Detecting long-memory: Monte Carlo simulations and application to daily streamflow processes

W. Wang, P. H. A. J. M. van Gelder, J. K. Vrijling, X. Chen

**Hydrology and
Earth System
Sciences**

Detecting long-memory: Monte Carlo simulations and application to daily streamflow processes

W. Wang[1,2], **P. H. A. J. M. Van Gelder**[2], **J. K. Vrijling**[2], **and X. Chen**[1]

[1]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, 210098, China
[2]Faculty of Civil Engineering & Geosciences, Section of Hydraulic Engineering, Delft University of Technology, 2628 CN Delft, Netherlands

Abstract. The Lo's modified rescaled adjusted range test (R/S test) (Lo, 1991), GPH test (Geweke and Porter-Hudak, 1983) and two approximate maximum likelihood estimation methods, i.e., Whittle's estimator (W-MLE) and another one implemented in S-Plus (S-MLE) based on the algorithm of Haslett and Raftery (1989) are evaluated through intensive Monte Carlo simulations for detecting the existence of long-memory. It is shown that it is difficult to find an appropriate lag q for Lo's test for different short-memory autoregressive (AR) and fractionally integrated autoregressive and moving average (ARFIMA) processes, which makes the use of Lo's test very tricky. In general, the GPH test outperforms the Lo's test, but for cases where a strong short-range dependence exists (e.g., AR(1) processes with ϕ=0.95 or even 0.99), the GPH test gets useless, even for time series of large data size. On the other hand, the estimates of d given by S-MLE and W-MLE seem to give a good indication of whether or not the long-memory is present. The simulation results show that data size has a significant impact on the power of all the four methods because the availability of larger samples allows one to inspect the asymptotical properties better. Generally, the power of Lo's test and GPH test increases with increasing data size, and the estimates of d with GPH method, S-MLE method and W-MLE method converge with increasing data size. If no large enough data set is available, we should be aware of the possible bias of the estimates.

The four methods are applied to daily average discharge series recorded at 31 gauging stations with different drainage areas in eight river basins in Europe, Canada and USA to detect the existence of long-memory. The results show that the presence of long-memory in 29 daily series is confirmed by at least three methods, whereas the other two series are indicated to be long-memory processes with two methods. The intensity of long-memory in daily streamflow processes has only a very weak positive relationship with the scale of watershed.

Correspondence to: W. Wang
(w.wang@126.com)

1 Introduction

Long-memory, or long-range dependence, refers to a not negligible dependence between distant observations in a time series. Long-memory processes can be expressed either in the time domain or in the frequency domain. In the time domain, long-memory is characterized by a hyperbolically decaying autocorrelation function. In fact, it decays so slowly that the autocorrelations can not be summed. For a stationary discrete long-memory time series process, its autocorrelation function $\rho(k)$ at lag k satisfies (Hosking, 1981)

$$\rho(k) \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1}, \text{ as } k \to \infty \qquad (1)$$

where, d is the long-memory parameter (or fractional differencing parameter), and $0<|d|<0.5$; $\Gamma(\bullet)$ is the Gamma function.

Since the early work of Hurst (1951), it has been well recognized that many time series, in diverse fields of application, such as financial time series (e.g., Lo, 1991; Meade and Maier, 2003), meteorological time series (e.g., Haslett and Raftery, 1989; Bloomfield, 1992; Hussain and Elbergali, 1999) and internet traffic time series (see Karagiannis et al., 2004), etc., may exhibit the phenomenon of long-memory. A number of models have been proposed to describe the long-memory feature of time series. The Fractional Gaussian Noise model is the first model with long-range dependence introduced by Mandelbrot and Wallis (1969). Then Hosking (1981) and Granger and Joyeux (1980) proposed the fractionally integrated autoregressive and moving average model, denoted by ARFIMA (p, d, q) that is defined by

$$\phi(B)(1-B)^d x_t = \theta(B)\varepsilon_t \qquad (2)$$

where $\{x_t\}$, t=1, 2, ..., n, is the time series; B is the backshift operator, that is, $Bx_t=x_{t-1}$; $\phi(B)$=1–$\phi_1 B$–...– $\phi_p B^P$ and $\theta(B)$=1–$\theta_1 B$ – ... – $\theta_q B^q$ represent the ordinary autoregressive and moving average components; ε_t is a white noise process with zero mean and variance σ^2. When –0.5<d<0.5, the ARFIMA (p, d, q) process is stationary,

and if $0<d<0.5$ the process presents long-memory behavior. Instead of using d, we may use $H=d+0.5 \in (0.5, 1)$ which is known as the Hurst coefficient (see Hurst, 1951) to measure the long memory in $\{x_t\}$. The larger d or H, the longer the memory the stationary process has. When $d=0$, the ARFIMA model is reduced to an ARMA(p,q) model. If $q=0$, the ARMA(p,q) model is further reduced to an AR(p) model.

In the hydrology community, many studies have been carried out on the test and modeling for long-memory in hydrological processes. Montanari et al. (1997) applied the ARFIMA model to the monthly and daily inflows of Lake Maggiore, Italy. Rao and Bhattacharya (1999) explored some monthly and annual hydrologic time series, including average monthly streamflow, maximum monthly streamflow, average monthly temperature and monthly precipitation, at various stations in the mid-western United States. They stated that there is little evidence of long-term memory in monthly hydrologic series, and for annual series the evidence for lack of long-term memory is inconclusive. Montanari et al. (2000) introduced seasonal ARFIMA model and applied it to the Nile River's monthly flows at Aswan. The resulting model indicates that nonseasonal long-memory is not present in the data. At approximately the same time, Ooms and Franses (2001) documented that monthly river flow data displays long-memory, in addition to pronounced seasonality based on simple time series plots and periodic sample autocorrelations. Wang et al. (2005) investigated the existence of long-memory in two daily streamflow series of the Yellow River in China, and found that both daily streamflow processes exhibit a strong long-memory.

This study seeks to evaluate several methods for detecting the presence of long-memory in time series and investigate the possible relationship between the intensity of long-memory in daily streamflow processes and the watershed scales. In Sect. 2, four methods used in the present study to detect long-memory will be described briefly. Simulation results with these methods are presented in Sect. 3. Then, the four methods are applied to 31 daily streamflow series to detect the existence of long-memory in Sect. 4, and finally, some conclusions are drawn in Sect. 5.

2 Methods for detecting the existence of long-memory

Many methods are available for detecting the existence of long-memory and estimating the fractional differencing parameter d. Many of them are well described in the monograph of Beran (1994). These techniques include graphical methods (e.g., classic rescaled adjusted range analysis, i.e., R/S analysis; aggregated variance method etc.), parametric methods (e.g., Whittle maximum likelihood estimation method) and semi-parametric method (e.g., GPH method and local Whittle method). Graphical methods are useful to heuristically test if there exists a long-range dependence in

the data and to find a first estimate of d, but they generally are inaccurate and sensitive to short range serial correlations. The parametric methods obtain consistent estimates of d via maximum likelihood estimation of parametric long-memory models. They give a more accurate estimate of d, but generally require knowledge of the true model which is in fact always unknown. Semi-parametric methods, such as the GPH method (Geweke and Porter-Hudak, 1983), seek to estimate d under few prior assumptions concerning the spectral density of a time series and, in particular, without specifying a finite parameter model for the d-th difference of the time series. In the present study, two statistic tests: Lo's modified R/S test, which is a modified version of classical R/S analysis, and the GPH test will be used to test for the null hypothesis of no presence of long-memory. Besides, two approximate maximum likelihood estimation methods are used to estimate the fractional differencing parameter d, but without testing for the significance level of the estimate.

2.1 Lo's modified R/S test

In classical R/S analysis, for a given time series $\{x_t\}$, $t = 1, 2, \ldots, n$, with the jth partial sum $Y_j = \sum_{i=1}^{j} x_i$, $j = 1, 2, \ldots, n$, and the sample variance $S_j^2 = j^{-1} \sum_{i=1}^{j} (x_i - j^{-1}Y_j)^2$, $j = 1, 2, \ldots, n$, the rescaled adjusted range statistic or R/S-statistic is defined by

$$R/S(j) = \frac{1}{S_j} \left[\max_{0 \le t \le j} \left(Y_t - \frac{t}{j}Y_j \right) - \min_{0 \le t \le j} \left(Y_t - \frac{t}{j}Y_j \right) \right], \quad j = 1, 2, ..., n \quad (3)$$

The classical R/S analysis is sensitive to the presence of explicit short-range dependence structures, and lacks a distribution theory for the underlying statistic. To overcome these shortcomings, Lo (1991) proposed a modified R/S statistic that is obtained by replacing the denominator S_j in Eq. (3), i.e., the sample standard deviation, by a modified standard deviation S_q which takes into account the autocovariances of the first q lags, in order to discount the influence of the short-range dependence structure that might be present in the data. Instead of considering multiple lags as in Eq. (3), only focus on lag $j = n$. The S_q is defined as

$$S_q = \left(\frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x}_n)^2 + \frac{2}{n} \sum_{j=1}^{q} \omega_j(q) \left[\sum_{i=j+1}^{n} (x_i - \bar{x}_n)(x_{i-j} - \bar{x}_n) \right] \right)^{1/2} \quad (4)$$

where $\bar{x}_n$ denotes the sample mean of the time series, and the weights $\omega_j(q)$ are given by $w_j(q)=1-j/(q+1)$, $q<n$. Then the Lo's modified R/S statistic is defined by

$$Q_{n,q} = \frac{1}{S_q} \left\{ \max_{0 \le i \le n} \sum_{j=1}^{i} (x_j - \bar{x}_n) - \min_{0 \le i \le n} \sum_{j=1}^{i} (x_j - \bar{x}_n) \right\} \quad (5)$$

If a series has no long-range dependence, Lo (1991) showed that given the right choice of q, the distribution of $n^{-1/2}Q_{n,q}$ is asymptotic to that of

$$W = \max_{0 \le r \le 1} V(r) - \min_{0 \le t \le 1} V(r),$$

where V is a standard Brownian bridge, that is, $V(r)=B(r)-rB(1)$, where B denotes standard Brownian motion. Since the distribution of the random variable W is known as

$$P(W \le x) = 1 + 2\sum_{j=1}^{\infty}(1 - 4x^2j^2)e^{-2x^2j^2}, \qquad (6)$$

Lo gave the critical values of x for hypothesis testing at sixteen significance levels using Eq. (6), which can be used for testing the null hypothesis H_0 that there is only short-term memory in a time series at a significance level α.

2.2 GPH Test

Geweke and Porter-Hudak (1983) proposed a semiparametric approach to the testing for long-memory. Given a fractionally integrated process $\{x_t\}$, its spectral density is given by:

$$f(\omega) = [2\sin(\omega/2)]^{-2d}f_u(\omega)$$

where ω is the Fourier frequency, $f_u(\omega)$ is the spectral density corresponding to u_t, and u_t is a stationary short memory disturbance with a zero mean. Consider the set of harmonic frequencies $\omega_j=(2\pi j/n)$, $j=0, 1, \ldots, n/2$, where n is the sample size. By taking the logarithm of the spectral density $f(\omega)$ we have

$$\ln f(\omega_j) = \ln f_u(\omega_j) - d\ln\left[4\sin^2\left(\omega_j/2\right)\right]$$

which may be written in the alternative form

$$\ln f(\omega_j) = \ln f_u(0) - d\ln\left[4\sin^2(\omega_j/2)\right] + \ln\left[f_u(\omega_j)/f_u(0)\right] \quad (7)$$

The fractional difference parameter d can be estimated by the regression equations constructed from Eq. (7). Geweke and Porter-Hudak (1983) showed that using a periodogram estimate of $f(\omega_j)$, if the number of frequencies used in the regression Eq. (7) is a function $g(n)$ (a positive integer) of the sample size n where $g(n)=n^\alpha$ with $0<\alpha<1$, the least squares estimate $\hat{d}$ using the above regression is asymptotically normally distributed in large samples:

$$\hat{d} \sim N(d, \frac{\pi^2}{6\sum_{j=1}^{g(n)}(U_j - \overline{U})^2})$$

where $U_j=\ln[4\sin^2(\omega_j/2)]$ and $\overline{U}$ is the sample mean of U_j, $j=1,\cdot\quad\cdot\quad\cdot, g(n)$. Under the null hypothesis of no long-memory ($d=0$), the t-statistic

$$t_{d=0} = \hat{d}\cdot\left(\frac{\pi^2}{6\sum_{j=1}^{g(n)}(U_j - \overline{U})^2}\right)^{-1/2}$$

has a limiting standard normal distribution.

2.3 Maximum likelihood estimation (MLE) of fractional differencing parameter d

The Gaussian log-likelihood of a long-memory ARFIMA process X defined by Eq. (2) is given by

$$\log L(\eta, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}X^t\Sigma^{-1}X \qquad (8)$$

where $\eta = (\phi_1, \ldots, \phi_p; d; \theta_1, \ldots, \theta_q)$ is the parameter vector; Σ denotes the $n\times n$ covariance matrix of X depending on η and σ^2, and $|\Sigma|$ denotes the determinant of Σ. The maximum likelihood estimators $\hat{\eta}$ and $\hat{\sigma}^2$ can be found by maximizing $\log L(\eta, \sigma^2)$ with respect to η and σ^2.

Due to the computational problems (Beran, 1994, 108–109), some approximate MLE's are needed. In this study, two approximate maximum likelihood estimators are applied: (1) the maximum likelihood estimation method implemented in S-Plus version 6 (referred to as S-MLE) is used to estimate the fractional differencing parameter d. S-MLE is implemented based on the approximate Gaussian maximum likelihood algorithm of Haslett and Raftery (1989); (2) Whittle's estimator (referred to as W-MLE), which is the value of the vector that minimizes the function:

$$L_W(\eta) = \int_{-\pi}^{\pi}\frac{I(\lambda)}{f(\lambda; \eta)}d\lambda + \int_{-\pi}^{\pi}\log f(\lambda; \eta)d\lambda,$$

where the subscript W stands for Whittle; $f(\lambda; \eta)$ is the spectral density and $I(\lambda)$ is the periodogram of the process. The S-Plus code of Whittle's estimator is available in the book of Beran (1994). A modified version by M. S. Taqqu is available at http://math.bu.edu/people/murad/methods/whittle/. If the estimated d obtained with S-MLE or W-MLE is significantly greater than zero, we consider it an evidence of the presence of long-memory.

3 Monte Carlo simulations

3.1 Implementation of Monte Carlo simulations

An extensive Monte Carlo investigation is performed in order to find out how reliable the Lo's test, the GPH test, the S-MLE method and the W-MLE method are for detecting the long-memory property of AR and ARFIMA processes. We consider five AR(1) and six ARFIMA(1,d,0) processes. All AR(1) models are of the form $(1-\phi B)x_t=\varepsilon_t$, and all ARFIMA(1,d,0) of form $(1-B)^d(1-\phi B)x_t=\varepsilon_t$, where $\{\varepsilon_t\}$ are i.i.d standard normal, and B is the backshift operator. For the AR models, large autoregressive coefficients, i.e., $\phi=0.5, 0.8, 0.9, 0.95$ and 0.99, are considered, because these are commonly seen cases in streamflow processes. For the ARFIMA models, $\phi=0, 0.5, 0.9$ and $d=0.3, 0.45$. We generate 500 simulated realizations of size 500, 1000, 3000, 10 000 and 20 000, respectively, for each model. The AR series and the ARFIMA series are produced with the *arima.sim* and *arima.fracdiff.sim* function built in S-Plus version 6.

www.hydrol-earth-syst-sci.net/11/851/2007/

Hydrol. Earth Syst. Sci., 11, 851–862, 2007

For Lo's modified R/S test, the right choice of q is essential. It must be chosen with some consideration of the data at hand. Some simulation studies (Lo, 1991; Teverovsky et al., 1999) have shown that the probability of accepting the null hypothesis varied significantly with q. In general, the larger the q, the less likely the null hypothesis is to be rejected. One appealing data-driven formula (Andrew, 1991) for choosing q based on the assumption that the true model is an AR(1) model is given by

$$q = \left[\left(\frac{3n}{2} \right)^{1/3} \left(\frac{2\hat{\rho}}{1 - \hat{\rho}^2} \right)^{2/3} \right]$$

where $[\bullet]$ denotes the greatest integer function, n is the length of the data, $\hat{\rho}$ is the estimated first-order autocorrelation coefficient. However, our simulation for AR processes and ARFIMA processes with different intensity of dependence indicates that this data-driven formula is too conservative in rejecting the null hypothesis of no long-memory, especially for cases where autocorrelations at lag 1 are high. After a trial-and-error procedure, we use the following modified formula to choose the lag q:

$$q = \left[\left(\frac{n}{10} \right)^{1/4} \left(\frac{2\hat{\rho}}{1 - \hat{\rho}^2} \right)^{2/3} \right], \qquad (9)$$

where $\hat{\rho}$ is the autocorrelation at lag 1, i.e., ACF(1). This modified formula is a trade-off between lowering the probability of wrongly rejecting the null hypothesis of no long-memory for AR processes, and reserving the power of correctly rejecting the null hypothesis for ARFIMA processes. The null hypothesis of no long-memory is rejected at a 5% significance level if $Q_{n,q}$ is not contained in the interval [0.809, 1.862] (Lo, 1991).

Similarly to the case with Lo's test, there is a choice of the number of frequencies $g(n)$ used in the regression Eq. (7) for the GPH test. This choice entails a bias-variance trade-off. For a given sample size, as $g(n)$ is increased from 1, the variance of the d estimate decreases, but this decrease is typically offset by the increase in bias due to non-constancy of $f_u(\omega)$. Geweke and Porter-Hudak (1983) found that choosing $g(n)=n^{0.5}$ gave good results in simulation. We adopt such a criterion in our Monte Carlo simulation study. The periodogram used for calculating GPH test statistic is smoothed with a modified Daniell smoother of length 5. The null hypothesis of no long-memory ($d=0$) is rejected at a 5% significance level if the t-statistic is not contained in the interval [−1.960, 1.960].

When estimating the parameter d with S-MLE method and W-MLE method, the order p of the AR component for a given time series process is usually unknown. One way to estimate the order p of the AR component of an ARFIMA process is using the AIC criterion (Akaike, 1973). But in the present study, we assume that the order p of the AR component for each simulated ARFIMA process is known

(here p=1) in order to make the comparison between S-MLE method and W-MLE method clear. In the S-MLE method, the d is optimized in the range (0, 0.5).

3.2 Results of simulations

The results of detecting long-memory in simulated AR and ARFIMA processes of sizes ranging from 500 to 20 000 with Lo's test, GPH test, S-MLE and W-MLE method based on 500 Monte Carlo simulations are reported in Table 1. For Lo's test, we list the average values of the lags chosen with the data-driven Eq. (9) (denoted as "average lag"), the standard deviations of the lags ("SD of lag"), and the number of acceptance of the null hypothesis for 500 simulations. For GPH test, we list the average values of the estimates of d ("average d"), the standard deviations of the estimates ("SD of d"), and the number of acceptance of the null hypothesis for 500 simulations. For both the S-MLE method and the W-MLE method, the averages and standard deviations of the estimates of d ("average d" and "SD of d") are reported. According to the results of simulated AR and ARFIMA processes, shown in Table 1, we have the following findings:

(1) For AR(1) processes, when the autocorrelation is less than 0.9, both the Lo's R/S test and the GPH test work well, but the GPH test has a better performance. However, when the autoregressive coefficient is higher than 0.9, the probability of committing Type I error with the GPH test increases very fast, and the GPH test gets useless for the cases where ϕ is above 0.97 (to save space, the results with ϕ=0.97 are not presented in Table 1), even for the size of 20 000 points. In contrast, the probability of committing Type I error with the Lo's R/S test is still considerably low even for AR processes with a ϕ of as high as 0.99.

(2) For ARFIMA(1,d,0) processes, the GPH technique yields negatively biased estimates of d when an AR component of low autoregressive coefficient value (e.g., $\phi \leq 0.5$) is present, whereas it yields positively biased estimates of d when an AR component of high autoregressive coefficient value (e.g., ϕ=0.9) is present. This seems to be not in agreement with the results of Sowell (1992), who showed that, when the sample length is small, the GPH technique yields upwardly biased estimates of d when AR and MA terms are present. On the other hand, the power of GPH test increases with the increase of data size, the intensity of long-memory, and autocorrelations of their AR components. For cases where the data size is over 10 000, the probability of committing Type II error, i.e., false acceptance of the null hypothesis of no long-memory, by the GPH test is close to zero. In contrast, the Lo's test only performs slightly better than the GPH test when the intensity of long-memory is not strong and the value of ϕ in the AR component is low, but for the cases of strong intensity of long-memory and with an AR component of strong autocorrelation, the Lo's test performs far less powerful than the GPH test.

Table 1. Long-memory test results for simulated AR and ARFIMA series.

Model	Data size	Lo's R/S test				GPH test			S-MLE		W-MLE	
		average lag	SD of lag	accepted	average d	SD of d	accepted	average d	SD	average d	SD
AR(1)	500	2.8	0.5	464	−0.0167	0.1302	495	0.0178	0.0378	−0.1060	0.1490
ar = .5	1000	3.2	0.4	454	−0.0123	0.1141	490	0.0146	0.0280	−0.0612	0.1071
	3000	4.6	0.5	468	−0.0124	0.0772	490	0.0101	0.0183	−0.0198	0.0492
	10 000	6.1	0.2	455	−0.0119	0.0607	490	0.0067	0.0112	−0.0069	0.0238
	20 000	7.8	0.4	469	−0.0078	0.0479	488	0.0047	0.0078	-0.0043	0.0159
AR(1)	500	6.7	0.8	428	0.1220	0.1388	470	0.0395	0.0733	0.0010	0.1127
ar = .8	1000	8.0	0.7	442	0.0637	0.1110	489	0.0272	0.0517	−0.0032	0.0776
	3000	10.8	0.5	441	0.0163	0.0827	490	0.0159	0.0271	−0.0027	0.0431
	10 000	14.7	0.5	441	−0.0016	0.0605	490	0.0084	0.0131	−0.0016	0.0226
	20 000	17.6	0.5	454	−0.0036	0.0511	483	0.0062	0.0099	−0.0010	0.0165
AR(1)	500	11.3	1.6	431	0.3252	0.1342	268	0.0341	0.0702	0.0711	0.2062
ar = .9	1000	13.5	1.4	408	0.2189	0.1135	326	0.0193	0.0379	0.0550	0.1902
	3000	18.1	1.1	414	0.0957	0.0851	436	0.0098	0.0162	0.0105	0.0969
	10 000	24.6	0.8	441	0.0273	0.0600	483	0.0054	0.0082	0.0001	0.0303
	20 000	29.4	0.7	457	0.0107	0.0500	489	0.0040	0.0063	−0.0005	0.0105
AR(1)	500	18.7	3.6	451	0.5739	0.1395	24	0.0286	0.0510	0.0082	0.0614
ar = .95	1000	22.4	3.1	429	0.4488	0.1154	34	0.0158	0.0225	0.0022	0.0364
	3000	29.6	2.4	426	0.2594	0.0800	91	0.0102	0.0139	0.0023	0.0225
	10 000	40.3	1.8	416	0.1201	0.0601	300	0.0051	0.0076	0.0004	0.0122
	20 000	47.9	1.6	416	0.0665	0.0475	409	0.0034	0.0051	0.0000	0.0084
AR(1)	500	52.9	20.3	494	0.9122	0.1617	0	0.0207	0.0268	0.0082	0.0427
ar = .99	1000	65.3	19.3	484	0.8530	0.1226	0	0.0124	0.0170	0.0030	0.0284
	3000	86.8	14.7	399	0.7297	0.0826	0	0.0068	0.0094	0.0005	0.0168
	10 000	119.7	11.9	389	0.5555	0.0583	0	0.0037	0.0052	0.0002	0.0091
	20 000	142.4	9.5	380	0.4478	0.0477	0	0.0025	0.0037	−0.0001	0.0065
ARFIMA	500	2.2	0.5	129	0.2587	0.1360	353	0.2663	0.0638	0.2720	0.0784
d = 0.3	1000	2.8	0.5	61	0.2749	0.1157	228	0.2852	0.0427	0.2890	0.0437
	3000	3.8	0.5	15	0.2821	0.0826	68	0.2971	0.0233	0.2969	0.0233
	10 000	5.2	0.4	0	0.2884	0.0572	2	0.3018	0.0128	0.2985	0.0124
	20 000	6.3	0.5	0	0.2900	0.0470	0	0.3035	0.0093	0.2992	0.0089
ARFIMA	500	7.1	1.4	255	0.2729	0.1402	333	0.1951	0.1258	0.1964	0.1615
ar = 0.5	1000	8.6	1.3	139	0.2783	0.1130	233	0.2441	0.0928	0.2473	0.1086
d = 0.3	3000	11.4	1.2	63	0.2878	0.0919	83	0.2881	0.0496	0.2864	0.0502
	10 000	15.6	1.0	8	0.2934	0.0604	4	0.3027	0.0241	0.2944	0.0229
	20 000	18.6	0.9	5	0.2955	0.0493	0	0.3083	0.0172	0.2973	0.0160
ARFIMA	500	41.1	12.2	493	0.6375	0.1513	16	0.2820	0.0620	0.3795	0.2251
ar = 0.9	1000	49.4	11.6	478	0.5213	0.1123	6	0.2903	0.0460	0.3686	0.2078
d = 0.3	3000	65.4	11.2	345	0.3964	0.0881	5	0.2974	0.0278	0.3360	0.1536
	10 000	89.4	9.2	155	0.3316	0.0627	2	0.3008	0.0158	0.3261	0.1270
	20 000	106.6	8.3	78	0.3145	0.0512	0	0.3014	0.0108	0.3099	0.0776
ARFIMA	500	7.0	4.0	130	0.4077	0.1506	157	0.3926	0.0544	0.4228	0.0676
d = 0.45	1000	8.5	4.4	56	0.4274	0.1237	53	0.4192	0.0362	0.4372	0.0431
	3000	11.2	5.2	11	0.4371	0.0873	0	0.4414	0.0218	0.4468	0.0235
	10 000	15.4	6.0	0	0.4373	0.0613	0	0.4542	0.0137	0.4493	0.0126
	20 000	18.6	7.0	0	0.4371	0.0489	0	0.4588	0.0111	0.4498	0.0094
ARFIMA	500	19.1	10.1	346	0.4331	0.1515	133	0.2874	0.1370	0.3452	0.1636
ar = 0.5	1000	22.9	10.6	204	0.4385	0.1164	33	0.3627	0.0976	0.3973	0.1142
d = 0.45	3000	31.0	12.2	66	0.4404	0.0893	3	0.4290	0.0425	0.4397	0.0496
	10 000	42.4	14.6	11	0.4429	0.0635	0	0.4579	0.0218	0.4482	0.0239
	20 000	50.2	16.2	4	0.4459	0.0507	2	0.4683	0.0162	0.4498	0.0157
ARFIMA	500	135.0	78.5	493	0.7956	0.1394	2	0.3806	0.1040	0.5035	0.2649
ar = 0.9	1000	163.4	90.2	495	0.6733	0.1172	1	0.4230	0.0775	0.5087	0.2384
d = 0.45	3000	222.9	116.2	472	0.5539	0.0878	0	0.4651	0.0386	0.4629	0.1533
	10 000	299.5	138.7	273	0.4856	0.0599	0	0.4690	0.0262	0.4465	0.0654
	20 000	361.8	158.0	140	0.4666	0.0491	0	0.4695	0.0225	0.4485	0.0502

(3) It seems difficult to choose an appropriate lag for Lo's test that is valid for all cases. For the cases where the data sizes are less than 3000, while the lag chosen by Eq. (9) seems to be already very large and cannot get larger without a high probability of wrongly rejecting the null hypothesis of no long-memory for AR processes, the lag seems not
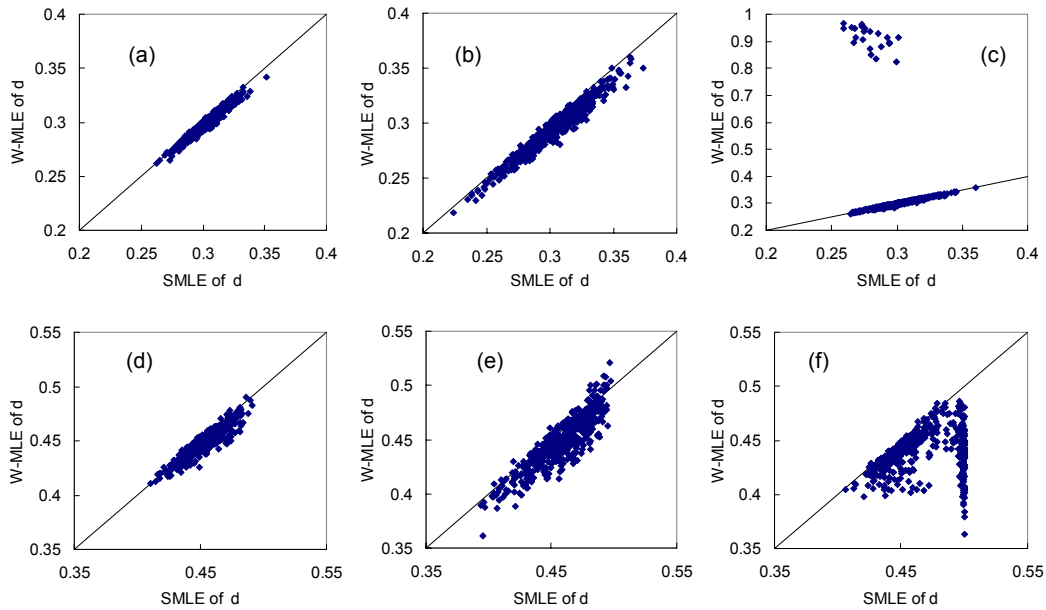
Fig. 1. Estimates of d with S-MLE versus W-MLE for 500 simulations of **(a)** ARFIMA (0,0.3,0); **(b)** ARFIMA(0.5,0.3,0); **(c)** ARFIMA(0.9,0.3,0) **(d)** ARFIMA (0,0.45,0); **(e)** ARFIMA(0.5,0.45,0); **(f)** ARFIMA(0.9,0.45,0) (Note: the straight line has a slope of 0.5).

to be large enough to avoid the high probability of wrong acceptance of the null hypothesis for ARFIMA processes. The good news is that the lag chosen by Eq. (9) works well when the data size is over 10^4, especially when the value of ϕ in the AR component is low (e.g., $\phi \leq 0.5$). But for AR(1) processes with high autoregressive coefficients and ARFIMA(1,d,0) processes with high values of ϕ in their AR components, the lag chosen by Eq. (9) seems to be too small for AR series of large size, but not large enough for ARFIMA processes. Namely, no good trade-off can be achieved in choosing an appropriate lag for the Lo's test. This result further substantiates the limitation of the use of the Lo's test, which has been shown in the previous study of Teverovsky et al. (1999).

(4) The estimates of d given by both the S-MLE method and the W-MLE method seem to give a good indication of whether or not the long-memory is present, and the estimates given by the two MLE methods have less standard deviation than those given by the GPH method in general (even taking the fact that d is optimized in the range (0, 0.5) in S-MLE into account). Therefore, the estimate of d given by S-MLE or W-MLE could be used to detect the presence of long-memory, and the estimates given by the two MLE methods are more accurate than the GPH method. It is shown by our simulation study that:

1. For AR(1) processes, both S-MLE and W-MLE give basically correct estimates of d, i.e., d=0, even when the autoregressive coefficients are very high. Although the estimates tend to be slightly positively biased when the

data size is small (e.g., 500 points) with both methods, the estimates get more accurate (according to the averages) and more stable (according to the standard deviations) with the increase of sample size.

2. For ARFIMA processes, S-MLE provides significantly negatively biased estimates when the data size is small (e.g., less than 10^3). The estimates of d given by S-MLE increase with increasing sample size and are basically correct when the data size is close to 10^4. But the estimates of d get upwardly biased when the data size is too big (say, $>10^4$). This is in contradiction with the result of Kendziorski (1999), who showed that S-MLE provided unbiased estimates of d for ARFIMA(0,d,0) processes of length 2^{11} (2048) or greater.

3. The estimates given by W-MLE are also significantly biased when the data size is small (e.g., less than 3×10^3) for ARFIMA processes. But the direction of bias may depend on the autoregressive coefficient of the AR component of the ARFIMA processes. For an ARFIMA process with d=0.3 and 0.45, when the autoregressive coefficient of the AR component in the ARFIMA process is equal to or less than 0.5, the estimates of d are negatively biased, whereas they are positively biased when the autoregressive coefficient is high (e.g., 0.9).

(5) We plotted the estimates of S-MLE versus the estimates with W-MLE for 500 simulations of 6 different

ARFIMA models in Fig. 1. The estimates are all based on the data size of 10^4. Comparing the estimates of S-MLE and W-MLE, we see that, although the estimates given by S-MLE are in general slightly larger than by W-MLE, the estimates of the two estimators are generally consistent when the autoregressive coefficient ϕ of the AR component in the ARFIMA process is not high (i.e., $\leq$0.5), especially when the size of d is moderate (i.e., d=0.3). But the estimation given by either S-MLE or W-MLE often goes wrong for ϕ = 0.9 (see Figs. 1c, f). The estimates given by S-MLE tend to be close to 0.5 for ARFIMA(0.9, 0.45, 0) processes (notice that the range of d is optimized between 0 and 0.5 with S-MLE), whereas the estimates given by W-MLE sometimes are larger than 0.8 for ARFIMA (0.9, 0.3, 0) processes. This indicates that a value close to 0.5 given by S-MLE or a value higher than 0.5 given by W-MLE may be due to the presence of strong short-range dependence in the process. On the other hand, according to the results for ARFIMA(0.9, 0.45, 0) processes, shown in Fig. 1f, estimates of d given by W-MLE are concentrated in the region around 0.45, whereas many estimates of d given by S-MLE are close to 0.5, indicating that W-MLE seems to work better than S-MLE for cases where both the strong short-range dependence and the strong long-range dependence exist.

(6) Data size has a significant impact on the power of all the four methods. Generally, the power of Lo's test and GPH test increases with the increase of data size, and the estimates of d with GPH method, S-MLE method and W-MLE method converge with the increase of data size. This is a result that can be proven through theoretical reasoning because the availability of larger samples allows one to inspect the asymptotical properties better, but is sometimes not well-recognized in previous literature. For instance, Agiakloglou et al. (1993) found that GPH estimators performed poorly for AR(1) processes with ϕ=0.9 for sample size of 100 to 900. The simulation results of Hurvich and Beltrao (1993) also showed the poor performance of the GPH estimator when ϕ=0.9 for not only AR(1) processes but also ARFIMA(1,d,0) processes. In our simulation study, it is shown that, on one hand, the power of GPH test does decrease with the increase of the autoregressive coefficient; on the other hand, its power increases with the increase of sample size. If we use a sample size larger than 10^4 points, GPH test still performs very well for AR(1) processes with ϕ=0.9. But the use of GPH test is useless when ϕ is larger than 0.95, even with a data size of larger than 10^4. One possible solution could be to choose the number of frequencies used in the regression Eq. (7) more carefully (Giraitis et al., 1997; Hurvich and Deo, 1999). But the effectiveness of these methods seems to be limited. For example, when Hurvich and Deo (1999) proposed the plug-in method to choose the number of frequencies $g(n)$ in the GPH test, they also showed that as ϕ increases, the estimates of d using the number of frequencies $g(n)$ selected by the plug-in method are much more positively biased than simply using $g(n)$=$n^{1/2}$.

On the basis of above findings, we have two suggestions to obtain reliable test results on detecting the presence of long-memory: Firstly, use a large enough data set (e.g., of size 3000$\sim$10 000) when detecting the existence of long-memory, especially for the cases where data exhibit strong serial dependence. If no large enough data set available, we should be aware of the possible bias of the estimates. For example, the estimate with S-MLE has a significant negative bias when the data size is less than 3000, and the estimates with W-MLE as well as GPH method may be negatively/positively biased when the autoregressive coefficient of the AR component in the ARFIMA process is small (e.g., $\leq$0.5)/big (e.g., 0.9). But notice that S-MLE has a positive bias in general when the data size is greater than 10^4. Therefore, the most appropriate date size for estimating d with S-MLE may be slightly less than 10^4. Secondly, use the methods in combination with each other for detecting the existence of long-memory. Here we consider the combined use of Lo's test, GPH-test, S-MLE and W-MLE. According to the simulation results, the combined use of these four methods produces the following alternatives:

1. Failure to reject by both the Lo's test and the GPH-test, and low values of estimated d (e.g., <0.1) with S-MLE and W-MLE, provide evidence in favour of no existence of long-memory;

2. Rejection by both Lo's test and GPH test, and high values of estimated d(e.g., >0.2) with S-MLE and W-MLE support that the series is a long-memory process;

3. For other cases, the data are not sufficiently informative with respect to the long-memory properties of the series. But if the GPH test, S-MLE and W-MLE give positive results in detecting the existence of long-memory, then we may consider the long-memory is present whatever the result given by the Lo's test, and vice versa.

4 Results with daily streamflow data

4.1 Daily streamflow data used

Daily average discharge series recorded at 31 gauging stations in eight basins in Europe, Canada and USA are analyzed in the present study. The data come from Global Runoff Data Centre (GRDC) (http://grdc.bafg.de), US Geological Survey Water Watch (http://water.usgs.gov/waterwatch), and Water Survey of Canada (http://www.wsc.ec.gc.ca). We generally have the following three rules to select stations in each basin:

1. The selection of basins covers different geographical and climatic regions;

Table 2. Description of selected daily streamflow time series.

No.	Basin	Location of gauging stations	Area (km^2)	Latitude.	Longitude	Elevation (m)	Period	Average discharge (m^3/s)
Color-1	Colorado	Colorado River At Lees Ferry	289,400	36.865	−111.588	946.8	1922–1951	489.1
Color-2		Colorado River Near Cisco	62 390	38.811	−109.293	1246.6	1923–1952	222.3
Color-3		Colorado River Near Kremmling	6167	40.037	−106.439	2231.1	1904-1918	52.3
Color-4		Williams Fork Near Parshall	476	40.000	−106.179	2380.2	1904–1924	4.9
Colum-1	Columbia	Columbia River At The Dalles	613,565	45.108	−121.006	0.0	1880-1909	6065.7
Colum-2		Columbia River at Trail	88 100	49.094	−117.698	–	1914–1936	2029.4
Colum-3		Columbia River at Nicholson	6660	51.244	−116.912	–	1933–1962	107.5
Colum-4		Columbia River Near Fairmont Hot Springs	891	50.324	−115.863	–	1946–1975	11.1
Danu-1	Danube	Danube river at Orsova	576232.	44.700	22.420	44	1901–1930	5711.9
Danu-2		Danube river at Achleiten	76653.	48.582	13.504	288	1901–1930	1427.0
Danu-3		Inn river at Martinsbruck	1945.	46.890	10.470	–	1904-1933	57.8
Fras-1	Fraser	Fraser River at Hope	217,000	49.381	−121.451	–	1913–1942	2648.8
Fras-2		Fraser River at Shelley	32 400	54.011	−122.617	–	1950-1979	825.3
Fras-3		Fraser River at Mcbride	6890	53.286	−120.113	–	1959–1988	197.3
Fras-4		Canoe River below Kimmel Creek	298	52.728	−119.408	–	1972–1994	14.5
Missi-1	Mississippi	Mississippi River At Vicksburg	2962,974	32.315	−90.906	14.1	1932–1961	16 003.1
Missi-2		Mississippi River at Clinton	221 608	41.781	−90.252	171.5	1874–1903	1477.3
Missi-3		Minnesota River At Mankato	38,574	44.169	−94.000	228.0	1943–1972	94.9
Missi-4		Minnesota River At Ortonville	3003	45.296	−96.444	291.5	1943–1972	3.4
Misso-1	Missouri	Missouri River at Hermann	1 353 000	38.710	−91.439	146.8	1929-1958	2162.0
Misso-2		Missouri River at Bismarck,	482 776	46.814	−100.821	493.0	1929–1953	604.6
Misso-3		Missouri River at Fort Benton	64,070	47.818	−110.666	796.8	1891–1920	219.7
Misso-4		Madison River near McAllister	5,659	45.490	−111.633	1429.2	1943–1972	50.5
Ohio-1	Ohio	Ohio River At Metropolis	525 500	37.148	-88.741	84.2	1943–1972	7567.5
Ohio-2		Ohio River at Sewickley	50 480	40.549	−80.206	207.3	1943–1972	922.4
Ohio-3		Tygart Valley River At Colfax	3529	39.435	−80.133	261.0	1940–1969	72.4
Ohio-4		Tygart Valley River Near Dailey	479	38.809	−79.882	591.3	1940–1969	9.2
Rhine-1	Rhine	Rhine at Lobith	160 800	51.840	6.110	8.5	1911–1940	2217.8
Rhine-2		Rhine at Rheinfelden	34 550	47.561	7.799	259.6	1931–1960	1017.3
Rhine-3		Rhine at Domat/Ems	3229	46.840	9.460	562.0	1911–1940	126.9
Rhine-4		Emme River at Emmenmatt	443	46.960	7.740	–	1915-1944	12.0

2. The drainage area of each station is basically within 5 different watershed scales, namely, $<10^3$ km^2; $10^3 \sim 10^4$ km^2; $10^4 \sim 10^5$ km^2; $10^5 \sim 10^6$ km^2; $>10^6$ km^2;

3. The stations are located in the main river channel of the river if possible so that river flows come from the same origin. When stations at the main channel are not available, stations at major tributaries are used.

For each station, we select a segment of historical daily streamflow records of mostly 30 years long. However, because of data limitations, the shortest series covers a period of only 14 years. The segments are chosen with the following criteria:

(1) The series should be approximately stationary, as least by visual inspection. Stationarity is our primary criterion because, when certain types of non-stationarity are present, many long-memory parameter estimators may fail (Klemes, 1974).

(2) The recording period of the data should be as early in time as possible, assuming that the influence of human intervention would be less intensive in an earlier period (in

early 20th century or even late 19th century) than in later period.

(3) The temporal spans of streamflow series at different locations in one basin should be as close as possible, so as to avoid possible impacts of regional low-frequency climatic variations.

The description of selected stations and their corresponding daily streamflow series is listed in Table 2.

4.2 Results

The Lo's modified R/S test and the GPH test are carried out with S+FinMetrics, a module of S-plus, for financial time series analysis (Zivot and Wang, 2003). To alleviate the impact of skewed distribution and the seasonality in mean values and standard deviations of daily discharge series, it is common to firstly log-transform the series and then deseasonalize the series by subtracting the daily means and dividing by the daily standard deviations (see e.g., Wang et al., 2006). The daily means and the daily standard deviations are calculated on the basis of each day over the year, and then smoothed with their first 8 Fourier harmonics.

Table 3. Results of long-memory detection for daily streamflow series.

No.	data size	ACF(1)	Lo's R/S test				GPH test		S-MLE	W-MLE
			Lag-1	Stat-1	Lag-2	Stat-2	d	Stat	d	d
Color-1	10957	0.9738	64	2.9566	50	3.2475	0.5125	7.5412	0.4478	0.4287
Color-2	10958	0.9627	50	3.4320	50	3.4320	0.4906	7.2192	0.4506	0.4386
Color-3	5113	0.9431	31	2.1437	50	1.8067	0.4766	5.6613	0.4863	0.5390
Color-4	7305	0.9549	40	1.1811	50	1.0826	0.4043	5.3169	0.0000	0.7390
Colum-1	10957	0.9910	132	1.5357	50	2.1519	0.5071	7.4617	0.4615	0.4433
Colum-2	8401	0.9966	238	1.1342	50	1.8357	0.4673	6.3838	0.4187	0.3642
Colum-3	10957	0.9778	72	3.1202	50	3.5159	0.3466	5.1010	0.4392	0.4042
Colum-4	10957	0.9676	55	1.8590	50	1.9213	0.3642	5.3600	0.4213	0.3939
Danu-1	10957	0.9931	158	1.5328	50	2.0899	0.3441	5.0639	0.2634	0.2001
Danu-2	10957	0.9577	46	1.9412	50	1.8957	0.3017	4.4398	0.3598	0.3557
Danu-3	10958	0.9326	33	3.1827	50	2.7771	0.3782	5.5651	0.4059	0.3834
Fras-1	10957	0.9772	70	1.5279	50	1.6994	0.3879	5.7077	0.3878	0.3318
Fras-2	10958	0.9734	63	2.9821	50	3.1849	0.2511	3.6952	0.3529	0.2720
Fras-3	10958	0.9582	47	2.3767	50	2.3411	0.2272	3.3430	0.1886	0.1436
Fras-4	8401	0.9294	30	2.2163	50	1.9096	0.2769	3.7833	0.3100	0.2912
Missi-1	10958	0.9961	232	1.8789	50	3.0163	0.4133	6.0813	0.3909	0.3594
Missi-2	10956	0.9921	144	2.6780	50	3.7589	0.3846	5.6601	0.4001	0.3654
Missi-3	10958	0.9917	139	1.8277	50	2.6476	0.5098	7.5018	0.4847	0.5463
Missi-4	10958	0.9563	45	2.7527	50	2.6345	0.5358	7.8847	0.0000	0.7818
Misso-1	10958	0.9711	60	3.6930	50	3.9396	0.4484	6.5985	0.4238	0.4033
Misso-2	9131	0.9805	75	3.6145	50	4.1707	0.4639	6.4915	0.4124	0.3678
Misso-3	10958	0.9165	29	5.1261	50	4.1325	0.4179	6.1498	0.0000	0.7054
Misso-4	10958	0.9522	42	3.2612	50	3.0869	0.2450	3.6050	0.0000	0.1070
Ohio-1	10958	0.9723	62	1.7652	50	1.8735	0.2910	4.2822	0.2983	0.2813
Ohio-2	10958	0.9547	44	2.1173	50	2.0477	0.2569	3.7810	0.2581	0.2450
Ohio-3	10958	0.9291	32	1.7894	50	1.6164	0.3289	4.8401	0.2263	0.2123
Ohio-4	10958	0.8985	25	1.9601	50	1.5937	0.3659	5.3839	0.3324	0.3353
Rhine-1	10957	0.9897	120	1.2813	50	1.6822	0.3787	5.5729	0.4254	0.4296
Rhine-2	10958	0.9715	61	2.0457	50	2.1880	0.3513	5.1699	0.0000	0.6647
Rhine-3	10958	0.9048	26	2.1554	50	1.7478	0.3792	5.5799	0.4176	0.4214
Rhine-4	10958	0.8739	21	2.2409	50	1.7306	0.2489	3.6627	0.3447	0.3300

Note: In the Lo's R/S test, lag-1 is determined by the data-driven formula, lag-2 is the fixed lag, and, stat-1 and stat-2 are their corresponding test statistics.

For Lo's modified R/S test, both a fixed lag (i.e., 50) and a lag determined by the data-driven formula (Eq. 9) are used. For GPH test, we choose $g(n)=n^{0.5}$ as suggested by Geweke and Porter-Hudak (1983). When using S-MLE to estimate the fractional differencing parameter d, the order p of the AR component in ARFIMA (p,d,q) model is determined by the AIC criteria (Akaike, 1973). The same AR order is used for the W-MLE method. The results of detecting long-memory in daily streamflow processes are reported in Table 3, which show the following:

1. The Lo's test indicates that about 1/3 (11 according to the data-driven lag, and 9 according to the fixed lag) of all the 31 streamflow series do not exhibit long-memory property; GPH test results are positive in detecting the presence of long-memory for all the streamflow series;

estimates of d given by S-MLE indicate that 5 out of all the series have d's of zero value, implying the absence of long-memory in these 5 series; estimates of d given by W-MLE are all significantly larger than 0, but one has a small value of 0.1, and four have values over 0.5. In summary, evidence of the existence of long-memory is found with all the four methods for about half of all the streamflow processes; except for one case of Williams Fork of the Colorado River basin near Parshall where both the results of Lo's test and S-MLE indicating the absence of long-memory, and another case of the Madison River of the Missouri River basin near McAllister where the estimate of d given by S-MLE is zero and the estimate of d by W-MLE is close to zero ($\approx$0.1), at least three methods applied here give
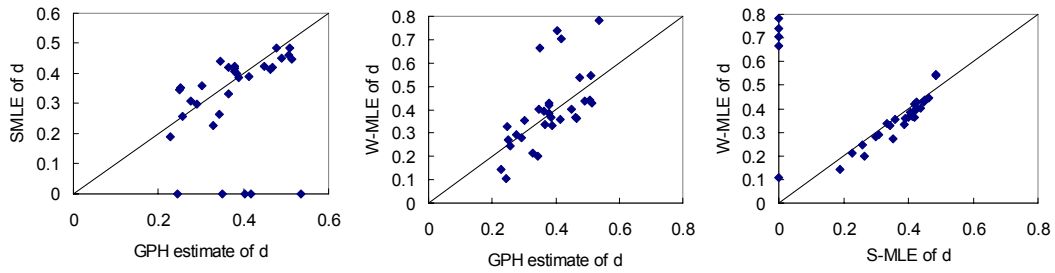
Fig. 2. Comparison of different estimates for observed daily streamflow processes (Note: the straight line has a slope of 0.5).
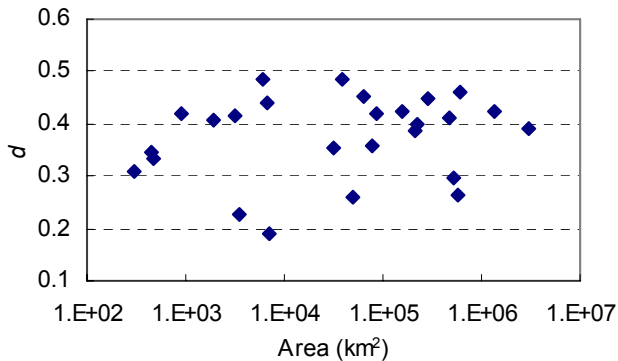

Fig. 3. The estimates of d given by S-MLE versus watershed scale for streamflow processes.

evidences of the existence of long-memory in the other 29 daily streamflow processes.

2. Teverovsky et al. (1999) pointed out that, picking a single value of q with Lo's test to determine whether or not to reject the null hypothesis of no long-range dependence in a given data set is highly problematic. Consequently, they recommended that one always relies on a wide range of different q-values, and does not use Lo's method in isolation, instead, always uses it in conjunction with other graphical and statistical techniques for checking for long-memory, especially when Lo's method results in accepting the null hypothesis of no long-range dependence. While we agree that we should not use Lo's method in isolation, it is doubtful that using a wide range of different q-values may improve the test reliability. With a wide range of q-values, we are still not sure which one gives the right answer, as shown here in the cases for detecting long-memory in daily streamflow series. In addition, Lo's test sometimes may give results that are not in agreement with all other methods with respect to the intensity of long-memory. For example, with either the data-driven value of lag q or the fixed value of lag q, the Lo's test indicates that the daily streamflow process of the Rhine River at Lobith

is a short-memory process, whereas all the other three methods indicate that it exhibits long-memory.

3. The estimates of d given by the GPH method, S-MLE method and W-MLE method are in good agreement, as shown in Fig. 2, especially between the estimates given by S-MLE and W-MLE, except for five series for which the estimates of d given by S-MLE are zero, whereas the estimates of d given by *W-MLE* are either higher than 0.6 or lower than 0.2. The estimates of zero given by S-MLE are possibly due to its erroneousness, so are the estimates over 0.5 given by W-MLE. In addition, similar to the results of Monte Carlo simulations, we see that the estimates given by S-MLE are generally larger than those by W-MLE.

4. The intensity of long-memory, denoted by the estimates of d given by S-MLE (with zero estimates removed) has little relationship with the watershed scale, as shown in Fig. 3. Only a very positive weak relationship can be established between the intensity of long-memory and the watershed scale, that is, the larger the watershed scale, the stronger the intensity of the long-memory.

5 Conclusions

Many studies have been carried out on detecting and modeling long-memory in various field since the seminal work of Hurst (1951). The Lo's modified rescaled adjusted range test (R/S test) (Lo, 1991), GPH test (Geweke and Porter-Hudak, 1983) and two approximate maximum likelihood estimation methods, i.e., Whittle's estimator (W-MLE) and another one implemented in S-Plus (S-MLE) based on the algorithm of Haslett and Raftery (1989) are evaluated through intensive Monte Carlo simulations for detecting the existence of long-memory. It is shown that it is difficult to find an appropriate lag q for Lo's test for different short-memory autoregressive (AR) and long-memory fractionally integrated autoregressive and moving average (ARFIMA) processes, which makes the use of Lo's test very tricky. In general, the GPH test outperforms the Lo's test, but for cases where there is a very strong autocorrelation such as AR(1) processes with $\phi =$

0.95 or even 0.99, the GPH test gets useless, even for time series of very large data size. On the other hand, the estimates of d given by S-MLE seem to be good indicators of the presence of long-memory. Data size has a significant impact on the power of all the four methods, which is in agreement with theoretical reasoning because the availability of larger samples allows one to better inspect the asymptotical properties. Generally, the power of Lo's test and GPH test increases with the increase of data size, and the estimates of d with GPH test and S-MLE converge with the increase of data size. If no large enough data set available, we should be aware of the possible bias of the estimates. For example, the estimate with S-MLE has a significant negative bias generally when the data size is less than 3000, and the estimates with W-MLE as well as GPH method may be negatively/positively biased when the autoregressive coefficient of the AR component in the ARFIMA process is small (e.g., ≤ 0.5)/big (e.g., 0.9). But S-MLE has a positive bias when the data size is above 10^4.

The four methods are applied to daily average discharge series recorded at 31 gauging stations with different drainage areas in eight river basins in Europe, Canada and USA to detect the existence of long-memory. The results show that, evidence of the existence of long-memory is found with all the four methods for about half of all the streamflow processes; except for one case where both the results of Lo's test and S-MLE indicating the absence of long-memory, and another case where the estimate of d given by S-MLE is zero and the estimate of d by W-MLE is close to zero (≈ 0.1), at least three methods applied here give evidence of the existence of long-memory in the other 29 daily streamflow processes; the intensity of long-memory in daily streamflow processes only has a very weak positive relationship with the scale of the watershed, that is, the larger the watershed scale, the stronger the intensity of the long-memory.

One limitation of the present study, especially for the analysis of the daily streamflow observations, is the restricted number of approaches considered here. The detection of long memory is affected by some uncertainty. It would be better to use several estimators so as to increase the reliability of the estimation. For instance, Koutsoyiannis (2002) recommended the use of the aggregated variance method (Beran, 1994), while in climate research many authors (see, e.g., Ribsky et al., 2006) used the detrended fluctuation analysis. These methods may not be very suitable for Monte Carlo simulations, but could make the detection of long-memory in observed time series more convincing if results of different methods are combined.

Edited by: D. Solomatine

References

Akaike, H.: Information theory and an extension of the maximum likelihood principle, in: Proceedings of the 2nd International Symposium on Information Theory, edited by: Petrov, B. N. and Csaki, F., Akademia Kiado, Budapest, 267–281, 1973.

Agiakloglou, C., Newbold, P., and Wohar, M.: Bias in an estimator of the fractional difference parameter, J. Time Ser. Anal., 14, 235–246, 1993.

Andrews, D. W. K.: Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, Econometrica, 59, 817–858, 1991.

Beran, J.: Statistics for Long-Memory Processes, Chapman & Hall, New York, 1994.

Bloomfield, P.: Trend in global temperature, Clim. Change, 21, 1–16, 1992.

Geweke, J. and Porter-Hudak, S.: The estimation and application of long memory time series models, J. Time Series Anal., 4, 221–238, 1983.

Giraitis, L., Robinson, P. M., and Samarov, A.: Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence, J. Time Ser. Anal., 18, 49–60, 1997.

Granger, C. W. J. and Joyeux, R.: An Introduction to Long-Memory Time Series Models and Fractional Differencing, J. Time Series Anal., 1, 15–29, 1980.

Haslett, J. and Raftery, A. E.: Space-time modeling with long-memory dependence: assessing Ireland's wind power resource (with discussion), Appl. Stat., 38, 1–50, 1989.

Henry, M.: Robust automatic bandwidth for long-memory, J. Time Ser. Anal., 22(3), 293–316, 2001.

Hosking, J. R. M.: Fractional Differencing, Biometrika, 68, 165–176, 1981.

Hurst, H. E.: Long-term storage capacity of reservoirs, Transactions of the American Society of Civil Engineers, 116, 770–808, 1951.

Hurvich C. M. and Deo, R. S.: Plug-in Selection of the Number of Frequencies in Regression Estimates of the Memory Parameter of a Long-memory Time Series, J. Time Ser. Anal., 20(3), 331–341, 1999.

Hurvich, C. M. and Bellrao, K.: Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series, J. Time Ser. Anal., 14, 455–472, 1993.

Hussain, S. and Elbergali, A.: Fractional order estimation and testing, application to Swedish temperature data, Environmetrics, 10, 339–349, 1999.

Karagiannis, T., Molle, M., and Faloutsos, M.: Long-range dependence: ten years of internet traffic modeling, IEEE Internet Computing, 8(5), 57–64, 2004.

Kendziorskia, C. M., Bassingthwaighteb, J. B., and Tonellato, P. J.: Evaluating maximum likelihood estimation methods to determine the Hurst coefficient, Physica A, 273, 439–451, 1999.

Klemes, V.: The Hurst phenomenon: A Puzzle?, Water Resour. Res., 10, 675–688, 1974.

Koutsoyiannis, D.: The Hurst phenomenon and fractional Gaussian noise made easy, Hydrol. Sci. J., 47(4), 573–595, 2002.

Lo, A. L.: Long-term memory in stock market prices, Econometrica, 59(5), 1279–1313, 1991.

Mandelbrot, B. B. and Wallis, J. R.: Computer experiments with fractional Gaussian noises. Part 1, 2 and 3, Water Resour. Res., 5, 228–267, 1969.

Meade, N. and Maier, M. R.: Evidence of long-memory in short-term interest rates, J. Forecast., 22, 553–568, 2003.

Montanari, A., Rosso, R., and Taqqu, M. S.: Fractionally differenced ARIMA models applied to hydrological time series: Identification, estimation, and simulation, Water Resour. Res., 33(5), 1035–1044, 1997.

Montanari, A., Rosso, R., and Taqqu, M. S.: A seasonal fractional ARIMA model applied to the Nile River monthly at Aswan, Water Resour. Res., 36, 1249–1259, 2000.

Ooms, M. and Franses, P. H.: A seasonal periodic long memory model for monthly river flows, Environ. Modell. Software, 16, 559–569, 2001.

Rao, A. R. and Bhattacharya, D.: Hypothesis testing for long-term memory in hydrological series, J. Hydrol., 216(3–4), 183–196, 1999.

Rybski, D., Bunde, A., Havlin, S., and von Storch, H.: Long-term persistence in climate and the detection problem, Geophys. Res. Lett., 33, L06718, doi:10.1029/2005GL025591, 2006.

Sowell, F.: Modeling Long-Run Behavior with the Fractional ARIMA Model, J. Monetary Economics, 29, 277–302, 1992.

Teverovsky, V., Taqqu, M. S., and Willinger, W.: A critical look at Lo's modified R/S statistic, Journal of Statistical Planning and Inference, 80, 211–227, 1999.

Wang, W., van Gelder, P. H. A. J. M., and Vrijling, J. K.: Long-memory properties of streamflow processes of the Yellow River. Proceedings of the International Conference on Water Economics, Statistics and Finance, 1, Rethymno-Crete, Greece, 481–490, 8–10 July, 2005,.

Wang, W., van Gelder, P. H. A. J. M., Vrijling, J. K., Ma, J.: Forecasting daily streamflow using hybrid ANN models, J. Hydrol., 324, 383–399, 2006.

Zivot, E. and Wang, J.: Modeling Financial Time Series with S-Plus. New York: Springer Verlag, 2003.