

## Surrogate Data – A Qualitative and Quantitative Analysis

Thomas Maiwald<sup>1</sup>, Enno Mammen<sup>2</sup>, Swagata Nandi<sup>3</sup>, and Jens Timmer<sup>1</sup>

<sup>1</sup> FDM, Freiburg Center for Data Analysis and Modeling, University of Freiburg, Eckerstr. 1, 79104 Freiburg, Germany

maiwald@fdm.uni-freiburg.de, jeti@fdm.uni-freiburg.de

<sup>2</sup> Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany

emammen@rumms.uni-mannheim.de

<sup>3</sup> Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 7, S.J.S. Sansanwal Marg, New Delhi 110016, India

nandi@isid.ac.in

### 2.1 Motivation

The surrogates approach was suggested as a means to distinguish linear from nonlinear stochastic or deterministic processes. The numerical implementation is straightforward, but the statistical interpretation depends strongly on the stochastic process under consideration and the used test statistic. In the first part, we present quantitative investigations of level accuracy under the null hypothesis, power analysis for several violations, properties of phase randomization, and examine the assumption of uniformly distributed phases. In the second part we focus on level accuracy and power characteristics of Amplitude Adjusted Fourier–Transformed (AAFT) and improved AAFT (IAAFT) algorithms. In our study AAFT outperforms IAAFT. The latter method has a similar performance in many setups but it is not stable in general. We will see some examples where it breaks down.

### 2.2 Introduction

In a statistical test, for a given data set a hypothesis is formulated, whose validity has to be examined. This *null hypothesis* cannot be verified or falsified with 100% accuracy. Instead, only an upper bound for the probability is given, that the null hypothesis is rejected although it holds true. For this a real valued *test statistic*  $T$  is derived from the data, with known distributions under the possible specifications of the null hypothesis. If the null hypothesis contains only one specification (*simple hypothesis*) and the value of  $T$  is for example

larger than the  $1 - \alpha$  quantile of the null distribution, the null hypothesis will be rejected and a false rejection will occur only with probability  $\alpha$ . Often, the null distribution of  $T$  is not simple and the distribution of the underlying specification has to be estimated and/or approximated. One possibility is to use *surrogate* methods. They have been suggested for the null hypothesis of a linear Gaussian process transformed by an invertible nonlinear function, see [20].

This chapter examines the different approaches to generate surrogate data and their properties. The basic idea of all surrogate methods is to randomize the Fourier phases of the underlying process. In Sect. 2.3 we illustrate how the nature of a process changes if the phases are randomized. For this purpose we show plots where the amount of randomization is continuously increased. Section 2.4 summarizes all suggested surrogate approaches and illustrates their qualitative behavior. Section 2.5 presents a simulation study where the level accuracy and power performance of surrogate data tests is compared for different processes, test statistics and surrogate methods. Section 2.6 takes a closer look on two methods for generating surrogate data. It compares the AAFT and the IAAFT approach. The chapter ends with short conclusions in Sect. 2.7.

## 2.3 Phase Randomization – Effects and Assumptions

The key procedure of all surrogate methods is to randomize the Fourier phases. It is argued that linear Gaussian processes do not possess asymptotically any information in the Fourier phases, since they are comprehensively determined by their mean and autocovariance function, which corresponds one-to-one via the Fourier transformation to the power spectrum. Hence, realizations of a linear Gaussian process should differ essentially only in their Fourier phases which is utilized by the surrogates approach: New realization of a linear Gaussian process based on one realized time series can be obtained by drawing new values for the Fourier phases. Since no information is expected in the Fourier phases, the underlying distribution for the new phase values is the uniform distribution on  $[0, 2\pi]$ .

This section investigates the assumption of uniformly distributed Fourier phases and takes a qualitative look on the effects of phase randomization on linear and nonlinear time series.

### 2.3.1 Effects of Phase Randomization

Let  $\mathbf{x} = (x_1, \dots, x_N)$  be a given time series of length  $N$  with mean 0. Its Fourier transformation is

$$f(\omega) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^N e^{-i\omega t} x_t, \quad -\pi \leq \omega \leq \pi. \quad (2.1)$$

If the transformation is calculated for discrete frequencies  $\omega_j = 2\pi/N$ ,  $j = 1, 2, \dots, N$ , the original time series can be regained by the back transformation:

$$x_t = \sqrt{\frac{2\pi}{N}} \sum_{j=1}^N e^{i\omega_j t} f(\omega_j), \quad t = 1, 2, \dots, N. \quad (2.2)$$

The FT-surrogates method constructs a new time series  $y_t$  with the same periodogram and otherwise statistically independent from  $x_t$  [19]. The Fourier amplitudes  $|f(\omega_j)|$  are fixed and the Fourier phases  $\varphi(\omega_j) = \arg(f(\omega_j))$  are replaced by uniform distributed random numbers  $\varphi_{\text{rand}}(\omega_j) \in [0, 2\pi]$ . The new realization is given by

$$y_t = \sqrt{\frac{2\pi}{N}} \sum_{j=1}^N e^{i\omega_j t} |f(\omega_j)| e^{i\varphi_{\text{rand}}(\omega_j)}. \quad (2.3)$$

To track the effect of phase randomization, we increase randomization strength continuously from 0% to 100%. For this random numbers

$$\varphi_{\text{rand}}(\omega_j) = \varphi(\omega_j) + a u(\omega_j)$$

with  $u(\omega_j) \in \mathcal{U}[-\pi, \pi]$  are drawn. Parameter  $a \in [0, 1]$  changes the phase randomization strength. Figure 2.1(a) shows, that three test statistics of Sect. 2.5, time invariance, prediction error and kurtosis do not change for the autoregressive process with increasing phase randomization. For the nonlinear stochastic Van-der-Pol oscillator already small changes of the Fourier phases lead to significant different values of the test statistics (b). Figure 2.1(c–f) exemplifies the influence of phase randomization on time series directly. The autoregressive process as a linear process does not change visually with increasing phase randomization (c). The nonlinear Lorenz system loses already for 20% phase randomization its characteristic oscillations (d). With 100% phase randomization, even the ear switch is not visible any more. The time series equals a realization of an autoregressive process with small coherence length. The deterministic Rössler system is shown in total length (e). It possesses a sharp harmonic component leading to higher harmonics in the power spectrum. If the phase relation of these harmonics is lost due to phase randomization, a beat occurs resulting in strong correlations of the time series up to the whole time series length. This holds even stronger for the zig-zag curve (f). The time series of a stationary, nonlinear, deterministic process becomes a time series which is indistinguishable from a realization of a non stationary, cyclic, linear process.

The correlation dimension in Fig. 2.2 differs for the nonlinear processes logistic map, Lorenz system and stochastic Van-der-Pol oscillator. Whereas for the logistic map already a 1% phase randomization leads to an unbounded correlation dimension on small scales, this happens for the Lorenz systems not until a randomization of 10%. The stochastic Van-der-Pol does not show any changes with increasing phase randomization – as a stochastic process it has an unbounded correlation dimension even for the original process.

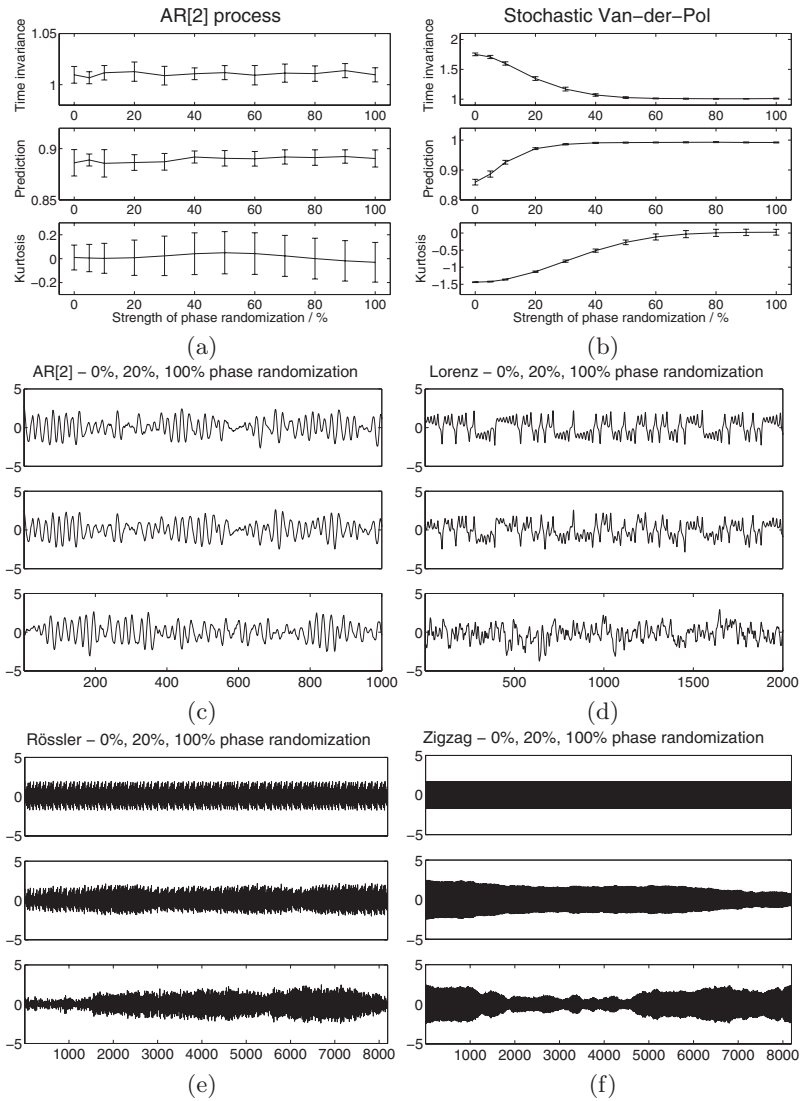
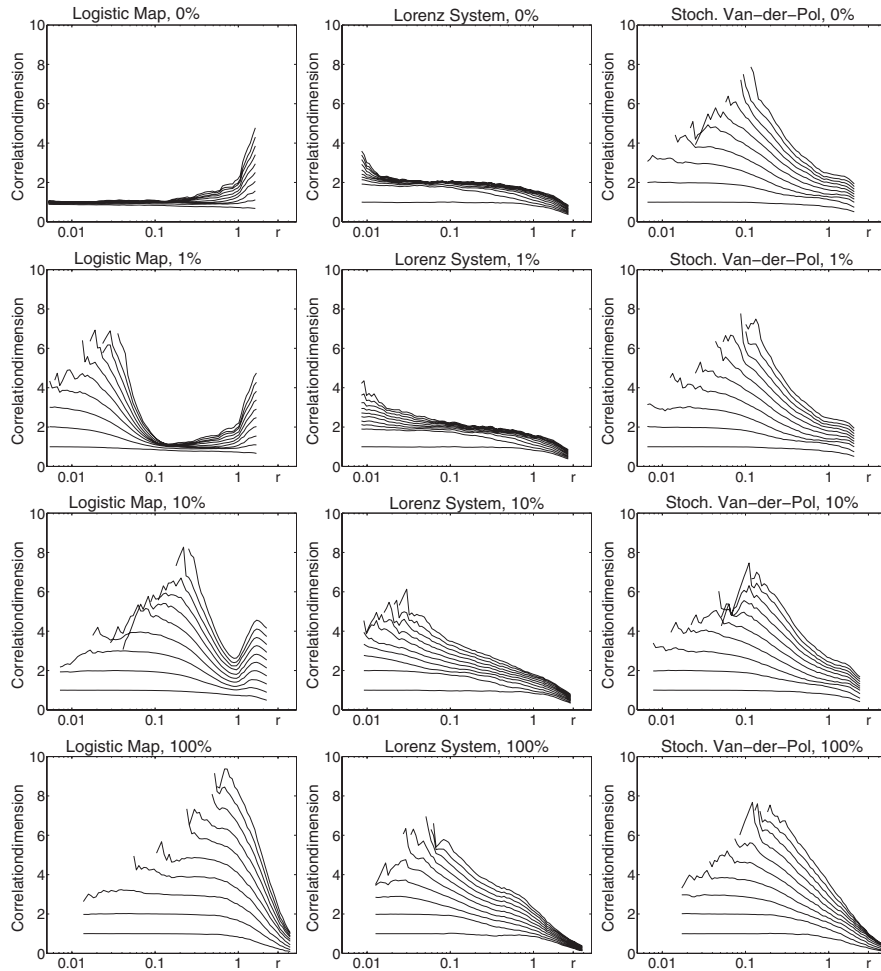


Fig. 2.1 Influence of increasing phase randomization

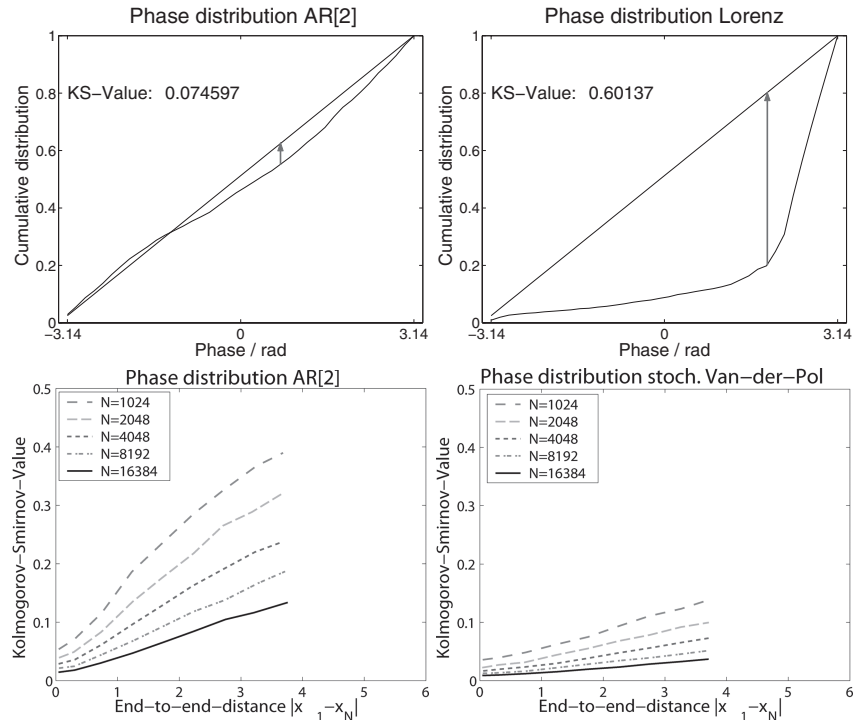
### 2.3.2 Distribution of Fourier Phases

Figure 2.3 shows the cumulative Fourier phase distribution for realizations of an AR[2] process and the Lorenz system. For the AR[2], the cumulative distribution is very close to the straight line corresponding to perfect uniform distribution. For the Lorenz system, most phases are larger than  $2\pi$  – the distribution is unbalanced. The deviation from a uniform distribution can



**Fig. 2.2** Influence on the correlation dimension for increasing phase randomization

be quantified by the Kolmogorov-Smirnov-statistic measuring here the largest distance between the cumulative distribution and the straight line. The AR[2] leads to a KS-value of 0.075, whereas the Lorenz system has a nearly ten times larger value of 0.6. To examine whether this result is representative, a simulation with 1000 AR[2] of length between 1024 and 16384 has been done and the KS-value calculated. Figure 2.3 exhibits a nearly linear relationship between the mean KS-value and the end-to-end-distance  $|x_N - x_1|$ . Since a similar simulation for the nonlinear stochastic Van-der-Pol oscillator shows the same effect, the phase distribution of linear and nonlinear systems depends strongly



**Fig. 2.3** Distribution of the Fourier phases. The Kolmogorov-Smirnov-statistic measures the deviation of the Fourier phase distribution from a uniform distribution. Interestingly, a nearly linear dependency on the end-to-end-mismatch is observed

on the end-to-end-distance and deviations from a uniform distributions are results of the finite size effect, not the nonlinearity. Only noise-free nonlinear systems like the Lorenz one always show a strong deviation from the uniform distribution.

## 2.4 Surrogates Methods

There exist two classes of surrogate data approaches. In surrogate data testing one reproduces all the linear properties of the process by preserving the second order properties of the observed data. This can be done by preserving the observed sample mean and sample auto-covariances (or by preserving the observed sample mean and periodogram values at Fourier frequencies  $\omega_j = 2\pi j/n$ ). The approach can be implemented by first Fourier transforming the data set, then randomizing the phases and finally inverting the transformed data. Then resamples will have the same linear properties as the data set, see Sect. 2.3 II. The term “surrogate data” was first introduced by [19]

and the method became popular after that. But the basic idea and related approaches were discussed in a number of earlier publications, see [6, 8].

Several other ways exist to generate surrogate data with asymptotically the same autocovariance function as a given time series. Essentially they differ in conserving the original periodogram by construction or generating a new one derived from the estimated power spectrum.

Fourier transformed surrogate data method generates resamples for the null hypothesis that the process is a stationary Gaussian linear process. This method is not suitable for the hypothesis that the data are not Gaussian. This can be encountered in several applications. In practice, the null hypothesis of linear Gaussian processes is rather restrictive as only very few data pass the test that they are normally distributed. Fourier-transformed surrogates are, by construction, asymptotically jointly normally distributed, thus surrogates will have a different distribution than the observed data, when the data deviate from normality. For such cases the more general null hypothesis has been proposed that the process is a linear stationary process transformed by a static (invertible) nonlinear function  $h(\cdot)$ . More explicitly, the observed process  $\{x_t\}$  is generated by a transformation:

$$x_t = h(z_t),$$

where  $z_t$  is a Gaussian stationary process. For this extended null hypothesis methods have been proposed that transform the original data to a Gaussian amplitude distribution before the surrogate method is applied. A back transformation to the original amplitudes realizes the nonlinear measurement function.

Classical Monte-Carlo-simulation is the counterpart of the parameter-free surrogate approach. Here, the problem arises to select the right model, e.g., the right order of an autoregressive process. If the right model is known, the Monte-Carlo-approach should be the used.

The basic method to generate Fourier transformation surrogates (FT) has been described in Sect. 2.3. For an overview on resampling methods that have been used in the statistical literature on time series analysis see also [12].

#### 2.4.1 Amplitude Adjusted FT-Surrogates (AFT)

If the data derive under an extended null hypothesis from a linear, Gaussian process measured via an invertible nonlinear function  $h$ , the Gaussian amplitude distribution is lost in general. Since only Gaussian distributed linear processes are uniquely given by their autocovariance function and mean value, the measurement function  $h$  has to be estimated in order to calculate surrogates for  $\hat{h}^{-1}(\mathbf{x})$ . The surrogate data will be *measured* with  $\hat{h}$  to be comparable to the original data. In detail:

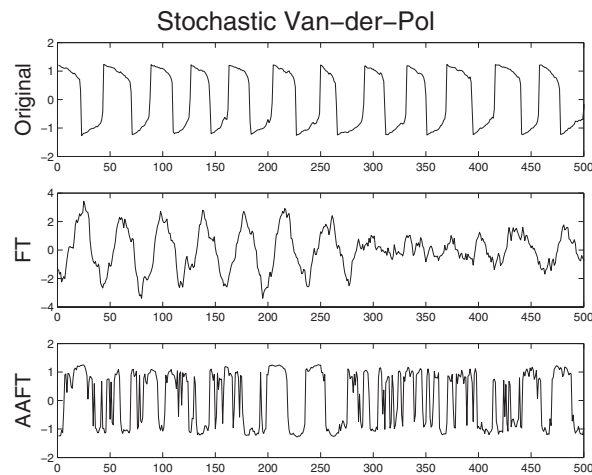
1. Ranking of the original data.
2. Generation of ranked, Gaussian distributed random numbers

3. The  $k$ -th value of the sorted original series is replaced with the  $k$ -th value of the sorted Gaussian data and the order of the original data is re-established.
4. The data are now Gaussian distributed and FT-Surrogates can be calculated.
5. The  $k$ -th largest value of the FT-Surrogates is replaced with the  $k$ -th largest value of the original time series. Note, that the original amplitude distribution is exactly maintained.

Asymptotically for  $N \rightarrow \infty$  the first three steps of the transformation are equivalent to the application of the true inverse function  $h^{-1}$  itself. The procedure was suggested as *Amplitude Adjusted Fourier Transform–Surrogates* [19] and its effect is illustrated in Fig. 2.4. Test statistics like the skewness or kurtosis based on the amplitude distribution have by construction exactly the same value for the original and the surrogate time-series. A more detailed description of the AAFT algorithm can be found in Sect. 2.5 where also the convergence of the fitted transformations is discussed.

For finite data sets a flattened power spectrum is observed for AAFT-Surrogates compared to original data – the spectrum “whitens”. The reason is, that the estimator of the inverse function  $\hat{h}^{-1}(\cdot)$  does not match exactly  $h^{-1}(\cdot)$  for finite  $N$  [17]. The differences  $\delta(x_t) = \hat{h}^{-1}(x_t) - h^{-1}(x_t)$  are essentially independent of  $t$  and possess as an uncorrelated process a white spectrum. Therefore, application of the estimated inverse function adds uncorrelated random numbers to the original data,

$$\hat{h}^{-1}(x_t) = h^{-1}(x_t) + \delta(x_t).$$



**Fig. 2.4** FT- and AAFT-Surrogates. The bimodal distribution of the stochastic Van-der-Pol oscillator is not kept by the FT-Surrogates method, but by the AAFT-method. Both methods yield time inverse invariant time series



### 2.4.2 Iterated Amplitude Adjusted FT–Surrogates (IAAFT)

To reduce the whitening effect of AAFT–surrogates, an iterative approach has been suggested, which asymptotically yields the same periodogram and amplitude distribution as the original data [16]. First, AAFT–Surrogates are generated. Then, in the Fourier domain, the periodogram values are replaced by the periodogram of the original process, the phases are kept. In time domain, the amplitudes are adjusted to the original process. These two steps are iterated until periodogram and amplitude distributions of original and generated time series are equal except for a given tolerance. A description of the IAAFT algorithm can be found in Sect. 2.5 where this approach is compared with the AAFT algorithm.

### 2.4.3 Digitally Filtered Shuffled Surrogates (DFS)

Some authors criticize that the FT-, AAFT- and IAAFT-surrogates methods conserve not only mean and autocovariance function of the underlying process, but also properties of the given realization itself, since only Fourier phases and not the Fourier amplitudes are changed. In this way, the new data possess less variability than new realizations of the original process [5]. Alternatively *Digitally Filtered Shuffled Surrogates* have been suggested.

The DFS–Surrogates approach is based on the following steps [5]:

1. Estimation of the power spectrum by averaging periodograms of several data segments.
2. Estimation of the autocovariance function by an inverse Fourier transformation of the estimated spectrum. This yields the “response”-function.
3. Convolution of a random permutation of the original time series with the response-function. This corresponds to a multiplication in the Fourier domain, where the transformed of the randomized original time series is multiplied to the estimated spectrum.
4. Adapting the amplitude distribution to the original one.

Similar to the AAFT–Surrogates, the amplitude distribution is adjusted in the last step leading to a whitened spectrum. Again, an iterative procedure, the *iDFS-method*, can be applied to reduce this effect.

(Fourier based) surrogate data tests, AAFT, IAAFT, DFS and iDFS belong to the class of constrained realization approaches. In constrained realization approaches one avoids fitting of model parameters and one does not assume any model equation. Instead of fitting a model (e.g., a finite order AR process), one generates resamples that have a certain set of properties in common with the observed data set. An alternative is based on fitting the model of the data and generating data from the fitted model.

### 2.4.4 New Periodogram from Estimated Spectrum (NPS)

We suggest a new method to generate surrogate data, which is based on the statistical properties of the periodogram

$$I(\omega) \sim \frac{1}{2}S(\omega)\chi_2^2,$$

i.e., the periodogram is for given frequency  $\omega$  distributed like a  $\chi^2$ -distribution with two degrees of freedom around the power spectrum  $S(\omega)$ . New realizations of the process can be obtained by

1. estimation of the power spectrum, e.g., via a smoothed periodogram,
2. generation of a periodogram realization by multiplication of  $\chi_2^2$ -distributed random numbers to the spectrum,
3. drawing new phases as uniform random numbers on  $[0, 2\pi]$ ,
4. calculation of an inverse Fourier transformation.

This approach differs slightly from frequency domain bootstrap that does not apply Step 3, see [4]. Again, an iterative approach is imaginable for the extended null hypothesis of a linear, Gaussian process with an invertible, nonlinear measurement function.

#### 2.4.5 Fixed Fourier Phases (FPH)

In a sense *orthogonal* to the FT-surrogates, one could keep the Fourier phases and draw only a new periodograms, like for the NPS-surrogates. This is actually no surrogates approach, but it illustrates the influence of Fourier phases and amplitudes.

#### 2.4.6 Classical Monte-Carlo-Simulation

If the model of the underlying stochastic process under the null hypothesis is known, new realizations of the process can be generated after estimating all parameters  $\theta$  by means of the given data set. The estimation minimizes the square error

$$\sum_{i=1}^N (x_i - y_i(\theta))^2$$

for the given realization  $\mathbf{x} = (x_1, \dots, x_N)$  and a parameterized time series  $\mathbf{y}(\theta) = (y_1(\theta), \dots, y_N(\theta))$ . The difficulty is to select the right model. Linear processes can be formulated as: A parameter  $\theta = (p, q, a_1, \dots, a_p, b_1, \dots, b_q)$  exist, for which

$$X_t = \sum_{i=1}^p a_i X_{t-i} + \sum_{i=0}^q b_i \varepsilon_{t-i}, \quad \varepsilon_t \sim \mathcal{N}(0, 1). \quad (2.4)$$

The model order  $p, q$  has to be determined. Since finite, invertible ARMA[ $p, q$ ] processes can be written as infinite ARMA[ $\infty, 0$ ] or ARMA[ $0, \infty$ ] processes, selection of the wrong model class can lead to infinite parameters which have to be determined. If the right model class is known, calculation of the partial

ACF or ACF for AR or MA processes respectively, can be used to determine the process order.

In this study we fit an AR[80] process to the original time series and generate AR-Fit-surrogates by

$$y_t = \sum_{i=1}^p a_i y_{t-i} + \varepsilon_t \quad (2.5)$$

with random start values and after the transient behavior is over.

## 2.5 Statistical Properties of the Surrogate Methods – A Comparative Simulation Study

This section contains simulation results on surrogates based hypothesis tests. The simulations have been carried out for a representative class of processes under the null hypothesis and under the alternative and for a variety of test statistics.

### 2.5.1 Processes Under the Null Hypothesis

Linear Gaussian processes can be written as [2, 14]

$$X_t = \sum_{i=1}^p a_i X_{t-i} + \sum_{i=0}^q b_i \varepsilon_{t-i}, \quad \varepsilon_t \sim \mathcal{N}(0, 1). \quad (2.6)$$

These processes are linear, but not cyclic, a property which is implicitly assumed by the Fourier transformation on a finite data set. We suggest a cyclic process with the periodogram of an autoregressive process:

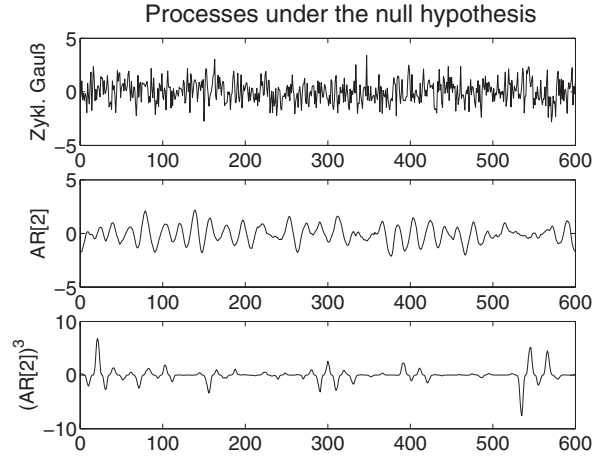
$$X_t = A + \sqrt{\frac{2\pi}{N}} \sum_{j=1}^M \sqrt{B_j^2 + C_j^2} \cos(\omega_t j + \theta_j), \quad t = 1, \dots, N$$

with  $A \sim \mathcal{N}(0, 1)$ ,  $B_j, C_j \sim \mathcal{N}(0, \sigma_j^2)$ ,  $\theta_j \sim \mathcal{U}[0, 2\pi]$  and  $\sigma_j^2 = I_Y(\omega_j)$ .  $I_Y$  is the periodogram of an AR[2] process.  $A$ ,  $B_j$ ,  $C_j$  and  $\theta_j$  are independent for  $j = 1, \dots, M$ . Besides,  $\omega_t = 2\pi t/N$ ,  $M = (N - 1)/2$  for odd  $N$  and  $M = (N - 2)/2$  for even  $N$ . Note, that

$$B_j^2 + C_j^2 \sim \sigma_j^2 \chi_2^2$$

with mean  $2\sigma_j^2$ . The periodogram of the process reads  $I_X(\omega_k) = (B_k^2 + C_k^2)/4$  for  $k = 1, \dots, M$ . The same approach with an exponentially decreasing periodogram with  $\sigma_j^2 = \exp(-j/M)$  leads to a cyclic Gaussian process (see Fig. 2.5).

The extension of the null hypothesis with invertible, nonlinear measurement function  $h$  is investigated in this study with  $h(x) = x^3$  and an AR[2].



**Fig. 2.5** Processes under the null hypothesis. Cyclic Gaussian process with exponentially decreasing spectrum, AR[2] process and AR[2] process measured via  $h(x) = x^3$

### 2.5.2 Processes Under the Alternative: Nonlinear Deterministic and Stochastic Systems

Deterministic systems can be written as differential systems oder difference equations. All consequent time points derive exactly from the initial values. If the distance between neighbored trajectories increases exponentially, the system is chaotic. Here we use the logistic map as example for a chaotic difference equation [7]

$$x_i = rx_{i-1}(1 - x_{i-1}), \quad 3.6... < r \leq 4.$$

For  $r = 4$  chaotic behavior occurs and the spectrum is not distinguishable from white noise. All information about the nonlinearities is saved in the phases, making this toy system interesting for our study (see Fig. 2.6).

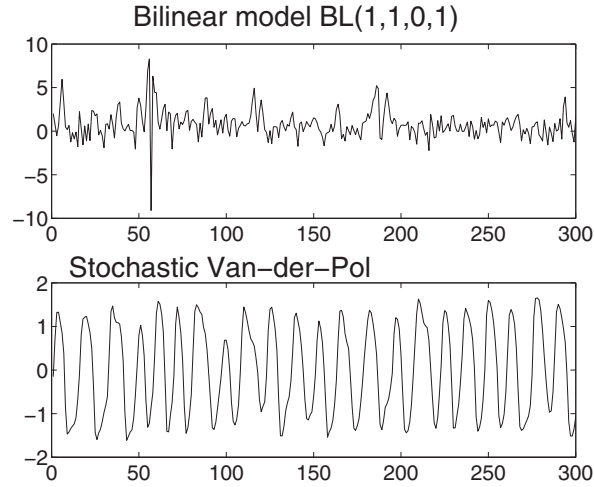
The most famous differential systems with chaotic behavior are the Lorenz and Rössler systems [11, 15]:

$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= -y + x(r - z) \\ \dot{z} &= xy - bz\end{aligned}$$

here with  $\sigma = 10$ ,  $b = 8/3$  and  $r = 40$ , and

$$\begin{aligned}\dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + (x - c)z\end{aligned}$$

here with  $a = 0.1$ ,  $b = 0.1$  and  $c = 18$ .



**Fig. 2.6** Bilinear model and stochastic Van-der-Pol oscillator

A stable limit cycle is given by the Van-der-Pol oscillator [22]

$$\ddot{x} = \mu(1 - x^2)\dot{x} - x, \quad \mu > 0.$$

Bilinear models are the extension of ARMA models. We use the bilinear model BL(1,0,1,1) [18]

$$X_i = aX_{i-1} + bX_{i-1}\varepsilon_{i-1} + \varepsilon_i,$$

with  $a = b = 0.4$ . The deterministic Lorenz, Rössler and Van-der-Pol systems can be disturbed with additive noise leading, e.g., to

$$\ddot{x} = \mu(1 - x^2)\dot{x} - x + \varepsilon, \quad \mu > 0.$$

### 2.5.3 Test Statistics

This subsection presents published and new developed test statistics, which are used to examine the qualitative and quantitative properties of the surrogate methods in the next main section. Since linear and nonlinear processes have to be distinguished from each other, emphasis is laid on test statistics which are sensitive for this difference.

#### Skewness and Kurtosis

Deviations from Gaussian amplitude distributions can be measured with the centered third and fourth moment, skewness and kurtosis:

$$\text{Skewness} = \frac{1}{N} \sum_{i=1}^N \left[ \frac{x_i - \bar{x}}{\sigma} \right]^3,$$

$$\text{Kurtosis} = \left\{ \frac{1}{N} \sum_{i=1}^N \left[ \frac{x_i - \bar{x}}{\sigma} \right]^4 \right\} - 3.$$

### Error of Nonlinear Prediction

More robust than the estimation of the correlation dimension is the the estimation of the nonlinear Prediction error. It does not need a certain scaling behavior and is easier to interpret. After embedding the time series to

$$\mathbf{x}_i = (x_i, x_{i-\tau}, x_{i-2\tau}, \dots, x_{i-(m-1)\tau})^T \quad (2.7)$$

the nonlinear prediction can be calculated as

$$F_{\varepsilon, \tau}(\mathbf{x}_i) = \frac{1}{N_{\varepsilon}(\mathbf{x}_i)} \sum_j \mathbf{x}_{j+\tau} \quad \forall j \neq i \text{ with } \|\mathbf{x}_i - \mathbf{x}_j\| < \varepsilon.$$

For every  $\mathbf{x}_i$ , the point  $\hat{\mathbf{x}}_{i+s} = F_{\varepsilon, s}(\mathbf{x}_i)$  is predicted.  $N_{\varepsilon}(\mathbf{x}_i)$  is the number of all neighbors of  $\mathbf{x}_i$  with a distance less than  $\varepsilon$ . As test statistic the well defined Prediction error is used:

$$\gamma(m, \tau, \varepsilon) = \left( \frac{1}{N} \sum \|\mathbf{x}(t + \tau) - F_{\varepsilon, \tau}(\mathbf{x}(t))\|^2 \right)^{1/2}.$$

### Time Invariance

We suggest a very powerful test statistic to measure the time invariance:

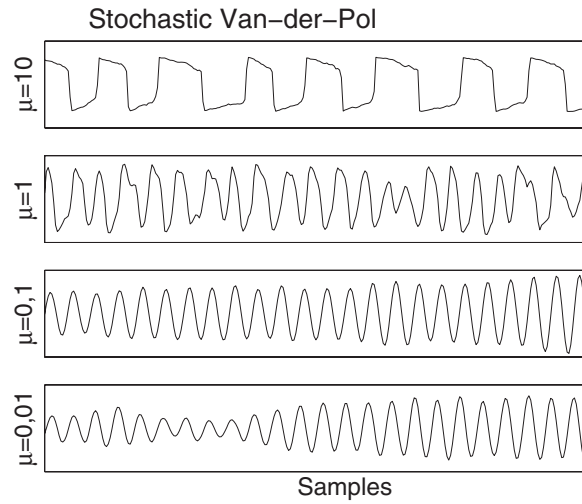
$$\max \left\{ \frac{\#\{|x_{i+1}| > |x_i|\}}{\#\{|x_{i+1}| < |x_i|\}}, \frac{\#\{|x_{i+1}| < |x_i|\}}{\#\{|x_{i+1}| > |x_i|\}} \right\}$$

after demeaning  $\mathbf{x}$ . Application to the stochastic Van-der-Pol oscillator illustrates the power of this test statistic (Fig. 2.7). Even with  $\mu = 0.01$  time invariance test statistic detects the nonlinearity whereas it is not visible for  $\mu = 0.1$ . Besides, this test statistic does not depend on the end-to-end-distance  $|x_1 - x_N|$ .

### Normalized End-to-End-Distance

The Fourier transform considers a finite time series as part of an infinite, periodical time series. A jump between the first and last data point  $|x_N - x_1|$  bigger than the mean point-to-point-distance leads to a sawtooth behavior of the infinite long time series. This needs an exact adjustment of the Fourier phases and also changes the Fourier amplitudes.

$$\text{Normalized end-to-end-distance} = \frac{|x_1 - x_N|}{\frac{1}{N-1} \sum_{i=2}^N |x_i - x_{i-1}|}.$$



**Fig. 2.7** Sensitivity of time invariance. Displayed are realizations of the stochastic Van-der-Pol oscillator with different values for  $\mu$ , corresponding to a different strength of the nonlinearity. The test statistic time invariance is able to detect the nonlinearity even for  $\mu = 0.01$

### Biggest Jump

The maximum distance between two consecutive points is given by

$$\max_{j=2,\dots,N} |x_{j+1} - x_j|.$$

### Smoothness

The smoothness of a time series can be quantified by:

$$\text{Smoothness} = \frac{\frac{1}{N-1} \sum_{i=2}^N |x_i - x_{i-1}|}{\frac{1}{N} \sum_{i=1}^N |x_i - \bar{x}|}.$$

### Normalized Distribution of Fourier Phases

The test statistic *distribution of Fourier phases* measured via the Kolmogorov-Smirnov-statistic has already been introduced in Sect. 2.3. To reduce the linear dependency on the end-to-end-distance, the KS-value is divided by the end-to-end-distance for the normalized version.

### 2.5.4 Qualitative Analysis Under the Null Hypothesis

The surrogates approach has been developed in order to estimate the distribution of a test statistic under the null hypothesis. In case of a false approximation the size and critical region of the corresponding test may change. Then the behavior of the corresponding test is unknown and hence the test useless. Furthermore, since a test is constructed for the detection of violations of the null hypothesis, the power is of high practical interest.

This section investigates numerically, whether the surrogate methods generate the correct distribution of a test statistic under the null hypothesis. We analyze qualitatively whether the surrogate methods are able to reproduce the distribution of a test statistic under the null hypothesis. The following hypotheses are investigated:

1. Linear cyclic Gaussian process with exponentially decreasing spectrum
2. Linear cyclic Gaussian process with spectrum of an AR[2] process
3. Linear Gaussian process, realized with an AR[2] process
4. Linear Gaussian process measured via a nonlinear invertible measurement function, here realized with an AR[2] process and  $h(x) = x^3$ .

For ten realizations of every process under null hypothesis 200 surrogate time series are generated and several test statistics are calculated yielding values  $t_1, \dots, t_N$ . Their cumulative distribution

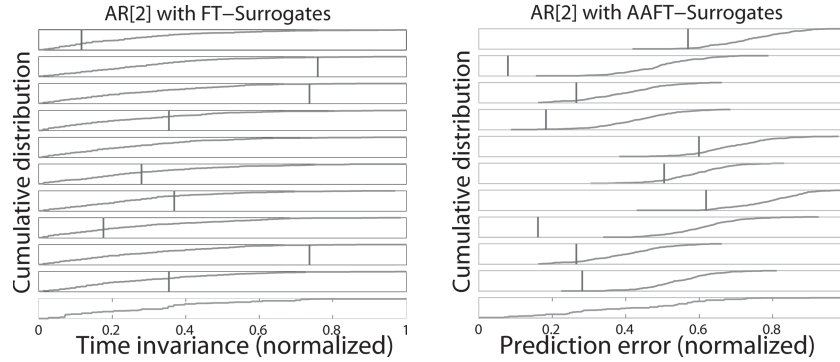
$$F(t) = \frac{\#\{t_i, i = 1, \dots, N \mid t_i \leq t\}}{N} \quad (2.8)$$

is qualitatively compared to the one of the test statistic based on 200 realizations of the underlying process itself. Figure 2.8 shows the result: For some combinations of null hypothesis, surrogate method and test statistic, the original distribution displayed in the bottom panel can be regained in the upper ten panels. Here, this occurs for the combination linear Gaussian process with FT-Surrogates and the test statistic *time invariance*. For the combination linear Gaussian process with AAFT-Surrogates and the test statistic *prediction error* the original distribution is not maintained but, even worse, the distribution depends on the value of the test statistic for the original time series marked in each panel by the vertical line.

The qualitative characterization has been done for six surrogate methods, eleven test statistic and four null hypothesis. The results are summarized in Tables 2.1–2.4. The  $\checkmark$  symbols a correct reproduction of the test statics distribution, a false reproduction is marked by  $\times$ , and  $\circ$  has been used for ambiguous cases.

The first null hypothesis leads to a correct distribution for most combinations. Except for a few cases the phase fixing FPH-method does not yield a correct distribution, as expected. The methods AAFT and DFS scale the amplitude distributions to the distribution of the original time series resulting in always the same value for the test statistics skewness and kurtosis.





**Fig. 2.8** Cumulative distributions of test statistic values under the null hypothesis. Surrogate methods should asymptotically be able to reproduce the complete cumulative distribution of any test statistic under the null hypothesis. However it turns out, that this depends strongly on the chosen null hypothesis, test statistic and surrogate method. The new developed time invariance measure on the left is able to reproduce the original cumulative distribution (panel 11 at the bottom) for each of ten AR[2] realizations (panels 1–10). Using the prediction error as test statistic and AAFT surrogates show a strong dependency of the test statistic value of the original data realization, marked by vertical lines

Since a constant value does not reflect the statistic fluctuations, the original distribution is not regained.

Similar results yield the simulations for the second null hypothesis, the cyclic Gaussian process with periodogram of a linear Gaussian process. Only the AR-Fit method has problems, which is based on the vanishing

**Table 2.1** Qualitative analysis under the null hypothesis 1: Gaussian cyclic linear process with exponentially decreasing spectrum

	FT	AAFT	DFS	NPS	FPH	AR-Fit
Skewness	✓	×	×	✓	×	✓
Kurtosis	✓	×	×	✓	×	✓
Prediction error	✓	✓	✓	✓	✓	✓
Time invariance	✓	✓	✓	✓	✓	✓
Mean distance	×	×	×	○	×	×
End-to-end-distance	✓	✓	✓	✓	×	✓
Norm. end-to-end-distance	✓	✓	✓	✓	×	✓
Biggest jump	✓	✓	✓	✓	✓	✓
Smoothness	×	×	×	○	×	×
Phase distribution	✓	✓	✓	×	×	✓
Norm. phase distribution	✓	✓	○	○	×	✓

**Table 2.2** Qualitative analysis under the null hypothesis 2: Gaussian cyclic linear process with spectrum of a linear Gaussian process

	FT	AAFT	DFS	NPS	FPH	AR-Fit
Skewness	✓	×	×	✓	✓	✓
Kurtosis	✓	×	×	✓	✓	✓
Prediction error	×	×	×	×	×	×
Time invariance	✓	✓	✓	✓	✓	✓
Mean distance	×	×	×	×	×	×
End-to-end-distance	✓	✓	✓	✓	×	×
Norm. end-to-end-distance	✓	✓	✓	✓	×	×
Biggest jump	✓	✓	✓	✓	○	✓
Smoothness	×	×	×	×	×	×
Phase distribution	✓	✓	✓	×	×	×
Norm. phase distribution	✓	○	✓	○	○	×

end-to-end-distance for the cyclic process. This behavior is not reproduced by the generated AR[80].

Very good results yield the AR-Fit method for the third null hypothesis, the linear Gaussian process realized via an autoregressive process, since the right model class was used for the Monte-Carlo-approach. The *actual* surrogates methods drop behind indicating their demand for cyclicity.

The fourth null hypothesis, the linear Gaussian process measured via an invertible, nonlinear function is for most combinations already part of the “alternative”, since a Gaussian distribution is assumed. Even the AAFT method, which was constructed for this setting, does not succeed in combination with most test statistics.

**Table 2.3** Qualitative analysis under the null hypothesis 3: Linear Gaussian process

	FT	AAFT	DFS	NPS	FPH	AR-Fit
Skewness	✓	×	×	✓	✓	✓
Kurtosis	✓	×	×	✓	×	✓
Prediction error	×	×	×	×	×	✓
Time invariance	✓	✓	✓	✓	✓	✓
Mean distance	○	×	×	×	×	×
End-to-end-distance	×	×	×	×	×	✓
Norm. end-to-end-distance	×	×	✓	×	×	✓
Biggest jump	✓	✓	×	✓	○	✓
Smoothness	×	×	×	○	×	×
Phase distribution	×	×	×	×	×	✓
Norm. phase distribution	○	○	○	×	○	✓

**Table 2.4** Qualitative analysis under the null hypothesis 4: Linear Gaussian process measured with nonlinear function

	FT	AAFT	DFS	NPS	FPH	AR-Fit
Skewness	×	×	×	×	×	×
Kurtosis	×	×	×	×	×	×
Prediction error	×	✓	×	×	×	×
Time invariance	✓	✓	✓	✓	✓	✓
Mean distance	×	○	×	×	×	×
End-to-end-distance	×	×	×	×	×	✓
Norm. end-to-end-distance	×	○	×	×	×	○
Biggest jump	×	×	×	×	×	×
Smoothness	×	×	×	×	×	×
Phase distribution	×	×	×	×	×	○
Norm. phase distribution	○	○	○	×	×	○

The qualitative analysis can be summarized as follows: The accuracy of the reproduction of the test statistics distribution under the null hypothesis depends *strongly* on the chosen combination of null hypothesis, surrogates method and test statistic. Only for one test statistic, the new developed time invariance test statistic, all surrogate methods and null hypotheses lead to a correct distribution. For the normalized end-to-end-distance *no* combination was successful.

### 2.5.5 Quantitative Analysis of Asymptotic, Size and Power

This section investigates numerically, which power the surrogate methods have for different alternatives. The power is analyzed depending on data length and strength of violation, in order to examine the asymptotic behavior and to establish a ranking of the different surrogate methods. A two-sided test with a significance level of 6% is constructed for the following simulations. For every original time series 99 surrogates are generated and the value of a given test statistic calculated. Is the value for the original time series under the first three or last three of the ranked 100 test statistic values, the null hypothesis is rejected. This procedure is repeated for 300 original time series.

The quantitative analysis of the asymptotic determines the power depending on the length of the time series. For the time invariance test statistic and time series of the logistic map, every surrogate method FT, AAFT, DFS, NPS, IAAFT and iDFS show a similar behavior (Fig. 2.9). For 256 or more data points, the null hypothesis is rejected in all cases and for all methods. The same time series evaluated with the test statistic smoothness lead to a ranking of the surrogate methods: The FT-surrogates reach maximum power already for 256 data points, NPS-surrogates for 1024, AAFT-surrogates for 8192, DFS- and iDFS-surrogates need about 10000 data points to reject the

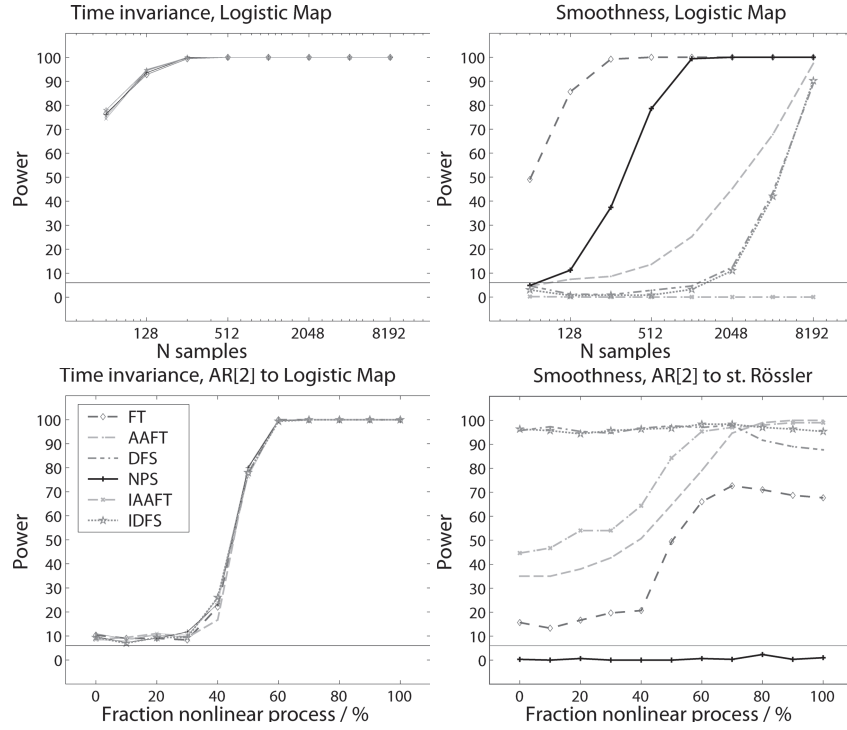


Fig. 2.9 Quantitative analysis of asymptotic, size and power

false null hypothesis with a power of 100%. The IAAFT-surrogates have no power at all. Unfortunately, this ranking is not reproduced in other, not shown combinations of alternatives and test statistics, indicating the irregular behavior of tests based on surrogate methods.

Necessary requirement for the usefulness of a test is on the one hand a correct size and on the other hand a good power. This section investigates the power of surrogates based tests for linearity with variable strength of the null hypothesis violation. First, two stochastic processes  $X_0$  and  $X_A$  are combined to a new process

$$X = (1 - a)X_0 + aX_A, \quad a \in [0; 1]. \quad (2.9)$$

The process  $X_A$  violates the null hypothesis, which is fulfilled by  $X_0$ . A representative selection of processes  $X_A$  is chosen, in order to investigate violations of all kind. The parameter  $a$  increases from 0 to 1 and consequently the influence of the nonlinear process increases from 0% to 100%.

Figure 2.9 shows two cases to discuss the behavior of surrogates based linearity tests. For the test statistic time invariance all surrogate methods lead to a similar behavior and reject the null hypothesis for 50%–60% of the nonlinear process. The size is a bit too high. For the test statistic smoothness, the size

is not correct for all surrogate methods except for the NPS method, which has on the other hand no power for the violation with the stochastic Rössler system. The worst case of nearly 100% rejection of the true null hypothesis based on the DFS- and iDFS-surrogates indicates again the irregular behavior of surrogates based tests.

### 2.5.6 Summary of the Simulation

The nonlinearity of the stochastic Van-der-Pol oscillator can be changed via the parameter  $\mu$ . The cases of Fig. 2.10 summarize the behavior of surrogates based linearity tests:

- For some test statistics like *time invariance*, every method works similarly and in a plausible way (a).
- Some test statistics have a different behavior which can be derived from the surrogates construction like for the *kurtosis* in combination with amplitude adjusting surrogates (b).

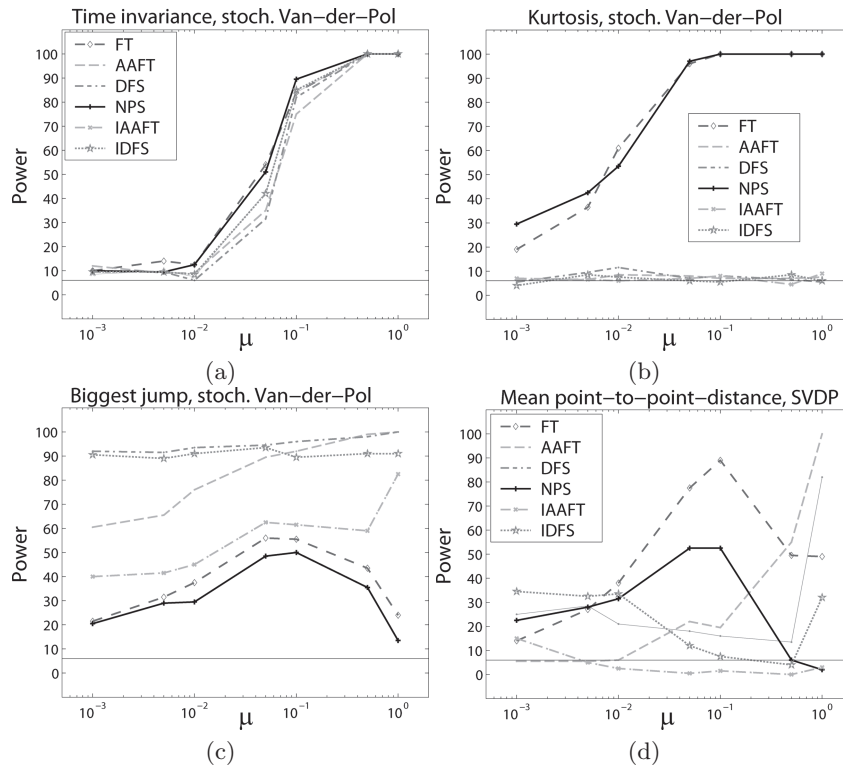


Fig. 2.10 Quantitative analysis of size and power

- For some test statistic, like for the *biggest jump*, a performance order of the methods can be established (c). Unfortunately, this order is not maintained but reversed in other combinations.
- And finally, for some test statistics like the *mean point to point distance*, no regular or reasonable behavior of the surrogates methods occurs (d). This would lead in a statistical test to spurious results and wrong consequences.

## 2.6 The AAFT and the IAAFT Approach – A Detailed Comparison

In this section, we study level accuracy and power properties of tests based on Amplitude Adjusted Fourier-Transformed (AAFT) surrogates and on Iterated AAFT (IAAFT) surrogates. We will see that both methods work rather reliable as long as the process do not have long coherence times. IAAFT is outperformed by AAFT: It has a similar performance in many setups but it is not stable in general. We will see some examples where it breaks down.

The AAFT method and the IAAFT approach are designed for the general null hypothesis of a linear Gaussian process that is transformed by a nonlinear invertible function. More explicitly, we will assume that the observed process  $\{x_t\}$  is generated by a transformation:

$$x_t = h(z_t),$$

where  $z_t$  is a Gaussian stationary process (e.g.,  $z_t$  is an autoregressive moving average (ARMA) process of order  $(p, q)$ :  $z_t = \sum_{j=1}^p a_j z_{t-j} + \sum_{j=0}^q b_j \varepsilon_{t-j}$ , where  $\{\varepsilon_t\}$  is a sequence of uncorrelated Gaussian random variables and  $b_0 = 1$ ). It has been argued that this process is linear as non-linearity is contained only in the invertible transformation function  $h(\cdot)$ . In this section we will discuss AAFT and IAAFT that both have been proposed for this model. We will give a more detailed description of both procedures below. The AAFT method was first discussed by [19]. Its basic idea is to apply the Fourier based surrogate data method after the data have been transformed to Gaussianity. An alternative procedure is the IAAFT approach of [16]. In this approach an iterative algorithm is used for the generation of the surrogate data. The iteration procedure uses alternatively two steps. In one step the surrogate data are transformed such that their marginal empirical distribution coincides with the marginal empirical distribution of the observed time series. In the second step phases are randomized similarly as in the generation of surrogate data. For a detailed description see again the next section.

Up to now there is no theory available for the performance of AAFT and IAAFT. In this section we present a detailed discussion of level accuracy of statistical tests based on AAFT and IAAFT. For a large class of hypothesis models and for a set of test statistics level accuracy will be checked. We will see that the methods can not guarantee a test that is valid for the large null

hypothesis of transformed Gaussian stationary processes. The methods do not perform well for time series with long coherence times. But we will also present examples of well behaved standard Gaussian linear models where IAAFT turn out as very unstable. We now give a detailed description of the AAFT and IAAFT algorithms.

### 2.6.1 The AAFT and the IAAFT Approach – Definition

The method of Amplitude Adjusted Fourier-Transformed (AAFT) proceeds as follows:

1. Let  $\mathbf{r} = (r_1, r_2, \dots, r_n)$  be the rank vector of an observational data vector  $(x_1, \dots, x_n)$  and  $(x_{(1)}, \dots, x_{(n)})$  be its ordered sample.
2. Generate a sample  $w_1, \dots, w_n$  of i.i.d. standard normal random variables and denote its ordered sample by  $(w_{(1)}, \dots, w_{(n)})$ . Put  $\hat{z}_t = w_{(r_t)}$ ,  $t = 1, \dots, n$ . Then  $\hat{z}_t$  has the same rank in  $\hat{z}_1, \dots, \hat{z}_n$  as  $x_t$  in  $x_1, \dots, x_n$ .
3. Obtain phase-randomized surrogates of  $\hat{z}_t$ , say  $z_t^*$ ,  $t = 1, \dots, n$ .
4. The surrogate data sample  $x_t^*$  of  $x_t$  is defined by  $x_t^* = x_{(r_t^*)}$ . Here  $r_t^*$  is the rank of  $z_t^*$  in the series  $z_1^*, \dots, z_n^*$ .

In Steps 1 and 2, the observed data are transformed to normal variables. This is done by using the random transformation  $\hat{z}_t = \hat{g}(x_t) = \hat{\Phi}^{-1}\hat{F}(x_t)$ . Here  $\hat{\Phi}$  is the empirical distribution function of the normal random variables  $w_1, \dots, w_n$  and  $\hat{F}$  is the empirical distribution function of  $x_1, \dots, x_n$ . In Step 3, the transformed data are phase randomized: Fourier-transformed surrogates are generated. In the final step, the surrogate data are transformed back to the original values. This is done by sorting the observed data according to the ranks of the Fourier transformed surrogates: The phase randomized data  $z_1^*, \dots, z_n^*$  are transformed by using the random transformation  $x_t^* = \hat{g}^{-}(z_t^*) = \hat{F}^{-1}\hat{G}(z_t^*)$ , where now  $\hat{G}$  is the empirical distribution function of  $z_1^*, \dots, z_n^*$ . The resulting data  $x_1^*, \dots, x_n^*$  are called AAFT surrogates. Note that in this procedure the transformation  $\hat{g}^{-}$  in Step 4 is typically not the inverse of the transformation  $\hat{g}$  used in Step 2. They differ because  $\hat{F}$  differs from  $\hat{G}$ . In particular, the transformation  $\hat{g}$  does not depend on the surrogate sample whereas  $\hat{g}^{-}$  does.

The basic model assumptions imply that the transformed data  $\hat{z}_t$  and the FT-based surrogates  $z_t^*$  follow approximately a linear Gaussian process. Thus, it may be reasonable directly to base the statistical inference on these time series and to check if  $\hat{z}_t$  follows a linear Gaussian process. This would suggest to calculate a test statistic for the transformed series  $\hat{z}_t$  and to calculate critical levels for this test by using the surrogates  $z_t^*$ . We will consider both type of tests, tests based on the original time series  $x_t$  and their AAFT surrogates  $x_t^*$  and tests based on  $\hat{z}_t$  and  $z_t^*$ .

Iterative AAFT (IAAFT) is an iterative algorithm based on AAFT. It was proposed by [16] as an improved algorithm of AAFT. This algorithm is used to generate resamples for the same hypothesis as AAFT, i.e., the hypothesis of a

transformed linear Gaussian process. IAAFT is a method to produce surrogate data which have the same power spectrum and the same empirical distribution as the observed data. Note that this aim is not achieved by AAFIT. The IAAFT algorithm proceeds iteratively. It iteratively corrects deviations in the spectrum and deviations in the empirical distribution (between the surrogates and the original time series). The algorithm proceeds in the following steps:

1. Generate an ordered list of the sample ( $x_{(1)} \leq \dots \leq x_{(n)}$ ) and calculate the periodogram  $I_k^2 = |\sum_{t=0}^{n-1} x_t e^{it\omega_k}|^2$ ,  $\omega_k = 2\pi k/n$ ,  $k = 1, \dots, n$ .
2. Initialization step: generate a random permutation (sample without replacement)  $\{x_t^{a,(0)}\}$  of the data  $\{x_t\}$ .
3. Iteration steps:
  - (i) At the  $j$ -th iteration, take the Fourier transform of  $\{x_t^{a,(j)}\}$ , replace the squared amplitudes by  $I_k^2$  (without changing the phases) and transform back by application of the inverse Fourier transform:

$$x_t^{b,(j)} = 1/\sqrt{n} \sum_{s=0}^{n-1} \frac{\widehat{x}_s^{a,(j)}}{|\widehat{x}_s^{a,(j)}|} |I_s| \exp(-it\omega_s),$$

where  $\widehat{x}_t^{a,(j)} = 1/\sqrt{n} \sum_{s=0}^{n-1} x_s^{a,(j)} \exp(-it\omega_s)$  is the discrete Fourier transform of  $\{x_s^{a,(j)}\}$ .

- (ii) The resulting series in (i) is rescaled back to the original data:

$$x_t^{a,(j+1)} = x_{(r_t^j)},$$

where  $r_t^j$  is the rank of  $x_t^{b,(j)}$  in  $x_1^{b,(j)}, \dots, x_n^{b,(j)}$ .

4. Repeat (i) and (ii) in Step 3 until the relative difference in the power spectrum is sufficiently small. The limiting values  $x_t^*$  of  $x_t^{b,(j)}$  (or  $x_t^{a,(j)}$ ) are called IAAFT surrogates.

### 2.6.2 Comparison of AAFIT and IAAFT – Numerical Experiments

In this section, results based on numerical experiments are presented. We have conducted simulations for the following models.

- $M_1$  :  $x_t$  is an i.i.d. sequence with distribution  $\chi_1^2$ .
- $M_2$  :  $x_t$  is an i.i.d. sequence with distribution  $U[0, 1]$ .
- $M_3$  :  $x_t = z_t^3$ ,  $z_t$  is a circular process with  $\sigma_j^2 = \exp(-j/m)$ .
- $M_4$  :  $x_t = z_t$ ,  $z_t$  is a circular process with  $\sigma_j^2 = \exp(-j/8)$ .
- $M_5$  :  $x_t = z_t^3$ ,  $z_t$  is a circular process with  $\sigma_j^2 = \exp(-j/8)$ .
- $M_6$  :  $x_t = z_t^3$ ,  $z_t = 1.4z_{t-1} - 0.48z_{t-2} + \varepsilon_t$ .
- $M_7$  :  $x_t = z_t^3$ ,  $z_t = 1.8z_{t-1} - 0.81z_{t-2} + \varepsilon_t$ .



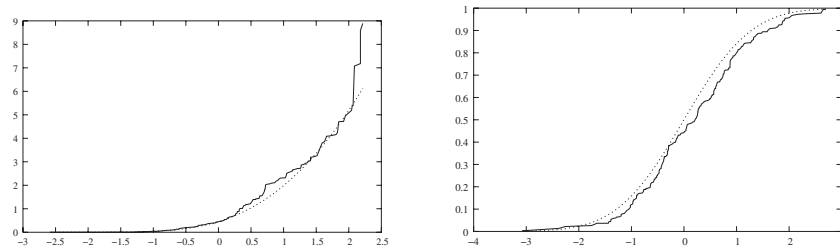
In all models the residuals  $\varepsilon_t$  are i.i.d. and have a standard normal distribution. The circular processes in  $M_4$ ,  $M_5$  and  $M_6$  are generated as follows:

$$z_t = A + c \sum_{j=1}^m \sqrt{B_j^2 + C_j^2} \cos(\omega_j t + \theta_j), \quad t = 1, \dots, n.$$

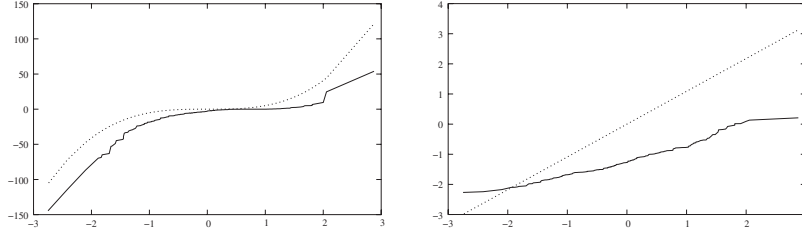
Here  $A \sim \mathcal{N}(0, \sigma^2)$ ,  $B_j, C_j \sim \mathcal{N}(0, \sigma_j^2)$ ,  $\theta_j \sim U[0, 2\pi]$  and  $A$ ,  $B_j$ ,  $C_j$ , and  $\theta_j$  are independent. Furthermore, we used the notation  $\omega_j = 2\pi j/n$ ,  $c = \sqrt{2\pi/n}$ ,  $m = (n-2)/2$ . Note that after application of a transformation the circular Gaussian processes are still circular but in general not Gaussian distributed.

All models are transformed stationary Gaussian processes. Models  $M_1$  and  $M_2$  are i.i.d. processes. Trivially, both can be written as  $x_t = h(z_t)$  with  $z_t$  i.i.d. standard gaussian.  $M_3$  is a transformed Gaussian circular process. For this circular process  $\sigma_j^2 = \exp(-j/m)$ ,  $m = n-1/2$  and it exhibits a performance near to an i.i.d. sequence. Models  $M_4$  and  $M_5$  are circular processes with  $\sigma_j^2 = \exp(-j/8)$ . Model  $M_4$  is a Gaussian circular process and  $M_5$  is a nonlinear transformation of  $M_4$ . The fast decay of  $\sigma_j^2$  leads to long coherence times. Model  $M_6$  is a transformed AR(2) process transformed by  $h(x) = x^3$ . The underlying AR(2) process has all roots inside the unit circle (0.6 and 0.8).  $M_7$  represents a transformed Gaussian AR process of order two. The underlying AR(2) process has roots 0.9 close to unity.

In the implementation of the AAFT algorithm the transformation function  $h(\cdot)$  and its inverse  $h^{-1}(\cdot)$  are estimated. One may expect that AAFT works as well as a classical surrogate data test if these estimates are accurate. For this reason, we checked the accuracy of these estimates for models  $M_1 - M_7$ . We consider a single realization  $(x_1, \dots, x_n)$  of each model with  $n = 256$ . In Figs. 2.11–2.14 the series are plotted against  $(\hat{z}_1, \dots, \hat{z}_n)$  (horizontal axis) as solid lines. In the same plots, the true function  $h(\cdot)$  as a function of  $z_t$  is plotted as dotted line. For models  $M_3$ ,  $M_5$ – $M_7$ ,  $h(x) = x^3$  and for  $M_4$ ,  $h(x) = x$ .  $M_1$  and  $M_2$  are i.i.d. processes. So for these i.i.d. processes, the true transformation function is given as  $h(x) = F^{-1}\Phi(x)$ , where  $\Phi(\cdot)$  denote the distribution function of the standard normal distribution and  $F$  denotes



**Fig. 2.11** The estimated (solid line) and true (dotted line) transformation functions for models  $M_1$  and  $M_2$



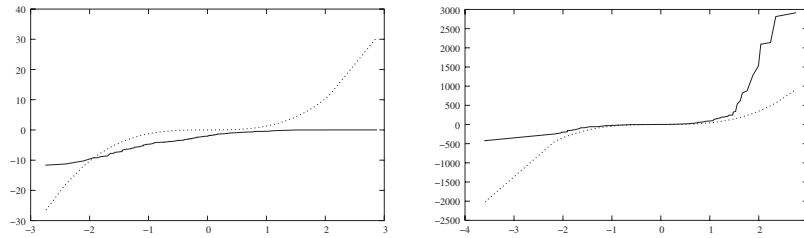
**Fig. 2.12** The estimated (solid line) and true (dotted line) transformation functions for models  $M_3$  and  $M_4$

the distribution function of  $\chi_1^2$  or  $U[0, 1]$ , respectively. We observe that for the two i.i.d. processes  $M_1$  and  $M_2$ , the transformation function is estimated very accurately. Approximately, this also holds for  $M_3$  and  $M_6$  (note that the  $z$ -axes differ and always they include extreme points of a standard normal law). For the other models the estimate behaves poorly, in particular in the tails.

We have performed numerical experiments for the following test statistics:

$$\begin{aligned}
 T_1 &= \frac{1}{n} \sum_{t=1}^{n-1} (X_t X_{t+1}^2 - X_t^2 X_{t+1}), \\
 S_1 &= \frac{\#\{X_t > X_{t+1}\}}{n}, \quad S_2 = n - 1 - S_1, \\
 T_2 &= S_1, \\
 T_3 &= \frac{|S_2 - S_1|}{S_1 + S_2}, \\
 T_4 &= \max_{\tau} Q(\tau), \quad Q(\tau) = \frac{\sum_{t=\tau+1}^n (X_{t-\tau} - X_t)^3}{[\sum_{t=\tau+1}^n (X_{t-\tau} - X_t)^2]^{3/2}}, \\
 T_5 &= \frac{1}{n} \sum_{t=1}^{n-2} \prod_{k=0}^2 (X_{t+k} - \bar{X}), \\
 T_6 &= \frac{1}{n} \sum_{t=1}^{n-4} \prod_{k=0}^4 (X_{t+k} - \bar{X}), \\
 T_7 &= \max \left\{ \frac{\#\{|X_{t+1} - \bar{X}| > |X_t - \bar{X}|\}}{\#\{|X_{t+1} - \bar{X}| < |X_t - \bar{X}|\}}, \frac{\#\{|X_{t+1} - \bar{X}| < |X_t - \bar{X}|\}}{\#\{|X_{t+1} - \bar{X}| > |X_t - \bar{X}|\}} \right\} \\
 T_8 &= C_n(r), \quad C_n(r) = \frac{\sum_{i=2}^n \sum_{j=1}^i I(\|\mathbf{X}_i^\nu - \mathbf{X}_j^\nu\| < r)}{n(n-1)/2}.
 \end{aligned}$$

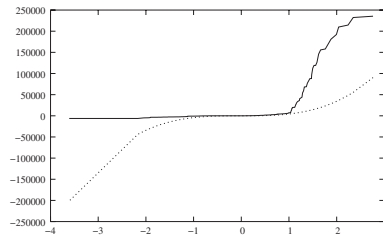
Here,  $I$  is the indicator function and  $\|\mathbf{X}\| = \max_k |X_k|$ . The vector  $\mathbf{X}_i^\nu = (X_{i-(\nu-1)d}, X_{i-(\nu-2)d}, \dots, X_i)^T$  is an element of the phase space with embedding dimension  $\nu$  and delay time  $d$ . We have used delay time  $d = 2$ .



**Fig. 2.13** The estimated (solid line) and true (dotted line) transformation functions for models  $M_5$  and  $M_6$ .

The results are reported for embedding dimension  $\nu = 4$ . We have used different values of  $r$  for different models. The test statistics  $T_1, \dots, T_4$  and  $T_7$  are statistics for checking time asymmetry. The tests  $T_5$  and  $T_6$  are based on joint higher order central moments and they test the nonlinearity of the dynamics.  $T_8$  is the correlation sum. It is the sample analog of the correlation integral. Other physically meaningful measures (for example, the correlation dimension, the maximum Lyapunov exponent, etc.) have been proposed to check the nonlinear chaotic behavior of a data generating process. But there is no automatic implementation of these test statistics. Thus, it is difficult to incorporate these statistics in a simulation study. This is the reason why we have considered correlation sums which can be computed by an automatic scheme for different values of the threshold parameter  $r$ .

In the simulations we generated data  $\mathbf{x}_n = (x_1, \dots, x_n)$  from Models  $M_1$ – $M_7$  and we calculated the test statistics  $T_j(\mathbf{x}_n)$  for  $j = 1, \dots, 6$ . We used sample size  $n = 512$ . For each simulated  $\mathbf{x}_n$ , 1000 surrogate resamples  $\mathbf{x}_n^*$  were generated. For each of these resamples, we calculated test statistics  $T_j(\mathbf{x}_n^*)$ ,  $j = 1, \dots, 6$ . The surrogate data test rejects the hypothesis of a linear stationary Gaussian process, if  $T_j(\mathbf{x}_n) > k_{j\alpha}^*$ , where  $k_{j\alpha}^*$  denotes the  $(1-\alpha)$ -th quantile of  $T_j(\mathbf{x}_n^*)$ . The first aim of our simulations is to check the level accuracy of this test, i.e., to check if the rejection probability on the hypothesis is approximately equal to the nominal level  $\alpha_{\text{nom}}$ :  $P[T_j(\mathbf{x}_n) > k_{j\alpha}] \approx \alpha_{\text{nom}}$ .



**Fig. 2.14** The estimated (solid line) and true (dotted line) transformation function for  $M_7$ .

For this check, the whole procedure was repeated 1000 times. The empirical fraction  $\hat{\alpha}$  of  $T_j(\mathbf{x}_n) > k_{j\alpha}$  is a Monte-Carlo-approximation of the level of the test. The simulation results are given in Tables 2.5–2.7. We have used the nominal value  $\alpha_{\text{nom}} = .05$ . The tables report how much  $\hat{\alpha}$  differs from  $\alpha_{\text{nom}}$ .

We used different categories in Tables 2.5–2.9 for the level accuracy of the tests, see the caption of Table 2.5. In category “– –” the test always rejects, in category “++++” the test always accepts. Both cases indicate a total break down of the test. In particular, the latter case indicates a totally erroneous performance of the test. Category “–” indicates that the test is too conservative. In particular, the test will have poor power for neighbored alternatives. We call level accuracies “–”, “ok” and “+” reasonable. Categories “++” and “+++” again indicate a break down of the test. The test behaves like a blind test: it has a performance comparable to a random number that is produced without looking at the data.

Tables 2.5 and 2.6 summarize the results for the AAFT surrogate data tests. Table 2.5 gives the results when the test statistics have been applied to the untransformed data  $x_t$  (AAFT I) and Table 2.6 shows the performance when the test statistics have been calculated for the transformed data  $\hat{z}_t$  (AAFT II). Both tests work quite well for Models  $M_1$ – $M_6$ . There is no big difference in the level accuracy of AAFT I and AAFT II. In Models  $M_1$ – $M_6$ , AAFT I is too liberal (“++”, “+++”) in two cases and too conservative in one case (“–”). AAFT II has a slightly poorer performance: it is too liberal in one case and too conservative in six cases. Both procedures outperform IAAFT, see Table 2.7. In Models  $M_1$ – $M_6$ , IAAFT is in five cases too liberal and in 9 cases too conservative. Furthermore, in 8 of these cases it totally breaks down: it always rejects or it always accepts. On the other hand, the AAFT I procedure never totally breaks down and the AAFT II procedure only in one case. This suggests that IAAFT is not stable. This may be caused by the iterative nature of the algorithm. The algorithm sometimes runs into

**Table 2.5** *Level accuracy for AAFT surrogate data tests based on  $(x_t, x_t^*)$ . The level accuracy is marked by “–” if the Monte-Carlo estimate  $\hat{\alpha}$  of the level is 0.0, “–” if  $0 < \hat{\alpha} \leq .015$ , “–” if  $.015 < \hat{\alpha} \leq .03$ , “ok” if  $.03 < \hat{\alpha} \leq .075$ , “+” if  $.075 < \hat{\alpha} \leq .125$ , “++” if  $.125 < \hat{\alpha} \leq .250$ , “+++” if  $.250 < \hat{\alpha} < 1$ , and “++++” if  $\hat{\alpha} = 1$ . The nominal level is .05*

Model	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$T_6$	$T_7$	$T_8$	$r$
$M_1$	ok	ok	ok	ok	ok	ok	ok	ok	.1
$M_2$	ok	ok	ok	ok	ok	ok	ok	ok	.2
$M_3$	ok	ok	ok	ok	ok	ok	ok	ok	.2
$M_4$	–	ok	ok	+	+	+	ok	+++	.1
$M_5$	–	ok	ok	ok	+	ok	ok	+	.1
$M_6$	–	ok	ok	+	ok	ok	ok	++	.2
$M_7$	ok	++	++	+	++	++	++	ok	.5

**Table 2.6** Level accuracy for AAFT surrogate data tests based on  $(\hat{z}_t, \hat{z}_t^*)$  with nominal level 0.05. The level accuracy is marked as in Table 2.5

Model	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$T_6$	$T_7$	$T_8$	$r$
$M_1$	ok	ok	ok	ok	ok	ok	ok	ok	.1
$M_2$	ok	ok	ok	ok	ok	ok	ok	ok	.2
$M_3$	ok	ok	ok	ok	ok	ok	ok	ok	.2
$M_4$	ok	ok	ok	ok	–	–	ok	+	.1
$M_5$	ok	ok	ok	ok	–	–	ok	+	.2
$M_6$	++	ok	ok	ok	–	–	ok	+	.5
$M_7$	+++	++	++	+	–	–	++	+++	.2

a totally misleading generation of surrogate data. Even for i.i.d. data (Models  $M_1$  and  $M_2$ ) IAAFT breaks totally down, here in five out of 16 cases. This means that here IAAFT does not capture the transformation of the data: in case of large deviations from normality the basic idea of IAAFT to have the same power spectrum as the observed data forces the method to biased estimates of the transformation.

Both, AAFT I and AAFT II, work perfectly well for i.i.d. data and for Model  $M_3$ . This performance can be easily theoretically verified. If the underlying process  $x_1, \dots, x_n$  is an i.i.d. sequence then their ranks  $(r_1, \dots, r_n)$  have a uniform distribution on the set of all permutations of  $(1, \dots, n)$ . Thus  $\hat{z}_1, \dots, \hat{z}_n$  is a random permutation of an i.i.d. sample of standard normal variables. Thus it is also an i.i.d. sample of standard normal variables, and in particular it is a circular stationary linear Gaussian process. It has been shown in [3] that the method of phase randomized surrogates has exact finite sample level for circular stationary processes, see also [13]. This immediately implies that AAFT I has exact finite sample level if the underlying process is i.i.d.. Furthermore, it also implies that AAFT II has exact finite sample level. This can be seen as follows. Consider first a deterministic sequence  $u_1, \dots, u_n$  and a test

**Table 2.7** Level accuracy for IAAFT surrogate data tests with nominal level 0.05. The level accuracy is marked as in Table 2.5

Model	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$T_6$	$T_7$	$T_8$
$M_1$	+	ok	ok	ok	ok	ok	–	–
$M_2$	–	–	–	–	ok	ok	+++	–
$M_3$	+	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	–	–	–	++
$M_7$	+	++	++	ok	–	–	–	–

statistic  $T$ . Order  $u_1, \dots, u_n$  in the same order as  $\widehat{z}_1, \dots, \widehat{z}_n$  and as its surrogates  $z_1^*, \dots, z_n^*$ . This gives two samples. Calculate now  $T$  for the two samples. This can be interpreted as AAFT I applied to a modified test statistic. Thus it achieves exact finite sample level. The procedure would be exactly equal to AAFT II if  $u_1, \dots, u_n$  would be chosen as  $x_1, \dots, x_n$ . This choice is nonrandom but the argument carries over because the order statistic  $x_{(1)} \leq \dots \leq x_{(n)}$  is independent of  $\widehat{z}_1, \dots, \widehat{z}_n$ .

Model  $M_3$  is a circular stationary model with short coherence time. Both AAFT procedures work well for this model. This does not extend to circular stationary model with long coherence times. This can be seen in the results for Models  $M_4$  and  $M_5$ . In some cases, the AAFT procedures have very poor level accuracies for these models. In particular this shows that in this respect AAFT behaves differently as (phase randomized) surrogate data testing that has exact level for all circular stationary (Gaussian) processes. The additional estimation of the transformation  $h$  that is incorporated in AAFT can lead to level inaccuracies in case of long coherence times. The poor performance of this estimation was illustrated in the plots for Models  $M_4$  and  $M_5$  in Figs. 2.12 and 2.13.

All procedures, AAFT and IAAFT performed poorly in Model  $M_7$ . AAFT I and IAAFT achieved reasonable results in 3 out of 8 cases, AAFT II only in one case.  $M_7$  is a near to unit root process. Reference [13] used simulations for the same process without transformation to illustrate a poor performance of phase randomized surrogate data tests for near to unit root processes. They considered two modifications of surrogate data testing. One was based on taking subsamples of the series with small gap between the first and last observation. The end to end mismatch correction leads to drastic improvements. Motivated by this result, we tried the same modification for both AAFT procedures for Model  $M_7$ . The results are reported in Tables 2.8 and 2.9. In Table 2.8 the mismatch correction procedure is applied to the observed series  $x_t$ , whereas in Table 2.9 it is applied to the transformed series  $\widehat{z}_t$ . For these experiments, we used the sample size  $N = 512$  and  $K_1 = K_2 = 40$ . (The starting point of the subseries is chosen among the first  $K_1$  observations, the last point among the last  $K_2$  observations, for details see [13]). There was no big difference if the method was applied to  $x_t$  or to  $\widehat{z}_t$ . There was also a gain after

**Table 2.8** Level accuracy for AAFT surrogate data tests with end to end correction. The end to end correction was applied to the untransformed data  $x_t$ . The nominal level is 0.05. The level accuracy is marked as in Table 2.5

	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$T_6$	$T_7$	$T_8$	$r$
$(x_t, x^*)$	--	+	+	ok	++	++	+	ok	.5
$(\widehat{z}_t, z^*)$	+	+	+	+	---	---	ok	+++	.2

**Table 2.9** Level accuracy for AAFT surrogate data tests with end to end correction. The end to end correction was applied to the transformed data  $\widehat{z}_t$ . The nominal level is 0.05. The level accuracy is marked as in Table 2.5

	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$T_6$	$T_7$	$T_8$	$r$
$(x_t, x^*)$	--	ok	+	ok	++	++	+	ok	.5
$(\widehat{z}_t, z^*)$	+	ok	+	+	---	---	+	+++	.2

application of the end to end correction, but not quite so impressive as for surrogate data tests. Now both AAFT methods worked for 5 out of 8 cases.

We also checked the performance of AAFT for the test statistics  $T_1, \dots, T_8$  on the following alternatives:

- $A_1$ : Logistic map:  $x_{t+1} = 4x_t(1 - x_t)$ ,  $x_0 \in (0, 1)$ .
- $A_2$ : Reversal logistic map:  $x_t = 0.5[1 + \varepsilon_t(1 - x_{t-1})^{\frac{1}{2}}]$  where  $\varepsilon_t$  is equal to 1 or  $-1$  with probability  $\frac{1}{2}$  and  $x_0 \sim \text{Beta}(.5, .5)$ .
- $A_3$ : Tent map:  $x_{t+1} = \begin{cases} 2x_t & \text{if } 0 \leq x_t \leq .5 \\ 2(1 - x_t) & \text{if } .5 \leq x_t \leq 1 \end{cases}$
- $A_4$ : Reversal tent map:  $x_t = .5(1 - \varepsilon_t + \varepsilon_t x_{t-1})$  where  $\varepsilon_t$  is equal to 1 or  $-1$  with probability  $\frac{1}{2}$  and  $x_0 \sim U[0, 1]$ .

Here we have considered two pairs of models  $(A_1, A_2)$  and  $(A_3, A_4)$ .  $A_1$  and  $A_3$  are two purely deterministic models and  $A_2$  and  $A_4$  are their stochastic counterparts. After time reversion the stochastic models  $A_2$  and  $A_4$  are identical to  $A_1$  or  $A_3$ , respectively. For a discussion of stochastic processes with time reversal deterministic processes see [1, 9, 10, 21].

We performed similar simulations for  $A_i$ ,  $i = 1, \dots, 4$  as for Models  $M_1, \dots, M_8$ . The results of the simulations are summarized in Tables 2.10 and 2.11. We see that both AAFT tests, the test based on  $(x_t, x_t^*)$  and the test based on  $(\widehat{z}_t, z_t^*)$  have a quite similar performance. For the test statistics  $T_1 - T_4$ ,  $T_7$  and  $T_8$  they have nearly the same power. As above, they show a different performance for  $T_5$  and  $T_6$ . The tests show quite different behavior. Three tests ( $T_3$ ,  $T_7$  and  $T_8$ ) always have a power near to one. The tests based

**Table 2.10** Estimated power of AAFT tests based on  $(x_t, x_t^*)$  with nominal level 0.05

Model	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$T_6$	$T_7$	$T_8$	$r$
$A_1$	1.00	.000	1.00	.999	.093	.043	.983	1.00	.1
$A_2$	.000	1.00	1.00	.059	.082	.027	.998	1.00	.1
$A_3$	1.00	.000	1.00	1.00	.235	.073	1.00	1.00	.1
$A_4$	.000	1.00	1.00	.067	.145	.175	1.00	1.00	.1

**Table 2.11** Estimated power of AAFT tests based on  $(\hat{z}_t, z_t^*)$  with nominal level 0.05

Model	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$T_6$	$T_7$	$T_8$	$r$
$A_1$	.994	.000	1.00	.999	.321	.693	.980	1.00	.1
$A_2$	.000	1.00	1.00	.054	.333	.718	1.00	1.00	.1
$A_3$	1.00	.000	1.00	1.00	.479	.664	1.00	1.00	.1
$A_4$	.000	1.00	1.00	.068	.368	.739	1.00	1.00	.1

on  $T_1$  and  $T_4$  have power near to one for the deterministic models  $A_1$  and  $A_3$  and have rejection probability less than the nominal level for the stochastic models  $A_2$  and  $A_4$ . The test statistic  $T_2$  behaves the other way around. It rejects always for  $A_2$  and  $A_4$  and it never rejects for  $A_1$  and  $A_3$ . The test statistics  $T_5$  and  $T_6$  are not stable in their performances.

### 2.6.3 Comparison of AAFT and IAAFT – A Summary

In the previous subsection we investigated level accuracy and power characteristics of Amplitude Adjusted Fourier–Transformed (AAFT) and improved AAFT (IAAFT) algorithms. Both approaches are methods to get a statistical test for the hypothesis of a transformed stationary linear Gaussian processes. In both methods surrogate data are generated and tests are based on the comparison of test statistics evaluated with original observations and with surrogate data. In our study AAFT outperforms IAAFT in keeping the level on the hypothesis. AAFT works quite well as long as the coherence time is not too long. IAAFT is not stable in general. In many cases it totally breaks down: it always rejects or it always accepts on the hypothesis. In case of long coherence times the performance of AAFT can be improved by using end to end corrections. But the improvements are not so impressive as for phase randomized surrogate data.

## 2.7 Conclusions

This chapter contains a detailed description of the performance for a range of surrogate methods in a variety of settings. In a simulation study with a more general focus we showed that the performance strongly depends on the chosen combination of test statistic, resampling method and null hypothesis. For one test statistic, the new introduced time invariance test statistic, all surrogate methods lead to accurate levels. In a more detailed comparison of AAFT and IAAFT, both methods perform well as long as the coherence time of the process is not too large. In this case, AAFT has a more reliable and more accurate level.



## References

- [1] M.S. Bartlett. Chances or chaos. *Journal of Royal Statistical Society, Series A*, 153:321–347, 1990.
- [2] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer, New York, 1987.
- [3] K.S. Chan. On the validity of the method of surrogate data. *Fields Institute Communications*, 11:77–97, 1997.
- [4] R. Dahlhaus and D. Janas. A frequency domain bootstrap for ratio statistics in time series. *Annals of Statistics*, 24:1934–1963, 1996.
- [5] K. Dolan and M.L. Spano. Surrogates for nonlinear time series analysis. *Physical Review E*, 64, 2001.
- [6] S. Elgar, R.T. Guza, and R.J. Seymour. Groups of waves in shallow water. *Journal of Geophysical Research*, 89:3623–3634, 1984.
- [7] M.J. Feigenbaum. Quantitative universality in a class of nonlinear transitions. *Journal of Statistical Physics*, 19:25–52, 1978.
- [8] P. Grassberger. Do climatic attractors exist? *Nature*, 323:609, 1986.
- [9] A.J. Lawrance. Chaos: but not in both directions! *Statistics and Computing*, 11:213–216, 2001.
- [10] A.J. Lawrance and N.M. Spencer. Curved chaotic map time series models and their stochastic reversals. *Scandinavian Journal of Statistics*, 25:371–382, 1998.
- [11] E.N. Lorenz. Deterministic aperiodic flow. *Journal of the Atmospheric Sciences*, 20:130, 1963.
- [12] E. Mammen and S. Nandi. Bootstrap and resampling. In Y. Mori J. E. Gentle, W. Härdle, editor, *Handbook of Computational Statistics*, pages 467–496. Springer, New York, 2004.
- [13] E. Mammen and S. Nandi. Change of the nature of a test when surrogate data are applied. *Physical Review E*, 70:16121–32, 2004.
- [14] M.B. Priestley. *Spectral analysis and time series*. Academic Press, London, 1989.
- [15] O.E. Roessler. An equation for continuous chaos. *Physical Letters A*, 57:397–381, 1976.
- [16] T. Schreiber and A. Schmitz. Improved surrogate data for nonlinearity tests. *Physical Review Letters*, 77:635–638, 1996.
- [17] T. Schreiber and A. Schmitz. Surrogate time series. *Physica D*, 142:346–382, 2000.
- [18] T. Subba Rao and M.M. Gabr. *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Springer, New York, 1984.
- [19] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J.D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58:77–94, 1992.
- [20] J. Timmer. What can be inferred from surrogate data testing? *Physical Review Letters*, 85:2647, 2000.

- [21] H. Tong and B. Cheng. A note on the one-dimensional chaotic maps under time reversal. *Advances in Applied Probability*, 24:219–220, 1992.
- [22] B. van der Pol. On oscillation-hysteresis in a simple triode generator. *Philosophical Magazine*, 43:700–719, 1922.