Fewer Jumps, Less Memory: Homogenized Temperature Records and Long Memory

H. W. Rust

Physics Institute, University of Potsdam, Germany

O. Mestre

Ecole Nationale de la Météorologie, Toulouse, France

V. K. C. Venema

Meteorological Institute, University of Bonn, Germany

Abstract. Air temperature records are commonly subjected to inhomogeneities, e.g., sudden jumps caused by a relocation of the measurement station or by installing a new type of shelter. We study the effect of these inhomogeneities on the estimation of the Hurst exponent and show that they bias the estimates towards larger values. The Hurst exponent is a parameter to measure long-range dependence (LRD) – a characteristic frequently used to describe the natural variability of temperature records. Analyzing a set of temperature time series before and after homogenization with respect to LRD we find that the average Hurst exponent is clearly reduced for the homogenized series. To test whether a) jumps cause this positive bias and b) the homogenization does not artificially reduce the Hurst exponent estimates, we perform a simulation study. This test shows that inhomogeneities in form of jumps bias the Hurst exponent estimation and the homogenization procedure is able to remove this bias, leaving the Hurst exponent unchanged. This result holds for FARIMA-based, as well as for DFA based estimation. We conclude that the use of homogenized series is necessary to prevent misleading conclusions about the dependence structure and thus about subsequent analysis such as trend tests.

1. Introduction

Many long instrumental climate records are available and may provide useful information for climate research. Numerous records have been incorporated into comprehensive data sets [e.g., *Jones et al.*, 2006], which are used for policy advice [e.g., *Trenberth et al.*, 2007]. Others have been studied directly, mostly with respect to climate change matters [e.g., *Alexandrov et al.*, 2004; *Brunetti et al.*, 2004; *Schär et al.*, 2004; *Begert et al.*, 2005; *Drogue et al.*, 2005; *Auer et al.*, 2006].

A typical problem in such studies is to discriminate an actual change in the climate from a) natural climate variability or b) temperature shifts due to inhomogeneities in the observations. Inhomogeneities in the order of magnitude of the expected change in the signal directly affect climate change studies. A different problem is to discriminate changes in the system from natural variability. This depends crucially on the assumption on the process assumed to cause this natural variability. This assumption is usually motivated by the observed data and we will show that it can also be influenced by the presence of inhomogeneities. In other words, climate change studies are affected by the presence of inhomogeneities in temperature records in two ways: 1) directly, via artificial jumps which might be mistaken for or mask climate change signals and 2) indirectly, via the overestimation of the natural variability due to inhomogeneous observations. The second indirect effect is subject of the present work.

Natural temperature variability has been frequently described as a stochastic process with long-range dependence (LRD) or long-memory [e.g., *Smith*, 1993; *Fraedrich and Blender*, 2003; *Cohn and Lins*, 2005]. Such processes show outstanding long excursions from a constant mean value, which might be easily mistaken for deterministic trends, i.e. actual changes in the mean [*Koutsoyiannis*, 2003]. From a statistical viewpoint, LRD is a property of a stochastic process resulting from an algebraic, i.e. slow, decay of the autocorrelation function (ACF). The ACF describes the dependence of two observations separated by a time lag τ, Fig. 1 exemplarily shows the ACF of an independent, a short-range dependent (SRD) and a LRD process.

A decay of the ACF slower than exponential leads to an increase in the uncertainty of estimates obtained from a sample, it thus alters error bars and significance levels of statistical tests. Typically error bars are larger under the assumption of LRD than for SRD or independence [e.g., *Bloomfield*, 1992; *Beran*, 1994; *Koutsoyiannis*, 2003; *Cohn and Lins*, 2005]. This means a slope of a linear trend might be significantly different from zero under the assumption of SRD or independence but not under the assumption of LRD. The schematic plot in Fig. 2 shows the confidence bands under the assumption of independence (dark grey shadow) and LRD (light grey shadow); only under the latter assumption the slope is compatible with zero. This example highlights how important it is for a reliable statistical trend test to have information about the dependence structure of the data, in particular about the presence and strength of LRD. Its confident detection and quantification, in turn, requires a homogeneous data record, as we will show in Sec. 4. Temperature records are, however, affected by changes in the measurement conditions, e.g., relocation of the weather stations, modernization of the instrumentation, changes in observation rules, automation, etc. Therefore obtaining information about the dependence structure from the raw records without prior homogenization is likely to yield unreliable results.

0148-0227/08/$9.00

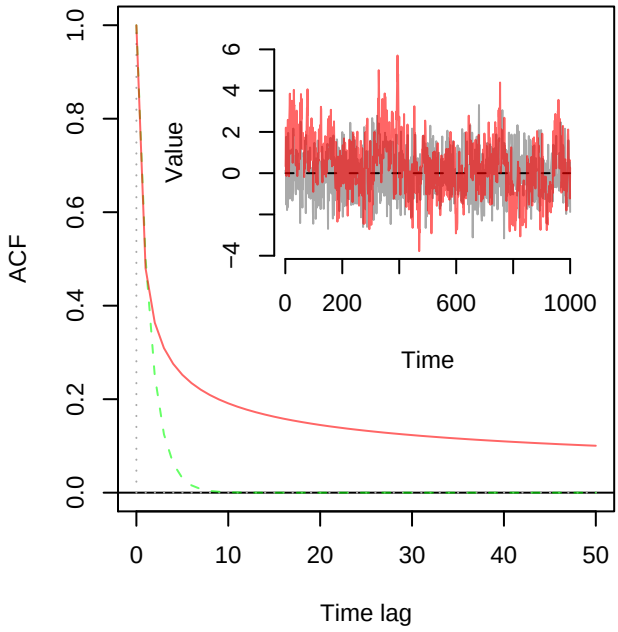

Figure 1. ACF of an independent (dotted, grey), SRD (dashed, green) and LRD (solid, red) process. The SRD process shows an exponential decay, the tail of the LRD process decays slower. Inset: example of a time series from an independent (dotted, grey) and a LRD (solid, red) process. The latter shows longer departures from the mean.

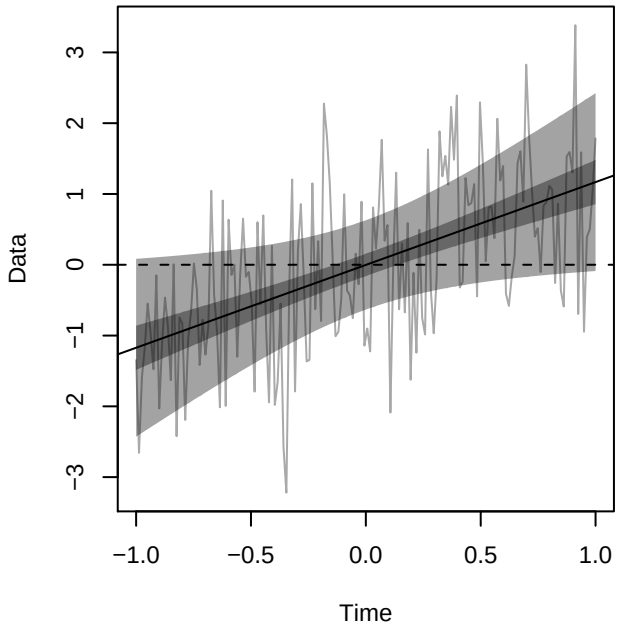

Figure 2. Example for the sensitivity of uncertainty intervals on the assumption on the dependence structure. A white noise series plus a linear trend (grey) and estimated linear trend (solid black) with 95% uncertainty intervals under the assumption of independence (light grey) and LRD (dark grey). Under the latter assumption, the estimated trend is compatible with a constant zero mean, i.e. no trend.

A literature survey reveals that most studies on long-range dependence of climate records were probably performed with data that was not homogenized, or at least, that most authors were not aware of the importance of homogenization. We searched the "Science Citation Index" database for articles containing keywords related to long-range correlations and climate records. This search was restricted to a range of physics and climate journals and produced 40 articles of which 24 analyzed observed climatological time series. Most, that is 18 articles, give no information on the quality of the data. Two papers assert that the data is high quality or consists of selected records that are supposed to be homogeneous. One article indicates that part of the data could be inhomogeneous [*Blender and Fraedrich*, 2006]. Two articles analyzed two different datasets of which only one was homogenized [*Weber and Talkner*, 2001; *Caballero et al.*, 2002]. One article used only homogenized data [*Talkner and Weber*, 2000], but part of the inhomogeneities was detected by eye [*Weber et al.*, 1994].

In the following, we show that inhomogeneities cause a positive bias in the LRD parameter estimation. This bias can be reduced with a suitable homogenization procedure. In Sec. 2 we introduce the homogenization algorithm and the two estimation strategies for the LRD parameter. The analysis of the set of temperature records is described in Sec. 3. This is followed by a simulation experiment in Sec. 4 meant to corroborate the conclusions from the empirical data. Findings are summarized in Sec. 5. Detailed information about the two estimation procedures and a discussion about undetected inhomogeneities are given in the appendix.

2. Methods

2.1. Homogenization procedure

In most cases inhomogeneities in monthly mean-values are step-like changes which typically alter only the mean, leaving the higher moments of the distribution unchanged [*Alexandersson*, 1986]. The aim of a homogenization procedure is to detect and correct these changes. Inhomogeneities of this kind can be detected as jumps in the time series obtained as difference in observations from two nearby stations. The measurements of neighboring stations are usually strongly correlated and jumps in the difference of these measurements indicate a change in the conditions of one station. By analyzing a larger network of stations, these jumps can in general be attributed to a single station. Once detected, shifts have to be corrected. In our study, the correction is performed by means of a two factor ANOVA model. It is based on the assumption that temperature series belonging to the same climatic area are more or less affected by the the same climatic conditions at a given time. This assumption is realistic when considering monthly or annual observations at a regional scale. Details of the detection algorithm and the correction method discussed below are given in *Caussinus and Mestre* [2004].

We assume that each series of observations is the sum of a climate effect μ_t (at time t), a station effect v_j (for station j) and random white noise. To ensure parameter identification, the μ_t's are assumed to have zero mean (we thus investigate temperature anomalies), but no further assumptions are made on climate effect variations. The station effect v_j is assumed to be constant if the series is reliable. If not, the station effect is then piecewise constant between two shifts, that is v_j becomes $v_{j,h}$, h being the index of the sub-period between consecutive shifts h and $h+1$.

Given a set of series, and assuming the dates where jumps occur are known from earlier studies (based on change-point

detection procedures and the use of archives and meta data which are not described here), each climate effect μ_t and station effect $v_{j,h}$ may be estimated by means of standard least squares. Let $v_{j,e}$ be the most recent station effect for station j. Correction (or homogenization) is then performed for each series by setting all $v_{j,h}$ equal to the last one, that is adding $v_{j,e} - v_{j,h}$ for every data between consecutive shifts h and $h+1$.

It is verified on observations that, conditioned on a climate signal μ_t, independence and normality assumptions for the residuals can be accepted. The latter is consistent with the knowledge of climatologists [see, e.g., *Alexandersson*, 1986]. Temporal auto-correlation that exists in the instrumental series is well taken into account through the climate effects μ_t's.

2.2. Estimating the LRD Parameter

A characteristic of LRD processes are long departures from a mean value. Contrary to deterministic trends, the process return regularly to this constant mean. A parameter to quantify the strength of LRD can be defined in multiple ways: for example using the Hurst exponent H [*Hurst*, 1951; *Beran*, 1994], or a power-law exponent describing the decay of a pole in the spectral density [*Geweke and Porter-Hudak*, 1983; *Beran*, 1994; *Robinson*, 1995]. A very general definition of LRD involves the concept of a stationary stochastic process X_t with ACF $\rho(\tau)$. X_t is long-range dependent (LRD) or has long-memory if $\sum_{\tau=-\infty}^{\infty} \rho(\tau) = \infty$, i.e. the sum of the ACF – the memory – is infinite. This is the case for a process with an algebraically, i.e. slowly, decaying ACF as shown in Fig. 1 (red line),

$$\rho(\tau) \propto \tau^{(2H-2)},\ \tau \to \infty. \tag{1}$$

H denotes the Hurst exponent, and is used as a parameter to quantify the strength of long-range dependence. For stationary long-range dependent processes $0.5 < H < 1$. Alternatively, a pole in the spectral density as the frequency approaches zero indicates a LRD process [e.g., *Beran*, 1994].

If the ACF is summable ($\sum_{\tau=-\infty}^{\infty} \rho(\tau) = c < \infty$) the memory is finite and the process is short-range dependent (SRD). This behavior is found for most commonly used time series models as, for example, auto-regressive moving average (ARMA) processes [*Brockwell and Davis*, 1991]. Short-range dependent processes have an exponentially decaying ACF. As well as independent processes, they have a trivial value for the Hurst exponent: $H = 0.5$.

There are basically two modi operandi for estimating a long-range dependence parameter: a parametric and a semi-parametric approach. The parametric approach requires specification of a parameterized model, i.e. a stochastic process. The model parameters (including H) are then estimated with standard procedures such as maximum-likelihood from the empirical time series [*Beran*, 1994].

Semi-parametric approaches do not require a complete parameterization of the process. Instead only suitable asymptotic relations such as Eq. (1) or a similar equation in the spectral domain are used in a regime where this asymptotic behavior is assumed to hold [*Geweke and Porter-Hudak*, 1983; *Robinson*, 1995]. Proceeding this way is useful if it is difficult to find an adequate parametric model. However, the problem of choosing a parametric model is avoided at the costs of having to choose a suitable regime for the asymptotic relation to hold.

Different estimation methods, parametric, as well as semi-parametric have been discussed and compared by *Beran* [1994]; *Taqqu et al.* [1995]; *Doukhan et al.* [2003] and *Robinson* [2003]. Many semi-parametric methods are "heuristic", i.e. they yield an estimator for which no limiting distribution has been derived. Thus, confidence intervals cannot be obtained straightforwardly and statistical inference about the LRD parameter is not possible. These estimators should be avoided.

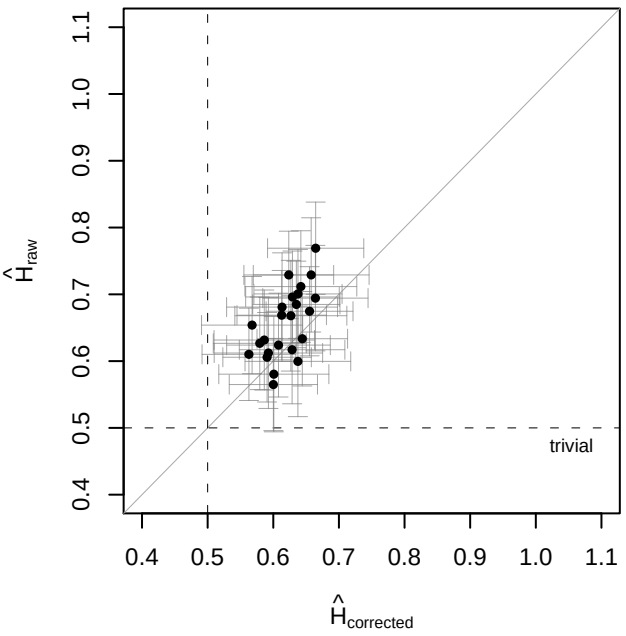

Figure 3. Estimates of the LRD parameter using a FARIMA$[1, d, 0]$ model for the homogenized series series (abscissae) versus the respective raw series (ordinate). The whiskers give 95% confidence intervals, the dashed lines mark the $H = 0.5$ as expected for short-range or independent series.

Table 1. Hurst exponent estimates averaged for the 24 raw and corrected temperature anomaly series. FARIMA-based estimation with standard errors in parentheses and DFA-based estimation for three different bandwidth $s_{\min} < s < s_{\max}$: $\log s_{\min}^{(1)} \gtrsim 2$, $\log s_{\min}^{(2)} \gtrsim 1.5$, and $\log s_{\min}^{(3)} \gtrsim 1$; $s_{\max} \approx N/4$, as shown in Fig. 6.

	FARIMA		DFA		
Series	$[1,d,0]$	$[1,d,1]$	$s^{(1)}$	$s^{(2)}$	$s^{(3)}$
Raw	0.657(0.008)	0.707(0.012)	0.675	0.707	0.751
Corrected	0.619(0.008)	0.641(0.011)	0.634	0.651	0.690

In the following, we use two approaches: a parametric approach involving the flexible class of FARIMA$[p, d, q]$ models (App. A, [*Beran*, 1994]), and detrended fluctuation analysis (DFA) (App. B, [*Kantelhardt et al.*, 2001]). Although the latter is a heuristic approach we include it in the present work because it has become popular in the geosciences. We particularly point out that it is affected by the inhomogeneities typically occurring in temperature series in the same way as other methods are.

3. Analyzing Temperature Series

We study monthly maximum surface air temperature from 24 stations in France. The raw data is extracted from the BDCLIM (Meteo-France climatological database). The set of series is carefully built, analyzed and homogenized in *Caussinus and Mestre* [2004]. Most of those series date back to the late XIXth century. In the following, we consider the temperature anomalies (deviation of a long-term mean obtained for a specific month) of the raw series and their corresponding homogenized series.

In order to investigate the influence of homogenization, we estimate the Hurst exponent for both types of anomaly records, from the raw and the corrected series. Estimates are obtained using the Whittle estimator for a FARIMA$[1, d, 0]$ and a FARIMA$[1, d, 1]$ model (Sec. A). Both models are a reasonable choice according to the Hannan-Quinn information criterion (HIC) [*Beran et al.*, 1998]. Additionally, DFA-based estimates are obtained for the same series.

The scatter plot in Fig. 3 depicts the Hurst exponent estimates for the raw series ($\hat{H}_{\mathrm{raw}}$) against the corrected series ($\hat{H}_{\mathrm{corrected}}$) obtained using the FARIMA$[1, d, 0]$ model. Estimates are more likely to be reduced for the corrected series than increased: for the FARIMA$[1, d, 0]$ model 19 out of 24 (80%) estimates are reduced for the corrected series. The mean Hurst exponent estimates for two variants of FARIMA-based and three variants of DFA-based estimation are shown in Table 1. In all cases we found a reduction of the estimate.

Although the estimates are reduced, all but two FARIMA-based estimates are on a 95% level significantly different from the trivial value ($H = 0.5$) and are thus not compatible with an SRD process. According to the argumentation in App. C, we expect that undetected small inhomogeneities are not responsible for $H > 0.5$. However, we can not exclude network inhomogeneities causing a positive bias; jumps occurring simultaneously in the whole network of stations will not be detected but we consider them as unlikely. Furthermore, a positive bias due to a deterministic trend or due to an inadequate model can not be excluded.

4. Simulation Experiment

In the following, we design a simulation experiment to show that a) jumps introduced in the series cause a positive bias of the Hurst exponent estimate and b) the homogenization algorithm used in Sec. 3 properly corrects the series and this bias almost vanishes. On the basis of a known process, we can construct an artificial set of "temperature" series with similar properties as the empirical data set studied in Sec. 3. The main characteristics of interest, which we pursued to reproduce in the following, are the LRD in the anomalies and the spatial correlation of records measured at two different locations.

We start with a FARIMA$[1, d, 0]$ process with parameters $d = 0.119, a = 0.061$ and $\sigma_\eta = 3.414$. The spectral properties are similar to those of the empirical anomaly series, in fact d, a, and σ_η are the arithmetic means of the corresponding FARIMA$[1, d, 0]$ parameter estimates found for these empirical temperature anomalies. In the following, we use realizations T_i^A of this process with $i = 1, \ldots, N = 12 \cdot 150$ data points, i.e. 150 years of monthly values, and the superscript A indicating temperature anomalies.

A periodic signal representing the annual cycle is added to T_i^A. It was obtained by averaging the mean annual cycle of all 24 temperature records used in Sec. 3. This results in an artificial "temperature" series T_i^0 which represents the mean temperature of a regional network.

To obtain a set of multiple "observed temperature" series T_i^j, $j = 1, \ldots, 10$, one for each station j in the regional network, ten Gaussian white noise series $\epsilon_i^j \sim WN(0, \sigma_\epsilon^2)$ with standard deviation $\sigma_\epsilon = 0.25$ are added to T_i^0. This yields a set of ten mutually correlated (in space) "temperature" series T_i^j which resemble a collection of records taken at ten different locations with slightly different local influences. We refer to these data sets as the "original" sets. The unit of the signal will be °C in the following.

To obtain a set of inhomogeneous (or "perturbed") series, step-like changes are added randomly with a probability of occurrence of 1/15 per year. The amplitudes of the shifts are randomly chosen as ±0.6°C, ±0.8°C or ±1°C. Note that, in order to prevent poor estimation of the correction constants, i.e. the sizes of the breaks, we ensure that two consecutive shifts are separated by at least three years. According to this scheme, we produced an ensemble of ten "perturbed" sets of simulated "temperature" records, which can be thought of as ten different regions, each with measurements at ten different locations.

The ten randomly perturbed series in each region are corrected using the two factors correction model (Sec. 2.1),

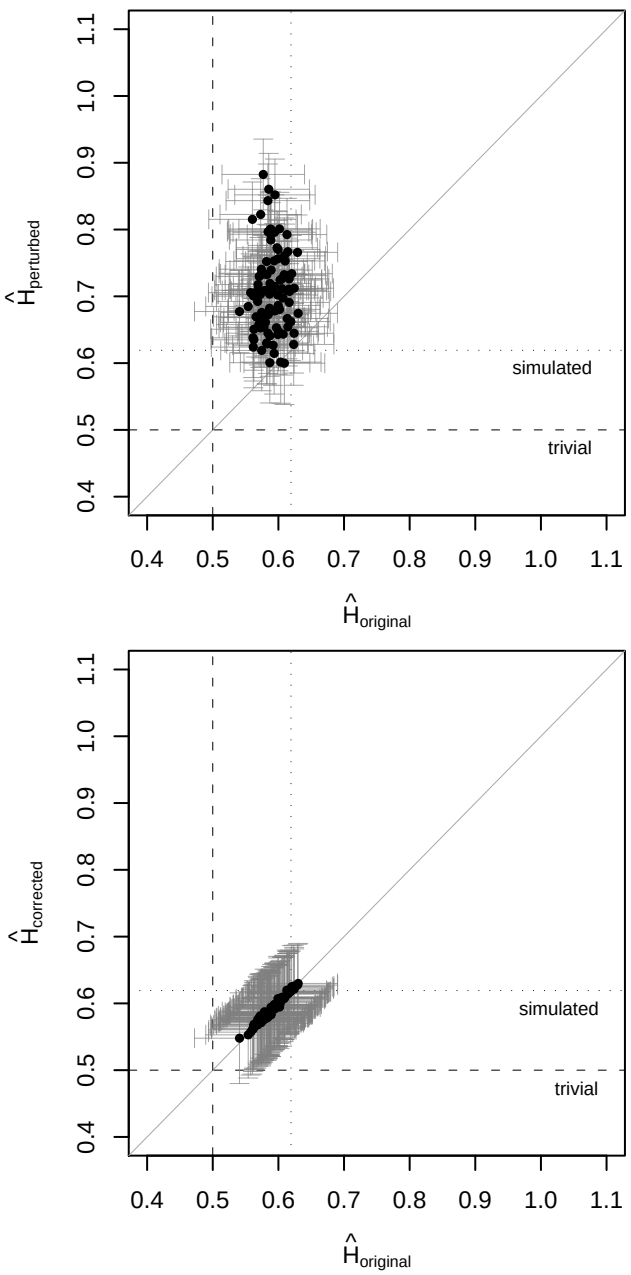

Figure 4. Estimates of the LRD parameter using a FARIMA$[1, d, 0]$ model for the original series series (abscissae) versus the respective series including breakpoints (ordinate, upper panel) and for the series with breakpoints removed (ordinate, lower panel). The whiskers give 95% confidence intervals. The dashed lines mark the $H = 0.5$ as expected for short-range or independent series, the dotted lines mark the LRD parameter used for simulation.

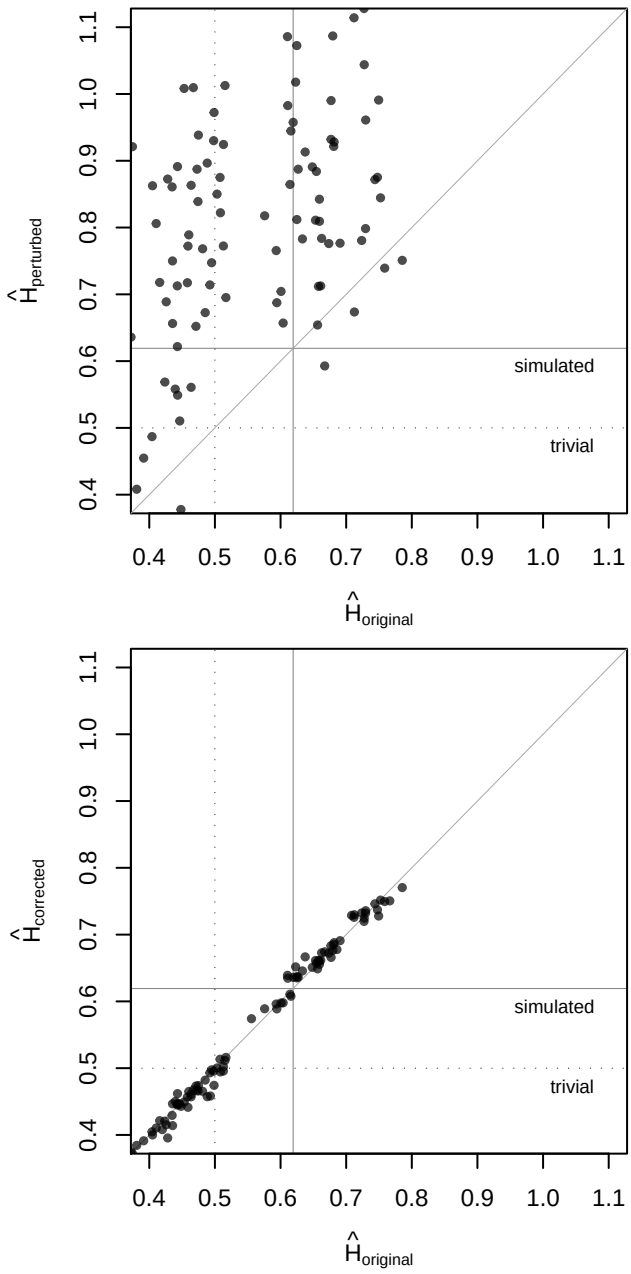

Figure 5. Estimate of the LRD parameter using DFA and scales $\log s_{\min} \gtrsim 1.5$ for the straight line fit to the logarithmic fluctuation function. The estimates for the original series are on the abscissae. The upper panel shows the estimates for series with breakpoints (ordinate) and the lower panel the result for the series with the breakpoints removed (ordinate). The dashed and the dotted lines are as in Fig. 4.

yielding the "corrected" data set. The position of the shifts is assumed to be known, which is realistic: in practical applications, only weak changes of amplitude smaller than one noise standard deviation remain poorly detected. For a discussion on undetected inhomogeneities see also App. C.

These three artificial sets of temperature series, the original, the perturbed, and the corrected are then analyzed in two steps according to the procedure used for the empirical series: the annual cycle is estimated and subtracted and a Hurst exponent is estimated from the residual (anomaly) series.

4.1. Estimation using a FARIMA Model

We separately analyze the three different data sets: "original", "perturbed" and "corrected". For their corresponding anomaly series, we estimate the Hurst exponent H as a measure for LRD using a FARIMA$[1, d, 0]$ model.

Figure 4 (upper panel) shows a scatter plot of the estimates for the set of perturbed records (ordinate) plotted against the original records (abscissae). Estimates for the perturbed series are in general larger than those for the corresponding original series. This indicates that perturbations (inhomogeneities) induce a positive bias to the estimator. This bias vanishes for the corrected data sets, Figure 4 (lower panel). Estimates from the corrected sets and the corresponding original sets are on the bisector, i.e. are identical. The majority of estimates $\hat{H}$ are larger than the trivial value $H = 0.5$ (dashed line) and smaller than the value $H = 0.62$ (dotted line), which was used for simulating the network anomaly temperature T^A. The latter is a consequence of adding white noise to the simulated series.

4.2. Estimation using DFA

The same analysis is repeated using the DFA-based estimator for the Hurst exponent. Figure 5 (upper panel) shows a scatter plot of the estimates from the perturbed series versus the corresponding original records. Similar to Fig. 4 (upper panel), also the DFA estimates for the perturbed sets are mostly larger than those for the corresponding original series, indicating a positive bias for DFA as well. The lower panel shows estimates of the corrected sets versus the corresponding original series. Again, we find that the correction of the data largely removes the bias in the estimates of the Hurst exponent.

Different from Fig. 4 (lower panel), we find many estimates above the parameter used for simulation ($H = 0.62$). Bearing in mind, that we added white noise to the simulations and expect a reduced estimate, this suggest a positive bias for the DFA estimator already for the unperturbed series. This is the result of a bias-variance trade-off made when choosing the range for the straight-line fit. To account for this issue, we chose three different ranges in the same way as in Fig. 6 in App. B: $s_{\min,1} = 10$, $s_{\min,2} = 44$, $s_{\min,3} = 110$, with $s_{\max} \approx N/4 = 450$. As expected, the variability of the estimate increases with $s_{\min}$. Figure 5 (lower panel) shows the results for $s_{\min,2} = 44$, which we consider as a reasonable compromise.

5. Conclusion

Temperature observations are subjected to changes in the measurement conditions (inhomogeneities) which may cause perturbations to their analysis. We study the effect of these inhomogeneities on the estimation of a LRD parameter (or Hurst exponent) for a set of temperature records from France. Comparing Hurst exponent estimates obtained from the set of raw (inhomogeneous) temperature anomalies to their corrected (homogenized) counterparts, we observe larger estimates for the inhomogeneous records. This indicates that inhomogeneities cause a positive bias to the Hurst exponent estimation with the following consequences: the Hurst exponent gives valuable information about the dependence structure of the underlying process, a misspecification might result in misleading conclusions about this dependence structure which a) hinders the understanding of the system and b) leads to too large uncertainty bounds or significance levels for subsequent analyzes such as trend tests.

To corroborate this hypothesis and exclude the possibility that the reduction in the Hurst exponent is an artefact of the homogenization algorithm, we repeat the same analysis for a set of simulated temperature records. For both estimation variants, a full parametric estimator based on FARIMA$[p, d, q]$ models and DFA as a heuristic estimator, the Hurst exponent estimates are larger for inhomogeneous records than for the original data; estimates obtained from the corrected records are, instead, almost identical with those obtained from the original sets. This supports our hypothesis that inhomogeneities cause a positive bias to the Hurst exponent estimation. It shows further that the homogenization algorithm discussed, whose efficiency in trend correction is well known, is also an effective way to compensate for the bias in the Hurst exponent estimation. We recommend that future studies investigating for LRD are performed only with homogenized data.

It is further interesting to note that the FARIMA-based, as well as the DFA-based estimation is affected. Especially the latter may seem surprising because extensive simulation studies with fractional Gaussian noise processes have been conducted to show that DFA is robust against some influences [*Chen et al.*, 2002; *Hu et al.*, 2001]. However, break inhomogeneities, which are ubiquitous in temperature records, can not be accounted for by a DFA-based estimator. The problem discussed is not only an issue of temperature records. *Klemes* [1974], for example, showed that sudden shifts can be mistaken for fractional Brownian noise in rescaled range analysis of river discharge. Such shift could occur due to re-calibration (after floods) of the rating curve between gauge height and discharge.

Identifying or ignoring LRD in temperature, river discharge or other geophysical records may have a tremendous influence on subsequent analysis as, e.g., trend test [e.g., *Cohn and Lins*, 2005; *Kallache et al.*, 2005; *Rust*, 2007; *Kallache*, 2007]. For a reliable detection of LRD it is, according to this study, necessary to account for influences of inhomogeneities.

Appendix A: Fractional ARIMA (FARIMA) Modeling

A FARIMA$[p, d, q]$ process is a linear stochastic process X_t satisfying

$$\Phi(B)(1-B)^d X_t = \Psi(B)\epsilon_t, \tag{A1}$$

with $BX_t = X_{t-1}$ (back-shift operator), auto-regressive and moving average polynomial $\Phi(z)$ and $\Psi(z)$ of order p and q, respectively; ϵ_t is a zero mean Gaussian white noise process with variance σ_ϵ^2, $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. For $0 < d < 0.5$ the process is LRD with $H = d + 0.5$. Setting $d = 0$ yields the widely used class of ARMA$[p, q]$ models [*Brockwell and Davis*, 1991].

Model parameters can be estimated using, e.g., exact maximum-likelihood or, as we did here, the Whittle estimator [*Beran*, 1994]. Figure 6 (upper panel) shows the periodogram of a simulated data set (Sec. 4) together with the estimated spectral densities of a FARIMA$[1, d, 0]$ (green) and a FARIMA$[1, d, 1]$ (red) process.

As statistical means to estimate the model orders p and q we use selection criteria as the Hannah-Quinn information criterion (HIC) [*Beran et al.*, 1998] which is similar to the popular Akaike Information Criterion (AIC). Alternatively, one can discriminate nested models using the likelihood-ratio test. A discussion on model selection among FARIMA models can be found in *Rust* [2007]. In the present study, we use HIC as an orientation but require one common model for all stations.

Appendix B: Detrended Fluctuation Analysis

Detrended fluctuation analysis (DFA, or "residuals of regression") has become popular as an estimator of the Hurst exponent H. The method is described in detail in *Kantelhardt et al.* [2001]. DFA works with the running sum (integral) of the data series within windows of size s. In these segments a local polynomial fit is subtracted and the variance $F^2(s)$ of the residuals is estimated. The slope of a straight-line fit to $\log F(s)$ vs. $\log s$ for large scales s yields the Hurst exponent estimator $\hat{H}_{\mathrm{DFA}}$ [*Taqqu et al.*, 1995].

As well as for other semi-parametric estimators, the choice of the range for this fit is crucial. Different ranges $s_{\min} < s < s_{\max}$ might yield different estimates; see Fig. 6 (lower panel). *Taqqu et al.* [1995] derived the expectation value for the DFA-based estimator $\hat{H}_{\mathrm{DFA}}$ in the limit of large scales s for fractional Gaussian noise and FARIMA processes and show that, indeed, the slope of the straight line fit yields an asymptotically unbiased estimator for H. The effect of certain types of perturbations, such as trends, outliers and missing values has been investigated for self-similar processes [*Chen et al.*, 2002; *Hu et al.*, 2001].

A major drawback of DFA is that, as for all heuristic estimators, there is no straightforward way to obtain confidence intervals for the estimates. This renders statistical inference impossible, i.e. there is no rigorous way to distinguish an estimate $\hat{H}$ from a trivial value $H = 0.5$. Because estimates are very unlikely to exactly attain the trivial value, incautious use might lead to false conclusions about LRD [*Metzler*, 2003; *Maraun et al.*, 2004; *Rust*, 2007].

Appendix C: Undetected Inhomogeneities

Statistical homogenization algorithms can not find all inhomogeneities. Especially small jumps appearing shortly before or after larger ones are difficult to detect. This is reflected in the distribution of the size of detected inhomogeneities which displays a local minimum around zero. Fitting a Gaussian to the tails of the detected inhomogeneity frequency-size distribution, we estimate the frequency of small values and thus the fraction of undetected inhomogeneities. In the worst case we find 25% of undetected inhomogeneities for the BDCLIM data set. Their size is about half the size of the detected jump. We thus expect the remaining undetected inhomogeneities in the homogenized BDCLIM dataset to make only a minor contribution to the LRD parameter estimates.

Acknowledgments. We would like to thank Alice Kapala for bringing us together. Henning Rust acknowledges support from the EC Project "Extreme Events: Causes and Consequences (E2-C2), Contract No. 12975 (NEST)" and the Collaborative Research Center SFB 555 by the DFG. Olivier Mestre acknowledges support from "COST-ACTION ES0601: Advances in homogenization methods of climate series: an integrated approach (HOME)". Victor Venema was supported by the project Large Scale Climate Changes and their Environmental Relevance funded by the North Rhine-Westphalia Academy of Science. We thank the R Development Core Team and the community for providing `R`, a free language and environment for statistical computing [*R Development Core Team*, 2004]. Furthermore, we thank two anonymous reviewers for very helpful comments.

References

Alexandersson, H., A homogeneity test applied to precipitation data, *J. Clim.*, *6*(6), 661–675, 1986.

Alexandrov, V., M. S. and E. Koleva, and J.-M. Moisselin, Climate variability and change in Bulgaria during the 20th century, *Theor. and Appl. Climatol.*, *79*, 133–149, 2004.

Auer, I., et al., HISTALP - historical instrumental climatological surface time series of the greater alpine region, *Int. J. Climatol.*, *27*(1), 17–40, doi:10.1002/joc.1377, 2006.

Begert, M., T. Schlegel, and W. Kirchhofer, Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000, *Int. J. Climatol.*, *25*, 65–80, 2005.

Beran, J., *Statistics for Long-Memory Processes*, Monographs on Statistics and Applied Probability, Chapman & Hall, 1994.

Beran, J., R. J. Bhansali, and D. Ocker, On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes, *Biometrika*, *85*(4), 921–934, 1998.

Blender, R., and K. Fraedrich, Long-term memory of the hydrological cycle and river runoffs in China in a high-resolution climate model, *Int. J. Climatol.*, *26*, 1547–1565, 2006.

Bloomfield, P., Trends in global temperature, *Climatic Change*, pp. 1–16, 1992.

Brockwell, P. J., and R. A. Davis, *Time series: Theory and Methods*, Springer Series in Statistics, Springer, Berlin, 1991.

Brunetti, M., M. Maugeri, F. Monti, and T. Nanni, Changes in daily precipitation frequency and distribution in Italy over the last 120 years, *J. Geophys. Res.*, *109*(D05102), doi: 10.10209/2003JD004296, 2004.

Caballero, R., S. Jewson, and A. Brix, Long memory in surface air temperature: detection, modeling, and application to weather derivative valuation, *Clim. Res.*, *21*, 127–140, 2002.

Caussinus, H., and O. Mestre, Detection and correction of artificial shifts in climate series, *Appl. Statist.*, *53*(3), 405–425, 2004.

Chen, Z., P. C. Ivanov, K. Hu, and H. E. Stanley, Effects of nonstationarities on detrended fluctuation analysis, *Phys. Rev. E*, *65*, 041,107, 2002.

Cohn, T. A., and H. F. Lins, Nature's style: Naturally trendy, *Geophys. Res. Lett.*, *32*(L23402), doi:10.1029/2005GL024476, 2005.

Doukhan, P., G. Oppenheim, and M. S. Taqqu (Eds.), *Theory and Application of Long-Range Dependence*, Birkhäuser, Boston, 2003.

Drogue, G., O. Mestre, L. Hoffmann, J.-F. Iffly, and L. Pfister, Recent warming in a small region with semi-oceanic climate, 1949-1998: what is the ground truth?, *Theor. and Appl. Climatol.*, *81*, 1–10, 2005.

Fraedrich, K., and R. Blender, Scaling of atmosphere and ocean temperature correlations in observations and climate models, *Phys. Rev. Lett.*, *90*, 108,501, 2003.

Geweke, J., and S. Porter-Hudak, The estimation and application of long memory time series models, *J. Time Ser. Anal.*, *4*, 221–237, 1983.

Hu, K., P. C. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley, Effects of trends on detrended fluctuation analysis, *Phys. Rev. E*, *64*, 011,114, 2001.

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

Jones, P. D., D. E. Parker, T. J. Osborn, and K. R. Briffa, Global and hemispheric temperature anomalies–land and marine instrumental records, in *Trends: A Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A., 2006.

Kallache, M., Trends in the mean and extreme values of hydro-meteorological time series, Ph.D. thesis, University of Bayreuth, Bayreuth, 2007.

Kallache, M., H. W. Rust, and J. Kropp, Trend assessment: Applications for hydrology and climate, *Nonlin. Proc. Geoph.*, *2*, 201–210, 2005.

Kantelhardt, J. W., E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, and A. Bunde, Detecting long-range correlations with detrended fluctuation analysis, *Physica A*, *295*, 441–454, 2001.

Klemes, V., The Hurst phenomenon: A puzzle?, *Water Resour. Res.*, *10*(4), 675–688, 1974.

Koutsoyiannis, D., Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrol. Sci. J.*, *48*(1), 2003.

Maraun, D., H. W. Rust, and J. Timmer, Tempting long-memory - on the interpretation of DFA results, *Nonlin. Proc. Geophys.*, *11*, 495–503, 2004.

Metzler, R., Comment on "Power-law correlations in the southern-oscillation-index fluctuations characterizing El Niño", *Phys. Rev. E*, *67*(018201), 2003.

R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, 2004.

Robinson, P. M., Log-periodogram regression of time-series with long-range dependence, *Ann. Statist.*, *23*(3), 1048–1072, 1995.

Robinson, P. M. (Ed.), *Time Series with Long Memory*, Advanced Texts in Econometrics, Oxford University Press, 2003.

Rust, H. W., Detection of long-range dependence – applications in climatology and hydrology, Ph.D. thesis, Potsdam University, Potsdam, 2007.

Schär, C., P. Vidale, D. L. C. Frei, C. Häberli, M. Lininger, and C. Appenzeller, The role of increasing temperature variability in European summer heatwaves, *Nature*, *427*(6972), 332–336, 2004.

Smith, R. L., Long-range dependence and global warming, *Stat. Env.*, 1993.

Talkner, P., and R. O. Weber, Power spectrum and detrended fluctuation analysis: Application to daily temperatures, *Phys. Rev. E*, *62*(1), 150–160, 2000.

Taqqu, M. S., V. Teverovsky, and W. Willinger, Estimators for long-range dependence: An empirical study, *Fractals*, *3*(4), 785–798, 1995.

Trenberth, K. E., et al., Observations: Surface and atmospheric climate change, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Weber, R. O., and P. Talkner, Spectra and correlation of climate data from days to decades, *J. Geophys. Res.*, *106*, 20,131–20,144, 2001.

Weber, R. O., P. Talkner, and G. Stefanicki, Asymmetric diurnal temperature change in the Alpine region, *Geophys. Res. Lett.*, *21*, 673–676, 1994.

H. W. Rust, Institute of Physics, University of Potsdam, Am Neuen Palais 10, D-14469 Potsdam, Germany. (hrust@unipotsdam.de)

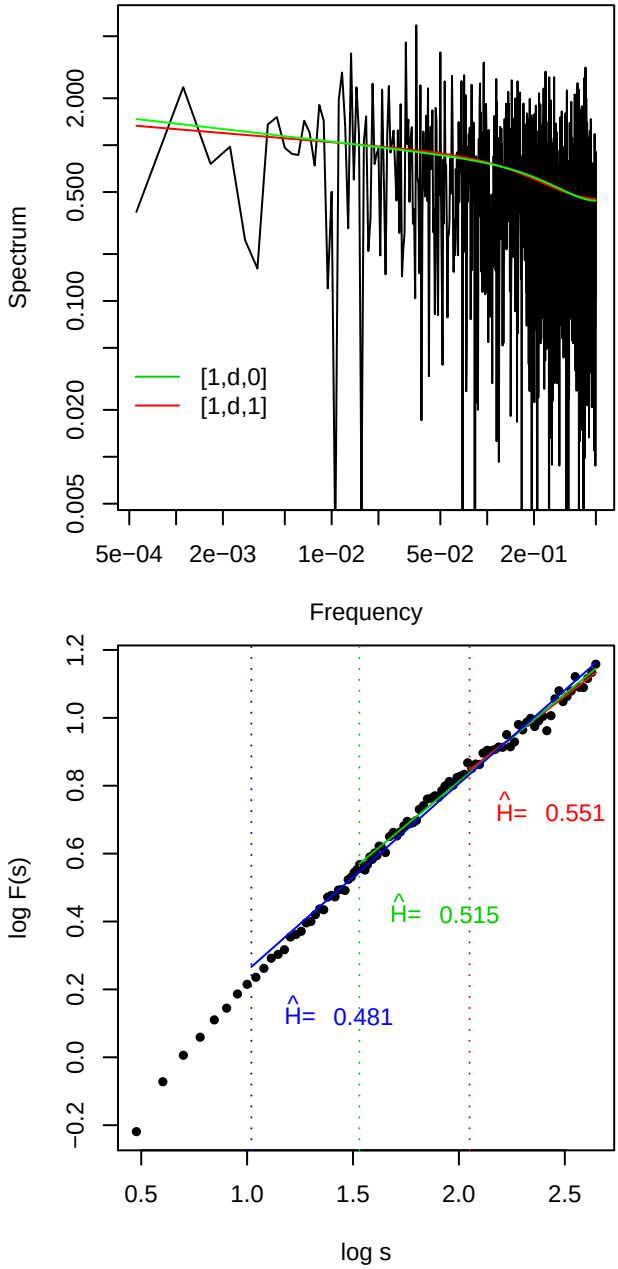

Figure 6. Upper panel: Periodogram of a data set from the simulation study together with the spectral density estimates obtained for a FARIMA$[1, d, 0]$ (green) and a FARIMA$[1, d, 1]$ (red) model, double-logarithmic representation. Lower panel: DFA fluctuation function obtained for a data set from the simulation study, double-logarithmic representation. A slope $\hat{H}_{\mathrm{DFA}}$ is estimated using least squares and the assumption of a linear relationship for $\log F(s)$ on $\log s$ for three different ranges of s: $\log s_{\min} \gtrsim 2$ (red), $\log s_{\min} \gtrsim 1.5$ (green) and $\log s_{\min} \gtrsim 1$; $s_{\max} \approx N/4 = 450$