Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Hydrology 277 (2003) 89–99

Journal of Hydrology
www.elsevier.com/locate/jhydrol

Linearity analysis on stationary segments of hydrologic time series

Huey-Long Chen[a,*], A. Ramachandra Rao[b,1]

[a]Department of Environmental Engineering, Lan-Yang Institute of Technology, Toucheng, Ilan 26141, Taiwan, ROC
[b]School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA

Received 25 January 2002; accepted 28 February 2003

Abstract

The rainfall-runoff process is widely perceived as being non-linear; however, the degree of non-linearity might not be significant in hydrologic time series. Evidence of non-linearity was reported in the past in the detrended time series of daily precipitation, but found not to be significant in annual series. The objective of this study is to detect non-linearity in monthly hydrologic time series by applying the Hinish tests for Gaussianity and linearity to selected stationary segments of four kinds of such series, namely, streamflow, temperature, precipitation and Palmer's drought severity index. The results indicate that all of the stationary segments of standardized monthly temperature and precipitation series are found to be either Gaussian or linear. Some of the standardized monthly streamflow and Palmer's drought severity index are found to be non-linear.

Keywords: Stationarity; Linearity; Gaussianity; Hydrologic time series; Hinich tests

1. Introduction

Linear models and Gaussian distributed variables are usually used in time series analysis mainly due to the convenience in studying relevant statistical properties. For models such as linear autoregressive (AR) and autoregressive-moving average (ARMA) models, procedures for model identification and parameter estimation have been well formalized based on Gaussianity and linearity (Box and Jenkins, 1976; Priestley, 1981; Hamilton, 1994). Both Gaussianity and linearity also play important roles in a recently developed signal processing technique—wavelet analysis (Chen and Rao, 2002; Donoho and Johnstone, 1994; Hu, 1994; Ogden and Richwine, 1996; Wang, 1995). However, non-linear mechanisms are often encountered in physical sciences. Non-Gaussian stationary time series may be generated as a result from a specific non-linear operation on a Gaussian input process. Therefore, non-linear modeling approaches have gained increasing attention from time series analysts (Subba Rao and Gabr, 1984; Nikias and Petropulu, 1993; Priestley, 1988).

Applications of linear models to hydrologic time series have been widely and thoroughly discussed in the literature (Kashyap and Rao, 1976; Salas et al., 1985). Meanwhile, investigations on non-linearity and application of non-linear models in hydrology have also received attention from researchers (Kember

* Corresponding author. Tel.: +886-3-977-1997x206.
E-mail addreses: hueylong@mail.fit.edu.tw (H.-L. Chen)
[1] Tel.: +1-765-494-2176; fax: +1-765-496-1988.

0022-1694/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0022-1694(03)00086-6

et al., 1993; Prasad, 1967; Rao and Yu, 1990; Rao et al., 1971; Rogers, 1980, 1982; Rogers and Zia, 1982; Tong, 1990). Although the rainfall-runoff process is widely perceived as being non-linear, the signatures of non-linearity are not all recognizable in hydrologic time series. By using Hinich's (1982) test, non-linearity is detected in daily hydrologic time series by Rao and Yu (1990), but not in annual series.

Monthly hydrologic time series is seasonal with a cycle of 12 months. A standardization procedure is often applied to monthly hydrologic time series (Hipel and McLeod, 1994). However, this standardization procedure does not assure stationarity in the transformed series (Salas, 1993). A segmentation algorithm to identify and partition non-stationary time series into stationary segments was applied by Chen and Rao (2002) to the standardized monthly series, and the results indicate that the majority of the investigated monthly streamflow and Palmer's drought severity index (Palmer, 1965) series are identified as non-stationary and the majority of the investigated monthly precipitation series are stationary.

The objective of this study is to further investigate the signature of non-linearity in monthly hydrologic time series. Hinich's (1982) Gaussianity and linearity tests are applied only to the stationary segments of hydrologic time series discussed in Chen and Rao (2002).

2. Tests for Gaussianity and linearity

The second-order cumulants (sample autocovariances) and spectra do not contain enough information to characterize non-linear or non-Gaussian time series (Subba Rao and Indukumar, 1996). Standard whiteness tests cannot detect non-linear serial dependence in the residuals obtained from fitting a linear model, because these tests mostly rely on the second-order cumulants. Tests based on the sample estimate of the third-order spectrum (bispectrum) for Gaussianity and linearity were proposed by Subba Rao and Gabr (1980). Hinich (1982) modified Subba Rao and Gabr's (1980) approach by using the asymptotic properties of the sample bispectrum. It is shown in the simulations of Ashley et al. (1986) that Hinich's (1982) test has substantial power for many non-linear models. Both their empirical and theoretical results also show that the test is equally powerful in detecting non-linearity either in source or in residual series. In this section, the theoretical aspects of Hinich's (1982) tests are presented.

Suppose that a series $\{Y_t,\ t = 1, 2, \ldots, N\}$ is linear, that is,

$$Y_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \tag{1}$$

where $\{b_j\}$ are weights, and $\{\varepsilon_j\}$ are assumed to be independently identically distributed. The bispectrum of $\{Y_t\}$ at frequency pair (f_1, f_2) is given by

$$S_{3Y}(f_1, f_2) = \mu_3 B(f_1) B(f_2) B^*(f_1 + f_2), \tag{2}$$

where $\mu_3 = E[\varepsilon_t^3]$,

$$B(f) = \sum_{j=0}^{\infty} b_j \exp(-2\pi i f j), \qquad i = \sqrt{-1},$$

and $B^*(f)$ is its complex conjugate. $S_{3Y}(f_1, f_2)$ is a spatially periodic function of which the principle domain is the triangular set $\Omega = \{0 \le f_1 \le \frac{1}{2}, f_2 \le f_1, 2f_1 + f_2 \le 1\}$. A normalized bispectrum is defined as having the form of the squared bicoherence

$$\psi^2(f_1, f_2) = \frac{|S_{3Y}(f_1, f_2)|^2}{S_Y(f_1) S_Y(f_2) S_Y(f_1 + f_2)} = \frac{\mu_3^2}{\sigma_\varepsilon^6}, \tag{3}$$

where

$$S_Y(f) = \sigma_\varepsilon^2 |B(f)|^2,$$

is the spectrum of $\{Y_t\}$. The squared bicoherence is equivalent to the square of the skewness function of $\{Y_t\}$, which is constant if $\{Y_t\}$ is linear. If $\{Y_t\}$ is Gaussian distributed, the squared bicoherence is zero. An estimator of the bispectrum S_{3Y} is defined as follows. Let

$$F(j, k) = \frac{Y\left(\frac{j}{N}\right) Y\left(\frac{k}{N}\right) Y^*\left(\frac{j+k}{N}\right)}{N}, \tag{4}$$

where

$$Y\left(\frac{j}{N}\right) = \sum_{t=0}^{N-1} Y_t \exp\left(\frac{-2\pi i t j}{N}\right). \tag{5}$$

The principle domain of $F(j,k)$ is the triangular grid set $D = \{0 < j \le (N/2), 0 < k \le j, 2j + k \le N\}$. A consistent estimator at the frequency pair

$$\left(\frac{m-\frac{1}{2}}{N}, \frac{n-\frac{1}{2}}{N}\right),$$

is obtained by averaging $F(j,k)$ over M^2 adjacent frequency pairs in the domain as

$$\hat{S}_{3Y}(m,n) = \frac{\sum_{j=(m-1)M}^{mM-1} \sum_{k=(n-1)M}^{nM-1} F(j,k)}{M^2}, \tag{6}$$

where M is an integer greater than $(N)^{0.5}$. The center of the M^2 frequency pairs is defined by the lattice

$$L = \left\{ \frac{(2m-1)M}{2}, \frac{(2n-1)M}{2} : \right.$$
$$\left. m = 1, \ldots, n \quad \text{and } m \le \frac{N}{2M} - \frac{n}{2} + \frac{3}{4} \right\}.$$

It follows from Eq. (3) that an estimator of the bicoherency is defined as

$$\hat{\psi}_{m,n} = \frac{\hat{S}_{3Y}(m,n)}{\left(\frac{N}{M^2}\right)^{1/2} \sqrt{\hat{S}_Y(g_m)\hat{S}_Y(g_n)\hat{S}_Y(g_{m+n})}}, \tag{7}$$

where

$$g_j = \frac{(2j-1)M}{2N},$$

and $\hat{S}_Y$ is a smoothed estimator of the periodogram. $\hat{\psi}_{m,n}$ is complex Gaussian distributed with unit variance. Under the null hypothesis of Gaussianity, the bicoherence of $\{Y_t\}$ is zero:

$\mathbf{H_{0G}}$: $\psi^2(f_1,f_2) = 0$.

A test statistic is defined as

$$T_\psi = 2\sum_m \sum_n |\hat{\psi}_{m,n}|^2, \tag{8}$$

is asymptotically chi-square distributed with $2P$ degrees of freedom, where P denotes the number of frequency pairs

$$\left(\frac{m-\frac{1}{2}}{N}, \frac{n-\frac{1}{2}}{N}\right),$$

with the entire lattice square within the principle domain. Consequently, the decision rule is

$$\begin{cases} T_\psi < \chi^2_\alpha(2P), & \text{accept } \mathbf{H_{0G}} \\ T_\psi \ge \chi^2_\alpha(2P), & \text{do not accept } \mathbf{H_{0G}} \end{cases}.$$

Under the null hypothesis of linearity, the squared bicoherence of $\{Y_t\}$ is constant. If $\{Y_t\}$ is Gaussian (i.e. the normalized bispectrum is zero), it cannot be concluded whether or not the process is linear based on the bispectrum alone. Therefore, the hypothesis of linearity is tested only if $\{Y_t\}$ is deemed to be non-Gaussian. Assuming $\{Y_t\}$ is non-Gaussian, the null hypothesis of linearity is

$\mathbf{H_{0L}}$: $\psi^2(f_1,f_2)$

is constant. Under $\mathbf{H_{0L}}$, $2|\hat{\psi}_{m,n}|^2$ are asymptotically non-central chi-square distributed ($\chi^2(2, \lambda_{m,n})$) with two degrees of freedom and the noncentrality parameter has the form

$$\lambda_{m,n} = \frac{2M^2}{N} \frac{|S_{3Y}(f_1,f_2)|^2}{S_Y(f_1)S_Y(f_2)S_Y(f_1+f_2)}. \tag{9}$$

A consistent estimator of $\lambda_{m,n}$ is

$$\hat{\lambda}_0 = \frac{T_\psi}{P} - 2. \tag{10}$$

A robust test statistic based on the sample interquartile range, $\varepsilon_3 - \varepsilon_1$ (where ε_3 is third quartile and ε_1 is the first quartile), is asymptotically Gaussian distributed as $N(\xi_3 - \xi_1, \sigma_0^2)$, where

$$\sigma_0^2 = \frac{(3f^{-2}(\xi_1) - 2f^{-1}(\xi_1)f^{-1}(\xi_3) + 3f^{-2}(\xi_3))}{16P},$$

ξ_1 and ξ_3 are the true quartiles, and f is the density function of $\chi^2(2, \lambda_{m,n})$. The decision rule for this case is

$$\begin{cases} N_{1-(\alpha/2)}(0,1) < \dfrac{(\varepsilon_3-\varepsilon_1)-(\xi_3-\xi_1)}{\sigma_0} < N_{\alpha/2}(0,1), & \text{accept } \mathbf{H_{0L}}, \\ \dfrac{(\varepsilon_3-\varepsilon_1)-(\xi_3-\xi_1)}{\sigma_0} \le N_{1-(\alpha/2)}(0,1) \text{ or } \dfrac{(\varepsilon_3-\varepsilon_1)-(\xi_3-\xi_1)}{\sigma_0} \ge N_{\alpha/2}(0,1), & \text{do not accept } \mathbf{H_{0L}}. \end{cases}$$

The threshold is chosen corresponding to a given level of significance α.

Two stages of hypothesis testing are performed for testing Gaussianity and linearity. First, the hypothesis of Gaussianity ($\mathbf{H_{0G}}$) is tested. If $\mathbf{H_{0G}}$ holds, the assumption of Gaussianity is accepted. Since the bispectrum is zero when $\mathbf{H_{0G}}$ holds, linearity cannot be detected from the bispectrum. Next, if $\mathbf{H_{0G}}$ fails, the hypothesis of linearity ($\mathbf{H_{0L}}$) is tested. If $\mathbf{H_{0L}}$ holds, the assumption of linearity is accepted. The averaging parameter M in Eq. (6) is specified by rounding off $(N)^{0.51}$. Both hypotheses are tested at the significance level $\alpha = 0.05$.

3. Data used in this study

Four types of hydrologic and climate related monthly time series recorded in the midwestern United States are analyzed in this study, namely, streamflow, temperature, precipitation and Palmer's drought severity index (PDSI) series. PDSI has been the most commonly used drought indicator in the United States, which is a dimensionless index derived from measurements of precipitation, air temperature, and local soil moisture. Values of PDSI range from -4.0 (extreme drought) to 4.0 (extreme wet conditions). Details of the statistical characteristics of these series can be found in Bhattacharya (1996a,b,c). A standardization procedure is used to transform these monthly hydrologic time series to remove the seasonality in the mean and variance. The possible trend of each standardized series is removed by differencing. Both the standardized series and the first-order differences of the standardized series are tested for linearity. The standardization procedure adopted is given below:

1. The average value of the hydrologic variable Y_{ij} is computed for each particular month as

$$\bar{Y}_i = \frac{\sum_{j=1}^{T} Y_{ij}}{T} \qquad \text{for } i = 1,2,\ldots,12, \tag{11}$$

where i represents the month, j represents the year and T is the number of years in total.
2. The standard deviation of Y_{ij} is computed for each particular month by using

$$\hat{s}_i = \sqrt{\frac{\sum_{j=1}^{T}(Y_{ij}-\bar{Y}_i)^2}{T-1}} \qquad \text{for } i = 1,2,\ldots,12. \tag{12}$$

3. The standardized variable of Y_{ij} is defined as

$$X_t = \frac{Y_{ij}-\bar{Y}_i}{\hat{s}_i}, \tag{13}$$

where $t = 12(j-1)+i$.

The hydrologic and climate related series analyzed in this study are summarized as follows:

The monthly streamflow series at five stations (EarthInfo, 1993) are analyzed. The monthly streamflow is computed by summing each average daily flow and dividing by the number of days corresponding to that particular month. The statistical characteristics of each monthly streamflow series are summarized in Table 1. The skewness coefficient and the autocorrelation coefficient for the first lag (ACF(1)) are computed from the standardized series. It is shown that most of the standardized monthly streamflow series are highly positively skewed, and the values of the ACF(1) are significant.

The monthly temperature series at five stations (EarthInfo, 1993) are analyzed. The monthly

Table 1
Summary of the hydrologic time series used in this study

Streamflow at Station	Period of record	Mean (cfs)	Standard deviation (cfs)	Skewness coefficient	ACF(1)
Minnesota River at Clinton, IA	1874–1993	47 936	30 354	1.213	0.645
White River near Alpine, IN	1928–1992	560	568	2.176	0.389
Kalamazoo River at Fennville, MI	1929–1993	1472	657	0.920	0.597
Missouri River at Hermann, MO	1929–1993	77 177	52 031	1.531	0.628
Wisconsin River at Merrill, WI	1903–1991	2673	1513	0.963	0.614
Temperature at Station	**Period of record**	**Mean (°F)**	**Standard deviation (°F)**	**Skewness coefficient**	**ACF(1)**
Urbana, IL	1902–1992	61	19	−0.050	0.190
Aledo, IL	1901–1989	61	20	0.054	0.211
Ft. Wayne, IN	1948–1992	60	19	0.071	0.239
Evansville, IN	1950–1992	64	19	0.177	0.226
Minneapolis, MN	1891–1992	54	22	0.049	0.228
Precipitation at Station	**Period of record**	**Mean (in.)**	**Standard deviation (in.)**	**Skewness coefficient**	**ACF(1)**
Urbana, IL	1903–1992	3.12	1.96	1.009	0.037
Aledo, IL	1901–1989	2.87	2.10	0.945	0.065
Ft. Wayne, IN	1948–1992	3.02	1.65	1.010	0.063
Indianapolis, IN	1948–1992	3.29	1.87	1.004	0.043
Minneapolis, MN	1891–1992	2.28	1.90	1.197	0.056
PDSI at Station	**Period of record**	**Mean**	**Standard deviation**	**Skewness coefficient**	**ACF(1)**
Region 2, IL	1895–1993	0.06	2.12	−0.544	0.895
Region 8, IL	1895–1993	−0.16	2.15	−0.123	0.896
Region 1, IN	1895–1993	0.07	2.34	−0.307	0.840
Region 7, IN	1895–1993	−0.02	2.26	−0.207	0.890
Region 2, OH	1895–1993	−0.08	2.38	−0.454	0.904
Region 9, OH	1895–1993	−0.08	2.12	−0.144	0.871

temperature is computed as the sum of each maximum daily temperature divided by the number of days corresponding to that particular month. It is shown in Table 1 that most of the skewness coefficients of the standardized monthly temperature series are close to zero, and that the values of the ACF(1) are significant but less than those of the monthly streamflow series.

The monthly precipitation series at five stations (EarthInfo, 1993) are analyzed. The total monthly precipitation is computed by summing each daily precipitation in that particular month. Most of the standardized monthly precipitation series are highly positively skewed, and the values of the ACF(1) are insignificant (Table 1).

The monthly PDSI series analyzed in this study comprise six series from stations in Indiana, Illinois, and Ohio. These PDSI data were originally obtained from the United States National Climatic Data Center. It is shown in Table 1 that most of the values of the ACF(1) are highly significant.

4. Segmentation of stationary segments

Both the standardized series and the first-order differences of the standardized series were tested for non-stationarity and partitioned into stationary segments by Chen and Rao (2002). The stationary segments were obtained using a segmentation algorithm with three statistical tests (Chen (1999); de Souza and Thomson (1982); Tsay (1988)). The segmentation algorithm consists of three stages, the procedures for which can be found in Chen and Rao (2002). The preliminary ending segment boundary (t_e in Fig. 1) is detected in the first stage, the optimal ending boundary is determined in the second stage (t_{eo} in Fig. 1), and the optimal starting boundary is determined in the third stage (t_{so} in Fig. 1).

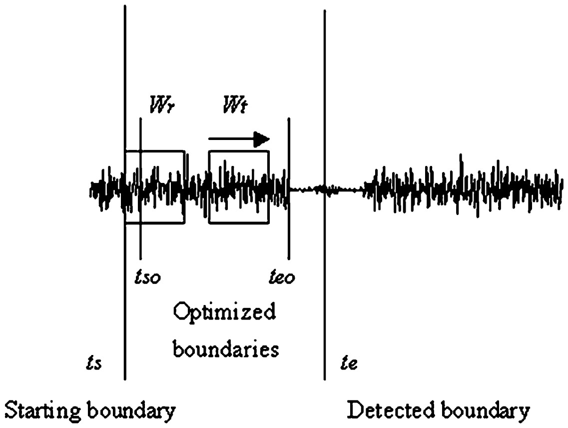

Fig. 1. A stationary segment between optimized boundaries t_{so} and t_{eo} determined by the segmentation algorithm.

In the first stage of the segmentation algorithm, a reference window (W_r in Fig. 1) is fixed at a starting point (t_s in Fig. 1) of the series. A test window (W_t in Fig. 1) next to the reference window slides along the series until the preliminary ending boundary t_e is detected at the observation where the test statistic computed from the two windows fails the null hypothesis of 'no change'. In the second stage, the optimal ending boundary t_{eo} is determined at the observation corresponding to the local critical value of the test statistic within the stopped test window in the first stage. In the third stage, the optimal starting boundary t_{so} is determined at the observation corresponding to the local critical value of the test statistic within the reference window used in the first stage. The segmentation algorithm has been used to partition the time series of temperature gradient measured in lakes into stationary segments (Chen et al., 2002). By fitting the spectrum of a stationary temperature gradient segment to the theoretical one, the turbulence kinetic energy dissipation rate can be estimated (Chen et al., 2001).

The tests of de Souza and Thomson (1982) and Tsay (1988) are based on AR models, which require specifying the AR order p. The test of Chen (1999) is based on wavelet analysis. According to Chen (1999), higher AR orders used in the tests of de Souza and Thomson (1982) and Tsay (1988) lead to higher rates of false identification; the wavelet-based test of Chen (1999) renders comparable detection results to these AR tests with the AR order p appropriately specified.

The numbers of stationary segments partitioned from differenced standardized and standardized hydrologic monthly series are given in Tables 2 and 3, respectively. Each of the tests (Chen (1999); de Souza and Thomson (1982); Tsay (1988)) is applied to both differenced standardized and standardized monthly series. The stationary segments partitioned from a monthly temperature series are given in Fig. 2. The stationary segments obtained corresponding to the three statistical tests (Chen (1999); de Souza and Thomson (1982); Tsay (1988)) are denoted as segmentations 1, 2 and 3, respectively, in Tables 2 and 3. The denominator in each cell indicates the number of stationary segments partitioned from each series. The segmentation results indicate that the majority of monthly streamflow and PDSI series are identified as being non-stationary, while the majority of monthly precipitation series are found to be stationary (Tables 2 and 3). According to Chen and Rao (2002), the change points in these hydrologic series, either differenced or not, are distributed in a similar pattern commonly observed during two periods, one between 1960 and 1970, and the other between 1930 and 1940. This suggests the possibility that a common non-stationary mechanism (e.g. periodicity) has an influence on these hydrologic series (Chen and Rao, 1998). Change points in 1960s and 1970s have been found in hydrologic related data by Bardossy and Caspary (1990), Hurrell (1995), Perreault et al. (1999) and Rodionov and Krovin (1992). Change points in the late 1930s have been reported by Quinn (1981).

5. Results of testing for Gaussianity and linearity using the Hinich tests

Since the standardized hydrologic time series are not always stationary, it does not make sense in testing for linearity on the entire length of each series. In the following analyses, the stationary segments are obtained from Chen and Rao (2002) using the segmentation algorithm with the above-mentioned three statistical tests denoted as segmentations 1, 2 and 3. The results of Gaussianity and linearity tests on stationary segments partitioned from differenced standardized and standardized monthly series are shown in Tables 2 and 3, respectively. For each series,

Table 2
Ratios of Gaussian, linear and non-linear segments partitioned from differenced standardized monthly series (G: Gaussian; L: Linear; NL: Non-Linear)

Streamflow at Station	Segmentation 1			Segmentation 2			Segmentation 3		
	G	L	NL	G	L	NL	G	L	NL
Minnesota River at Clinton, IA	1/3	2/3		1/3	2/3		3/4	1/4	
White River near Alpine, IN		1/1			1/1				1/1
Kalamazoo River at Fennville, MI	1/1				2/2			2/2	
Missouri River at Hermann, MO		2/2			2/2			2/2	
Wisconsin River at Merrill, WI	1/1	0/1		1/3	2/3		1/2	1/2	

Temperature at Station	Segmentation 1			Segmentation 2			Segmentation 3		
	G	L	NL	G	L	NL	G	L	NL
Urbana, IL	2/2			2/2			1/3	2/3	
Aledo, IL	1/1			2/2			1/1		
Ft. Wayne, IN	1/1			2/2			1/1		
Evansville, IN	1/1			2/2			1/1		
Minneapolis, MN	2/2			1/1			2/2		

Precipitation at Station	Segmentation 1			Segmentation 2			Segmentation 3		
	G	L	NL	G	L	NL	G	L	NL
Urbana, IL	1/1			1/1			1/1		
Aledo, IL	1/1			1/1			1/1		
Ft. Wayne, IN	1/1			1/1			2/2		
Indianapolis, IN	1/2	1/2		1/1			1/1		
Minneapolis, MN	1/1			2/3	1/3		2/4	2/4	

PDSI at Station	Segmentation 1			Segmentation 2			Segmentation 3		
	G	L	NL	G	L	NL	G	L	NL
Region 2, IL		1/1		2/3	1/3		1/2	1/2	
Region 8, IL	1/1			2/4	2/4		3/4	1/4	
Region 1, IN	1/1			3/4	1/4		2/3	1/3	
Region 7, IN	1/1			3/3			1/1		
Region 2, OH	1/1			3/4	1/4		2/2		
Region 9, OH	1/1			2/2			1/1		

Summary

Series Type	Total no.	Gaussian	Linear	Non-linear
Streamflow	30	9	20	1
Temperature	24	22	2	
Precipitation	22	18	4	
PDSI	39	30	9	

the results of testing for Gaussianity and linearity, using the Hinich (1982) tests, corresponding to a specific segmentation results, are given in three cells: Gaussian (G), linear (L) or non-linear (NL). The denominator in each cell indicates the number of stationary segments partitioned from each series, and the numerator represents the number of segments is identified as being Gaussian (G), linear (L) or non-

Table 3
Ratios of Gaussian, linear and non-linear segments partitioned from standardized monthly series (G: Gaussian; L: Linear; NL: Non-Linear)

Streamflow at Station	Segmentation 1			Segmentation 2			Segmentation 3		
	G	L	NL	G	L	NL	G	L	NL
Minnesota River at Clinton, IA		2/3	1/3		3/4	1/4		1/3	2/3
White River near Alpine, IN		1/1			1/1			2/2	
Kalamazoo River at Fennville, MI		1/1		1/2	1/2			2/2	
Missouri River at Hermann, MO		1/2	1/2		1/3	2/3		1/2	1/2
Wisconsin River at Merrill, WI	1/1				3/3			2/3	1/3
Temperature at station	Segmentation 1			Segmentation 2			Segmentation 3		
	G	L	NL	G	L	NL	G	L	NL
Urbana, IL	1/1			2/2			2/2		
Aledo, IL	1/1			2/2			1/1		
Ft. Wayne, IN	1/1			1/1			1/1		
Evansville, IN	1/1			2/2			1/1		
Minneapolis, MN	1/1			1/1			2/2		
Precipitation at station	Segmentation 1			Segmentation 2			Segmentation 3		
	G	L	NL	G	L	NL	G	L	NL
Urbana, IL	1/1			1/1			1/1		
Aledo, IL	1/1			1/1			1/1		
Ft. Wayne, IN	1/1			1/1			1/1		
Indianapolis, IN		2/2		1/1			1/1		
Minneapolis, MN	1/1			2/2			1/3	2/3	
PDSI at station	Segmentation 1			Segmentation 2			Segmentation 3		
	G	L	NL	G	L	NL	G	L	NL
Region 2, IL		1/1		1/2	1/2			2/3	1/3
Region 8, IL	1/2	1/2		1/3	1/3	1/3	1/4	2/4	1/4
Region 1, IN	1/1			1/4	3/4		1/4	3/4	
Region 7, IN	1/1			2/5	3/5		1/4	3/4	
Region 2, OH	2/2			1/3	2/3		1/4	2/4	1/4
Region 9, OH	1/1			2/4	2/4		2/3	1/3	

Summary

Series type	Total no.	Gaussian	Linear	Non-Linear
Streamflow	33	2	22	9
Temperature	20	20		
Precipitation	19	15	4	
PDSI	51	20	27	4

linear (NL). For example, the streamflow series of the Minnesota River at Clinton, IA (Table 2) is partitioned into three stationary segments, including one segment detected as being Gaussian (G), two as linear (L), and none as non-linear (NL). The results of testing the four types of hydrologic series (streamflow, temperature, precipitation and PDSI), using the Hinich (1982) tests, are discussed below.

According to the results of testing on the segments of the streamflow series, all but one of

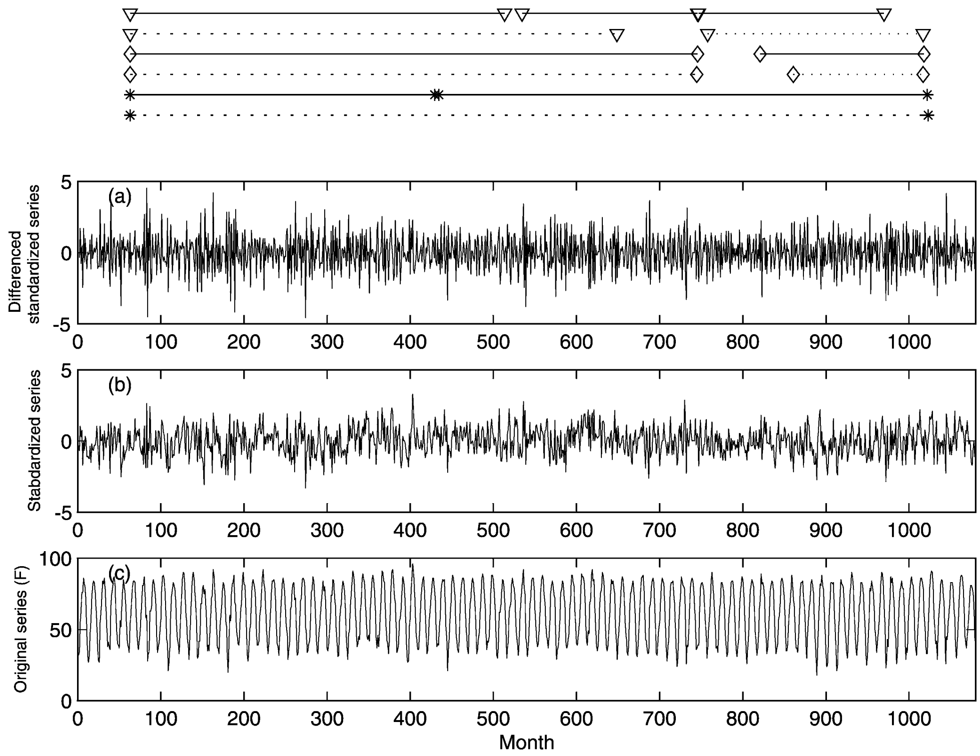

Fig. 2. Segmentation results from monthly temperature of Urbana, IL. Test used in the segmentation algorithm: ($*$) test 1; (◇) test 2; (△) test 3; (--) segments from standardized series; (—) segments from differenced standardized series.

the 30 segments of the differenced standardized series are identified as being either Gaussian or linear, the one segment being non-linear (Table 2); two segments of the 33 standardized series are found to be Gaussian, 22 segments are linear and 9 are non-linear (Table 3).

For the temperature series, 22 of the 24 segments of the differenced standardized series are Gaussian, two segments being linear (Table 2); all of the 20 segments of the standardized series are found to be Gaussian (Table 3). No segment is non-linear.

Eighteen of the 22 segments of the differenced standardized precipitation series are Gaussian, two being linear (Table 2); of the 19 standardized segments (Table 3), 15 are Gaussian and four are linear. No segment is non-linear.

For the PDSI series, all of the 39 segments of the differenced standardized series are either Gaussian or linear (Table 2); of the 51 segments of the standardized series, 47 are either Gaussian or linear, with four segments being detected as non-linear (Table 3).

With one exception, the segments of differenced standardized monthly series (Table 2) are found to be either Gaussian or linear, with one segment of streamflow series partitioned by the algorithm using test 3 being identified as non-linear (Table 2). The results of testing the standardized monthly series (without differencing; Table 3) indicate that nearly 90% of the segments are either Gaussian or linear, except for 13 segments from six streamflow and PDSI series being identified as non-linear (Table 3). It is also shown in the results that all of the segments of monthly temperature and precipitation series, either differenced standardized or standardized (without differencing), are identified as being either Gaussian or linear, i.e. non-linearity is not detected in the segments of monthly temperature and precipitation series, either differenced standardized or standardized.

6. Conclusions

The rainfall-runoff process is widely perceived as being non-linear; however, the degree of non-linearity might not be significant enough for detection in hydrologic time series. Evidence of non-linearity in the detrended daily precipitation series was reported by Rao and Yu (1990), but found not to be significant in the annual series. Since the monthly series have the time scale in between, the question of linearity naturally came to attention. For monthly hydrologic time series, a standardization procedure can be used to remove the monthly cycle (Hipel and McLeod, 1994). Chen and Rao (2002) tested the standardized monthly series for stationarity, and the results indicate that the majority of the investigated monthly streamflow and PDSI series are identified as non-stationary and the majority of the investigated monthly precipitation series are stationary. Given that the standardized monthly series is not stationary, Hinich's (1982) Gaussianity and linearity tests cannot be directly applied to.

In this study, Hinich's (1982) tests are performed on the stationary segments of monthly hydrologic time series partitioned by using a segmentation algorithm (Chen and Rao, 2002). In general, the results indicate that stationary segments of standardized monthly temperature and precipitation series are Gaussian or linear, while the conventional assumption of linear models, i.e. that of Gaussian-distributed variables, may not be valid for all standardized monthly streamflow and PDSI series.

References

Ashley, R.A., Patterson, D.M., Hinich, M.J., 1986. A diagnostic test for nonlinear serial dependence in time series fitting errors. J. Time Series Anal. 7 (3), 165–178.

Bardossy, A., Caspary, H., 1990. Detection of climate change in Europe by analysing European atmospheric circulation patterns from 1881 to 1989. Theor. Appl. Climatol. 42, 155–167.

Bhattacharya, D., 1996a. Rescaled range characteristics of monthly hydrologic series-I. Technical Report CE-EHE-96-3(a), Purdue University, West Lafayette

Bhattacharya, D., 1996b. Rescaled range characteristics of monthly hydrologic series-II. Technical Report CE-EHE-96-3(b), Purdue University, West Lafayette

Bhattacharya, D., 1996c. Empirical orthogonal functions-IV: monthly PDSI series. Technical Report CE-EHE-96-4(d), Purdue University, West Lafayette

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control, second ed, Holden-Day, Francisco.

Chen, H.-L., 1999. Stochastic characteristics of environmental and hydrologic time series. PhD Thesis, Purdue University, West Lafayette

Chen, H.-L., Rao, A.R., 1998. Periodicity in midwest US climatic data: the Hale cycle or the luni-solar signal? Stochast. Hydrol. Hydraul. 12 (4), 205–222.

Chen, H.-L., Rao, A.R., 2002. Testing hydrologic time series for stationarity. ASCE J. Hydrol. Engng 7 (2), 129–136.

Chen, H.-L., Hondzo, M., Rao, A.R., 2001. Estimation of turbulent kinetic energy dissipation. Water Resour. Res. 37 (6), 1761–1769.

Chen, H.-L., Hondzo, M., Rao, A.R., 2002. Segmentation of temperature microstructure in lakes. J. Geophys. Res., Oceans 107 (C12) 3211.

Donoho, D.L., Johnstone, I.M., 1994. Ideal spatial adaptation via wavelet shrinkage. Biometrika 81, 425–455.

EarthInfo, 1993. EarthInfo Earthinfo CD^2 USGS daily values, CD ROM, EarthInfo Inc, Boulder.

Hamilton, J.D., 1994. Time Series Analysis, Princeton University Press, Princeton.

Hinich, M.J., 1982. Testing for Gaussianity and linearity of a stationary time series. J. Time Series Anal. 3 (3), 169–176.

Hipel, K.W., McLeod, A.I., 1994. Time Series Modelling of Water Resources and Environmental Systems, Elsevier, Amsterdam, pp. 463–465.

Hu, Y.-S., 1994. Wavelet approach to change-point detection with application to density estimation. PhD Thesis, Texas AandM University, College Station

Hurrell, J., 1995. Decadal trends in the north atlantic oscillation: regional temperature and precipitation. Science 269, 676–679.

Kashyap, R.L., Rao, A.R., 1976. Dynamic stochastic models for empirical data, Academic Press, New York.

Kember, G., Flower, A.C., Holubeshen, J., 1993. Forecasting river flow using nonlinear dynamics. Stoch. Hydrol. Hydraul. 7, 205–212.

Nikias, C.L., Petropulu, A.P., 1993. Higher-Order Spectral Analysis, Prentice-Hall, Englewood Cliffs.

Ogden, R.T., Richwine, J., 1996. Wavelets in Bayesian change-point analysis. Technical Report, University of South Carolina, Columbia

Palmer, W.C., 1965. Meteorological drought. US Weather Bureau, Re. Paper No. 45

Perreault, L., Haché, M., Slivitzky, M., Bobée, B., 1999. Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach. Stochast. Environ. Res. Risk Assessment 13, 201–216.

Prasad, R., 1967. A non-linear hydrologic system response model. Proc. ASCE, J. Hydraul. Div. 93 (HY4), 201–202.

Priestley, M.B., 1981. Spectial Analysis and Time Series, Academic Press, London.
Priestley, M.B., 1988. Non-Linear and Non-Stationary Time Series Analysis, Academic Press, London.
Quinn, F.H., 1981. Secular changes in annual and seasonal Great Lakes precipitation, 1854–1979, and their implications for Great Lakes water resource studies. Water Resour. Res. 17 (6), 1619–1624.
Rao, A.R., Yu, G.-H., 1990. Gaussianity and linearity tests of hydrologic time series. Stochast. Hydrol. Hydraul. 4, 121–134.
Rao, A.R., Delleur, J.W., Sarma, P.B.S., 1971. Conceptual hydrologic models for urbanizing basins. Proc. ASCE, J. Hydraul. Div. 98 (HY7), 1205–1220.
Rodionov, S.N., Krovnin, A.S., 1992. The 1980s in the Context of Climate Changes in the North Atlantic Region. ICES Mar. Sci. Symp. 195, 86–95.
Rogers, W.F., 1980. A practical method for linear and non-linear runoff. J. Hydrol. 46, 51–78.
Rogers, W.F., 1982. Some characteristics and implications of drainage basin linearity and non-linearity. J. Hydrol. 55, 247–265.
Rogers, W.F., Zia, H.A., 1982. Linear and non-linear runoff from large drain basins. J. Hydrol. 55, 267–278.
Salas, J.D., 1993. Analysis and modeling of hydrologic time series. In: Maidment, D.R., (Ed.), The McGraw Hill Handbook of Hydrology, pp. 19.5–19.9.
Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L., 1985. Applied Modeling of Hydrologic Time Series, second ed, Water Resources Publications, Michigan.
de Souza, P.V., Thomson, P.J., 1982. LPC distance measures and statistical tests with particular reference to the likelihood ratio. IEEE Trans. ASSP 30 (2), 304–315.
Subba Rao, T., Gabr, M.M., 1980. A test for linearity of stationary time series. J. Time Series Anal. 1 (1), 145–158.
Subba Rao, T., Gabr, M.M., 1984. An Introduction to Bispectral Analysis and Bilinear Time Series Models, Springer Lectur Notes in Statistics 24, Springer Verlag, New York.
Subba Rao, T., Indukumar, K.C., 1996. Spectral and wavelet methods for the analysis of nonlinear and nonstationary time series. J. Franklin Inst. 333 (B3), 425–452.
Tong, H., 1990. Non-linear Time Series: A Dynamical Systems Approach, Clarendon Press/Oxford University Press, Oxford.
Tsay, R.S., 1988. Outliers, level shifts, and variance changes in time series. J. Forecast. 7, 1–20.
Wang, Y., 1995. Jump and sharp cusp detection by wavelets. Biometrika 82, 385–397.