Hydrol. Earth Syst. Sci. Discuss., 3, 1603–1627, 2006
www.hydrol-earth-syst-sci-discuss.net/3/1603/2006/

Hydrology and Earth System Sciences Discussions

Papers published in *Hydrology and Earth System Sciences Discussions* are under open-access review for the journal *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, The Netherlands

Received: 15 May 2006 – Accepted: 12 June 2006 – Published: 14 July 2006

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

Abstract

The Lo's R/S tests (Lo, 1991), GPH test (Geweke and Porter-Hudak, 1983) and the maximum likelihood estimation method implemented in S-Plus (S-MLE) are evaluated through intensive Mote 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 AR and 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 strong autocorrelations (e.g., AR(1) processes with ϕ=0.97 or even 0.99), the GPH test is totally useless, even for time series of large data size. Although S-MLE method does not provide a statistic test for the existence of long-memory, the estimates of d given by S-MLE seems to give a good indication of whether or not the long-memory is present. Data size has a significant impact on the power of all the three 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 test and S-MLE converge with the increase of data size.

According to the results with the Lo's R/S test (Lo, 1991), GPH test (Geweke and Porter-Hudak, 1983) and the S-MLE method, all daily flow series exhibit long-memory. The intensity of long-memory in daily streamflow processes has only a very weak positive relationship with the scale of watershed.

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 are not summable. For a stationary discrete long-memory time series process, its autocorrelation function $\rho(k)$ at lag k

satisfies (Hosking, 1981)

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

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

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 fractional integrated autoregressive and moving average model, denoted by ARFIMA (p, d, q). 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 behaviour.

In the hydrology community, many studies have been carried out on the test for long-memory in hydrological processes. Montanari et al. (1997) applied fractionally integrated autoregressive moving average (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 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 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, three methods used in the present study to detect long-memory will be briefly described. Simulation results with the three methods are presented in Sect. 3. Then, the three methods will be applied to 31 daily streamflow series to detect the existence of long-memory in Sect. 4, and some discussions are given in Sect. 5. Finally, some conclusions are drawn in Sect. 6.

2 Methods of detecting the existence of long-memory

Many methods are available for detecting for 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 R/S analysis; aggregated variance method etc.), parametric methods (e.g., Whittle maximum likelihood estimation method) and semiparametric 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 or H, but they are generally not accurate and are sensitive to short range serial correlations. The parametric methods obtain consistent estimators of d or H via maximum likelihood estimation of parametric long-memory models. They give more accurate estimate of d or H, but generally require knowledge of the true model which is in fact always unknown. Semiparametric 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 GPH test which is a semiparametric method will be used to test for the null hypothesis of no presence of long-memory. Besides, an approximate maximum likelihood estimation method is 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, ..., N, with the j-th partial sum $Y_j=\sum_{i=1}^{j} x_i$, j=1, 2, ..., N, and the sample variance $S_j^2=j^{-1}\sum_{i=1}^{j}(x_i-j^{-1}Y_j)^2$, j=1, 2, ..., N, the rescaled adjusted range statistic or R/S-statistic is defined by

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

The classical R/S analysis is sensitive to the presence of explicit short-range dependence structures, and lacks of 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. (2), i.e., the sample standard deviation, by a modified standard deviation S_q which takes into account the autocovariances of the first q lags, so as 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. (1), only focus on lag j=N. The S_q is defined as

$$S_q=\left(\frac{1}{N}\sum_{j=1}^{N}(x_j-\bar{x}_N)^2+\frac{2}{N}\sum_{j=1}^{q}\varpi_j(q)\left[\sum_{i=j+1}^{N}(x_i-\bar{x}_N)(x_{i-j}-\bar{x}_N)\right]\right)^{1/2} \tag{3}$$

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

$$Q_{N,q}=\frac{1}{S_q}\left\{\max_{0\le i\le N}\sum_{j=1}^{i}(x_j-\bar{x}_N)-\min_{0\le i\le N}\sum_{j=1}^{i}(x_j-\bar{x}_N)\right\} \tag{4}$$

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

$$P(W\le x)=1+2\sum_{j=1}^{\infty}(1-4x^2j^2)e^{-2x^2j^2}, \tag{5}$$

Lo gave the critical values of x for hypothesis testing at sixteen significance levels using Eq. (4), 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 semi-parametric approach to testing for long-memory. Given a fractionally integrated process $\{x_t\}$, its spectral density is given by:

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

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

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

which may be written in the alternative form

$$\ln f(\varpi_j) = \ln f_u(0) - d \ln\left[4\sin^2(\varpi_j/2)\right] + \ln\left[f_u(\varpi_j)/f_u(0)\right] \quad (6)$$

The fractional difference parameter d can be estimated by the regression equations constructed from Eq. (5). Geweke and Porter-Hudak (1983) showed that using a periodogram estimate of $f(\varpi_j)$, if the number of frequencies used in the regression Eq. (6) 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 $U_j=\ln[4\sin^2(\varpi_j/2)]$ and $\overline{U}$ is the sample mean of U_j, $j=1, \cdots, g(n)$. 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 of fractional differencing parameter d

Suppose that the observation $X=(x_1, \ldots, x_n)^t$ is an ARFIMA (p, d, q) process defined by

$$\phi(B)(1-B)^d(x_t-\mu) = \theta(B)\varepsilon_t \quad (7)$$

where B is the backshift operator, that is, $Bx_t=x_{t-1}$; $\phi(B)=1-\phi_1 B-\ldots-\phi_p B^P$ and $\theta(B)=1-\theta_1 B-\ldots-\theta_q B^q$ represent the ordinary autoregressive and moving average components; ε_t is a white noise process with zero mean and variance σ^2.

The Gaussian log-likelihood of X for the process (7) is given by

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

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, $|\Sigma|$ denote 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.

In this study, 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). If the estimated d is significantly greater than zero, we consider it an evidence of the presence of long-memory.

3 Monte Carlo simulations

We perform an extensive Monte Carlo investigation in order to find out how reliable the Lo's test, the GPH test and the S-MLE are with 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., ϕ=0.5, 0.8, 0.9, 0.95, 0.99 are considered, because these are the cases commonly seen in streamflow processes. For the ARFIMA models, ϕ=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 by the *arima.sim* and *arima.fracdiff.sim* function built in S-Plus version 6.

For Lo's modified R/S test, the right choice of q in Lo's method 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 is the null hypothesis 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 [•] 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:

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

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, for the GPH test, there is a choice of the number of frequencies $g(n)$ used in the regression Eq. (6). This choice entails a bias-variance tradeoff. For a given sample size, as $g(n)$ is increased from 1, the variance of the d esti-

mate decreases, but this decrease is typically offset by the increase in bias due to non-constancy of $f_u(\varpi)$. 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, we assume that the order p of the AR component for each simulated ARFIMA process is unknown before hand. Instead, we estimate the order p of AR component by using the AIC criterion (Akaike, 1973).

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 and the S-MLE estimates of d 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 the S-MLE method, the averages and standard deviations of the estimates of d ("average d" and "SD of d") are reported. According to the results with 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, and the GPH test has a better performance. But when the autoregressive coefficient is higher than 0.9, the probability of committing Type I error with the GPH test increase very fast, and the GPH test gets useless for the cases when ϕ is above 0.97 (for saving 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 downwardly biased estimates of d when an AR term of low autoregressive coefficient value (e.g., $\phi \leq 0.5$) is present, whereas yields upwardly biased estimates of d when an AR term 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 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.

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 so as to avoid the high probability of wrongly rejecting the null hypothesis of no long-memory for AR processes, the lag seems to be not 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 value of ϕ in their AR components, the lag chosen by Eq. (9) seems too short for AR series of big size, but not large enough for ARFIMA processes. Namely, no good tradeoff can be achieved in choosing an appropriate

lag value for the Lo's test. This result further substantiate the limitation of the use of the Lo's test, which has been shown in the previous study of Teverovsky et al. (1999).

4. Although S-MLE method does not provide a statistic test for the existence of long-memory, the estimates of d seems to give a good indication of whether or not the long-memory is present. It is shown by our simulation study that:

   a) For AR(1) processes, S-MLE gives basically correct estimates of d, i.e., d=0, even when the autoregressive coefficients are very high, although the estimates are slightly positively biased when the data size is small (e.g., 500 points). The estimates get more accurate (according to the averages) and more stable (according to the standard deviations) with the increase of sample size.

   b) For ARFIMA processes, S-MLE provides significantly downwardly biased estimates when the data size is small (e.g., less than 10^3). The values of 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.

5. Data size has a significant impact on the power of all the three 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 test and S-MLE converge with the increase of data size. 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, the power of GPH test increases with the increase of sample size. If we use a sample size of larger than 10^4 points, GPH test still has very good performance for AR(1) processes with ϕ=0.9. But the use of GPH test is helpless 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. (6) 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 the above findings, to obtain reliable test results on detecting the presence of long-memory, we have two suggestions: Firstly, as we see that the power of Lo's test and GPH test increases with the increase of data size, and the estimates of d with the GPH-test and S-MLE converge as the sample size increase, therefore, use test data of enough data size (e.g., 3000~10^4) when detecting the existence of long-memory. Secondly, the estimate of d given by the S-MLE is recommended to be used as and indicator of the intensity of long-memory, but notice that the estimate with S-MLE would be biased downwardly significantly when the data size is less than 3000 and biased upwardly when the data size is above 10^4. Therefore, the most appropriate date size for estimating d with S-MLE is slightly less than 10^4. Thirdly, 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 and S-MLE. As shown in Table 1, the combined use of these three methods produces the following alternatives:

a) 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, provide evidence in favour of no existence of long-memory

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

c) In other cases, the data are not sufficiently informative with respect to the long-memory properties of the series. But if both the GPH test and S-MLE give positive result 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), U.S. 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;
2. The drainage area of each station is basically within 5 different watershed scales, namely, >10^6 km^2; 10^6~10^5 km^2; 10^5~10^4 km^2; 10^4~10^3 km^2; <10^3 km^2;
3. The stations are located in the main river channel of the river if possible. 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 limitation, the shortest series covers a period of only 14 years. The segments are chosen with following criteria:

1. The series should be approximately stationary, as least by visual inspection. We have stationarity as our primary data criterion because, when certain types of nonstationarity are present, many longmemory parameter estimators may fail (Klemes, 1974).

2. The data should be recorded as early as possible, so as to limit the influence of human intervention to the minimum.

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 module of statistical analysis package S-plus (Zivot and Wang, 2003). To alleviate the impact of seasonality, all the series are deseasonalized by subtracting the daily means and dividing by the daily standard deviations.

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 results of detecting long-memory in daily streamflow processes are reported in Table 3, which show the following:

1. The GPH estimates and the S-MLE estimates are in good agreement, as shown in Fig. 1, except for four series for which the estimates of d given by S-MLE are zero. According to such a consistency, we believe that the estimates of zero given by S-MLE probably are resulted from its erroneousness. Therefore, when using

S-MLE method to estimate the fractional differencing parameter d, caution must be taken if an estimate of zero is given by S-MLE. At the same time, we notice that the estimates given by GPH test are generally slightly larger than those given by the S-MLE method for cases where the estimates of d are greater than 0.4.

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. In consequence, they recommended that one always relies on a wide range of different q-values, and does not use Lo's method in isolation, instead, uses it always 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, the results given by Lo's test are not in agreement with those of the GPH test and S-MLE method 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 of the Rhine River at Lobith is a short-memory process, whereas both the GPH test and the S-MLE method indicate that the streamflow process of the Rhine River at Lobith exhibits long-memory.

3. 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, whereas the estimates of S-MLE show that 4 out of all the series have d's of zero value. But the results of Lo's test and S-MLE are not in agreement (except for one case of Color-4), namely, those series with zero estimated d given by S-MLE seem to exhibit significant long-memory according to Lo's test (with either data-driven lag or fixed lag). On the other hand, GPH test

tells us that all the series exhibit long-memory. Therefore, for each series, at least two methods applied here give evidences of the existence of long-memory in all the daily streamflow processes (except for one case of Color-4). Because of the unreliability of the Lo's test, and possible erroneousness of the S-MLE estimates, we conclude that all the streamflow series have long-memory.

4. The intensity of long-memory, denoted by the estimates of d given by S-MLE (the zeros are removed), has little relationship with the watershed scale, as shown in Fig. 2. Only a very weak positive 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

The Lo's R/S tests (Lo, 1991), GPH test (Geweke and Porter-Hudak, 1983) and the maximum likelihood estimation method implemented in S-Plus (S-MLE) are evaluated through intensive Mote 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 AR and 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 strong autocorrelations (e.g., AR(1) processes with ϕ=0.97 or even 0.99), the GPH test is totally useless, even for time series of large data size. Although S-MLE method does not provide a statistic test for the existence of long-memory, the estimates of d given by S-MLE seems to give a good indication of whether or not the long-memory is present. Data size has a significant impact on the power of all the three 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 test and S-MLE converge with the increase of data size.

Although the S-MLE method perform very well for simulated AR(1) and ARFIMA(1, d, 0) processes, when applying the S-MLE method to the observed stream-

flow series, it is found that it may wrongly give zero estimates of d for several series even for large data size (e.g., around 10^4). Therefore, caution must be taken in case the S-MLE gives a zero estimate of d.

According to results with the Lo's R/S tests (Lo, 1991), GPH test (Geweke and Porter-Hudak, 1983) and the S-MLE method, all daily flow series exhibit long-memory. The intensity of long-memory in daily streamflow processes has only 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.

Acknowledgements. The financial support from the Program for New Century Excellent Talents in University, China (NCET-04-0492) is gratefully acknowledged.

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, Climatic Change, 21, 1–16, 1992.

Geweke, J. and Porter-Hudak, S.: The estimation and application of long memory time series models, J. Time Ser. 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 Ser. 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, Trans. Amer. Soc. Civil Eng., 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.

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, Environmental Modelling & 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.

Sowell, F.: Modeling Long-Run Behavior with the Fractional ARIMA Model, Journal of Monetary

Economics, 29, 277–302, 1992.

Teverovsky, V., Taqqu, M. S., and Willinger, W.: A critical look at Lo's modified R/S statistic, J. Stat. 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, vol. 1, Rethymno-Crete, Greece, 8–10 July 2005, 481–490, 2005.

Zivot, E. and Wang, J.: Modelling Financial Time Series with S-Plus, New York: Springer Verlag, 2002.

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

Model	Data size	Lo's R/S test			GPH test			S-MLE	
		average lag	SD of lag	accepted	average d	SD of d	accepted	average d	SD of d
AR(1)	500	2.8	0.5	464	−0.0167	0.1302	495	0.0149	0.0350
ar=.5	1000	3.2	0.4	454	−0.0123	0.1141	490	0.0189	0.0325
	3000	4.6	0.5	468	−0.0124	0.0772	490	0.0136	0.0220
	10 000	6.1	0.2	455	−0.0119	0.0607	490	0.0093	0.0132
	20 000	7.8	0.4	469	−0.0078	0.0479	488	0.0057	0.0100
AR(1)	500	6.7	0.8	428	0.1220	0.1388	470	0.0269	0.0669
ar=.8	1000	8.0	0.7	442	0.0637	0.1110	489	0.0209	0.0419
	3000	10.8	0.5	441	0.0163	0.0827	490	0.0199	0.0322
	10 000	14.7	0.5	441	−0.0016	0.0605	490	0.0114	0.0207
	20 000	17.6	0.5	454	−0.0036	0.0511	483	0.0079	0.0149
AR(1)	500	11.3	1.6	431	0.3252	0.1342	268	0.0290	0.0566
ar=.9	1000	13.5	1.4	408	0.2189	0.1135	326	0.0296	0.0632
	3000	18.1	1.1	414	0.0957	0.0851	436	0.0240	0.0488
	10 000	24.6	0.8	441	0.0273	0.0600	483	0.0132	0.0236
	20 000	29.4	0.7	457	0.0107	0.0500	489	0.0081	0.0150
AR(1)	500	18.7	3.6	451	0.5739	0.1395	24	0.0302	0.0497
ar=.95	1000	22.4	3.1	429	0.4488	0.1154	34	0.0390	0.0801
	3000	29.6	2.4	426	0.2594	0.0800	91	0.0270	0.0535
	10 000	40.3	1.8	416	0.1201	0.0601	300	0.0117	0.0284
	20 000	47.9	1.6	416	0.0665	0.0475	409	0.0065	0.0160
AR(1)	500	52.9	20.3	494	0.9122	0.1617	0	0.0482	0.0674
ar=.99	1000	65.3	19.3	484	0.8530	0.1226	0	0.0431	0.0780
	3000	86.8	14.7	399	0.7297	0.0826	0	0.0231	0.0442
	10 000	119.7	11.9	389	0.5555	0.0583	0	0.0093	0.0211
	20 000	142.4	9.5	380	0.4478	0.0477	0	0.0068	0.0148
ARFIMA	500	2.2	0.5	129	0.2587	0.1360	353	0.2144	0.1100
d=0.3	1000	2.8	0.5	61	0.2749	0.1157	228	0.2571	0.0829
	3000	3.8	0.5	15	0.2821	0.0826	68	0.2786	0.0646
	10 000	5.2	0.4	0	0.2884	0.0572	2	0.3043	0.0201
	20 000	6.3	0.5	0	0.2900	0.0470	0	0.3072	0.0162
ARFIMA	500	7.1	1.4	255	0.2729	0.1402	333	0.1728	0.1346
ar=0.5	1000	8.6	1.3	139	0.2783	0.1130	233	0.2126	0.1165
d=0.3	3000	11.4	1.2	63	0.2878	0.0919	83	0.2849	0.0675
	10 000	15.6	1.0	8	0.2934	0.0604	4	0.3049	0.0363
	20 000	18.6	0.9	5	0.2955	0.0493	0	0.3102	0.0202
ARFIMA	500	41.1	12.2	493	0.6375	0.1513	16	0.1683	0.1451
ar=0.9	1000	49.4	11.6	478	0.5213	0.1123	6	0.2035	0.1333
d=0.3	3000	65.4	11.2	345	0.3964	0.0881	5	0.2397	0.1243
	10 000	89.4	9.2	155	0.3316	0.0627	2	0.3103	0.0678
	20 000	106.6	8.3	78	0.3145	0.0512	0	0.3281	0.0501
ARFIMA	500	7.0	4.0	130	0.4077	0.1506	157	0.3092	0.1572
d=0.45	1000	8.5	4.4	56	0.4274	0.1237	53	0.3616	0.1309
	3000	11.2	5.2	11	0.4371	0.0873	0	0.4238	0.0620
	10 000	15.4	6.0	0	0.4373	0.0613	0	0.4589	0.0173
	20 000	18.6	7.0	0	0.4371	0.0489	0	0.4676	0.0164
ARFIMA	500	19.1	10.1	346	0.4331	0.1515	133	0.2355	0.1628
ar=0.5	1000	22.9	10.6	204	0.4385	0.1164	33	0.3328	0.1311
d=0.45	3000	31.0	12.2	66	0.4404	0.0893	3	0.4226	0.0668
	10 000	42.4	14.6	11	0.4429	0.0635	0	0.4608	0.0228
	20 000	50.2	16.2	4	0.4459	0.0507	0	0.4718	0.0170
ARFIMA	500	135.0	78.5	493	0.7956	0.1394	2	0.1306	0.1757
ar=0.9	1000	163.4	90.2	495	0.6733	0.1172	1	0.1712	0.1828
d=0.45	3000	222.9	116.2	472	0.5539	0.0878	0	0.3128	0.1665
	10 000	299.5	138.7	273	0.4856	0.0599	0	0.4464	0.0577
	20 000	361.8	158.0	140	0.4666	0.0491	0	0.4748	0.0226

Note: The results of the Lo's R/S test and the GPH test are based on 500 replications. The results of the S-MLE estimate of *d* are based on 100 replications.

Table 2. Description of selected daily streamflow time series.

No.	Basin	Location of gauging stations	Area (km^2)	Latitude	Longitude	Elevation (m)	Period	Avg. discharge (cms)
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	2 962 974	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	5659	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

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
			Lag-1	Stat-1	Lag-2	Stat-2	*d*	Stat	*d*
Color-1	10 957	0.9738	64	2.9566	50	3.2475	0.5125	7.5412	0.4478
Color-2	10 958	0.9627	50	3.4320	50	3.4320	0.4906	7.2192	0.4506
Color-3	5113	0.9431	31	2.1437	50	1.8067	0.4766	5.6613	0.4863
Color-4	7305	0.9549	40	1.1811	50	1.0826	0.4043	5.3169	0.0000
Colum-1	10 957	0.9910	132	1.5357	50	2.1519	0.5071	7.4617	0.4615
Colum-2	8401	0.9966	238	1.1342	50	1.8357	0.4673	6.3838	0.4187
Colum-3	10 957	0.9778	72	3.1202	50	3.5159	0.3466	5.101	0.4392
Colum-4	10 957	0.9676	55	1.8590	50	1.9213	0.3642	5.36	0.4213
Danu-1	10 957	0.9931	158	1.5328	50	2.0899	0.3441	5.0639	0.2634
Danu-2	10 957	0.9577	46	1.9412	50	1.8957	0.3017	4.4398	0.3598
Danu-3	10 958	0.9326	33	3.1827	50	2.7771	0.3782	5.5651	0.4059
Fras-1	10 957	0.9772	70	1.5279	50	1.6994	0.3879	5.7077	0.3878
Fras-2	10 958	0.9734	63	2.9821	50	3.1849	0.2511	3.6952	0.3529
Fras-3	10 958	0.9582	47	2.3767	50	2.3411	0.2272	3.343	0.1886
Fras-4	8401	0.9294	30	2.2163	50	1.9096	0.2769	3.7833	0.3100
Missi-1	10 958	0.9961	232	1.8789	50	3.0163	0.4133	6.0813	0.3909
Missi-2	10 956	0.9921	144	2.6780	50	3.7589	0.3846	5.6601	0.4001
Missi-3	10 958	0.9917	139	1.8277	50	2.6476	0.5098	7.5018	0.4847
Missi-4	10 958	0.9563	45	2.7527	50	2.6345	0.5358	7.8847	0.0000
Misso-1	10 958	0.9711	60	3.6930	50	3.9396	0.4484	6.5985	0.4238
Misso-2	9131	0.9805	75	3.6145	50	4.1707	0.4639	6.4915	0.4124
Misso-3	10 958	0.9165	29	5.1261	50	4.1325	0.4179	6.1498	0.0000
Misso-4	10 958	0.9522	42	3.2612	50	3.0869	0.2450	3.605	0.0000
Ohio-1	10 958	0.9723	62	1.7652	50	1.8735	0.2910	4.2822	0.2983
Ohio-2	10 958	0.9547	44	2.1173	50	2.0477	0.2569	3.781	0.2581
Ohio-3	10 958	0.9291	32	1.7894	50	1.6164	0.3289	4.8401	0.2263
Ohio-4	10 958	0.8985	25	1.9601	50	1.5937	0.3659	5.3839	0.3324
Rhine-1	10 957	0.9897	120	1.2813	50	1.6822	0.3787	5.5729	0.4254
Rhine-2	10 958	0.9715	61	2.0457	50	2.1880	0.3513	5.1699	0.0000
Rhine-3	10 958	0.9048	26	2.1554	50	1.7478	0.3792	5.5799	0.4176
Rhine-4	10 958	0.8739	21	2.2409	50	1.7306	0.2489	3.6627	0.3447

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.

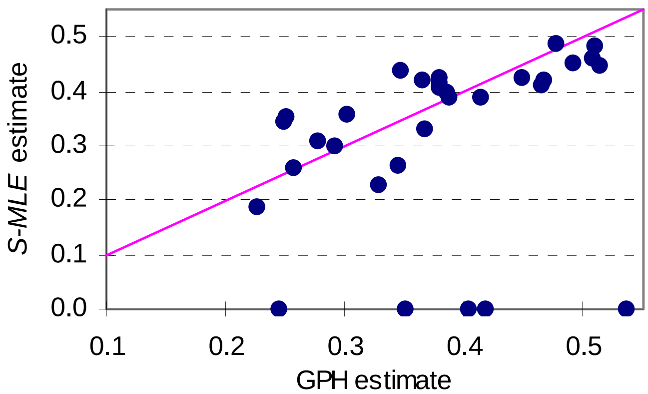

Fig. 1. The GPH estimate versus the S-MLE estimate of *d* (note: the straight line has a slope of 0.5).

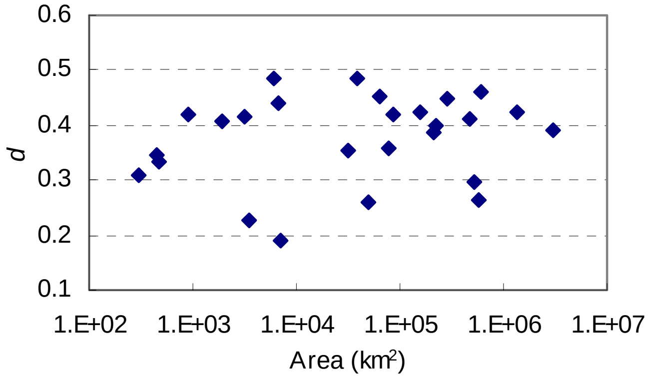

Fig. 2. Estimated *d* versus watershed scale for streamflow processes.