STATISTICAL ESTIMATION METHODS FOR EXTREME HYDROLOGICAL EVENTS

P.H.A.J.M. VAN GELDER*, W. WANG, AND
J. K. VRIJLING
TU Delft, the Netherlands

Abstract- In this paper an overview is given of the statistical methods which are needed to analyse observed environmetric data with a particular interest for the extreme values. The methods for trend analysis, stationarity tests, seasonality analysis, long-memory studies will be presented, critically reviewed, applied to some existing datasets, and compared.

Keywords: Trend analysis, stationarity tests, seasonality analysis, and long-memory studies.

1. Introduction

In designing civil engineering structures use is made of probabilistic calculation methods. Stress and load parameters are described by statistical distribution functions. The parameters of these distribution functions can be estimated by various methods. The main point of interest is the behaviour of each method for predicting p-quantiles (the value which is exceeded by the random variable with probability p), where $p \ll 1$. The estimation of extreme quantiles corresponding to a small probability of exceedance is commonly required in the risk analysis of hydraulic structures. Such extreme quantiles may represent design values of environmental loads (wind, waves, snow, earthquake), river discharges, and flood levels specified by design codes and regulations.

In civil engineering practice many parameter estimation methods for probability distribution functions are in circulation. Well known methods are for example:

- the method of moments,
- the method of maximum likelihood,

O.F. Vasiliev et al. (eds.), Extreme Hydrological Events: New Concepts for Security, 199–252.

- the method of least squares (on the original or on the linearized data),
- the method of Bayesian estimation,
- the method of minimum cross entropy,
- the method of probability weighted moments,
- the method of L-moments.

These methods have been judged on their performance and critically reviewed in for instance, Van Gelder, 1999. It has been investigated which estimation method is preferable for the parameter estimation of a particular probability distribution in order to obtain a reliable estimate of the p-quantiles. Particularly attention was paid to the performance of the parameter estimation method with respect to three different criteria; (i) based on the relative bias and (ii) root mean squared error (RMSE), (iii) based on the over- and underdesign.

It is desirable that the quantile estimate be unbiased, that is, its expected value should be equal to the true value. It is also desirable that an unbiased estimate be efficient, i.e., its variance should be as small as possible. The problem of unbiased and efficient estimation of extreme quantiles from small samples is commonly encountered in the civil engineering practice. For example, annual flood discharge data may be available for past 50 to 100 years and on that basis one may have to estimate a design flood level corresponding to a 1,000 to 10,000 years return period.

This paper will concentrate on the steps before fitting an analytical probability distribution to represent adequately the sample observations. These steps involve trend analysis, stationarity tests, seasonality analysis, and long-memory studies. After those steps, the distribution type can be judged from the data and parameters of the selected distribution type can be estimated. Since the bias and efficiency of quantile estimates are sensitive to the distribution type, the development of simple and robust criteria for fitting a representative distribution to small samples of observations has been an active area of research. Van Gelder (1999) gives an overview of such considerations. This paper will start with the issues on trend analysis in Section 2, followed by stationarity tests in Section 3. Most environmetric data show seasonality behaviour. Methods to take this into account are discussed in Section 4. The last part of the paper (Section 5) is devoted to long memory studies of environmetric data. The paper ends with a summary and list of references.

2. Trend Analysis

Many hydrological time series exhibit trending behavior or nonstationarity. In fact, the trending behavior is a type of nonstationarity. But in this present study, they are treated separately. The purpose of a trend test is to determine if the values of a series have a general increase or decrease with the time increase, whereas the purpose of stationarity test is to determine if the distribution of a series is dependent on the time.

An important task in hydrological modeling is to determine if there is the existence of any trend in the data and how to achieve stationarity when the data is nonstationary. On the other hand, the possible effects of global warming on water resources have been the topic of many recent studies (e.g., Lettenmaier et al., 1999; Jain and Lall, 2001; Kundzewicz et al., 2004). Thus, detecting the trend and stationarity in a hydrological time series may help us to understand the possible links between hydrological processes and changes in the global environment. The focus of the trend analysis and stationarity test in this study is not to detect the changes of regional or world-wide streamflow processes. As a matter of fact, the presence of trends and nonstationarity is undesirable in further analysis. Therefore, we should make sure whether there is the presence of trend and nonstationarity or not, and if the presence of trend and nonstationarity is detected, the appropriate pre-processing procedure should be applied. In this section the issue of trend analysis is studied, and the nonstationarity problem will be addressed in the following section.

Non-parametric trend detection methods are less sensitive to outliers (extremes) than are parametric statistics such as Pearson's correlation coefficient. In addition, nonparametric test can test for a trend in a time series without specifying whether the trend is linear or nonlinear. Therefore, A rank-based nonparametric method, the Mann-Kendall's test (Kendall, 1938; Mann, 1945), is applied in this study to annual and monthly series.

2.1. TREND TEST FOR ANNUAL STREAMFLOW SERIES

First of all, we test for the trend in annual series so as to get an overall view of the possible changes in streamflow processes.

2.1.1. *Mann-Kendall test*

Kendall (1938) proposed a measure *tau* to measure the strength of the monotonic relationship between x and y. Mann (1945) suggested using the test for the significance of Kendall's *tau* where one of the variables is time as a test for trend. The test is well known as the Mann-Kendall's test (referred to as MK test hereafter), which is powerful for uncovering deterministic trends. Under the null hypothesis H_0, that a series $\{x_1, \ldots, x_N\}$ come from a population where the random variables are independent and identically distributed, the MK test statistic is

$$S = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \operatorname{sgn}(x_j - x_i) \text{ ,where } \operatorname{sgn}(x) = \begin{cases} +1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases} \quad (1)$$

and *tau* is estimated as:

$$\tau = \frac{2S}{N(N-1)}. \quad (2)$$

The general principles of hypothesis testing are explained in Appendix 1.

Kendall (1975) showed that the variance of S, $Var(S)$, for the situation where there may be ties (i.e., equal values) in the x values, is given by

$$\sigma_S^2 = \frac{1}{18}\left[N(N-1)(2N+5) - \sum_{i=1}^{m} t_i(t_i-1)(2t_i+5) \right], \quad (3)$$

where m is the number of tied groups in the data set and t_i is the number of data points in the ith tied group.

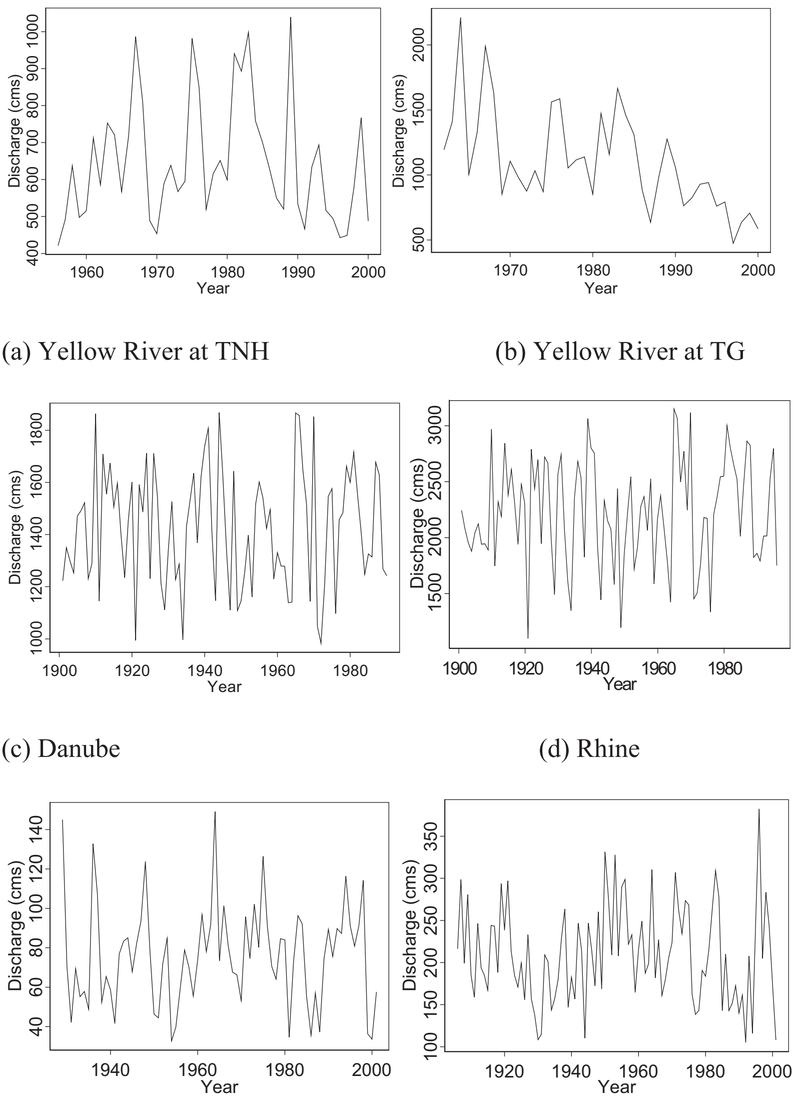
Under the null hypothesis, the quantity z defined in the following equation is approximately standard normally distributed even for the sample size $N = 10$:

$$z = \begin{cases} (S-1)/\sigma_S & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ (S+1)/\sigma_S & \text{if } S < 0 \end{cases}. \quad (4)$$

It has been found that the positive serial correlation inflates the variance of the MK statistic S and hence increases the possibility of rejecting the null hypothesis of no trend (von Storch, 1995). In order to reduce the impact of serial correlations, it is common to prewhiten the time series by removing the serial correlation from the series through $y_t = x_t - \phi x_{t-1}$, where y_t is the prewhitened series value, x_t is the original time series value, and ϕ is the estimated lag 1 serial correlation coefficient. The pre-whitening approach has been adopted in many trend-detection studies (e.g., Douglas et al., 2000; Zhang et al., 2001; Burn and Hag Elnur, 2002).

2.1.2. *MK test results*

The first step in time series analysis is visually inspecting the data. Significant changes in level or slope usually are obvious. The annual average streamflow series of the Yellow River at TNH and TG, the Rhine River at Lobith, the Umpqua River near Elkton and the Ocmulgee River at Macon are shown:



(a) Yellow River at TNH

(b) Yellow River at TG

(c) Danube

(d) Rhine

(e) Ocmulgee

(f) Umpqua

Figure 1. Annual average discharge series of the five rivers at six sites

From the visual inspection, it seems that except for the annual flow series of the Yellow River at TG which exhibits obvious downward trend, other annual series have no obvious trend. The MK test results are displayed in Table 1. The results are in agreement with the heuristic result by the visual examination.

TABLE 1. Mann-Kendall tests on Annual average discharge series

Streamflow	*Tau*	z statistic	p-value
TNH	-0.1015	-0.9609	0.3366
TG	-0.3144	-2.7658	0.0057
Danube	0	0	1
Rhine	0.0467	0.6710	0.5022
Ocmulgee	0.1025	1.2688	0.2045
Umpqua	-0.0258	-0.3665	0.7140
Null hypothesis: *tau* = 0			

2.2. TREND TEST FOR MONTHLY STREAMFLOW SERIES

The trend test for annual series gives us an overall view of the change in streamflow processes. To examine the possible changes occur in smaller timescale, we need to investigate the monthly flow series. Monthly streamflows usually exhibit strong seasonality. Trend test techniques for dealing with seasonality of univariate time series fall into three major categories (Helsel and Hirsh, 1992, pp 337-343): (1) fully nonparametric method, i.e., seasonal Kendall test; (2) mixed procedure, i.e., regression of deseasonalized series on time; (3) parametric method, i.e., regression of original series on time and seasonal terms. The first approach, namely, seasonal Kendall test will be used here.

2.2.1. *Seasonal Kendall test*

Hirsch et al. (1982) introduced a modification of the MK test, referred to as the seasonal Kendall test that allows for seasonality in observations collected over time by computing the Mann-kendall test on each of p seasons separately, and then combining the results. Compute the following overall statistic S':

$$S' = \sum_{j=1}^{p} S_j \, , \qquad (5)$$

where S_j is simply the S-statistic in the MK test for season j (j = 1, 2, ..., p) (see Eq. 1). When no serial dependence exhibit in the time series, the variance of S' is defined as

$$\sigma_{S'}^2 = \sum_{j=1}^{p} Var(S_j) \, . \qquad (6)$$

When serial correlation is present, as in the case of monthly streamflow processes, the variance of *S'* is defined as (Hirsch and Slack, 1984)

$$\sigma_{S'}^2 = \sum_{j=1}^{p} Var(S_j) + \sum_{g=1}^{p-1} \sum_{h=g+1}^{p} \sigma_{gh} , \tag{7}$$

where σ_{gh} denotes the covariance between the MK statistic for season *g* and the MK statistic for season *h*. The covariance is estimated with the following procedures.

Let the matrix

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \tag{8}$$

denote a sequence of observations taken over *p* seasons for *n* years. Let the matrix

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1p} \\ R_{21} & R_{22} & \cdots & R_{2p} \\ \vdots & \vdots & & \vdots \\ R_{n1} & R_{n2} & \cdots & R_{np} \end{bmatrix} \tag{9}$$

denote the ranks corresponding to the observations in *X*, where the *n* observations for each season are ranked among themselves, that is,

$$R_{ij} = \frac{1}{2}\left[n + 1 + \sum_{k=1}^{n} \text{sgn}(x_{ij} - x_{kj}) \right]. \tag{10}$$

Hirsch and Slack (1984) suggest using the following formula, given by Dietz and Killeen (1981), to estimate σ_{gh} in the case where there are no missing values:

$$\hat{\sigma}_{gh} = \frac{1}{3}\left[K_{gh} + 4\sum_{i=1}^{n} R_{ig} R_{ih} - n(n+1)^2 \right], \tag{11}$$

where

$$K_{gh} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sgn}\left[(X_{jg} - X_{ig})(X_{jh} - X_{ih}) \right]. \tag{12}$$

If there are missing values,

$$R_{ij} = \frac{1}{2}\left[n_j + 1 + \sum_{k=1}^{n_j} \mathrm{sgn}(X_{ij} - X_{kj}) \right], \tag{13}$$

where n_j denotes the number of observations without missing values for season j. And the covariance between the MK statistic for season g and season h is estimated as

$$\hat{\sigma}_{gh} = \frac{1}{3}\left[K_{gh} + 4\sum_{i=1}^{n} R_{ig} R_{ih} - n(n_g + 1)(n_h + 1) \right]. \tag{14}$$

Then the quantity z' defined in the following equation is approximately standard normally distributed:

$$z' = \begin{cases} (S'-1)/\sigma_{S'} & \text{if } S' > 0 \\ 0 & \text{if } S' = 0 \\ (S'+1)/\sigma_{S'} & \text{if } S' < 0 \end{cases}. \tag{15}$$

The overall *tau* is the weighted average of the *p* seasonal τ's, defined as

$$\tau = \sum_{j=1}^{p} n_j \tau_j \bigg/ \sum_{j=1}^{p} n_j , \tag{16}$$

where τ_j is the *tau* for season *j*, estimated with Eq. 2.

Seasonal Kendall test is appropriate for testing for trend in each season when the trend is always in the same direction across all seasons. However, the trend may have different directions in different seasons. Van Belle and Hughes (1984) suggested using the following statistic to test for heterogeneity in trend

$$\chi^2_{het} = \sum_{j=1}^{p} z_j^2 - p\bar{z}^2 , \tag{17}$$

where z_j denotes the z-statistic for the jth season computed as

$$z_j = \frac{S_j}{(Var(S_j))^{1/2}} , \tag{18}$$

and

$$\bar{z} = \frac{1}{p}\sum_{j=1}^{p} z_j . \tag{19}$$

Under the null hypothesis of no trend in any season, the statistic defined in Eq. 17 is approximately distributed as a chi-square random variable with p - 1 degrees of freedom.

2.2.2. *Seasonal Kendall test results*

The six monthly streamflow processes are tested for the trend with the seasonal Kendall test which allows fort the serial dependence. And the heterogeneity in trend is also tested. The results are shown in Table 2. The results give the same conclusion as the test for annual series, that is, among 5 series, only the streamflow of the Yellow River at TG exhibits significant downward trend. Meanwhile, it is found that while the streamflow processes at TG present downward trend in general, the trend directions of every month are heterogeneous.

TABLE 2. Seasonal Kendall Tests on Monthly Series

Streamflow	*tau*	z statistic	trend p-value	Het p-value
TNH	-0.0178	-0.2732	0.7847	0.3705
TG	-0.2431	-3.5561	0.0057	0.0039
Danube	-0.0084	-0.2010	0.8407	0.2558
Rhine	0.0089	0.2047	0.8378	0.5125
Ocmulgee	-0.0101	-0.2078	0.8354	0.5105
Umpqua	-0.0129	-0.3120	0.7550	0.8185
Null hypothesis of trend test: *tau* = 0				
Null hypothesis of trend homogeneity test: *tau* of all seasons are equal to 0. “Het” denotes the van Belle and Hughes heterogeneity test.				

Therefore, the trend of streamflows at TG in each month is further investigated with the MK test. The results are shown in Table 3. It is seen that, for the streamflows of the Yellow River at TG, the trends in December to April, and in June, are not significant, whereas in other months, there are obvious downward trends. This indicates that the discharges at TG in the summer and autumn are significantly decreased, but in winter, the change is not significant. One reason for such kind of behaviour is the similar change pattern in the monthly rainfall in the area along middles reaches of the Yellow River (Fu et al., 2004). Another reason may be the runoff regulation of about 10 dams over the main channel and thousands of reservoirs along the tributaries in this basin, which were mainly built over the last 50 years.

TABLE 3. Mann-Kendall Tests for streamflows at TG in each month

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Tau	-.15	.031	.018	-.14	-.51	-.18	-.35	-.29	-.33	-.46	-.48	-.07
p	.183	.790	.885	.200	.000	.110	.002	.008	.003	.000	.000	.561

3. Stationarity Test

In most applications of hydrological modelling, we have an assumption of stationarity. It is thus necessary to test for stationarity for the justification of using those models. On the other hand, sometimes the investigation of nonstationarity may give us some insights into the underlying physical mechanism, especially in the context of global changes. Therefore, testing for stationarity is an important topic in the field of hydrology.

3.1. TEST METHODS

There are roughly two groups of methods for testing stationarity. The first group is based on the idea of analyzing the statistical differences of different segments of a time series (e.g., Chen and Rao, 2002). If the observed variations in a certain parameter of different segments are found to be significant, that is, outside the expected statistical fluctuations, the time series is regarded as nonstationary. Another group of stationarity tests is based on statistics for the full sequence. We adopt the second approach here.

The stationarity test is carried out with two methods in this present study. The first one is the augmented Dickey-Fuller (ADF) unit root test that is first proposed by Dickey and Fuller (1979) and then modified by Said and Dickey (1984). It tests for the presence of unit roots in the series (difference stationarity). The other one is the KPSS test proposed by Kwiatkowski et al. (1992), which tests for the stationarity around a deterministic trend (trend stationarity) and the stationarity around a fixed level (level stationarity). KPSS test can also be modified to be used as a unit root test, but it was shown by Shin and Schmidt (1992) that the KPSS statistic, designed for use as a test for stationarity, was not as good a unit root test as other standard test. In particular, its power is noticeably less than the power of the Dickey-Fuller test (or other similar tests) against stationary alternatives.

3.1.1. *ADF test*

Dickey–Fuller unit-root tests are conducted through the ordinary least squares (OLS) estimation of regression models incorporating either an intercept or a linear trend. Consider the autoregressive AR (1) model

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad t = 1, 2, ..., N, \tag{20}$$

where $x_0 = 0$; $|\rho| \leq 1$ and ε_t is a real valued sequence of independent random variables with mean zero and variance σ^2. If $\rho = 1$, the process $\{x_t\}$ is nonstationary and it is known as a random walk process. In contrast, if $|\rho| <$ 1, the process $\{x_t\}$ is stationary. The maximum likelihood estimator of ρ is the OLS estimator

$$\hat{\rho} = \left(\sum_{t=2}^{N} x_{t-1}^2 \right)^{-1} \sum_{t=2}^{N} x_t x_{t-1} . \tag{21}$$

Under the null hypothesis that $\rho = 1$, Dickey and Fuller (1979) showed that $\hat{\rho}$ is characterized by

$$N(\hat{\rho}-1) = \frac{N\sum_{t=2}^{N}(x_t x_{t-1} - x_{t-1}^2)}{\sum_{t=2}^{N} x_{t-1}^2} \xrightarrow{D} \frac{\Lambda^2 - 1}{2\Gamma}, \tag{22}$$

where

$$(\Gamma, \Lambda) = (\sum_{i=1}^{\infty} \gamma_i^2 Z_i^2, \sum_{i=1}^{\infty} 2^{1/2} \gamma_i Z_i), \tag{23}$$

with

$$\gamma_i = 2(-1)^{i+1}/[(2i - 1)\pi], \tag{24}$$

and the Z_i are i.i.d N(0,1) distributed random variables.

The result with Eq. 22 allows the point estimate $\hat{\rho}$ to be used by itself to test the null hypothesis of a unit root. Another popular statistic for testing the null hypothesis that $\rho = 1$ is based on the usual OLS *t*-test of this hypothesis,

$$t = \frac{\hat{\rho} - 1}{\hat{\sigma}_{\hat{\rho}}}, \tag{25}$$

where $\hat{\sigma}_{\hat{\rho}}$ is the usual OLS standard error for the estimated coefficient,

$$\hat{\sigma}_{\hat{\rho}} = s_e \left(\sum_{t=2}^{N} x_{t-1}^2 \right)^{-1/2} , \tag{26}$$

and s_e denotes the standard deviation of the OLS estimate of the residuals in the regression model with Eq. 20, estimated as

$$s_e^2 = \frac{1}{N-2} \sum_{t=2}^{N} (x_t^2 - \hat{\rho} x_{t-1})^2 . \tag{27}$$

Dickey and Fuller (1979) derived the limiting distribution of the statistic t under the null hypothesis that $\rho = 1$ as

$$t \xrightarrow{D} 2\Gamma^{-1/2}(\Lambda^2 - 1) . \tag{28}$$

A set of tables of the percentiles of the limiting distribution of the statistic t under $\rho = 1$ is available in Fuller (1976, pp. 371, 373). The test rejects $\rho = 1$ when t is “too negative”.

The unit root test described above is valid if the time series $\{x_t\}$ is well characterized by an AR(1) with white noise errors. Many hydrological time series, however, have a more complicated dynamic structure than is captured by a simple AR(1) model. The basic autoregressive unit root test can be augmented (referred to as ADF test) to accommodate general ARMA(p, q) models with unknown orders (Said and Dickey, 1984; Hamilton, 1994, pp 516-530). The ADF test is based on estimating the test regression

$$x_t = \beta D_t + \phi x_{t-1} + \sum_{j=1}^{p} \psi_j \nabla x_{t-j} + \varepsilon_t , \;\; t = 1, 2, ..., N , \tag{29}$$

where D_t is a vector of deterministic terms (constant, trend, etc.). The p lagged difference terms, ∇x_{t-j}, are used to approximate the ARMA structure of the errors, and the value of p is set so that the error ε_t is serially uncorrelated. Said and Dickey (1984) show that the Dickey-Fuller procedure, which was originally developed for autoregressive representations of known order, remains valid asymptotically for a general ARIMA(p, 1, q) process in which p and q are unknown orders.

3.1.2. *KPSS test*

Let $\{x_t\}$, $t = 1, 2, \ldots, N$, be the observed series for which we wish to test stationarity. Assume that we can decomposes the series into the sum of a deterministic trend, a random walk, and a stationary error with the following linear regression model

$$x_t = r_t + \beta t + \varepsilon_t \tag{30}$$

where r_t is a random walk, i.e., $r_t = r_{t-1} + u_t$, and u_t is iid $N(0, \sigma_u^2)$; βt is a deterministic trend; ε_t is a stationary error.

To test in this model if x_t is a trend stationary process, namely, the series is stationary around a deterministic trend, the null hypothesis will be $\sigma_u^2 = 0$, which means that the intercept is a fixed element, against the alternative of a positive σ_u^2. In another stationarity case, the level stationarity, namely, the series is stationary around a fixed level, the null hypothesis will be $\beta = 0$. So that, under the null hypothesis, in the case of trend stationary, the residuals e_t ($t = 1, 2, \ldots, N$) are from the regression of x on an intercept and time trend, $e_t = \varepsilon_t$; whereas in the case of level stationarity, the residuals e_t are from a regression of x on intercept only, that is $e_t = x_t - \overline{x}$.

Let the partial sum process of the e_t as

$$S_t = \sum\nolimits_{j=1}^{t} e_j \,, \tag{31}$$

and σ^2 be the long-run variance of e_t, which is defined as

$$\sigma^2 = \lim N^{-1} E\left[S_N^2 \right]. \tag{32}$$

The consistent estimator of σ^2 can be constructed from the residuals e_t by (Newey and West, 1987)

$$\hat{\sigma}^2(p) = \frac{1}{N}\sum_{t=1}^{N} e_t^2 + \frac{2}{N}\sum_{j=1}^{p} w_j(p) \sum_{t=j+1}^{N} e_t e_{t-j} \,, \tag{33}$$

where p is the truncation lag, $w_j(p)$ is an optional weighting function that corresponds to the choice of a special window, e.g., Bartlett window (Bartlett, 1950) $w_j(p) = 1 - j/(p+1)$.

Then the KPSS test statistic is given by

$$KPSS = N^{-2} \sum_{t=1}^{N} S_t^2 \big/ \hat{\sigma}^2(p). \tag{34}$$

Under the null hypothesis of level stationary,

$$KPSS \rightarrow \int_0^1 V_1(r)^2 dr \,, \tag{35}$$

where $V_1(r)$ is a standard Brownian bridge: $V_1(r) = B(r) - rB(1)$ and $B(r)$ is a Brownian motion process on $r \in [0, 1]$. Under the null hypothesis of trend stationary,

$$KPSS \rightarrow \int_0^1 V_2(r)^2 dr\,, \tag{36}$$

where $V_2(r)$ is the second level Brownian bridge, given by

$$V_2(r) = B(r) + (2r - 3r^2)B(1) + (-6r + 6r^2)\int_0^1 B(s)ds\,. \tag{37}$$

The upper tail critical values of the asymptotic distribution of the KPSS statistic are listed in Table 4, given by Kwiatkowski et al. (1992).

TABLE 4. Upper tail critical values for the KPSS test statistic asymptotic distribution

Distribution	Upper tail percentiles			
	0.1	0.05	0.025	0.01
$\int_0^1 V_1(r)^2 dr$	0.347	0.463	0.574	0.739
$\int_0^1 V_2(r)^2 dr$	0.119	0.146	0.176	0.216

3.2. RESULTS OF STATIONARITY TESTS RESULTS OF STATIONARITY TESTS

Because on the one hand both the ADF test and the KPSS test are based on the linear regression, which assumes a normal distribution; on the other hand, the log-transformation can convert an exponential trend possibly present in the data into a linear trend, therefore, it is common to take logs of the data before applying the ADF test and the KPSS test (e.g., Gimeno et al., 1999). In this study, the streamflow data are also log-transformed before the stationarity tests.

An important practical issue for the implementation of the ADF test as well as the KPSS test is the specification of the truncation lag values of p in Eqs. 29 and 33. The KPSS test statistics are fairly sensitive to the choice of p, and in fact for every series the value of the test statistic decreases as p increases (Kwiatkowski et al., 1992). If p is too small then the remaining serial correlation in the errors will bias the test toward rejecting the null hypothesis. If p is too large then the power of the test will suffer. The larger the p, the less likely was the null hypothesis to be rejected. Following Schwert (1989), Kwiatkowski et al. (1992) and some others, the number of lag length is subjectively chosen as $p = \text{int}[x(N/100)^{1/4}]$, with $x = 4$, 12 in the present study for streamflow processes at from monthly to daily timescales. For annual series, because the autocorrelation at lag one is very low, so it is generally enough to exclude the serial correlation by choosing p = 1. The function *unitroot* and *stationaryTest* implemented in S+FinMetrics version 1.0 (Zivot and Wang, 2002) are used to do the ADF test and KPSS test. Table 5 shows the results.

The test results show that, except for the streamflow process of the Yellow River at TG which has significant downward trend at different timescales, all the other streamflow series appear to be stationary, since we cannot accept the unit root hypothesis with ADF test at 1% significance level and cannot reject the level stationarity hypothesis with KPSS test mostly at the 10% level or at least at the 2.5% level. In fact, the level stationarity is a major criterion in selecting streamflow series in the present study, while the use of the streamflow series at TG is for the purpose of comparison. For some series (such as the daily series of Rhine at Lobith, etc.) the hypothesis of trend stationarity is rejected by the KPSS test or just accepted at a low significance level, especially when the lag p is small. But this seems to be unreasonable, because the level stationarity can also be interpreted as the stationarity around a deterministic trend with a slope of zero. Therefore, we still consider these series stationary.

TABLE 5. Stationarity test results for log-transformed streamflow series

Station	Series	Lag	KPSS level stationary		KPSS trend stationary		ADF unit root	
			results	p-value	results	p-value	Results	p-value
Yellow	Daily	14	0.366	>0.05	0.366	<0.01	-7.6	4.03E-11
(TNH)		42	0.138	>0.1	0.138	>0.05	-10.89	2.18E-23
	1/3-montly	8	0.078	>0.1	0.078	>0.1	-15.16	1.88E-40
		24	0.113	>0.1	0.113	>0.1	-8.369	2.49E-13
	Monthly	6	0.084	>0.1	0.084	>0.1	-14.2	1.26E-31
		18	0.115	>0.1	0.115	>0.1	-5.982	2.11E-06
	Annual	1	0.186	>0.1	0.1797	>0.01	-4.689	2.53E-03
Yellow	Daily	13	8.6673	<0.01	0.6473	<0.01	-13.38	1.12E-34
(TG)		41	3.5895	<0.01	0.2744	<0.01	-12.4	4.83E-30
	1/3-montly	7	2.3768	<0.01	0.1861	>0.01	-12.55	5.86E-29
		23	1.7241	<0.01	0.166	>0.025	-7.774	2.11E-11
	Monthly	5	1.8194	<0.01	0.1567	>0.025	-8.661	2.08E-13
		17	1.0985	<0.01	0.1239	>0.05	-4.7	7.69E-04
	Annual	1	1.0367	<0.01	0.1277	>0.05	-4.665	3.17E-03
Danube	Daily	17	0.173	>0.1	0.1699	>0.025	-16.96	6.18E-53
(Achleiten)		51	0.0737	>0.1	0.0724	>0.1	-14.32	1.93E-39

Station	Series	Lag	KPSS level stationary		KPSS trend stationary		ADF unit root	
			results	p-value	results	p-value	results	p-value
	1/3-montly	9	0.0486	>0.1	0.048	>0.1	-15.71	9.98E-45
		28	0.0539	>0.1	0.0533	>0.1	-9.263	1.20E-16
	Monthly	7	0.0478	>0.1	0.0472	>0.1	-14.44	5.57E-36
		21	0.0445	>0.1	0.0441	>0.1	-7.056	3.01E-09
	Annual	1	0.0347	>0.1	0.0335	>0.1	-8.465	7.02E-10
Rhine	Daily	17	0.413	>0.05	0.394	<0.01	-19.23	6.58E-65
(Lobith)		51	0.186	>0.1	0.178	>0.01	-14.54	1.64E-40
	1/3-montly	9	0.119	>0.1	0.114	>0.1	-13.45	3.49E-34
		29	0.076	>0.1	0.073	>0.1	-8.13	1.01E-12
	Monthly	7	0.088	>0.1	0.081	>0.1	-10.18	1.97E-19
		22	0.064	>0.1	0.059	>0.1	-6.573	5.71E-08
	Annual	1	0.0702	>0.1	0.0496	>0.1	-8.57	3.23E-10
Ocmulgee	Daily	16	0.543	>0.025	0.408	<0.01	-16.21	5.26E-49
(Macon)		48	0.228	>0.1	0.171	>0.025	-11.61	1.39E-26
	1/3-montly	9	0.128	>0.1	0.1	>0.1	-13.5	4.11E-34
		27	0.121	>0.1	0.095	>0.1	-8.515	5.91E-14
	Monthly	6	0.097	>0.1	0.086	>0.1	-13.73	2.19E-32
		20	0.081	>0.1	0.072	>0.1	-5.473	2.33E-05
	Annual	1	0.0773	>0.1	0.0749	>0.1	-6.311	6.27E-06
Umpqua	Daily	17	0.254	>0.1	0.242	<0.01	-13.47	2.85E-35
(Elkton)		51	0.101	>0.1	0.096	>0.1	-15.23	5.27E-44
	1/3-montly	9	0.061	>0.1	0.059	>0.1	-21.25	3.99E-71
		29	0.136	>0.1	0.133	>0.1	-9.894	4.94E-19
	Monthly	7	0.079	>0.1	0.08	>0.1	-20.1	2.66E-58
		22	0.133	>0.1	0.132	>0.05	-5.856	3.19E-06
	Annual	1	0.1334	>0.1	0.1328	>0.05	-7.124	1.11E-07

Two issues should be noticed. Firstly, although no significant cycle with a period longer than one year is detected with spectral analysis for any streamflow series in the study (results are not shown here for saving space), as we will see later in Section 4, streamflow processes normally exhibit strong seasonality, therefore, have periodic stationarity, rather than the stationarity we talk about normally. According to the results shown in Table 5, KPSS test is not powerful enough to distinguish the periodic stationarity from the stationarity in normal sense. Secondly, it is not clear how the presence of seasonality impacts the test of stationarity. Besides testing for nonstationarity in log-transfomed series, we have also tested the stationarity for the deseasonalized streamflow series. The deseasonalization is conducted by firstly taking log-transformation, then subtracting the seasonal (daily, 1/3-monthly or monthly) mean values and dividing by seasonal standard deviations. The results are presented in Table 6, which show that all the test results are generally larger for KPSS test and "less negative" for ADF test. In consequence, the *p*-values decrease for KPSS test, indicating the increase of the probability of rejecting the hypothesis of stationariy, and increase for ADF test, indicating the increase of the probability (though still very small) of accepting the hypothesis of unit root. That is, the removal of seasonality in the mean and variance tends to make the streamflow series less stationary, or at least from the point of view of the KPSS test. This is an issue open for future investigation.

TABLE 6. Stationarity test results for log-transformed and deseasonalized streamflow series

Station	Series	Lag	KPSS level stationary		KPSS trend stationary		ADF unit root	
			results	p-value	results	p-value	results	p-value
Yellow	Daily	14	2.4024	<0.01	2.3961	<0.01	-11.940	5.40E-28
(TNH)		42	0.9972	<0.01	0.9946	<00.01	-8.869	2.05E-15
	1/3-montly	8	0.5581	>0.025	0.5579	<00.01	-6.842	9.75E-09
		24	0.266	>0.1	0.2659	<00.01	-5.133	1.06E-04
	Monthly	6	0.298	>0.1	0.297	<00.01	-5.005	2.14E-04
		18	0.1742	>0.1	0.1737	>0.025	-5.123	1.29E-04
Yellow	Daily	13	12.5127	<0.01	1.2348	<00.01	-16.66	4.60E-51
(TG)		41	5.5909	<0.01	0.5794	<00.01	-11.220	8.30E-25
	1/3-montly	7	3.3145	<0.01	0.3638	<00.01	-8.294	4.89E-13
		23	1.5704	<0.01	0.1904	>0.01	-5.043	1.61E-04
	Monthly	5	1.7948	<0.01	0.2086	>0.01	-6.385	2.66E-07
		17	0.8977	<0.01	0.1203	>0.05	-4.833	4.53E-04
Danube	Daily	17	0.2835	>0.1	0.2892	<0.01	-21.79	3.07E-78
(Achleiten)		51	0.1366	>0.1	0.1394	>0.05	-15.71	1.82E-46
	1/3-montly	9	0.0934	>0.1	0.0951	>0.1	-12.59	2.94E-30
		28	0.054	>0.1	0.0549	>0.1	-8.038	2.05E-12
	Monthly	7	0.0577	>0.1	0.0589	>0.1	-8.712	2.81E-14
		21	0.0407	>0.1	0.0415	>0.1	-6.307	2.73E-07
Rhine	Daily	17	0.5229	>0.025	0.5085	<0.01	-19.25	4.93E-65
(Lobith)		51	0.2347	>0.1	0.2282	<0.01	-14.01	6.76E-38
	1/3-montly	9	0.15	>0.1	0.1452	>0.05	-11.93	2.21E-27
		29	0.0797	>0.1	0.0772	>0.1	-8.065	1.65E-12
	Monthly	7	0.0966	>0.1	0.0897	>0.1	-8.369	3.38E-13
		22	0.0657	>0.1	0.0611	>0.1	-6.463	1.09E-07

Station	Series	Lag	KPSS level stationary		KPSS trend stationary		ADF unit root	
			results	p-value	results	p-value	results	p-value
Ocmulgee	Daily	16	0.9316	<0.01	0.7318	<0.01	-21.160	7.05E-75
(Macon)		48	0.4387	>0.05	0.3449	<0.01	-12.460	1.92E-30
	1/3-montly	9	0.2471	>0.1	0.2002	>0.01	-9.924	4.76E-19
		27	0.1347	>0.1	0.1092	>0.1	-7.563	6.93E-11
	Monthly	6	0.1366	>0.1	0.1253	>0.05	-8.025	5.45E-12
		20	0.0836	>0.1	0.0766	>0.1	-4.864	3.65E-04
Umpqua	Daily	17	1.2445	<0.01	1.2436	<0.01	-21.170	4.38E-75
(Elkton)		51	0.5697	>0.025	0.5694	<0.01	-14.500	2.38E-40
	1/3-montly	9	0.3536	>0.05	0.355	<0.01	-12.830	2.24E-31
		29	0.2	>0.1	0.2009	>0.01	-7.936	4.31E-12
	Monthly	7	0.2109	>0.1	0.2151	>0.01	-9.110	1.19E-15
		22	0.1365	>0.1	0.1392	>0.05	-5.166	9.44E-05

4. Seasonality Analysis

4.1. SEASONALITY IN MEAN AND VARIANCE

The dynamics of streamflow are often dominated by annual variations. How well the seasonality is captured is a very important criterion for assessing a stochastic model for streamflow. The seasonality of hydrological processes is often described in terms of the mean values, the variances, the extrema, and the probability distribution of the variable in each season (in general, a season may denote a day, a month, etc.). We will use the daily streamflow series to present the approaches we adopt here for analyzing the seasonality. The same approaches can be easily adapted to the cases of 1/3-monthly series and monthly series.

To make it convenient to analyze the seasonality of a daily flow series of N years, we rewrite it as the following matrix form:

$$X = \begin{Bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,365} \\ x_{2,1} & x_{2,2} & \dots & x_{2,365} \\ \vdots & \vdots & x_{j,i} & \vdots \\ x_{N,1} & x_{N,2} & \dots & x_{N,365} \end{Bmatrix}, \tag{38}$$

where the rows denote year 1 ~ N, the columns denote day 1 ~ 365. For simplicity, the 366th days of leap years are omitted. This would not introduce major errors when analyzing seasonality of daily flows.

Consequently, the mean value, standard deviation and coefficient of variation of each column of the matrix are the daily mean discharge, standard deviation and coefficient of variation (CV) of daily discharges for each day over the year. They are easily calculated as follows:

Mean value:

$$\overline{x}_i = \frac{1}{N}\sum_{j=i}^{N} x_{j,i} \; ; \tag{39}$$

Standard deviation:

$$s_i = \left(\frac{1}{N}\sum_{j=1}^{N} (x_{j,i} - \overline{x}_i)^2 \right)^{1/2} \; ; \tag{40}$$

Coefficient of variation:

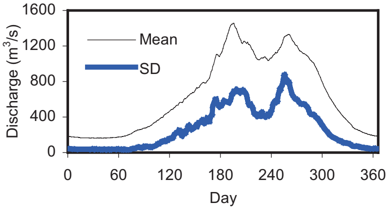
$$CV_i = \frac{s_i}{\overline{x}_i} \; . \tag{41}$$

Daily mean values and standard deviations of the six streamflow processes are shown in Figure 2 (a ~ f), and the daily variations in CVs are shown in Figure 3 (a ~c). It is shown that, days with high mean values have also high standard deviations, this is a property which has been well recognized (e.g., Mitosek, 2000). But two exceptional cases here are the streamflow processes of Danube and Ocmulgee. Danube has a clear seasonality in mean values, but no clear seasonality in variances. In consequence, it has a similar seasonal pattern in CVs to the Rhine River, as shown in Figure 3 (b). Ocmulgee has no clear seasonal variations in CVs although it has clear seasonality in means and variances. In June, thunderstorm activity results in high CV values in the daily streamflow process of Ocmulgee, as shown in Figure 3 (c).

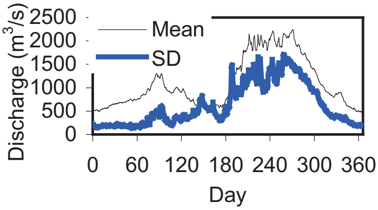
Two special points should be noted about the variations in streamflow processes of the Yellow River:

1. Streamflow processes of the Yellow River at both TNH and TG are characterized by a bimodal distribution. Extrema occur in July and September at TNH and in late March to early April and August at TG. However, the causes of bimodality of the two streamflow processes are different. The bimodality of the streamflow process at TNH exits mainly in response to the bimodal distribution of rainfall, whereas the first peaks of the streamflow process at TG is caused by snowmelt water and the break-up of the river-ice jam in spring and the second peak is due to concentrated rainfall.

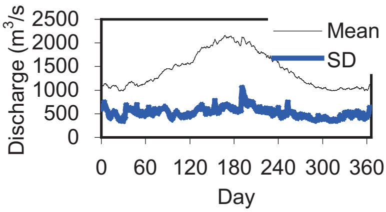
2. Although the contributing area of TG is about as 5 times larger as that of TNH, the streamflow process at TNH changes much smoother than that at TG, as indicated by CVs shown in Figure 3. This is mainly because of less rainfall variability and much less anthropogenic disturbances in the watershed above TNH.



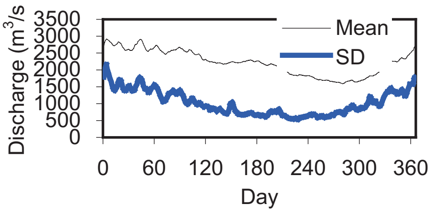
(a) Yellow River at TNH



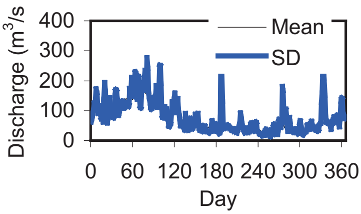
(b) Yellow River at TG



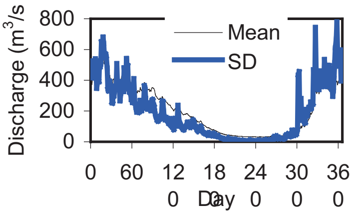
(c) Danube River at Achleiten



(d) Rhine River at Lobith

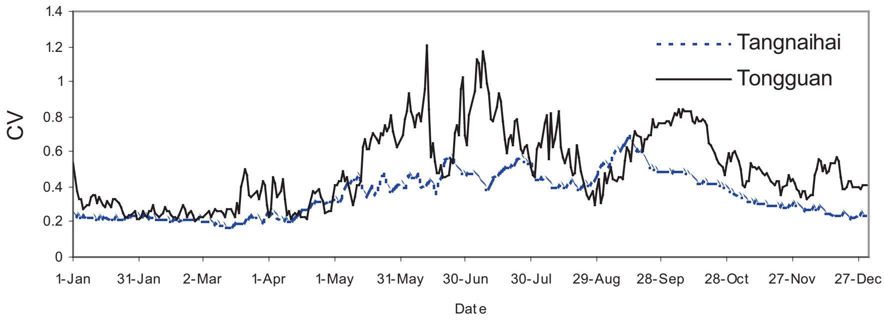


(e) Ocmulgee River at Macon

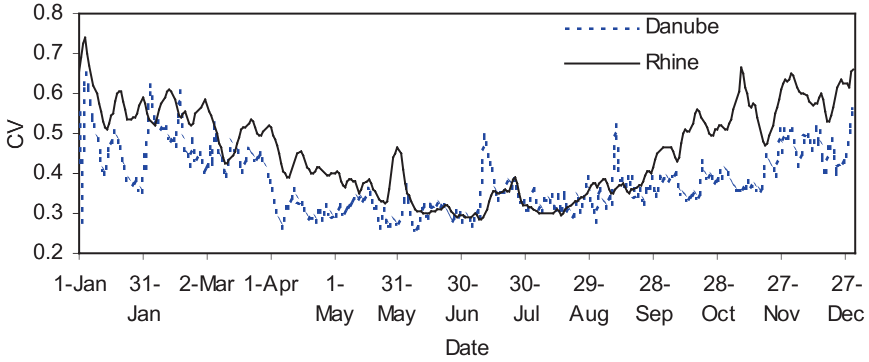


(f) Umpqua River near Elkton

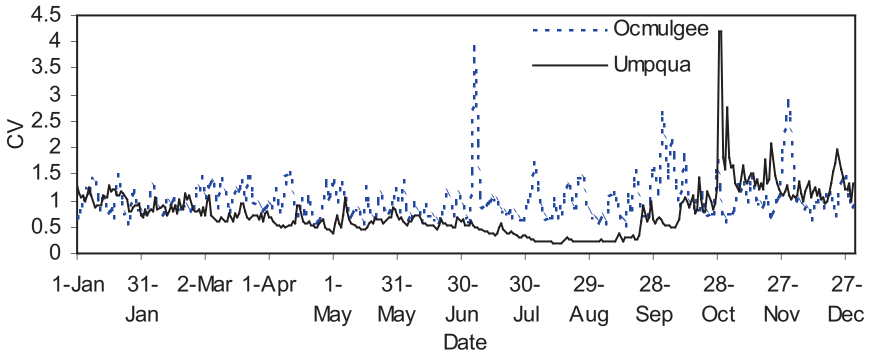
Figure 2. Variation in daily mean and standard deviation of streamflow processes



(a) Yellow River at TNH and TG



(b) Danube River at Achleiten and Rhine River at Lobith



(c) Ocmulgee River at Macon and Umpqua River near Elkton

Figure 3. Seasonal variation in CVs of streamflow processes

4.2. DETREND, NORMALIZATION AND DESEASONALIZATION

After trend analysis and seasonality analysis, we can remove the trend component and seasonal components out of the original river flow series, and get an approximately stationary process, then further analyse autocorrelation properties and long-memory properties.

Because streamflow series are skewed and heavily tailed, whereas many models, such as regression models or autoregressive moving average (ARMA) models, require the time series data to be normally distributed, it is thus necessary to normalize the data to make them at least approximately normally distributed. The most popular approach is the Box-Cox transformation (Box and Cox, 1964):

$$x = \begin{cases} \lambda^{-1}[(x+c)^{\lambda} - 1] & \lambda \neq 0 \\ \ln(x+c) & \lambda = 0 \end{cases} . \tag{42}$$

Usually we simply take logarithm to normalize the data. After log-transformation, we can estimate the trend by fitting a regression models if the trend is present, and then subtract it out of the original series.

The deseasonalization can be viewed as the standardization for each season (in the case of daily streamflow series, each season means each day). To do this, we use the daily mean $\overline{x}_i$, standard deviation s_i given by Eqs. (39) and (40), then apply to each element $x_{j,i}$ in matrix (38) the following standardization transformation:

$$y_{j,i} = \frac{x_{j,i} - \overline{x}_i}{s_i} . \tag{43}$$

With the above pre-processing procedure, the seasonality in mean values and standard deviations in the streamflow series is removed. With such deseasonalized series, we go further to make autocorrelation analysis.

4.3. SEASONALITY IN AUTOCORRELATION STRUCTURES

Given a time series $\{x_i\}$, $i = 1, 2, \ldots, n$, the autocorrelation function (ACF) at lag k for the time series is given by (Box and Jenkins, 1976)

$$\hat{\rho}(k) = c_k / c_0 , \tag{44}$$

where $k = 0, 1, 2, \ldots$, and

$$c_k = \frac{1}{n-k}\sum_{i=1}^{n-k}(x_i - \bar{x})(x_{i+k} - \bar{x}) . \tag{45}$$

The ACF obtained in this way takes the whole time series into consideration, which reflects the overall autocorrelation property for the time series, but to examine the seasonal variation in the autocorrelation structure of a daily streamflow series, we need to calculate values of the autocorrelation coefficient between column vector X_i and X_{i+k} of matrix (Eq.38), where i = 1, 2, …, 365 and k = 0, 1, 2, …, $k_{\max}$, ($k_{\max} \le 365$) (Mitosek, 2000), namely,

$$\hat{\rho}_i(k) = \begin{cases} \dfrac{\frac{1}{N}\sum_{j=1}^{N}(x_{j,i} - \bar{x}_i)(x_{j,i+k} - \bar{x}_{i+k})}{s_i s_{i+k}}, & i+k \le 365 \\ \dfrac{\frac{1}{N-1}\sum_{j=1}^{N-1}(x_{j,i} - \bar{x}_i)(x_{j+1,i+k-365} - \bar{x}_{i+k-365})}{s_i s_{i+k-365}}, & i+k > 365 \end{cases} , \tag{46}$$

where $\bar{x}_i$ and s_i are the same as in Equation (39) and (40), N is the number of years, and

$$\bar{x}_{i+k-365} = \frac{1}{N-1}\sum_{j=i}^{N-1} x_{j+1,i+k-365} , \tag{47}$$

$$s_{i+k-365} = \left(\frac{1}{N-1}\sum_{j=1}^{N} x_{j+1,i+k-365} - \bar{x}_{i+k-365})^2\right)^{1/2} . \tag{48}$$

The result obtained by Equation 46 is the autocorrelation function on a day-by-day basis, referred to as daily autocorrelation function here. It is calculated after detrending (only for the case of the streamflow of the Yellow River at TG), log-transforming and deseasonalizing the raw series. The daily autocorrelations at different lags for the six daily streamflow processes are displayed in Figure 4 to 9.

Similarly, we can deseasonalize the 1/3-monthly and monthly streamflow series, and then calculate their autocorrelations at different lags for the six 1/3-monthly and six monthly streamflow processes, as shown in Figure 10 and 11.

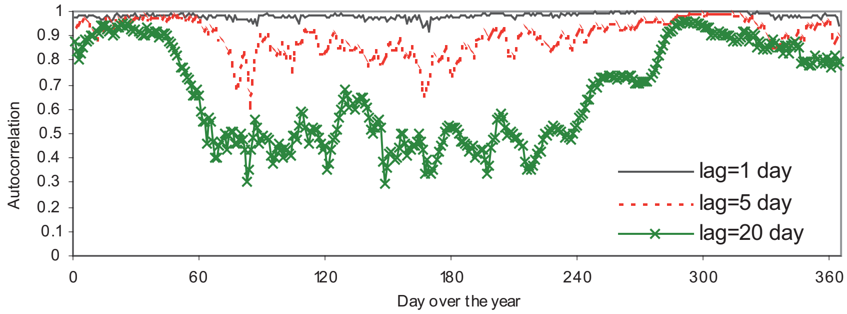


Figure 4. Daily autocorrelations at different lag days for daily flow series of Yellow River at TNH

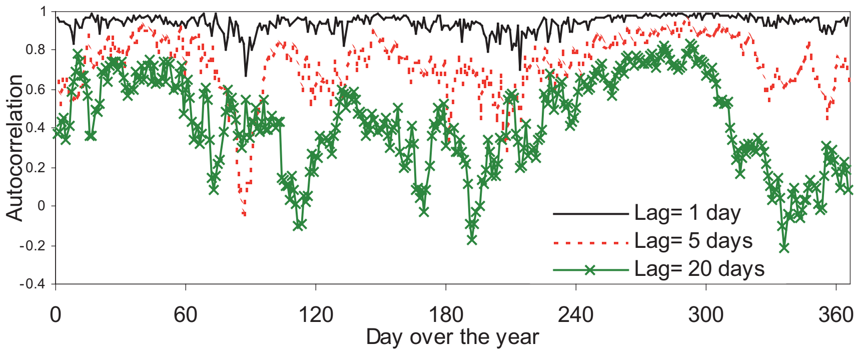


Figure 5. Daily autocorrelations at different lag days for daily flow series of Yellow River at TG

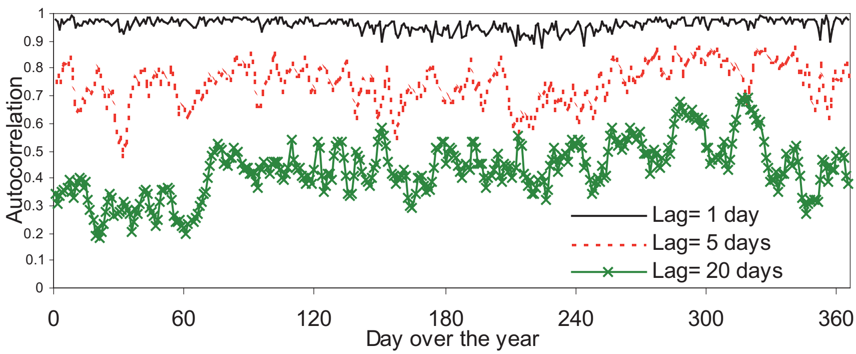


Figure 6. Daily autocorrelations at different lag days for daily flow series of the Danube

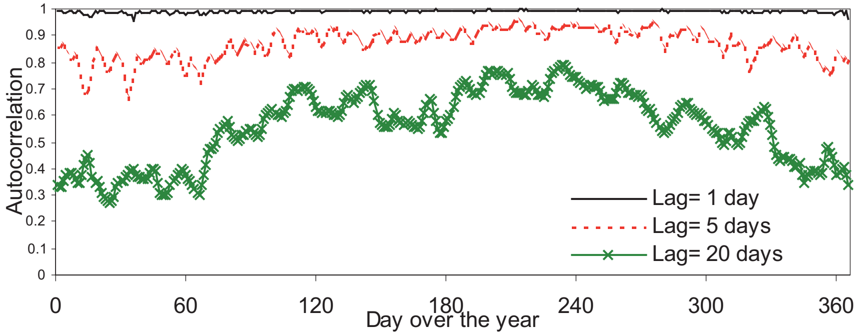


Figure 7. Daily autocorrelations at different lag days for daily flow series of the Rhine

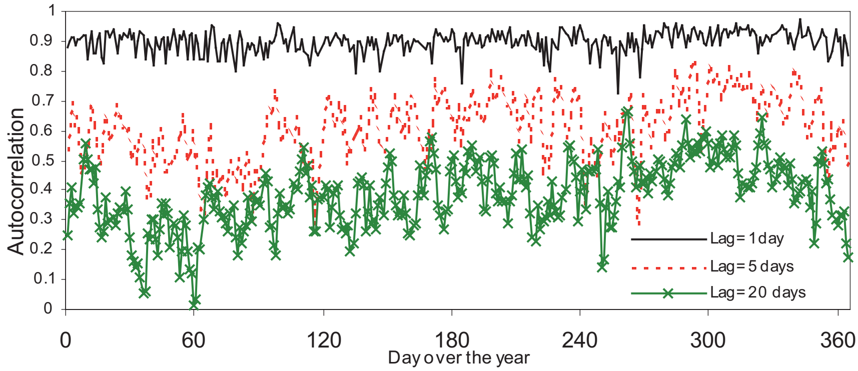


Figure 8. Daily autocorrelations at different lag days for daily flow series of the Ocmulgee

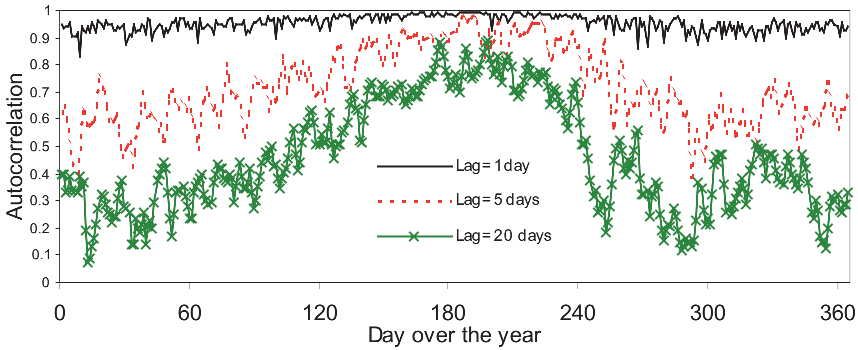


Figure 9. Daily autocorrelations at different lag days for daily flow series of the Umpqua

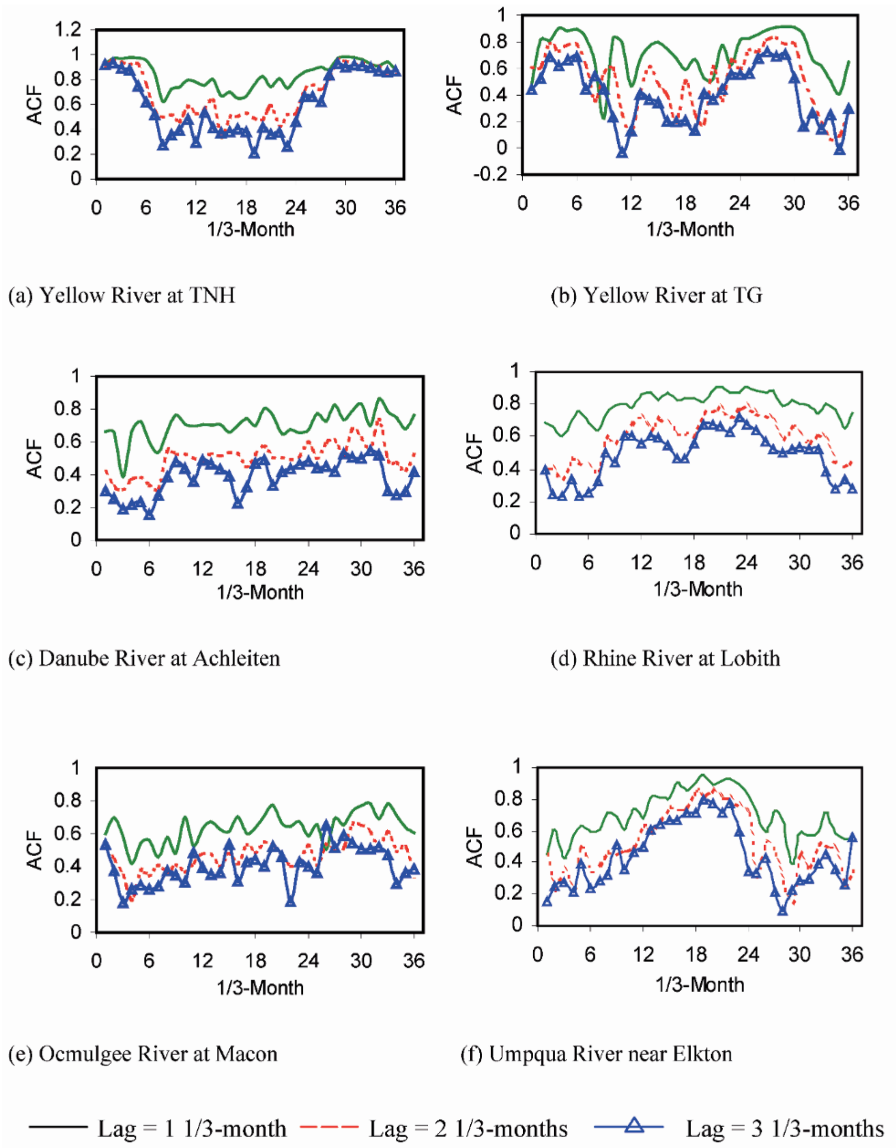


Figure 10. 1/3-monthly autocorrelations at different lags for 1/3-monthly flow series

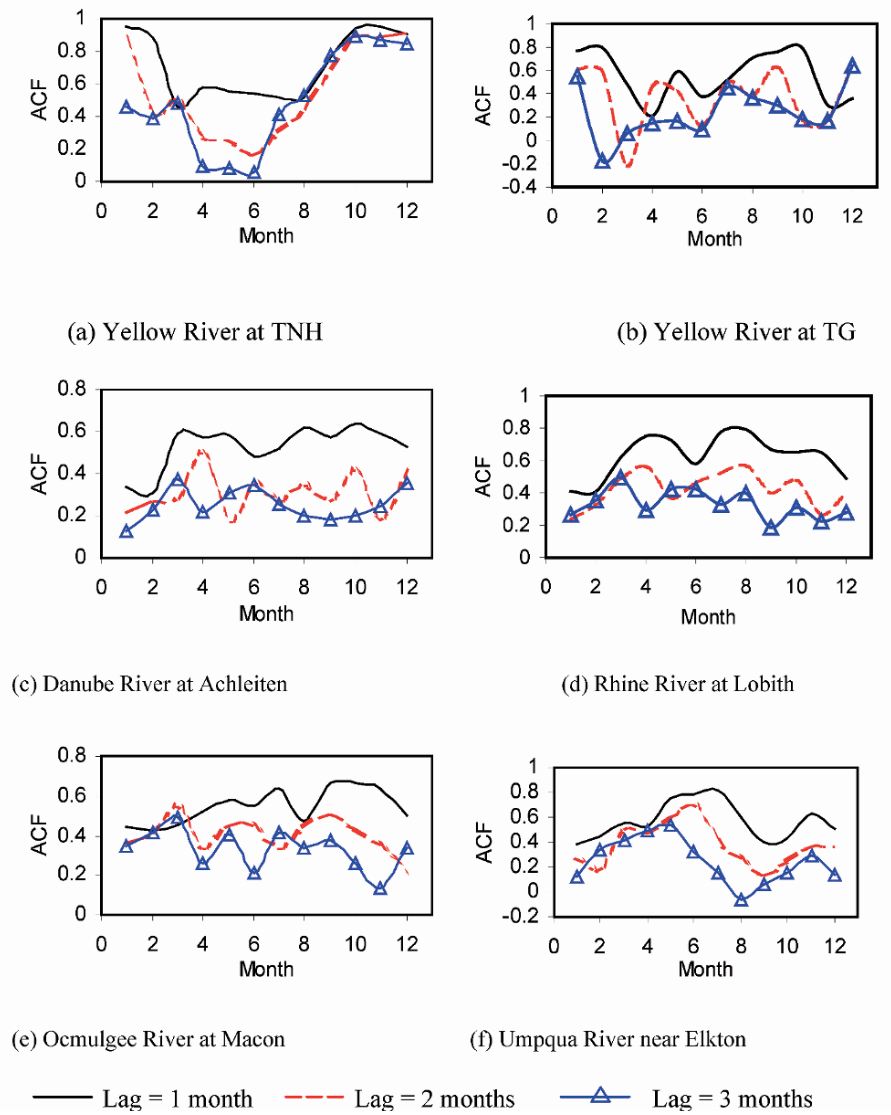


Figure 11. Monthly autocorrelations at different lags for monthly flow series

By a visual inspection of Figure 4 to 11, we see that:

1. There are more or less seasonal variations in the autocorrelation structures of all the daily, 1/3-monthly and monthly streamflow processes. In general, the autocorrelation is high for low-flow seasons and low for high flow seasons. However, there are some exceptions. For example, the daily flows of the Yellow River at TG have lower autocorrelations in late

November and December when discharges are lower than those in August to October. For the Danube, the autocorrelations in January and February are lower than those in June and July, although the flows are lower rather than those in June and July. In fact, the seasonal variation in the autocorrelation functions of streamflows processes has been observed by many researchers (e.g., Vecchia and Ballerini, 1991; Mcleod, 1994). With such kind of season dependence of the autocorrelation structure, the streamflow processes are not second-order stationary. Instead, they are periodic stationary (see the definition of the periodic stationarity in Appendix 2).

2. Daily autocorrelations of the Yellow River at TNH are generally much higher than those at TG. In the period from the end of January to February and in November, the daily autocorrelations at TNH are especially high, which can still be as high as 0.9 at a lag of 20 days. In March, June and July, daily autocorrelations at TNH are low because of large volume of snowmelt water and heavy rainfall respectively. Daily autocorrelations at TG are generally much lower because the streamflow process changes much more irregularly than that at TNH. The daily autocorrelations at TG are especially low in March because river ice-jam breakup and in July and August because of over-concentrated rainfall. In these two periods, the autocorrelations between adjacent days are very low, for instance, lower than 0.5 in the end of March and the beginning of April, and lower than 0.6 in the end of August.

5. Long-Memory Analysis

5.1. INTRODUCTION TO LONG-MEMORY

Long-memory, or long-range dependence, refers to a not negligible dependence between distant observations in a time series. 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 or long-range dependence. In the hydrology community, many studies have been carried out on the test for long-memory in streamflow 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 to detect whether long-memory is present. The resulting model also 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.

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).

$$\rho(k) \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} \quad , \tag{49}$$

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

In frequency domain, long-memory manifests itself as an unbounded spectral density at zero frequency. For a stationary discrete long-memory time series process, its spectral density at zero frequency satisfies

$$f(\lambda) \sim C\lambda^{1-2H}, \tag{50}$$

as $\lambda \rightarrow 0+$, for a positive, finite C. H is called the Hurst coefficient (or self-similarity parameter), as originally defined by Hurst (1951), and it represents the classical parameter characterizing long-memory. H is related to the fractional differencing parameter d with a relationship: $d = H - 0.5$.

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 (1969a). 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.

Many methods are available for testing for the existence of long-memory and estimating the Hurst coefficient *H* or the fractional differencing parameter *d*. Many of them are well described in the monograph of Beran (1994). These techniques include graphical methods (e.g., classical 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). Heuristic methods are useful to test if a long-range dependence exists in the data and to find a first estimate of *d* or *H*, but they are generally not accurate and not robust. The parametric methods obtain consistent estimators of *d* or *H* via maximum likelihood estimation (MLE) of parametric long-memory models. They give more accurate estimates 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.

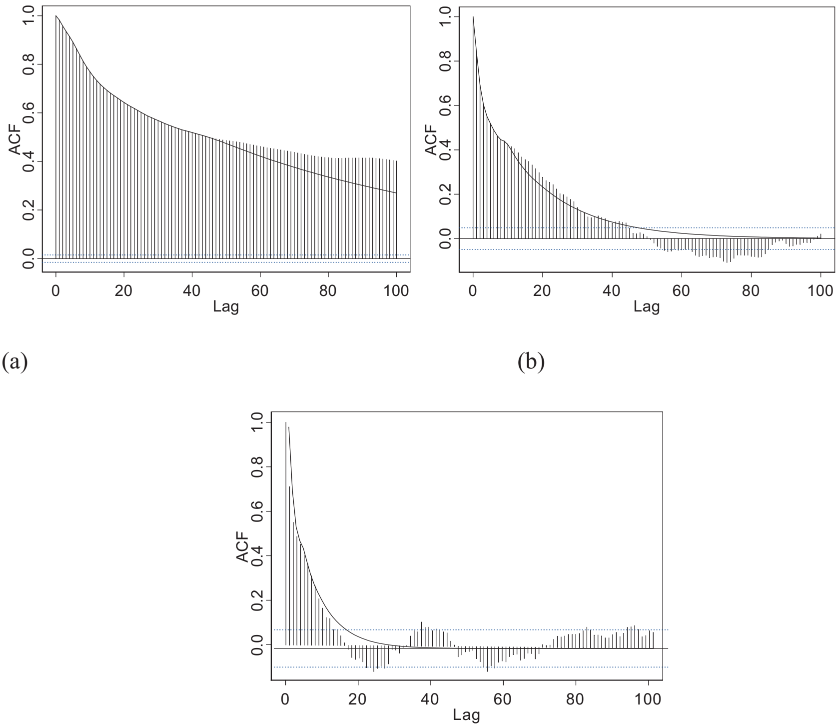
In Section 5.2, we will use three heuristic methods, i.e., autocorrelation function analysis, classical *R*/*S* analysis, and the aggregated variance method to detect the existence of long-memory in the streamflow processes of the upper and middle Yellow River at TNH and TG (To save space, other streamflow processes are not analysed with heuristic methods). Then in the Section 5.3, two statistical test methods, i.e., Lo's modified R/S test (Lo, 1991) and the GPH test (Geweke and Porter-Hudak, 1983), to test for the existence of long-memory in the streamflow processes of all the five rivers, and the maximum likelihood estimates of the fractional differencing parameter *d* will be made as well. To verify the validity of these statistical test and estimation methods, some Monte Carlo simulation results will also be presented in Section 5.3.

5.2. DETECTING LONG-MEMORY WITH HEURISTIC METHODS

5.2.1. *Autocorrelation function analysis*

In the presence of long-memory, the autocorrelation function (ACF) of a time series decreases to 0 at a much slower rate than the exponential rate implied by an short-memory ARMA model. So we can compare the sample ACF of the observed time series under investigation with the theoretical ACF (McLeod, 1975) of the ARMA model fitted to the time series. If the sample ACF of the observed series decays much slower than the ACF of the fitted ARMA model, then it probably indicates the existence of long-memory.

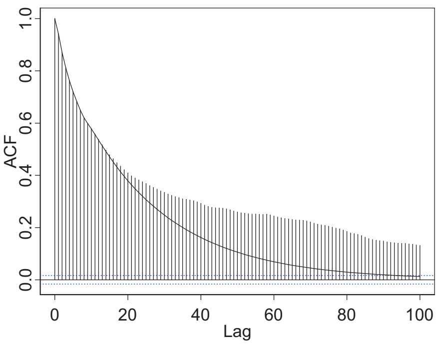
First, we select the best fitting AR models for the streamflow series using the Akaike Information Criterion (AIC) (Akaike, 1973), which turns out to be an AR(38), AR(9) and AR(4) model for the daily, 1/3-monthly, and monthly streamflow series at TNH, and an AR(9), AR(5) and AR(15) model for the daily, 1/3-monthly, and monthly streamflow series at TG. The high autoregressive order for monthly series at TG arises from the remaining seasonality that has not been fully removed with the deseasonalization procedure. The sample ACF of the streamflow series and the theoretical ACF of the fitted models from lag 1 to lag 100 are plotted in Figure 12 and 13.



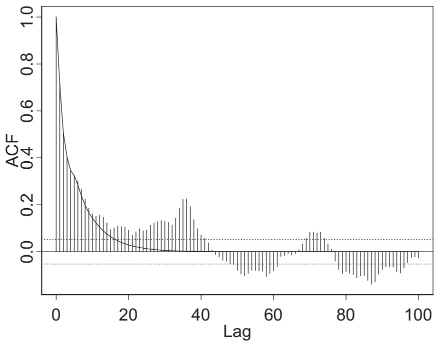
(a) (b)

(c)

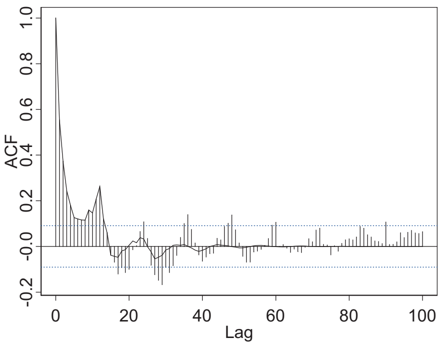
Figure 12. Sample ACF (vertical lines) and the theoretical ACF (curve line) of fitted AR models for (a) daily, (b) 1/3-monthly and (c) monthly streamflow at TNH



(a)



(b)



(c)

Figure 13. Sample ACF (vertical lines) and the theoretical ACF (curve line) of fitted AR models for (a) daily, (b) 1/3-monthly and (c) monthly streamflow at TG

Comparing the theoretical ACF of the fitted AR models with the sample ACF of the observed streamflow series, we can find that:

1.The daily streamflow process is highly persistent and the autocorrelation remains significant from zero at lag 100. The theoretical autocorrelation closely matches the sample autocorrelation at short lags. However, for large lags, the sample ACF decays much slower than the theoretical ACF.

2. The 1/3-monthly and monthly streamflow processes are much less persistent. For both 1/3-monthly flow series at TNH and TG, the sample autocorrelations are slightly larger than the theoretical autocorrelations for large lags. But for the monthly flow series, the sample ACF is basically at the same level as the theoretical ACF.

5.2.2. *Classical R/S analysis*

The *R/S* statistic, or the "rescaled adjusted range" statistic, is the adjusted range of partial sums of deviations of a times series from its mean, rescaled by its standard deviation. It was developed by Hurst (1951) in his studies of river discharges, and suggested by Mandelbrot and Wallis (1969b) using the R/S statistic to detect long-range dependence. Consider a time series $\{x_t\}$, $t = 1, 2, \ldots, N$, and define the jth partial sum as

$$Y_j = \sum_{i=1}^{j} x_i , \tag{51}$$

$j = 1, 2, \ldots, N$. Suppose to calculate the storage range of a reservoir between time t and $t+k$, and assume that: (1) the storage at time t and $t+k$ is the same; (2) the outflow during time t and $t+k$ is the same; and (3) there is no any loss of storage. Then the rescaled adjusted range, i.e., *R/S* statistic, is defined as (Beran, 1994):

$$R/S_{(t,k)} = \frac{1}{S_{(t,k)}}\left\{\max_{0\le i\le k}\left[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)\right] - \min_{0\le i\le k}\left[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)\right]\right\}, \tag{52}$$

where

$$S_{(t,k)} = \sqrt{k^{-1}\sum_{j=t+1}^{t+k}(x_j - \overline{x}_{t,k})^2} , \tag{53}$$

and

$$\overline{x}_{t,k} = k^{-1}\sum\nolimits_{j=t+1}^{t+k} x_j . \tag{54}$$

The *R/S* statistic varies with the time span k. Hurst (1951) found that the *R/S* statistic for many geophysical records is well described by the following empirical relation: $E[R/S] \sim c_1 k^H$, as $k \to \infty$, with typical values of H (the Hurst coefficient) in the interval (0.5, 1.0), and c_1 a finite positive constant that does not depend on k.

The classical *R/S* analysis is based on a heuristic graphical approach. Compute the *R/S*-statistic in Equation 52 at many different lags k and for a number of different points, and plots the resulting estimates versus the lags on log-log scale. The logarithm of k should scatter along a straight line having a slope equal to H. The value of H can be estimated by a simple least-squares fit. An H value equal to 0.5 means absence of long-memory. The higher the H is, the higher the intensity of long-memory.

The log-log plots of R/S versus different lags k for streamflow processes at both TNH and T are displayed in Figure 14 and 15. The slopes of the fitted lines are the estimates of values of H.

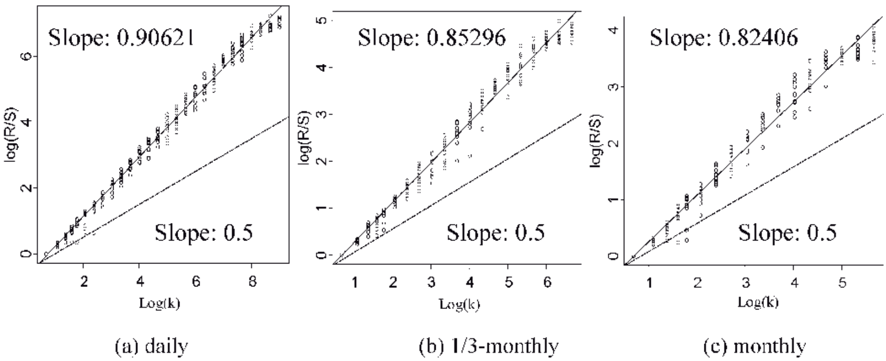


Figure 14. *R/S* plot of (a) daily, (b) 1/3-monthly and (c) monthly flow series at TNH

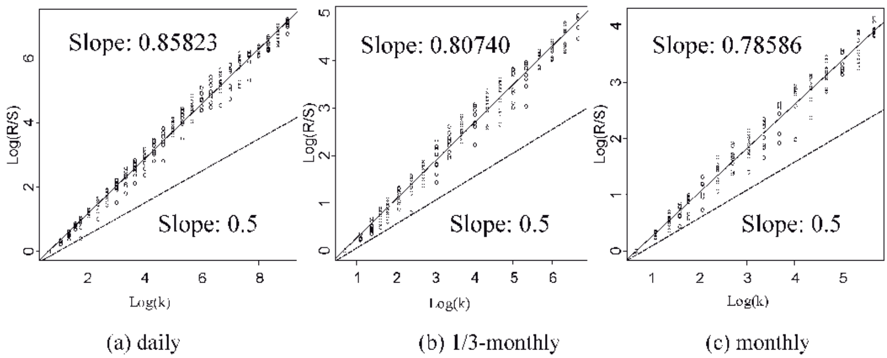


Figure 15. *R/S* plot of (a) daily, (b) 1/3-monthly and (c) monthly flow series at TG

According to the R/S statistics obtained with the graphical approach, all the streamflow series have values of H larger than 0.5, indicating the presence of long-memory in all these streamflow series. The H values, which indicate the intensity of long-memory, decrease with the increase of timescales. Furthermore, at each timescale, the intensity of long-memory of the streamflow process at TNH is stronger than that at TG.

To check the effectiveness of the R/S analysis for detecting long-memory, we generate ten simulations of an AR(1) model, ten simulations of an ARFIMA(0,*d*,0) model, and ten simulations of an ARFIMA(1,*d*,0) model. The AR(1) model is of the form $(1-\phi B)x_t = \varepsilon_t$ with ϕ =0.9, the ARFIMA(0,d,0) of form $(1-B)^d x_t = \varepsilon_t$ with $d = 0.3$, and the ARFIMA(1,d,0) of form $(1-\phi B)(1-B)^d x_t = \varepsilon_t$ with ϕ =0.9 and $d = 0.3$, where $\{\varepsilon_t\}$ are i.i.d standard normal, B is the backshift operator, i.e., $Bx_t = x_{t-1}$. Each of them has a size of 3000 points. The AR series and the ARFIMA series are produced by the *arima.sim* and *arima.fracdiff.sim* function built in S-Plus version 6 (Insightful Corporation, 2001). The estimated values of H are listed in Table 7.

TABLE 7. Estimated H values with classical R/S analysis for simulated series

Simulation	AR(1)	ARFIMA(0,d,0)	ARFIMA(1,d,0)
1	0.83789	0.75434	0.91157
2	0.79296	0.76044	0.89271
3	0.78578	0.73048	0.90742
4	0.78821	0.77499	0.87063
5	0.82238	0.75269	0.88660
6	0.82636	0.73367	0.87649
7	0.77678	0.81083	0.89122
8	0.83730	0.77748	0.91854
9	0.77904	0.76316	0.89593
10	0.83119	0.77612	0.90586
Average	0.80779	0.76342	0.89570

The simulation results show that, for a pure fractionally integrated process ARFIMA (0, d, 0), the estimate of H is very close to its true value 0.8 (i.e., d + 0.5). But when a process is a mixture of short memory and long-memory, as the ARFIMA(1, d, 0) process, then the estimates of H are biased upwardly. Furthermore, classical R/S analysis gives estimated H values (= d + 0.5) higher than 0.5 even for short memory AR (1) processes, which indicates its sensitivity to the presence of explicit short-range dependence.

5.2.3. *Aggregated Variance Method*

For independent random variables $x_1, \ldots, x_N$, the variance of sample mean is equal to $\mathrm{var}(\bar{x}) = \sigma^2 N^{-1}$. But in the presence of long-memory, Beran (1994) proved that the variance of the sample mean could be expressed by

$$\mathrm{var}(\bar{x}) \approx cN^{2H-2}, \tag{55}$$

where c>0 and H is the Hurst coefficient. Correspondingly, Beran (1994) suggested the following method for estimating the Hurst coefficient H.

1. Take a sufficient number (say m) of subseries of length k ($2 \le k \le N/2$), calculate the sample means $\bar{x}_1(k), \bar{x}_2(k), ..., \bar{x}_m(k)$ and the overall mean

$$\bar{x}(k) = m^{-1} \sum_{j=1}^{m} \bar{x}_j(k) ; \tag{56}$$

2. For each k, calculate the sample variance $s^2(k)$ of the m sample means:

$$s^2(k) = (m-1)^{-1} \sum_{j=1}^{m} (\bar{x}_j(k) - \bar{x}(k))^2 \; ; \tag{57}$$

3. Plot log $s^2(k)$ against log$k$. For large values of k, the points in this plot are expected to be scattered around a straight line with negative slope $2H - 2$. The slope is steeper (more negative) for short-memory processes. In the case of independence, the ultimate slope is -1.

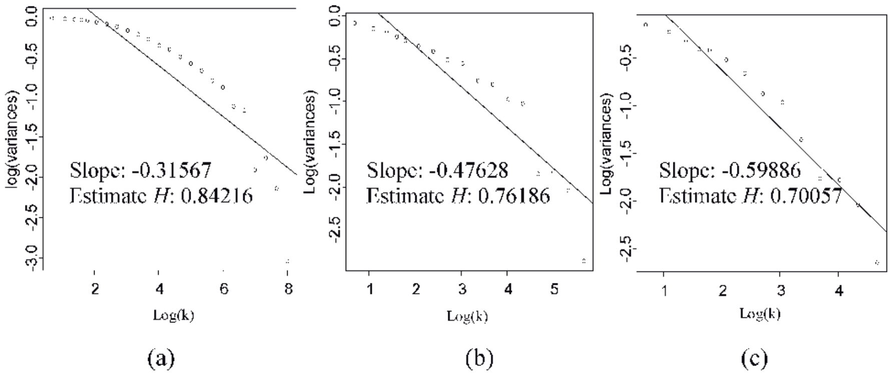


Figure 16. Variance plot of (a) daily, (b) 1/3-monthly and (c) monthly flow series at TNH

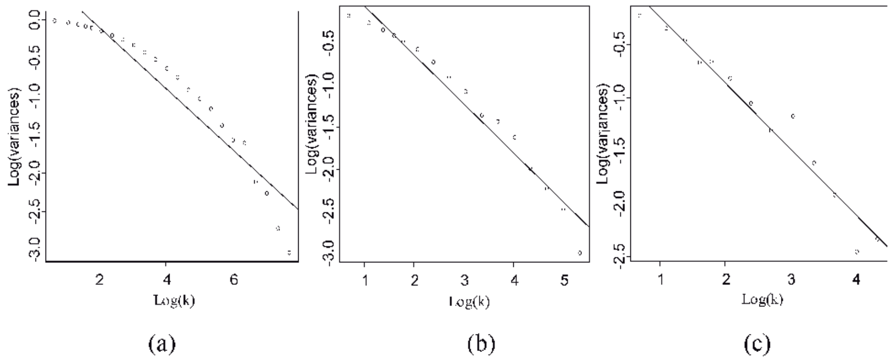


Figure 17. Variance plot of (a) daily, (b) 1/3-monthly and (c) monthly flow series at TG

Comparing the variance plot for the streamflow processes at TNH and TG, displayed in Figure 16 and 17, we can find that the slopes of the fitted lines get more negative as the timescale increases (from day to month) for the streamflow processes at both TNH and TG, which indicates that, from the view of time series themselves, the *H* values, namely the intensity of long-memory, decreases with the increase of timescales. Furthermore, at each timescale, the intensity of long-memory in streamflow process at TNH is stronger than that at TG.

Similarly to the assessment of the effectiveness of classical *R*/*S* analysis, we assess the effectiveness of variance analysis for detecting the long-memory by estimating the *H* values for the generated simulations of the AR(1) model, ARFIMA(0,d,0) model and ARFIMA(1,d,0) model. The estimated *H* values are listed in Table 8. The results show that, variance analysis is also sensitive to the presence of explicit short-range dependence, and generally it gives smaller estimate than the classical *R*/*S* analysis.

TABLE 8. Estimated *H* values with variance analysis for simulated series Estimated *H* values with variance analysis for simulated series

Simulation	AR(1)	ARFIMA(0,d,0)	ARFIMA(1,d,0)
1	0.69158	0.78782	0.83284
2	0.64412	0.71964	0.77195
3	0.66903	0.67128	0.84894
4	0.64130	0.80683	0.79878
5	0.65846	0.78597	0.87033
6	0.71512	0.71407	0.87689
7	0.68218	0.80170	0.80999
8	0.69148	0.72700	0.80335
9	0.59842	0.64447	0.82691
10	0.71557	0.72315	0.78931
Average	0.67073	0.73819	0.82293

Because both the *R*/*S* analysis method and variance plot method are sensitive to the presence of explicit short-range dependence, whereas the ACF analysis only gives us a heuristic suggestion without quantitative estimations, we need some formal statistical techniques for detecting long-memory in the streamflow series.

5.3. DETECTING LONG-MEMORY WITH STATISTICAL TEST METHOD AND MLE METHOD

In this section, we will detect the presence of long-memory in the streamflow processes of streamflow processes with two statistical test techniques, i.e., the Lo's modified *R/S* test (Lo, 1991), and the GPH test (Geweke and Porter-Hudak, 1983). In addition we will try to detect the presence of long-memory by estimating the fractional differencing parameter *d*.

5.3.1. *Lo's modified R/S analysis*

As having been shown in Section 5.2, the classical *R/S* analysis is sensitive to the presence of explicit short-range dependence structures, and it 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 $S_{(t,\ k)}$ in Eq. 52, 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. 52, only focus on lag $k = N$. The S_q is defined as

$$S_q = \left(\frac{1}{N} \sum_{j=1}^{N} (x_j - \bar{x}_N)^2 + \frac{2}{N} \sum_{j=1}^{q} \omega_j(q) \left[\sum_{i=j+1}^{N} (x_i - \bar{x}_N)(x_{i-j} - \bar{x}_N) \right] \right)^{1/2} , \tag{58}$$

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

$$Q_{N,q} = \frac{1}{S_q} \left\{ \max_{0 \le i \le N} \sum_{j=1}^{i} (x_j - \bar{x}_N) - \min_{0 \le i \le N} \sum_{j=1}^{i} (x_j - \bar{x}_N) \right\} . \tag{59}$$

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

$$W = \max_{0 \le r \le 1} V(r) - \min_{0 \le t \le 1} V(r) , \tag{60}$$

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

$$P(W \le x) = 1 + 2 \sum_{j=1}^{\infty} (1 - 4x^2 j^2) e^{-2x^2 j^2} . \tag{61}$$

Lo gave the critical values of x for hypothesis testing at sixteen significance levels using Eq. 61, 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.

5.3.2. *GPH test*

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

$$f(\omega) = \left[2\sin(\omega/2)\right]^{-2d} f_u(\omega), \tag{62}$$

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 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

$$\ln f(\omega_j) = \ln f_u(\omega_j) - d\ln\left[4\sin^2\left(\omega_j/2\right)\right], \tag{63}$$

which may be written in the alternative form

$$\ln f(\omega_j) = \ln f_u(0) - d\ln\left[4\sin^2(\omega_j/2)\right] + \ln\left[f_u(\omega_j)/f_u(0)\right]. \tag{63}$$

The fractional difference parameter d can be estimated by the regression equations constructed from Eq. 63. 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. 63 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:

$$\hat{d} \sim N\left(d, \frac{\pi^2}{6\sum_{j=1}^{g(n)}(U_j - \overline{U})^2}\right), \tag{64}$$

where

$$U_j = \ln[4\sin^2(\omega_j/2)] \tag{65}$$

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

$$t_{d=0} = \hat{d} \cdot \left(\frac{\pi^2}{6\sum_{j=1}^{g(n)} (U_j - \overline{U})^2} \right)^{-1/2} \tag{66}$$

has a limiting standard normal distribution.

5.3.3. *Maximum likelihood estimation of fractional differencing parameter d*

Let the observation $X = (x_1, \ldots, x_n)^t$ be the ARFIMA(p,d,q) process defined by

$$\phi(B)(1-B)^d (x_t - \mu) = \theta(B)\varepsilon_t , \tag{67}$$

where B is the backshift operator; $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^P$ 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 (Eq. 67) is given by

$$\log L(\mu, \eta, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2} X^t \Sigma^{-1} X , \tag{68}$$

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.

5.3.4. *Monte Carlo simulation results for long-memory detection*

Before applying the Lo's test, GPH test and S-MLE method to the streamflow processes, 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 $(1-\phi B)(1-B)^d x_t = \varepsilon_t$, where $\{\varepsilon_t\}$ are i.i.d standard normal, B is the backshift operator. For the AR models, large autoregressive coefficients, i.e., $\phi = 0.5, 0.8, 0.9, 0.95, 0.99$, because these are the cases commonly seen 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 with size 500, 1000, 3000, 10000 and 20000, 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 (Insightful Corporation, 2001).

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 have shown (Lo, 1991; Teverovsky et al., 1999) that, for any of these series, the probability of accepting the null hypothesis varied significantly with q. In general, for the larger sample lengths, the larger the q, the less likely was the null hypothesis to be rejected. One appealing data-driven formula (Andrews and Donald WK, 1991) for choosing q based on the assumption that the true model is an AR(1) model is given by

$$q = \left[\left(\frac{3n}{2} \right)^{1/3} \left(\frac{2\hat{\rho}}{1-\hat{\rho}^2} \right)^{2/3} \right], \tag{69}$$

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 indicate 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:

$$q = \left[\left(\frac{n}{10} \right)^{1/4} \left(\frac{2\hat{\rho}}{1-\hat{\rho}^2} \right)^{2/3} \right]. \tag{70}$$

where $\hat{\rho}$ is the autoregressive function 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. 63. 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 the 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 t-statistic is not contained in the interval [-1,960, 1.960].

When estimating the parameter d with the *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 the 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 20000 with Lo's test, GPH test and the *S-MLE* estimates of d are reported in Table 9. For Lo's test, we list the average values of the lags chosen with the data-driven formula (Eq. 70), their standard deviations (denoted as SD of lag), and the number of acceptance of the null hypothesis for 500 simulations. For GPH test, we list the average of the estimates of d, their standard deviations (denoted as SD of lag), and the number of acceptance of the null hypothesis for 500 simulations. For the S-MLE method, we give the averages and standard deviations (SD) of the estimates of d. According to the results with simulated AR and ARFIMA processes, shown in Table 9, we have the following findings:

(1) For AR 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 as high as 0.95 or above, even for the size of 20000 points. In contrast, the probability of committing Type I error with the Lo's *R/S* test still considerably lower even for AR processes with a ϕ of as high as 0.99. But it seems that the lag chosen with formula (Eq. 8) tends to be too small for series of big size, whereas a little bit too large for series of small size for AR processes with large values of ϕ.

(2) For ARFIMA processes, the GPH technique yields downwardly biased estimates of d when an AR term of low autoregressive coefficient value (e.g., ≤ 0.5) is present, whereas yields upwardly biased estimates of d when an AR term of high autoregressive coefficient value (e.g., = 0.9) is present. This seems to be in contradiction with the results of Sowell (1992), who showed that, when the sample length is small, the GPH technique yieldsupwardly 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 component. For cases where the data size is over 10000, the percentage 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 a autoregressive component of strong autocorrelation, the Lo's performs far less powerful than the GPH test.

(3) 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 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 Kendziorskia et al (1999), who showed that S-MLE provided unbiased estimates of d for ARFIMA(0,d,0) processes of length 2^{11} (2048) or greater.

(4) Data size has a significant impact on the power of all the three methods. 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 $\phi = 0.9$ 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 $\phi = 0.9$ 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 $\phi = 0.9$. 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. (63) more carefully (Giraitis et al., 1997; Hurvich and Deo, 1999). But the effectiveness of these methods is yet to be confirmed.

For example, as ϕ increases, the estimates of d using the number of frequencies $g(n)$ selected by the plug-in method proposed by Hurvich and Deo (1999) are much more positively biased than simply using $g(n) = n^{1/2}$.

(5) 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 recommend 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, you are still not sure which one gives the right answer.

TABLE 9. Long-memory detection 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) ar = .5	500	2.8	0.5	464	-0.0167	0.1302	495	0.0149	0.0350
	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
	10000	6.1	0.2	455	-0.0119	0.0607	490	0.0093	0.0132
	20000	7.8	0.4	469	-0.0078	0.0479	488	0.0057	0.0100
AR(1) ar = .8	500	6.7	0.8	428	0.1220	0.1388	470	0.0269	0.0669
	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
	10000	14.7	0.5	441	-0.0016	0.0605	490	0.0114	0.0207
	20000	17.6	0.5	454	-0.0036	0.0511	483	0.0079	0.0149
AR(1) ar = .9	500	11.3	1.6	431	0.3252	0.1342	268	0.0290	0.0566
	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
	10000	24.6	0.8	441	0.0273	0.0600	483	0.0132	0.0236
	20000	29.4	0.7	457	0.0107	0.0500	489	0.0081	0.0150
AR(1) ar = .95	500	18.7	3.6	451	0.5739	0.1395	24	0.0302	0.0497
	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
	10000	40.3	1.8	416	0.1201	0.0601	300	0.0117	0.0284
	20000	47.9	1.6	416	0.0665	0.0475	409	0.0065	0.0160

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) ar = .99	500	52.9	20.3	494	0.9122	0.1617	0	0.0482	0.0674
	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
	10000	119.7	11.9	389	0.5555	0.0583	0	0.0093	0.0211
	20000	142.4	9.5	380	0.4478	0.0477	0	0.0068	0.0148
ARFIMA d=0.3	500	2.2	0.5	129	0.2587	0.1360	353	0.2144	0.1100
	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
	10000	5.2	0.4	0	0.2884	0.0572	2	0.3043	0.0201
	20000	6.3	0.5	0	0.2900	0.0470	0	0.3072	0.0162
ARFIMA ar=0.5 d=0.3	500	7.1	1.4	255	0.2729	0.1402	333	0.1728	0.1346
	1000	8.6	1.3	139	0.2783	0.1130	233	0.2126	0.1165
	3000	11.4	1.2	63	0.2878	0.0919	83	0.2849	0.0675
	10000	15.6	1.0	8	0.2934	0.0604	4	0.3049	0.0363
	20000	18.6	0.9	5	0.2955	0.0493	0	0.3102	0.0202
ARFIMA ar=0.9 d=0.3	500	41.1	12.2	493	0.6375	0.1513	16	0.1683	0.1451
	1000	49.4	11.6	478	0.5213	0.1123	6	0.2035	0.1333
	3000	65.4	11.2	345	0.3964	0.0881	5	0.2397	0.1243
	10000	89.4	9.2	155	0.3316	0.0627	2	0.3103	0.0678
	20000	106.6	8.3	78	0.3145	0.0512	0	0.3281	0.0501
ARFIMA d=0.45	500	7.0	4.0	130	0.4077	0.1506	157	0.3092	0.1572
	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
	10000	15.4	6.0	0	0.4373	0.0613	0	0.4589	0.0173
	20000	18.6	7.0	0	0.4371	0.0489	0	0.4676	0.0164
ARFIMA ar=0.5 d=0.45	500	19.1	10.1	346	0.4331	0.1515	133	0.2355	0.1628
	1000	22.9	10.6	204	0.4385	0.1164	33	0.3328	0.1311
	3000	31.0	12.2	66	0.4404	0.0893	3	0.4226	0.0668
	10000	42.4	14.6	11	0.4429	0.0635	0	0.4608	0.0228
	20000	50.2	16.2	4	0.4459	0.0507	0	0.4718	0.0170
ARFIMA ar=0.9 d=0.45	500	135.0	78.5	493	0.7956	0.1394	2	0.1306	0.1757
	1000	163.4	90.2	495	0.6733	0.1172	1	0.1712	0.1828
	3000	222.9	116.2	472	0.5539	0.0878	0	0.3128	0.1665
	10000	299.5	138.7	273	0.4856	0.0599	0	0.4464	0.0577
	20000	361.8	158.0	140	0.4666	0.0491	0	0.4748	0.0226

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

On the basis of the above findings, to obtain reliable test results on detecting the presence of long-memory, we have two suggestions: First, increase the size of test data, 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 increases, but notice that the estimate with *S-MLE* would be biased upwardly when the data size is above 10^4; Second, use the detection results in combination with each other, as have been suggested by Teverovsky et al. (1999). Here we consider the combined use of Lo's test, GPH-test and *S-MLE*. As shown in Table 9, 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 suggests, 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.

We especially recommend the combined use of GPH test and *S-MLE* to detect the existence of long-memory, and the most appropriate date size for estimating *d* with *S-MLE* is slightly less than 10^4.

5.3.5. *Long-memory test for streamflow processes*

According to what we found with the Monte Carlo simulations, we use the Lo's *R*/*S* test, GPH test and *S-MLE* method jointly to detect the existence of long-memory in streamflow processes in this study. For Lo's modified *R*/*S* test, we adopt the data-driven formula (Eq. 70) to choose the lag q. For GPH test, we choose the number of frequencies $g(n) = n^{-0.5}$ as suggested by Geweke and Porter-Hudak (1983). The null hypothesis of no long-term dependence is rejected if $Q_{N,q}$ in Lo's test is not contained in the interval [0.809, 1.862] (Lo, 1991), or if *t*-statistic in GPH test is not contained in the interval [-1,960, 1.960]. With the *S-MLE* method, we assume that the processes are ARFIMA(p,d,0) processes, and the order p of AR component is determined by using the AIC criterion (Akaike, 1973).

All the streamflow series are log-transformed and deseasonalized. In addition, the streamflow series of the Yellow River at TG are detrended first. The test results for all streamflow series are listed in Table 10. The results show that, the intensity of long-memory decreases with the increase of timescale according to the results of all the three methods. All daily flow series exhibit strong long-memory, because the presence of long-memory is confirmed by all the three methods for 4 cases, and in another two cases (Danube and Rhine), it is confirmed by the GPH test and S-MLE method.

The presence of long-memory in 1/3-monthly series is confirmed in three cases (Yellow at TNH, Yellow at TG, and Umpqua), rejected in two cases (Danube and Rhine), and not conclusive in one case (Ocmulgee). For monthly series, the existence long-memory is rejected by both the GPH test and S-MLE method for four cases because the hypothesis of no long-memory is accepted by the GPH test, and S-MLE give a estimate of *d* less than 0.2. But the monthly series of Yellow River at TG and Umpqua may exhibit long-memory.

TABLE 10. Detecting the existence in streamflow series with Lo's modified *R*/*S* test, GPH test and S-MLE method

River (station)	Timescale	Lo's test		GPH test		S-MLE
		Lag	Statistic	statistic	d	d
Yellow (TNH)	Daily	94	2.5111*	7.7853 *	0.4720	0.4922
	1/3-monthly	11	2.2910 *	3.277 0*	0.3854	0.4518
	Monthly	5	1.9644 *	1.4233	0.2357	0.0000
Yellow (TG)	Daily	39	3.06975 *	5.4234 *	0.3422	0.3561
	1/3-monthly	7	2.4679 *	2.9501 *	0.3636	0.3194
	Monthly	3	2.1437 *	1.3756	0.2415	0.3400
Danube (Achleiten)	Daily	63	1.7273	5.4564 *	0.2742	0.3865
	1/3-monthly	8	1.5512	0.8792	0.0848	0.0001
	Monthly	4	1.3044	-0.1307	-0.0176	0.0000
Rhine (Lobith)	Daily	170	1.5288	6.5402 *	0.3229	0.4167
	1/3-monthly	11	1.6853	1.1944	0.1129	0.0000
	Monthly	5	1.4853	0.1528	0.0202	0.0000
Ocmulgee (Macon)	Daily	31	2.7826 *	7.1878 *	0.3821	0.4663
	1/3-monthly	6	2.0852 *	1.8812	0.1916	0.1956
	Monthly	4	1.6260	1.4253	0.2039	0.1368
Umpqua (Elkton)	Daily	58	3.1110 *	5.6400 *	0.2785	0.4445
	1/3-monthly	8	2.6341 *	2.3899 *	0.2259	0.2189
	Monthly	4	2.2376 *	2.5076 *	0.3313	0.1799

Note: An asterisk indicates the rejection of the null hypothesis of no long-memory at the 0.05 significance level.

A special concern here is the value of d for the daily streamflow process of the Yellow River at TNH, because we will model and forecast this streamflow process later. The estimates of d given by the GPH test and S-MLE are 0.472 and 0.4922, respectively. In addition, with S-MLE, we know that the process also has an AR component of high autoregressive coefficients. The size of the series is 16437 points. As we know from the results for simulation ARFIMA series, for a series of this size and strong autocorrelations, both GPH method and S-MLE method give positively biased estimates of *d*. Therefore, taking the results from the heuristic methods into account, we will consider a *d* of less than 0.4 when modeling the daily process at TNH.

6. Conclusions

There is no obvious trend in the average annual flow process of the upper Yellow River at TNH over the period 1956 to 2000, whereas the discharges recorded at a downstream station TG exhibit significantly declining trend. Fu et al. (2004) investigated the trend of annual runoffs at another three stations along the mainstream of the Yellow River. Put together the results, we find that the lower the reaches of the Yellow River, the more significant the declining trend. However, generally, there is no significant decline in the precipitation processes in the Yellow River basin (Fu et al., 2004). The phenomenon that the lower the reaches of the Yellow River, the more significant the downward trend is a clear indication of anthropogenic influence, because the lower the reaches, the more human intervention the river would suffer, including the expansion of irrigation areas, the regulation of thousands of reservoirs in both the main channel and tributaries, and the increase of water consumption with the fast growing industry and population. Although the impacts of global warming on water supply are widely concerned, in the case of the Yellow River basin, the impacts of warming on the river flow processes of the Yellow River seem far less significant than anthropogenic influences.

The Augmented Dickey-Fuller (ADF) unit root test (Dickey and Fuller, 1979; Said and Dickey, 1984) and KPSS test (Kwiatkowski et al., 1992) are introduced to test for the nonstationarity in streamflow time series. It is shown that the smaller the timescale of the streamflow process, the more likely it tends to be nonstationary.

Seasonal variations in the autocorrelation structures are present in all the deseasonalized daily, 1/3-monthly and monthly streamflow processes, albeit that such seasonal variation is less obvious for the streamflow of the Danube and the Ocmulgee. This indicates that, even after the deseasonalization procedure, the seasonality still shows itself, not in the mean and variance, but in the autocorrelation structure.

The investigation of the long-memory phenomenon in streamflow processes at different timescales shows that, with the increase of timescale, the intensity of long-memory decreases. According to 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 version 6 (*S-MLE*), all daily flow series exhibit strong long-memory. Out of six 1/3-monthly series, three series (Yellow River at TNH and TG, and Umpqua) exhibit long-memory, whereas two 1/3-monthly series (Danube and Rhine) seem to be short-memory series. Only one monthly flow series (Umpqua) may exhibit long-memory.

Comparing the stationarity test results and the long-memory test results, we find that these two types of test are more or less linked, not only in that the test results have similar timescale pattern, but also in that there is a general tendency that the stronger the nonstationarity, the more intense the long-memory. In fact, there are some attempts to use KPSS stationarity test to test for the existence of long-memory (e.g., Lee and Schmidt, 1996). It is worthwhile to further the investigation on the issue of the relationship between nonstationarity and the long-memory in the future.

References

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

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. *In Proceedings of the 2nd International Symposium on Information Theory*, B.N. Petrov and F. Csaki (eds.). Akademia Kiado, Budapest, 267-281.

Andrews, Donald, W.K, 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59, 817-858.

Bartlett, M.S., 1950. Periodogram analysis and continuous spectra. *Biometrika*, 37, 1-16.

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

Bloomfield, P., 1992. Trend in global temperature. *Climatic Change*, 21, 1-16.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. *Journal of the Royal Statistical Society,* Series B, 26:211-252.

Box, G.E.P., Jenkins, G.M., 1976. *Time Series Analysis: Forecasting and Control* (2nd edition). San Francisco: Holden-Day.

Burn, D.H., Hag Elnur, M.A., 2002. Detection of hydrological trend and variability. *J. Hydrol.* 255, 107–122.

Chen, H.L., Rao, A.R., 2002. Testing hydrological time series for stastionarity. *J Hydrol. Eng.*, 7(2), 129-136.

Dickey, D.A., Fuller,W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* 74, 423–431.

Dietz, E.J., Killeen., T.J., 1981. A Nonparametric Multivariate Test for Monotonic Trend With Pharmaceutical Applications. *J. Am. Stat. Assoc.* 76(373), 169-174.

Douglas, E. M., Vogel, R. M., and Kroll, C. N., 2000. Trends in floods and low flows in the United States: Impact of spatial correlation, *J. Hydrol.*, 240, 90–105.

Fu, G., Chen, S., Liu, C., Shepard, D., 2004. Hydro-climatic trends of the Yellow river basin for the last 50 years. *Climatic Change* 65, 149–178.

Fuller, W.A., 1976. *Introduction to Statistical Time Series*. New York: Wiley.
Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. *J. Time Ser. Anal*., 4:221-238.
Gimeno, R., Manchado, B., Minguez, R., 1999. Stationarity tests for financial time series. *Physica* A 269, 72-78.
Giraitis, L., Robinson, P.M., Samarov, A., 1997. Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. *J. Time Ser. Anal*., 18 49-60.
Granger, C. W. J., Joyeux, R., 1980. An Introduction to Long-Memory Time Series Models and Fractional Differencing, *J. Time Ser. Anal.,* 1, 15-29.
Hamilton, J.D., 1994. *Time Series Analysis*, Princeton University Press, Princeton.
Haslett, J., Raftery, A. E., 1989. Space-time modeling with long-memory dependence: assessing Ireland's wind power resource (with discussion). *Applied Statistics* 38, 1-50.
Helsel, D.R., Hirsch, R.M., 1992. *Statistical Methods in Water Resources*. Amsterdam: Elsevier.
Hirsch, R.M., Slack, J.R., 1984. A Nonparametric Trend test for seasonal data with serial dependence. *Water Resources Research*, 20(6), 727-732.
Hirsch, R.M., Slack, J.R., Smith, R.A., 1982. Techniques of trend analysis for monthly water quality data. *Water Resources Research*, 18(1), 107-121.
Hosking, J.R.M., 1981. Fractional differencing, *Biometrika,* 68, 165-176.
Hurst, H.E., 1951. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers,* 116:770-808.
Hurvich, C.M., Bellrao, K., 1993. Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series. *J. Time Ser. Anal*., 14, 455-472.
Hurvich, C.M., Deo, R.S., 1999. 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.
Hussain, S., Elbergali, A., 1999. Fractional order estimation and testing, application to Swedish temperature data. *Environmetrics*, 10, 339-349.
Insightful Corpration, 2001. Chapter 27 " Analyzing time series and signals" of the S-PLUS 6 for Windows Guide to Statistics, Volume 2, *Insightful Corporation*, Seattle, WA.
Jain, S., Lall, U., 2001. Floods in a changing climate: Does the past represent the future? *Water Resources Research*, 37(12), 3193-3205.
Karagiannis, T., Molle, M., Faloutsos, M., 2004. Long-range dependence: ten years of internet traffic modeling. *IEEE Internet Computing*, 8(5): 57-64.
Kendall, M.G., 1938. A new measure of rank correlation. *Biometrika*, 30, 81-93.
Kendall, M.G., 1975. *Rank Correlation Methods*, Charles Griffin, London.
Kendziorskia, C.M., Bassingthwaighteb, J.B., Tonellato, P.J., 1999. Evaluating maximum likelihood estimation methods to determine the Hurst coefficient. *Physica* A, 273, 439-451.
Kundzewicz, Z.W., Graczyk, D., Maurer, T., Przymusinska, I., Radziejewski, M., Svensson, C., Szwed, M., 2004. Detection of change in world-wide hydrological time series of maximum annual flow. *GRDC Report 32*, GRDC, Koblenz, Germany.
Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *J. Econometrics* 54, 159–178.

Lee, D., Schmidt, P., 1996. On the power of the KPSS test of stationarity against fractionally-integrated alternatives. *J. Econometrics*, 73(1), 285-302.

Lettenmaier, D.P., Wood, E.F. Chapter 26 Hydrological forecasting. In: D.R. Maidment (ed.), *Handbook of Hydrology*. McGraw-Hill, New York, 199326.1-26.30.

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

Mandelbrot, B.B., Wallis, J.R., 1969a. Computer experiments with fractional Gaussian noises. Part 1, 2 and 3. *Water Resources Research* 5, 228–267.

Mandelbrot, B.B., Wallis, J.R., 1969b. Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence. *Water Resources Research,* 5, 967-988.

Mann, H.B., 1945. Nonparametric test against trend. *Econometrica*, 13, 245-259.

McLeod, A.I., 1994. Diagnostic Checking of Periodic Autoregression. *J. Time Ser. Anal.,* 15(2), 221–233.

McLeod, A.I., 1975. Derivation of the theoretical autocovariance function of autoregressive-moving average time series. *Applied Statistics*, 24(2), 255-256.

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

Mitosek, H.T., 2000. On stochastic properties of daily river flow processes. *Journal of Hydrology*, 228, 188-205.

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

Montanari, A., Rosso, R., Taqqu, M. S., 2000. A seasonal fractional ARIMA model applied to the Nile River monthly at Aswan. *Water Resources Research*, 36:1249-1259.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

Ooms, M., Franses, P.H., 2001. A seasonal periodic long memory model for monthly river flows. *Environmental Modelling & Software*, 16, 559-569.

Priestley, M.B., 1988. *Non-linear and Non-stationary Time Series Analysis*. London: Academic Press.

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

Said, S.E., Dickey, D., 1984. Testing for unit roots in autoregressive moving-average models with unknown order. *Biometrika,* 71, 599-607.

Schwert, G.W., 1989. Tests for unit roots: A Monte Carlo investigation. *J. Business Eco. Stat*. 7,147-159.

Shin, Y., Schmidt, P., 1992. The KPSS stationarity test as a unit root test. *Economics Letters*, 38, 387-392.

Sowell, F., 1992. Modeling Long-Run Behavior with the Fractional ARIMA Model, *Journal of Monetary Economics*, 29, 277-302.

Van Belle, G., Hughes, J. P., 1984. Nonparametric tests for trend in water quality. *Water Resources Research*, 20(1), 127-136.

Van Gelder, P.H.A.J.M., 1999. Statistical Methods for the Risk-Based Design of Civil Structures, *Communications on Hydraulic and Geotechnical Engineering*, ISSN:0169-6548 00-1, 248pages.

Vecchia, A.V., Ballerini, R., 1991. Testing for periodic autocorrelations in seasonal time series data, *Biometrika*, 78, 53-63.

Vecchia, A.V., 1985. Maximum Likelihood Estimation for periodic autoregressive moving average models. *Technometrics,* 27, 375-384.

Von Storch, H., 1995. Misuses of statistical analysis in climate research, *In: Analysis of Climate Variability: Applications of Statistical Techniques*, edited by H. V. Storch and A. Navarra, Springer-Verlag, New York, pp. 11– 26.

Zhang, X., Harvey, K. D., Hogg, W. D., Yuzyk, T. R., 2001. Trends in Canadian streamflow, *Water Resources Research*, 37(4), 987–998.

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

APPENDIX 1 HYPOTHESIS TESTING

Setting up and testing hypotheses is an essential part of statistical inference. In order to carry out a statistical test, it is necessary to define the null and alternative hypotheses; which describe what the test is investigating. In each problem considered, the question of interest is simplified into two competing claims / hypotheses between which we have a choice; the null hypothesis, denoted H_0 (e.g., there is no significant change in the annual maximum flow series), against the alternative hypothesis, denoted H_1 (e.g., the annual maximum flow is changing over time). In carrying out a statistical test one starts by assuming that the null hypothesis is true, and then checks whether the observed data are consistent with this hypothesis. The null hypothesis is rejected if the data are not consistent with H_0.

To compare between the null and the alternative hypotheses a test statistic is selected and then its significance is evaluated, based on the available evidence. A test statistic is a quantity calculated from our sample of data subject to testing. Its value is used to decide whether or not the null hypothesis should be rejected in our hypothesis test. The choice of a test statistic will depend on the assumed probability model and the hypotheses under question.

The significance level of a statistical hypothesis test is a fixed probability of wrongly rejecting the null hypothesis H_0, if it is in fact true. It is the probability of a type I error. Usually, the significance level is chosen to be 0.05.

The probability value (p-value) of a statistical hypothesis test is the probability of getting a value of the test statistic as extreme as or more extreme than that observed by chance alone, if the null hypothesis H_0 is true. It is equal to the significance level of the test for which we would only just reject the null hypothesis. Small p-values suggest that the null hypothesis is unlikely to be true. The smaller it is, the more convincing is the rejection of the null hypothesis.

The diagram below represents four outcomes of the decisions we make, in terms of whether or not the null is true, and whether we reject the null or not.

Decision	Truth of Null	
	True	**Not True**
Reject Null	**TYPE I**	**POWER**
FTR Null	**CORRECT**	**TYPE II**

As you see, FTR (failed to reject) the null when the null is true is a correct decision. However, we're usually interested in trying to find true differences, and therefore look to reject null hypotheses. Rejecting the null when it is really not true is a correct decision as well. More specifically, the probability a test has to do this is referred to as *power*. Power may be defined as the probability of correctly rejecting the null hypothesis. In other words, it is the probability of rejecting the null hypothesis given that the null is incorrect. Some people also refer to power as *precision* or *sensitivity.*

APPENDIX 2 STATIONARITY AND PERIODIC STATIONARITY

Let $\{x_t\}$, $t = 1, \ldots, N$, be N consecutive observations of a seasonal time series with seasonal period s. For simplicity, assume that $N/s = n$ is an integer. In other words, there are n full years of data available. The time index parameter t may be written $t = t(r\text{-}m) = (r\text{-}1)s + m$, where $r = 1, \ldots, n$ and $m = 1, \ldots, s$. In the case of monthly data s = 12 and r and m denote the year and month.

If

$$\mu_m = E(z_{t(r,m)})$$

and

$$\gamma_{l,m} = \mathrm{cov}(z_{t(r,m)}, z_{t(r,m)-l})$$

exist and depend only on l and m, z_t is said to be periodically correlated or periodic stationary (Gladysev, 1961). Note that the case where μ_m and $\gamma_{t,m}$ do not depend on m reduces to an ordinary covariance stationary time series.

A series $\{x_t\}$ is called stationary if, loosely speaking, its statistical properties do not change with time. More precisely, $\{x_t\}$ is said to be completely stationary if, for any integer k, the joint probability distribution of x_t, x_{t+1}, …, $x_{t+\mathrm{k}-1}$ is independent on the time index t (see e.g., Priestley, 1988, pp. 4-5).