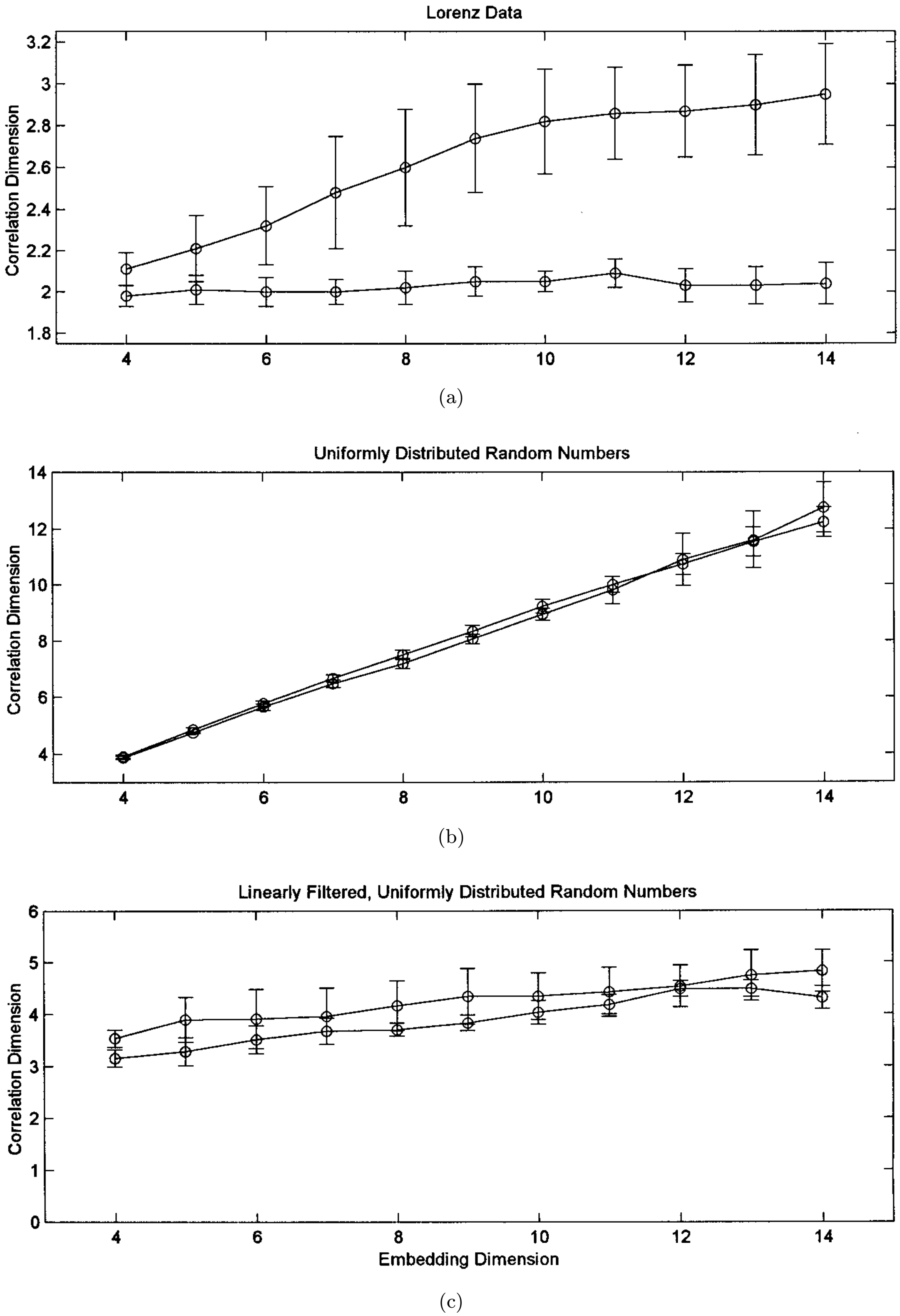
$$Z = \frac{|M_{\text{Orig}} - \langle M_{\text{Surr}} \rangle|}{\sigma_{\text{Surr}}}$$

where σ_{Surr} is the standard deviation of $\langle M_{\text{Surr}} \rangle$. Given the assumption of a normal distribution, for a dynamical measure such as dimension where there is *a priori* expectation that the value obtained with a deterministic signal is less than the value obtained with its surrogates, the probability of the null hypothesis, P_Z, is given by

$$P_Z = \frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{Z}{\sqrt{2}}\right)\right)$$

For the one-tailed test, the null hypothesis is rejected at the 5% confidence level if $Z \le 1.65$ and at the 1% level if $Z \le 2.33$

An additional question must be addressed: for any given time series and dynamical measure, how many surrogates should be used? There is no universally applicable answer to this question. The



Lorenz Data

(a)

Uniformly Distributed Random Numbers

(b)

Linearly Filtered, Uniformly Distributed Random Numbers

(c)

Fig. 2. Estimated correlation dimension for the original time series compared with the average correlation dimension estimate obtained with 50 random phase surrogates. (a) In the case of the Lorenz data the divergence of dimension estimates obtained with the original data (lower function) and its surrogates (upper function) results in a rejection of the null hypothesis. (b) In the cases of uniformly distributed random numbers and (c) linearly filtered, uniformly distributed random numbers, the superposition or near superposition of estimates obtained with the original data and its surrogates results in the expected failure to reject the surrogate null hypothesis.

number of surrogates required will depend on the quality of the original signal and on the numerical stability of the applied dynamical measure. An empirical assessment can be made by calculating P_M or Z as a function of the number of surrogates. N_{Surr} should be increased until a stable value is obtained. As described in the discussion of the results presented in Table 1, a comparison was made using 10 and 50 surrogates. Six different data sets were examined. The applied measure was the correlation dimension. On average, the value of Z obtained with 10 surrogates was within 13.9% of the Z value obtained with 50 surrogates. All results reported in this paper are based on 50 surrogate calculations.

The three statistical criteria described above were applied to the three example systems of Fig. 2 (Lorenz data, uniformly distributed random numbers, and linearly filtered random numbers). In each case embedding dimensions $4, 5, \ldots, 14$ were calculated with 50 surrogates. In the case of Lorenz data, the null hypothesis was rejected using the nonparametric criterion for all embedding dimension except $m = 5$ and $m = 8$. In the case of $m = 5$ two of the 50 surrogates gave values of dimension less than D_{Orig}. In the case of $m = 8$, one surrogate of 50 gave a value of dimension less than D_{Orig}. This argues against the unconsidered use of the nonparametric criterion.

For the Lorenz data, the value of P_M, the Monte Carlo probability of the null hypothesis, averaged over the eleven embeddings was $P_M = 0.025$. Recall that the lowest possible value of P_M is determined by the number of surrogates. If $N_{\text{Surr}} = 50$, then the lowest possible value of P_M is $P_M = 0.020$. Averaged over eleven embeddings, the average value of Z was found to be $Z = 2.61$, which corresponds to a probability of the null hypothesis of less that 1%. This is seen in Fig 2. The values of dimension obtained with the original signal and with its surrogates diverge. In aggregate, the calculations provide, as they should, a convincing rejection of the surrogate null hypothesis. We can conclude that the Lorenz system can be distinguished from linearly filtered noise.

The results obtained with uniformly distributed random numbers and random phase surrogates also coincide with expectations. Under the nonparametric criterion, there was a failure to reject the null hypothesis in all eleven embeddings. The Monte Carlo criterion also failed to reject the null hypothesis with an average value of $P_M = 0.315$. Similarly, the average value of Z was $Z = 1.04$. As seen in Fig. 2, the values of dimension obtained with the original data set and its surrogates coincide. The random nature of the signal was correctly identified.

Table 1. Average Z scores and corresponding one-tailed probabilities obtained with 50 surrogates.

	Random Phase Surrogates	Gaussian Scaled Surrogates	Iterative Surrogates
Hénon	32.29 $p < 10^{-6}$	33.33 $p < 10^{-6}$	30.16 $p < 10^{-6}$
Gaussian Distributed Random	1.01 $p = 0.16$	0.79 $p = 0.21$	0.87 $p = 0.19$
Linearly Filtered Gaussian Distributed	1.27 $p = 0.10$	0.97 $p = 0.17$	1.86 $p = 0.03$
Uniformly Distributed Random	0.94 $p = 0.17$	0.98 $p = 0.16$	0.88 $p = 0.19$
Linearly Filtered Uniformly Distributed Random	0.50 $p = 0.31$	0.96 $p = 0.17$	0.71 $p = 0.24$
Nonlinearly Transformed, Linearly Filtered, Uniformly Distributed Random	6.25 $p < 10^{-6}$	0.85 $p = 0.20$	2.17 $p = 0.02$

The analysis of the filtered random signal using random phase surrogates also correctly identified the underlying random structure of the process. Based on the nonparametric criterion, there was a failure to reject the null hypothesis for all embeddings except $m = 4$. The average value of the Monte Carlo probability of the null hypothesis was $P_M = 0.280$, and the value of Z averaged over the eleven embeddings was $Z = 0.896$. The values of dimension obtained with the original signal and with its surrogates very nearly coincide (Fig. 2).

At this stage of the investigation it might seem that the algorithm used to generate random phase surrogates has been successfully validated. The null hypothesis is rejected when a deterministic signal is examined, and there is a failure to reject the null hypothesis when random signals, including filtered random numbers, are evaluated. However, an example constructed in the next section indicates that this is not the case.

2. False-Positive Rejection of the Surrogate Null Hypothesis: An Example

Pradhan and his colleagues [Pradhan *et al.*, 1997] have published a computational example of the false-positive rejection of the random phase surrogate null hypothesis. Based on this example, they have questioned the value of surrogate data in the examination of biological time series. In this section we present a similar example that follows their presentation.

We began with 16,384 normally distributed random numbers with a mean of zero and a variance of one. The Fourier transform of this data set was calculated. The coefficients of the jth harmonic were multiplied by one for $j = 1, \ldots, 192$ and by zero for $j = 193, \ldots, 8{,}192$. The inverse transform was then calculated. A filtered noise data set was constructed by extracting the middle 8,192

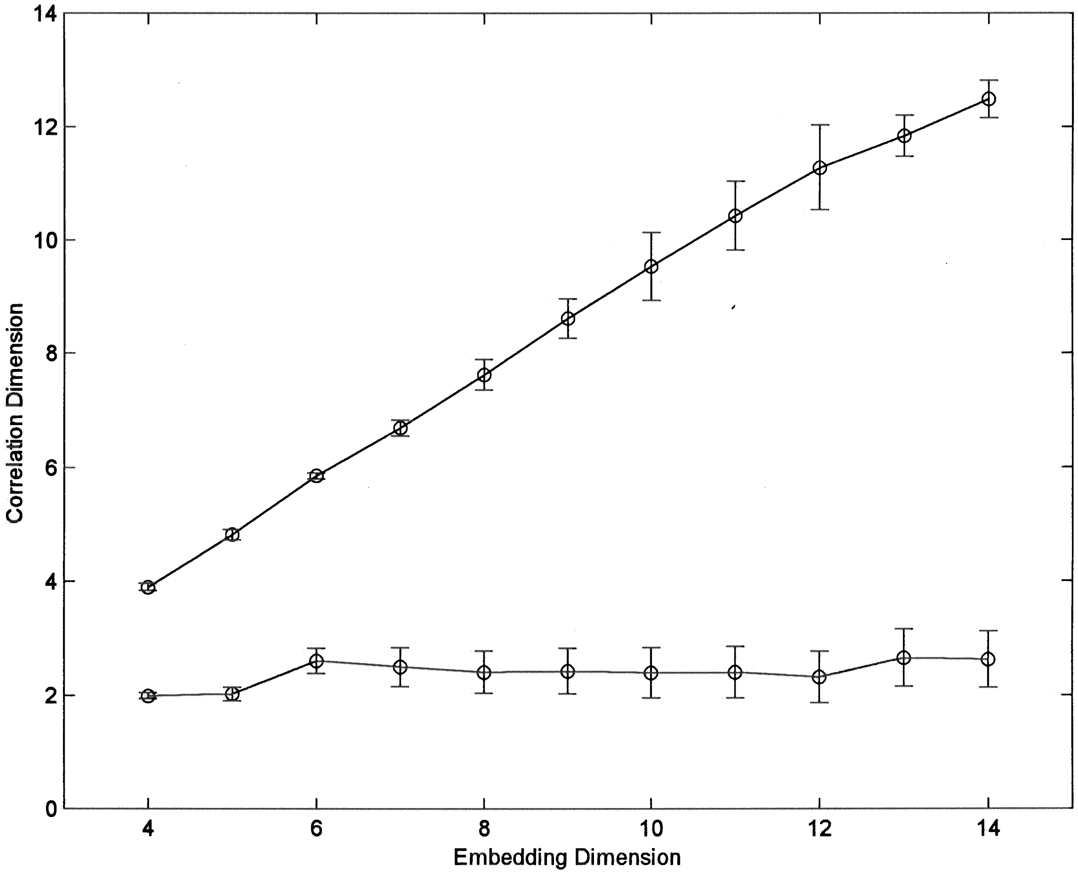
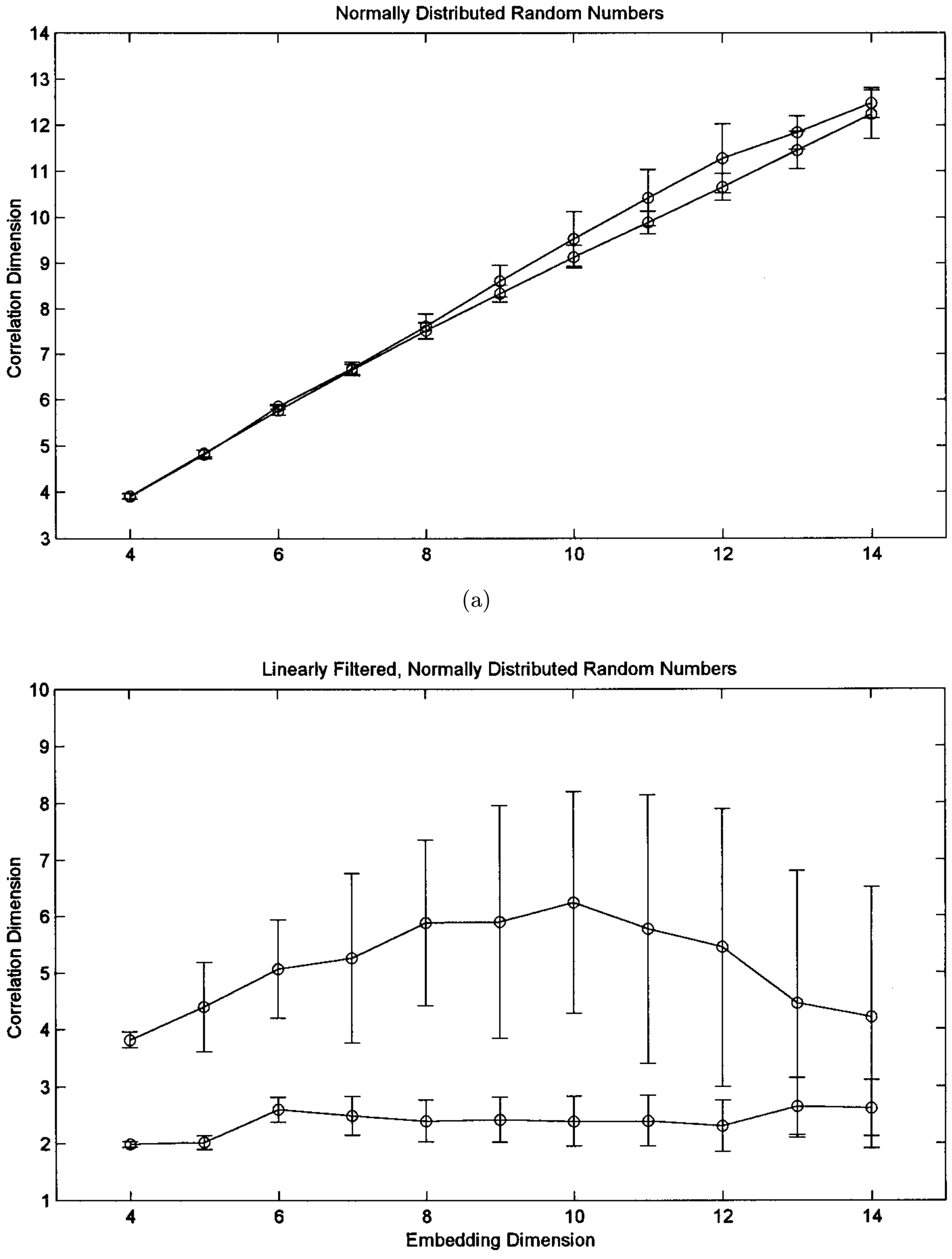


Fig. 3. Estimated correlation dimension for two time series as a function of embedding dimension. The two time series are normally distributed random numbers (the upper function) and a time series produced by linearly filtering this random data (lower function). The stability of the estimated correlation dimension as a function of embedding dimension in the case of filtered noise indicates that this cannot be used as a criterion for low dimensional, deterministic structure.

points from the inverse transform. The first 8,192 normally distributed points of the original data set were used as an unfiltered comparison set. This filtered data set differs from the filtered data set of the previous section in the degree of filtering applied to the random numbers. In the previous case, the first 1,643 coefficients made nonzero contributions to the inverse transform. In the second case, only 192 coefficients were retained.

The values of the correlation dimension as a function of embedding dimension are shown in Fig. 3. The value of dimension obtained with



(a)

(b)

Fig. 4. Estimated correlation dimension for the original time series and the average correlation dimension estimates obtained with 50 random phase surrogates as a function of embedding dimension. (a) In the case of normally distributed random numbers superposition of dimension estimates obtained with the original data and with its surrogates results in the anticipated failure to reject the surrogate null hypothesis. (b) In the case of filtered random numbers the divergence of values obtained with the filtered signal and its surrogates produces a false-positive rejection of the surrogate null hypothesis.

normally distributed random numbers increases with embedding dimension. In the case of the filtered random numbers, a stable value of dimension versus embedding dimension is obtained.

The comparison of these dimension estimates with 50 random phase surrogates is shown in Fig. 4. In the case of the unfiltered, normally distributed numbers there is a failure to reject the surrogate null hypothesis in all embeddings when the nonparametric criterion is applied. The average Monte Carlo probability is $P_M = 0.777$, and the average value of Z is 1.02. As shown in Fig. 4, the values of dimension obtained with the original data set and its surrogates coincide. Thus the unfiltered, normally distributed random numbers fail to reject the null hypothesis. This is as it should be.

This contrasts with the results obtained with the heavily filtered random signal and its random phase surrogates. When the very demanding nonparametric criterion as used, the null hypothesis was rejected only when $m = 4$. The average Monte Carlo probability was $P_M = 0.144$. However, when Z scores were computed, it was found that the average Z score was $Z = 2.81$. This is reflected in Fig. 4. In the case of filtered random numbers, there is a separation of the dimension values obtained with the filtered signal and with its surrogates. That is, there is a false-positive rejection of surrogate null hypothesis for this time series. When the correlation dimension is used as the discriminating metric, the method suggests the presence of an underlying deterministic structure in a signal produced by linearly filtering random numbers. The resolution of the problem, at least for this time series and dynamical measure, is constructed in the next section.

3. Resolution of the Example of the False-Positive Rejection of the Surrogate Null Hypothesis

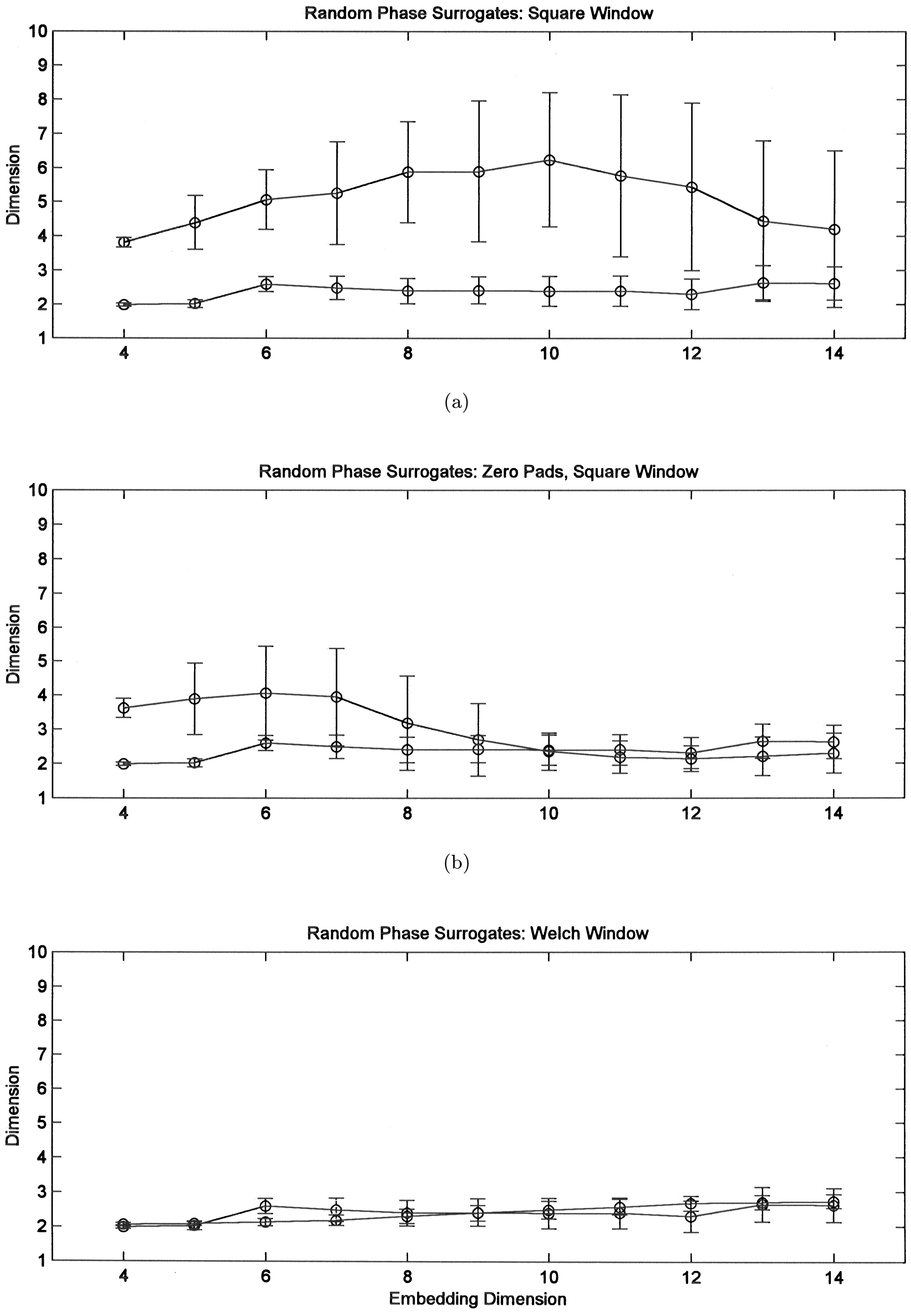
Calculation of the Fourier transform is the central step in the construction of a random phase surrogate. If a pathological surrogate result is obtained, an examination of the calculation of the transform is indicated. A crude but often successful procedure for improving the robustness of a Fourier transform calculation is to introduce zero pads. Zero pads are strings of zeros introduced at the beginning and end of the time series before the Fourier transform is calculated. They suppress wrap-around effects. The reader is referred to [Press *et al.*, 1992] for a more systematic presentation. In this implementation, the first and last 200 points of the 8,192 point time series were set equal to zero before random phase surrogates were constructed. Though this is a very primitive implementation of zero padding, it was nonetheless effective in eliminating the false-positive rejection of the surrogate null hypothesis. These results are shown in Fig. 5. Using the nonparametric criterion, the null hypothesis is rejected for embedding dimension $m = 4$. The hypothesis is not rejected for dimensions $m = 5, 6, \ldots, 14$. The average value of the Monte Carlo probability is $P_M = 0.513$, and the average value of Z is $Z = 1.17$. The introduction of minimal zero pads thus eliminated the previously observed false-positive rejection of the random phase null hypothesis.

A more systematic and mathematically satisfying approach to improve the quality of a computed Fourier transform is to introduce windowing. Let $\{x_j, j = 1, \ldots, N\}$ denote the original time series. A window is introduced by multiplying x_j by w_j, where w_j is zero, or close to it, at $j = 1$ and N, and equal to 1 at $j = N/2$. Several alternative windows have been introduced [Press *et al.*, 1992]. We have followed the Press *et al.* recommendation and used the Welch window.

$$w_j = 1 - \left[\frac{j - \frac{N}{2}}{\frac{N}{2}}\right]^2$$

The results are shown in Fig. 5. With the Welch window in place, there is a nonparametric failure to reject the null hypothesis in all embeddings. The average value of P_M is $P_M = 0.415$, and the average value of Z is $Z = 1.27$. As seen in Fig. 5, the introduction of the Welch window in the construction of surrogates results in surrogate dimension values that coincide with the values obtained with the original data set.

A comparison of random phase surrogates constructed without the Welch window (this is called a square window in the spectral analysis literature) and with a Welch window using the same seed to the random number generator is given in Fig. 6. This comparison makes it possible to identify the cause of the false-positive result obtained with the square window. In the calculation using the square window, numerical errors in the calculated transform introduce high frequency artifacts into the inverse transform. The artifacts resulted in aberrantly high value of surrogate dimension estimates,

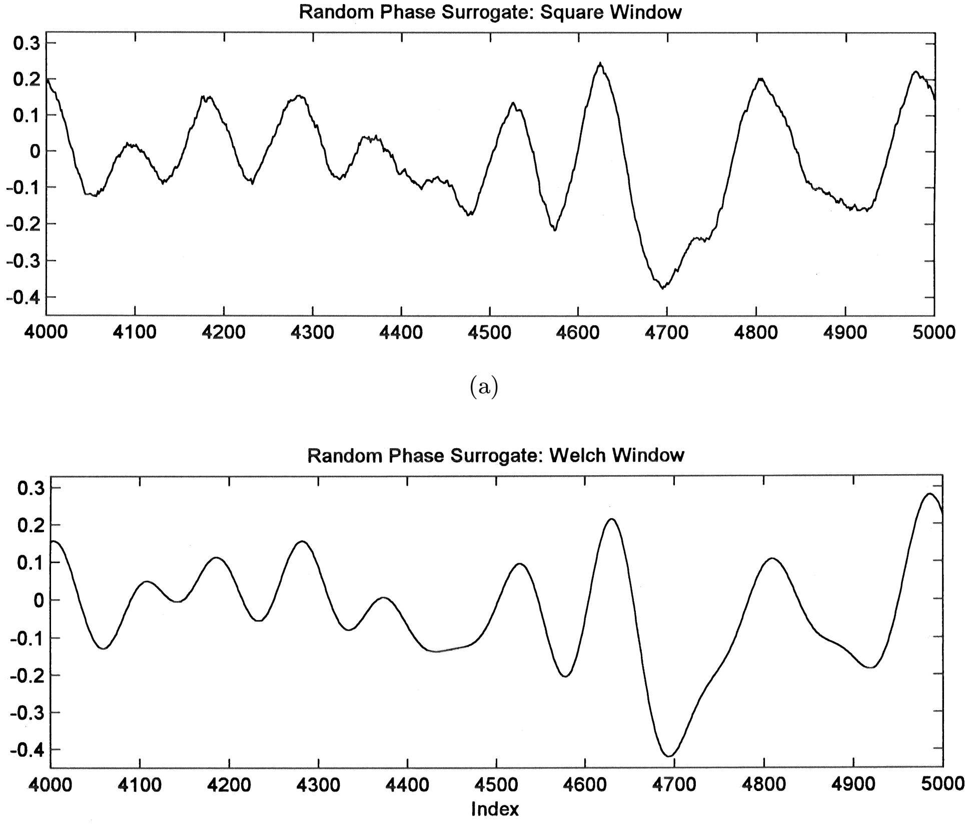


(a)

(b)

(c)

Fig. 5. Dimension estimates obtained from linearly filtered, normally distributed random numbers compared against the average dimension estimates obtained with 50 random phase surrogates. The surrogates were calculated with three implementations of the algorithm. (a) A square window was used. (b) A square window and zero pads were used. (c) A Welch window was used.



(b)

Fig. 6. Comparison of random phase surrogates calculated using different implementations of the algorithm. (a) A square window was used. (b) A Welch window was used.

which produced the fallacious rejection of the surrogate null hypothesis. In all subsequent calculations, the Welch window is used when random phase surrogates are constructed.

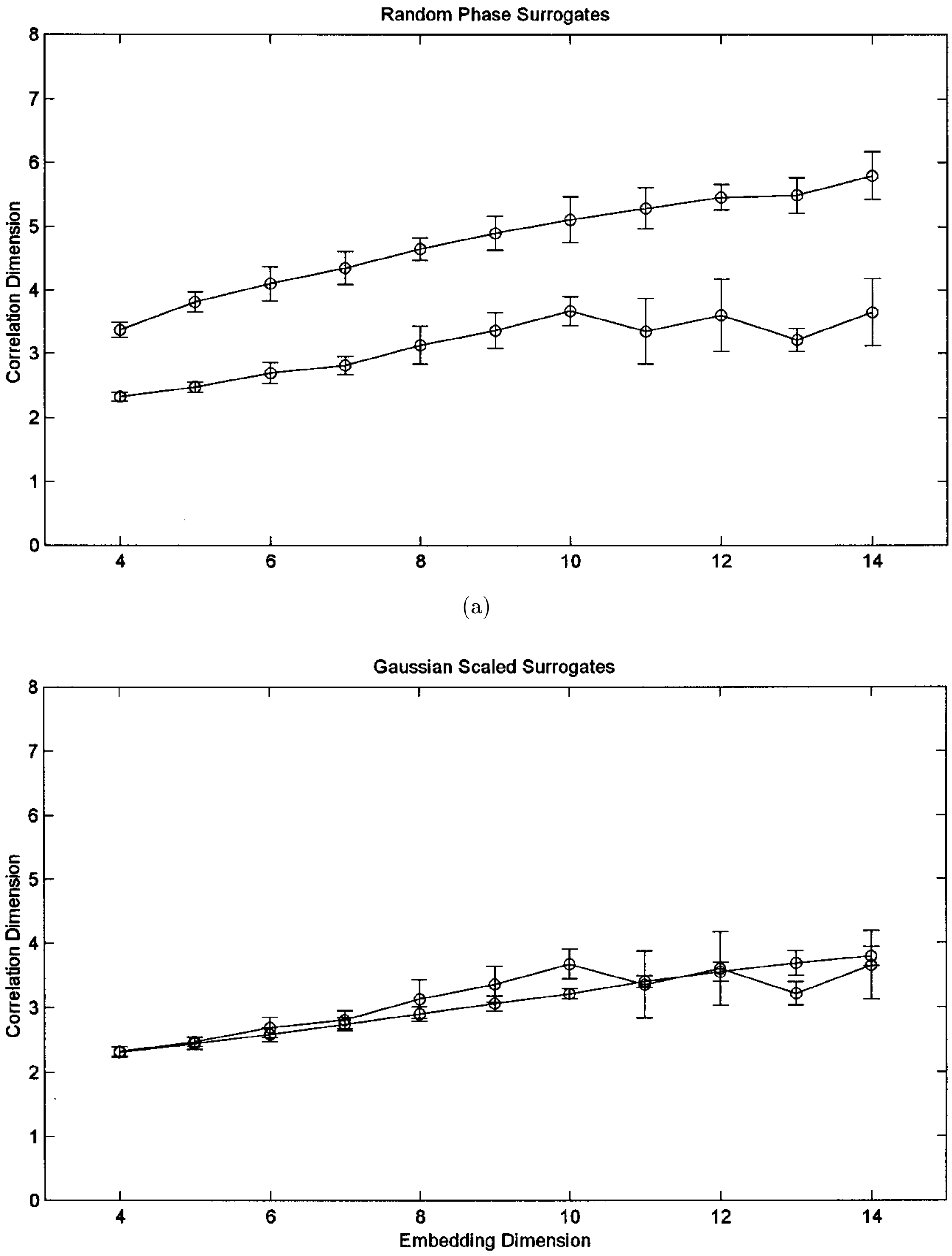
4. Nonlinear Transformations of Linearly Filtered Noise

The limitations of the random phase surrogate null hypothesis should be recognized explicitly. The null hypothesis states that under the applied metric, the original signal is indistinguishable from linearly filtered noise. It is sometimes supposed that rejection of this null hypothesis indicates that the signal has an underlying deterministic structure. This is not the case. We consider here in greater detail an example considered in an earlier contribution [Rapp *et al.*, 1993]. Consider the following possibility. Suppose that the signal examined in Fig. 1 that was constructed by linearly filtering uniformly distributed random numbers is transformed by a static, monotonic nonlinear function $h(x)$.

$$h(x) = \frac{\left[\dfrac{x - x_{\text{min}} - 0.0001}{x_{\text{max}} - x + 0.0001}\right]^{\rho}}{1 + \left[\dfrac{x - x_{\text{min}} - 0.0001}{x_{\text{max}} - x + 0.0001}\right]^{\rho}}$$

where x_{min} and x_{max} are the minimum and maximum value of x in the original time series, and $\rho = 3$. As before, the applied measure is the correlation dimension.

Random phase surrogates were constructed using a Welch window, and the corresponding values of dimension were calculated. As shown in Fig. 7, a definitive rejection of the null hypothesis was obtained. The null hypothesis is rejected with the nonparametric criterion in all embeddings. The average value of P_M is $P_M = 0.02$, which is the lowest value that can be obtained when 50 surrogates are calculated. The average value of Z is $Z = 6.24$. However, the procedure has not failed. This does not constitute a false-positive rejection of the null hypothesis. The random phase null hypothesis



(a)

(b)

Fig. 7. Estimated correlation dimension for nonlinearly transformed, linearly filtered, uniformly distributed random numbers compared against the average dimension estimate obtained with 50 surrogates. (a) Random phase surrogates calculated with a Welch window were used. The null hypothesis was rejected. (b) Gaussian scaled surrogates calculated with a Welch window were used. The null hypothesis was not rejected.

supposes that the signal was generated by linearly filtering random numbers. The application of the nonlinear transformation $h(x)$ violates that null hypothesis.

Amplitude adjusted or Gaussian scaled surrogates address a more demanding null hypothesis: under the applied measure, the signal is indistinguishable from linearly filtered noise that has been transformed by a static, monotone nonlinearity. The procedure for constructing surrogates of this class [Theiler *et al.*, 1992] is outlined in the second appendix. With this class of surrogate, the underlying random structure of the signal was correctly identified. Using the nonparametric criterion, there was a failure to reject the null hypothesis in all eleven embeddings. The average value of P_M was $P_M = 0.71$ and the average value of Z was $Z = 0.85$. The dimension values obtained with surrogates coincide with the dimension values obtained with the original time series (Fig. 7).

5. Comparison Tests of Random Phase, Gaussian Scaled and Iterative Surrogates

Schreiber and Schmitz [1996] have proposed an iterative algorithm for constructing surrogates that have the same power spectrum and the same distribution as the original time series. As in the preceding sections of this paper, the objective is to avoid false-positive identification of nonlinear deterministic structure in the data. A specification of the algorithm is given in Appendix C.

Three surrogate algorithms were compared and tested against six test signals. The algorithms were random phase surrogates, Gaussian scaled surrogates, and an independent implementation of the Schreiber–Schmitz iterative surrogates. The six test signals included a time series generated by the Hénon map ($x_{n+1} = 1 - ax_n^2 + y_n$, $y_{n+1} bx_n$, $a = 1.4$, $b = 0.3$) and the time series described in earlier sections of this paper. Each time series contained 8,192 points. Fifty surrogates of each type were constructed. The discriminating measure was the correlation dimension. Calculations were performed for embedding dimensions $m = 4$ to $m = 14$. The results are shown in Table 1. A Z score is calculated for each embedding dimension. In the table, the average Z score ($m = 4$ to $m = 14$) is displayed along with the corresponding single-tailed probability of the null hypothesis.

Expected results were obtained in the case of the Hénon map. The surrogate null hypothesis was rejected using all three classes of surrogate. In the case of normally distributed random numbers, all three classes of surrogate failed to reject the null hypothesis. This is also as it should be.

More interested results were obtained with the signal produced by heavily filtering the normally distributed random numbers using the filter described in Sec. 2. Based on an examination of the average Z score, there is a false-positive rejection of the null hypothesis when iterative surrogates are used. However, this result should be examined with care. The average Monte Carlo probability of the null hypothesis is $P_M = 0.160$. The high average Z score is due to high scores obtained with lower embedding dimensions. The average Z score for $m = 7$ to $m = 14$ is $Z = 1.23$. A convincing rejection of the null hypothesis is, therefore, not obtained, which is consistent with the stochastic origin of the signal.

The results obtained with uniformly distributed random numbers are consistent with expectations. A failure to reject the null hypothesis is observed. This is also the case when linearly filtered, uniformly distributed random numbers were tested. (In this case, the filter is the filter described in Sec. 1.)

The final entries in Table 1 present the results obtained when the nonlinear transformation of the linearly filtered random numbers were examined. As previously discussed, the rejection of the random phase null hypothesis is expected and does not constitute a surrogate failure since, by construction, the signal violates the random-phase null hypothesis. The failure to reject the null hypothesis when Gaussian scaled surrogates are used is also the expected result. Using this type of surrogate, the stochastic origin of the signal was correctly identified. On first examination, the iterative surrogates would seem to produce a false-positive rejection of the null hypothesis, but as in the previous case, this only occurs in lower embedding dimension when the estimated value of dimension is nearly equal to the embedding dimension. Under these circumstances, the calculation of the correlation dimension is of uncertain reliability. The average Z score of embeddings $m = 7$ to $m = 14$ is $Z = 1.60$.

6. Conclusions

A number of conclusions can be drawn from the calculations. First, it was found that it is essential to validate surrogate software with examples where the null hypothesis should not be rejected as well as with examples that result in a legitimate rejection of the surrogate null hypothesis. Second, care must be exercised when estimating a Fourier transform. Numerical errors that can be tolerated in some applications can have unacceptable consequences when they are incorporated into surrogates. Third, it is advisable to use more than one type of surrogate. If results obtained with different classes of surrogate are inconsistent, a more detailed examination, for example an assessment of the calculation's dependence on embedding dimension, should be made. Fourth, as addressed in a previous publication [Rapp *et al.*, 1994], a sufficiently large number of surrogates should be used. If anomalous results are obtained, the number of surrogates used should be increased.

Finally, we wish to emphasize the limitations of the conclusions that can be supported even by carefully validated surrogate calculations. Invalid generalizations are sometimes made when there is a failure to reject the surrogate null hypothesis. Let us suppose that a systematic calculation results in a well-documented failure to reject the random phase surrogate null hypothesis. This does not mean that a deterministic nonlinear structure is not present in the signal. Rather, it can only be concluded that the discriminating measure used failed to detect such a structure. It always remains possible that calculations using the same data, the same surrogates but a different measure could result in a convincing rejection of the null hypothesis.

Acknowledgments

We would like to acknowledge support from the Controlled Biological Systems Program at DARPA from the U.S. Department of Education Award No. H235J000001 to the Krasnow Institute from the Janssen Pharmaceutica Research Foundation, and from the Bristol–Myers Squibb Pharmaceutical Research Institute. The encouragement and support of R. J. Josiassen, Director of the Arthur Noyes Clinical Research Center is gratefully acknowledged.

References

Barnard, G. A. [1963] "Discussion on the spectral analysis of point processes by M. S. Bartlett," *J. R. Statist. Soc.* **B25**, p. 294.

Grassberger, P. & Procaccia, I. [1983a] "Measuring the strangeness of strange attractors," *Physica.* **D9**, 189–208.

Grassberger, P. & Procaccia, I. [1983b] "Characterization of strange attractors," *Phys. Rev. Lett.* **50**, 346–349.

Hope, A. C. A. [1968] "A simplified Monte Carlo significance test procedure," *J. R. Statist. Soc.* **B30**, 582–598.

Pradhan, N. & Sadasivan, P. K. [1997] "Validity of dimensional complexity measures of EEG signals," *Int. J. Bifurcation and Chaos* **7**, 173–186.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. [1992] *Numerical Recipes. The Art of Scientific Computing*, 2nd edition (Cambridge University Press, Cambridge).

Rapp, P. E., Albano, A. M., Schmah, T. I. & Farwell, L. A. [1993] "Filtered noise can mimic low dimensional chaotic attractors," *Phys. Rev.* **E47**, 2289–2297.

Rapp, P. E., Albano, A. M., Zimmerman, I. D. & Jiménez-Montaño, M. A. [1994] "Phase-randomized surrogates can produce spurious identifications of non-random structure," *Phys. Lett.* **A192**, 27–33

Schreiber, T. & Schmitz, A. [1996] "Improved surrogate data for nonlinearity tests," *Phys. Rev. Lett.* **77**, 635–638.

Schreiber, T. & Schmitz, A. [2000] "Surrogate time series," *Physica* **D142**, 346–382.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. [1992] "Testing for nonlinearity in time series: The method of surrogate data," *Physica* **D58**, 77–94.

Appendix A
Estimation of the Correlation Dimension

The mathematical basis for this procedure is given in [Grassberger & Procaccia, 1983a, 1983b]. Let $\{x_j,\, j = 1, \ldots, N\}$ denote a single channel time series. These values are embedded in $\Re^m$ to produce points Z_j, where

$$Z_j = (x_j,\, x_{j+L}, \ldots,\, x_{j+(m-1)L})$$

K denotes the number of points in the embedded set; $K = N - (m-1)L$. The correlation integral is then calculated.

$$C_m(r) = \frac{1}{N_p} \sum_{i=1}^{K-B} \sum_{j=i+B}^{K} \Theta(r - |Z_i - Z_j|)$$

Θ is the Heaviside function. In our implementation the Euclidean metric is used. Correlations between consecutively measured points can give false indications of low dimensional structure [Theiler, 1986; Theiler & Rapp, 1996]. It is possible to eliminate these correlations by incorporating parameter B into the expression for the correlation integral [Theiler, 1986]. In these calculations, B is at least twice the autocorrelation time of the time series. N_p is the number of distinct pairs of points used in the calculation of the correlation integral.

$$N_p = (K - B + 1)(K - B)/2$$

The logarithm $\ln C_m(r)$ is examined as a function of $\ln(r)$. An attempt is made to find a linear scaling region. The slope of $\ln C_m(r)$ versus $\ln(r)$ in the scaling region is the correlation dimension [Grassberger & Procaccia, 1983a, 1983b]. The upper and lower bounds of the scaling region were determined numerically. The derivative $dC_m(r)/d(\ln(r))$ was calculated. The longest interval on the $\ln(r)$ axis such that the variation of the derivative is ±0.35 was found. The slope of $\ln C_m(r)$ in that interval became one of the estimates of the correlation dimension. This process was repeated for scaling regions where the variation of the derivative is ±0.30, ±0.25, ±0.20, ±0.15, ±0.10 and ±0.05. The final estimate of the correlation dimension is the mean of these seven estimates. The standard deviation of that mean is a measure of the uncertainty of the reported value of the correlation dimension.

In the calculations presented here $L = 1$ and $m = 4, \ldots, 14$. A stable value of dimension as a function of embedding dimension m is an important, but, as demonstrated, not definitive indication of low dimensional, deterministic structure in the time series.

Appendix B
Gaussian Scaled Surrogates

Again let $\{x_j,\, j = 1, \ldots, N\}$ denote the original time series. The object is to produce $\{x'_j,\, j = 1, \ldots, N\}$ a Gaussian scaled surrogate of $\{x\}$.

1. $\{y_j,\, j = 1, \ldots, N\}$ is produced by taking a normally distributed set of random numbers and arranging the elements in the same order as the elements of $\{x\}$; that is if x_j is the kth largest value of $\{x\}$, then y_j is the kth largest value of $\{y\}$.
2. A random phase surrogate of $\{y\}$ is constructed. Calculate the Fourier series of $\{y\}$. In our implementation of the algorithm, a Welch window is used. Superscript Y is used to denote the resulting Fourier coefficients: $a_0^Y, a_1^Y, \ldots, a_{N/2}^Y$, $b_1^Y, \ldots, b_{N/2}^Y$. A temporary Fourier series that has the same amplitude but random phase is constructed.
$$a_j^T = [(a_j^Y)^2 + (b_j^Y)^2]^{1/2} \cos\phi$$
$$b_j^T = [(a_j^Y)^2 + (b_j^Y)^2]^{1/2} \sin\phi$$
where ϕ is drawn from a uniformly distributed set of random numbers on $[0, 2\pi]$. $\{y'\}$ is the corresponding inverse transform.
3. $\{x'\}$, the Gaussian scaled surrogate of $\{x\}$, is constructed by reordering the elements of $\{x\}$ so that it has the same rank structure as $\{y'\}$. A justification of the algorithm is given in [Theiler *et al.*, 1992] and in [Rapp *et al.*, 1993].

Appendix C
Schreiber–Schmitz Iterative Surrogates

Let $\{x\}$ denote the original time series of N elements. The procedure for constructing $\{x^U\}$, the first iterate of the surrogate, is presented.

1. As an implementation detail, construct $\{x^{\text{Sort}}\}$ which is the data vector $\{x\}$ ordered from the smallest to the largest element. This is done here so that is not unnecessarily incorporated into the iterative process.
2. Calculate the Fourier series of the original data. Superscript D is used to denote this Fourier series: $a_0^D, a_1^D, a_2^D, \ldots a_{N/2}^D, b_1^D, b_2^D, \ldots b_{N/2}^D$. (Note that throughout, a_j denotes the real cosine coefficient. This is in contrast with Krantz and Schreiber who use a_j to denote the amplitude of the jth harmonic.)
3. Construct $\{x^S\}$ a random shuffle of $\{x\}$. Different surrogates are constructed using different initial random shuffles.
4. Calculate the Fourier series of the shuffled time series, denoted $a_0^S, a_1^S, a_2^S, \ldots a_{N/2}^S, b_1^S$, $b_2^S, \ldots b_{N/2}^S$.
5. Construct a temporary Fourier series that has the amplitude of the Fourier series of $\{x\}$ and the phase of the Fourier series of $\{x^S\}$, denoted as follows $a_0^T, a_1^T, a_2^T, \ldots a_{N/2}^T, b_1^T, b_2^T, \ldots b_{N/2}^T$.
$$a_j^T = [(a_j^D)^2 + (b_j^D)^2]^{1/2} \left\{ \frac{a_j^S}{[(a_j^S)^2 + (b_j^S)^2]^{1/2}} \right\}$$
$$b_j^T = [(a_j^D)^2 + (b_j^D)^2]^{1/2} \left\{ \frac{b_j^S}{[(a_j^S)^2 + (b_j^S)^2]^{1/2}} \right\}$$
6. Construct temporary time series $\{x^T\}$ by calculating the inverse transform of the Fourier series, denoted $a_0^T, a_1^T, a_2^T, \ldots a_{N/2}^T, b_1^T, b_2^T, \ldots b_{N/2}^T$.
7. Construct the first iterate of the surrogate, denoted $\{x^U\}$, by re-ordering $\{x\}$ the original data so that it has the same rank structure as

the temporary time series $\{x^T\}$. The previously constructed vector $\{x^{\mathrm{Sort}}\}$ is used in this step.

The next iterate is produced by setting $\{x^S\} = \{x^U\}$ and returning to Step 4. The process is continued until a convergence criterion is satisfied. In our implementation of the procedure, it was found that the convergence criterion proposed by Schreiber and Schmitz was satisfied in all cases when 30 iterations were performed.