Chinese Journal of Oceanology and Limnology
Vol. 26 No. 3, P. 313-322, 2008
DOI: 10.1007/s00343-008-0313-z

Analysis of long-term dependence phenomenon in Benue River flow process and its hypothesis testing*

Martins Y. OTACHE**, LI Zhijia (李致家), Mohammad BAKIR
State Key Laboratory of Hydrology-Water Resources & Hydraulic Engineering, Hohai University, Nanjing 210098, China

Received June 22, 2007; revision accepted Oct. 29, 2007

Abstract In this paper, the long-term dependence phenomenon (the Hurst Effect) which characterizes hydrological and other geophysical times series is studied. The long-term memory is analysed for both daily and monthly streamflow series of the Benue River at Makurdi, Nigeria by using heuristic methods and testing specifically the null hypothesis of short-term memory in the monthly flow series. Results obtained by applying heuristic procedures indicated that there may be the presence of long-term memory component in mean daily flow series but there is no discernible reason to suspect the presence in both average monthly and maximum monthly flow series (extreme event). Hypothesis testing was conducted by using original and modified versions of rescaled range statistic. When the modified rescaled range, which accounts for short-term memory in the series, is used, the null hypothesis is accepted for both the average monthly and maximum monthly flow series, indicating little or no probable presence of long-term memory in the series. An identical conclusion is also arrived at when second null hypothesis for independence of the monthly flow series is tested. Therefore, apart from the mean daily flow series, there is little evidence of long-term dependence in the Benue River streamflow series at Makurdi. However, considering the limited length of data used, the results are inconclusive.

Keyword: Hurst exponent; hydrological method; time series; river flow; long memory; *Nigeria*

1 INTRODUCTION

Long-range dependent processes provide an elegant explanation and interpretation of an empirical law, which is commonly referred to as the *Hurst Effect*. The phenomenon of long-range dependence has a long history; it remains a topic of active research in the study of economic and financial time series, and has been extensively documented in hydrology, meteorology and geophysics.

Recent results have led to re-awakening of the need by hydrologists to further analyze long-term dependence in temporal series of hydrologic data. This quest is aimed at developing suitable methods for estimating and modelling the intensity of long-term dependence in time series, as well as providing insight to what might be the reasons for the Hurst phenomenon. Based on the concept of long-term dependence, a stationary process X_t possesses long memory if there exists a real number $H \in (0.5,1)$, called the "Hurst exponent", and a constant c_H such that

$$\rho(k) : c_H k^{2H-2} \quad (1)$$

as $k \to \infty$, where $\rho(k)$ is the autocorrelation coefficient of the process at lag *k*. As a result, long memory implies that the autocorrelation function of the process decreases slowly, like a power function (Montanari et al., 2000). The exponent *H*, in a hydrological context, is called the intensity of long memory. A value of the Hurst exponent equal to 0.5 means absence of long memory, the greater the *H*, the higher the intensity of long memory; small *H* less than 0.5 corresponds to short-term negative time dependence, which is rarely encountered in the analysis of hydrologic data (Montanari et al., 2000). Generally, the existence of long memory has been

* Supported by the Post Graduate Research Fund from Federal Government of Nigeria under BEA Scholarship Program (No. CSC2001566010),
** Corresponding author: martinso3@yahoo.com

explained either by pre-asymptotic behaviour of the rescaled adjusted range R_n^*, certain classes of non-stationarity in time series, and infinite memory, or erroneous estimation of the Hurst exponent (Mesa and Poveda, 1993).

To test for statistically significant long-term memory in a hydrologic time series, a precise distinction between short-term and long-term memory must be made (Rao and Bhattacharya, 1999). The concept of short-term dependence is based upon the notion of “strong mixing” (Rosenblatt, 1956) which is a measure of the correlation between two events separated by successively increasing time lags (Rao and Bhattacharya, 1999). Based upon the definition of “strong-mixing”, the difference between long-term and short-term memory is based on the rate at which the correlation between events separated by increasing time lags decreases (Rao and Bhattacharya, 1999). It has been noted that long-term memory processes behave differently from the short-term one; for instance, spectral density of the FGN (Fractional Gaussian Noise) model, a model for long-term memory processes is infinite at zero frequency while it is finite at zero frequency for short-term memory processes.

Against this background, the primary objective of this study is to investigate if the streamflow series of the Benue River, at daily and monthly time scales are characterised by long-term dependence. If it is present, then it connotes a significant serial correlation among observations which are far apart in time. Toward this end, heuristic methods like the classical R/S analysis, modified R/S analysis, and the Aggregated Variance Method shall be used for detecting the presence of long-term dependence in the data series. To further understand the long-range dependence phenomenon, hypothesis testing shall also be performed for the monthly data. The paper is organised as follows: in section 2, the heuristic methods used here are outlined. The data used in the study and the underlying pre-processing approach are discussed in section 3, while the results for both long-term memory and hypothesis testing are presented and discussed in section 4. Finally, in section 5, a set of conclusions is presented.

2 DETECTING LONG-MEMORY IN STREAMFLOW SERIES OF THE BENUE RIVER

The presence of long memory in a time series can be detected by estimating the value of the Hurst exponent *H*. This can be done by applying different heuristic methods. Though robust in many cases, they cannot provide any additional information about the spectral density of the flow data (Montanari et al., 2000). Flows of the Benue River have never been subjected to stochastic analysis of this form in the recent past. Thus, the dearth of continuous data or availability of limited data sample will create indeterminacy in any form of stochastic studies. It is noted that reliable long-memory estimation can be performed only when sample size of available data is large enough for asymptotic properties to hold. The question here is: What size of data sample is judged large enough?

This seemingly potential lack of reliability of the range/static (R/S) method when dealing with short sample sizes constitutes a stiff problem. The fact that no appropriate length of sample size is advocated in literature informed the basis of this study; the primary focus being to inspect in detail whether long-term dependence is present in the Benue River flows. Considering all these, and coupled with the fact that long-term dependence is in-variant with respect to data aggregation (Montanari et al., 2000), the long-term dependence estimation is performed on both the daily and monthly series. Because of the limited sample size of the monthly series, hypothesis testing for the presence of long-term dependence and serial independence are performed on it in order to be better disposed to draw objective conclusions.

2.1 Rescaled range statistic (R/S) method

The classical R/S-analysis aims at inferring from an empirical record the value of the Hurst parameter for long-range dependent process that presumably generated the record at hand. In practice, classical R/S-analysis is based on a heuristic graphical approach where the resulting R/S values are plotted against the lag in a log-log plot (pox plot) to yield a straight line with slope equal to *H* (i.e., the Hurst exponent). In this context, the R/S statistic is the range of partial sums of deviations (R) of times series from its mean, rescaled by its standard deviation (S).

Thus, given a sample of N observations and $2 \le n \le N$, one can define

$$< x >_n = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (2)$$

$$\mathrm{X}(i,n)=\sum_{j=1}^{i}\left[x_j-<x>_n\right] \tag{3}$$

$$R(n)=\max_{1\le i\le n} X(i,n)-\min_{1\le i\le n} X(i,n) \tag{4}$$

$$S(n)=\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i-<x>_n\right)^2\right]^{1/2} \tag{5}$$

Hurst (1951) found that

$$E[R(n)/S(n)]:\left(\frac{n}{2}\right)^H \tag{6}$$

where the ratio of $R(n)/S(n)$ is defined here as $Q(n)$, and H is the Hurst exponent while n is the number of segments or blocks; similarly, X and x represent the demeaned and mean of the data points in each blocks, respectively. To this end, for any lag n and $2\le n\le N$, there are Int [N/n] estimates of $R(n)$ and $S(n)$ where the eventual value of $R(n)/S(n)$ is averaged over all the estimates of $R(n)/S(n)$, precisely over all Int [N/n].

In this study, Eqs.2–6 are estimated for both the daily and monthly streamflow series. In doing so, suppose w is a series of natural integers such as 1, 2, 3, and N is total observations, n series is set as 2, 3, ... , $w\le N/2$ for both daily and monthly flow series while the R/S statistic was calculated for 157 and 155 logarithmically spaced values of n respectively.

2.2 Modified R/S analysis

To overcome the shortcomings of the classical R/S-analysis, Lo (1991) proposed a modified R/S statistic by replacing $S(n)$ in Eq.5 with $S(N, q)$. Both $S(n)$ and $S(N, q)$ measure the standard deviation of the partial sum, but in $S(N, q)$'s regime, the variance of the partial sum is not just simply the sum of the variance of individual observations, but also includes the weighted autocovariances up to lag q which is aimed at correcting the bias probably caused by the existence of short memory as well as heteroscedasticity.

The $S(N, q)$ is defined as

$$S(N,q)=\left(\frac{1}{N}\sum_{j=1}^{N}\left(x_j-\bar{x}_N\right)^2+\frac{2}{N}\sum_{j=1}^{q}\omega_j(q)\left[\sum_{i=j+1}^{N}\left(x_i-\bar{x}_N\right)\left(x_{i-j}-\bar{x}_N\right)\right]\right)^{1/2} \tag{7}$$

where $\bar{x}_N$ denotes the sample mean and the weights $\omega_j(q)$ are given by

$$\omega_j(q)=1-j/(q+1),\ q<N. \tag{8}$$

As a result, the Lo's modified R/S statistic is defined as

$$Q_{N,q}=\frac{1}{S(N,q)}\left\{\left[\max_{0\le i\le N}\sum_{j=1}^{i}\left(x_j-\bar{x}_N\right)-\min_{0\le i\le N}\sum_{j=1}^{i}\left(x_j-\bar{x}_N\right)\right\}\right. \tag{9}$$

In practice, the null hypothesis of no long-range dependence is rejected at 5% significance level if $Q_{N,q}$ is not contained in the interval [0.809, 1.862]; i.e., the 95% (asymptotic) acceptance region. However, for the implementation, a right choice of q is very essential.

For purposes of the analysis, in this study, the data-driven formula for choosing q given as

$$q_{\text{opt}}=\left[\left(\frac{3N}{2}\right)^{1/3}\left(\frac{2\hat{\rho}}{1-\hat{\rho}^2}\right)^{2/3}\right] \tag{10}$$

and

$$q_{\text{opt}}=\left[\left(\frac{N}{10}\right)^{1/4}\left(\frac{2\hat{\rho}}{1-\hat{\rho}^2}\right)^{2/3}\right] \tag{11}$$

by Wang (2006) are adopted. Here, N is the length of the data, $\hat{\rho}$ is the estimated first-order autocorrelation coefficient, and [ ] is the greatest integer function.

2.3 Aggregated variance (AV) method

The aggregated series is obtained by dividing the streamflow series, here of length N, into blocks of length m and averaging the series over each block. The sample mean is computed as:

$$\bar{X}_m(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \quad k = 1,2,..., N/m \tag{12}$$

and the sample variance as

$$S^2(m) = (N/m - 1)^{-1} \sum_{k=1}^{N/m} \left[\bar{X}_m((k) - \bar{X}_N) \right]^2 \tag{13}$$

where $\bar{X}_N$ denotes the overall mean.

For successive values of *m*, the sample variance of the aggregated series is plotted versus *m* on a log-log plot. It has been proved that in the presence of long-range dependence the variance,

$$Var(\bar{X}_m) : cm^{2H-2} \tag{14}$$

for a large *N/m* and *m*, where *c*>0, and $\bar{X}_m$ is the sample mean (Beran, 1989). To implement this scheme, 155 logarithmically spaced values of *m* were chosen respectively for the daily and monthly flow series.

2.4 Hypothesis testing for long-term dependence in the streamflow series

Memory in hydrologic time series is usually analysed by testing the null hypothesis for the existence of short-term memory. Hence, any empirical investigation of long-term dependence must first account for short-term memory or high frequency autocorrelation.

Toward this end, the objectives here are to test the null hypothesis that there is only short-term memory and to determine whether there is serial independence in the monthly flow series of the Benue River at Makurdi. The decision to confine this test only to the monthly series is informed by its limited sample size since it is reported in literature that the reliability of the long-term memory analysis is affected by sample size.

The null hypothesis test for short-term memory is done by using the rescaled range, $Q(n)$ as in Eq.6 and the modified rescaled $Q_{N,q}$ statistic in Eq.9. Since the modified rescaled range statistic $Q_{N,q}$ accounts for short-term memory in the series, when it is used, the null hypothesis is tested after accounting for short-term memory or high frequency autocorrelation in the series. To test the null hypothesis of no long-term dependence, the following test statistic is computed for both the average monthly and maximum monthly flow series,

$$V_n = \frac{1}{\sqrt{n}} Q(n). \tag{15}$$

In order to compare the rescaled range, the statistic which is defined by Eq.16 is computed

$$V_n(q) = \frac{1}{\sqrt{n}} \hat{Q}(n,q), \tag{16}$$

where $\hat{Q}(n,q)$ becomes the modified rescaled range and is computed as the mean $R(n)/\hat{S}_n(q)$ value of sample length *n*. Similarly for $Q_{N,q}$, the null hypothesis is tested at 5% significance level and V_n is rejected if it is not contained in the interval [0.809, 1.862], which assigns equal probability to the upper and lower tail of V on the assumptions of asymptotic theory (Rao and Bhattacharya, 1999; Lo, 1991). To check the sensitivity of the statistic to the truncation lag, $V_n(q)$ is computed for three values of *q*, namely *n*/10, *n*/20, *n*/30 while the bias in V_n is also evaluated to account for variance in the range $R(n)$ arising from autocovariance in the series, as it indicates the correction factor associated with $Q(n)$.

On the other hand, the test for serial independence in the monthly flow series is based on comparing values of the Hurst exponent (Hurst, 1951) with those that could arise by chance from a series of independent and identically distributed data (Lye and Lin, 1994). Here, the Hurst exponent, denoted by *K* (Rao and Bhattacharya, 1999; Wallis and Matalas, 1970) is computed by $K = \log Q(n)/\log(n/2)$, where $Q(n)$ is the rescaled range. Using $\hat{Q}(n,q)$ instead of $Q(n)$, the modified Hurst exponent K_1 is computed as

$$K_1 = \frac{\log\left[R(n)/\hat{\sigma}_n(q)\right]}{\log[n/2]} = \frac{\log \hat{Q}(n,q)}{\log(n/2)} \tag{17}$$

Percentage points for testing the statistical significance of K and K_1 exist only for a series of length n = 200 (Table 4) thus, K and K_1 are evaluated for a sample length of n = 200. To do this, if the computed value of K_1 for a hydrologic time series is greater than K at a given sample size, then a particular series is concluded to be serially dependent (Rao and Bhattacharya, 1999; Lye and Lin, 1994).

3 DATA AND PRE-PROCESSING

In this study, historical time series of the Benue River flow at Makurdi (centre at 7°44′N, 8°32′E) location is used; it is based on a total of 26 years (1974–2000) of continuous average daily discharge data provided by the Federal Inland Waterways, and Benue State Water Board, Makurdi Office, Nigeria. The Benue River is the major tributary of the Niger River, and is approximately 1 400km long; it is almost navigable during the wet (rainy) season, precisely from July to October. Its headwaters rise in the Adamawa Plateau of the Northern Cameroon, then into Nigeria. The Benue River is characterised by definite wet and dry seasons which give rise to changes in river flow and salinity regimes. To investigate the long-range dependence phenomenon in the flow series, the average daily flows are aggregated to mean monthly streamflows by summing the average daily flow over the total number of days in the month. Similarly, the maximum monthly streamflow is the maximum of all average daily flows in a month.

Hydrologic time series observed at time intervals shorter than a year, generally exhibit periodic patterns caused by the oscillations of the weather in an annual cycle. As a consequence, seasonal or periodic characteristics are not constant through the year (Montanari et al., 2000). Thus, for long-range dependence analysis, the seasonality must be removed. Both the daily and monthly flow series show presence of seasonality (Figs.1, 2). To remove the seasonality in the daily flow series, the average daily flow series are log-transformed to normalize the data and then deseasonalised; the deseasonalisation is done by applying

$$m(j,i) = \frac{x_{j,i} - \bar{X}}{s_i} \tag{18}$$

where $\bar{X}$, is the daily mean, s_i the standard deviation and $x_{j,i}$ is the flow data matrix.

Because of the highly visible periodic patterns in the Benue River series (Fig.2), the periodicities in the mean and variance are removed by taking a preliminary transformation of the monthly flow data. Here, the classical harmonic regression method

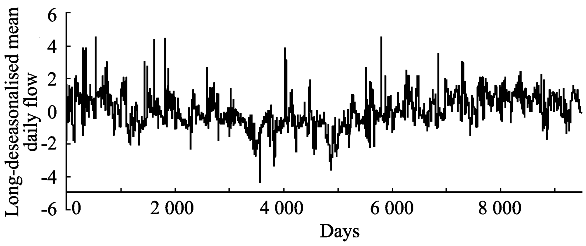

Fig.1 Deseasonalised logarithmic transformed daily flow series of the Benue River at Makurdi

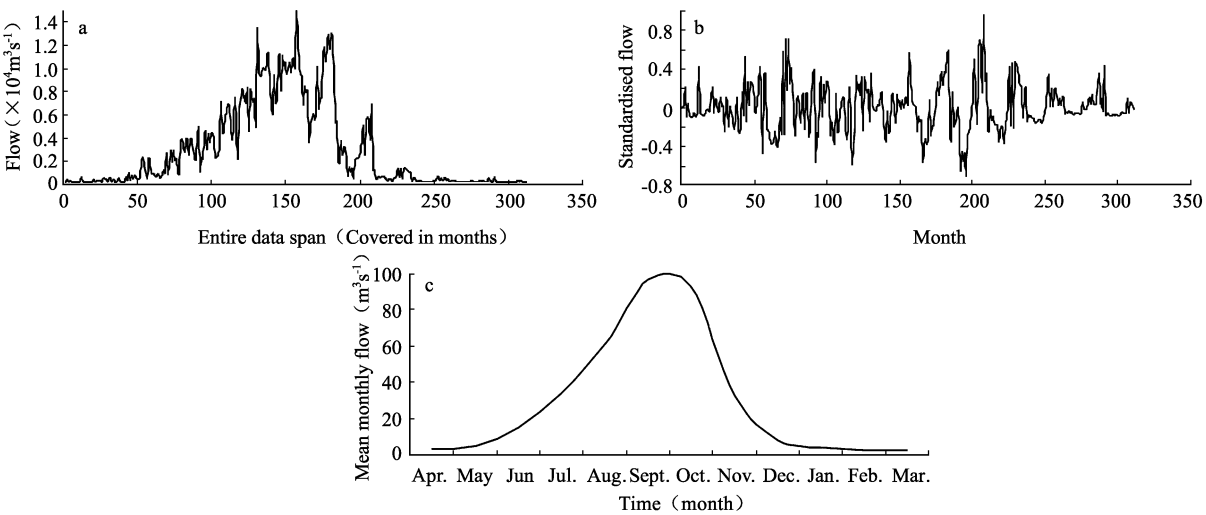

Fig.2 Average monthly flow in the Benue River at Makurdi

(a) original time series; (b) harmonic standardised time series; (c) yearly zero-mean periodic component (April 1974 – March 2000)

(Kottegoda, 1980) is adopted. This is done by computing first the η, periodic yearly averages of the series,

$$m_\tau = n^{-1}\sum_{i=1}^{n} x_{\tau+\eta}(i-1),\ 1 \le \tau \le n \qquad (19)$$

where η is the time span of the periodicity and n, the number of years of data.

Here, n = 26, η = 12, and τ = 1, . . ., 12. This is followed by representing the η values of m_τ through the sum of a finite number l of harmonics given as

$$m_\tau = \sum_{i=1}^{l} \lambda_i \sin\left(\frac{2\pi\tau}{\eta} i + \phi_i\right) + \xi_t. \qquad (20)$$

The amplitudes (λ_i) and phases (ϕ_i) are estimated by minimizing the sum of squares of ξ_t, whereas the number of harmonics needed is found through the analysis of the variance explained by each harmonic. For monthly data, it is sufficient to fit a maximum of six harmonics of periodic parameter for a time series but should be tested for the significance (Patra, 2001). Thus, based on the analysis of the variance for significance, six harmonics are used in the smoothing of the monthly m_τ, displayed in Figure 2c, as an estimate of the periodic component of the flow series; this is also extended to the standard deviations. The periodic component estimated is subtracted from the series and divided by the periodic component of the standard deviations to obtain the deseasonalised series which is then used for further analysis.

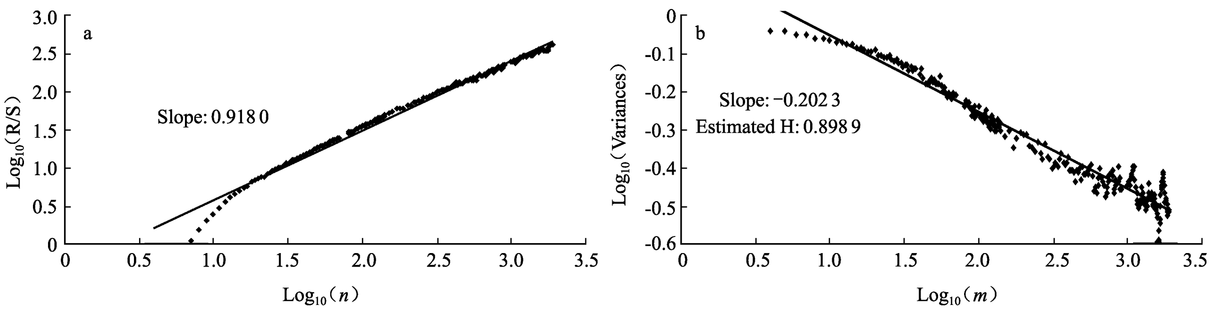

Fig.3 Pox plots of the mean daily flow series

(a) R/S-analysis; (b) aggregated variance method

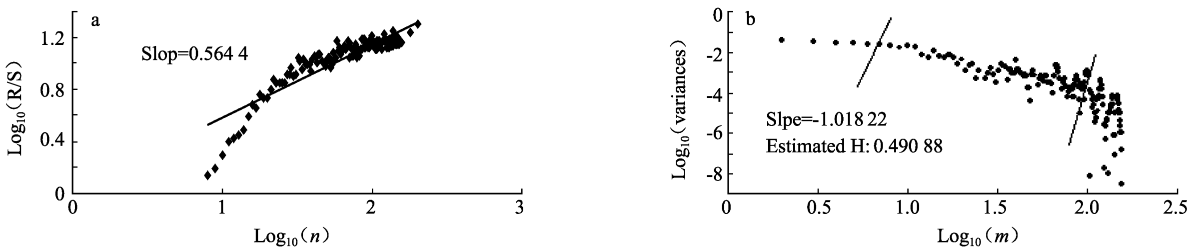

Fig.4 Pox plots of the average monthly flow series

(a) R/S-analysis; (b) aggregated variance method, *n* and *m* are number of blocks in the respective methods

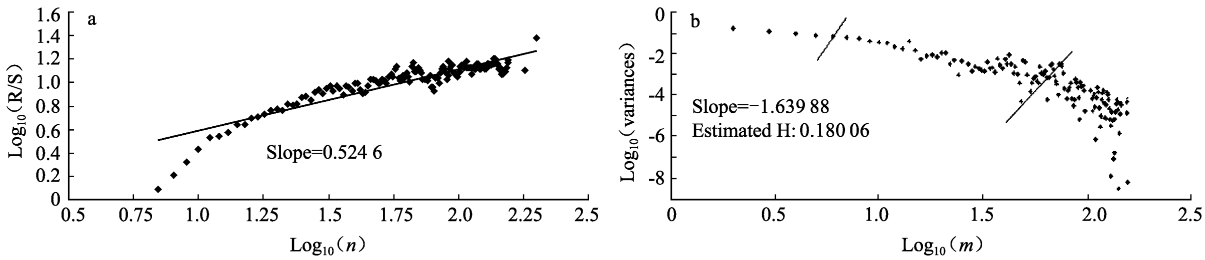

Fig.5 Pox plots of the maximum monthly flow series

(a) R/S-analysis; (b) aggregated variance method, *n* and *m* are number of blocks in the respective methods

4 RESULTS AND DISCUSSION

The results of the analysis in the preceding section are discussed in detail to provide insight to the dynamics of the flow regimes of the Benue River at Makurdi.

4.1 Long-term dependence

The three heuristic methods described in section 2 are used to evaluate the Hurst exponent in the flow series. Both the R/S and the aggregated variance pox plots suggest that a strong long-memory component may be present in the mean daily flow series (Figs.3a, 3b). Table 1 also shows, in terms of H values that by using the modified rescaled range statistic, there is indication of the probable presence of high intensity long-memory in the flow series. By applying both the R/S and aggregated variance statistics, the values of H were obtained by least-square fit.

But the scenario presented above is completely different for the average monthly flow series as well as in the maximum monthly flow series (extreme event). The values of H obtained by least-square fit by applying the R/S statistic respectively for average monthly flow series and maximum monthly flow series indicate the presence of low intensity long-memory component in the monthly flow series (Figs.4a, 5a). The aggregated variance (Figs.4b, 5b) and modified rescaled range (Table1) values of H show a contrasting phenomenon. Considering the R/S, modified rescaled range and aggregated variance statistics in this case, there is little or no discernible evidence to suspect the presence of long-term memory in the monthly flow series but rather, the values of H are indicative of short-term memory; these results accord with the findings of Rao and Bhattacharya (1999) in their analysis of long-term memory in hydrologic time series. Be that as it may, because of the relatively short length of the data series and its probable variability, there is visible presence of associated problems. This highly manifest in nonlinearity and excessive scatter which may cause uncertainty in the slopes and thus render the computed value of the Hurst exponent unrealistic (Figs. 3b; 4a, 4b; 5a, 5b) especially with the aggregated variance method.

Though the modified rescaled range statistic (Table 1), using Lo's (1991) $q_{opt,}$ is in agreement with this, if Wang 's (2006) q_{opt} formulation is used in evaluating the rescaled range statistic (Table 1), the null hypothesis of no long-range dependence is rejected at 5% significance level, i.e., the interval [0.809, 1.862]. This development could be attributable to variability in data (as noticeable in the tail part of Figs.4b, 5b) and as such, when Wang 's (2006) q_{opt} formulation is used in the modified rescaled range statistic, a balance must be found so as to determine the most appropriate formula for q_{opt}, while Wang's formulation gives an economical value of q, Lo's is prohibitively high. Considering this therefore, it is imperative to note that picking a single value of q to determine, on the basis of $Q_{N,q}$, whether or not to reject or accept the null hypothesis of no long-range dependence in a given data set is highly problematic and may lead to erroneous conclusions; doing this, requires a delicate trade-off. As shown in Table 1, Figs 4a, 5a, it is worrisome to note that Lo's method has the strong preference of accepting the null hypothesis of no long-range dependence irrespective of whether long-range dependence is present in the data or not. But against the backdrop of the short sequence of data used here, it is difficult to really contest this question.

On the other hand, the deseasonalisation of both the mean daily and monthly flow series does greatly explains the variance in the original data (Figs.1, 2b). Fig.2c displays the yearly-zero mean periodic component, as expected, it reflects the weather pattern, with one peak corresponding to the raining season.

Table 1 Modified rescaled R/S analysis

Time series	ACF (1)	Lo's $V_q(n)$	Q_{opt}*	Wang's $V_q(n)$	Q_{opt}**
Deseasonalised log mean daily flow	0.9472	5.9623	169	9.6586	39
Deseasonalised average monthly flow	0.3725	0.8734	7	0.7281	2
Deseasonalised max. monthly flow	0.0873	0.9555	2	0.6679	1

* q_{opt} is computed according to Eq. 11; ** Lo's R/S statistic is computed with Wang's q as in Eq. 11

4. 2 Hypothesis testing for short-term memory

The values of V_n for $n = N/4$, $n = N/2$ for the time series considered are listed in Table 2. By using both $n = N/2$, and $n = N/4$, the hypothesis testing is performed with two and four realisations of the series, respectively. For the harmonic standardised average monthly flow, the bias in $Q(n)$, indicated in Table 2 when $q = n/10$ for $n = N/4$, is 1.353%, though not significant but for $n = N/2$, it is –46.57% indicating that $Q(n)$ is downwardly biased; so is the situation with maximum monthly flow. The only plausible explanation for this could be probably data quality problem as well as its length. How much the length of the data and its quality may have impacted negatively or otherwise is another issue that has to be further investigated. In general, the bias in V_n for the maximum monthly flow series is relatively smaller than that of the average monthly flow series though the $V_n(q)$ values when $q = n/10$ for $n = N/2$ are significant; but on the whole, changing q influences the estimates of V_n but not substantially.

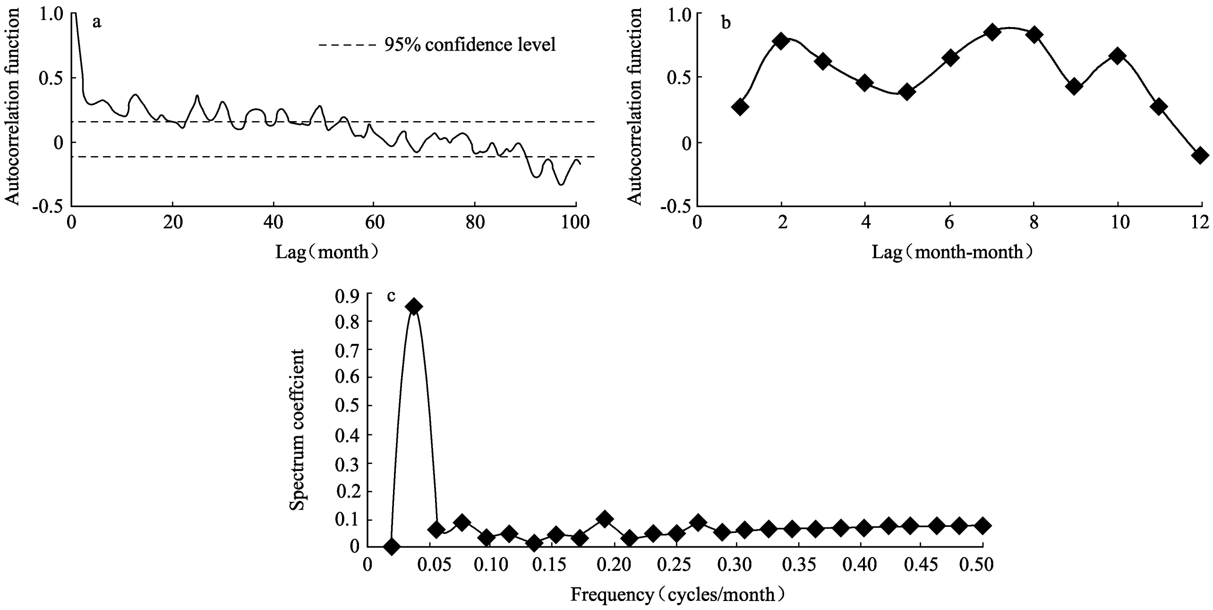

Fig.6 Autocorrelation functions of the standardised average monthly flow series

(a) Over time; (b) lag1 month-to-month over realization; (c) sample spectrum

Table 2 Comparison of test statistics

Time Series	V_n	$V_n(q)$ $q = n/10$	%Bias	$V_n(q)$ $q = n/20$	%Bias	$V_n(q)$ $q = n/30$	%Bias
n = N/4							
Mean monthly flow	1.570	1.550	1.353	1.331	18.008	1.243	26.30
Max. monthly flow	1.300	1.610	−19.28	1.391	−6.54	1.261	3.07
n = N/2							
Mean monthly flow	1.149	2.152*	−46.57	1.239	−7.18	1.087	5.77
Max. monthly flow	1.188	2.353*	−49.50	1.492	−20.34	1.309	−9.22

$\% \text{ Bias} = \left(\frac{V_n}{V_n(q)} - 1 \right) \times 100$; * values are significant

Table 3 Comparison of Hurst exponent estimates for monthly Series

Time series	K	$K_l(q = n/20)$	$K_l(q = n/5)$	$K_l(q = n/4)$	$K_l(q = n/2)$
Mean monthly flow	0.646	0.103	0.218	0.251	0.379
Max. monthly flow	0.690**	0.125	0.219	0.277	0.343

** Marginally significant

Table 4 Empirical percentage points for Hurst's *K* for i.i.d data (*n* = 50—200) (Lye and Lin, 1994)

Sample	Significance levels				
size	1%	5%	10%	20%	50%
50	0.777	0.739	0.720	0.691	0.629
55	0.772	0.735	0.715	0.687	0.647
60	0.768	0.731	0.711	0.683	0.625
65	0.765	0.727	0.708	0.680	0.623
70	0.762	0.724	0.704	0.678	0.622
75	0.759	0.721	0.702	0.675	0.621
80	0.757	0.718	0.699	0.673	0.619
85	0.754	0.716	0.696	0.671	0.618
90	0.752	0.714	0.694	0.669	0.617
95	0.749	0.711	0.692	0.667	0.616
100	0.747	0.709	0.690	0.666	0.616
110	0.742	0.706	0.687	0.663	0.614
120	0.738	0.703	0.684	0.660	0.613
130	0.734	0.700	0.681	0.658	0.611
140	0.630	0.698	0.679	0.656	0.610
150	0.727	0.695	0.676	0.654	0.609
160	0.725	0.693	0.75	0.653	0.608
170	0.722	0.690	0.673	0.651	0.607
180	0.720	0.688	0.671	0.649	0.606
190	0.718	0.686	0.670	0.648	0.605
200	0.715	0.685	0.668	0.647	0.604

In the analysis for test of serial independence in the flow series, the modified rescaled range statistic was computed by using autocovariance of the series up to truncation lags of $q = n/20$, $n/5$, $n/4$ and $n/2$, though chosen subjectively. The mean values of K and K_l for both the average monthly and maximum monthly streamflow series are listed in Table 3; these values were then compared with significant levels of Hurst exponent as shown in Table 4. Results in Table 3 indicate that K values for both average monthly and maximum monthly flows do not show any meaningful significance at 5%, 10% and 20% levels while those of K_l are highly insignificant. This is explained by the relatively weak autocorrelation structure (Fig.6a, b). In both cases, the variation in K_l is sparingly marginal as the truncation lag q is increased. But considering the small magnitudes of the values of K_l and its dependence on q, it becomes obvious that short length of data series may impact negatively on the outcome thereby rendering it statistically insignificant to be meaningful.

Further analysis show that the average monthly flow series is still affected by some non-stationarity in its autocovariance structure (Fig.6a). The autocorrelation functions of each calendar month, i.e., lag 1 month-month correlations, showing low-lag autocorrelation coefficients involving some months do not differ significantly at 95% confidence level, unlike their averages on all months (Fig.6a, 6b). This goes to say that the removal of seasonality in the mean and variance through deseasonalisation does not entirely free its residual series from seasonality; Fig.6b typifies this complexity in autocorrelation function. However, it may still be present in the covariance structure as is generally the case for seasonal streamflow series (Salas, 1993). On the whole, the monthly flow series shows evidence of low-frequency components, which are manifested in a slow decaying, periodic correlogram and sample spectrum (Kottegoda, 1980) with visible high values at frequencies near 0.02 and 0.19 cycles per month; however, Fig.6b emphasizes the distortion in the autocorrelation function when it is computed "over time" rather than "over realisations".

5 CONCLUSIONS

Based on the heuristic tests carried out on the streamflow series, using the classical R/S, modified rescaled range statistic and the aggregated variance method, in terms of the Hurst exponent values, one comes to the conclusion that the mean daily flow series may display long-term memory although it is difficult to estimate precisely the measure of persistence. The other streamflow series, i.e., the average monthly and maximum monthly flows examined show that there is no discernible reason to suspect the presence of long-term memory, indicating that there is no significant serial correlations in the series. Though this might seem to be the case here, in the face of short data sequence, it is recommended that bootstrapping should be done in order to be able to objectively choose between Lo's and Wang's formulations of q_{opt} for adoption.

Other analyses on the monthly flow series led to the acceptance of the null hypothesis that there is only short-term memory in the series examined when both the V_n and $V_n(q)$ are used. The inference from these results is that after accounting for autocorrelation up to lag q, there is little evidence of long-term memory in the monthly flow series. In addition, the effect of varying the truncation lag q is marginal though noticeable. Similarly too, the null hypothesis that both the average monthly and maximum monthly flow series are independent is accepted based on the test statistics. This leads to the conclusion that after accounting for autocorrelation up to lag q, there is high tendency of the series being serially independent. These results conform to those previously found and documented in literature (Rao and Bhattacharya, 1999; Kashyap and Rao, 1976; Hipel and McLeod, 1994).

With these results though, since the autocovariance structure of the deseasonalised average monthly flow series still looks periodic (Fig.6a), one may conclude that a non-deterministic periodic component is present. Therefore, a seasonal model is needed to fit the monthly flow data. Generally, because of the limited length of the sample data used for this study, the results obtained here are inconclusive and thus subject to further study.

References

Beran, J., 1989. A test of location for data with slow decaying serial correlations. *Biometrika* **76**: 261-269

Hipel, K. W. and A. I. McLeod, 1994. Time Series Modelling of Water Resources and Environmental Systems. Elsevier, Amsterdam.

Hurst, H., 1951. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **116**: 770-799.

Kashyap, R. L. and A. R. Rao, 1976. Dynamic Stochastic Models from Empirical Data. Academic Press, New York.

Kottegoda, N. T., 1980. *Stochastic Water Resources Technology*. Macmillan, New York. p. 34-41.

Lo, A. L., 1991. Long-term memory in stock market prices. *Econometrica* **59**(5): 1 279-1 313.

Lye, L. M. and Y. Lin, 1994. Long-term dependence in annual peak flows of Canadian rivers. *J. Hydrology* **160**: 89-103.

Mesa, O. J. and G.. Poveda, 1993. The Hurst effect: the scale of fluctuations approach. *Water Resources Research*, **29**(12): 3 995-4 002.

Montanari. A., R. Renzo, and S. M. Taqqu, 2000. A seasonal fractional ARIMA Model applied to the Nile River monthly flows at Aswan. *Water Resources Research* **36**(5): 1 249-1 259.

Patra, K. C., 2001. Hydrology and Water Resources Engineering. Alpha Science International Ltd, UK. p. 68-89.

Rao, A. R. and D. Bhattacharya, 1999. Hypothesis testing for long-term memory in hydrologic series. *J. Hydrology* **216**: 183-196.

Rosenblatt, M., 1956. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences* **42**: 27-60.

Salas, J. D., 1993. Analysis and Modelling of Hydrologic Time series. *In*: D. R. Maidement (Ed.), Handbook of Hydrology, McGraw-Hill, New York, Chapter 19.

Wallis, J. R. and N. C. Matalas, 1970. Small sample properties of H and K-estimates of Hurst's coefficients h. *Water Resources Research* **6**(6): 1 583-1 594

Wang, W., 2006. Stochasticity, non-linearity and forecasting of streamflow processes. Unpublished Ph.D Dissertation, Technische Universiteit Defts, Netherlands. p. 45-54.