水文時間序列LC持續及LC週期模式之一致性估算

Consistent Estimation of LC-persistent and LC-periodic Models of Hydrologic Time Series

淡江大學水資源研究所副教授

虞　國　興

Gwo-Hsing Yu

摘　　要

ARIMA (p, d, q) 模式廣泛被運用於預估上，同時證明爲一族十分有效之模式，然而當 d 大於零時，該模式却可能無法用於產生合成資料，這限制了其在水文學上的運用。

爲克服此困難，近年來在水文學上，分數 ARIMA(p d_f, q) 模式被提出，此模式 d_f值可能爲一分數，且較ARIMA (p, d, q)模式更具彈性，並可用來產生合成資料。

本文係採用 Kashyap和 Eom (1985) 所提之 LC-persistent 及 LC-periodic 模式來推估d_f值，並與最大概似法 (Maximum Likelihood Method) 和對數波譜法 (Log-Spectrum Method) 所推估之 d_f 值比較。

Abstract

ARIMA(p, d, q) models have been shown to be a powerful class of models which are being widely used for forecasting, With d greater than zero, it may not be possible to use these models for synthetic data genevation. This property limits their use in fields such as hydrology where designs are often based on generated data. In developing these models, it is usually assumed that the parameter d is restricted to take integer values.

Recently, a class of models called fractional difference models have been proposed for use in hydrology. In these models the parameter d is relaxed to include fractional values. ARIMA models denoted by ARIMA (p, d_f, q) are fitted to these fractional differenced series. The fractional differenced ARIMA models are more flexible than integer differenced ARIMA (p, d, q) models. ARIMA (p, d_f, q) medels can also be used to generate data.

Estimation of the fractional difference parameter d_f is of obvious importance in using the fractional difference models. A consistent parameter estimation method for LC-persistent and LC-periodic models proposed by Kashyap and Eom (1985) is used. The parameters of aunual hydrologic series are estimated and the results are presented. The results are then compared with those obtained by maximum likelihood and log-spectrum methods.

I. INTRODUCTION

Recently, a new approach of estimation of fractional difference parameter d_f has been proposed by Kashyap and Eom (1984). This new method is based on the least squares estimation method in the frequency domain. The estimator has been shown to be unbiased and consistent in the mean square sense with convergence rate $0(1/N)$. Kashyap and Lapsa (1982) generalized the first order long memory time series (LC-persistent series) to second order long memory time series (LC-periodic series). The spectral density of this second order series is represented by $f_z(\omega)$.

$$f_Z(\omega) = \left| 2(\cos\omega - \cos\omega_0) \right|^{-d_f} f_W(\omega)$$

Similar to the LC-persistent series, this series has indefinite spectral density at $\omega = \omega_0$, and the spectral density decays as $(\omega - \omega_0)^{-2d_f}$ as ω goes far from ω_0. The characteristics of the spectral density of LC-periodic series depends not only on the differencing parameter d_f and noise spectral density but also on resonant frequency ω_0. The spectral density of LC-periodic series is the same of LC-persistent series if $\omega_0 = 0$. Thus, the LC-persistent model is a special case of the LC-periodic model when $\omega_0 = 0$.

II. DATA USED IN THE STUDY

Two sets of data are used in the present study. The first set of data consists of Sunspot number series, Central England mean annual temperature series, and Eastern American mean annual temperature series. Some information about these three series are given below.

(a) Annual Sunspot series from 1660 to 1975 with 316 observations with a mean value of 9.13 and a variance of 0.362.

(b) Central England mean annual temperature series has been compiled and reported in Lamb (1977, Table App. V. 10). This series starts in 1659 and ends in 1976 with 318 observations. The mean value of this series is 9.15°C and variance is 0.38.

(c) Eastern American mean annual temperature series has also collected by Lamb (1977, Table App. V. 11). This series covers the period 1738 to 1967, with some missing data (for years 1741, 1764, 1778, 1779 and 1780). The average value of this series is used for these five missing values. The series has 230 observations with a mean value of 12.4°C and a variance of 0.29.

The second set of data consists of four annual flow series from St. Lawrence, Gota, Blacksmith and Gunpowder rivers. Some information about these four river flow series are given in Table 1.

Table 1. Some Statistical Characteristics of the Data Used in the Study.

River	Location	Period	N(1)	AVG(2)	VAR(2)	Source
St. Lawrence	Ogdensburg, New York	1860-1957	97	0.9999	0.00753	3
Gota	S jotorp-Vanersburg, Sweden	1807-1957	150	1.0	0.03276	3
Blacksmith	Hyrum, Utah	1913-1979	66	1.0	0.112	4
Gunpowder	Loch Raver, Maryland	1883-1978	95	1.0	0.10895	5

1. Number of data points (series is given as the modular coefficients of observed and computed annual flows for water year).
2. AVG, VAR are the mean value and variance of the series.
3. Yevdjevich (1963).
4. Data from 1913 to 1957 are from Yevdjevich (1963), data from 1957 to 1970 are from United Geological Supply, Water Supply Paper, and data from 1970 to 1979 are from Utah Power and Light Company, Salt Lake City, Utah.
5. Data from 1883 to 1963 are from Eastman (1972). The rest of the data are from the Department of Public Works, Bureau of Water and Waste Water, Baltimore, Maryland.

III. THEORETICAL ASPECTS

The estimation method for LC-persistent and LC-periodic series proposed by Kashyap and Eom (1984) is as follows.

3.1 Estimation in LC-persistent Model

The long memory time series, LC-persistent series, is expressed in time domain as follows

$$Z_t = (1-z^{-1})^{-d_f} W_t, \quad t = 0,1,\ldots,N-1 \qquad [1]$$

where z^{-1} is the unit delay operator and W_t is a zero mean white Gaussian noise with variance ρ. The relation between the spectral density of Z_t denoted by $f_z(\omega)$, and that of W_t, represented by $f_w(\omega)$, can be written as follows.

$$f_Z(\omega) = [2(1-\cos\omega)]^{-d_f} f_W(\omega) \qquad [2]$$

The discrete Fourier transform (DFT) of eq. (1) is

$$Z(k) = F[Z_t] \qquad [3]$$

$$Z(k) = (1-e^{\frac{-i2\pi k}{N}})^{-d_f} W(k) \qquad [4]$$

$$Z(k) = e^{\frac{-i\pi k d_f}{N}} [2i\sin(\frac{\pi k}{N})]^{-d_f} W(k) \qquad [5]$$

for $k = 0, 1, \ldots, N-1$. In eq. 3 Z(k) and W(k) are discrete Fourier transforms of Z_t and W_t, respectively. The magnitude of Z(k) is given in eq. 4.

$$|Z(k)| = [2|\sin(\frac{\pi k}{N})|]^{d_f} |W(k)|,$$

$$\text{for } k = 0,1,\ldots,N-1 \qquad [6]$$

By applying logarithm operator to eq. 4, a linear equation can be obtained in parameter d_f.

$$\log|Z(k)| = -d_f \log|2\sin(\frac{\pi k}{N})| + \log|W(k)| \qquad [7]$$

for $k = 0, 1, \ldots, N-1$.

The relationship Z(k) and d_f is given in eq. 7 as a signal plus noise.

Kashyap and Eom (1984) have shown that the sequence $\{|W(k)|, k = 0, 1, \ldots, \frac{N}{2}\}$, the magnitude of a discrete Fourier transformed white Gaussian noise sequence, is also a white sequence with Rayleigh densities.

$$f_{|W(k)|}(\omega) = \begin{cases} 2\frac{\omega}{\rho N}\exp\{-\frac{\omega^2}{\rho N}\} & \omega \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad [8]$$

Then, they showed that the transformed frequency domain noise sequence $\{\log|W(k)|, k = 0, 1, ..., \frac{N}{2}\}$ is a white sequence

$$\text{with} \quad E[\log|W(k)|] = -\frac{\gamma}{2} + \frac{1}{2}\log(\rho N)$$

$$\text{and} \quad Var[\log|W(k)|] = \frac{\pi^2}{24} \qquad [9]$$

where γ is Euler's constant (= 0.5772157) and ρ is the variance of time domain noise sequence W(t). Let $Z = \{|Z(0)|, |Z(1)|, ..., |Z(N-1)|\}$. Then, it can be shown that Z has the following log-likelihood function for $\underset{\sim}{\theta} = (d_f, \rho)$.

$$L(Z;\theta) = 2d_f \sum_{k=1}^{\frac{N}{2}} \log(2|\sin(\frac{\pi k}{N})|) + \sum_{k=0}^{\frac{N}{2}-1} \log|Z(k)|$$

$$-\frac{N}{2}\log(\frac{\rho N}{2}) - \sum_{k=1}^{\frac{N}{2}} \frac{|2\sin(\frac{\pi k}{N})|^{2d_f}|Z(k)|^2}{\rho N} \qquad [10]$$

Now, define $\alpha = -E[\log|W(k)|]$ and $V(k) = \log|W(k)| + \alpha$. Then, V(k) is a zero mean white noise sequence. Eq. 5 can be rewritten as follows.

$$\log|Z(k)| = -d_f\log|2\sin(\frac{\pi k}{N})| - \alpha + V(k) \qquad [11]$$

Eq. 11 expresses log|Z(k)| in terms of deterministic trend term and additive zero mean white noise. Thus, the least square estimation algorithm can be applied to the above equation. Let us define the parameter vector $\underset{\sim}{\theta} = (d_f, \alpha)^T$. Then the least square estimator θ is the argument which minimizes the following cost function $J(\underset{\sim}{\theta})$,

$$J(\underset{\sim}{\theta}) = \sum_{k=1}^{\frac{N}{2}} (\log|Z(k)| + d_f\log|2\sin(\frac{\pi k}{N})| + \alpha)^2$$

$$= \sum_{k=1}^{\frac{N}{2}} (\log|Z(k)| - \underset{\sim}{\theta}^T \underset{\sim}{A}(k))^2 \qquad [12]$$

where

$$\underset{\sim}{A}^T(k) = (-\log|2\sin(\frac{\pi k}{N})| \quad -1)$$

The cost function is summed from k = 1 to $\frac{N}{2}$, because |Z(N−k)| = |Z(k)| for k = 1, 2, ..., $\frac{N}{2}$ and $\log|2\sin(\frac{\pi k}{N})| = -\infty$ for k = 0. Then, the least square estimator $\hat{\underset{\sim}{\theta}}$ is obtained by the standard formula (Kashyap and Rao, 1976, p. 137).

$$\hat{\underset{\sim}{\theta}} = (d_f, \alpha)^T = (\sum_{k=1}^{\frac{N}{2}} \underset{\sim}{A}(k)$$

$$\underset{\sim}{A}^T(k))^{-1}(\sum_{k=1}^{\frac{n}{2}} \underset{\sim}{A}(k)\log|Z(k)|) \qquad [13]$$

Kashyap and Eom (1984) have shown that the estimators d_f and α are unbiased and consistent in the mean square sense with variance $\frac{1}{N}$ and $(\frac{\pi^2}{12N})$ for large N. They also showed that ρ is unbiased and consistent in the mean square sense with variance $\frac{\rho^2\pi^2}{3N}$, where

$$\rho = \frac{1}{N}\exp(\gamma - 2\alpha - \frac{\pi^2}{6N}) \qquad [14]$$

3.2 Estimation in the LC-Periodic Model

The LC-periodic series is expressed in time domain as follows

$$Z_t = (1 - 2\cos\omega_0 z^{-1} + z^{-1})^{\frac{-d_f}{2}} W_t \qquad [15]$$

for t = 0, 1, ..., N−1

where z^{-1} is unit delay operator and W_t is zero mean white Gaussian noise with variance ρ. The spectral density of Z_t is given by

$$f_Z(\omega) = \rho[2(\cos\omega - \cos\omega_0)]^{-d_f} \quad [16]$$

The discrete Fourier transform of eq. 15 is

$$Z(k) = F[Z_t]$$

$$Z(k) = (1 - 2\cos\omega_0\, e^{-\frac{i2\pi k}{N}} + e^{-\frac{i4\pi k}{N}})^{-\frac{d_f}{2}} W(k)$$

$$= e^{-\frac{i\pi k d_f}{N}} [2(\cos(\frac{2\pi k}{N}) - \cos\omega_0)]^{-\frac{d_f}{2}} W(k) \quad [17]$$

for k = 0, 1, ..., N−1

Then, equation 18 follows from eq. 17

$$|Z(k)| = |2(\cos(\frac{2\pi k}{N}) - \cos\omega_0)|^{-\frac{d_f}{2}} |W(k)| \quad [18]$$

for k = 0, 1, ..., N−1

Let $\underset{\sim}{Z} = \{ |Z(0)|, |Z(1)|, ..., |Z(N-1)| \}$. Kashyap and Eom (1984) have shown that $\underset{\sim}{Z}$ has the following log-likelihood function for $\theta = (d_f, \rho, \omega_0)$

$$L(\underset{\sim}{Z};\underset{\sim}{\theta}) = d_f \sum_{k=0}^{\frac{N}{2}} \log|2(\cos(\frac{2\pi k}{N}) - \cos\omega_0)|$$

$$+ \sum_{k=0}^{\frac{N}{2}} \log|Z(k)| \; - \sum_{k=0}^{\frac{N}{2}} \log(\frac{\rho N}{2})$$

$$- \sum_{k=0}^{\frac{N}{2}} \frac{|2(\cos(\frac{2\pi k}{N}) - \cos\omega_0)|^{d_f} |Z(k)|^2}{\rho N} \quad [19]$$

As discussed in the LC-persistent model, the cost function $(J\underset{\sim}{\theta})$ of LC-periodic model can be easily obtained as in eq. 20.

$$J(\underset{\sim}{\theta}) = \sum_{k=0}^{\frac{N}{2}} (\log|Z(k)|$$

$$+ d_f \log|2(\cos(\frac{2\pi k}{N}) - \cos\omega_0)| + \alpha)^2 \quad [20]$$

In this case, the three parameters are estimated by minimizing the cost function by numerical methods. Consequently parameter estimation in LC-periodic model is more difficult than in LC-persistent model, because in the former the resonant frequency ω_0 must also be estimated. However, eq. 20 is linear with respect to parameters if the resonant frequency ω_0 is known. Kashyap and Eom (1984) have proposed an algorithm which is a hybrid of LS and ML estimation methods. The parameters d_f and ρ are estimated by LS method, and ω_0 is estimated by ML method. The procedures of this hybrid algorithm are described as below.

a. Guess resonant frequency ω_0 in the range $(0, \frac{\pi}{2})$.
b. Estimate d_f and α by using LS estimation algorithm.

$$(d_f, \alpha) = (\sum_{k=0}^{\frac{N}{2}} \underset{\sim}{A}(k)\underset{\sim}{A}^T(k))^{-1} (\sum_{k=0}^{\frac{N}{2}} \underset{\sim}{A}(k) \log|Z(k)|) \quad [21]$$

where

$$A^T(k) = (-\frac{1}{2}\log|2(\cos(\frac{2\pi k}{N})-\cos\omega_0)| \quad -1)$$

c. Compute the estimate of ρ by using eq. 14.

d. Using the estimates d_f and ρ found in (b) and (c), maximize eq. 19 with respect to ω_0.

e. Using the estimated resonant frequency ω_0 from (d), repeat (b) and (c).

f. Repeat (b), (c), (d) and (e) until all three estimates have no significant change in successive iterations.

IV. RESULTS AND CONCLUSIONS

Both the LC-persistent and LC-periodic methods were used to estimate the parameters d_f of some hydrologic and sunspot number series. These estimates are given in Tables 2 and 3. It is interesting to note $\hat{\omega}_0$ changes $\hat{d}_f$ considerably. Obviously for the nonperiodic series such as Gota, St. Lawrence and Blackswith river flow, and Central England and Eastern American mean annual temperature series the LC-persistent method is preferred.

The results obtained by LC-persistent and LC-periodic methods are then compared with those obtained by maximum likelihood and log-spectrum methods. For further details about maximum likelihood and log-spectrum methods, reader is referred to Rao and Yu (1984). The following conclusions may be presented on the basis of the results of the present study and Rao and Yu (1984):

(1) The statistical characteristics of the ARIMA (p, d, q) and ARIMA (p, d_f, q) are close to each other.

(2) The rescaled range characteristics of the data generated by ARIMA (p, d, q) and ARIMA (p, d_f, q) models are quite similar to each other.

(3) The forecasting abilities of ARIMA (p, d_f, q) models are of the same caliber.

(4) The theoretical advantages claimed for fractional difference models are not so obviously present in the characteristics of synthetic data traces.

(5) d_f value obtained by LC-persistent method is less than that obtained by LC-periodic method when the process is nonperiodic.

(6) When the process is nonperiodic, the LC-persistent method is preferred because d_f value is close to that obtained by maximum likelihood method.

Table 2. LC-Persistent Model Estimates

Series	d_f	ρ
Gota River Flow	0.45	0.0265
Gunpowder River Flow	0.28	0.1277
St. Lawrence River Flow	0.72	0.0039
Blacksmith River Flow	0.61	0.0926
Sunspot Number Series	0.85	294.5
Central England Annual Temp.	0.23	0.3706
Eastern American Annual Temp.	0.38	0.2352

(文轉第91頁)