

J. R. Statist. Soc. A (2020)

How does temperature vary over time?: evidence on the stationary and fractal nature of temperature fluctuations

John K. Dagsvik,
Statistics Norway, Oslo, Norway

Mariachiara Fortuna
Vanlog, Turin, Italy

and Sigmund Hov Moen
Oslo, Norway

[Received September 2016. Final revision January 2020]

Summary. The paper analyses temperature data from 96 selected weather stations world wide, and from reconstructed northern hemisphere temperature data over the last two millennia. Using a non-parametric test, we find that the stationarity hypothesis is not rejected by the data. Subsequently, we investigate further properties of the data by means of a statistical model known as the fractional Gaussian noise (FGN) model. Under stationarity FGN follows from the fact that the observed data are obtained as temporal aggregates of data generated at a finer (basic) timescale where temporal aggregation is taken over a 'large' number of basic units. The FGN process exhibits long-range dependence. Several tests show that both the reconstructed and most of the observed data are consistent with the FGN model.

Keywords: Fractional Gaussian noise; Hurst index; Observed and reconstructed temperatures; Self-similar processes; Temperature change

1. Introduction

Debate about whether or not a change in temperature levels is taking place has been fierce over the past few decades. It seems to be widely accepted now that there has been a systematic change in the temperature process in recent years. However, this is not as easy to demonstrate as might be expected. Recorded series of temperatures typically exhibit local trends and cycles that sometimes seem to persist for up to several decades. The difficulty of settling this matter relates to the fact that the lengths of the available observed time series of temperature data are limited. Few of the recorded temperature series are longer than 250 years. Although 250 years may seem a long time, it is not when it comes to examining properties such as stationarity and long-range dependence.

Recently, there have been attempts to reconstruct temperature data from other sources. One important data set has been obtained by Moberg *et al.* (2005a), who have reconstructed annual temperatures from 1979 back to 1 AD based on information from ice core drillings, tree rings, lake sediments and other sources. The advantage of the reconstructed data sets is that they

Address for correspondence: John K. Dagsvik, PO Box 2633, St Hanshaugen, NO-0131 Oslo, Norway.
E-mail: john.dagsvik@ssb.no

© 2020 The Authors. Journal of the Royal Statistical Society: Series A (Statistics in Society) 0964–1998/20/183000
Published by John Wiley & Sons Ltd on behalf of the Royal Statistical Society. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

cover quite a long time range. However, since they are reconstructions one cannot rule out the possibility that the extent of measurement error might be substantial.

Whereas existing physical models provide good short-term temperature forecasts, they are not reliable for forecasts more than a few days ahead. In the absence of a physical model that is capable of explaining weather dynamics precisely over a longer time range, an interesting question is whether temperature fluctuations are consistent with outcomes of an underlying *stationary* stochastic mechanism (process), and what the features of such a stochastic process might be. In this paper we address several challenges. First, we investigate whether the temperature process is stationary by applying a non-parametric test on observed as well as reconstructed data. Second, we apply theoretical arguments to justify a specific stochastic model for the temperature process. Third, we estimate the parameters of the model by using both observed and reconstructed data. Finally, we apply different approaches to test whether the observed and reconstructed data are consistent with our model.

Statistical time series analysis based on actual recorded temperature time series includes Baillie and Chung (2002), Bloomfield (1992), Bloomfield and Nychka (1992), Galbraith and Green (1992), Koenker and Schorfheide (1994), Harvey and Mills (2001), Kärner (2002), Gil-Alana (2003, 2008a, b), Gay-Garcia and Estrada (2009), Mills (2007, 2010), Estrada *et al.* (2010), Kaufmann *et al.* (2010), Schmith *et al.* (2012), Davidson *et al.* (2016) and Holt and Teräsvirta (2020). Mills (2007) analysed the reconstructed data that were obtained by Moberg *et al.* (2005a).

With a few exceptions, those references applied univariate statistical tools for analysing time series without any *a priori* restrictions derived from theoretical considerations. When the choice between statistical models is based primarily on goodness-of-fit criteria, choosing between specifications that yield virtually the same fit but entail different interpretations and produce substantially different out-of-sample predictions is highly problematic. This difficulty is illustrated by discussions in the recent literature (Kaufmann *et al.* (2010) and Mills (2010), and their references) about whether the temperature process exhibits systematic trends. This difficulty is not likely to be resolved by statistical tests alone. If a precise physical theory were available one might imagine applying a combination of *a priori* physical assumptions and statistical modelling techniques, as has been done by, for example, Aldrin *et al.* (2012), Humlum *et al.* (2013), Schmith *et al.* (2012) and Solheim *et al.* (2011). However, such a strategy has serious drawbacks because existing physical theories are not sufficiently precise to provide reliable quantitative relationships to explain observed temperature fluctuations. Climate models are rough deterministic representations of very complex phenomena and so cannot be expected to give precise answers. Since they are deterministic, they are also not very helpful for choosing between competing statistical formulations.

Motivated by the results from the non-parametric stationarity tests, which show that stationarity is not rejected in most cases, we take a theoretical approach that draws on recent advances in the theory of stochastic processes to justify the model structure before the model is confronted with the data. Our approach is simple in the sense that no explanatory variable is introduced and no deeper explanations of the temperature variations are provided. The justification is based on a key feature of the data, namely that the observed temperature series are obtained as temporal (or can be interpreted as) aggregates of data generated at a finer (basic) timescale where temporal aggregation is taken over a ‘large’ number of basic units. For example, weekly temperatures are obtained as averages of daily (average) temperatures, monthly temperatures as averages of weekly (average) temperatures, etc. This feature is very important because it has a crucial bearing on the structure of the statistical model which is supposed to have generated these data. Under stationarity and specific regularity conditions, it follows from Giraitis *et al.* (2012), pages 77 and 79, that the observed temperature data have, approximately, the properties of a so-called

fractional Gaussian noise (FGN) process (Mandelbrot and van Ness, 1968; Mandelbrot and Wallis, 1968, 1969). Recall that the stationarity assumption is reasonable since the first-stage non-parametric tests show that stationarity is not rejected by the reconstructed—and most of the observed—temperature series. As will be explained later, other tests of the properties of the model (including stationarity) are also applied. As demonstrated by Mandelbrot and Wallis (1969), the FGN model allows for highly irregular patterns of observations. It implies a specific structure of the serial correlation pattern. Thus, our approach avoids *ad hoc* parameterizations of the serial correlation structure that are typical in statistical time series analysis. In our case, where the timescale is based on either ‘month’ or ‘year’ as the unit, although the basic timescale may be ‘day’ (or fractions of a day), we can safely say that temporal aggregation is taken over a large number of units at the basic scale.

We have used data from 96 weather stations and data reconstructed by Moberg *et al.* (2005a, 2006) to test the stationarity hypothesis and to estimate and test the FGN model. We have used different estimation methods and tests. These tests seem to provide convincing evidence to support the assertion that the FGN model is consistent with the major part of the data. What is particularly striking is that the reconstructed data also appear to be consistent with the FGN model.

The paper is organized as follows. In Section 2 the data are described. Section 3 contains a discussion on the justification of the modelling assumptions. In Section 4 we discuss methods for estimation and testing, whereas Section 5 contains results from the estimation and testing. Section 6 is devoted to an analysis of the reconstructed data from the last two millennia (Moberg *et al.*, 2005b).

Additional results are given in four on-line supplementary appendices. The first file contains appendices with further descriptions of the data and estimation results and figures. The second file contains the R code that was used to produce the results in the first file. The third file contains the estimation results of the paper together with the R code that was used to produce the results. The fourth file contains all the functions that were developed for this paper by Mariachiara Fortuna written as plain text.

2. Data

Data on observed temperatures were obtained from various sources (Hov Moen; www.rimfrost.no). These sources are the National Aeronautics and Space Administration Goddard Institute for Space Studies, European climate assessment and data and national meteorological institutes, such as the Swedish Meteorological and Hydrological Institute and the Norwegian Meteorological Institute. The data, certified by the National Aeronautics and Space Administration, comprise time series of temperatures for 1258 weather stations in cities or towns in more than 100 countries. The time series are available as yearly and monthly figures. The lengths of the time series vary greatly across stations. Some, such as Uppsala, contain data for 290 years, with monthly data from 1722 to 2012, apart from some missing observations. Other time series are shorter than two decades. Some of the time series have several periods of missing data. After various selection criteria were applied we ended up with our sample consisting of 96 time series from 32 countries. The time series were selected on the basis of the quality of data, such as the length and number of missing observations. More information about the selection process is available in appendix F in the first on-line resource file.

The first on-line resource file contains plots of the temperature data for nine selected weather stations (Fig. B1) as well as summary information for the 96 weather stations (Table B1). The nine weather stations were selected because they have the longest temperature series with the fewest interruptions. We have used annual data as well as monthly data adjusted for seasonal

variations. Under the Gaussian hypothesis, seasonality can affect only the monthly mean, the variance and possibly the auto-correlation function. By assuming that the auto-correlation function does not depend on seasonal variations, we have adjusted temperatures for seasonal variations by subtracting the respective monthly means from the observations and dividing the observations by the respective monthly standard deviations. The figures of the temperature series show strong cycles and local trends (Fig. 1B and appendix E in the first on-line resource file). (It is likely that there are measurement errors in these temperature series. Some errors are due to the location of the measurement sites, which initially were often to be found in central urban areas. For example, the thermometer that was used for temperature recordings in New York City was previously in Central Park and only recently moved outside the city. It has been documented that the temperatures tend to rise with increasing urbanization. There may also have been variations in the quality of the thermometers that were used over time.)

We have also analysed the data obtained by Moberg *et al.* (2005a, 2006). They reconstructed temperatures for the northern hemisphere from 1 AD to 1979 by using data from borehole drillings, tree rings and lake sediments: see Moberg *et al.* (2005a, b) for a detailed discussion and description of their data; see also the discussion by Moberg (2012). These data show considerable variations over time: there are several cycles, with a high swing occurring from 1000 to 1100 and a low swing occurring from 1500 to 1600: see Fig. 4 in Section 6.

3. Fractional Brownian motion and fractional Gaussian noise

Let $\{X(t), t \geq 0\}$ denote the (observed) temperature process, which we view as a stochastic process in continuous time t . In this section, we provide a more precise formulation of our model and its justification. In what follows, let N denote the set of integers, including 0, and let $[t]$ denote the integer part of t .

Property 1. The process $\{X(t), t \geq 0\}$ is generated as averages of observations at a finer (basic) timescale, i.e.

$$X(t) = \frac{1}{m} \sum_{[mt-m]+1}^{[mt]} \tilde{X}(q)$$

where $\{\tilde{X}(q), q \in N\}$ denotes the process defined at the basic timescale.

Suppose, for example, that the unit of the basic timescale is ‘day’ and that t indexes ‘month’. Then $\tilde{X}(q)$ is the temperature recorded in day q , $m \cong 30$, and $X(t)$ is defined as the temperature of month t . Here, it is understood that the enumeration of the days goes from 1 to mT where T is the length of the time series when using the timescale that corresponds to the observed data, which in this case is month. Alternatively, if instead t indexes year, then $X(t)$ is defined as the temperature of year t . In our case property 1 is simply a formalized statement of how the observed time series (suitably adjusted for seasonal variations) have been constructed.

Hypothesis 1. The process $\{\tilde{X}(q), q \in N\}$ defined at the basic timescale is weakly stationary with constant mean.

Note, however, that in this paper we *do not* assume at the outset that the temperature process is weakly stationary. Thus, hypothesis 1 is a property which we shall test.

To provide an *a priori* theoretical justification for our model, namely the FGN process, let

$$\tau_m^2 = \text{var} \left(\sum_{q=1}^m [\tilde{X}(q) - E\{\tilde{X}(q)\}] \right).$$

If the process $\{\tilde{X}(q), q \in N\}$ is stationary and satisfies some regularity conditions, then it follows from proposition 4.4.1, page 77, and proposition 4.4.4, page 79, in Giraitis *et al.* (2012) that the process $\{V(t), t \geq 0\}$ given by

$$V(t) = \lim_{m \rightarrow \infty} \tau_m^{-1} \sum_{q=1}^{[tm]} [\tilde{X}(q) - E\{\tilde{X}(q)\}]$$

is fractional Brownian motion (FBM): see theorem 3 in Appendix A. The FBM process is Gaussian and has the property known as self-similarity. Remember that a self-similar stochastic process (in continuous time) $\{V(t), t \geq 0\}$ has the property that the joint distribution of $(V(bt_1), V(bt_2), \dots, V(bt_n))$ is equal to the joint distribution of $(b^H V(t_1), b^H V(t_2), \dots, b^H V(t_n))$ for any set of time periods (t_1, t_2, \dots, t_n) , for any integer n , and any positive constant b where H satisfies, $0 < H < 1$ (Samorodnitsky and Taquq (1994), chapter 7). The parameter H is the so-called Hurst index, named after the British engineer Harold Edwin Hurst (1880–1978). One way of explaining the self-similar feature (or fractal feature in a stochastic context) is to say that the distribution of the average process (normalized to have zero mean) up to time bt is, apart from a change of scale, the same as the distribution of the average process up to time t . In other words, the time span over which the average is taken is not essential for the qualitative properties of the probability law of the process. Thus, under self-similarity, a change in the input scale by a factor b will leave the process invariant up to a change in the ‘output’ scale by the factor b^H . The FGN process is defined by $\{V(t) - V(t-1), t \geq 1\}$ where $\{V(t), t \geq 0\}$ is FBM. It follows immediately from the definition above that

$$V(t) - V(t-1) = \lim_{m \rightarrow \infty} \tau_m^{-1} \sum_{q=[mt-m]+1}^{[tm]} [\tilde{X}(q) - E\{\tilde{X}(q)\}]$$

so that, when m is ‘large’, property 1 implies that

$$\tau_m^{-1} \sum_{q=[mt-m]+1}^{[tm]} [\tilde{X}(q) - E\{\tilde{X}(q)\}]$$

is approximately FGN provided that it is weakly stationary. The FGN process has autocovariance function given by

$$\text{cov}\{X(t), X(s)\} = 0.5\sigma^2\{(|t-s|+1)^{2H} - 2|t-s|^{2H} + ||t-s|-1|^{2H}\} \quad (3.1)$$

where $\sigma^2 = \text{var}\{X(1)\}$. We see that the autocovariance function is determined by σ and the Hurst index H . Note that equation (3.1) implies that the auto-correlation function is determined by only one unknown parameter H . It follows from equation (3.1) that when $H=0.5$ the autocovariance of the FGN is 0 and FGN reduces to a process with independent realizations. One can prove (see Samorodnitsky and Taquq (1994), proposition 7.2.10, page 335) that

$$\text{cov}\{X(s), X(t)\} \cong \sigma^2 H(2H-1)|t-s|^{2H-2} \quad (3.2)$$

when $|t-s|$ is large. (When $H \in [0.7, 0.9]$ the auto-correlations obtained from expression (3.2) differ from the corresponding exact values by less than 0.01 when $|t-s| \geq 4$.) This restrictive implication of our approach contrasts with the conventional approach to statistical time series analysis, where no or few *a priori* theoretical principles are invoked. Furthermore, expression

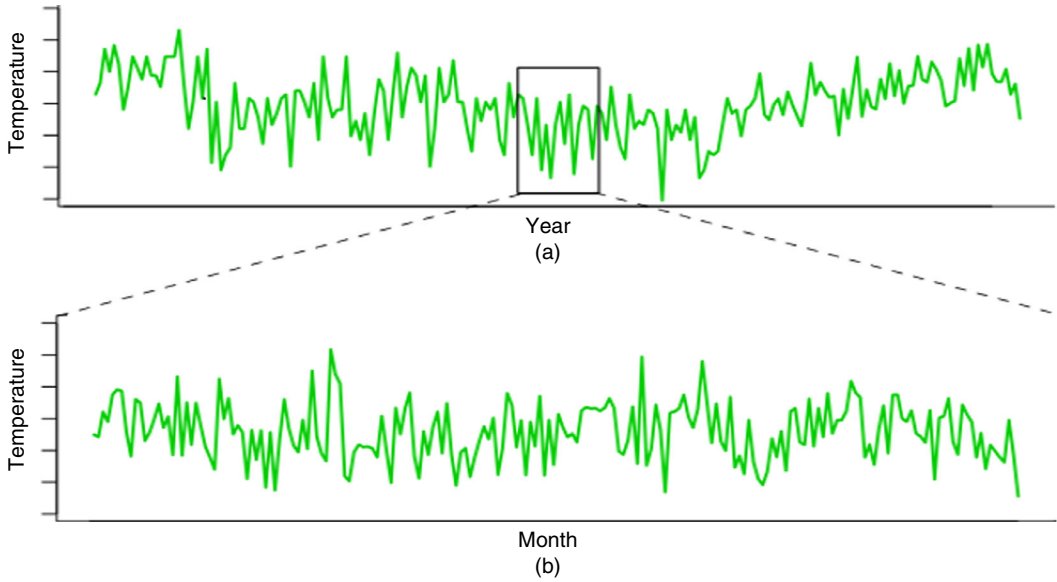


Fig. 1. Illustration of statistical self-similarity: Paris, 1757–2009

(3.2) shows that the auto-correlation function decreases slowly according to a power function (asymptotically). This means that FGN exhibits long-range dependence. In contrast, conventional time series models typically have exponential (or linear combinations of) auto-correlation functions. Moreover, since expression (3.2) implies that the auto-correlation function becomes independent of the timescale we conclude that the FGN process is approximately self-similar.

Fig. 1 provides an intuitive illustration of the self-similar feature of the temperature process for Paris. Fig. 1(a) shows the total series of annual recorded temperatures for Paris, from 1757 to 2009 (253 years). Fig. 1(b) shows the monthly temperatures over 253 months, suitably rescaled. We note that although the two graphs are different they nevertheless appear to have quite similar patterns (fractal feature). (Beran *et al.* (2013) and Mandelbrot and van Ness (1968) have discussed aspects of the fractal feature of self-similar processes and FGN.)

Now we turn to another property of the FGN process that will be useful for interpreting the reconstructed temperature data. The reconstructed temperature data that were obtained by Moberg *et al.* (2005a, b) are based on several sources of data. A key question is therefore whether or not a linear combination of self-similar processes is approximately a self-similar process. For this let $W = r_1 W_1 + r_2 W_2$ where W_1 and W_2 are independent self-similar processes with Hurst indices H_j and r_j , $j = 1, 2$, are weights. We have the following result.

Theorem 1. Suppose that $W = r_1 W_1 + r_2 W_2$ where r_1 and r_2 are constants and W_1 and W_2 are independent self-similar processes with Hurst indices H_1 and H_2 respectively, with $H_1 \geq H_2$. Then $\{W(bt)b^{-H_1}, t \geq 0\}$ converges weakly towards W_1 as $b \rightarrow \infty$.

The proof of theorem 1 is given in Appendix A. The result of theorem 1 means that a linear combination of independent self-similar processes will be approximately self-similar with Hurst index equal to the largest Hurst index of the processes. Accordingly, the sum of FGN processes will also be approximately FGN with Hurst index equal to the largest Hurst index of the original processes.

4. Methods of statistical inference

In the previous section we presented theoretical support for FGN as a model representation of the temperature process under the hypothesis of stationarity (hypothesis 1). Thus, it is of fundamental importance to test hypothesis 1. The theoretical result in support for the FGN model is valid only for large samples (asymptotic results) so it is of interest also to test the properties of the FGN even if hypothesis 1 has been found to hold. There are several methods that can be applied to analyse long memory processes; see Beran *et al.* (2013). In this paper we have used a non-parametric test (Cho, 2016) to test for stationarity, a method based on the empirical characteristic function, the Whittle method and the wavelet lifting method (Knight *et al.*, 2017).

4.1. Method based on the characteristic function

In this section we outline a specific regression approach based on the empirical characteristic function that is used for estimation and for carrying out graphical tests of whether the observed temperature process is consistent with the FGN model. This method is an extension of Koutrouvelis (1980) and Koutrouvelis and Bauer (1982) to the time series setting. In the time series setting an important question is whether the empirical characteristic function is consistent for the corresponding theoretical characteristic function. Feuerverger (1990) has proved strong consistency of the empirical characteristic function in stationary time series under a specific condition on the autocovariance function. This condition does not, however, hold for long memory time series. It is therefore of interest to establish consistency of the empirical characteristic function in our setting. The following theorem yields strong consistency of the empirical characteristic function for long memory Gaussian processes.

Theorem 2. Assume that $\{Z(t), t \in N\}$ is a Gaussian process in discrete time with the property that

$$|\text{cov}\{Z(s), Z(t)\}| \leq K|t - s|^\delta$$

for all $s, t \in N$ where $K > 0$ and $\delta < 0$ are constants and let

$$\hat{\psi}(s, s + T; \lambda) = \frac{1}{T} \sum_{r=s+1}^{s+T} \exp\{i\lambda Z(r)\}$$

and

$$\psi(s, s + T; \lambda) = \frac{1}{T} \sum_{r=s+1}^{s+T} E \exp\{i\lambda Z(r)\}$$

where $i = \sqrt{-1}$. Then

$$\hat{\psi}(s, s + T; \lambda) - \psi(s, s + T; \lambda) \rightarrow 0$$

with probability 1 as $T \rightarrow \infty$.

The proof of theorem 2 is given in Appendix A. Theorem 2 shows that the empirical characteristic function for Gaussian processes is a strongly consistent estimator of the mean characteristic function even when the process $\{Z(t), t \in N\}$ is non-stationary. Regarding the FGN process it follows from expression (3.2) with $\delta = 2H - 2$ that the condition of theorem 2 holds. In the stationary case $\psi(s, s + T; \lambda)$ reduces to $\psi(s, s + T; \lambda) = E[\exp\{i\lambda Z(s)\}]$.

Let $\{Y(t), t \in N\}$ denote the aggregate process in discrete time defined by

$$Y(t) = \sum_{k=1}^t [X(k) - E\{X(k)\}]$$

and let

$$\varphi(t-s; \lambda) = E\left(\exp\left[\frac{i\lambda\{Y(t) - Y(s)\}}{\sqrt{|t-s|}}\right]\right)$$

be the characteristic function of $Y(t) - Y(s)$ for real λ . The reason why we divide the exponent above by $\sqrt{|t-s|}$ is that this will enable us to calculate reliable estimates of the corresponding empirical characteristic function for higher values of λ than otherwise. Define the corresponding empirical counterpart of the characteristic function above by

$$\hat{\varphi}(t-s; \lambda) = \frac{1}{T-|t-s|} \sum_{k=1}^{T-|t-s|} \exp\left[\frac{i\lambda\{Y(|t-s|+k) - Y(k)\}}{\sqrt{|t-s|}}\right]. \quad (4.1)$$

As explained in Appendix A, it follows from equation (4.1) that

$$\begin{aligned} |\hat{\varphi}(t-s; \lambda)| &= \{\hat{\varphi}(t-s; \lambda)\hat{\varphi}(t-s; -\lambda)\}^{1/2} \\ &= \frac{1}{T-|t-s|} \left(\sum_{k=1}^{T-|t-s|} \sum_{r=1}^{T-|t-s|} \cos\left[\frac{\lambda\{Y(|t-s|+k) - Y(k) - Y(|t-s|+r) + Y(r)\}}{\sqrt{|t-s|}}\right] \right)^{1/2}. \end{aligned} \quad (4.2)$$

Formula (4.2) is more convenient than formula (4.1) for calculating numerical values.

Note next that if Z is a normally distributed random variable with mean b and standard deviation σ it is well known that for any real or complex c we have

$$E\{\exp(cZ)\} = \exp(bc + \sigma^2 c^2/2). \quad (4.3)$$

When $\{Y(t), t \in N\}$ is FBM it follows by using equation (4.3) and the self-similarity property that

$$\begin{aligned} \varphi(t-s; \lambda) &= E\left[\exp\left\{\frac{i\lambda Y(|t-s|)}{\sqrt{|t-s|}}\right\}\right] = E[\exp\{i\lambda|t-s|^{H-0.5}Y(1)\}] \\ &= \exp(-0.5\sigma^2|t-s|^{2H-1}\lambda^2). \end{aligned} \quad (4.4)$$

From equation (4.4) it follows that

$$\log\{-\log|\varphi(t-s; \lambda)|\} = (2H-1)\log|t-s| + 2\log|\lambda| + \log(0.5\sigma^2). \quad (4.5)$$

By replacing the theoretical characteristic function with the corresponding empirical characteristic function we obtain from equation (4.5) that

$$\log\{-\log|\hat{\varphi}(t-s; \lambda)|\} = (2H-1)\log|t-s| + 2\log|\lambda| + \log(0.5\sigma^2) + \varepsilon(s, t) \quad (4.6)$$

where $\varepsilon(s, t)$ is an error term that is small when the number of observations is large. Note that the right-hand side of equation (4.6) is linear in $\log|t-s|$ and in $\log|\lambda|$. Thus, equation (4.6) implies that H and σ can be estimated by regression analysis by treating $\log|t-s|$ and $\log|\lambda|$ as independent variables with suitable values of $t-s$ and λ . Koutrovelis (1980) has given a look-up table with recommended values of λ and $t-s$. Kogon and Williams (1998) proposed the use of 10 equally spaced points, ranging from $\lambda=0.1$ to $\lambda=1$. Because the normal distribution is symmetric around zero this is equivalent to using values from $\lambda=1$ to $\lambda=10$. In

our estimation we have used $t_j - s_j = 1, 2, \dots, 20$, for the nine selected cities that are reported below and $t_j - s_j = 1, 2, \dots, 10$, for the remaining weather stations. As regards λ we have chosen the points $\lambda_j = 0.1, 0.2, \dots, 1$, for $j = 1, 2, \dots, 10$. If the temperature process is not stationary $\log\{-\log|\hat{\varphi}(t-s; \lambda)|\}$ will in general not be approximately linear in $\log|t-s|$ and $\log|\lambda|$. When $\{Y(t), t \geq 0\}$ is a self-similar and stable process with stationary stable increments one can show in a similar way that

$$\log\{-\log|\hat{\varphi}(t-s; \lambda)|\} = \alpha(H-0.5) \log|t-s| + \alpha \log|\lambda| + \log(\tau^\alpha) + \varepsilon(s, t) \quad (4.7)$$

where τ is the scale parameter and $\alpha \in (0, 2]$ is the index parameter representing tail thickness in the stable distribution. Thus, if the slope that is associated with $\log|\lambda|$ is found to be less than 2 it indicates that the process is not Gaussian but stable. Under the stationarity hypothesis let $\mu = E\{X(t)\}$. In Appendix A we derive the estimator $\hat{\mu}$ of μ , given in equation (A.10), based on the characteristic-function approach.

Kogon and Williams (1998) have examined the properties of the Koutrouvelis method further in the context of estimating the parameters of stable distributions and found that it performs quite well. However, to the best of our knowledge the characteristic-function-based method that was outlined above has not been applied for estimation of self-similar processes. The characteristic function approach can also be used to carry out graphical tests. This is done by plotting the expression on the left-hand side of equation (4.6) (or (4.7)). If the FGN model is correct this plot will be linear in $\log|t-s|$ and in $\log|\lambda|$, with the coefficient that is associated with $\log|\lambda|$ equal to 2. When producing the plots we have used the same values of λ and $t-s$ as during the estimation.

Recall that even if all one-dimensional marginal distributions are normal it does not necessarily follow that the corresponding joint distribution is multivariate normal. The tests based on the empirical characteristic function that was discussed above can be extended to test whether the joint distribution of the temperatures at several points in time is multivariate normal.

We have conducted a series of bootstrap simulations to check whether or not the distributions of the estimates \hat{H} , $\hat{\sigma}$ and $\hat{\mu}$ that are obtained by the characteristic function procedure are asymptotically normal and unbiased. For this we have computed corresponding Q - Q -plots, which are obtained by bootstrapping based on 1000 simulated time series with length 2000 (Figs D1 and D2 in the first on-line resource file). These plots seem to be almost perfectly linear and clearly indicate that the estimates are normally distributed. However, the estimates of H seem to be downward biased when H is greater than 0.8.

4.2. Test of goodness of fit

We have also used a χ^2 -type statistic, here called the Q -statistic, to examine whether the data are consistent with the FGN model. (Note that the Ljung-Box test cannot be used. According to Chen and Deo (2004), the distribution of this test is not known when the data exhibit long-range dependence.) Let $\{Z(t), t \in N\}$ denote a Gaussian stochastic process with zero mean and let $Z_T = (Z(T), \dots, Z(2), Z(1))'$ be a random Gaussian vector. Let $\Omega_T(H)$ be the covariance matrix of Z_T when $\{Z(t), t \in N\}$ is FGN. In this case it is well known that $Z_T' \Omega_T^{-1}(H) Z_T$ is χ^2 distributed with T degrees of freedom and therefore $Z_T' \Omega_T^{-1}(H) Z_T$ has the same distribution as

$$\sum_{j=1}^T U_{jT}$$

where U_{jT} , $j = 1, 2, \dots, T$, are independent χ^2 -distributed random variables with 1 degree of freedom, $E(U_{jT}) = 1$ and $\text{var}(U_{jT}) = 2$. Consequently, the central limit theorem applies and implies that the distribution of the statistic Q given by

$$Q_T(H) := \frac{Z_T' \Omega_T^{-1}(H) Z_T - T}{\sqrt{(2T)}} \quad (4.8)$$

converges to a standard normal distribution as $T \rightarrow \infty$. Consider the hypothesis H_0 that $\{Z(t), t \in N\}$ is an FGN process against the alternative that the vector Z_T is not Gaussian and/or the covariance matrix is different from $\Omega_T(H)$. Then, H_0 will be rejected if $|Q_T(H)| > \kappa$ where κ is the critical value that corresponds to the selected level of significance.

To check whether $Q_T(H)$ is normally distributed in finite samples with high values of H we have run bootstrap simulations based on the FGN model with sample size $T = 2000$. The results are displayed in Fig. D6 in appendix D in the first on-line resource file. Similarly, we have also used simulations to investigate whether the distribution of $Q_T(H)$ is affected by missing observations. The simulations show that the distribution of $Q_T(H)$, even for high values of H , is still very close to a standard normal distribution under the null hypothesis that the FGN model is the true model and, furthermore, the occurrence of a moderate number of randomly dispersed missing data does not affect the distribution of $Q_T(H)$ (Fig. D5 in appendix D in the first on-line resource file).

5. Empirical results based on the observed temperature data

5.1. Non-parametric testing of stationarity

Several non-parametric tests of stationarity are available. As mentioned above, we have applied the test that was developed recently by Cho (2016). In this method the test statistic is obtained from measuring and maximizing the difference in the second-order structure over pairs of randomly drawn intervals. Cho (2016) proved asymptotic normality of the statistic under specific conditions. These conditions are satisfied for Gaussian and many non-Gaussian processes. Cho (2016) demonstrated that his test has good finite sample performance on both simulated and real data. (We have chosen to apply the test of Cho (2016) because it seems to have good properties and is easy to use. R code is available and is also easy to use.)

For the monthly time series, stationarity is rejected for between about 14 and 18 out of the total of 96 series at a 5% significance level. (The variability in the number of rejections stems from the fact that different subsamples are selected each time that the test procedure is applied.) The number of rejections might vary slightly from one run to another because the intervals are drawn randomly. In contrast, for the annual time series, stationarity is rejected for only one series (Djupivogur, Iceland) at 5% significance level. Note that the monthly and annual data cover the same time range. The reason why stationarity is rejected for some of the monthly series (compared with the annual series) might be that those series are quite long, or, alternatively, that the control for seasonality is insufficient (Tables C6 and C7 in appendix C in the first on-line resource file).

We have also applied the stationary test of Cho (2016) to the Moberg data (Moberg *et al.*, 2005a, 2006). We find that stationarity is not rejected at the 5% significance level (Table C5 in appendix C in the first on-line resource file).

5.2. Inference based on the maximum likelihood, the characteristic function regression and the wavelet lifting methods

In this section we report inference results for nine selected weather stations. These were chosen because they are among those with the longest series and have the fewest missing observations (appendix F in the first on-line resource file). The estimates for the remaining stations are given in Tables C1 and C2 in appendix C in the first on-line resource file. Recall that when the time

series are Gaussian processes the likelihood function is fully identified by the autocovariance function and the mean. Recall that the FGN model is approximately invariant under choice of timescale, i.e. whether one uses observations on annual or monthly data the structure of the auto-correlation function is approximately the same (with the same H). In Table 1 we show estimates of H based on the characteristic function approach, H_C and estimates obtained by the Whittle method H_W : see Beran (1994), chapter 6.1. The Whittle method, which is based on an approximation of the likelihood function, has been found to perform well (Rea *et al.*, 2013). (For long time series the determinant of the corresponding autocovariance matrix of the underlying stochastic process may be close to zero. In such cases exact maximum likelihood estimation often does not work, which is the case in our application. Rea *et al.* (2013) have studied the performance of several estimators of long memory processes and have found that the Whittle method is among the better estimators of those examined. Also, Abry *et al.* (2000) have found that the Whittle estimator may be heavily affected by trends.) The estimates of the standard errors of the characteristic function estimates are obtained by bootstrap simulations. For this we have used 1000 simulations of FGN time series of length 2000 (appendix D in the first on-line resource file). A striking feature of these estimates is that the Hurst index does not vary much across weather stations.

The estimates based on monthly data are quite precise because we have long time series. We note that the maximum likelihood estimates, and the estimates obtained by the characteristic function regression procedure, are quite similar. Since our method for seasonal adjustment may not be optimal, this may produce additional ‘noise’ in the data. One might therefore expect estimates based on annual data to yield higher estimates of H because the presence of random noise might weaken the serial dependence in the seasonally adjusted monthly temperature series. This is in fact what we find, namely that the estimates of H based on annual data often are significantly higher than the estimates based on monthly data.

In Tables D1, D2 and D3 (appendix D in the first on-line resource file) we also report results from bootstrap simulations of the properties of various estimators. These results show that the characteristic function regression estimator seems to be downward biased when H is greater than 0.8, whereas the Whittle estimator appears to be unbiased even for large H , such as $H = 0.95$. We note that for values of H greater than 0.8 all the estimators seem to underestimate σ .

To test for normality of the time series, it is possible to use the Kolmogorov–Smirnov test (Beran (1994), page 199). In this paper we have used the characteristic function approach that was outlined above for this. Recall that under the assumptions of theorem 3 in appendix A it

Table 1. Estimates of the Hurst parameter for selected cities based on characteristic function regression and the Whittle maximum likelihood estimation method: monthly and annual data†

City	H_C	H_W	H_C	H_W
Germany, Berlin	0.664 (0.009)	0.662 (0.012)	0.726 (0.019)	0.712 (0.041)
Switzerland, Geneva	0.693 (0.001)	0.667 (0.012)	0.845 (0.021)	0.818 (0.042)
Switzerland, Basel	0.625 (0.011)	0.622 (0.012)	0.664 (0.019)	0.720 (0.042)
France, Paris	0.733 (0.010)	0.672 (0.012)	0.873 (0.023)	0.802 (0.042)
Sweden, Stockholm	0.681 (0.015)	0.721 (0.012)	0.614 (0.017)	0.632 (0.041)
Italy, Milan	0.724 (0.019)	0.709 (0.012)	0.851 (0.023)	0.826 (0.043)
Czech Republic, Prague	0.684 (0.015)	0.670 (0.012)	0.745 (0.019)	0.716 (0.043)
Hungary, Budapest	0.627 (0.011)	0.645 (0.012)	0.682 (0.019)	0.663 (0.043)
Denmark, Copenhagen	0.755 (0.051)	0.758 (0.013)	0.817 (0.021)	0.753 (0.045)

†Standard errors are in parentheses.

follows that the temperature process is asymptotically Gaussian. It is still of interest to test for normality since we do not know to what extent the assumptions are met in finite samples and with finite m . Under the FGN hypothesis it follows from equation (4.7) that if we select values λ_j as described above and plot the left-hand side of equation (4.7) against $\log(\lambda_j)$ the plot will be linear with the slope (tail index α) close to 2. In Fig. 2 we show corresponding plots for selected cities. From Fig. 2 we see that the plots are almost perfectly linear, with most slopes between 1.99 and 2.01. In one case (Basel) the plot is linear with the slope equal to 1.98, which indicates a stable distribution (with slightly heavier tails than a normal distribution). More results are given in appendix E in the first on-line resource file.

As discussed above, the characteristic function regression approach can also be applied to check whether data indicate that $\{Y(t), t \geq 0\}$ is self-similar with stationary increments, and whether the increments are Gaussian. We have plotted the left-hand side of equation (4.7), for $t = 1, 2, \dots, 10$, against $\log |t - s|$. The resulting plots are shown in Fig. 2 for the cities selected. Additional results are given in appendix E in the first on-line resource file. We note that in most cases the plots are approximately linear. (There is a problem with this test because the calculation of the empirical characteristic function yields imprecise results when λ is large.) We have subsequently applied the Q -statistics defined in equation (4.8) to test whether the FGN model is consistent with the temperature data. Recall that when H is known the test statistic $Q_T(H)$ is, asymptotically, standard normally distributed, which implies that the corresponding 5% confidence interval is equal to $(-1.96, 1.96)$. In Table 2 we present results for the nine cities selected.

We see from Table 2 that when H is estimated by the characteristic function regression method the model passes the test for all nine weather stations, whereas in two cases (Paris and Milan) models that were estimated by the Whittle maximum likelihood method are rejected. For the total data set we have found that in only two cases (Buenos Aires, Argentina, and Cap Otway, Australia) out of the 96 time series are the models that were estimated by the characteristic function regression method inconsistent with the data, whereas in eight cases the models that were estimated by the Whittle maximum likelihood method are inconsistent with the data (Tables C1 and C2 in appendix C in the first on-line resource file). When H is replaced by its corresponding estimate \hat{H} (say), then it is not known what the corresponding distribution of $Q_T(\hat{H})$ is. For this, we have conducted a series of simulation experiments under the assumption that the FGN model is true. These simulations indicate that when H is estimated by the Whittle method the distribution of $Q_T(\hat{H})$ is stable with zero mean and maximally skew to the right. It follows that when we compute confidence intervals based on the unconditional (stable) distribution of $Q_T(\hat{H})$ the model is rejected in only one case (Adelaide) (Fig. D3 in appendix D in the first on-line resource file).

We have also applied the wavelet lifting method to estimate H on the basis of the annual data. Table C8 in appendix C in the first on-line resource file contains estimates of H and of the $Q_T(\hat{H})$ statistic. According to Table C8 the model is rejected for 36 out of the 96 time series. Fig. C1 in appendix C in the first on-line resource file shows that the wavelet lifting method tends to produce lower estimates of H than does the Whittle method. This might be the reason for the relatively high proportion of rejections.

Some of the observed time series suffer from missing observations. To investigate whether the distribution of $Q_T(H)$ is affected by missing observations we have carried out 1000 simulations of FGN processes with $H = 0.8$ and length 2000 and with 70 missing and randomly dispersed data points. This amount of missing data is typical for the 96 selected time series. The plots in Fig. D5 in appendix D in the first on-line resource file show that missing data have no effect in this case.

When looking at some of the observed temperature graphs which exhibit local trends of considerable length it may seem puzzling that a stationary model such as FGN with only three

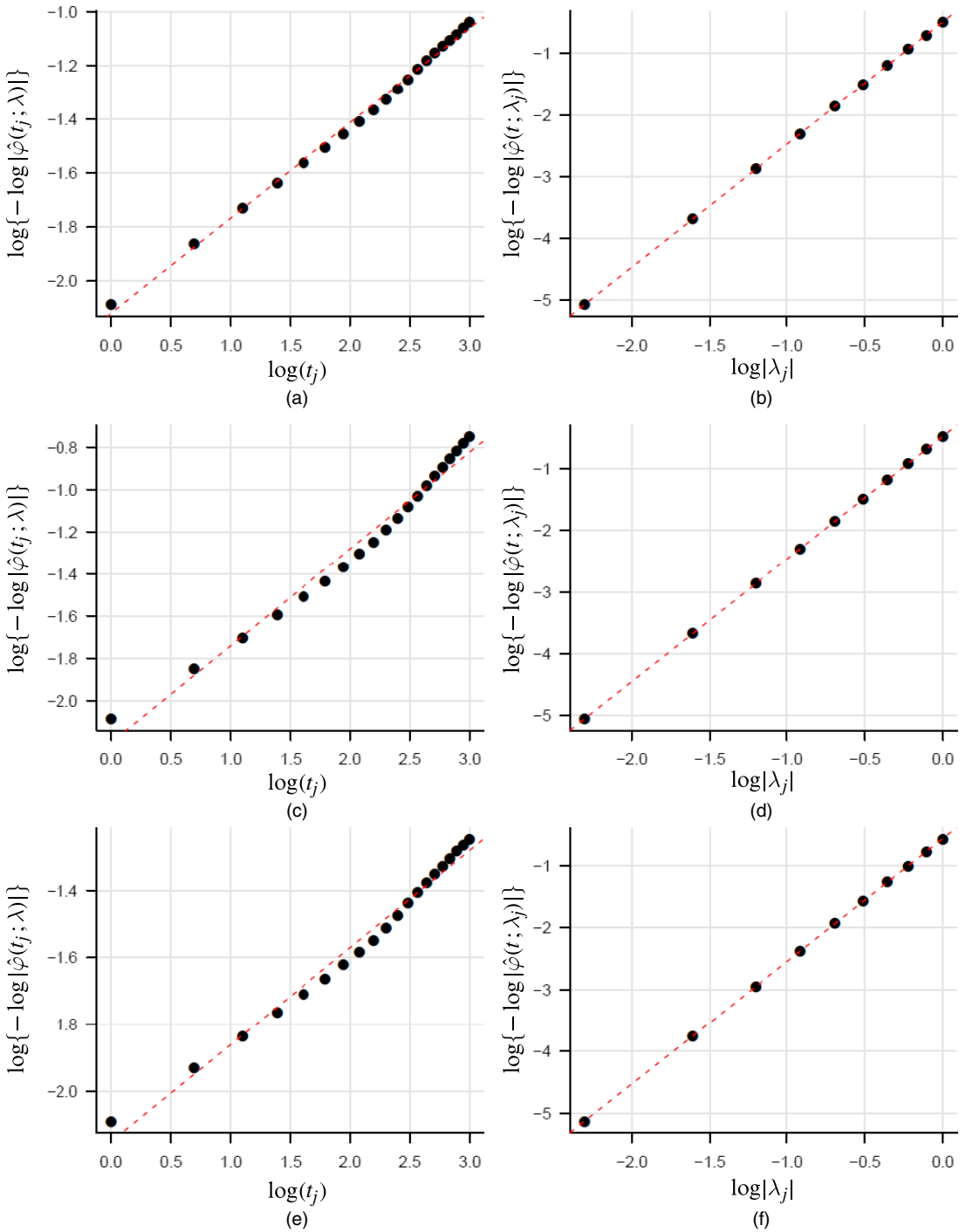


Fig. 2. Graphical tests of the FGN model: (a) self-similarity test, Germany, Berlin; (b) normality test, Germany, Berlin (estimated $\alpha = 1.99$); (c) self-similarity test, Switzerland, Geneva; (d) normality test, Switzerland, Geneva (estimated $\alpha = 1.99$); (e) self-similarity test, Switzerland, Basel; (f) normality test, Switzerland, Basel (estimated $\alpha = 1.98$)

Table 2. Q -statistic under the FGN hypothesis for selected cities

City	$Q(H_C)$	$Q(H_W)$
Germany, Berlin	-0.517	-0.625
Switzerland, Geneva	-0.222	-1.621
Switzerland, Basel	-0.408	-0.483
France, Paris	1.285	-2.790
Sweden, Stockholm	-1.209	1.097
Italy, Milan	-1.338	-2.334
Czech Republic, Prague	-0.197	-0.855
Hungary, Budapest	-0.819	-0.254
Denmark, Copenhagen	-1.006	-0.693

parameters can fit the data. The explanation is that, in the presence of long-range dependence, such patterns are indeed possible. To illustrate the extent of local trends and cycles of the FGN process we have carried out simulations of the FGN process for various values of H : see Fig. 3. Recall that the auto-correlation function of FGN is approximately invariant under choice of timescale: the timescale in Fig. 3 may be 1 year, 10 years or 100 years. However, the corresponding amplitudes will be affected by a change in time unit. If, for example, the timescale is changed from 1 year to 10 years the corresponding standard deviation is found by dividing the standard deviation based on annual data by 10^{1-H} . Recall also that most of the estimates of H based on annual data have orders of magnitude in the interval (0.7, 0.9). With $H = 0.7$ we see from Fig. 3 that from about 625 to about 720 time units there seems to be a decreasing trend, whereas from about 260 to about 330 time units there is an increasing trend. When $H = 0.8$ and $H = 0.9$ this type of pattern seems to be more pronounced. In these cases we note that the local trends may be several hundred time units long.

To investigate whether it is possible to detect departures from stationarity given that the ‘core’ stationary process is FGN, we have conducted a simulation experiment. We have simulated 180 years of the following process: during the first 120 years the temperature process is assumed to be FGN, with zero mean and unit variance; during the last 60 years the temperature process is assumed to be FGN, plus a linear trend with positive slope, starting at 0 in year 120. We used the test based on the Q -statistic to see how steep the trend must be before the FGN hypothesis is rejected. It turns out that the trend must be equivalent to an increase of at least about 2.16°C over 60 years before departure from stationarity can be seen, when $H = 0.7$. When $H > 0.7$ the slope had to be even steeper for departures from the FGN model to be revealed by the Q -test. Thus, this result indicates that the Q -test might have low power for alternatives in the class of models where each member of the class can be broken down into an FGN process plus a deterministic trend. This is hardly surprising given the properties of the FGN process. This stems from the fact that, as H increases, the FGN model exhibits increasingly complex patterns with long stochastic trends and cycles, as seen from the graphs in Fig. 3. However, these are local features that disappear in the long run.

6. Empirical analysis based on two millennia of temperature constructions

In this section we discuss how the FGN model fits the reconstructed data (Moberg *et al.*, 2005a, b). Fortunately, the argument that was used to support the hypothesis that the temperature

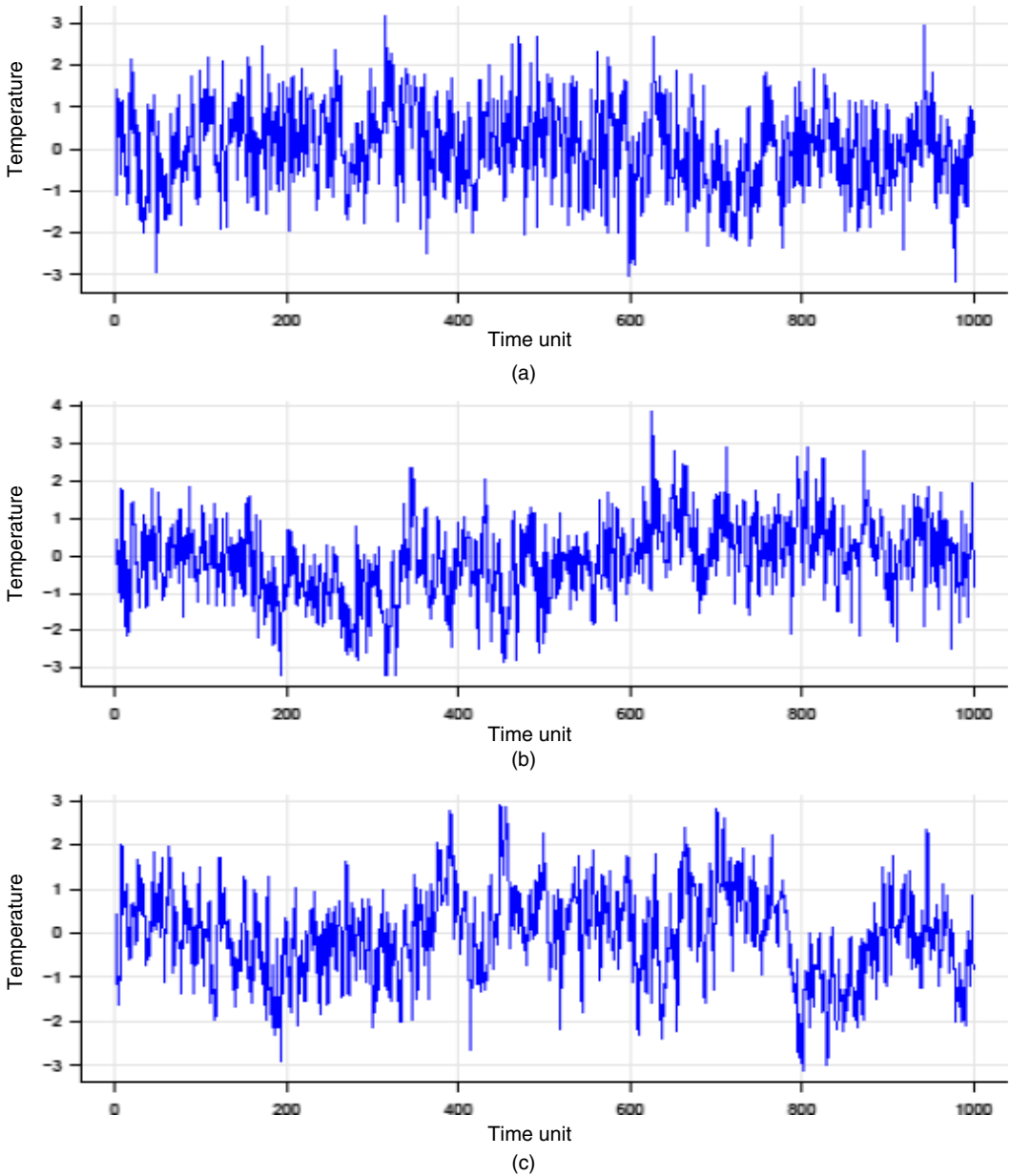


Fig. 3. Simulated FGN process with zero mean and unit variance: (a) $H = 0.7$; (b) $H = 0.8$; (c) $H = 0.9$

process is FGN under stationarity also applies in this case. Recall that the reconstructions of Moberg *et al.* (2005a,b) were based on temperature proxies that were obtained from several sources, such as tree rings, lake sediments and ice core borehole data. According to the technology of temperature reconstructions from tree rings and ice core samples, these data yield annual temperatures which may be interpreted as temporal aggregates of data generated at a finer (continuous) timescale. Consequently, it seems reasonable to assume that property 1 continues to hold also in this case. This may be true even if the observed data contain measurement errors.

Hence, provided that the observed reconstructed series is generated by a stationary process, it follows that it must be approximately an FGN process.

Fig. 4 displays the reconstructed temperatures from 1 AD to 1979. Recall that when applying Cho’s stationarity test the hypothesis of stationarity was not rejected. We have applied the characteristic function regression approach and the Whittle maximum likelihood estimation procedure, as well as tests based on the graphical method. Both the plots of the normality test and the plots of the self-similarity test are close to being perfectly linear and are therefore consistent with the FGN hypothesis: see Fig. 5. The estimate of H by the characteristic function regression method is $\hat{H} = 0.917$. Recall that according to the simulations in Table D2 in appendix D in the first on-line resource file the characteristic function regression estimator underestimates H when $H > 0.8$. The estimate that is based on the Whittle method yields $\hat{H} = 0.990$ with standard error 0.015. By using a test based on the unconditional Q -statistic we find, however, that both the estimate that is based on the characteristic function and the Whittle estimate imply that the

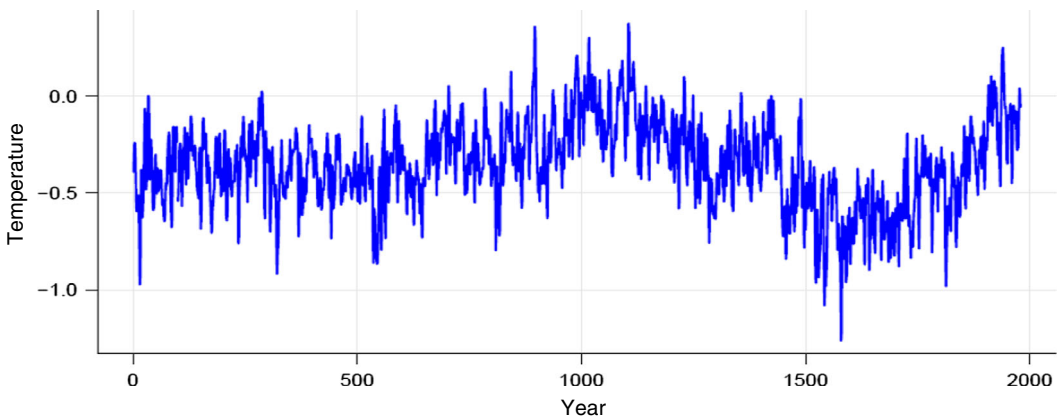


Fig. 4. Reconstructed temperature data by Moberg *et al.* (2005a)

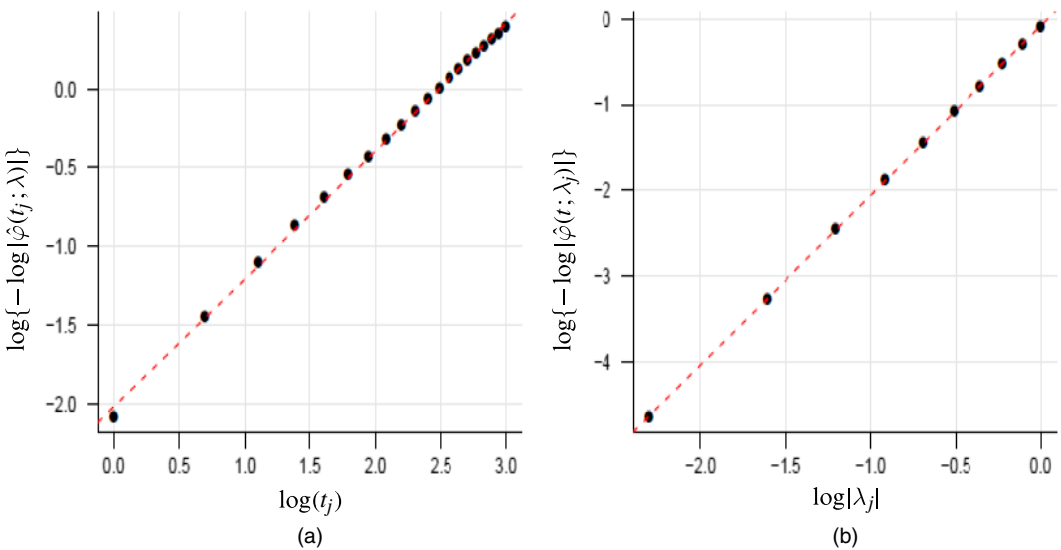


Fig. 5. Tests of the FGN model: (a) self-similarity test (estimated H by regression, 0.92); (b) normality test, (estimated $\alpha = 1.98$)

FGN model is rejected. It turns out that only when H is close to 0.95 is the value of $Q_T(H)$ within the confidence interval $(-1.96, 1.96)$. Thus, only when H is close to 0.95 is the FGN model not rejected, which means that $\hat{H} = 0.95$ might be the better estimate. Recall that when H is replaced by the Whittle estimate \hat{H} the distribution of $Q_T(\hat{H})$ is stable with zero mean, maximally skew to the right with $\alpha = 1.99$ when $H = 0.7$ and $H = 0.8$, and $\alpha = 1.96$ when $H = 0.9$ (Table D3 in appendix D in the first on-line resource file). When $\hat{H} = 0.95$ the 95% confidence interval for $Q_T(\hat{H})$ is approximately $(-6, 12)$. Computing Q for $H = 0.94, 0.95, 0.96$ respectively yields values equal to $-5.27, -0.94$ and 5.60 , which are all within the confidence interval $(-6, 12)$.

In Fig. 6 we have plotted the empirical and the FGN auto-correlation function (with $H = 0.95$) by using the approximation formula (A.15). This formula is a modified version of the formula that was provided by Hosking (1996). The formula that was given by Hosking (1996) does not seem to perform well with high values of H , whereas we have found that the formula that is given in expression (A.15) seems to work well when H is close to 0.95 (appendix A and Fig. D4 in appendix D in the first on-line resource file). We have also provided a confidence band with 2 standard deviations on either side. We see from Fig. 6 that when $H = 0.95$ the FGN model tends to underpredict the empirical auto-correlations of the Moberg data but the differences between the theoretical and empirical auto-correlations are less than 2 standard deviations except for the auto-correlation of lag 1. The standard deviations are obtained by the bootstrap method, based on 1000 simulated time series data from the FGN model with length 2000. From the auto-correlation plot we see that the auto-correlation of lag 1 is significantly higher than that predicted by the FGN model with $H = 0.95$. One possible explanation why the auto-correlation of lag 1 is so high may be measurement error. To realize this, suppose that the reconstructed temperature observations can be expressed as the true temperature values plus measurement error. If the auto-correlation function of the true temperature process is denoted ρ_k and the auto-correlation function of the measurement error process $\{\xi(t), t \geq 0\}$ (say) is denoted γ_k then if the

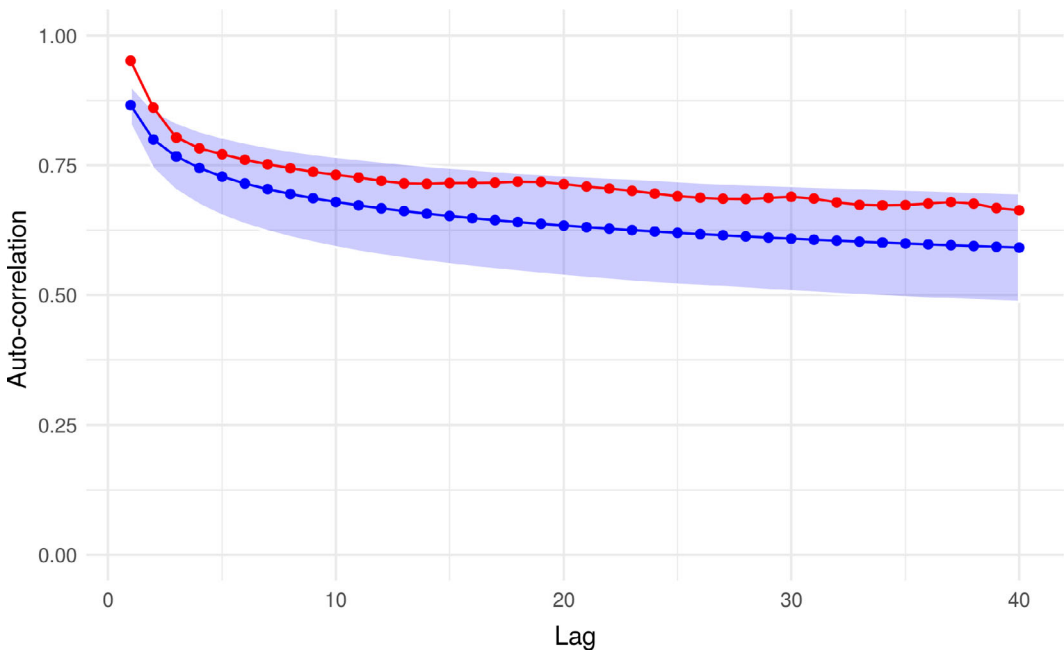


Fig. 6. Empirical and theoretical auto-correlation plots: ●, empirical; ●, FGN model

measurement error process is independent of the true temperature process the auto-correlation of the observed process, $\hat{\rho}_k$, is given by $\hat{\rho}_k = \rho_k + \beta(\gamma_k - \rho_k)$ where

$$\beta = \frac{\text{var}\{\xi(t)\}}{\text{var}\{\xi(t)\} + \text{var}\{X(t)\}}.$$

So if γ_1 is very high, whereas $\gamma_k \cong \rho_k$ for $k > 1$, then we realize that the auto-correlations of lag 1 of the true FGN model and the observed reconstructed temperature process will differ whereas the auto-correlations for higher lags will be approximately equal.

The Whittle method may not be robust with respect to departures from the assumed model, especially when the Hurst index is close to 1 (Abry *et al.*, 2000). To check whether the Whittle estimation method is sensitive to a very high auto-correlation of lag 1 we have estimated the FGN model by using only data for half of the sample, namely for the even and odd years respectively. Since the Whittle method is based on the spectral representation of the model we therefore needed to transform the spectral density of the FGN model to the case where the time unit for the observed series is twice the unit of the original series, whereas we maintain the model that is defined at the original time unit. By using the fact that the autocovariance function is the Fourier transform of the spectral density it follows easily that, if $g(\omega)$ is the spectral density of the FGN process, then $0.5g(0.5\omega)$ will be the spectral density for the model in the case where either the even or the odd years of observations are used. We found that when using the odd years we obtained the estimate $\hat{H} = 0.919$ with standard error equal to 0.021, and when using the even years we obtained the estimate $\hat{H} = 0.932$ with standard error equal to 0.022. We note that the last estimate differs from $H = 0.95$ by only about 1 standard error. These estimation results clearly demonstrate that the Whittle method is not robust against this type of departure from the maintained model.

Next, we shall put forward a possible explanation for why the estimated Hurst index is found to be so high. The argument goes as follows: if the measurement error process and the true (latent) temperature process are independent and additive stationary processes, then it follows from property 1 as discussed above that they are both approximate FGN processes. Hence, theorem 1 implies that the observed process is also approximately FGN with the Hurst parameter equal to the maximum of the Hurst parameters of the true process and the measurement error process. The measurement error process might be highly persistent because of systematic biases in the applied physical methodology linking the observed tree ring and borehole data to the unobserved temperature process. In addition, Moberg *et al.* (2005a) have used wavelet transform techniques to standardize, smooth and interpolate the observed proxies that were obtained from the respective sources to obtain temperature reconstructions. As a result, it may be that the Hurst exponent of the measurement error process is higher than the Hurst index of the latent temperature process and therefore, according to theorem 1, the Hurst index of the reconstructed series will be higher than the Hurst index of the observed temperature series. In addition, the use of interpolation may be one reason why the auto-correlation of lag 1 is particularly high in these data.

Mills (2007) has also analysed the data set that was obtained by Moberg *et al.* (2005a). As with our analysis, he found that these data are consistent with a model that has long memory properties and can be represented by an auto-regressive fractionally integrated moving average process that is both stationary and mean reverting. He also showed that, if we allow for a smoothly varying underlying trend function, then long memory disappears, implying that recent centuries have been characterized by trend temperatures moving upwards. Thus, his analysis provides another example of the difficulty of discriminating between competing models by statistical arguments alone: see Mills (2010).

7. Concluding remarks

In this paper we have provided evidence for why the FGN process is a good representation of the temperature process with the Hurst parameter ranging from about 0.63 to 0.95. Both theoretical arguments and empirical tests that were carried out support our claim. A key feature of the FGN model is long-range dependence. The implication of long-range dependence is that temperature series may exhibit long (non-systematic) trends and cycles. The most extreme case is found in the reconstructed data of Moberg *et al.* (2005a), where some trends last for several hundred years (Fig. 4) but still turn out to be consistent with the FGN model. Furthermore, long-range dependence implies that temperature levels are highly persistent. For example, in the case where the Hurst parameter is as high as 0.9, the correlation between temperatures 100 years apart will be about 0.29. A consequence of this feature of the temperature process is that we cannot achieve reliable claims about systematic changes in temperature levels by using simple statistical trend analyses or *ad hoc* statistical models.

However, it cannot be ruled out that temperature data coupled with other types of data or information, and models based on geophysical processes, might result in a different picture. Hence, it may be that a systematic change in the temperature levels is under way but that our statistical methods are unable to distinguish such changes from natural temperature variation.

Finally, we have conducted a simulation experiment based on a modification of the FGN model allowing for a linear trend with a positive slope during the last 60 years. The goal was to detect how steep the trend must be before stationarity is rejected when using the test based on the Q -statistics.

Acknowledgements

We are grateful for help and support from Mads Greaker, for comments by Arvid Raknerud, Terje Skjerpen and Bjart Holtsmark, and for technical assistance from Gabriele Orlando. We also thank participants in seminars at Statistics Norway, the Frisch Centre for Economic Research, the Department of Mathematics, University of Oslo, and the Department of Biostatistics, University of Oslo.

Appendix A

To make this paper self-contained we restate a very important result which is given in Giraitis *et al.* (2012), page 77 and page 79. For this let N be the set of integers including 0 and let $\{\tilde{X}(q), q \in N\}$ be a weakly stationary process. Define

$$\tau_m^2 = \text{var} \left(\sum_{q=1}^m [\tilde{X}(q) - E\{\tilde{X}(q)\}] \right)$$

and

$$S_m(t) = \sum_{q=1}^{\lfloor mt \rfloor} [\tilde{X}(q) - E\{\tilde{X}(q)\}].$$

Remember that according to Wold's representation theorem every weakly stationary process $\{U(t), t \in N\}$ can be expressed as the sum of a deterministic process and a causal linear process, i.e.

$$U(t) = \xi_t + \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$$

where $\{\xi_t, t \in N\}$ is a deterministic process, $\{\varepsilon_t, t \in N\}$ is a white noise process and $\{a_k\}$ is a sequence with the property

$$\sum_{k=1}^{\infty} a_k^2 < \infty.$$

If $\{U(t), t \in N\}$ is purely stochastic then ξ_t is equal to a constant. Recall also that a function $L(x)$ (say) is slowly varying at ∞ if it has the property that

$$L(tx)/L(x) \xrightarrow{x \rightarrow \infty} 1;$$

see Resnick (1987), page 13.

Theorem 3. Suppose that $\{\tilde{X}(q), q \in N\}$ is a causal linear process that satisfies

$$\tau_m^2 \xrightarrow{m \rightarrow \infty} \infty$$

and

$$\tau_m^2 = m^{2H} L(m)$$

for some constant $0 < H < 1$ where L is a slowly varying function at ∞ . Then the finite dimensional distributions of the process $\{\tau_m^{-1} S_m(t), t \geq 0\}$ converge towards the finite dimensional distributions of FBM with unit variance.

Theorem 3 is a slightly modified version of proposition 4.4.1, page 77, given in Giraitis *et al.* (2012), where also the proof is given. Provided that $H > 0.5$ (which is the case in our analysis) then proposition 4.4.4, page 79, in Giraitis *et al.* (2012) states that $\{\tau_m^{-1} S_m(t), t \geq 0\}$ converges weakly towards the standard FBM with the uniform metric. The FBM has autocovariance function given by

$$\text{cov}\{V(t), V(s)\} = 0.5\sigma^2(t^{2H} + s^{2H} - |t - s|^{2H}). \tag{A.1}$$

A.1. Proof of theorem 1

Let $W(t) = r_1 W_1(t) + r_2 W_2(t)$ where r_1 and r_2 are constants and let $\varphi_j(\lambda_1, \lambda_2; t_1, t_2)$ be the characteristic function of $(W_j(t_1), W_j(t_2))$ and $\varphi(\lambda_1, \lambda_2; t_1, t_2)$ the characteristic function of $(W(t_1), W(t_2))$. If $H_1 = H_2$ then obviously W will be self-similar. Assume next that $H_1 > H_2$. Then it follows that the characteristic function of $(W(bt_1)b^{-H_1}, W(bt_2)b^{-H_1})$ is equal to

$$\begin{aligned} \varphi(b^{-H_1} \lambda_1, b^{-H_1} \lambda_2; bt_1, bt_2) &= E[\exp\{ir_1 b^{-H_1} \lambda_1 W_1(bt_1) + ir_1 b^{-H_1} \lambda_2 W_1(bt_2)\}] E[\exp\{ir_2 b^{-H_2} \lambda_1 W_2(bt_1) \\ &\quad + ir_2 b^{-H_2} \lambda_2 W_2(bt_2)\}] \\ &= E[\exp\{ir_1 \lambda_1 W_1(t_1) + ir_1 \lambda_2 W_1(t_2)\}] E[\exp\{ir_2 b^{-H_1+H_2} \lambda_1 W_2(t_1) \\ &\quad + ir_2 b^{-H_1+H_2} \lambda_2 W_2(t_2)\}] \\ &= \varphi_1(r_1 \lambda_1, r_1 \lambda_2; t_1, t_2) \varphi_2(r_2 b^{H_2-H_1} \lambda_1, b^{H_2-H_1} \lambda_2; t_1, t_2). \end{aligned}$$

When $b \rightarrow \infty$ the last expression tends towards

$$\varphi_1(r_1 \lambda_1, r_1 \lambda_2; t_1, t_2) \varphi_2(0, 0; t_1, t_2) = \varphi_1(r_1 \lambda_1, r_1 \lambda_2; t_1, t_2).$$

Thus we have proved that $(W(bt_1)b^{-H_1}, W(bt_2)b^{-H_1})$ is approximately distributed as $(W_1(t_1), W_1(t_2))$ when b is large. Above we have considered only the bivariate characteristic function of $(W(t_1), W(t_2))$. The argument in the corresponding multivariate case is similar.

A.2. Proof of theorem 2

Let $\mu(t) = E\{Z(t)\}$ and $\text{cov}\{Z(s), Z(t)\} = c(s, t)$. Let

$$X_r = \frac{1}{2}(\exp\{i\lambda Z(r)\} - E[\exp\{i\lambda Z(r)\}]).$$

It follows that

$$|X_r| = \frac{1}{2} |\exp\{i\lambda Z(r)\} - E[\exp\{i\lambda Z(r)\}]| \leq \frac{1}{2} |\exp\{i\lambda Z(r)\}| + E|\exp\{i\lambda Z(r)\}| \leq 1. \quad (\text{A.2})$$

Using equation (4.3) it follows that

$$E[\exp\{i\lambda Z(s) + i\lambda Z(t)\}] = \exp[i\lambda\{\mu(s) + \mu(t)\} - 0.5\lambda^2\{c(s, s) + c(t, t) + 2c(s, t)\}]$$

implying that

$$\begin{aligned} |\text{cov}(X_r, X_t)| &= \frac{1}{4} \exp[-0.5\lambda^2\{c(r, r) + c(t, t)\}] |1 - \exp\{-\lambda^2 c(r, t)\}| \\ &\leq \exp[-0.5\lambda^2\{c(r, r) + c(t, t) - 2|c(r, t)|\}] |1 - \exp\{-\lambda^2 |c(r, t)|\}| \\ &\leq \exp[-0.5\lambda^2 \text{var}\{Z(r) - Z(t)\}] |1 - \exp\{-\lambda^2 |c(r, t)|\}| \\ &\leq 1 - \exp\{-\lambda^2 |c(r, t)|\}. \end{aligned} \quad (\text{A.3})$$

It is easily verified that when x is positive then $1 - \exp(-x) \leq x$. Accordingly, inequality (A.3) and the assumption of theorem 2 imply that

$$\begin{aligned} |\text{cov}(X_r, X_{r+m})| &= \frac{1}{4} |\text{cov}(\exp[i\lambda Z(r)], \exp\{i\lambda Z(r+m)\})| \leq 1 - \exp\{-\lambda^2 |c(r, r+m)|\} \\ &\leq \lambda^2 |c(r, r+m)| \leq \lambda^2 K |m|^\delta. \end{aligned} \quad (\text{A.4})$$

Moreover, since $\delta < 0$ by assumption we have that

$$\sum_{m>1} \frac{\lambda^2 K m^\delta}{m} = \lambda^2 K \sum_{m>1} \frac{1}{m^{1-\delta}} < \infty. \quad (\text{A.5})$$

Consequently, equations (A.2) and (A.5) imply that the conditions of corollary 4 in Lyons (1988) are fulfilled and thus

$$P\left(\sum_{r=s}^{s+T} X_r \xrightarrow[T \rightarrow \infty]{} 0\right) = 1,$$

which is the desired result.

A.3. Details of the regression approach based on the empirical characteristic function

Here, we provide the derivation and justification of the regression procedure based on the empirical characteristic function. Let $\{Y(t), t \geq 0\}$ be a self-similar process with stationary stable increments. Define the characteristic function $\varphi(t-s; \lambda)$ by

$$\varphi(t-s; \lambda) = E\left(\exp\left[\frac{i\lambda\{Y(t) - Y(s)\}}{\sqrt{|t-s|}}\right]\right)$$

for real λ where $i = \sqrt{-1}$. Define the corresponding empirical counterpart of the characteristic function above by

$$\hat{\varphi}(t-s; \lambda) = \frac{1}{T-|t-s|} \sum_{k=1}^{T-|t-s|} \exp\left[\frac{i\lambda\{Y(|t-s|+k) - Y(k)\}}{\sqrt{|t-s|}}\right]. \quad (\text{A.6})$$

Clearly, under self-similarity it follows that

$$E\{\hat{\varphi}(t-s; \lambda)\} = \frac{1}{T-|t-s|} \sum_{k=1}^{T-|t-s|} E\left(\exp\left[\frac{i\lambda\{Y(|t-s|+k) - Y(k)\}}{\sqrt{|t-s|}}\right]\right) = \varphi(t-s; \lambda)$$

which shows that the empirical characteristic function defined in equation (A.6) is an unbiased estimator of the corresponding theoretical characteristic function.

From equation (A.6) it follows readily that

$$\begin{aligned} |\hat{\varphi}(t-s; \lambda)|^2 &= \hat{\varphi}(t-s; \lambda)\hat{\varphi}(t-s; -\lambda) \\ &= \frac{1}{(T-|t-s|)^2} \sum_{k=1}^{T-|t-s|} \sum_{r=1}^{T-|t-s|} \left(\exp \left[\frac{i\lambda\{Y(|t-s|+k) - Y(k)\}}{\sqrt{|t-s|}} \right] - \frac{i\lambda\{Y(|t-s|+r) - Y(r)\}}{\sqrt{|t-s|}} \right) \\ &= \frac{1}{(T-|t-s|)^2} \left(\sum_{k=1}^{T-|t-s|} \sum_{r=1}^{T-|t-s|} \cos \frac{[\lambda\{Y(|t-s|+k) - Y(k)\} - Y(|t-s|+r) + Y(r)]}{\sqrt{|t-s|}} \right). \end{aligned}$$

Since $\{Y(t), t \geq 0\}$ is self-similar with stationary stable increments it follows that

$$\begin{aligned} \varphi(t-s; \lambda) &= E \left[\exp \left\{ \frac{i\lambda Y(|t-s|)}{\sqrt{|t-s|}} \right\} \right] = E[\exp\{i\lambda|t-s|^{H-0.5}Y(1)\}] \\ &= \exp(-\tau^\alpha |t-s|^{\alpha(H-0.5)} \lambda^\alpha) \end{aligned} \tag{A.7}$$

where τ is a dispersion parameter. Equation (A.7) is equivalent to

$$\log\{-\log|\varphi(t-s; \lambda)|\} = \alpha(H-0.5) \log|t-s| + \alpha \log|\lambda| + \log(\tau^\alpha). \tag{A.8}$$

By replacing the theoretical characteristic function by the corresponding empirical characteristic function in equation (A.8) we obtain that

$$\log\{-\log|\hat{\varphi}(t-s; \lambda)|\} = \alpha(H-0.5) \log|t-s| + \alpha \log|\lambda| + \log(\tau^\alpha) + \varepsilon(s, t)$$

where $\varepsilon(s, t)$ is an error term that is small when the number of observation is large.

A.4. Estimate of $\mu = E\{X(t)\}$ when $\alpha = 2$

Let

$$S(\lambda) = E[\sin\{\lambda X(t)\}]$$

and

$$C(\lambda) = E[\cos\{\lambda X(t)\}].$$

Since

$$C(\lambda) + iS(\lambda) = E[\exp\{i\lambda X(t)\}] = \exp(-0.5\lambda^2 + i\lambda\mu) = \exp(-0.5\sigma^2\lambda^2)\{\cos(\mu\lambda) + i \sin(\mu\lambda)\}$$

it follows that

$$C(\lambda) = \exp(-0.5\sigma^2\lambda^2) \cos(\mu\lambda)$$

and

$$S(\lambda) = \exp(-0.5\sigma^2\lambda^2) \sin(\mu\lambda)$$

which yield

$$\tan^{-1}\{S(\lambda)/C(\lambda)\} = \lambda\mu. \tag{A.9}$$

Thus, an estimator of μ can be obtained as follows (Koutrouvelis, 1980). Let

$$\hat{S}(\lambda) = \frac{1}{T} \sum_{k=1}^T \sin\{\lambda X(k)\}$$

and

$$\hat{C}(\lambda) = \frac{1}{T} \sum_{k=1}^T \cos\{\lambda X(k)\}.$$

Evidently, $\hat{S}(\lambda)$ and $\hat{C}(\lambda)$ are consistent estimators for respectively $S(\lambda)$ and $C(\lambda)$. Hence, for a suitable choice of λ , equation (A.9) implies that

$$\hat{\mu}_C = \frac{1}{\lambda} \tan^{-1} \left\{ \frac{\hat{S}(\lambda)}{\hat{C}(\lambda)} \right\} \quad (\text{A.10})$$

is a consistent estimator for μ .

A.5. Maximum likelihood and the Whittle estimator

Under the assumption that the temperature series is a Gaussian process we can also apply the method of maximum likelihood for estimation. Specifically, given the FGN model then the autocovariance function will depend on only two parameters, namely H and σ . Let $\Omega(H)$ be the matrix with elements

$$\Omega_{st}(H) = 0.5(|t-s|+1)^{2H} - 2|t-s|^{2H} + (|t-s|-1)^{2H}.$$

Furthermore, let

$$\mathbf{X}_T = (X(1), X(2), \dots, X(T))'$$

and

$$\mathbf{1} = (1, 1, \dots, 1)'$$

Then the covariance matrix of \mathbf{X}_T can be expressed as $\sigma^2 \Omega(H)$. The log-likelihood function can be written (apart from an additive constant)

$$\log\{L(\mu, \sigma, H)\} = -(\mathbf{X}_T - \mathbf{1}\mu)' \Omega^{-1}(H) (\mathbf{X}_T - \mathbf{1}\mu) / 2\sigma^2 - 0.5T \log(\sigma^2) - 0.5 \log |\det\{\Omega(H)\}| \quad (\text{A.11})$$

where $\mu = E\{X(t)\}$.

In the case where T is large it may be complicated to compute the likelihood function. However, Whittle has demonstrated (see Beran (1994)) that the likelihood can be approximated. This approximate likelihood converges to the exact likelihood as T increases without bounds.

Given H it follows readily from equation (A.11) that the corresponding conditional maximum likelihood estimates of μ and σ are given by

$$\hat{\mu}_{\text{ML}} = \frac{\sum_{k=1}^T \sum_{r=1}^T \Omega_{kr}^{-1}(H) X(r)}{\sum_{k=1}^T \sum_{r=1}^T \Omega_{kr}^{-1}(H)}$$

and

$$\hat{\sigma}_{\text{ML}}^2 = \frac{1}{T} \sum_{r=1}^T \{X(r) - \hat{\mu}\}^2.$$

A.6. Estimation of the auto-correlation function

In the presence of long-range dependence the usual estimator for the auto-correlation function may be seriously biased even if long time series data are available. Let r_k be the usual estimator for the auto-correlation ρ_k as given by

$$r_k = \frac{\sum_{t=1}^{T-k} \{X(t+k) - \bar{X}_T\} \{X(t) - \bar{X}_T\}}{\sum_{t=1}^T \{X(t) - \bar{X}_T\}^2}$$

where \bar{X}_T is the sample mean. Under the condition $\rho_k \cong ck^{-\gamma}$, for large values of k where c is a positive constant and $0 < \gamma < \frac{1}{2}$, Hosking (1996) has obtained that

$$E(r_k) - \rho_k \cong \frac{-2(1 - \rho_k)c}{(1 - \gamma)(2 - \gamma)T^\gamma}. \tag{A.12}$$

In our setting it follows from equation (3.2) that Hosking's condition given above is satisfied and that $c = \sigma^2 H(2H - 1)$ and $\gamma = 2 - 2H$. The approximation in expression (A.12) becomes better with increasing sample size T . Hence, by inserting these values into approximation (A.12) we obtain that

$$E(r_k) - \rho_k \cong \frac{\rho_k - 1}{T^{2-2H}}. \tag{A.13}$$

Expression (A.13) can be applied to obtain an asymptotic unbiased estimator for ρ_k .

Let

$$\hat{\rho}_k = \frac{r_k + T^{2H-2}}{1 + T^{2H-2}}. \tag{A.14}$$

Then it follows from expression (A.13) and (A.14) that $E(\hat{\rho}_k) \cong \rho_k$, which means that $\hat{\rho}_k$ is approximately an unbiased estimator for ρ_k . However, by conducting simulation experiments we have found that formula (A.14) does not produce unbiased estimates when H is large (Fig. D in appendix D in the first on-line resource file). Specifically, we have found that when $H = 0.95$ the estimator

$$\hat{\rho}_k = \frac{r_k + 0.9}{1.9} \tag{A.15}$$

yields approximately unbiased estimates.

A.7. Bootstrap estimation and graphical test for the distributed of $Q_T(\hat{H})$ when \hat{H} is stochastic

When \hat{H} is stochastic it is not known what the distribution of $Q_T(\hat{H})$ is. However, we can apply the Koutrouvelis (1980) approach to conduct a graphical test for the hypothesis that $Q_T(\hat{H})$ follows a stable distribution when \hat{H} is normally distributed. For this let \hat{H}_j be the bootstrap estimate that is obtained from the j th simulated sample, $j = 1, 2, \dots, M$. For given real λ , define the empirical characteristic function of

$$\hat{\psi}(\lambda) = \frac{1}{M} \sum_{j=1}^M \exp\{i\lambda Q_T(\hat{H}_j)\}.$$

Similarly to equation (4.7) it follows from Koutrouvelis (1980) that when $Q_T(\hat{H}_j)$, $j = 1, 2, \dots, M$, are independent and identically distributed according to a stable distribution then

$$\log\{-\log|\hat{\psi}(\lambda)|\} = \alpha \log|\lambda| + \log(\sigma^\alpha) + \varepsilon(\lambda) \tag{A.16}$$

where $\varepsilon(\lambda)$ is an error term that vanishes when M increases without bounds, α is the index parameter, σ is the scale parameter and $|\hat{\psi}(\lambda)|$ can conveniently be computed by using the relationship

$$|\hat{\psi}(\lambda)| = \frac{1}{M} \left[\sum_{k=1}^M \sum_{r=1}^M \cos\{\lambda(Q_T(\hat{H}_k) - Q_T(\hat{H}_r))\} \right]^{1/2}.$$

Similarly to Section 4.1 we can use regression analysis to estimate α and σ . Furthermore, one can use equation (A.16) to conduct graphical tests of the hypothesis that the Q -statistic is stable.

References

Abrý, P., Flandrin, P., Taqqu, M. S. and Veitch, D. (2000) Wavelets for the analysis, estimation and synthesis of scaling data. In *Self-similar Network Traffic and Performance Evaluation* (eds K. Park and W. Willinger), pp. 39–88. Chichester: Wiley.

- Aldrin, M., Holden, M., Guttorp, P., Skeie, R. B., Myhre, G. and Berntsen, T. K. (2012) Bayesian estimation of climate sensitivity based on a simple climate model fitted to observations of hemispheric temperatures and global ocean heat content. *Environmetrics*, **23**, 253–271.
- Baillie, R. T. and Chung, S.-K. (2002) Modeling and forecasting from trend-stationary long memory models with applications to climatology. *Int. J. Forecast.*, **18**, 215–236.
- Beran, J. (1994) *Statistics for Long-memory Processes*. London: Chapman and Hall.
- Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013) *Long-memory Processes*. Berlin: Springer.
- Bloomfield, P. (1992) Trends in global temperatures. *Clim. Change*, **21**, 275–287.
- Bloomfield, P. and Nychka, D. (1992) Climate spectra and detecting climate change. *Clim. Change*, **21**, 1–16.
- Chen, W. W. and Deo, R. S. (2004) A generalized portmanteau goodness-of-fit test for time series models. *Econometr. Theory*, **20**, 382–416.
- Cho, H. (2016) A test of second-order stationarity of time series based on unsystematic sub-samples. *Stat.*, **5**, 262–277.
- Davidson, J. E. H., Stephenson, D. B. and Turasie, A. A. (2016) Time series modeling of paleoclimate data. *Environmetrics*, **27**, 55–65.
- Estrada, F., Gay, G. and Sánchez, A. (2010) A reply to “Does temperature contain a stochastic trend?: Evaluating conflicting statistical results” by R. K. Kaufmann *et al.* *Clim. Change*, **101**, 407–414.
- Feuerverger, A. (1990) An efficiency result for the empirical characteristic function in stationary time-series models. *Can. J. Statist.*, **18**, 155–161.
- Galbraith, J. and Green, C. (1992) Inference about trends in global temperature data. *Clim. Change*, **22**, 209–221.
- Gay-Garcia, C. and Estrada, F. (2009) Global and hemispheric temperatures revisited. *Clim. Change*, **94**, 333–349.
- Gil-Alana, L. A. (2003) Estimating the degree of dependence in the temperatures in the northern hemisphere using semi-parametric techniques. *J. Appl. Statist.*, **30**, 1021–1031.
- Gil-Alana, L. A. (2008a) Time trend estimation with breaks in temperature series. *Clim. Change*, **89**, 325–337.
- Gil-Alana, L. A. (2008b) Warming break trends and fractional integration in the northern, southern, and global temperature anomaly series. *J. Atmos. Ocean. Technol.*, **25**, 570–578.
- Giraitis, L., Koul, H. L. and Surgailis, D. (2012) *Large Sample Inferences for Long Memory Processes*. London: Imperial College Press.
- Harvey, D. I. and Mills, T. C. (2001) Modelling global temperatures using cointegration and smooth transitions. *Statist. Modelling*, **1**, 143–159.
- Holt, M. T. and Teräsvirta, T. (2020) Global hemisphere temperature trends and co-shifting: a vector shifting-mean autoregressive analysis. *J. Econometr.*, **214**, 198–215.
- Hosking, J. R. M. (1996) Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *J. Econometr.*, **73**, 261–284.
- Humlum, O., Stordahl, K. and Solheim, J.-E. (2013) The phase relation between atmospheric carbon and global temperature. *Globl Planet. Change*, **100**, 51–69.
- Kärner, O. (2002) On stationarity and antipersistence in global temperature series. *J. Geophys. Res. D*, **107**, 1–11.
- Kaufmann, R. K., Kauppi, K. and Stock, J. H. (2010) Does temperature contain a stochastic trend?: Evaluating conflicting statistical results. *Clim. Change*, **101**, 395–405.
- Knight, M. I., Nason, G. P. and Nunes, M. A. (2017) A wavelet lifting approach to long-memory estimation. *Statist. Comput.*, **27**, 1453–1471.
- Koenker, R. and Schorfheide, F. (1994) Quantile spline models for global temperature change. *Clim. Change*, **28**, 394–404.
- Kogon, S. M. and Williams, D. B. (1998) Characteristic function based estimation of stable distribution parameters. In *A Practical Guide to Heavy Tails* (eds R. J. Adler, R. E. Feldman and M. S. Taqqu). Boston: Birkhäuser.
- Koutrouvelis, I. A. (1980) Regression-type estimation of the parameters of the stable laws. *J. Am. Statist. Ass.*, **75**, 918–928.
- Koutrouvelis, I. A. and Bauer, D. P. (1982) Asymptotic distribution of regression-type estimators of parameters of stable laws. *Commun. Statist. Theory Meth.*, **11**, 2715–2730.
- Lyons, R. (1988) Strong laws of large numbers for weakly correlated random variables. *Mich. Math. J.*, **35**, 353–359.
- Mandelbrot, B. B. and van Ness, J. W. (1968) Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, **10**, 422–437.
- Mandelbrot, B. B. and Wallis, J. R. (1968) Noah, Joseph and operational hydrology. *Wat. Resour. Res.*, **4**, 909–918.
- Mandelbrot, B. B. and Wallis, J. R. (1969) Some long-run properties of geophysical records. *Wat. Resour. Res.*, **5**, 321–340.
- Mills, T. C. (2007) Time series modelling of two millennia of northern hemisphere temperatures: long memory or shifting trends? *J. R. Statist. Soc. A*, **170**, 83–94.
- Mills, T. C. (2010) ‘Skinning a cat’: alternative models of representing temperature trends. *Clim. Change*, **101**, 415–426.
- Moberg, A. (2012) Comments on “Reconstruction of the extratropical NH mean temperature over the last millennium with a method that preserves low-frequency variability”. *J. Clim.*, **25**, 7991–7997.
- Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M. and Karlén, W. (2005a) Highly variable northern hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature*, **433**, 613–617.

- Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M. and Karlén, W. (2005b) 2,000-year northern hemisphere temperature reconstruction. *Data Contribution Series 2005-019*. World Data Center for Paleoclimatology, Boulder.
- Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., Karlén, W. and Lauritzen, S. E. (2006) Corrigendum: Highly variable northern hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature*, **439**, 1014.
- Rea, W., Oxley, L., Reale, M. and Brown, J. (2013) Not all estimators are born equal: the empirical properties of some estimators of long memory. *Math. Comput. Simulns.*, **93**, 29–42.
- Resnick, S. I. (1987) *Extreme Values, Regular Variation, and Point Processes*. Berlin: Springer.
- Samorodnitsky, G. and Taqqu, M. S. (1994) *Stable Non-Gaussian Random Processes*. London: Chapman and Hall.
- Schmith, T., Johansen, S. and Thejll, P. (2012) Statistical analysis of global surface temperature and sea level using cointegration methods. *J. Clim.*, **25**, 7822–7833.
- Solheim, J.-E., Stordahl, K. and Humlum, O. (2012) The long sunspot cycle 23 predicts a significant temperature decrease in cycle 24. *J. Atmos. Sol. Terrestr. Phys.*, **80**, 267–284.

Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Online resource 01; Online resource 02; Online resource 03; Online resource 04—supplementary appendices’.