**Nonlinear Processes
in Geophysics**

Statistical characteristics of surrogate data based on geophysical measurements

V. Venema[1]**, S. Bachner**[1]**, H. W. Rust**[2]**, and C. Simmer**[1]

[1]Meteorologisches Institut, Universität Bonn, Germany
[2]Potsdam Institute for Climate Impact Research, Germany

Abstract. In this study, the statistical properties of a range of measurements are compared with those of their surrogate time series. Seven different records are studied, amongst others, historical time series of mean daily temperature, daily rain sums and runoff from two rivers, and cloud measurements. Seven different algorithms are used to generate the surrogate time series. The best-known method is the iterative amplitude adjusted Fourier transform (IAAFT) algorithm, which is able to reproduce the measured distribution as well as the power spectrum. Using this setup, the measurements and their surrogates are compared with respect to their power spectrum, increment distribution, structure functions, annual percentiles and return values. It is found that the surrogates that reproduce the power spectrum and the distribution of the measurements are able to closely match the increment distributions and the structure functions of the measurements, but this often does not hold for surrogates that only mimic the power spectrum of the measurement. However, even the best performing surrogates do not have asymmetric increment distributions, i.e., they cannot reproduce nonlinear dynamical processes that are asymmetric in time. Furthermore, we have found deviations of the structure functions on small scales.

1 Introduction

Part of the beauty of nature lies in its mix of predictability and surprise. Such a pattern challenges us to discover regularities. This beauty is found in the structure of clouds, the cycles and the novelty of the weather, the deltas of rivers and mountain ranges. What the mentioned complex geophysical systems have in common, is their variability on a large range of temporal and spatial scales.

As a full sampling of the entire field over all scales is often infeasible, statistical generation of geophysical fields is frequently needed. For their statistical description, one still needs measurements on a range of scales, but a much smaller sample is often sufficient. Such statistical fields can be used to study and parameterise processes which depend on structure, such as radiative transfer (Scheirer and Macke, 2001; Pincus et al., 2005) or groundwater flow (Molz et al., 2004). Statistical fields or time series can also be necessary when few or no similar measurements are available, or because multiple records cannot be measured, such as for solar activity or runoff records of large rivers. Furthermore, interesting geophysical structures with known (statistical) properties are needed to test various algorithms used in the geosciences, e.g. analysis and error-detection algorithms. Moreover, such fields and time series can be used for nonlinearity testing (Schreiber and Schmitz, 1996), estimation of confidence intervals within a bootstrap framework (Rust et al., 2006[1]), trend testing (Radziejewski, 2000), or as boundary data for dynamical models.

Currently, the dominant framework for statistical modelling of structures is the multifractal paradigm using scale-free models (Mandelbrot, 1982; Turcotte, 1997). These models assume that the (statistical) properties of structures have a power law dependence on scale. For example, the variance of the wind or the kinetic energy of turbulence (E_k) in fractal models is related to its wavenumber (k) as $E_k \sim k^b$, where b is the scaling exponent (Frisch, 1995). An important contribution of fractal modelling was that it sensitised the geophysical sciences to the existence of structure on all scales. Before this era, many models assumed the existence of one or a few dominant scales (Mandelbrot, 1982).

Correspondence to: V. Venema
(victor.venema@uni-bonn.de)

[1]Rust, H., Kallache, M., Kropp, J., and Schellnhuber, H.-J.: Confidence intervals for return level estimation using a bootstrap approach, in preparation, 2006.

Even if the fractal approximation is able to describe a large part of the structures found in nature, important deviations do exist. Scale breaks, i.e. scales at which the scaling exponent changes, have been observed and have to exist as the earth is finite. The total power of a signal with a power law power spectrum without a scale break would diverge, for the exponent $b \geq -1$ at small scales, and for $b < -1$ at large scales.

Other important non-fractal patterns are waves and cycles. Gravity waves in the atmosphere can be revealed by wave patterns in clouds. Important cycles are the diurnal and annual cycle, and the El Niño/Southern Oscillation and the North Atlantic Oscillation. Such cycles are normally removed before fractal analysis.

Next to these two main classes, there are many more examples of non-fractal structures in the geosciences. As a result, a new paradigm is developing for the statistical modelling of geophysical fields, for example, in wind engineering (Masters and Gurley, 2003), hydrology (Montanari 1997; Lohre et al., 2003; Kallache et al., 2005), earth sciences and soil engineering (Christakos, 1992; Molz et al., 2004), and in the cloud sciences (Evans and Wiscombe, 2004; Scheirer and Schmidt, 2005; Venema et al., 2006a). In this approach to statistical modelling, the presence of variability at all scales is taken for granted, but the fractal approximation is not employed. Instead, especially in the examples from the cloud sciences, one tries to stay as close as possible to the measured structure, for instance, by using the measured spectrum instead of a power law fit to the spectrum.

In this new approach, one is normally not so much interested in the average structure of all fields, but rather in a statistical reconstruction or replication of a certain field, which cannot be measured fully or repeatedly. The reconstructed field is thus a surrogate for the measured field. Therefore, this approach is called the surrogate data paradigm.

The word paradigm (Kuhn, 1963) is appropriate in the sense that the fractal and the surrogate communities have their own way of seeing the world, their own set of questions and problems, their own algorithms, and find it hard to communicate with each other. However, in Kuhn's "The structure of scientific revolutions", a new paradigm replaces the older one. If paradigms coexist, they only do so during a usually short revolution. Sometimes, one has to wait until the proponents of the old paradigm leave academia. The process of paradigm change from assuming a dominant scale to the multifractal paradigm is still ongoing. In case of the surrogate data framework, we expect that it can peacefully coexist with the fractal one. Depending on the scientific question, one can choose the most appropriate methodology, just as meteorologists have a range of different atmospheric models (intermediate complexity models, global and regional climate models, global and limited area weather models), from which one can choose the most appropriate one. It may be argued that the existence of multiple methodologies is typical for complex system sciences.

The statistical parameters used to generate surrogate fields are rather simple compared to multifractal modelling. Initially algorithms only used power spectra (autocorrelations) to describe the structures. State-of-the-art algorithms additionally use the measured distribution of values. In this paper, we want to show that the combination of these two statistics is surprisingly effective for generating structures. We do so by comparing empirical records and their surrogates with respect to more sophisticated statistical measures such as structure functions, and increment distributions, which are used in the multifractal community, and annual statistics and return values, estimated using extreme value theory.

In Sect. 2, we shortly review the various algorithms to generate surrogate time series from these measurements. In Sect. 3, we introduce the meteorological and hydrological measurements we have used. The ways that we compared the surrogates with the measurements and the statistical properties we calculate are discussed in Sect. 4. The results of the comparison are presented in Sect. 5. The paper finishes with discussions, conclusions and an outlook.

2   Generators

The name surrogate time series and many of the algorithms originate from the nonlinear dynamics community. They are utilised for statistical tests, where the surrogate time series with well known statistical properties represent the null hypothesis.

In nonlinearity tests, the surrogates represent a linear dynamical process and are utilised to test whether a measurement can be assumed to originate from a nonlinear system (Theiler et al., 1992; Theiler and Prichard, 1996; Kugiumtzis, 1999; Kantz and Schreiber, 1999).

2.1   PDF surrogates

A simple, but useful, surrogate type is the PDF (probability density function) surrogate (Theiler et al., 1992). PDF surrogates have the same value distribution as the measurement, but on average no autocorrelations. Random indices can be produced by sorting white noise while keeping track of the indices. These random indices can then be used to randomly permute the measured values.

2.2   Fourier surrogates

Another simple surrogate time series is the Fourier surrogate (Theiler et al., 1992). The only stationary solutions of a linear dynamical system are cosines and a constant. As the Fourier transform constitutes a decomposition of a signal into cosine functions, the magnitudes of the Fourier spectrum completely describe a stationary linear system. The difference between a linear and a nonlinear stationary signal must thus lie in their Fourier phases.

These surrogates have the same Fourier spectrum as the measurements they are based on, but random Fourier phases. The Fourier generator computes the complex Fourier coefficients ($re^{i\varphi}$) of these surrogates by combining the Fourier phases (φ) of white noise with the magnitudes (r) of the Fourier coefficients of the measurement. By randomizing the phases, all nonlinearity is destroyed. Fourier surrogates are thus used to represent a linear dynamical system. A Fourier surrogate with a sufficient number of significant Fourier components will, by the central limit theorem, have a Gaussian distribution.

2.3 AAFT surrogates

It is possible that the measurement is nonlinear due to a static nonlinearity. Thus it may be that a record with a non-Gaussian distribution does not originate from a nonlinear dynamical system, but rather from a linear system which is measured by a static nonlinear measurement function. As the nonlinear dynamics community is not interested in static nonlinearities, they have developed a number of algorithms that include both the Fourier spectrum (for the linear dynamics) as well as the distribution (for the static nonlinearity). We will only describe the ones used in this study.

The first algorithm to generate such signals was the Amplitude Adjusted Fourier Transform (AAFT) algorithm (Theiler et al., 1992). This algorithm is similar to the Fourier method. First, the measured time series is converted to a time series with a Gaussian distribution, then the Fourier method is applied and finally the Gaussian distribution of the surrogate is converted back to the measured distribution. This computationally efficient algorithm has two weaknesses. Firstly, the nonlinear static function used for the conversions needs to be invertible. This is problematic, e.g., in the case of rain and cumulus measurements, as they have many zero values. Secondly, the spectrum of the surrogates is not the same as the one of the measurement, but has a bias towards a white noise spectrum (Schreiber and Schmitz, 2000).

2.4 IAAFT surrogates

As an alternative, Schreiber and Schmitz (1996, 2000) developed an iterative AAFT (IAAFT) algorithm. It works with the spectrum of the measurement, not with the spectrum of its Gaussian counterpart as the AAFT algorithm does. In the first step, it adjusts the Fourier spectrum in the same way as the Fourier generator above, i.e. it adjusts the magnitudes of the Fourier coefficients, but keeps the phases intact. In the second step, it adjusts the distribution to the measured one, thus changing the spectrum somewhat. Therefore, one has to iterate to get high-quality surrogates. If one stops iterating after the second step, i.e. after adjusting the distribution, one has a surrogate with exactly the same values as the measurement and almost the same power spectrum.

In the engineering community, the IAAFT algorithm was discovered independently (Masters and Gurley, 2003) to simulate wind pressure fields and offshore waves. Masters and Gurley compared the IAAFT algorithm to older methods used in the engineering community and found the IAAFT algorithm to be more accurate.

2.5 SIAAFT surrogates

Unfortunately, the IAAFT algorithm normally gets trapped in a local minimum after which it no longer converges. In the Stochastic IAAFT (SIAAFT), developed by Venema et al. (2006b), the adjustment of the distribution is made stochastically. As a consequence, the algorithm gets stuck in local minima less easily, and achieves a better approximation to the Fourier spectrum. Otherwise, the algorithm is the same as the IAAFT algorithm.

2.6 FARIMA surrogates

The last two algorithms used here involve a parametric approach to represent the spectral properties of the measurements as a family of Gaussian stochastic processes. Those processes play a prevailing role in the linear time series analysis and are known as fractional integrated autoregressive moving average (FARIMA) processes (Brockwell and Davis, 1991; Beran, 1994). FARIMA processes are frequently used also in geophysical contexts to describe complex signals in cases where models based on physical processes are not feasible or explicitly not asked for. Besides being very flexible in reproducing the autocorrelation function (ACF), FARIMA processes are able to reproduce the Hurst phenomenon (Hurst, 1951) observed in certain types of geophysical time series (Montanari, 1997; Koutsoyiannis, 2002). A record is said to exhibit the Hurst phenomenon if it shows a slow (algebraic) decay of the autocorrelation function for large lags.

We estimate the fractional differencing parameter d and the autoregressive $\phi_1, \phi_2, ... \phi_p$ and moving-average parameters $\psi_1, \psi_2, ..., \psi_q$ of a FARIMA$[p, d, q]$ processes using the Whittle approximation to the maximum likelihood (ML) estimator, which is numerically more efficient than the exact ML estimator (Haslett and Raftery, 1989) while retaining the asymptotic distribution. A detailed description of FARIMA processes and the associated parameter estimation can be found in Beran (1994).

The orders of the autoregressive (p) and moving average component (q) are a priori unknown and can be inferred using the Hannan-Quinn Information Criterion (HIC) which is advocated for FARIMA$[p, d, q]$ processes (Hannan and Quinn, 1979; Beran et al., 1998; Bisaglia, 2002):

$$\text{HIC} = N \log \hat{\sigma}_\eta^2 + 2c \log \log N(p + q + 1), \tag{1}$$

Table 1. Parameters of the FARIMA processes of the various signals. A dash for the log-transform constant (α) means that no log transform was performed.

	α	Process	Parameters
p-model	1.39	[1,d,1]	d=0.38±0.01, ϕ_1=−0.20±0.09, ψ_1=0.27±0.09
Rain Potsdam	1.82	[2,0,1]	ϕ_1=0.47±0.03, ϕ_2=−0.13±0.01, ψ_1=0.74±0.03
Runoff Burghausen	0	[2,d,2]	d=0.30±0.02, ϕ_1=1.19±0.04, ϕ_2=−0.28±0.02, ψ_1=−0.56±0.02, ψ_2=−0.16±0.01
Runoff Cologne	0	[3,d,1]	d=0.41±0.02, ϕ_1=1.46±0.07, ϕ_2=−0.68±0.08, ϕ_3=0.12±0.03, ψ_1=−0.15±0.06
Cumulus	$5 \cdot 10^{-4}$	[1,d,0]	d=0.20±0.07 and ϕ_1=0.63±0.07
Stratocumulus	–	[1,d,1]	d=0.62±0.04, ϕ_1=−0.24±0.13, ψ_1=0.51±0.10
Temperature	–	[3,d,0]	d=0.14±0.01, ϕ_1=0.74±0.02, ϕ_2=−0.15±0.01, ϕ_3=0.06±0.01

with $c>1$ and $\hat{\sigma}_\eta^2$ the ML estimate of the variance of the driving noise σ_η^2. We choose the orders p and q such that the HIC with c=1.0001 takes a minimum.

There are several algorithms to simulate data from a FARIMA process (Bardet et al., 2003). Here, we use a method based on the inverse Fourier transform described in Timmer and König (1995) for simulating self-similar processes which can be straightforwardly extended to FARIMA processes. A set of routines written in the computer language R with the algorithms for FARIMA parameter estimation (based on the code from Beran, 1994), model selection (HIC) and the simulation can be obtained from Rust (2006).

Since a FARIMA process is a Gaussian process, a time series simulated from it has a normal marginal distribution unlike most of the empirical geophysical records. Thus, it is convenient to transform the observed record to get closer to a normal distribution before the parameter estimation and to transform it back afterwards. For many records, a log-transform

$$y = \log(x + \alpha) \qquad (2)$$

turned out to be convenient, with α being a positive constant to avoid problems with negative values. This transformation is applied to all data sets studied except the Liquid Water Content (LWC) of the stratocumulus and the Potsdam temperature series. The parameter α is chosen in a way that, for the transformed series, the Kolmogorov-Smirnov statistic (maximum deviation of the empirical cumulative distribution function from a normal distribution) is minimized. The parameters of the FARIMA processes for the various records are listed in Table 1. Additionally, we removed the seasonal cycle after the log transform for the runoff and temperature records by subtracting the mean of, for example, 1 January over all years from all data for each 1 January. Further, we divide by its standard deviation to ensure a homogeneous variance throughout the year. This transformation is applied in reverse for the surrogate records generated with the FARIMA model.

2.7 FARIMA + IAAFT surrogates

As an extension to the conventional FARIMA modelling, we suggest adjusting the value distribution of a simulated series according to the observed record using the IAAFT algorithm. The IAAFT procedure is analogous to the one described before, only instead of adjusting the periodogram of the surrogate to the one from the empirical series, we adjust to the periodogram of a FARIMA realisation. Hence, for every surrogate we use the same value distribution, while the periodogram is different for every surrogate and stems from a realization of the chosen FARIMA process. This type of surrogate is denoted with FARIMA + IAAFT, while surrogates from the standard method are referred to as FARIMA.

2.8 Algorithm groups

To facilitate the presentation of the results, we group the above methods in the following as:

– Fourier-based surrogates: Fourier, AAFT, IAAFT, SIAAFT;

– FARIMA-based: FARIMA, FARIMA + IAAFT;

– Two-statistic surrogates: AAFT, IAAFT, SIAAFT, FARIMA + IAAFT.

3 Measurements and surrogates

As input, we use historical data, cloud measurements and a realization of a theoretical multifractal. In Fig. 1, the structure of all seven time series is illustrated by displaying a typical interval. The historical data are the time series of daily mean temperature and daily precipitation sums from a station in Germany, and the daily river runoff from a large and a small catchment. The high-resolution cloud liquid water measurements were obtained from a cumulus and a stratocumulus field. The realization of the theoretical multifractal originates from the p-model; see Sect. 3.1.
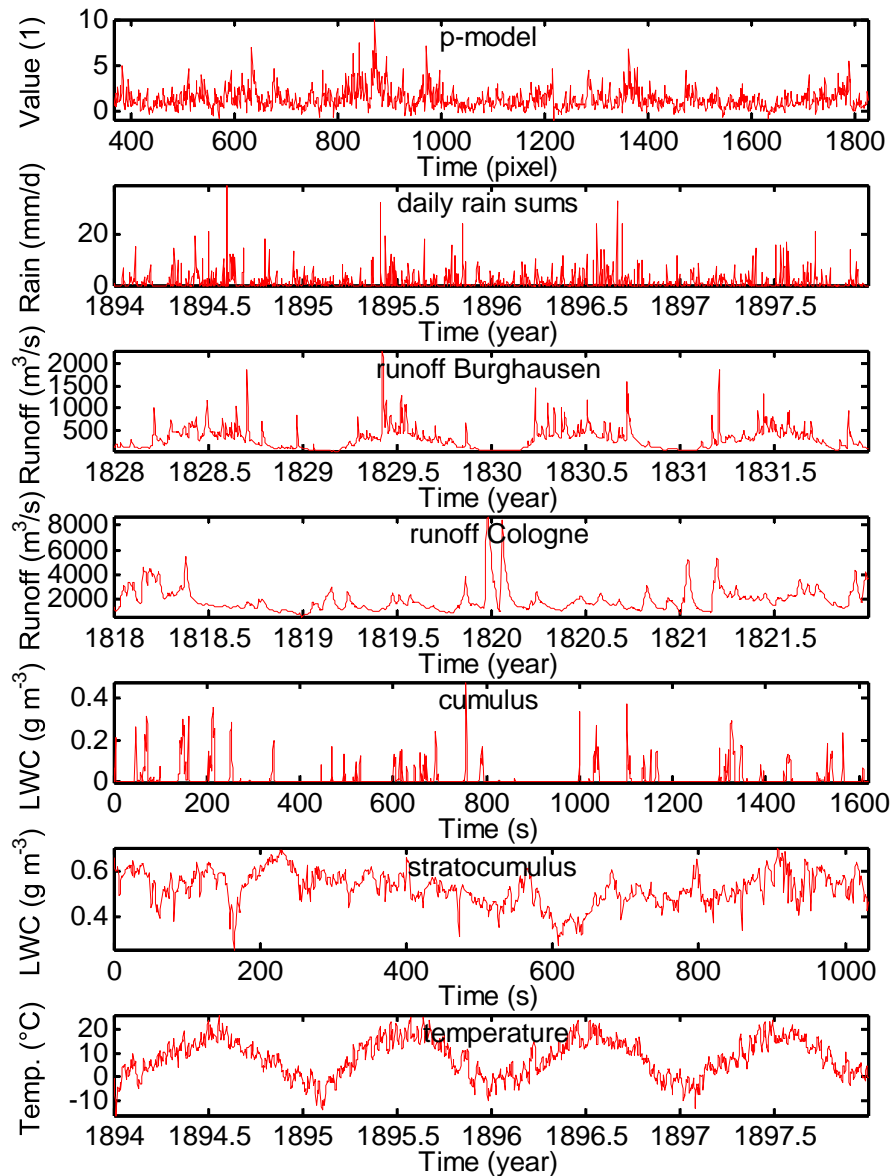
Fig. 1. The structure of the measurements we have used as input. Except for the cloud measurements, only a small part of the time series is shown, as otherwise the interesting small scale structure would not be visible.

These time series were selected because they contain a variety of features that are also typical for other geophysical time series. The temperature measurements have a strong annual cycle, the runoff measurements a weaker one. Furthermore, the dynamics of the river runoff is asymmetric in time, with fast swells and slower decreases. The rain and cumulus time series display burst-like intermittency. The stratocumulus measurement has a nearly fractal structure. The multifractal was chosen as an idealised geophysical model with strong intermittency, and because multifractal modelling is often applied in the geosciences. Furthermore, these time series were chosen for their length. Long records improve the performance of the structure function analysis.

3.1 P-model

Apart from empirical records, we study a time series simulated with the fractional integrated multifractal p-model (Davis et al., 1997; Wilson et al., 1991), which is a simple Fourier filtered multiplicative cascade model. With the parameter values $p=0.6$ (where this p is different to the one used in FARIMA modelling) and spectral exponent $s=-0.8$, we generate a time series with 32 768 values. We will treat this time series as one of the historical measurements, i.e. a record of 90 years of daily values of an arbitrary variable. However, when we refer to "real measurements", the realisation of the p-model is not included.

3.2   Daily rain sums and mean daily temperature

Records of precipitation and temperature data from the long-term meteorological station in Potsdam were acquired freely from the German weather service; this data was measured daily from 1893 to 2003 and has no gaps (DWD, 2006). The daily records have not been examined for homogeneity, but the observations at Potsdam are believed to be of good quality, in part due to the fact that the station has never been moved (Lehmann, 2001).

The precipitation time series is characterised by strong burst-like intermittency: about 51% of the values are zero due to days without significant rainfall. In order to benefit from the efficient fast Fourier transform (FFT), the temperature and precipitation time series were truncated to 40 500 values. Since the temperature data show strong seasonality, it is convenient to remove this periodic cycle prior to FARIMA modelling as described in Sect. 2.

3.3   Runoff

The runoff data were provided by the Global Runoff Data Centre (GRDC, 2006); we have chosen daily records from two gauge stations of catchments with very different sizes: (1) station Cologne on the river Rhine in the western part of Germany with a catchment size of $144\,232\,km^2$ and (2) station Burghausen on the river Salzach in southern Germany with a catchment area of $6649\,km^2$. The runoff time series span the years 1817 to 2001 for Cologne and 1827 to 1991 for Burghausen. These records have also been selected because of their completeness.

In order to be able to benefit from the efficient fast Fourier transform (FFT), the runoff time series at Cologne was truncated to 67 500 values. The runoff time series at Burghausen was extended by adding 100 average values at the end to reach a total of 60 000 datapoint.

3.4   Cloud measurements

The two cloud measurements are taken from the INSPEC-TRO measurement campaign performed on the east coast of the UK in 2002 (Kylling et al., 2005). The liquid water content (in $g\,m^{-3}$) measurements were obtained using a Particle Volume Measurement (PVM) probe. This optical laser cloud probe has an integration time of 1 s. It was mounted on a small aircraft that flew through the clouds.

We have selected long tracks during which the airplane flew horizontally in about the middle of the clouds. The stratocumulus case was measured on 14 September and contains 1024 values. The cumulus cloud was measured on the 22nd and has 1600 data points. The differencing parameter d is larger than 0.5 for the stratocumulus measurement and thus not in the stationary domain, implying that not the record itself, but the increment record can be modelled with a stationary FARIMA process.

4   Methodology

We want to demonstrate the usefulness of surrogate techniques in geophysical modelling by discussing several examples. Because surrogates do not reproduce all thinkable characteristics of a time series or a field, the appropriateness and accuracy of surrogate time series and fields (just as multifractal ones) has to be ascertained. This can only be shown with respect to a specific type of application. For example, Venema et al. (2006a) previously compared the optical properties of 3-D model liquid water clouds and their surrogates and found that they matched each other within the statistical uncertainty. This shows that surrogate clouds are likely to be useful to study the relationship between cloud structure and radiative transfer through clouds. In this study, we do not have such an application. Instead, we have chosen to investigate the statistical properties of the surrogates utilising a number of comprehensive statistical measures. The surprisingly good results for these statistics for a range of different kinds of data will hopefully stimulate other researchers working on structure to try to use surrogate techniques to solve their problems. As statistical measures, we have chosen the power spectrum, the increment distribution, structure functions, annual statistics (e.g. percentiles) and long-year return values.

The increment distributions and structure functions are of special interest for scientists working within the multifractal paradigm, as these scientists often view algorithms based on Fourier methods as monofractal. We will demonstrate that surrogates that allow for non-Gaussian distributions have a much richer structure than the Gaussian "monofractal" ones. In addition, these surrogates have the advantage that the fractal approximation or assumption is not needed.

For an honest comparison of the seven generators, we first generated the surrogates and then analysed their results; thus the algorithms were not tuned to the records to yield good results. To get an idea on the variability of the results, we generated a small ensemble of 10 surrogates for each measurement.

4.1   Power spectra

In the following, the measured time series is denoted by the vector $\{m_n\}$, with the index n={0,1,...,N-1}, where N is the number of values of the measurement. The Fourier coefficients (M_k) are calculated as:

$$M_k = \sum_{n=0}^{N-1} m_n e^{i2\pi kn/N}, \, k = \left[0, ..., \frac{N-1}{2}\right], \qquad (3)$$

where the [x] denotes the integer part of x. Squaring the Fourier coefficients yields the periodogram, which represents the variance of the time series on the scale corresponding to the wavenumber k. Since this estimate of the power spectrum is very noisy, we use octave binning in the figures (Davis et al., 1999). In octave binning, the mean variance (power) and

the mean wavenumber is calculated for eight bins per order of magnitude.

Because periodograms are noisy estimates, a direct comparison is not very useful. Instead, the difference between the periodograms of the original and surrogate time series is represented by the maximal deviation D of their cumulative periodograms P_j:

$$P_j = \sum_{k=0}^{j} M_k, \quad j = \left[0, .., \frac{N-1}{2}\right] \tag{4}$$

These cumulative periodograms are normalised to run from zero to unity to obtain a curve similar to an empirical cumulative distribution function. Analogue to the Kolmogorov-Smirnov (KS) distance, which is used to test whether two distributions are equal, we then calculate the maximum distance D between the two normalised cumulative periodograms:

$$D = \max_{j} \left| P_{\text{surrogate}}(j) - P_{\text{original}}(j) \right| \tag{5}$$

4.2 Increment distributions

A simple but very comprehensive statistic which is much used in turbulence research is the increment distribution. For lag (l), the increments (ρ_i) are given by:

$$\rho_i = m_{i+l} - m_i, \quad \forall \quad i \in \{0, 1, \ldots, N-l-1\}, \tag{6}$$

The increment distribution can, for example, reveal whether the variability at a certain scale is homogeneously distributed or caused by a few large increments (jumps) and many smaller increments. Also, asymmetries in the dynamics (e.g. steep increases and slow decreases) are visible here.

4.3 Structure functions

To facilitate the study of the shape of the increment distribution as a function of its lag (scale), it is common in multifractal analysis to calculate the average moments (q) of the absolute increments as a function of the lag l:

$$\text{SF}(q, l) = \left\langle |\rho_i(l)|^q \right\rangle \tag{7}$$

where $\langle \cdot \rangle$ denotes an average. This approach is called structure function (SF) analysis. In multifractal analysis, one would try to fit a power law to a SF as a function of the scale l. Because the absolute values of the increments are used, the information on the asymmetry of the dynamics is lost. In much work on turbulence, a definition of the SF is used without the modulus operator, so that asymmetries can be studied.

The second order SF (i.e. q=2) is equivalent to the autocorrelation function and thus also equivalent to the power spectrum. As a consequence, the surrogates (except the PDF surrogates) should reproduce the second order SF and the variance of the increment distribution well.

As a measure of how accurately the surrogates reproduce the structure functions, we compare the root mean square (RMS) relative difference of the octave binned 4th order coefficients of the data (M_k') and of the surrogate (S_k'):

$$\Delta \text{SF} = \sqrt{\frac{1}{N} \sum_k \left(M_k' - S_k'\right)^2 \Big/ M_k'^2} \tag{8}$$

This error measure was chosen to give equal weight to every order of magnitude. At large scales, the structure functions become noisy due to the finite length of the time series, and even the p-model shows significant deviations from a fractal spectrum. As a consequence we did not calculate the SF for the largest two orders of magnitude in the case of the cloud measurements, and the largest three orders of magnitude for the other time series.

4.4 Annual statistics

Some basic statistical measures of the annual distribution are calculated. Their distributions are compared between the original and surrogate time series. For temperature, runoff, and the p-model, the selected statistics are: mean value, maximum value, and the 5th and 95th percentile values. Those have been chosen in order to capture both the average and the extreme characteristics of the time series. Since precipitation data exhibit very special characteristics (e.g. limited to positive values, occurrence of zero values), different statistics were chosen. They are based on the definition of a "rain day" as a day where the precipitation sum is equal to or larger than 1 mm. These statistics are: number of rain days, mean intensity (= average precipitation on rainy days), maximum rainfall, and the 95th percentile value. The cloud measurements are not considered here because of their different temporal range. The distributions of the annual statistics between surrogates and original time series are compared by means of the KS statistic D, as described before.

4.5 Return values

Return values are calculated using extreme value statistics: a Generalised Extreme Value (GEV) distribution (Coles, 2001) with distribution function

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{9}$$

with parameters ξ, μ and σ, which are fitted to the annual maxima of all measurements and their surrogates using the EVIM package for Matlab (Gençai et al., 2001). Return values are then calculated for return periods ranging from 2 to 100 years. The deviations of the 50-year return levels (denoted by RV50) are examined more closely in Sect. 5.

Table 2. The difference in the estimates of the power spectrum as calculated by the Kolmogorov-Smirnov distance D; see Eq. (5). For each surrogate method, the quality of the approximation of the power spectrum is quantified by calculating the difference D between the average cumulative periodogram of the ensemble and the original periodogram (top). We assess the variability of the ensemble by calculating the mean (absolute) value from the individual distances D for each ensemble member (bottom).

	SIAAFT	IAAFT	AAFT	Fourier	PDF	FARIMA	FARIMA + IAAFT
Bias							
P-model	$1.3\,10^{-05}$	$5.8\,10^{-05}$	$5.4\,10^{-02}$	–	$4.4\,10^{-01}$	$2.0\,10^{-02}$	$2.4\,10^{-02}$
Rain Potsdam	$2.6\,10^{-03}$	$6.2\,10^{-03}$	$6.2\,10^{-02}$	–	$1.0\,10^{-01}$	$3.4\,10^{-02}$	$2.5\,10^{-02}$
Runoff Burghausen	$2.0\,10^{-04}$	$1.8\,10^{-03}$	$1.2\,10^{-01}$	–	$6.8\,10^{-01}$	$1.3\,10^{-01}$	$1.3\,10^{-01}$
Runoff Cologne	$4.8\,10^{-05}$	$3.4\,10^{-04}$	$9.0\,10^{-02}$	–	$8.1\,10^{-01}$	$1.8\,10^{-01}$	$1.8\,10^{-01}$
Cumulus	$1.0\,10^{-02}$	$1.9\,10^{-02}$	$1.3\,10^{-01}$	–	$5.2\,10^{-01}$	$7.9\,10^{-02}$	$1.1\,10^{-01}$
Stratocumulus	$9.8\,10^{-05}$	$8.1\,10^{-04}$	$1.6\,10^{-02}$	–	$7.2\,10^{-01}$	$1.8\,10^{-01}$	$1.8\,10^{-01}$
Temperature	$5.6\,10^{-05}$	$7.1\,10^{-04}$	$4.1\,10^{-02}$	–	$8.0\,10^{-01}$	$1.0\,10^{-02}$	$1.1\,10^{-02}$
Variabilty							
P-model	$1.3\,10^{-05}$	$5.8\,10^{-05}$	$5.5\,10^{-02}$	–	$4.4\,10^{-01}$	$4.0\,10^{-02}$	$4.1\,10^{-02}$
Rain Potsdam	$2.7\,10^{-03}$	$6.4\,10^{-03}$	$6.2\,10^{-02}$	–	$1.0\,10^{-01}$	$3.5\,10^{-02}$	$2.6\,10^{-02}$
Runoff Burghausen	$2.0\,10^{-04}$	$1.8\,10^{-03}$	$1.2\,10^{-01}$	–	$6.8\,10^{-01}$	$1.4\,10^{-01}$	$1.3\,10^{-01}$
Runoff Cologne	$5.2\,10^{-05}$	$3.5\,10^{-04}$	$9.2\,10^{-02}$	–	$8.1\,10^{-01}$	$1.8\,10^{-01}$	$1.8\,10^{-01}$
Cumulus	$1.3\,10^{-02}$	$2.4\,10^{-02}$	$1.5\,10^{-01}$	–	$5.2\,10^{-01}$	$9.9\,10^{-02}$	$1.2\,10^{-01}$
Stratocumulus	$1.4\,10^{-04}$	$9.5\,10^{-04}$	$1.9\,10^{-02}$	–	$7.2\,10^{-01}$	$2.1\,10^{-01}$	$2.1\,10^{-01}$
Temperature	$5.7\,10^{-05}$	$7.2\,10^{-04}$	$4.1\,10^{-02}$	–	$8.0\,10^{-01}$	$1.1\,10^{-02}$	$1.1\,10^{-02}$

5 Results

In this section, we will show how well the seven types of surrogates are able to reproduce the original time series. An example of the cumulus measurement and of its surrogates is shown in Fig. 2. Due to the large number of observations, types of surrogates and the many statistical measures, we cannot display all results. Instead, we will only show a number of figures that illustrate the most striking features. A comprehensive overview with automatically generated quicklook plots can be found in Venema (2006).

5.1 Power spectra

As all surrogates (except the PDF surrogates) try to reproduce spectral properties, we first look at the power spectra. Table 2 lists the maximum deviation D between the normalised cumulative periodograms of the surrogates and the measurements. Fourier based surrogates and FARIMA surrogates differ because the former generate a time series with exactly the same Fourier coefficients while the latter estimate the underlying – much smoother – power spectrum and based on this spectrum generate a noisy periodogram.

It can be seen that a large part of the difference of the FARIMA surrogates is due to bias rather than variability in the ensemble. This bias error is partly due to the log-transform that is applied to most records before FARIMA modelling. The log-transformed data are not sufficiently normally distributed. The time series generated from a FARIMA process are normally distributed sharing the variance of the log-transformed original record. However, after the inverse transform with an exponential function, the variances in the exponentially transformed time series and the original record are not the same.

The inclusion of the value distribution in the FARIMA scheme resulted in only a minimal loss of spectral accuracy. The SIAAFT method is more accurate in reproducing the periodogram than the IAAFT method which is again more accurate than the AAFT algorithm.

The power spectra of the p-model and its AAFT and SIAAFT surrogates are shown in Fig. 3. It can be seen that the SIAAFT surrogates reproduce the spectrum very accurately, whereas the slope of the AAFT surrogate is too flat, i.e. it has too much variance at small scales and too little at large scales. This is typical for AAFT surrogates and is known as the white noise bias (Schreiber and Schmitz, 2000).

Surprisingly, the spectra of the AAFT surrogates of the rain measurement and the runoff measurement in Burghausen (Fig. 4) have too little variance on small scales, contrary to the white noise bias. We speculate that, for rain, this deviation is caused by the conversion of the measurement data to a time series with a Gaussian distribution. In this step, the sorting algorithm keeps the original order in case values are exactly the same. In the rain measurement many equal values occur due to the many days without (significant) rain, i.e. 0 mm/d (51%). As a consequence, the rain-free days at the beginning of the time series get lower values in the Gaussian time series than rain-free days at the end of the series.
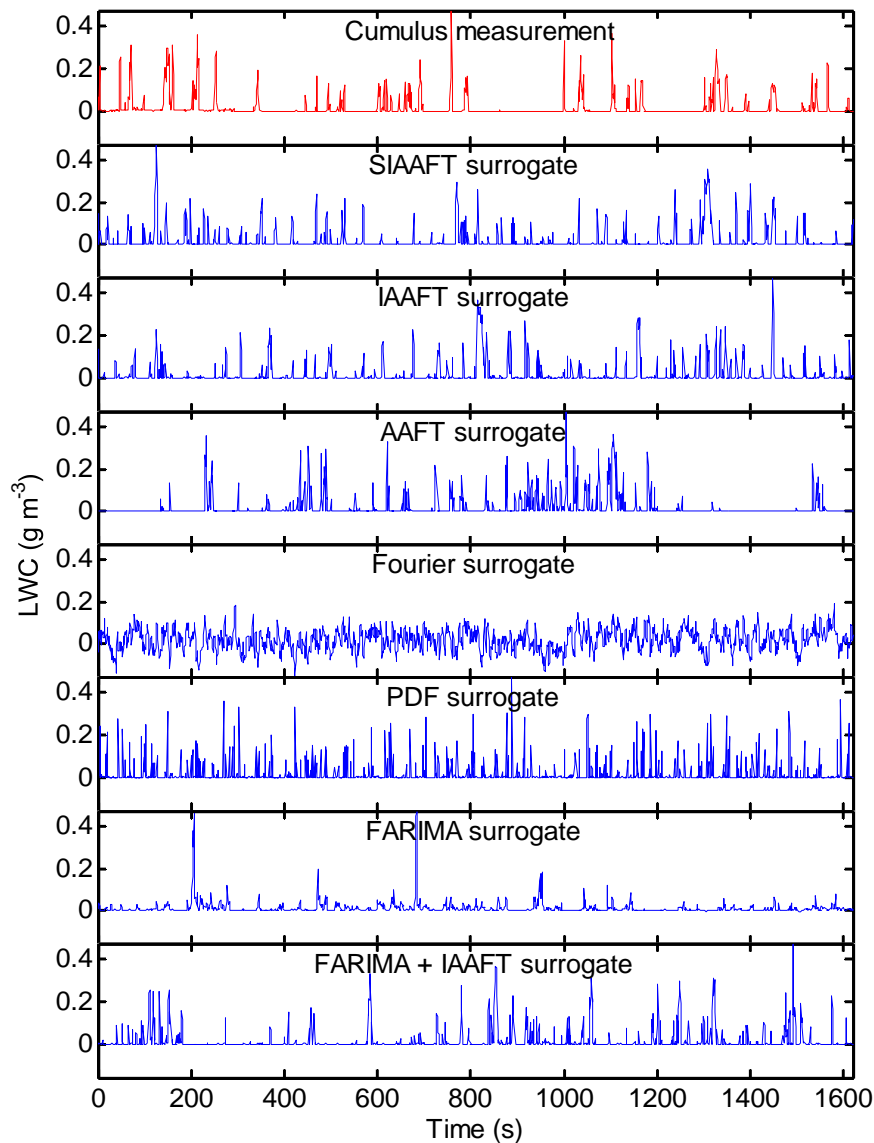
Fig. 2. A measurement of the liquid water content (LWC) of a cumulus cloud field and its seven surrogates.

This leads to a positive trend in this Gaussian time series and thus to enhanced variability on large scales, which consequently results in a lack of variance at small scales as the total variance is fixed. This idea is confirmed by the fact that the problem disappears when a little bit of noise is added, to make all values unique and spread the low values evenly over the Gaussian time series.

Adding some noise to the runoff measurement at Burghausen did not solve its red noise bias, however. This runoff measurement is also strongly discretised and as a consequence, every value occurs on average 44 times, but different from the rain measurement, there is not one value that dominates the distribution. We could not find a conclusive reason for this red noise bias.

5.2 Increment distributions

Figure 5 shows the increment distribution for a lag of one day of the temperature measurement and all its surrogates except for the PDF surrogate. The latter deviated too much to be displayed in the same plot as the others. The temperature increment distribution is slightly asymmetric. Temperature jumps between 5 and 10°C per day are more likely to go down than up. This is probably related to frontal activity. This asymmetry is not found in any of the surrogates.

The increment distribution of the Fourier surrogates is nearly Gaussian, while the increment distribution of the measurement shows a somewhat thicker tail. This tail is well reproduced by all surrogates, except the PDF and the Fourier surrogates. The number of small temperature jumps of less
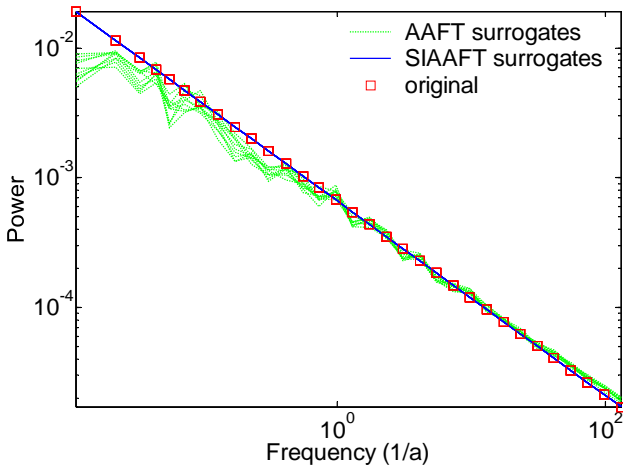
Fig. 3. The power spectra estimated by octave binning of the periodogram of the p-model and its ten AAFT surrogates and ten SIAAFT surrogates (which have no visible scatter). To enhance clarity, the values of the original series are plotted as squares, while the results of the surrogates are connected with lines.
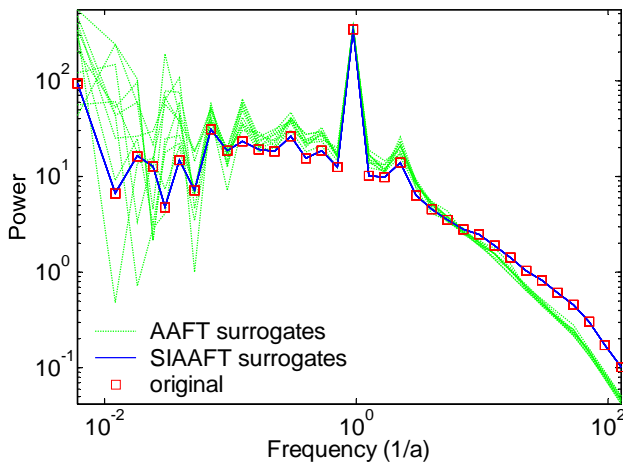

Fig. 4. The binned power spectra of the runoff at Burghausen and its ten AAFT and ten IAAFT surrogates.

than $2°C$ is underestimated by the surrogates (see inset of Fig. 5). The SIAAFT surrogates are the best in this respect. In general, the original increment distribution is best matched by the IAAFT, SIAAFT and FARIMA + IAAFT surrogates.

The asymmetry in the tail of the increment distribution is clearly visible in the runoff measurement of the Rhine (Fig. 6). Increments larger than $500\,m^3\,s^{-1}$ are more likely to be upward than downward. Increments of $+2000\,m^3\,s^{-1}$ are 35 times more likely than increments of $-2000\,m^3\,s^{-1}$. Because the mean increment has to be around zero, these large upward jumps result in a negative ($-20\,m^3\,s^{-1}$) most likely runoff increment, i.e., most days the water table of the river sinks, but when it rises it does so strongly. The increment
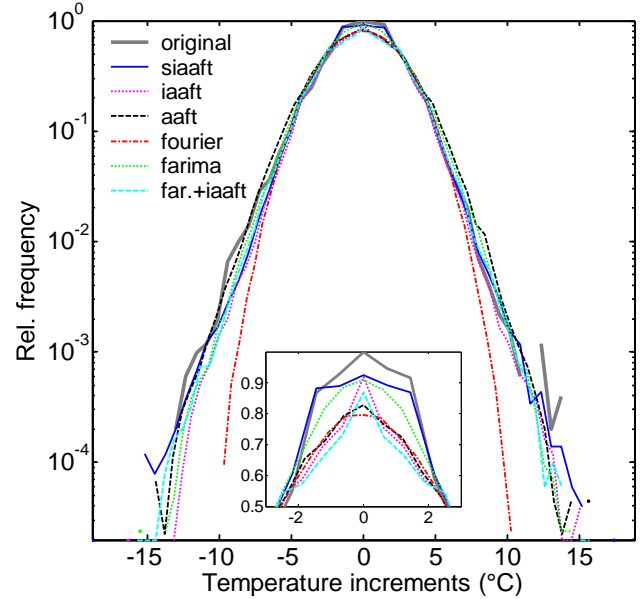

Fig. 5. Histograms of the increment distribution of the temperature measurement in Potsdam and its surrogates for a lag of one day. The increments have bins of $1°C$, and the histograms are normalised on the maximum number of bins of the measured temperature. The inset is an enlargement of the top of the distribution.
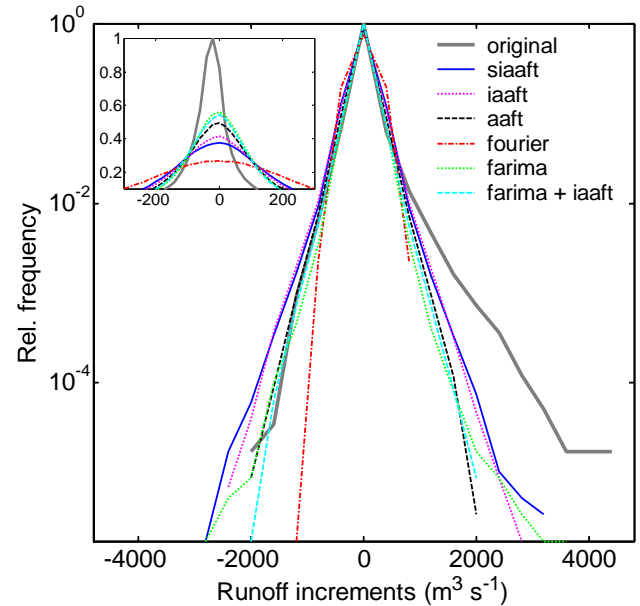

Fig. 6. Histograms of the increment distribution for the runoff of the Rhine at Cologne and its surrogates. For the main figure, the runoff has broad bins of $400\,m^3\,s^{-1}$, to decrease the noise for the rare jumps in the tails; the inset has fine bins of $20\,m^3\,s^{-1}$, to show the detailed distribution for the smallest increments. The histograms are normalised by the maximum number of bins of the measured runoff.

Table 3. The moments of the increment distribution for a lag of one day for the runoff records at Burghausen (Salzach) and Cologne (Rhine) and their surrogates.

	Record	SIAAFT	IAAFT	AAFT	Fourier	PDF	FARIMA	FARIMA + IAAFT
Burghausen								
std. deviation	91.4	91.4	91.7	64.7	91.4	263	57.6	63.5
skewness	2.15	−0.012	−0.019	0.014	0.004	−0.006	0.026	0.163
kurtosis	49.5	40.0	46.9	28.8	−0.2	6.1	6.8	92.7
Cologne								
std. deviation	210	211	211	181	211	1481	153	165
skewness	3.22	0.011	0.000	0.013	0.000	−0.006	0.061	0.029
kurtosis	32.6	7.5	7.5	6.9	−0.1	3.7	10.5	8.2

distributions of the surrogates are symmetrical and do not have this thick tail for positive increments.

In Table 3, we computed the unbiased standard deviation, skewness and kurtosis of the increments of the runoff records and their surrogates. The measured increments are strongly positively skewed, indicating more upward jumps, whereas the increments of the surrogates are minimally skewed, indicating symmetrical distributions.

To study the behaviour of the increments as a function of lag time, the skewness of the increment distributions of the runoff in Cologne was plotted for every lag (Fig. 7). This clearly reveals the skewness of the measured increments and the symmetry of the increments of the IAAFT surrogates. All other surrogates look similar, expect for the FARIMA surrogate, which is able to model some skewness due to its seasonal cycle. Also the runoff at Burghausen and the temperature record are characterised by skewed increment distributions, which are not modelled by any surrogate, expect partially for the FARIMA ones.

The number of small increments is much lower in the surrogates and is best replicated by the FARIMA-based surrogates (see inset in Fig. 6). On average, the tails are best reproduced by the SIAAFT and IAAFT surrogates. The rain measurement has a very thick-tailed non-Gaussian increment distribution (not shown), which is well captured by all surrogates, except the Fourier and FARIMA surrogates.

5.3 Structure functions

For an overview of the results for the structure functions, we calculated the difference between the binned 4th order structure function (ΔSF; Eq. 8) of the measurement and the surrogates in Table 4. Remembering that the 2nd order structure function is equivalent to the power spectrum, the 4th order SF was chosen, as it puts more emphasis on the tail of increment distribution.

The last column of this table is the difference of the 4th order SF to its linear fit in a log-log plot; this represents a first estimate of the difference one could expect to achieve with
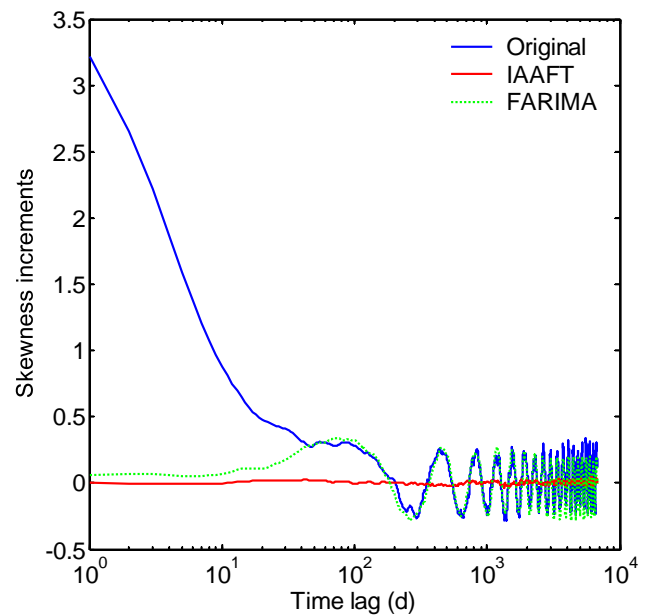

Fig. 7. The skewness of the increment distribution as a function of the lag of the increments.

a multifractal method (without subtracting cycles). For the fractal p-model and the rain records, this "multifractal generator" has the lowest deviation. The other measurements contain a stronger deviation from a power law fit and are more accurately reproduced by surrogate methods. The Fourier and the FARIMA surrogates have the largest deviations in the structure functions.

The structure functions of the LWC of the cumulus cloud are shown in Fig. 8 together with the mean SF over ten runs for PDF, Fourier and AAFT surrogates. The structure functions of the PDF surrogates are horizontal lines and match quite well at large scales where correlations are not strong, illustrating the link between the structure functions and the distribution. Fourier surrogates, instead, approximately reproduce the curved shape of the structure functions of the
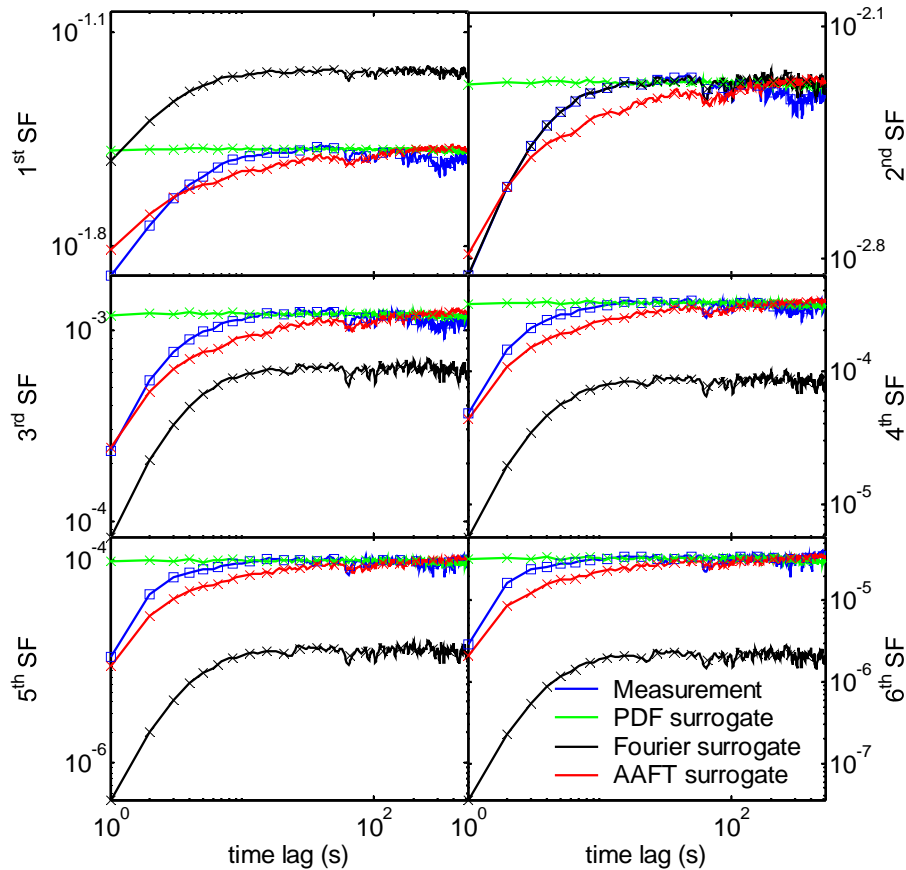
Fig. 8. The structure function spectra for the orders (q) 1 to 6 for the cumulus cloud measurement and its PDF, Fourier and AAFT surrogates. The values over octave bins are indicated by a square for the measurement or a cross for the surrogates.

measurement, although, as expected, they have an offset which gets larger with the moments deviating more from the 2nd moment of the structure function. The offset is negative for moments larger than the second and positive otherwise. The SF of the AAFT surrogates fit well at large scales, and have a similar curvature as the SF of the measurement.

Figure 9 allows a closer analysis of the SF for a number of surrogates by only displaying its 4th moment. It shows that the SIAAFT, IAAFT and FARIMA + IAAFT surrogates fit even better than the AAFT surrogate, this improvement is characteristic for all real measurements. The average SF of the FARIMA surrogates is a little closer to the original structure function than the SF of the Fourier surrogates (see Table 4); the log-transformation of the original and back-transformation of the simulated series already partially account for the non-Gaussian distribution. Systematic differences in the SIAAFT, IAAFT and FARIMA + IAAFT surrogates are mainly found at small scales, which is typical for all seven measurements.

The results achieved for the stratocumulus cloud (Fig. 10) are very similar, except that here, the AAFT surrogates are on par with their iterative counterparts. This measurement

shows a noticeable deviation from the FARIMA-based surrogates. Due to the parametric formulation, FARIMA-based surrogates favour a smooth spectrum and do not reproduce small fluctuations in the structure functions, such as the dip on scales around 100 s. The non-parametric Fourier-based surrogates are superior in this respect.

The structure functions of the runoff records are reproduced reasonably well by the IAAFT, SIAAFT and FARIMA + IAAFT surrogates (figure not shown). Fourier and FARIMA have more difficulties especially for the small catchment (Burghausen). Deviations are also found here on small scales, increasing with the SF moment. The seasonal cycle leads to periodic deviations at large scales which are comparably small for the large catchment (Cologne) but become relevant for Burghausen, the gauge exhibiting stronger seasonality.

5.4 Annual statistics

The annual statistics for the original records and the surrogates are compared in terms of their frequency distribution. The similarity of the distributions is assessed via the

Table 4. The difference (ΔSF) in the 4th order structure functions as calculated by the RMS relative difference over all lags; see Eq. (8). For each surrogate method, the quality of the approximation of the structure function is quantified by calculating the difference ΔSF between the average SF of the ensemble and the original SF (top). We assess the variability of the ensemble by calculating the mean value from the individual differences ΔSF for each ensemble member (bottom). The surrogate method which yields the smallest differences is indicated by a bold typeface. The last column shows ΔSF calculated for a linear fit in the log-log plot, which is the difference from an ideal multifractal model.

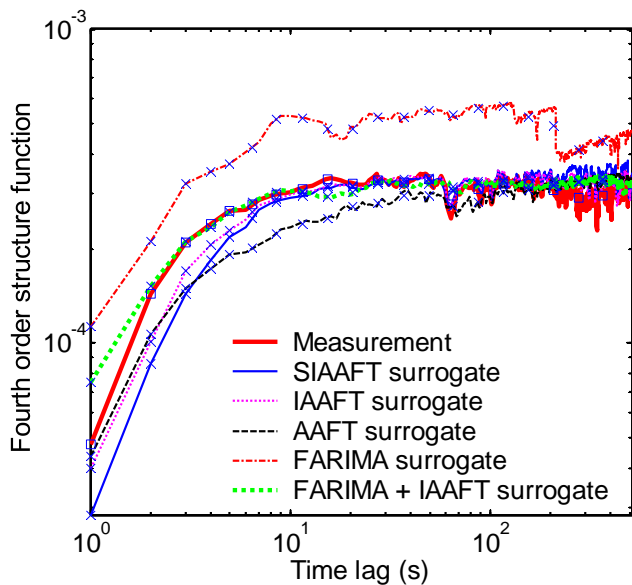|	SIAAFT	IAAFT	AAFT	Fourier	PDF	FARIMA	FARIMA + IAAFT	Multifractal
Bias								
P-model	0.018	0.019	**0.016**	0.071	0.065	0.068	0.020	0.017
Rain Potsdam	0.0027	0.0021	0.0038	0.093	0.0023	0.099	0.0029	**0.0020**
Runoff Burghausen	0.011	**0.0069**	0.029	0.076	0.081	0.076	0.023	0.028
Runoff Cologne	0.016	**0.016**	0.025	0.064	1.6	0.043	0.034	0.19
Cumulus	0.012	0.0080	0.016	0.076	0.044	0.063	**0.0070**	0.029
Stratocumulus	0.018	**0.017**	0.018	0.026	0.49	0.042	0.038	0.028
Temperature	**0.0027**	0.0037	0.016	0.0062	1.7	0.0060	0.0055	0.073
Variabilty								
P-model	0.18	0.19	**0.17**	0.71	0.65	0.68	0.20	0.17
Rain Potsdam	0.032	0.028	0.048	0.92	0.036	0.99	0.037	**0.020**
Runoff Burghausen	0.12	**0.081**	0.31	0.76	0.81	0.76	0.24	0.28
Runoff Cologne	0.16	**0.16**	0.26	0.64	16	0.45	0.35	1.9
Cumulus	0.14	**0.10**	0.20	0.76	0.45	1.5	0.13	0.29
Stratocumulus	0.19	**0.18**	0.19	0.26	4.9	0.46	0.40	0.28
Temperature	**0.034**	0.040	0.17	0.066	17	0.062	0.057	0.73

Fig. 9. The 4th order structure function for the cumulus cloud measurement and its SIAAFT, IAAFT, AAFT, FARIMA and FARIMA + IAAFT surrogates.
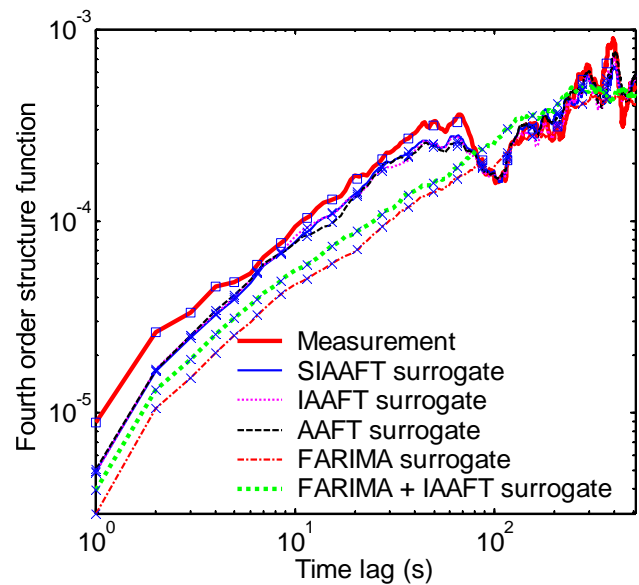

Fig. 10. The 4th order structure function for the stratocumulus cloud measurement and its SIAAFT, IAAFT, AAFT, FARIMA and FARIMA + IAAFT surrogates.

Kolmogorov-Smirnov distance D which is based on the maximum difference in the empirical distribution functions.

The differences D for the annual 95th percentile value are shown in Fig. 11. A small distance (small D value)

means that the two distributions are relatively similar. Those cases where the hypothesis of similarity is not rejected by the KS-test on the 5% level are marked by a green triangle. This is the case for four surrogates for the daily rain sums
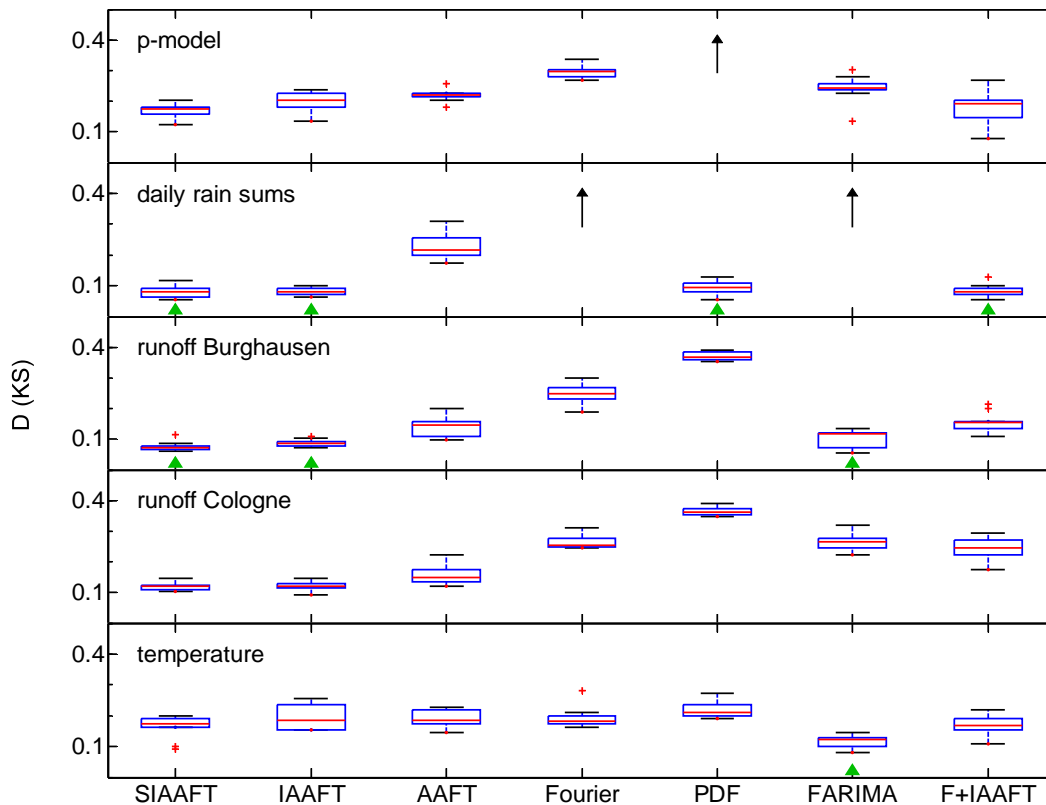
Fig. 11. The Kolmogorov-Smirnov distance of the distributions of the annual 95 percentile value between the measurement and its surrogates. The box plots show the spread of the ten surrogates. The red line in the box plots represents the median. The box indicates the 25 and 75 percentile levels; its size is called the interquartile range. The whiskers encompass all data points that fall in the range of 1.5 times the interquartile range. Outliers, i.e. data points beyond the whiskers, are indicated by a plus. In three cases, the distance D is so large that the box is outside the range of the plot, which is indicated by an upward arrow. This happens for the p-model because of a different standard deviation, for the rain cases because of the mean value. Additionally, a green triangle indicates the cases where the hypothesis of similarity of the distributions has not been rejected by the KS test.

(SIAAFT, IAAFT, PDF and FARIMA+IAAFT). SIAAFT and IAAFT also perform well for runoff in the small catchment (Burghausen), as does the FARIMA surrogate. Results for the Cologne runoff data are quite similar, but with slightly higher D-values so that in this case, the null hypothesis is rejected for all surrogates. For temperature, all surrogates behave similarly with only one non-rejection for the FARIMA surrogate. For the p-model, the distributions of the annual 95th percentile differ most between surrogates and originals.

Summarizing the results for all four statistics analyzed (figures not shown), the results for rainfall and temperature show that for precipitation the distribution is most important, while for temperature, the spectrum is most important; these results reflect the nature of those measurements. The agreement of the best surrogates is less good for temperature as compared to rain. For the FARIMA method, the distributions of the original time series and the surrogates are quite similar for temperature, but very different for precipitation.

The distributions of the yearly statistics of the p-model are quite different for all of the methods. Except for the PDF

surrogates, the distribution of the annual mean values is well matched for all surrogates. For the two runoff records, only minor differences are found between the smaller and larger river catchment size.

The distribution of the annual mean is the best reproduced statistic (except for the PDF method). In general, the annual statistics are best captured by the SIAAFT and the IAAFT method.

5.5 Return values

The variability for the estimated 50-year return values (RV50) in the small ensemble of surrogates is examined by using boxplots. Figure 12 displays the difference of the RV50 from each surrogate to that of the original time series. Ideally, the RV50 estimated from the original time series is within the range spanned by the surrogates, i.e. the boxplot of the surrogates includes zero.

The estimates of RV50 for precipitation are best with the two-statistic surrogates and also reasonable for the PDF
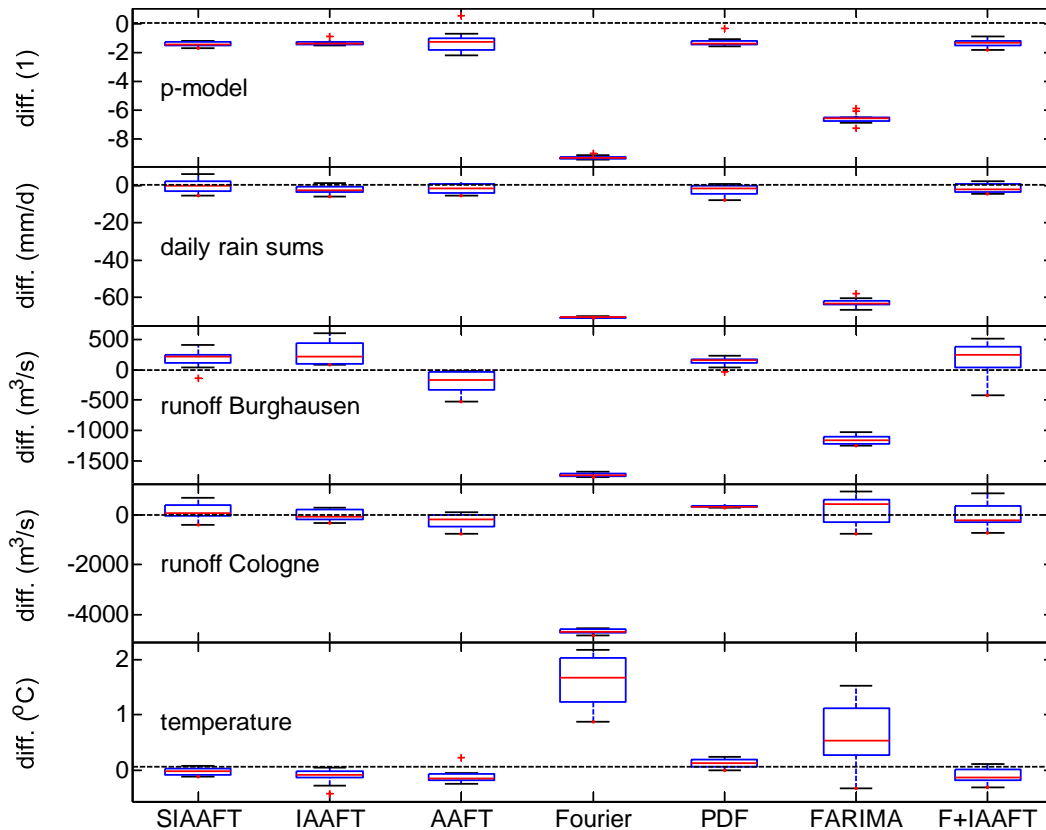
Fig. 12. Box plots of the deviations of the estimate of the 50-year return levels based on the measured data compared with their surrogates. For example, the negative differences for the p-model mean that the return levels of the surrogates is estimated to be too low.

surrogates. All surrogates (except the Fourier and FARIMA surrogates) are able to accurately estimate the return values for the Potsdam daily mean temperature.

The RV50 estimates for the p-model are not satisfying for any of the surrogates. They have a negative bias for the surrogates. In none of the methods zero is encompassed within the range of the whiskers, which maximally span 1.5 times the interquartile range.

For runoff, the differences are larger but there is still quite good agreement for the SIAAFT, IAAFT and the FARIMA + IAAFT methods. For the runoff data at Burghausen, we find SIAAFT, IAAFT, AAFT and FARIMA + IAAFT performing reasonably well. FARIMA and Fourier show a strong negative bias. At Cologne, the negative bias of the Fourier surrogate is even stronger.

The Fourier surrogates are out of range for all measurements. The return values of the PDF surrogates have a positive bias for the smoother measurements (runoff and temperature), because disregarding the autocorrelations decreases the probability of extremes being clustered together in one year; they are more evenly distributed among the years leading to more large values in the yearly maxima distribution.

In summary, the SIAAFT and IAAFT methods show reasonable results for all real measurements for the 50 year return value estimates, followed by the FARIMA + IAAFT, PDF, and AAFT methods. The Fourier and FARIMA algorithms are in general not satisfying in this respect.

6 Discussion

6.1 Measurements

The best surrogates (IAAFT, SIAAFT and FARIMA + IAAFT) based on the time series from the p-model reproduce the increment distribution and the structure functions reasonably well. Contrary to this, with respect to the annual statistics and return values this measurement was among of the most difficult ones.

We speculate that this contrast is connected with the strong intermittency, in the sense of the variance of the variance, of this time series. In years with a high variance, it is possible to have extreme events that one would not expect based on the mean variance over many years. The standard deviation of the annual standard deviation is about a third of

the total variability and twice as high in the original realisation as in, for example, its IAAFT surrogates. Conversely, the surrogates of the real historical measurements have more variability of the annual variability as their measurements. Furthermore, this deviation is only a few percent of the total variability.

The temperature measurement is characterised by a strong annual cycle. This cycle leads to a bimodal non-Gaussian value distribution. Even if the distribution of its Fourier surrogate is not closely matched, it is still bimodal. The increment distribution for a lag of 100 days (not shown) is also bimodal. As Fourier surrogates are generated by the addition of many independent sine signals, they are often said to have a Gaussian (increment) distribution because of the central limit theorem. Here we see that this is not always true. In this case, the annual cycle is so dominant that the distribution becomes bimodal.

The time series with daily rain sums was one of the easier records. Most of its statistics, including its fat tailed non-Gaussian increment distribution (not shown), were well captured by all surrogates, except the Fourier and FARIMA surrogates. Even the PDF surrogate was reasonably good, which reflects the relatively low autocorrelations found in this record.

Together with the p-model, the runoff measurement of the Rhine at Cologne was the most difficult time series. All surrogates showed large deviations in their structure function at scales less than 100 days and their return values. We expect that this is caused by the dynamics which is asymmetric in time, with fast swells and slower decreases. The (S)IAAFT surrogates are better than the FARIMA-based surrogates for this runoff measurement. The runoff of the Salzach at gauge Burghausen was easier to recreate, which may be due to its smaller asymmetry.

The power spectra of the surrogates of the cumulus cloud are considerably different from those of their corresponding measurement. These surrogates also display deviations on scales smaller than 10 s. The stratocumulus measurement was one of the easier data sets, except that a bump in the structure functions around a lag of 100 s was not picked up by the FARIMA-based surrogates.

6.2 Multifractal modelling

The advantage of multifractal modelling is often seen to be its ability to model non-Gaussian distributions and increment distributions, see, e.g., Davis (1999). In this article, we have shown that also the reverse can be true: that the inclusion of the distribution in a generator based on Fourier or FARIMA enables one to generate time series with almost the same increment distributions and structure functions. In this respect, it is to be expected that multifractal modelling and surrogate techniques are similarly powerful. Surrogate techniques have the advantage of enhanced accuracy for non-fractal signals: the structure functions of five of the six real measurements

were more accurately reproduced by the surrogates due to their non-fractal structure. Nonetheless, multifractal modelling has the benefit of providing a simple mathematical description and having a strong mathematical basis.

The link between the distribution and the structure functions suggests that the deviations found at small scales in the SF may be caused by the limited influence of these small scale variations on the distribution of the time series. This conjecture can be tested using more accurate techniques to generate surrogate time series, such as constrained randomization using global search algorithms such as simulated annealing (Schreiber, 1998) or evolutionary algorithms (Venema, 2003).

6.3 Algorithms

The main grouping that can be made is between algorithms that have only one statistic (PDF, Fourier and FARIMA surrogates) and the ones with both spectral and distribution statistics (SIAAFT, IAAFT, FARIMA + IAAFT and AAFT). The latter category clearly generates better surrogates, i.e. surrogates that fit better to the considered advanced statistics. Overall, SIAAFT, IAAFT and FARIMA + IAAFT have a similar quality. SIAAFT is slightly better than IAAFT for temperature with respect to the increment distributions, structure functions and 50-year return values. On the other hand, there are even cases where the AAFT algorithm performs best.

The choice of the best algorithm for a certain application will involve some trial and error. The Fourier-based algorithms are easily implemented in automatic data processing procedures and can be generalised to higher dimensions. The higher computational costs and the modest accuracy gains of the SIAAFT method compared to IAAFT limits the utilisation of the SIAAFT algorithm to applications where only a few high-quality time series or fields are needed. FARIMA-based approaches are most interesting in cases where longer records than the empirical ones are needed, because their full parametric form can easily be used for longer simulations.

7 Conclusions and outlook

In this study, the statistical properties of a range of measurements are compared with their surrogate time series. Seven different measurements are analysed: a realisation from the fractional integrated p-model (as an idealised record), historical time series of mean daily temperature, daily rain sums and runoff from two rivers, and cloud measurements from a stratocumulus and a cumulus field. Seven different algorithms were utilised to generate surrogate time series, and the most advanced ones were able to reproduce the empirical distribution exactly and the empirical power spectrum closely.

Using this setup, we studied the similarity of the measurements and their surrogates with respect to their power spectra, increment distributions, structure functions, annual statistics and return values.

The surrogates that reproduce both the power spectrum and the distribution of the measurement are able to accurately match the increment distributions and the structure functions of the measurement, but this often does not hold for surrogates that only mimic the power spectrum of the measurement. This illustrates the importance of using the distribution as well. Furthermore, an accurate reproduction of the distribution is important in many applications. For instance, for cloud fields, the distribution explains much of the radiative transfer through clouds, while the structure of the fields is mostly of second order importance (although still important).

However, even the best surrogates do not have asymmetric increment distributions, i.e., they cannot model dynamical processes that are asymmetric in time. Furthermore, an important inaccuracy is that the structure functions deviate on small scales.

Except for the deviations in the structure functions at small scales, the surrogates are thus at par with records from multifractal processes for modelling self similarity. When time series are not well characterised by fractal properties, the surrogate approach has the advantage of allowing for arbitrary spectra. For five of the six real measurements, we studied in this paper, the surrogates achieved a higher accuracy for the 4th order structure function than an ideal multifractal model would have realized without subtracting cycles. Multifractal modelling has the advantage of its strong analytical basis. The parametric (FARIMA) surrogate time series can be seen as a compromise between these two extremes; they are reasonably accurate, and have a clear mathematical model behind them.

The annual statistics and the return periods estimated by extreme value statistics are reproduced reasonably well. However, in many cases, the surrogates were statistically significantly different. Also here, the advantage of using the distribution as well as the spectrum was demonstrated.

Although the structure functions of the p-model realisation were reproduced with reasonable accuracy, its return values were much off. We speculate that for extreme value statistics, also the intermittency, in the sense of the variance of the variance at a certain scale, of the time series is important. Even if the better surrogates are able to handle the burst-like intermittency of the cumulus and the rain measurement, they do not reproduce the strong variance of the variance of the p-model, and these types of surrogates have more of this type of intermittency for the real measurements.

Even though the results in this study are encouraging, there is still much room for further developments in the flexible surrogate data framework. Schreiber and Schmitz (2000) presented a multivariate IAAFT algorithm. With such an algorithm, one could fill gaps or extend time series by us-

ing cross-correlations with other nearby measurements, estimating the spectra, cross spectra and distributions by using the available measurements. In addition, known cross-correlations with cycles such as the El Niño could be taken into account.

In Venema et al. (2006a), 3-dimensional surrogate cloud fields are generated where the cloud water distribution is a function of the height level. In the same way, it should be possible to generate surrogate time series where the distribution depends on the season or phase of a cycle.

We are developing a new algorithm which can also take the asymmetry of the dynamics into account. This will be especially important for modelling, for instance, runoff, turbulence and soil moisture. It seems to be possible, and should be worthwhile, to develop an IAAFT-type algorithm in which the wavelet transform is used instead of the Fourier transform. Such an algorithm would be superior in reproducing measurements from intermittent processes.

Edited by: M. Thiel
Reviewed by: three referees

References

Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A., and Taqqu, M. S.: Generators of long-range dependent processes: A survey, in: Theory and Applications of Long-Range Dependence, edited by: Oppenheim, G., Taqqu, M. S., and Doukha, P., Birkhäuse, Boston, USA, 579–623, 2003.

Beran, J.: Statistics for Long-Memory Processes, Monographs on Statistics and Applied Probability, Chapman & Hall, New York, USA, 1994.

Beran, J., Bhansali, R. J., and Ocker, D.: On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes, Biometrika, 85, 921–934, 1998.

Bisaglia, L.: Model selection for long-memory models, Quaderni di Statistica, 4, 2002.

Brockwell, P. J. and Davis, R. A.: Time series: theory and methods, Springer series in statistics, Springer, Berlin, Germany, 1991.

Christakos, G.: Random field models in earth sciences, Academic Press, San Diego, USA, 474 p., ISBN: 0-12-174230-X, 1992.

Davis, A., Marshak, A., Cahalan, R., and Wiscombe, W.: The LANDSAT scale break in stratocumulus as a three-dimensional radiative transfer effect: implications for cloud remote sensing, J. Atmos. Sci., 54, 241–260, 1997.

Davis, A. B., Marshak, A., Gerber, H., and Wiscombe, W. J.: Horizontal structure of marine boundary layer clouds from centimeter to kilometer scales, J. Geophys. Res., 104, 6123–6144, 1999.

Coles, S.: An introduction to statistical modeling of extreme values, Springer series in statistics, Springer, Berlin, Germany, 2001.

DWD, Deutscher Wetterdienst: http://www.dwd.de/en/FundE/Klima/KLIS, Offenbach, Germany, 2006.

GRDC, Global Runoff Data Centre of the Federal Institute of Hydrology: http://grdc.bafg.de, Koblenz, Germany, 2006.

Evans, K. F. and Wiscombe, W. J.: An algorithm for generating stochastic cloud fields from radar profile statistics, Atmos. Res., 72, 263–289, 2004.

Frisch, U.: Turbulence, the legacy of A.N. Kolmogorov, Cambridge University Press, UK, ISBN: 0521457130, 1995.

Gençai, R., Selçuk, F., and Ulugülyagci, A.: EVIM: A Software Package for Extreme Value Analysis in MATLAB, Studies in Nonlinear Dynamics and Econometrics 5, 213–239, 2001.

Kugiumtzis, D.: Test your surrogate data before you test for nonlinearity, Phys. Rev. E, 60, 2808–2816, 1999.

Hannan, E. J. and Quinn, B. G.: The determination of the order of an autoregression, J. Roy. Statist. Soc., 41, 190–195, 1979.

Haslett, J. and Raftery, A. E.: Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource, J. Appl. Statist., 38, 1–50, 1989.

Kallache, M., Rust, H., and Kropp, J.: Trend Assessment: Applications for Hydrology and Climate, Nonlin. Processes Geophys., 2, 201–210, 2005,
http://www.nonlin-processes-geophys.net/2/201/2005/.

Kantz, H. and Schreiber, T.: Nonlinear time series analysis, Cambridge Nonlinear Science Series No. 7, Cambridge University Press, Cambridge UK, 1997.

Koutsoyiannis, D.: The Hurst phenomenon and fractional Gaussian noise made easy, Hydrol. Sci. J., 47, 573–595, 2002.

Kuhn, Th.: The structure of scientific revolutions, Chicago University Press, USA, 1963.

Kylling, A., Webb, A. R., Kift, R., et al.: Spectral actinic flux in the lower troposphere: measurement and 1-D simulations for cloudless, broken cloud and overcast situations, Atmos. Chem. Phys., 5, 1975–1997, 2005,
http://www.atmos-chem-phys.net/5/1975/2005/.

Lehmann, A.: Die Säkulare Klimareihe von Potsdam, in: Klimastatusbericht 2001, DWD, Offenbach a.M., Germany, 227–239, 2001.

Lohre, M., Sibbertsen P., and Könnig, T.: Modeling water flow of the Rhine River using seasonal long memory, Water Resour. Res., 39, 1132, doi:10.1029/2002WR001697, 2003.

Hurst, H. E.: Long-term storage capacity of reservoirs, Trans. Am. Soc. Civil Eng., 116, 770–799, 1951.

Mandelbrot, B. B.: The Fractal Geometry of Nature, W H Freeman, New York, USA, ISBN: 0716711869, 460 p., 1982.

Masters, F. and Gurley, K. R.: Non-Gaussian Simulation: Cumulative Distribution Function Map-Based Spectral Correction, J. Eng. Mech., 1418–1428, 2003.

Montanari, A., Rosso, R., and Taqqu, M. S.: Fractional differenced ARIMA models applied to hydrologic time series: Identification, estimation and simulation, Water Resour. Res., 33, 1035, 97WR00043, 1997.

Molz, F. J., Rajaram, H., and Lu, S.: Stochastic fractal-based models of heterogeneity in subsurface hydrology: origins, applications, limitations, and future research questions, Rev. Geophys., 42, RG1002/2004, 2004.

Pincus, R., Hannay, C., and Evans, K. F.: The accuracy of determining three-dimensional radiative transfer effects in cumulus clouds using ground-based profiling instruments, J. Atmos. Sc., 62, 2284–2293, 2005.

Radziejewski, M., Bardossy, A., and Kundzewicz, S.: Detection of change in river flow using phase randomization, Hydrol. Sci. J., 45, 547–558, 2000.

Rust, H.: http://www.pik-potsdam.de/~hrust/tools.html, 2006.

Scheirer, R. and Macke, A.: On the accuracy of the independent column approximation in calculating the downward fluxes in the UVA, UVB and PAR spectral ranges, J. Geophys. Res., 106, 14 301–14 312, 2001.

Scheirer, R. and Schmidt, S.: CLABAUTAIR: a new algorithm for retrieving three-dimensional cloud structure from airborne microphysical measurements, Atmos. Chem. Phys., 5, 2333–2340, 2005,
http://www.atmos-chem-phys.net/5/2333/2005/.

Schreiber, T. and Schmitz, A.: Improved surrogate data for nonlinearity tests, Phys. Rev. Lett, 77, 635–638, 1996.

Schreiber, Th.: Constrained randomization of time series data, Phys. Rev. Lett., 80, 2105–2108, 1998.

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

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

Theiler, J. and Prichard, D.: Constrained-realization Monte-Carlo method for hypothesis testing, Physica D, 94, 221–235, 1996.

Timmer, J. and König, M.. On generating power law noise, A&A, 300, 707–710, 1995.

Turcotte, D. L.: Fractals and Chaos in Geology and Geophysics, Cambridge University Press, U.K., ISBN: 0521567335, 412, 1997.

Venema, V., Meyer, S., Gimeno García, S., Kniffka, A., Simmer, C., Crewell, S., Löhnert, U., Trautmann, U., and Macke, A.: Surrogate cloud fields generated with the Iterative Amplitude Adapted Fourier Transform algorithm, Tellus 58A, 104–120, 2006a.

Venema, V., Ament, F., and Simmer, C.: A Stochastic Iterative Amplitude Adjusted Fourier Transform Algorithm with improved accuracy, Nonlin. Processes Geophys., 13, 321–328,, 2006b.

Venema, V.: http://www.meteo.uni-bonn.de/mitarbeiter/venema/themes/surrogates/iaaft/statistics, 2006.

Venema, V.: An evolutionary search algorithm to generate 3D cloud fields with measured cloud boundary statistics, report, http://www.meteo.uni-bonn.de/victor/articles/2003/2003_report_cloud_boundaries_evol_search.pdf, 2003.

Wilson, J., Lovejoy, S., and Schertzer, D.: Continuous multiplicative cascade models of rain and clouds. Scaling, fractals and non-linear variability in geophysics, edited by: Schertzer, D. and Lovejoy, S., 185–208, Kluwer, Dordrecht, The Netherlands, 1991.