Available online at www.sciencedirect.com
ScienceDirect
Physica A 387 (2008) 296–302
PHYSICA A
www.elsevier.com/locate/physa

Anti-persistence in levels of Lake Naivasha: Assessing effect of human intervention through time-series analysis

Sujata Tarafdar[a,*], David Harper[b]

[a]Condensed Matter Physics Research Centre, Physics Department, Jadavpur University, Kolkata 700032, India
[b]Department of Zoology, University of Leicester, LE17RH Leicester, UK

Received 19 June 2007; received in revised form 5 August 2007
Available online 11 September 2007

Abstract

Lake Naivasha in Kenya is an important natural fresh water reserve, supporting surrounding wildlife as well as agriculture and industry. Uncontrolled use of the lake water for the past few decades is causing concern for environmentalists. In the present paper, fluctuations in the lake level for the last half century are analysed using standard tools for time-series analysis. The intervals 1951–1980 (period I) and 1981–2000 (period II) are treated separately, to look for any difference in their statistical patterns. From period II onwards, increased human consumption is believed to affect the level significantly. We analyse the data using three different approaches: (i) rescaled range analysis (R/S), (ii) roughness scaling analysis and (iii) a Lomb periodogram. R/S analysis shows no difference between the behavior in periods I and II, but the other methods reveal different fluctuation patterns for the two periods. The water level shows stronger fluctuations in period I compared to II. R/S analysis, however, shows an interesting anti-persistence with a Hurst exponent 0.44, which is not usually observed in natural time series.

PACS: 05.40.−a; 05.45.Tp; 05.45.Df

Keywords: Lake Naivasha; Time series analysis; Rescaled range; Lomb periodogram; Anti-persistence

1. Introduction

Variation in levels of lakes and rivers with time, is a classic problem widely studied since the earliest days of time series analysis [1]. Most observations record a *persistence* in the behaviour of natural time series [2,3], with Hurst exponents over 0.5. Lake Naivasha is an endorheic lake in Kenya. Records of the water level for the past hundred years is available [4], for the past 50 years, regular monthly readings have been recorded. This lake supports a rapidly growing agri-horticulture industry, as well as surrounding wildlife. A downward trend in the water level since about 1980, is worrying environmentalists [4,5]. Starting from this period, the observed lake level is falling significantly below the expected level simulated on the basis of flow rates of rivers which feed Lake Naivasha.

*Corresponding author. Tel.: +91 33 24237540.
E-mail address: sujata_tarafdar@hotmail.com (S. Tarafdar).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2007.08.050

The present work is an analysis of the lake level data, to search for any interesting feature. In particular, we compare the data for last 20 years with data prior to that period to see if human intervention has produced any noticeable difference in parameters characterising the temporal fluctuations of the lake level. We approach the problem from three different angles. (i) A rescaled range (R/S) analysis [2,6] has been done to determine the Hurst exponent H. Surprisingly, the Hurst exponent is lower than 0.5, indicating an *anti-persistence*. (ii) We find a power-law scaling behaviour for the difference in maximum and minimum levels $D(\tau)$ over a period of time τ. (iii) A Lomb periodogram [7] is drawn to look for periodicity in fluctuations of the level.

2. Fluctuations in Naivasha Lake level

Records of levels of Lake Naivasha starting from the year 1900 are available. From 1951 levels are recorded regularly at one month intervals, before that recording times are irregular and at larger intervals. Trends observed in these records and their deviation from the level predicted from water input from rivers flowing into the lake for the last 20 years are discussed in detail in Refs. [4,5]. The observed shortfall from the simulated level since the year 1980, is presumed to be due to excessive horticulture and other industrial development. There are no rivers flowing out from Naivasha, natural depletion in the water level is due only to evaporation and some seepage from subterranean streams. The lake level data L_n as a function of time is shown in Fig. 1.

We have done the entire analysis over 590 months starting from the year 1951, the actual number of data points is 560, as there are occasional readings taken at 2 or 3 month intervals. One of the objectives of this work is to look for any difference in statistical patterns during the first 30 years starting 1951, when only natural processes are at work, and the next 20 years when human intervention is believed to affect the lake level.

3. Analysis of the data

3.1. Rescaled range analysis

In the R/S analysis, the 'range' R is the cumulative departure from an initial time t_0 to time t, as if the successive data points are step lengths of a random walk. Obviously R increases with the interval $(t - t_0)$. It has been shown [1] that dividing the range by the standard deviation σ in the data for this time interval gives a robust parameter usually exhibiting a power-law with $(t - t_0)$. If the recorded variable $X(t)$, for successive observation times is totally uncorrelated, i.e. the data resembles a random Brownian motion, R/S varies as a

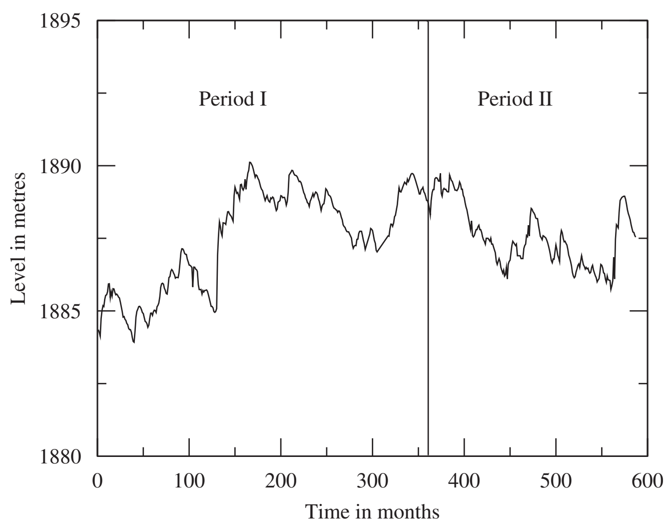

Fig. 1. Variation in levels of Lake Naivasha for the period 1951–2000 [4].

power-law with $(t - t_0)$, with the exponent 0.5:

$$R/S = A(t - t_0)^H. \tag{1}$$

A being a constant. The exponent $H = 0.5$ indicates a truly Brownian motion, while $H > 0.5$ indicates a persistence, that is an increasing trend in the previous step is more likely to be followed by an increase in the next step as well. $H < 0.5$, means there is anti-persistence in the variation of X, if there is an increasing trend over the mean in one step, it is more likely to be followed by a decrease in the following step.

We follow the method discussed in Refs. [6,8] $R_{i,\tau}/S_{i,\tau}$ is termed as the 'rescaled range', where $R_{i,\tau}$ is the cumulated range of a process i.e. the departure from the mean value, in the time interval $i+1$ to $i+\tau$ and $S_{i,\tau}$ is the standard deviation. For a series x_i in discrete integer valued time with X_i defined by $X_i = \sum_{j=1}^{i} x_j$ and given any lag $\tau > 1$

$$R_{i,\tau} = \max[(X_{i+j} - X_i) - (j/\tau)(X_{i+\tau} - X_i)] - \min[(X_{i+j} - X_i) - (j/\tau)(X_{i+\tau} - X_i)], \tag{2}$$

where max and min are taken over $0 < j < \tau$. The standard deviation $S_{i,\tau}$ is given by

$$S_{i,\tau}^2 = \tau^{-1} \sum_{j=1}^{\tau} x_{i+j}^2 - \left(\tau^{-1} \sum_{j=1}^{\tau} x_{i+j}\right)^2. \tag{3}$$

This is averaged over $i = 1, \ldots, n$ non-overlapping intervals of length τ to give the rescaled range

$$\mathbf{R}_\tau = \langle R_{i,\tau}/S_{i,\tau} \rangle. \tag{4}$$

Here we replace x_n by L_n. We assume all points for the data to be recorded at intervals of 1 month. This makes the analysis simpler without introducing any significant error. For R/S analysis, it is convenient to choose the lag in multiples of 2, so we start from $\tau = 0.25$ years (3 months), and continue to increase τ in multiples of 2 up to 16 years (192 months). Since we want to analyse the period 1951–1980 (period I) and 1981–2000 (period II) separately, it is not meaningful to work with any larger interval.

$\mathbf{R}_\tau$ has been evaluated averaging over as many time intervals that fit without overlapping in the periods I and II. We find the numerical values for each lag for periods I and II are identical up to the second decimal place, and values range from 3.75 to 13.97. It is therefore not necessary to show the results separately. Fig. 2 shows a log–log plot of $\mathbf{R}_\tau$ against the lag τ. The data give a reasonably good linear fit with a slope $H = 0.44$. It follows that the fractal dimension of the graph in Fig. 1 is 1.56 [2]. This means that if the plot in Fig. 1 is covered by a square grid, with boxes of varying size $l \times l$, the number of boxes $N(l)$ which cover the graph is

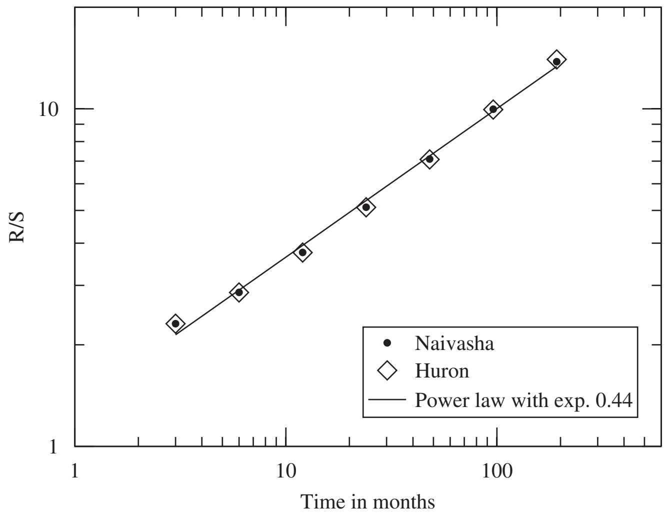

Fig. 2. Rescaled ranges for the lake levels of Naivasha and Huron. Data points for both lakes coincide and anti-correlation is observed with a Hurst exponent 0.44.

given by

$$N(l) \sim l^{(H-2)}. \tag{5}$$

So R/S analysis reveals no difference in long-range statistical correlations in period I and II. However, a very interesting result is that the Hurst exponent is less than 0.5, showing anti-persistence.

Natural processes usually show positive correlation i.e. persistence, earlier work on lake levels [2] also reports Hurst exponents close to 0.7. A natural question is whether the exponent 0.44 we obtain indicates some peculiarity specific to Naivasha. We looked for similar data for other lakes for comparison. Levels of Lake Huron/Michigan from [9] was analysed using the same formalism. Very surprisingly this data for a similar period of time gives not only exactly the same slope 0.44, but identical data points for the R/S values, agreeing up to the second decimal place in most cases. The average levels of the lakes are of course, quite different, being around 1890 m for Naivasha and 177 m for Huron. The data for Huron is also shown in Fig. 1 for comparison. This striking similarity may be a coincidence, but it would be interesting to study recent data (within the last 100 years) for other lakes as well.

3.2. Level difference $D(\tau)$

We have calculated a simpler quantity compared to R/S, which shows a difference in behaviour in the periods I and II. This is the quantity $D(\tau)$ defined as

$$D(\tau) = \max(L_n) - \min(L_n) \tag{6}$$

within the time interval i to $i + \tau$. This is evaluated as a function of τ for as many overlapping intervals shifted by one point, as can fit into the whole period under study, and averaged over all these for a specific τ. It is to be noted that here we allow overlap of the time intervals. Looking at the time series data as a rough surface, this quantity $D(\tau)$ is analogous to the surface width, determined over an interval. We find that $D(\tau)$ follows an approximate power law on the average for period I. For period II the log–log plot shows two different linear regions. Initially the points coincide with the points for region I, then for time intervals larger than about 5 years, deviate to a different average power-law. Writing the power-law as

$$D(\tau) = \tau^m. \tag{7}$$

The exponent m, which is a measure of the 'roughness' of the time-series 'surface' [10,11], is determined for the periods I and II. The exponent is significantly different in the periods I and the later section of period II. The values of m, for period I, is 0.58, while for the period II it is initially the same as I but later deviates to 0.361.

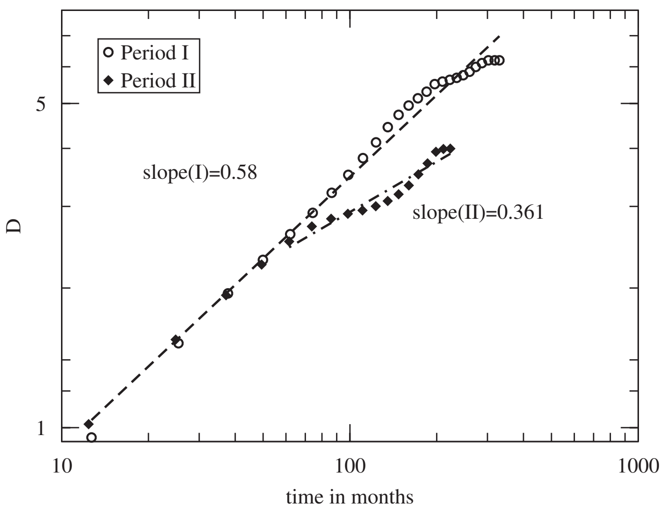

Fig. 3. The difference in maximum and minimum levels over a time interval, shows a scaling behaviour, with different power-laws for period I and II.

Fig. 3 shows the plots for the two periods. The value of m is much less in period II, compared to period I, signifying less fluctuations in levels, in this period. It is possible that this change reflects higher consumption due to more human intervention in period II, which varies less with time, than natural water supply and depletion processes. The water level expected in absence of human intervention can be simulated by monitoring the flux in rivers which feed the lake. Simulation data for period II show that during this time the observed level falls significantly below the simulated level [4], presumably due to increased demand for human consumption. However, the simulated levels for this period show an increase above the average. If the excess natural supply over the average is used up in agri-horticulture, m would decrease as observed here. But if their is a drought or decrease in water supply to the lake due to some natural cause, judicious regulation of human usage would continue to keep m low. An increase in m in such a case would be a cause for alarm indicating that the level is falling much below average due to natural and human effects adding up.

Another interesting feature seen is a smooth periodic variation decorating the power-law fit. It may indicate some discrete scale invariance (DSI) [12] in the data, signifying hidden deterministic fractal patterns [12,13]. Longer time intervals must be analysed to identify this more definitely.

3.3. The Lomb periodogram

Another aspect of a time series analysis is the occurrence of periodicities. Periodic cycles in a time series can be identified by Fourier transform or a Lomb periodogram. The Lomb periodogram [7] works for data recorded at irregular time intervals as well.

Suppose we have readings for a variable h_i at N data points which may correspond to unevenly spaced time instants t_i, the mean and variance of h are given, respectively, by

$$h_{av} = \frac{1}{N}\sum_{1}^{N} h_i, \tag{8}$$

$$\sigma^2 = \frac{1}{(N-1)}\sum_{1}^{N}(h_i - h_{av})^2. \tag{9}$$

The Lomb normalised periodogram, which is the spectral power as a function of the angular frequency $\omega = 2\pi f > 0$ is defined by

$$P_N(\omega) = \frac{1}{2\sigma^2}\left[\frac{\left(\sum_j (h_j - h_{av})\cos\omega(t_j - \tau)\right)^2}{\sum_j \cos^2\omega(t_j - \tau)} + \frac{\left(\sum_j (h_j - h_{av})\sin\omega(t_j - \tau)\right)^2}{\sum_j \sin^2\omega(t_j - \tau)}\right]. \tag{10}$$

τ is defined by

$$\tan(2\omega\tau) = \frac{\sum_j \sin(2\omega t_j)}{\sum_j \cos(2\omega t_j)}. \tag{11}$$

For N data points $2N$ values of ω are returned. One can identify prominent repeating cycles from the angular frequencies which have large amplitudes, i.e. show the highest peaks in the plot of $P_N(\omega)$ against ω. ω may be transformed into the frequency f and time period T, through

$$f = \omega/2\pi = 1/T.$$

In Fig. 4 the periodograms for the periods I, II separately and (I + II) together are shown in blue, red and black respectively. All three graphs show largest amplitudes in the range of frequencies corresponding to cycles of 10–8 years. In the combined black curve there are three separate peaks within this range, in period I a broad peak covering the whole range and in period II, a sharp peak for $T = 8$ years, and a smaller one for 10 years. Since the data for periods I and II, cover time intervals of about 30 and 20 years, respectively, we ignore peaks with larger T.

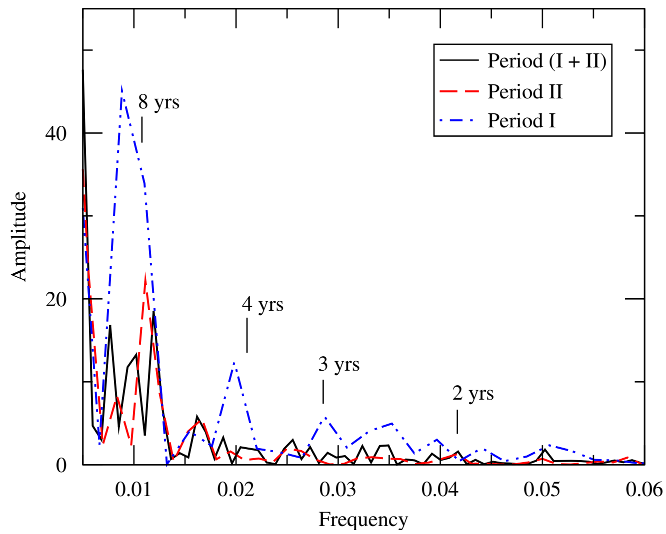

Fig. 4. Lomb periodogram for period I , period II and period (I + II).

A very prominent peak is seen around $T = 4$ years in period I, which is conspicuously absent in period II. A much smaller peak is seen for periods I as well as (I+II) near $T = 5$ years. An overall difference in the character of the time series data for I and II, is quite evident. The earlier period shows much more prominent cycles compared to the later period. This agrees with the 'roughness' exponent being larger for period I, in short, variation in levels is much less in the later period, this may be related to large amounts of water abstraction in period II for human utilisation. It may further be conjectured that the four year cycle is related to the 'El Nino' and the larger (of period 10 years) to sunspot cycles, which affect global temperatures [8].

4. Conclusions

In conclusion, the fluctuation in water levels of Lake Naivasha in Kenya for the period 1951–2000, has been analysed using different tools for time series analysis, namely R/S, roughness and the Lomb periodogram.

One significant feature is the observation of anti-persistence, with a Hurst exponent of 0.44 for the whole period in R/S analysis. This is surprising, considering that natural time series usually show persistence within such time scales. To see whether this result is peculiar to Naivasha, which has some unique features [5], such as being endorheic, we also analysed data for lake Huron/Michigan in USA. The two series turn out to have almost exactly same R/S values as shown in Fig. 2 and hence the same Hurst exponent. It is necessary to study this aspect further, with data for other lakes.

We look separately at the periods 1951–1980 and 1981–2000, in the later period simulations of the lake level do not agree with actual measured data [5]. So we search for any quantitative difference in statistical parameters for the two periods. R/S analysis does not reveal any difference, but the other methods do. The roughness of the lake level versus time curve, characterised by the exponent m, is significantly higher in the earlier period compared to the later period. In the periodogram also the earlier period shows prominent peaks, unlike the later period. All this points to strong natural fluctuations in the water level which may have been somewhat washed out by human intervention during the period following 1980. How this may affect the ecology will make an interesting study.

Acknowledgements

ST became interested in the problems of Lake Naivasha and its surroundings during a trip to Kenya as a volunteer for the Earthwatch project 'Lakes of the Rift Valley' (PI—David Harper). She thanks Barid

Tarafdar for introducing her to Earthwatch. Authors thank the team of scientists involved in the project—Rob Britton, M'bogo Kamau, Dominic Kamau and Laban Njoroge for their help and support.

References

[1] H.E. Hurst, R.P. Black, Y.M. Simaika, Long term Storage, Constable, London, 1965.
[2] J. Feder, Fractals, Plenum, New York, 1988.
[3] D.L. Turcotte, Fractals and Chaos in Geology and Geophysics, second ed., Cambridge University Press, Cambridge, 1997.
[4] R. Becht, D.M. Harper, Dev. Hydrobiol. 168 (2002) 1–11.
[5] D. Harper, K. Mavuti, Ecohydrol. Hydrobiol. 4 (2004) 255–266.
[6] B.B. Mandelbrot, The Fractal Geometry of Nature, Freeman, San Francisco, 1982.
[7] M. G. Kendall, A. Stuart, The Advanced Theory of Statistics, vol 3, second ed., Griffin, London 1968.
[8] O. Karner, J. Geophys. Res. 107 (2002) 4415–4425.
[9] ⟨ www.in.gov/dnr/water/lake_michigan/coastal/lake_level fluc.html⟩.
[10] A.L. Barabasi, H.E. Stanley, Fractal Concepts in Surface Growth, Cambridge University Press, Cambridge, 1995.
[11] R. Karmakar, T. Dutta, N. Lebovka, S. Tarafdar, Physica A 348 (2005) 236–244.
[12] D. Sornette, Phys. Rep. 297 (1998) 239.
[13] A. Roy, S. Roy, A.J. Bhattacharyya, S. Banerjee, S. Tarafdar, Eur. Phys. J. B 12 (1999) 1–3.