Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 355 (2005) 619–632
PHYSICA A
www.elsevier.com/locate/physa

A maximum likelihood estimator for long-range persistence

Alexandra Guerrero[a,1], Leonard A. Smith[a,b,*]

[a]*Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford OX1 3LB, UK*
[b]*Centre for the Analysis of Time Series, London School of Economics, Houghton St., London WC2A 2AE, UK*

Received 17 November 2004; received in revised form 17 February 2005
Available online 9 April 2005

Abstract

A wide variety of processes are thought to show "long-range persistence", specifically an autocorrelation function with power-law decay. A variety of methods have been proposed to quantify this power-law decay, and weather and climate systems, among others, have been claimed to show long-range persistence. In this paper we present a new approach, defining and illustrating a new maximum likelihood estimator of the persistence exponent H. This method provides estimates of H at each time scale considered, as well as meaningful uncertainty estimates. Several independent realisations of processes with a known degree of long-range persistence are used to test the accuracy of the new estimator in terms of spread and bias. The persistence exponent of temperature data is estimated and the problems of using observational data are addressed.

PACS: 02.50.−r; 72.70.+m; 74.40.+k; 02.70.Lq

Keywords: Long-range persistence; Long memory processes; Maximum likelihood estimators; Fractional Gaussian noise; Hurst coefficient; Scaling

*Corresponding author. Centre for the Analysis of Time Series, London School of Economics, Houghton St., London WC2A 2AE, UK.
E-mail addresses: martine1@maths.ox.ac.uk (A. Guerrero), L.Smith@lse.ac.uk (L.A. Smith).
[1]Present address: CCCma Canadian Centre for Climate Modelling and Analysis, University of Victoria, P.O. Box 1700, Victoria B.C., Canada V8W 2Y2.

0378-4371/$ - see front matter
doi:10.1016/j.physa.2005.03.002

1. Introduction

Persistence analysis has become a widely studied subject as it is claimed that a wide variety of physical and biological systems exhibit long-range persistence. Examples of these include geophysical data such as wind speed [1], temperature measurements [2,3], river flows [4,5], heart rate and blood pressure [6,7], DNA structure [8,9], among others. Long-range persistence is characterised by a power law autocorrelation function. This kind of autocorrelation function denotes a slow decay of the autocorrelations which indicates that observations well separated in time are linearly correlated. This property is not only called long-range persistence [3], but also known as long-range memory [10] and long-range dependence [11]. A power law autocorrelation function, $C(l)$, can be expressed in terms of the time lag l as,

$$C(l) \sim k l^{2(H-1)}, \quad 0.50 \leqslant H < 1.0\,, \tag{1}$$

where k is a constant and $\sim$ (read "scales as") reflects the fact that equality in this relationship holds only in theory. In practice, sample correlation functions, $C(l)$, can only "approximate" $kl^{2(H-1)}$. H, referred here as the *persistence exponent*, measures the strength of the autocorrelations in the time series and quantifies the decay. To give a point of reference, compare Eq. (1) to the autocorrelation function of a process with short memory, such as an autoregressive processes, where the autocorrelation function decays exponentially to zero, $C(l) \sim \exp(-l/l_c)$ (where l_c is a constant).

Direct estimation of the persistence exponent H from its autocorrelation function has been shown to be inaccurate [7,12,13], thus more "robust" methods have been adopted. These methods measure the correlations in the data as a function of time lag in a variety of different ways and have been applied to a wide range of data. For a review of some of these methods see [7,13,14].

Several of these methods estimate the persistence exponent by using a power law relation between the time lag l and some statistic, $F(l)$, that measures the degree of correlation in the data as a function of the time lag. The detailed form of $F(l)$ is determined by the method used to quantify the autocorrelation as a function of lag, but in general this relation has the form:

$$F(l) = k l^{g(H)}\,, \tag{2}$$

where k is a constant and g is a linear function of H. The persistence exponent H can be estimated from Eq. (2) by calculating the slope of the best fit line of $\log(l)$ vs. $\log(F(l))$. We refer to this approach as the *log–log approach*. While calculating the slope of the best fit line of $\log(l)$ vs. $\log(F(l))$ appears straight-forward, there are several problems with this approach; a major one being the difficulty of finding a scaling range. This is, of course, a long standing problem in log–log approaches to estimating other scaling exponents and dimensions [15,16]. A new method introduced in this paper overcomes some of these problems by using a maximum likelihood estimator (MLE) of H.

Section 2 presents this method. Section 3 shows some examples where H is estimated using this method and synthetic data where the true value of H is known.

The MLE is also used to estimate H for a temperature record at St. Petersburg and the issue of estimating H of 'real-life' data (hereafter, observational data) is addressed. A discussion of the advantages and disadvantages of the new method introduced are discussed in Section 4. Section 5 gives the final conclusions as well as some possibilities for future work.

2. A maximum likelihood estimator for H

According to random walk theory, for a time series with a power law correlation function $C(l)\sim l^{2(H-1)}$, the fluctuations $F(l)$ for a time lag of length l increase as the power law relation [11,17],

$$F(l) = cl^H, \quad 0.50 \leqslant H < 1.0\,, \tag{3}$$

where c is a constant and H is the persistence exponent. $F(l)$ can be estimated using a simple method referred as standard fluctuation analysis by Koscienly-Bunde et al. [3].

Let X_i, $i = 1,\ldots,N$ be a time series from a stationary process thought to display long-range persistence and with running sum $Y_m = \sum_{i=1}^{m}(X_i - \overline{X})$ where $\overline{X}$ is the sample mean. Firstly, divide the time series into bins of length lag l, $\{Y_1,\ldots,Y_{l+1}\}$, $\{Y_{l+2},\ldots,Y_{2(l+1)}\},\ldots$ and consider the observations at both ends of the non-overlapping bins, Y_t and Y_{t+l}. $F^2(l)$ (in Eq. (3)) is then estimated as the average of the squared differences of the observations at both ends of the bins,

$$\widehat{F}^2(l) = \frac{1}{n(l)}[(Y_1 - Y_{l+1})^2 + (Y_{l+2} - Y_{2(l+1)})^2 + \cdots]\,, \tag{4}$$

where $n(l)$ is the number of non-overlapping bins of length l and the differences $Y_t - Y_{t+l}$, $t = 1, l+2,\ldots$ are essentially the deviation of an l point moving average of $X(t)$ from the overall mean of $X(t)$: $Y_t - Y_{t+l} = l\overline{X} - \sum_{i=1}^{l} X_{t+i}$. H is then usually estimated as the slope of the best fit line of $\log(l)$ vs. $\log(\widehat{F}(l))$ [3,18]. The method now presented estimates H using a maximum likelihood approach.

Now consider a method which estimates H using a maximum likelihood approach. Let X_t be a time series from a process presumed to exhibit long-range persistence and Y_t be its running sum as defined above. For a fixed lag l, assume that the increments $Y_t - Y_{t+l}$, $t = 1,\ l+2,\ 2l+3,\ldots$ are independent and identically distributed as a Gaussian with mean zero and variance σ_l^2,

$$Z_{i,l} = Y_t - Y_{t+l} \sim N(0,\sigma_l^2), \quad i = 1,2,\ldots,n(l) \text{ for lag } l\,, \tag{5}$$

where σ_l^2 is the variance which can be estimated as,

$$\widehat{\sigma}_l^2 = \frac{1}{n(l)}\sum_{i=1}^{n(l)} Z_{i,l}^2 = \frac{1}{n(l)}[(Y_1 - Y_{l+1})^2 + (Y_{l+2} - Y_{2(l+1)})^2 + \cdots] \tag{6}$$

which for processes with long-range persistence (from Eqs. (3) and (4)) is

$$\sigma_l^2 = c^2 l^{2H}\,. \tag{7}$$

Define J_l as the squared sum of the $n(l)$ Gaussian independent and identically distributed random variables (the $Z_{1,l},\ldots,Z_{n(l),l}$ in Eq. (5) with expected mean zero) divided by their variance,

$$J_l = \frac{1}{\sigma_l^2}\sum_{i=1}^{n(l)}\frac{Z_{i,l}^2}{\sigma_l^2} = \frac{1}{\sigma_l^2}\sum_{i=1}^{n(l)} Z_{i,l}^2 = \frac{1}{c^2 l^{2H}}\sum_{i=1}^{n(l)} Z_{i,l}^2 \sim \chi^2_{n(l)-2}\,. \tag{8}$$

J_l is the sum of $n(l)$ χ^2 (chi-squared) random variables with one degree of freedom and thus has a χ^2 distribution with $n(l)-2$ degrees of freedom. Notice that we subtract two degrees of freedom as two parameters will be estimated [19].

Using this information it is possible to calculate the MLE of H for a fixed lag l. The likelihood function, which is the joint probability density function of the sample $z_{1,l},\ldots,z_{n(l),l}$, is the probability density function of a χ^2 variable J_l and is given by

$$P(z_{1,l},\ldots,z_{n(l),l}|H) = \frac{j_l^{((n(l)-2)/2)-1}\mathrm{e}^{-j_l/2}}{2^{(n(l)-2)/2}\Gamma(\frac{n(l)-2}{2})} \tag{9}$$

with $j_l = 1/c^2 l^{2H}\sum_{i=1}^{n(l)} z_{i,l}^2$ as in Eq. (8). For a given l, maximising the probability given in Eq. (9) yields the MLE of H at that lag l.

In the derivation above it is implicitly assumed that the value of the constant c in the expression $\sigma_l^2 = c^2 l^{2H}$ is known. In fact, it can be easily calculated using the observations at lag $l = 1$,

$$\widehat{\sigma_1}^2 = \frac{1}{n(1)}\sum_{t=1}^{n(1)}(Y_t - Y_{t+1})^2 = \widehat{c}^2 1^{2H} = \widehat{c}^2\,. \tag{10}$$

This MLE, $\widehat{H}$, for H has some important advantages. Firstly, it provides the estimate $\widehat{H}(l)$ (i.e. $\widehat{H}$ at lag l) which is the most likely value given the specific sample. Secondly, it provides an estimator for each time lag, l, which allows for the extent to which the long-range persistence extends to be evaluated. This avoids the need to find a scaling range. Finally, its main advantage is that with the probability density function Eq. (9) it is not only possible to obtain the value of the persistence exponent $\widehat{H}(l)$ that maximises this probability, but also a measure of the uncertainty of the estimates by providing a distribution of $\widehat{H}(l)$. This is particularly relevant in this context as the uncertainty of the estimators strongly depends on the time scale.

3. Results

In order to illustrate the maximum likelihood estimation approach presented in the previous section, we used synthetic data with a known degree of long-range persistence, that is, a known value of H. The time series consisted of 65,536 data points ($N = 2^{16}$) with $H = 0.65$. The method used to generate this data is a widely known and used spectral method [14]. After estimating the constant c using Eq. (10), we obtained the distribution of $\widehat{H}(l)$ for a few time lags $l = 10, 300, 3000, 10,000$ using Eq. (9). Note that the sample $z_{1,l},\ldots,z_{n(l),l}$ (and as a consequence j_l) is

influenced by the time lag l. The probability density function in Eq. (9) is a smooth, continuous function, thus it can be easily evaluated for $0.5 \leqslant H < 1.0$ to find the maximum and the distribution $Prob(\widehat{H}(l))$.

The results are shown in Fig. 1 for a few time lags. The tightest distribution corresponds to lag $l = 10$ and the distributions increase in spread as the lag increases. As expected the uncertainty associated with the estimates varies substantially as a function of the time scale. This is due to the fact that the size of the sample $z_{1,l}, \ldots, z_{n(l),l}$ decreases as the time lag increases because the number of points separated by increasing time lags decreases as the time lag increases. The maximum likelihood estimates of H (i.e., the values of $\widehat{H}$ with the greatest likelihood) also vary with the time lags but they all lie between 0.64 and 0.66. The spread of the estimators for longer lags (e.g. lag $l = 3000$ or 10,000) is considerably large and includes values far from the true value of $H = 0.65$. This is not discouraging, on the contrary one of the strong points of this method is that it provides a measure of the uncertainty associated with the estimator at each time lag: given only 2^{16} data points, the true value of H is not constrained more accurately than this.

In Fig. 1 the estimates of H for just a few time lags were shown, in Fig. 2 the estimates for all lags considered are shown. This plot shows the results for the same time series used in Fig. 1 but in this case, instead of all the distribution, only $\widehat{H}(l)$ that maximises the likelihood function Eq. (9) (middle line) for all the time lags as well as their 95% confidence interval (exterior lines) are plotted.

It is clear from this figure that the MLE correctly estimates H as $\widehat{H}(l)$ lies on or closely around the true value $H = 0.65$. Note that the 95% confidence intervals are not symmetrical. This is due to the fact that the likelihood function Eq. (9) is a χ^2

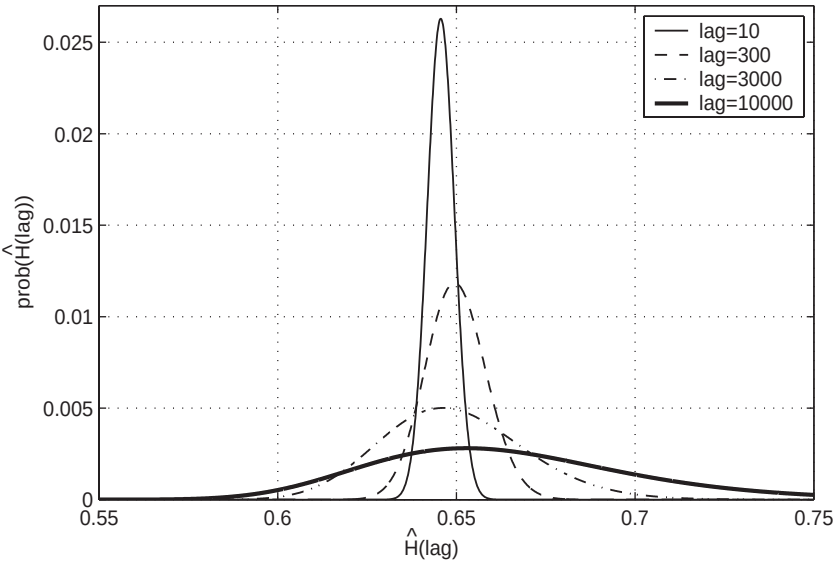

Fig. 1. Maximum likelihood function for a time series with $H = 0.65$. Each distribution corresponds to the M.L. function for lags $l = 10, 300, 3000, 10,000$.

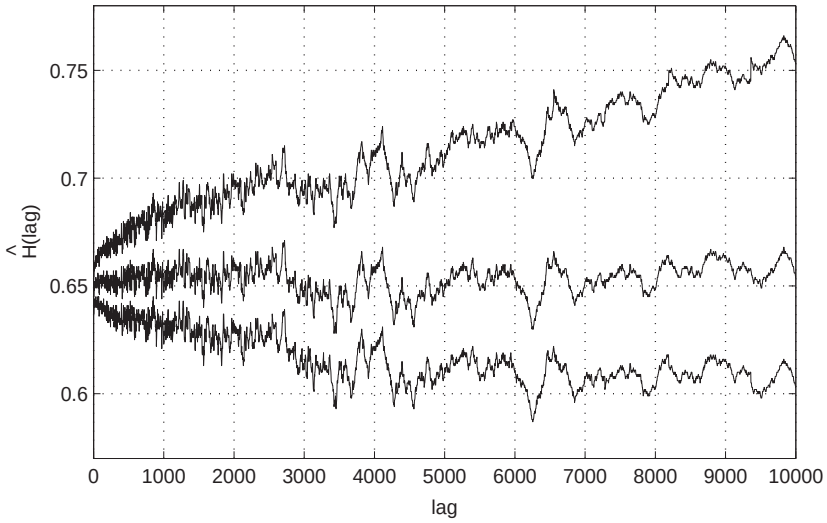

Fig. 2. $\widehat{H}(l)$ as a function of time lag (middle line) with 95% confidence intervals (exterior lines) for a time series with $H = 0.65$ and $N = 2^{16}$.

distribution (in general, this distribution is not symmetrical). In both figures the uncertainty associated with the estimators increases with the size of the time lags; this is simply because for any fixed value of N, the counting statistics will degrade as l increases. Thus, statistics of the longer lags have the greatest uncertainty. Unfortunately, these may in fact be the statistics that we are most interested in; those of large time lags as they are more relevant when measuring the degree of long-range persistence in a process as they quantify the degree of correlation between observations well separated in time. The time scales of most interest contain the least information. This problem is not exclusive to this method, any estimator which somehow measures the correlation between points that are increasingly further apart, will have fewer samples (when the size of the data set is fixed) thus the statistics will inevitably have greater uncertainty [16].

At the other end, the estimates for smaller lags have smaller uncertainty as the lag size becomes smaller since their estimation is based upon more observations. These statistics, however are less relevant in measuring long-range persistence. In some cases, for example, there exists a clear scaling for smaller length scales but not for larger time lags casting doubts on whether the departure from the scaling law for large lags is caused by lack of counting statistics or if there simply is not any long-range persistence. The MLE presented here provides an estimate of H for each time lag, thus it makes it possible to easily identify time scales where there is arguably no power law relation.

In order to test the accuracy of the MLE we use a Monte Carlo approach. We construct several independent realisations of a process with a known strength of long-range persistence and compare their ML estimates of H with the true value. Fig. 3 shows the results for $H = 0.5, 0.65, 0.75$ and 0.90. For each H, 64 independent

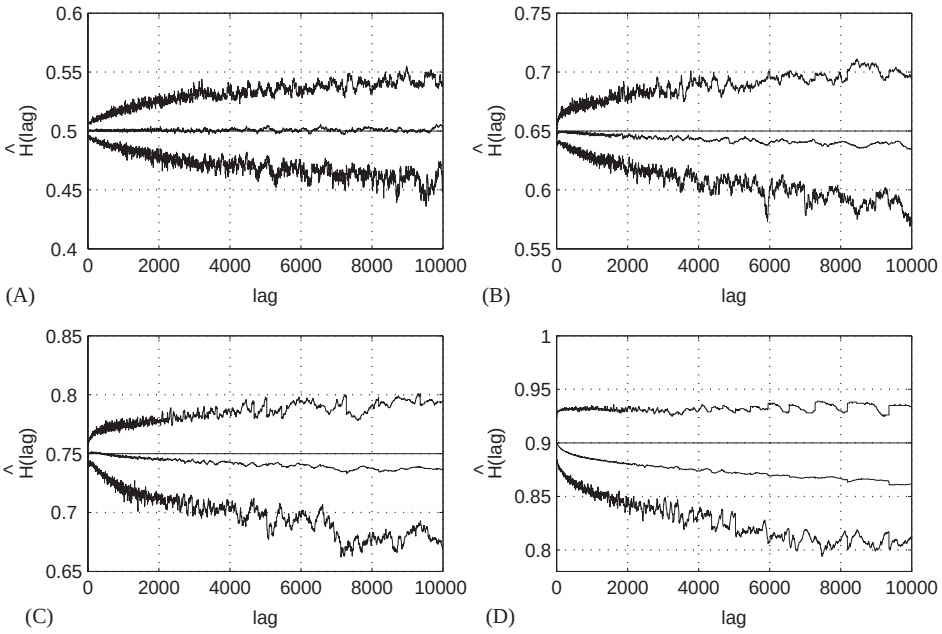

Fig. 3. Mean and 95% of $\widehat{H}(l)$ over 64 realisations for (A) $H = 0.50$, (B) $H = 0.65$, (C) $H = 0.75$ and (D) $H = 0.90$.

realisations of size 65,536 were used and $\widehat{H}(l)$ was computed for lags $l = 2, \ldots, 10,000$. The inner lines show the mean over 64 estimates as a function of time lag and the outer lines show the 95% of these estimates. The mean $\widehat{H}(l)$ over the 64 realisations correctly estimates the true value of H, particularly at small length scales. For the estimates of $H = 0.75$ and 0.90 there is a small bias for large lags; the mean over the 64 estimates underestimates the true value. This also causes the spread to be asymmetrical around the true value. This bias is in fact an artifact of how $\widehat{F}(l)$ is defined in Eq. (4) and not a deficiency of the MLE itself [18]. Despite the small bias for large values of H, the new method here proposed has many advantages which make it a valuable tool for characterising long-range persistence.

In previous examples we estimated the persistence exponent of synthetic data with a known degree of long-range persistence using the MLE proposed in Section 2. Now we focus our attention on some observational data.

3.1. St. Petersburg's temperature

The weather is claimed to be a system with long-range persistence and the persistence exponent has been estimated from temperature records at a variety of meteorological stations around the globe [2,3]. Here we analyse temperature data from St. Petersburg and discuss some of the problems encountered when using observational data. The data consist of daily mean temperatures for the years 1881–1994 ($N = 41,593$). We estimate H for the deseasonalised data using the MLE

and the results are shown in Fig. 4. The middle thick line corresponds to the ML estimates of H at each lag and the exterior thick lines indicate the 95% confidence intervals. Note that there is considerable variability in $\widehat{H}(l)$ and there are various reasons for this. This time series is observational data and thus observational noise is inevitable (in particular when compared to the synthetic data of previous examples). Also, the length of the time series is significantly smaller than the ones used in the previous examples, so greater uncertainty in $\widehat{H}(l)$ is expected. At small time scales the estimated value of $H(l)$ is higher than for larger lags. This indicates the presence of strong short range correlations [15].

Fig. 4 also shows that the 95% confidence intervals (thick exterior lines) for the estimates $\widehat{H}(l)$ expand over a wide range of values (i.e., $\widehat{H} \in [0.61, 0.85]$) which complicates the task of determining just one persistence exponent for the data. In order to illustrate this difficulty we also show $\widehat{H}(l)$ for 64 realisations of synthetic data with long-range persistence and known values $H = 0.65$ and $H = 0.80$ (thin lines). The length of these time series is the same as the temperature data ($N = 41,593$) and we only show the 95% of the estimates for the 64 realisations. It can be seen that the confidence intervals of the St. Petersburg data expand well inside the estimates of the synthetic data. This indicates that the observational data are consistent with a variety of processes with $H \in [0.65, 0.80]$ and that the persistence exponent of the temperature data is around this broad range. Thus, by comparing the temperature data with several realisations of synthetic data with two distant values of H (i.e., 0.65 and 0.80), it has been demonstrated the difficulty of estimating the persistence exponent with observational data. Furthermore, considering the

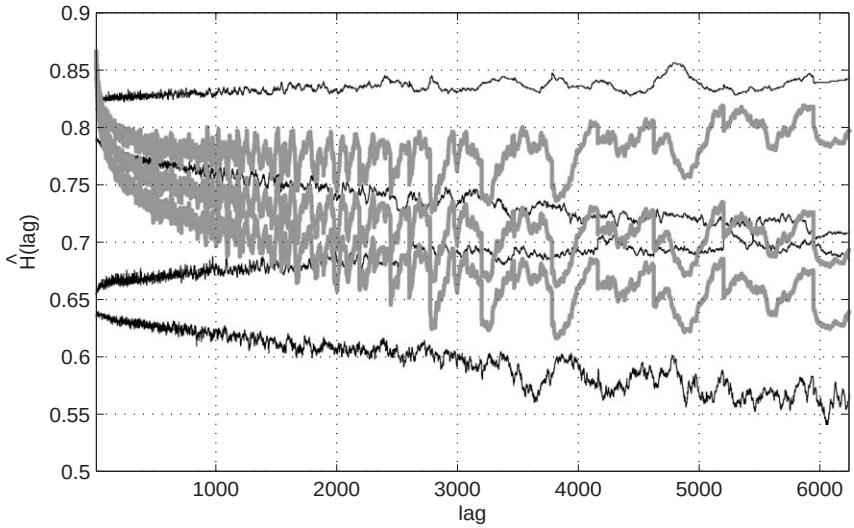

Fig. 4. $\widehat{H}(l)$ as a function of time lag (grey thick middle line) with 95% confidence intervals (grey thick exterior lines) for St. Petersburg data. Thin lines: 95% of $\widehat{H}(l)$ over 64 realisations for $H = 0.65$ and $H = 0.80$.

width of the confidence intervals and the great variability of $\widehat{H}(l)$ for different lags l, it is clear that it is meaningless to choose only one value as an estimate of the persistence of the data and that it is crucial to always provide a measure of the uncertainty of the estimates.

3.2. Detecting long-range persistence

In previous examples we dealt with the question of how to best estimate the persistence exponent H. This was done based on the assumption that the processes analysed indeed present long-range persistence. In practice, it is not always the case that one knows this to be true. Therefore, it is important that the methods for estimating the persistence exponent are able to detect whether a time series has long-range persistence or not. The kind of processes that can present a challenge are those which have strong short term correlations and the difficulty is to identify whether these correlations extend to longer time lags. A possible first step at assessing whether the process has long-range persistence is to examine the autocorrelation plot of the time series and look for significant correlations at large time lags. With observational (noisy) data and data with strong short term correlations it is unlikely that there will be a clear cut-off where short term correlations are present but long-range correlations are not significant. Nevertheless, this can be used as a first exploratory tool to identify the presence of long-range persistence.

In order to assess whether the MLE presented in Section 2 is capable of distinguishing processes with no long-range persistence, we compare the estimates $\widehat{H}(l)$ of long-range persistence processes with estimates of processes without long-range persistence but with short term correlations. We use an ensemble of independent realisations from autoregressive processes $X_t = \beta X_{t-1} + \varepsilon_t$ (where ε_t is Gaussian with zero mean and unit variance) which have short term correlations but no long-range persistence and estimate their persistence exponent H [20]. These estimates are then compared with those of the processes with known long-range persistence such as those shown in previous sections.

Fig. 5 displays the mean of the estimates $\widehat{H}(l)$ over 64 independent realisations of autoregressive processes $X_t = \beta X_{t-1} + \varepsilon_t$ with $\beta = 0.6$ and $\beta = 0.8$ (lower and upper thick lines respectively). This plot also shows the mean $\widehat{H}(l)$ over 64 realisations of processes with long-range persistence $H = 0.65$ and $H = 0.75$ (dashed lines) as in Fig. 3). And finally it shows the estimates for the St. Petersburg temperature data previously analysed (thin full line). For all the realisations in this plot, time series with $N = 41,593$ (which is the size of the St. Petersburg data set) were used. When comparing $\widehat{H}(l)$ of the autoregressive processes with the synthetic long-range persistence processes it is clear that they are different. The autoregressive processes have large values of $\widehat{H}(l)$ for short time lags which decrease rapidly as the time lag increases. This is not the case for the long-range persistence processes. The St. Petersburg data also shows this pattern, large $\widehat{H}(l)$ for short lags and decreasing estimates for longer lags. This data registers evidence of long-range persistence as well as strong short range correlations. Despite this, it is possible to see that the St. Petersburg estimates are different from the one corresponding to the autoregressive

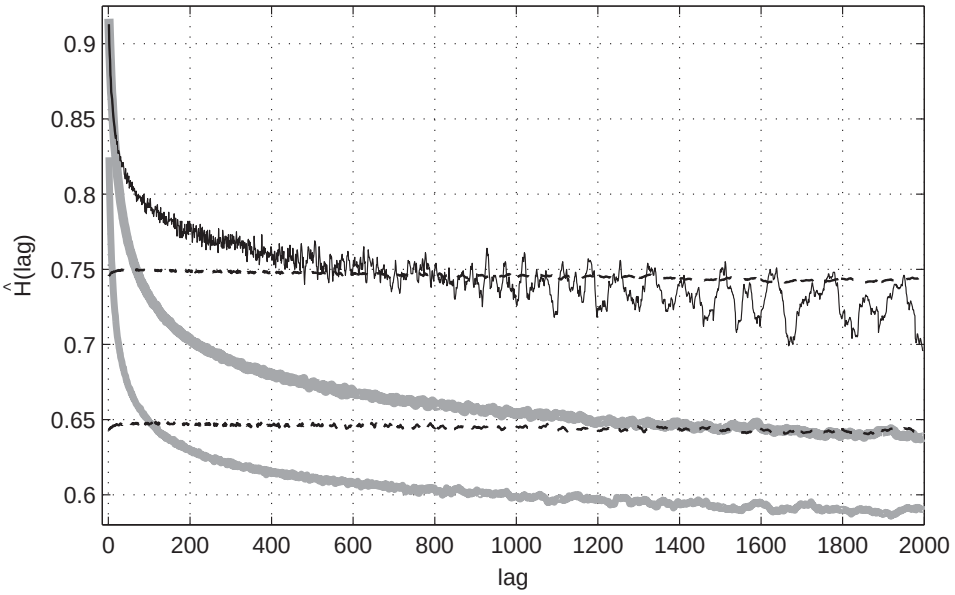

Fig. 5. Mean $\widehat{H}(l)$ over 64 realisations of: autoregressive process $X_t = \beta X_{t-1} + \varepsilon_t$ with $\beta = 0.6$ and $\beta = 0.8$ (lower and upper grey thick lines respectively), long-range persistence processes with $H = 0.65$ and $H = 0.75$ (dashed lines). $\widehat{H}(l)$ for St. Petersburg data (thin full line).

processes. The estimates $\widehat{H}(l)$ for the smallest lags considered have approximately the same values ($\widehat{H}(l) = 0.87$) for both the autoregressive processes (with $\beta = 0.8$) and the St. Petersburg data, however the rate at which $\widehat{H}(l)$ decreases as the time lag increases is different for these two kind of processes. The estimates $\widehat{H}(l)$ decrease very rapidly for the autoregressive process (as lag increases) which is not the case for St. Petersburg's which also decrease but at a much slower rate and most importantly they stabilise around $\widehat{H} = 0.70$ (see also Fig. 4). Thus, by examining the consistency of $\widehat{H}(l)$ over a range of time lags it is possible to determine the presence (and absence) of long-range persistence in a given time series.

In addition, by comparing the estimates of H for a variety of processes we have illustrated that the MLE is able to detect whether a time series presents or not long-range persistence. This is an important property when there is no previous knowledge on whether the data in question presents long-range persistence.

4. Discussion

In previous sections we presented a new approach and MLE of the persistence exponent H. There are some important advantages and disadvantages with this approach when compared to the most commonly used best fit line log–log approach. One of the main advantages is that there is no need to choose a scaling range where

the power law relation is valid, rather one can examine the consistency of a proposed value of $\widehat{H}$ over a range of time scales. In addition, having an estimate for each time lag allows identifying whether long-range persistence is present at large time scales and not only at small ones (just short term correlations). Using a MLE has also the essential advantage of generating a probability density function of the estimator, thus providing a meaningful measure of the uncertainty of the estimates.

It is also important to mention a potential problem with this MLE. When estimating $\widehat{H}(l)$ with Eq. (9) this is influenced by the value of the constant c (estimated using Eq. (10)) which reflects the correlations at lag $l = 1$. In signals where strong short-range correlations are present in addition to the long-range persistence, c can be overestimated which will have an effect on the estimates of H. One possible way to overcome this problem is to filter the short-range correlations prior to estimating H. However, this should be done with care since removing short-range correlations can destroy the scaling structure in the data. Another possibility is to simultaneously estimate c and H in Eq. (9) which can also overcome this problem.

4.1. Assumptions and necessary conditions

The MLE for H introduced in Section 2 is based on the assumption that the increments $Y_t - Y_{t+l}$, $t = 1, l+2, \ldots$ are independent and identically distributed as a Gaussian. Recall, Y_t is defined as the running sum of the original time series X_t with its mean subtracted. The assumption of Gaussianity is met as long as X_t is Gaussian, which is the case for most processes such as fractional Gaussian noises [14] and fractional autoregressive and moving average (fARIMA) processes [21].

On the other hand, the assumption of independence is clearly not met in this context. In fact, this is what the long-range persistence measures, the correlation or linear dependence between observations. Despite the fact that the independence assumption does not hold, likelihood-based estimators have proved useful in practice [19]. In addition, using non-overlapping bins as it is done in Eq. (4) diminishes the effects of violating this assumption. Moreover, a slight modification of the MLE allows for the assumption of independence not be violated. This can be done by only considering in Eq. (4) difference terms $(Y_t - Y_{t+l})$ that are not correlated. For each lag l, the time lag or interval (hereafter *intv*) at which the overlapping differences $(Y_t - Y_{t+l})$ $t = 1, \ldots, N - l$ can be considered linearly independent is determined and this information is used to calculate Eq. (4) only considering terms that are linearly independent. These terms have the form $(Y_1 - Y_{1+l})$, $(Y_{intv} - Y_{intv+l})$, $(Y_{2intv} - Y_{2intv+l}), \ldots$ (where *intv* is determined for each lag l).

Unlike the original method of Section 2 with this modification, depending on the data and the time lag l, the separations between bins will be different. For some l, the bins will overlap and for other l, there will be gaps between bins. This modification allows to have more realistic confidence intervals of $\widehat{H}(l)$ than by using non-overlapping bins which produces overly optimistic (too narrow) confidence intervals for small lags and overly pessimistic (too wide) for large lags.

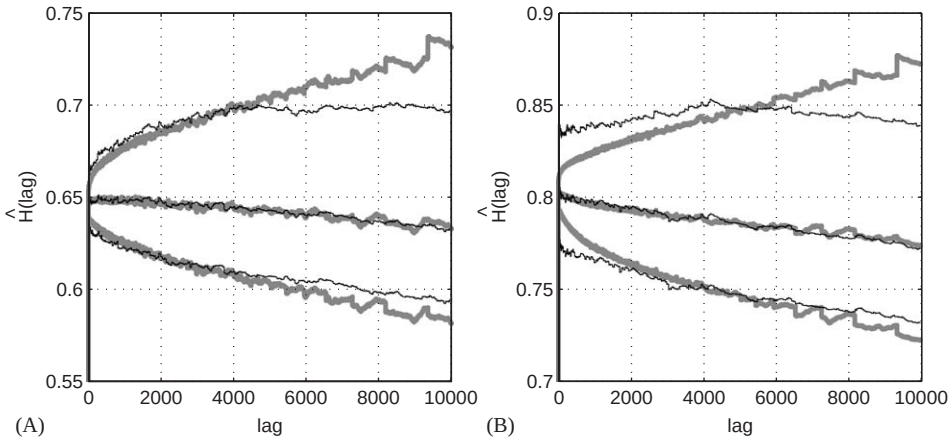

Fig. 6. Mean of $\widehat{H}(l)$ and mean of the upper and lower values of the 95% confidence intervals over 64 realisations with (A) $H = 0.65$, (B) $H = 0.80$. Thick grey lines correspond to the estimates using non-overlapping bins and black thin lines with independent bins.

In Fig. 6 we compare estimates of H using the original and the modified version of the MLE. This plot shows the mean estimates from 64 independent realisations of size 65,536 with known $H = 0.65$ and 0.80. The thick gray lines correspond to the estimates using the original MLE (with non-overlapping bins, as in Section 3) and the thin black lines correspond to the estimates using the modification mentioned above. In both cases, the middle lines show the mean over the 64 estimates as a function of time lag, and the bottom and top lines show the mean of the upper and lower value, respectively, of the 95% confidence intervals of these estimates. Note how the width of the confidence intervals is different for the 2 versions. The mean of $\widehat{H}(l)$ is not significantly different for the two versions and unfortunately the bias is still present at some lags.

5. Conclusions

Claims of long-range persistence in observed time series have received much attention in recent years. In this paper we presented a novel approach and maximum likelihood estimator of the persistence exponent H. In order to test the accuracy of this estimator in terms of spread and bias, we have generated several independent realisations of signals with a known persistence exponent and compared their estimates $\widehat{H}(l)$ with the real value of H. It has been found that in general this new approach correctly estimates H. Nevertheless, for large values of H ($H = 0.75, 0.90$), the estimates show a small bias for large lags. This is a small drawback on the method that in general provides very accurate estimates.

When estimating long-range dependence at different time lags, the uncertainty of those estimates increases as the time lag increases. Unfortunately, the estimates at

longer lags are those of most interest since they are more relevant to characterising long-range persistence. The new MLE presented deals with this problem by providing meaningful uncertainty estimates for $\widehat{H}$ at each time lag. In addition, by providing an estimate of H at each time lag, this new approach provides further insight on the extent to which long-range persistence is present at all lags considered. For the one atmospheric data considered, the range of uncertainty in $\widehat{H}$ is shown to be rather large; claims of equality in estimated H between many different data sets might be reexamined in the light of this analysis.

Acknowledgements

This work was supported by ONR DRI Grant N00014-99-1-0056 and by a CONACYT grant. The authors would like to thank the reviewers for their insightful comments.

References

[1] J. Haslett, A.E. Raftery, Space-time modelling with long-memory dependence: assessing Ireland's wind power resource, Appl. Stat. 38 (1) (1989) 1–50.
[2] K. Fraedrich, R. Blender, Scaling of atmospheric and ocean temperature correlations in observations and climate models, Phys. Rev. Lett. 90 (2003) 108501.
[3] E. Koscienly-Bunde, A. Bunde, S. Havlin, H.E. Roman, Y. Goldreich, H.J. Schellnhuber, Indication of a universal persistence law governing atmospheric variability, Phys. Rev. Lett. 81 (1998) 729–732.
[4] J.R.M. Hosking, Modeling persistence in hydrological time series using fractional differencing, Water Resour. Res. 20 (12) (1984) 1898–1908.
[5] H.E. Hurst, Long-term storage capacity of reservoirs, Trans. Amer. Soc. Civil Eng. 116 (1951) 770–808.
[6] P.Ch. Ivanov, L.A. Amaral, A.L. Goldberger, S. Havlin, M.G. Rosenblum, Z.R. Struzik, H.E. Stanley, Multifractality in human heartbeat dynamics, Nature 399 (1999) 461–465.
[7] B. Pilgram, D.T. Kaplan, A comparison of estimators for 1/f noise, Physica D 114 (1–2) (1998) 108–122.
[8] A. Arneodo, E. Bacry, P.V. Graves, J.F. Muzy, Characterizing long-range correlations in DNA sequences from Wavelet analysis, Phys. Rev. Lett. 74 (16) (1995) 3293–3296.
[9] C.K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Mosaic organization of DNA nucleotides, Phys. Rev. E 49 (2) (1994) 1685–1689.
[10] C.W.J. Granger, R. Joyeux, An introduction to long memory time series models and fractional differencing, J. Time Ser. Anal. 1 (1980) 15–29.
[11] M.S. Taqqu, Q. Teverovsky, W. Willinger, Estimators for long-range dependence: an empirical study, Fractals 3 (4) (1995) 185–798.
[12] A. Eke, P. Hernan, J.B. Bassingthwaighte, G.M. Raymond, D.B. Percival, M. Cannon, I. Balla, C. Ikrényi, Physiological time series: distinguishing fractal noises from motions, Eur. J. Physiol. 439 (2000) 403–415.
[13] J.C. Gallant, I.D. Moore, M.F. Hutchinson, P. Gessler, Estimating fractal dimension of profiles: a comparison of methods, Math. Geol. 26 (4) (1994) 455–481.
[14] B.D. Malamud, D.L. Turcotte, Self-affine time series: I. Generation and analyses, Adv. Geophys. 40 (1999) 1–90.
[15] A. Guerrero, Ph.D. Thesis, University of Oxford, 2002.

[16] A. Guerrero, L.A. Smith, Towards coherent estimation of correlation dimension, Phys. Lett. A 318 (2003) 373–379.
[17] A.-L. Barabasi, H.E. Stanley, Fractal Concepts in Surface Growth, Cambridge University Press, Cambridge, 1995.
[18] M.J. Cannon, D. Percival, D.C. Caccia, G. Raymond, J.B. Bassingthwaighte, Evaluating scaled windowed variance methods for estimating the Hurst exponent of time series, Physica A 241 (1997) 606–626.
[19] W.L. Hays, Statistics, Holt, Rinehart and Winston, Inc., New York, 1988.
[20] C. Chatfield, The Analysis of Time Series: An Introduction, Chapman & Hall, London, 1994.
[21] J.R.M. Hosking, Fractional differencing, Biometrika 68 (1981) 165–176.