ELSEVIER

Physica D 166 (2002) 147–154

PHYSICA D

www.elsevier.com/locate/physd

# Estimating $1/f^{\alpha}$ scaling exponents from short time-series

Octavio Miramontes [a,*], Pejman Rohani [b,c]

[a] Departamento de Sistemas Complejos, Instituto de Física, Universidad Nacional Autónoma de México, México 01000, D.F., Mexico
[b] Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK
[c] Institute of Ecology, University of Georgia, Athens, GA 30602-2202, USA

Received 5 September 2000; received in revised form 17 September 2001; accepted 14 March 2002
Communicated by C.K.R.T. Jones

**Abstract**

In recent years, there has been a concerted effort to develop methods for estimating the scaling exponents of time-series data, thus permitting a characterisation of their underlying dynamical behaviour. This task becomes rather inaccurate with data of limited length (less than 100 points), as is the case in many real studies where observation time is constrained. In this paper, we explore a novel method for accurately calculating the scaling exponents of short-term data, using what we term the multiple segmenting method (MSM). This approach relies on maximising the available information within a time-series by generating pseudo-replicates. We believe this method is potentially useful, especially when applied to biological data.


*PACS:* 72.70.+m; 73.50.Td; 74.40.+k; 87.23.−n

*Keywords:* Noise; Noise parameter estimation; $1/f$; Coloured noise; Time-series

## 1. Introduction

Any given time-series may exhibit a variety of auto-correlation structures. For example, successive terms may show strong ('brown noise'), moderate ('pink noise') or no ('white noise') positive correlation with previous terms. The strength of these correlations provides useful information about the inherent "memory" of the system. One approach for estimating this effect is to estimate the value of the scaling exponent ($\alpha$) in the power spectrum of the time-series ($P(f) = f^{\alpha}$). The exponent is determined by carrying out a linear regression on the log–log transformed Fourier Transform and estimating the slope of this straight line. This is clearly a frequency domain method.

Other temporal domain methods have also been developed in order to characterise the "colour" of time-series data. These include (i) estimating the Hurst Exponent [1], which quantifies the persistence of statistical behaviour in the time-series; (ii) establishing power–law relationships in the frequency and size of fluctuations in the data [2], which describe the probability distribution of fluctuations following perturbations; and (iii) the method of Iterated Function Systems [3], which provides a visual test for identifying possible scaling properties in data. Despite the increasing sophistication of available approaches, it is still unclear which method is optimal. In a recent study, Pilgrim and Kaplan [4] argued, after reviewing a number of these techniques for estimating $\alpha$, that

* Corresponding author.
*E-mail addresses:* octavio@fisica.unam.mx (O. Miramontes), rohani@uga.edu (P. Rohani).


PII: S0167-2789(02)00429-3

FFT regression remains one of the most accurate. This particular approach has the added advantage of being very straightforward to implement.

As with other methods for estimating nonlinear measures from data [5], the performance of the FFT regression approach is sensitive to the length of the time-series. In many applications, this may represent a significant problem. Pilgrim and Kaplan [4], showed that for most methods, typically 1000 data points were required to correctly identify the scaling exponent with equivalent accuracy. In this paper, however, we are interested in highlighting a novel method which works for much shorter time-series. Our method is not only reliable for series consisting of 400 data points, but is also shown to be extremely accurate for time-series with as few as 47 data points. As we outline below, our approach relies on segmenting a given time-series to obtain pseudo-replicates, the analysis of which would yield the true underlying scaling exponent.

## 2. Generating test data

To study this issue, first we generated a number of short, independent time-series of different colours (‘white’, ‘pink’ and ‘brown’). These were generated using two different procedures. The first set of data are generated according to a relaxation process and hence are termed relaxation process data (RPD). The algorithm for this procedure uses a return map to estimate the value of the variable $X$ at time $t+1$:

$$X_{t+1} = \beta X_t + (1-\beta)\epsilon_t, \tag{1}$$

where $\epsilon_t \sim N(0,1)$ is a Gaussian process and $\beta$ defines the strength of correlation or the colour of the data. More specifically, $\beta = 0.99$, 0.69 and 0.0 correspond to brown ($\alpha \approx -2$), pink ($\alpha \approx -1$) and white ($\alpha \approx 0$) noise, respectively.

The second method of data generation is perhaps the most intuitively obvious. It involves setting the amplitude ($A(f)$) at any given frequency ($f$) according to the following formula:

$$A(f) = \sqrt{\frac{1}{f^{\alpha}}}, \tag{2}$$

with $\alpha$ chosen according to the desired noise colour. This generates a spectrum with a “perfect” scaling exponent. The method, discussed by Pilgram and Kaplan [4], generates a time-series that is perfect in the sense that it scales as $1/f^{\alpha}$, but does not fluctuate as a real time-series usually would. There are at least two ways of dealing with this and both are discussed in [4]. The first approach consist of taking subsamples of the original long “perfect” series, while the second consists of perturbing the resulting time series by multiplying the spectral amplitude at each frequency by a random number. It is this last approach that we have chosen, we refer to it as the perfectly synthesised data (PSD).

In addition to these artificially generated data, we also explore the scaling characteristics of selected ecological, financial and meteorological time-series data. For all datasets, we detrend each time-series and subtract the mean. We then estimated scaling exponents with the multiple segmenting method (MSM). The rationale behind our method is that for any given process, there will be a precise underlying scaling exponent, $\alpha$. When an estimate of this exponent has to be made from relatively short data, however, there will be a significant reduction in the accuracy of the estimate. To overcome this limitation, we use segments of a time-series to produce a number of estimates for the scaling exponent. We conjecture that these estimates will have a Gaussian distribution with a mean value corresponding to the inherent $\alpha$. The standard deviation of these estimates will decline at a rate determined by the inverse square root of the length of the segments. This result is obvious for white noise processes and we provide a detailed analysis for pink and brown noises.

## 3. The multiple segmenting method (MSM)

Consider a time-series of length $N$, $\{X_i; i = 1, 2, \dots, N\}$. To estimate a scaling exponent, $\alpha$, we carry out an FFT on different segments of these data each of length $n$, where $n$ is a power of 2 ($n = 2^{\rho}$, $\rho \in \mathbf{N}$, $\rho > 3$). Hence, in general, it is possible to estimate an $\alpha$ for different sub-series of the data:

$\{X_1, X_2, \ldots, X_n\}$, $\{X_2, X_3, \ldots, X_{n+1}\} \cdots \{X_{N-n+1}, X_{N-n+2}, \ldots, X_N\}$. This way, we effectively have $N - n + 1$ pseudo-replicates of the exponent. The important quantities, therefore, are the means and standard deviations of these estimates.

Clearly, there is a cost to using this segmentation scheme: we lose information by discarding long-term correlations in the time-series as the segment size is shortened. However, we gain a statistical advantage since it is possible to estimate the scaling exponent a number of times for any given time-series (though these estimates are not independent). It is demonstrated that the *average* of the scaling exponents for the segments will accurately represent the true colour of the data.

To illustrate MSM in more detail, we present in Fig. 1 the results of our analysis for data generated by the relaxation process (Eq. (1)), with $N = 400$ and $\beta = 0.69$. The estimate of $\alpha$ calculated by Fourier Transforming the entire time-series is $-0.96$. For the MSM, we used segments varying in powers of two from 16 ($2^4$) to 256 ($2^8$) points and the number of 'replicates' for each length $n$ of the segment was $N - n + 1$, as explained above. The mean scaling exponent for all segment sizes was $\simeq -1.0$, as expected from a

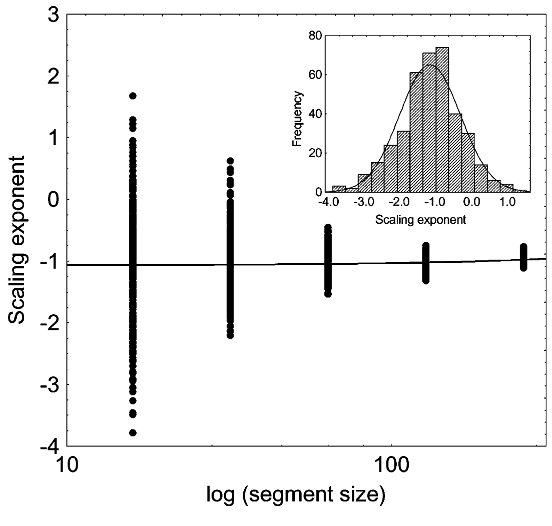



Fig. 1. Estimated scaling exponents versus $\log_{10}$(segment size) using the MSM. The data were generated using the relaxation process method and contained 44 points. Segments of sizes 256, 128, 64, 32 and 16 were considered. There are 145 ($N - n + 1$) segments of length 256 points and 385 segments of length 16 points. The variance in the calculated scaling exponent increased as the size of the segments decreased. Nevertheless the average value of the exponent remained almost constant, fluctuating closely to the value of $\alpha = -1.0$, corresponding to a $1/f$ process. The inset demonstrates the frequency distribution of exponents estimated for $n = 256$.

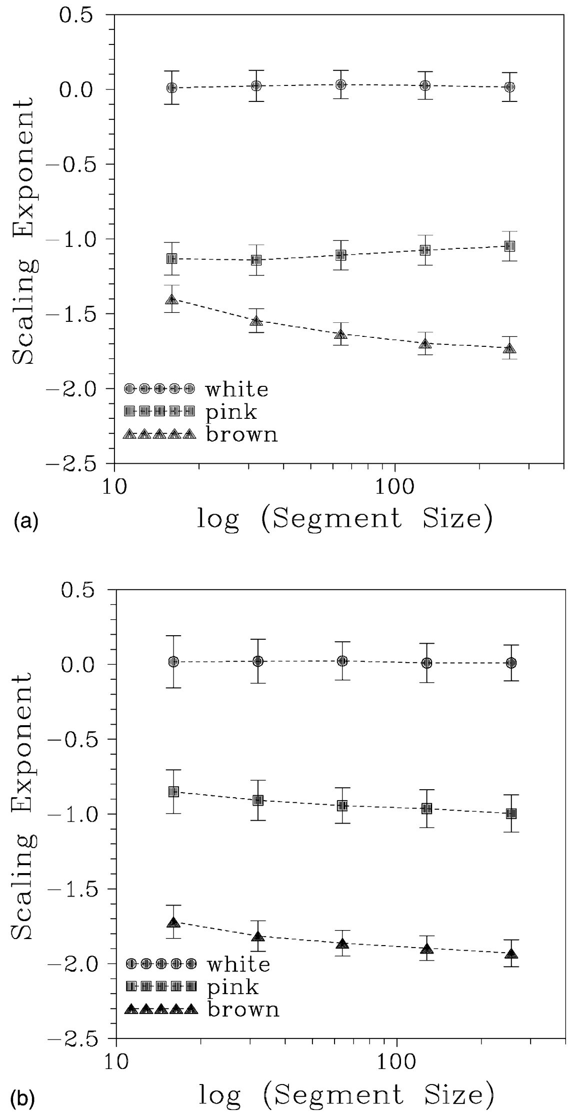



Fig. 2. The figure depicts the results of estimating the scaling exponents from 100 independent time-series of each colour using (a) the relaxation process and (b) the perfect synthesis methods. In both figures, the three classes of correlated data are clearly separated even for very short time-series (e.g. $n = 16$ or 32). For each plot, 300 complete time-series were generated, each of length 400. The error bars represent sample standard deviations.

$1/f$ process. As shown in the inset, the estimated exponents (for $n = 256$) are normally distributed around the true exponent. In Fig. 2a and b, we plot for each method of data generation, the mean estimated scaling exponent for 100 replicated time-series of each colour versus the size of segments ($n$) used to estimate the scaling exponent. The figure demonstrates a number of interesting findings. Most importantly, it is clearly possible to estimate accurately the underlying scaling exponent for very short time-series using this approach. This is especially true for a white noise process, where there is little significant variation in the estimates of $\alpha$ as segment size varies. However, another important finding is that estimates of $\alpha$ for pink and brown noises scale with segment size according to the following ansatz:

$$g(n) = a + \frac{b}{\sqrt{n}}, \tag{3}$$

where $a$ and $b$ are constants. Consider this scaling property for the original time-series data used to generate Fig. 1. The entire time-series ($n = 400$) has an estimated $\alpha \approx -0.96$. When a function $g(n) = a + b/\sqrt{n}$ is fitted to the data generated using the MSM, the values of the constants are $a \approx -1.01$ and $b \approx -0.22$. In this case, $g(400)$ gives an estimated value of $\approx -1.01$, which is almost identical to the *true* value calculated for the entire time-series (the absolute difference is 0.05 or 5%). Another example can be performed with a brown noise time-series ($n = 400$) with $\alpha \approx -1.86$. Again, fitting a function $g(n)$, the values of $a$ and $b$ were $\approx -2.01$ and $\approx 1.66$, respectively. With these values $g(400)$ yields $\approx -1.93$—again almost identical to the *true* estimated value (the absolute difference is 0.07 or 7%). Note that in each case, $\lim_{n\to\infty} g(n)$ approaches the inherent scaling exponent (i.e. $\lim_{n\to\infty} g(n) \to \alpha$). Hence, in the limit, $a = \alpha$.

A striking observation to emerge from Fig. 2a and b is that for a pink noise process, the average value of $\alpha$ has a tendency to decrease as the segment size decreases, albeit the decline is typically very small. For Brownian noise, however, it is difficult to estimate the true exponent from very short time-series ($N < 40$)—the departure from the expected value is considerable. There is a simple and intuitive explanation for this finding. For Brownian processes, there is a pronounced role of low frequency fluctuations in the power spectra. Hence, to estimate the correct exponent, the time-series must be sufficiently long to contain much of the long-period fluctuations. Otherwise, the low frequency fluctuations will be under-estimated, effectively flattening the power spectrum and increasing the value of the estimated exponent.

## 4. Real world examples

How does the MSM perform when applied to real data? In Fig. 3a, we show El Niño data from the eastern Pacific on the equator [3]. Our analyses of these data (Fig. 3b) revealed very strong auto-correlation in the data with $\alpha \approx -2.13$, estimated after fitting $g(n)$ (Eq. (3)) and setting $n \to \infty$. On the other hand, the scaling exponent estimated after carrying out a direct Fourier Transform on the entire time-series is $-2.1$. This *quantitatively* good match between the different estimation methods is reassuring.

Our next application of the MSM is to financial data. In particular, we explored the ratio of the mean daily exchange rate of the US Dollar against the British Sterling over a 5-year period (Fig. 4a). These data appear to exhibit very clear brown noise behaviour with $\alpha \approx -1.93$ when $n \to \infty$ in Eq. (3). Our estimate of $\alpha$ is once again extremely well matched with the exponent calculated from the Fourier Transform of the entire time-series (which gives a value of $-1.8$).

The final real-world application of this method we present is very illustrative. It is the case of a very short ecological time-series (consisting of only 47 data points) exhibiting an interesting interaction between biotic and abiotic variables. In Fig. 5a, we plot the population density of a bacterium in an English pond in the 1950s, together with the associated precipitation over the same period (source: Ref. [8]). In this system, bacterial growth rate is limited by rainfall and its mortality is largely determined by the population of a predatory epizoic protozoan (*Urceolaria mitra*). This analysis would therefore shed some light on the

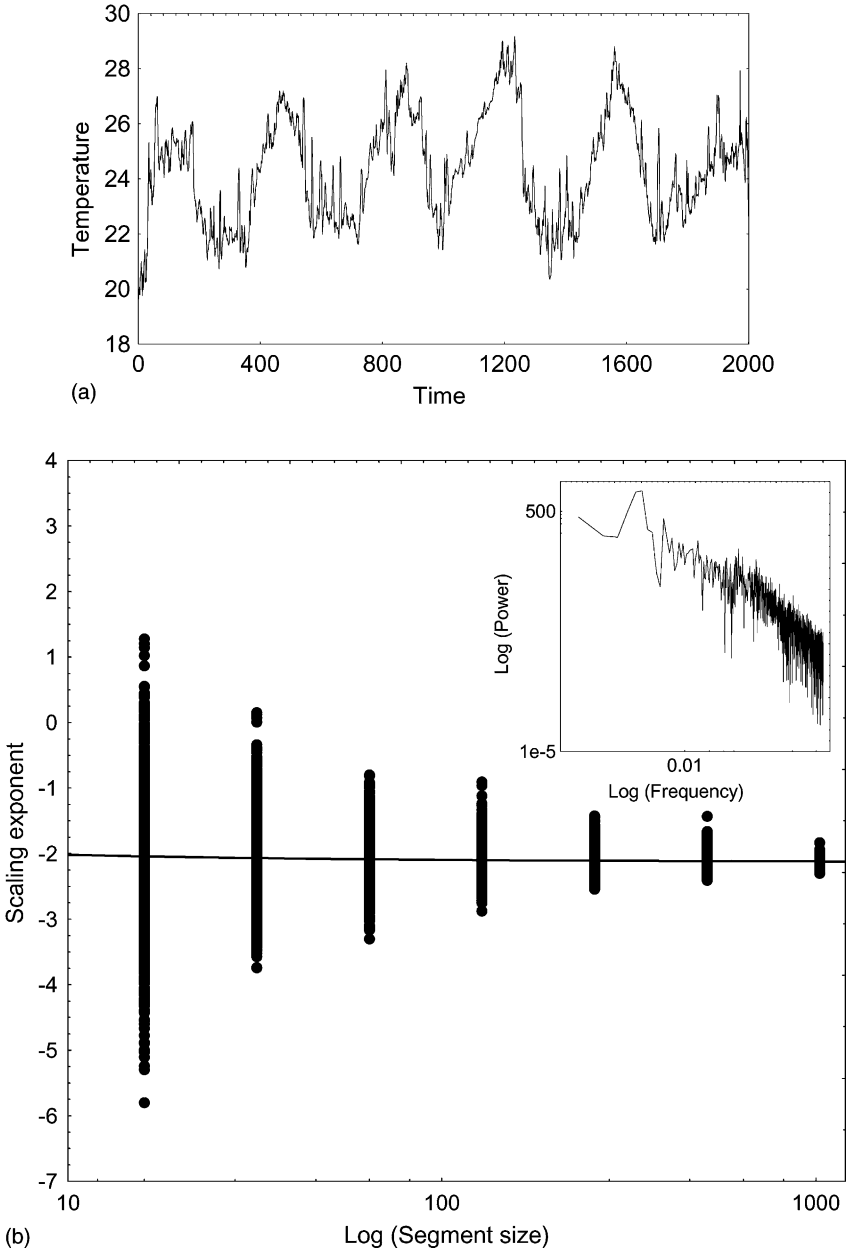
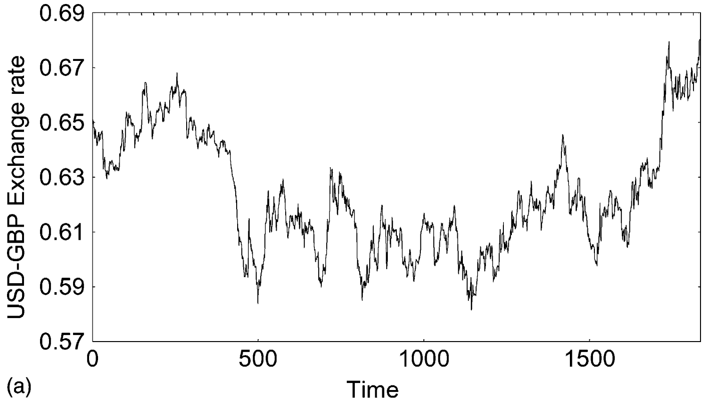



Fig. 3. (a) The mean ocean temperature data consisting of 2000 points, representing daily readings at a depth of 1 m between December 1988 and July 1994 taken from a buoy moored on the equator at 110°W. (b) The application of the MSM estimates an exponent of −2.13, with little variation as segment size increases. The inset shows the scaling behaviour of the power spectrum for $n = 1024$.

consequences of trophic interactions for the scaling behaviour of populations. As shown in Fig. 5c, the estimated scaling exponent using the MSM method for the bacterium population is −1.08, while the rainfall data appears undoubtedly to have a white spectrum, with $\alpha \approx 0.21$ (Fig. 5b). For the bacterium population, the MSM and the direct method provide starkly contrasting estimates. The exponent estimated directly



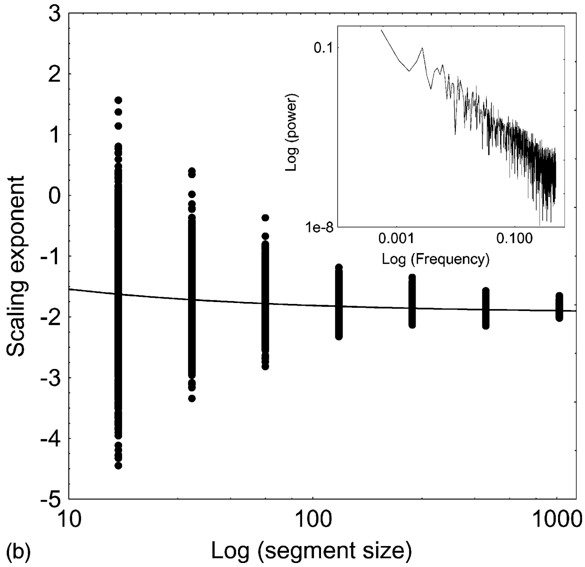




Fig. 4. (a) The mean daily US$-GB Sterling exchange rate. The data span the period 22/8/1995–22/8/2000 and were obtained from www.oanda.com. (b) The application of the MSM estimates an asymptotic exponent of $-1.93$. In these data, there is some variation as $n$ increases, with the larger segment sizes providing a smaller $\alpha$, as might be expected. The inset shows the scaling behaviour of the power spectrum for $n = 1024$.

from the Fourier Transform of the full time-series is $-0.16$. Hence, with the direct method, one would be tempted to conclude that the bacterium population fluctuations mimic those of the rainfall being essentially white noise.

These analyses of the ecological data with the MSM are especially interesting in light of recent debates in this field. Numerous real populations, including aquatic microorganisms, have been shown to have a "reddened" spectrum, however, most simple (and universally used) models of population processes seem incapable of reproducing this trait, often exhibiting white or blue noise spectra [9]. This has led some to argue that the redness must be due to environmental forcing (e.g. Ref. [10]). Yet our brief analysis here suggests that such broad generalisations may well be dangerous. The combination of environmental variability, inherent nonlinearities in population growth and trophic interactions can have subtle results [7].

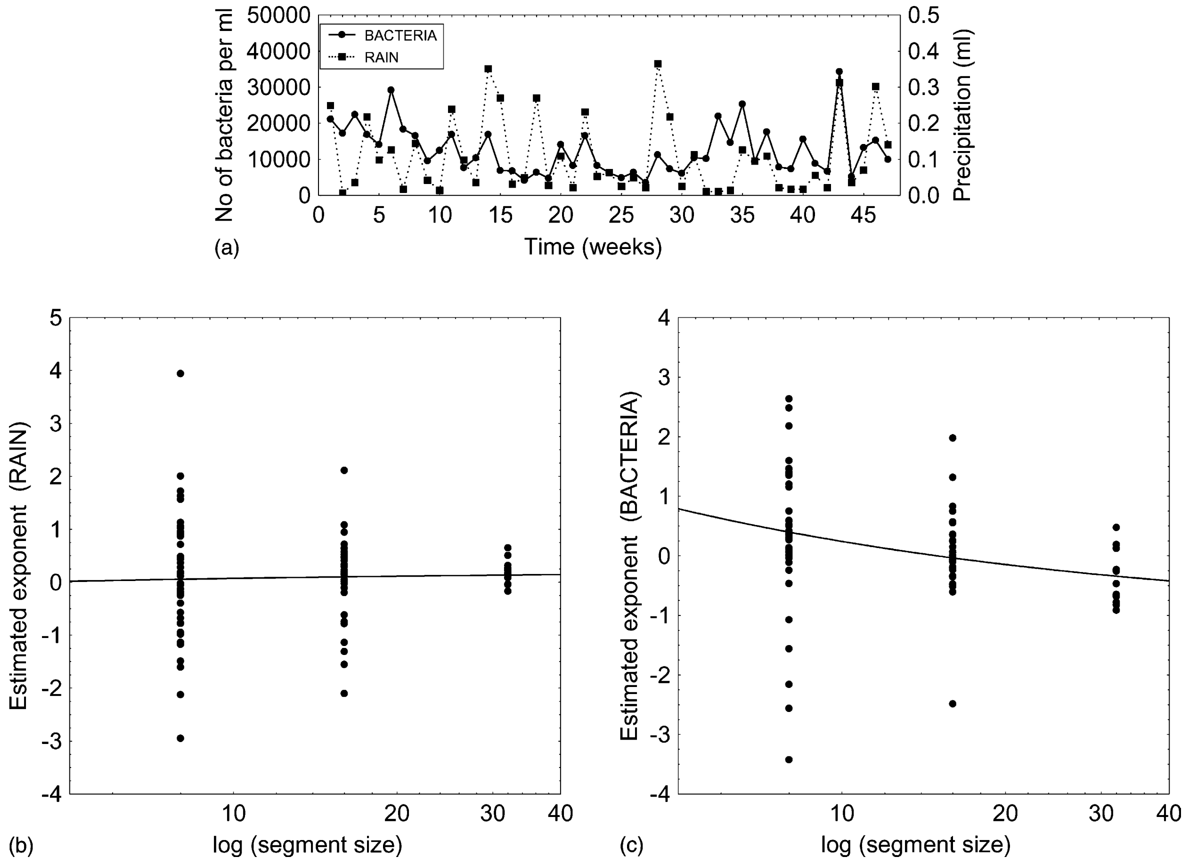



Fig. 5. (a) The bacterium population density (measured as number of bacteria per millilitre) and corresponding rainfall (ml) for the Reynoldson data. Despite the fact that the time-series contain only 47 points, the application of the MSM estimates asymptotic exponents of (b) 0.21 for the rainfall data and (c) $-1.08$ for the bacterium population. In this example, the MSM consist of segmenting the original 47 points series to produce 40, 32 and 15 subseries of 8, 16 and 32 points. When the scaling exponent of the subseries are plotted, it is possible to estimate the asymptotic exponent value of the original time-series after using Eq. (3).

## 5. Concluding remarks

Different degrees of correlation exist in time-series coming from many and diverse areas in physics, biology, economics, etc. It is becoming very important to find methods that may identify and measure accurately these correlations in the form of a scaling exponent in the frequency domain. This especially important when the time-series are short [4], a situation that is commonly present in some areas of the biological sciences [6,7]. We have in this article, described a novel method to estimate scaling exponents that may prove relevant to address the problem of accurately distinguishing the colour of noise in very short time-series. The method involves the use of replicas from segmentating the original time-series and their statistical analysis based on the behaviour of an scaling ansatz between the exponent values and the segment size. We think this method is potentially useful in the sense that it is an excellent tool to estimate a value for the scaling exponent $\alpha$ in short time-series and that, under these circumstances, it is superior to the most common method, namely the direct FFT method. Estimating the correct exponent is useful, in turn, to gain insight into what kind of dynamic processes may be involved in the generation of the time-series—this is specially important nowadays when there is a lack of adequate knowledge about the nature of the mechanisms generating $1/f$ like noise, particularly in biological phenomena.

## Acknowledgements

We are grateful to Pedro Miramontes and Stephen Cornell for stimulating discussion. This work was funded by DGAPA-UNAM (IN-111000), CONACYT (32453-E) and Royal Society of London-Academia Mexicana de Ciencias fellowships to OM and a Royal Society University Research Fellowship to PR.

## References

[1] B. Mandelbrot, The Fractal Geometry of Nature, Freeman, New York, 1983.
[2] P. Bak, C. Tang, K. Wiesenfeld, Phys. Rev. A 38 (1988) 364.
[3] J.C. Sprott, G. Rowlands, Chaos Data Analyzer, Physics Academic Software, American Institute of Physics, 1995.
[4] B. Pilgrim, D.T. Kaplan, Physica D 114 (1998) 108.
[5] J.P. Eckmann, D. Ruelle, Physica D 56 (1992) 185.
[6] O. Miramontes, E. Ceccon, Physica A 257 (1998) 439.
[7] O. Miramontes, P. Rohani, Proc. Roy. Soc. London B 265 (1998) 785.
[8] T.B. Reynoldson, J. Anim. Ecol. 24 (1955) 57.
[9] J.E. Cohen, Nature 378 (1995) 610.
[10] G. Sugihara, Nature 381 (1996) 199.