International Journal of Bifurcation and Chaos, Vol. 19, No. 1 (2009) 403–408

EFFECT OF JUMP DISCONTINUITY FOR PHASE-RANDOMIZED SURROGATE DATA TESTING*

ENNO MAMMEN
Department of Economics, University of Mannheim,
L 7, 3-5, 68131 Mannheim, Germany
emammen@rumms.uni-mannheim.de

SWAGATA NANDI
Indian Statistical Institute, 7, S.J.S. Sansanwal Marg
New Delhi 110016, India
nandi@isid.ac.in

THOMAS MAIWALD† and JENS TIMMER‡
FDM, Freiburg Center for Data Analysis and Modelling,
University of Freiburg, Eckerstr. 1,
79104 Freiburg, Germany
†*maiwald@fdm.uni-freiburg.de*
‡*jeti@fdm.uni-freiburg.de*

Received November 29, 2007; Revised April 14, 2008

In this paper we discuss two modifications of the surrogate data method based on phase randomization, see [Theiler *et al.*, 1992]. By construction, phase randomized surrogates are circular stationary. In this respect they differ from the original time series. This can cause level inaccuracies of surrogate data tests. We will illustrate this. These inaccuracies are caused by end to end mismatches of the original time series. In this paper we will discuss two approaches to remedy this problem: resampling from subsequences without end to end mismatches and data tapering. Both methods can be understood as attempts to make non-circular data approximately circular. We will show that the first method works quite well for a large range of applications whereas data tapering leads only to improvements in some examples but can be very unstable otherwise.

Keywords: Surrogate data; resampling; time series analysis.

1. Introduction

A classical algorithm to generate surrogate data is the method of phase randomization, see [Theiler *et al.*, 1992]. Phase randomized surrogates have the same linear properties as the observed data: the mean and the periodogram (at the Fourier frequencies) of the observed times series are preserved. This implies that the population auto-covariance function of the surrogate data vector is exactly equal to the circular auto-covariance function of the data, see [Chan, 1997]. But only the circular autocorrelation function is preserved, not the usual autocorrelation function. Moreover, for the hypothesis of circular stationary Gaussian processes, tests based on

*This study is supported by DFG project MA 1026/7-2 in the framework of priority program SPP-1114.

phase-randomized surrogates attain the exact level, see [Chan, 1997] for univariate time series, [Mammen & Nandi, 2008] for multivariate processes and [Mammen & Nandi, 2004] for consequences of this property on power properties. This is not true for the more general hypothesis of stationary Gaussian processes. Then tests based on Fourier-based surrogates may not work well. From a practical point of view, this is an essential drawback. The hypothesis of possibly noncircular processes is much more important. One reason why surrogate data fails for noncircular processes are end to end mismatches. Typically, the last point of the observed time series deviates from the starting point. These edge effects lead to periodicity artifacts. We will illustrate this below. In this paper, we will discuss two approaches that have been proposed in the literature to remedy this problem: choosing subsequences without end to end mismatches and data tapering, see [Theiler *et al.*, 1992]. Both methods are attempts to make noncircular data approximately circular.

In the first approach, one ignores some values of the time series at the beginning and at the end. This is done such that the first and the last data points are approximately equal. In the second approach one first applies tapering of the data before generating the surrogates (windowed Fourier-transform surrogates). For a detailed description of the methods, see the next section. This paper presents a thorough analysis of both approaches. We will check level accuracy of surrogate data tests for a class of noncircular stationary Gaussian processes and for a set of test statistics. We will show that the first method based on subsequences works quite well for a large range of applications whereas data tapering leads only to improvements in some examples but can be very unstable otherwise. The paper is organized as follows. In Sec. 2, we discuss the proposed methods. Numerical experiments are described and discussed in Sec. 3 and finally we conclude the paper in Sec. 4.

2. Correcting End to End Mismatches: Two Methods

We now give a detailed description of the two procedures to correct the effect of end to end mismatches. The method based on taking subsequences will be introduced in Sec. 2.1. Section 2.2 describes the method based on data tapering. We will introduce automatic implementations of the two procedures. This will allow us to study in the next section if these corrections improve the level accuracy of the modified surrogate data tests.

2.1. *Taking subsequences of the time series*

Suppose that a time series $X_1, \ldots, X_N$ is observed. For quantifying the mismatch between the start and the end of the time series we use a very simple measure. We consider the absolute difference of the two end points $|X_1 - X_N|$. In [Schreiber & Schmitz, 2000] it has also been proposed to measure the mismatch between the end points in their first derivatives. We consider the subseries procedure by the following steps:

1. Fix two integers K_1 and K_2 that are much smaller than the sample size N
2. Calculate the measure of mismatch $\tau(i,j)$ as

$$\tau(i,j) = |X_i - X_j|$$

   for all $K_1 \times K_2$ pairs (i,j) with $i = 1, \ldots, K_1$ and $j = N - K_2 + 1, \ldots, N$ and find that pair $(\bar{i}, \bar{j})$, for which $\tau(i,j)$ is minimized.
3. Take the subseries $\{X'_1, \ldots, X'_{\bar{j}-\bar{i}}\} = \{X_{\bar{i}+1}, \ldots, X_{\bar{j}}\}$ of size $\bar{j} - \bar{i}$ and base the whole data analysis on this subseries. In particular, use the subseries to generate the surrogate data.

The procedure drops at K_1 most data points in the beginning and then drops at $K_2 - 1$ most points at the end of the series. The corrected series will be at least of length $N - K_1 - K_2 + 1$.

2.2. *Data tapering*

Data tapering is the second method to be corrected for jump discontinuities between end points of the observed series. The tapering is applied before the surrogate resamples (FT-based surrogates) are generated. For the generation of the surrogate data, the sample mean of the original series and the periodogram values of the tapered data are used. In the tapering, the observed series is multiplied by a window function, say $u(t)$ which gradually decreases to zero at end points. We have used $u(t) = 1/2[1 + \cos(2\pi(t - (N/2))/(N+1))]$ as data taper, see [Brillinger, 1981, p. 55] for an extensive list of other data tapers.

3. Numerical Experiments

In this section, we describe our simulation experiment. The aim is to see how the end to end mismatch corrections affect the level of accuracy of the modified surrogate data tests. We have considered the following data generating processes:

- $M_1 : X_t = 0.9X_{t-1} + e_t$
- $M_2 : X_t = 0.99X_{t-1} + e_t$
- $M_3 : X_t = 1.4X_{t-1} - 0.48X_{t-2} + e_t$
- $M_4 : X_t = 1.7X_{t-1} - 0.72X_{t-2} + e_t$
- $M_5 : X_t = 1.8X_{t-1} - 0.81X_{t-2} + e_t$
- $M_6 : X_t = 2.4X_{t-1} - 1.91X_{t-1} + 0.504X_{t-3} + e_t$
- $M_7 : X_t = 2.6X_{t-1} - 2.25X_{t-1} + 0.648X_{t-3} + e_t$.

All these models are stationary Gaussian autoregressive (AR) models. They are of different orders; M_1 and M_2 are of order 1, $M_3 - M_5$ are of order 2 and M_6 and M_7 are of 3. All processes are stationary, but they are all near to unit root models.

For these processes, we have considered the following test statistics in our simulation study.

$$T_1 = \frac{1}{N}\sum_{t=1}^{N-1}(X_t X_{t+1}^2 - X_t^2 X_{t+1}),$$

$$S_1 = \frac{\#\{X_t > X_{t+1}\}}{N},$$

$$S_2 = N - 1 - S_1, \quad T_2 = S_1, \quad T_3 = \frac{|S_2 - S_1|}{S_1 + S_2}$$

$$T_4 = \max_\tau Q(\tau),$$

$$Q(\tau) = \frac{\sum_{t=\tau+1}^{N}(X_{t-\tau} - X_t)^3}{\left[\sum_{t=\tau+1}^{N}(X_{t-\tau} - X_t)^2\right]^{3/2}},$$

$$T_5 = \max\left\{\frac{\#\{|X_{t+1} - \overline{X}| > |X_t - \overline{X}|\}}{\#\{|X_{t+1} - \overline{X}| < |X_t - \overline{X}|\}}, \frac{\#\{|X_{t+1} - \overline{X}| < |X_t - \overline{X}|\}}{\#\{|X_{t+1} - \overline{X}| > |X_t - \overline{X}|\}}\right\},$$

$$T_6 = C_N(r),$$

$$C_N(r) = \frac{\sum_{i=2}^{N}\sum_{j=1}^{i} I(\|\mathbf{X}_i^\nu - \mathbf{X}_j^\nu\| < r)}{\frac{N(N-1)}{2}}.$$

Test statistics $T_1, \ldots, T_5$ are measures of time asymmetry. The statistic T_5 was proposed in [Maiwald *et al.*, 2008]. Test statistic $T_6 = C_N(r)$ is the correlation sum which is defined as the sample analogue of the correlation integral, see [Theiler *et al.*, 1992] who discusses correlation sums as discrimination statistics. In the definition of $C_N(r)$, I is the indicator function and $\|\mathbf{X}\| = \max_k |X_k|$. The vector $\mathbf{X}_i^\nu = (X_{i-(\nu-1)d}, X_{i-(\nu-2)d}, \ldots, X_i)^T$ belongs to the phase space with embedding dimension ν and delay time d. We have used delay time $d = 2$. The results are reported for embedding dimension $\nu = 4$. We have used different values of r for different models.

In the simulations, we generated data $\mathbf{X_N} = (X_1, \ldots, X_N)$ from the models listed in (1) and we calculated the test statistics $T_j(\mathbf{X_N})$ for $j = 1, \ldots, 6$. We used sample size $N = 512$. For generating surrogates we computed periodogram values (DFT) at Fourier frequencies $\omega_j = (2\pi j/N)$ for $1 \le j \le N/2$. For each simulated $\mathbf{X_N}$, 1000 surrogate resamples $\mathbf{X_N^*}$ were generated. For each of these resamples, we calculated test statistics $T_j(\mathbf{X_N^*})$, $j = 1, \ldots, 6$. We denote the $(1-\alpha)$th quantile of $T_j(\mathbf{X_N^*})$ by $k_{j\alpha}^*$. Then the surrogate data test rejects the hypothesis of a linear stationary Gaussian process, if $T_j(\mathbf{X_N}) > k_{j\alpha}^*$. The aim of our simulations is to check the level of accuracy of this test, i.e. to check if the rejection probability on the hypothesis is approximately equal to the nominal level α_{nom}: $P[T_j(\mathbf{X_N}) > k_{j\alpha}] \approx \alpha_{\text{nom}}$. For this check, the whole procedure was repeated 1000 times. The empirical fraction of $T_j(\mathbf{X_N}) > k_{j\alpha}$ gives an estimate of the level of the test, say $\hat{\alpha}$. The simulation results are given in Tables 1 and 2. We have used the nominal value $\alpha_{\text{nom}} = 0.05$.

Similar simulation experiments were carried out for obtaining the levels of the two mismatch corrected time series. For the first procedure, for each $\mathbf{X_N}$, we applied the procedure described in Sec. 2.1 to obtain the end corrected subseries, say $\mathbf{X'_{\overline{N}}}$. Here $\overline{N} = \overline{j} - \overline{i}$ is the sample size of the subseries. We used $K_1 = K_2 = 40$ for all models. For each $\mathbf{X'_{\overline{N}}}$, 1000 surrogates $\mathbf{X'^*_{\overline{N}}}$ were generated. Then the rejection probability $P[T_j(\mathbf{X'_{\overline{N}}}) > k'_{j\alpha}]$ is estimated similarly as for $T_j(\mathbf{X_N})$. Let us call it $\hat{\alpha}_{ec}$. Here $k'_{j\alpha}$ is the $(1-\alpha)$th quantile of $\mathbf{X'^*_{\overline{N}}}$. Note that in our implementation the test statistic is calculated for $\mathbf{X'_{\overline{N}}}$ and not for $\mathbf{X_N}$. Thus the test statistic is calculated for the same sample size as in the resampling. For small values of K_1 and K_2 one could also consider the use of $T_j(\mathbf{X_N}) > \mathbf{k'_{j\alpha}}$ as test criterion. We have not done this here. The (estimated) levels

Table 1. Level of accuracy of surrogate data tests using surrogates of the complete original series (uncorr.), of end to end corrected subseries (subseq.) and of tapered series (dt.). The level of accuracy is marked by "− − −" if the level (i.e. $\hat{\alpha}$, $\hat{\alpha}_{ec}$ or $\hat{\alpha}_{dt}$, respectively) is 0.0, "ok" if it lies between 0.03 and 0.075, "+" if it lies between 0.075 and 0.125, "++" if it lies between 0.125 and 0.25, and "+++" if it is larger than 0.25. The nominal level is 0.05. The table reports the results for test statistics T_1, T_2 and T_3.

	Test Statistics								
	T_1			T_2			T_3		
Models	uncorr.	subseq.	dt.	uncorr.	subseq.	dt.	uncorr.	subseq.	dt.
M_1	+	ok	+++	ok	ok	ok	ok	ok	ok
M_2	+++	+	+++	+	ok	ok	+	ok	+
M_3	++	ok	+++	ok	ok	ok	ok	ok	ok
M_4	+++	ok	+++	ok	ok	ok	+	ok	+
M_5	+++	ok	+++	+	ok	+	+	ok	+
M_6	+++	ok	+++	+	ok	+	++	ok	+
M_7	+++	+	+++	++	ok	+	++	ok	+

Table 2. Level of accuracy of surrogate data tests. The notation is as in Table 1. Now the results are reported for T_4, T_5 and T_6.

	Test Statistics								
	T_4			T_5			T_6		
Models	uncorr.	subseq.	dt.	uncorr.	subseq.	dt.	uncorr.	subseq.	dt.
M_1	ok	ok	ok	ok	ok	ok	ok	ok	− − −
M_2	++	+	+	ok	ok	ok	+++	+	− − −
M_3	ok	ok	ok	ok	ok	ok	ok	+	− − −
M_4	ok	ok	ok	ok	ok	ok	+++	ok	− − −
M_5	ok	ok	ok	+	ok	ok	+++	+	− − −
M_6	ok	ok	ok	+	ok	ok	+++	ok	− − −
M_7	ok	ok	ok	++	ok	ok	+++	+	− − −

of the second procedure are denoted by $\hat{\alpha}_{dt}$. Here, for each $\mathbf{X_N}$, we first obtained the tapered data, say $\mathbf{X_N^t}$. Using the method given in Sec. 2.2, 1000 surrogates, say $\mathbf{X_N^{t*}}$ for $\mathbf{X_N^t}$ was generated. Then, the $(1-\alpha)$th quantile for each test statistic $T_j(\mathbf{X_N^{t*}})$ was estimated as $k_{j\alpha}^t$. Finally the level $\hat{\alpha}_{dt}$ was estimated as the fraction of $T_j(\mathbf{X_N}) > k_{j\alpha}^t$. Here the test is based on $T_j(\mathbf{X_N})$, not on $T_j(\mathbf{X_N^t})$. The simulation results for $\hat{\alpha}_{ec}$ and $\hat{\alpha}_{dt}$ are summarized in Tables 1 and 2.

Before we discuss the results of the simulation study let us briefly visualize how the jump discontinuity affects the surrogate resamples. A realization of size $N = 512$ of the model M_7 has been generated and plotted in Fig. 1 (left plot). There is a big jump between the first and the last points of this series. A phased-randomized surrogate data with the same sample mean and the same periodogram values has been plotted on the right plot in Fig. 1. We observe that there are small artificial fluctuations visible in the surrogate series, that was not present in the original series. Now we apply the mismatch correction procedure given in Sec. 2.1 with $K_1 = 50$ and $K_2 = 50$. The procedure chooses an end to end corrected subseries of length 445. This sequence is plotted in Fig. 2 (left plot) and a surrogate resample of it is given in Fig. 2 (right plot). The small fluctuations which were observed for the surrogates in Fig. 1 are not present now. Figure 3 illustrates the method of data tapering. The tapered data are plotted in Fig. 3 (left plot) and one of its surrogate data are displayed in Fig. 3 (right plot).

We now compare the outcomes of our simulations, reported in Tables 1 and 2. The simulation shows that the estimated level of the uncorrected surrogate data test does not keep its level. The surrogate data test rejects too often in many cases. For the test statistics T_1 and T_6 the rejection probability is larger than 0.25 (+++) in five out of seven models. This means that the surrogate data test is breaking down here: the test outcome is the same as for a random number that is produced without looking at the data. The breakdown of the test in these cases can be easily explained by Fig. 1. The small bumps appearing in the right plots cause erroneous calculations of the two statistics that are both based on the quantitative evaluation of neighbored values of the time series. The other test statistics use only overall measures (T_4) or qualitative measures based on the qualitative comparison of neighbored values (T_2, T_3 and T_5). For the test statistics T_2–T_5, the uncorrected surrogate data test is unreliable in some cases for models M_2, M_6 and M_7. If we qualify the performance "$- - -$", "$++$" and "$+++$" as unreliable we get that the uncorrected surrogate data test is unreliable in 16 out of 42 cases. (There were no test outcomes with rejection probability between 0 and 0.03.) This is a quite unstable behavior. The performance is only slightly improved if data tapers are used before the surrogates are generated. This method is unreliable in 14 out of 42 cases: it breaks down in all models for test statistics T_1 and T_6. Thus for these two test statistics the performance is not improved although the little bumps have been removed for the surrogates (compare the right plots in Figs. 1 and 3). For T_6 the data taper test always rejects, in all 1000 runs of the simulations. Clearly, formally this results in a valid test that keeps its level. But it clearly indicates that the data taper surrogates are not able to capture the stochastic structure of the data and one cannot expect any reasonable power for this test with neighbored alternatives. For the other test statistics application of data tapers lead to an improvement of level accuracy. The data taper test is clearly outperformed by the end to end correction based on taking a subseries. The latter test has a reliable performance in all 42 cases. For the test statistics T_4 and T_5 the improvements are comparable to using data tapers but it clearly outperforms the data taper tests for all other test statistics. In particular, the performance is reliable for the test statistics T_1 and T_6 where the uncorrected test and the data taper test break down in most or all models.

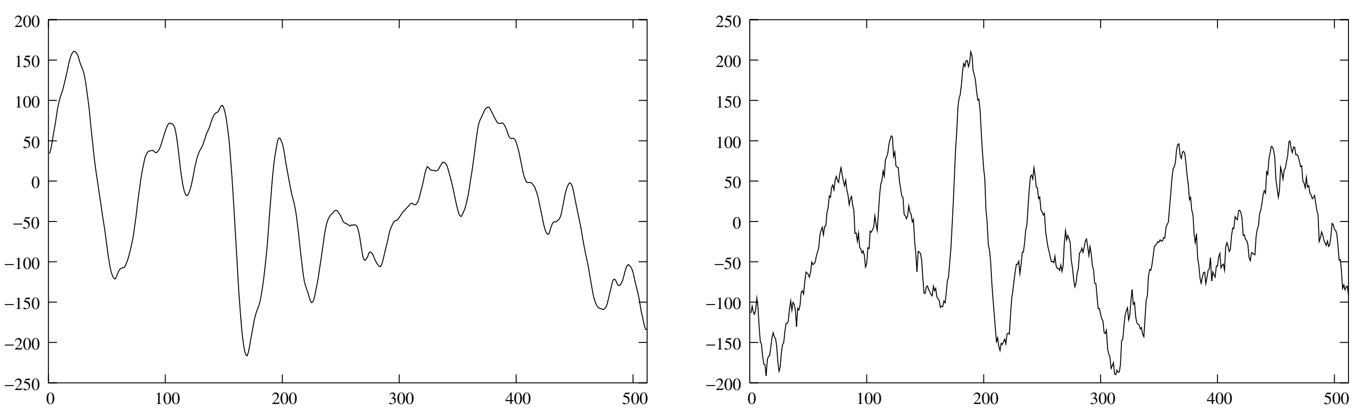


Fig. 1. A sample of size 512 of the process M_7 (left plot) with one Fourier-based surrogate (right plot).

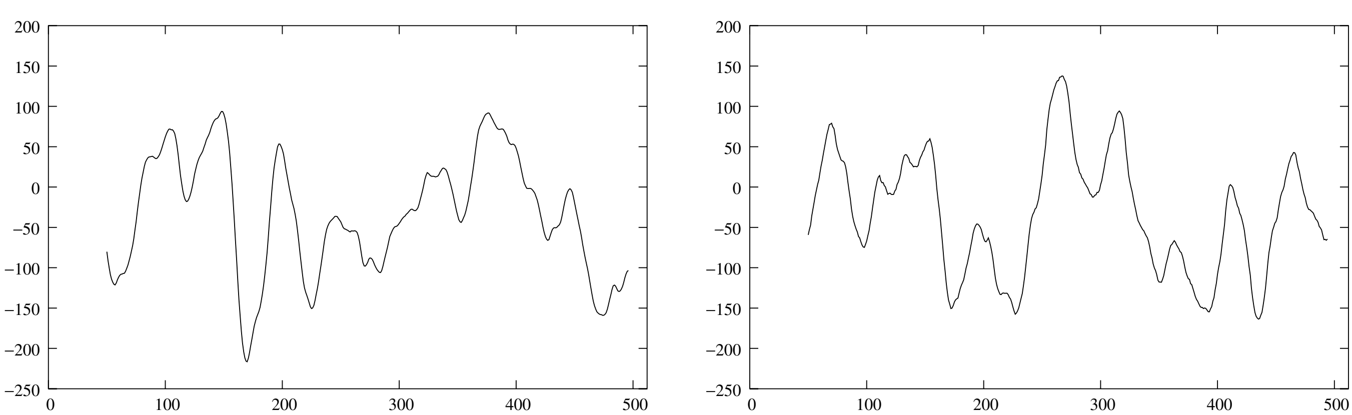
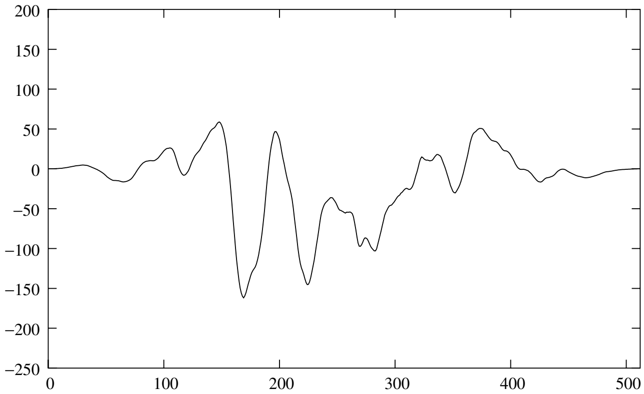


Fig. 2. A subseries of the series given in Fig. 1 without end to end mismatch (left plot) with one Fourier-based surrogate of the subseries (right plot).



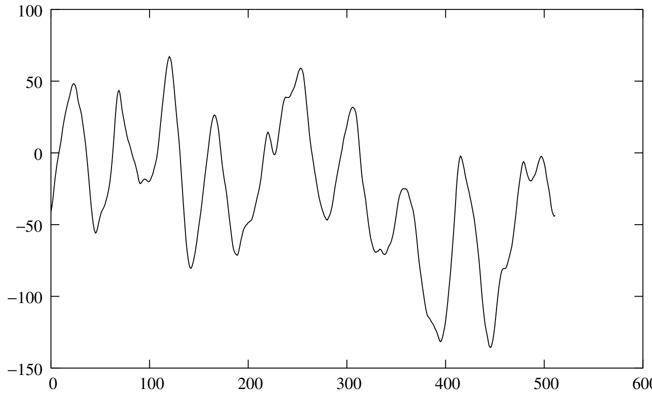


Fig. 3. Tapered data of the series given in Fig. 1 (left plot) with one Fourier-based surrogate of this series (right plot).

Thus, our simulations give a very clear picture. In the simulated cases data tapering only leads to moderate improvements. In particular, it helped mostly in cases where the unmodified surrogate data test was already quite reliable. In cases where the unmodified test broke down data tapering was not of any use. End to end corrections based on taking a subseries turned out to be very successful. It nearly always leads to improvements and it also produced a reliable test when the unmodified surrogate data tests (and its data taper modification) broke down.

4. Summary and Conclusions

In our numerical experiments, we have estimated the level of different tests for a class of linear stationary Gaussian processes. The critical values of the tests were fitted by using surrogate resamples, surrogate resamples for subseries without end to end mismatch or surrogate resamples generated for tapered data, respectively. We observed partially large level inaccuracies of the unmodified surrogate data test. The level inaccuracies were partially corrected by the application of data tapering. But data tapering does not always lead to improvements. For some test statistics it even produced totally misleading values. Surrogate resampling for subseries was quite reliable. It has not always produced totally accurate levels, but it has nearly always lead to improvements of the level accuracy and the improvements were partially drastic. The paper has shown that it is highly recommendable to repair mismatches between end values of time series before generating surrogate data. In this paper we only used the absolute difference of the end observations as mismatch criterion. This worked quite well for the considered test statistics. We conjecture that for more involved data analysis more complicated mismatch criterions should be used, e.g. for checking embeddings in higher dimensional spaces the mismatch of the tuple of the last values and the tuple of the first values could be checked. But this would only make sense for very long time series. Then it may be too complex to check the performance in a larger simulation study where a large number of surrogates must be generated for a large number of original samples.

References

Brillinger, D. R. [1981] *Time Series: Data Analysis and Theory* (McGraw-Hill Inc.).

Chan, K. S. [1997] "On the validity of the method of surrogate data," *Fields Instit. Commun.* **11**, 77–97.

Maiwald, T., Mammen, E., Nandi, S. & Timmer, J. [2008] "Surrogate data — a qualitative and quantitative analysis," *Mathematical Methods in Signal Processing and Digital Image Analysis*, eds. Dahlhaus, R., Kurths, J., Maass, P. & Timmer, J. (Springer, NY), pp. 41–74.

Mammen, E. & Nandi, S. [2004] "Change of the nature of a test when surrogate data are applied," *Phys. Rev. E* **70**, 16121–16132.

Mammen, E. & Nandi, S. [2008] "Some theoretical properties of phase randomized multivariate surrogates," *Statistics* **42**, 195–205.

Schreiber, T. & Schmitz, A. [2000] "Surrogate time series," *Physica D* **142**, 346–382.

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