

Box-Cox指數轉換之研究及其應用於月流量時序之分析

Studies on Box-Cox Power Transformation And Its Application on the Analysis of Monthly Streamflow Series

淡江大學水資源及環境工程研究所教授

淡江大學水資源及環境工程研究所研究助理

虞 國 興
Gwo-Hsing Yu

歐 立 中
Lih-Chung Ou

摘 要

水文時間序列上大都為非定常性非常態之時間序列，本研究利用Box andCox(1964)所提之指數轉換 (Power transformation) 將序列轉換後再分析其模式，研究中利用由Yu及Lin(1991)所提之部份自迴歸模式 (Subset autoregressive model簡稱SAR模式) 分別模擬轉換前後之序列。研究結果顯示，無論合成資料或實測資料，Box-Cox所提轉換參數之推估方法均能有效推求轉換參數，然當以非常態分布擾動項產生合成序列時，轉換後模式之擾動項大都無法通過常態分布檢定。實測月流量之研究結果發現大部份轉換後模式之擾動項通過常態分布檢定，顯示實測月流量序列大部份仍具備常態分布擾動項，同時，轉換後序列之SAR模式於低流量 (11-4月) 時其預測能力甚佳。本研究依據常態分布之特性，提出另一毋須經過模式模擬之轉換參數推估方法，結果顯示此方法能有效推估轉換參數，且能節省計算時間。

關鍵詞：Box-Cox指數轉換，定常性時序，非常態時序。

ABSTRACT

Most of the hydrologic time series are non-stationary and non-Gaussian process. The power transformation proposed by Box and Cox (1964) was used to transform the data in the present study. Then, the transformed data were analyzed and fitted the model. In this study, the subset autoregressive model, denoted SAR in brief, proposed by Yu and Lin (1991) was employed to fit the data and its transformed data. The results indicated that method proposed by Box-Cox can estimate the transformation parameter in accuracy. However, the noise term of transformed data could not pass the normality test in general when the series are generated with non-Gaussian noise. When the monthly streamflows were studied. It showed that most of the noise term of the transformed data pass the normality test. This results indicated that most of the observed monthly streamflows contain Gaussian noise terms. Meanwhile, SA Rmodel for transformed data can forecast the low flow (Nov.-Apr.) much better than the SAR model for the original data. Based on the properties of Gaussianity, a methodology for estimating the transformation parameter without modeling the data was proposed in the present study. The results indicated that the proposed method and efficiently estimate the transformation parameter and save lots of computation time.

Keywords : Box-Cox Power Transformation, Stationary Time Series, Non-Gaussian Time Series.

一、前言

時間序列可分為定常性序列(Stationary series)及非定常性序列(Non-stationary series)，而此兩類序列又可分为常態過程(Normal process)及非常態過程(Non-normal process)。在一般水文時間序列上多屬非定常性非常態時間序列，分析時常將序列轉換使其為定常性及常態。Box and Jenkins (1970)提出自迴歸—積分—移動平均模式(Autoregressive-integrated-moving average model, 簡稱ARIMA模式)，以差分方法(Difference)將非定常性序列轉換成定常性序列，再利用自迴歸—移動平均模式(Autoregressive-moving average model, 簡稱ARMA模式)加以分析。於序列常態過程部份，本研究利用Box及Cox(1964)提出之指數變換(power transformation)將非定常性非常態時序列轉換至常態後，再分析序列之模式。

另外，Box-Cox所提轉換參數之推估方法必須經過模式之選取，然可能因資料結構複雜，模式之選取將耗費相當之計算時間，為解決此缺失，本研究根據常態分布之特性，提出一無須經過模式模擬，直接推求轉換參數之方法。

Deaper及Cox(1969)針對Box-Cox所提之轉換參數推估方法所求得之轉換參數，無法有效的將資料轉換為常態，提出另一修正方法，唯此方法在某些時候僅能使資料變為較具對稱之型式，並無法保證資料具備常態分布特性。而Hopwood, Mckeown及Newbold(1984)將Box-Cox指數轉換應用於ARIMA模式時序列上，發現在某些序列上會獲得較佳之預測結果。

水文資料，常被視為對數常態分布(Log normal distribution)或加瑪分布(Gamma distribution)，Yevjevich(1972)直接以邊際機率分布(Marginal probability distribution)為加瑪分布及對數常態分布之AR(1)模式模擬水文資料，而不須將資料轉為常態。故合成序列之研究，本研究分別以擾動項為二參數對數常態分布及甘保(Gumbel)分布產生合成序列，另外並以擾動項為常態分布加以產生合成序列，並經逆轉換產生合成序列藉以探討轉換參數之推估方法。於實測資料上本研究使用台灣實測月流量資料，並利用由Yu及Lin(1991)所提之部份自迴歸模式(Subset autoregressive model簡稱SAR模

式)加以模擬月流量序列於轉換前後之模式。本研究之主題分述如下：

- (1)探討Box-Cox所提轉換參數推估方法之適切性。
- (2)提出一無須經過模式選取，直接推估轉換參數之方法，與Box-Cox所提之推估方法作比較。
- (3)以合成資料及實測資料比較轉換前後模式及特性之差異。
- (4)探討轉換後之資料是否具備常態分布。
- (5)探討分別以傳統方法及Bootstrap重新抽樣法於資料經轉換或不轉換兩種情況下，所產生之合成資料於原始資料統計特性保有性能力之優劣。
- (6)比較轉換前後模式之預測能力。

本論文之大綱如下：第二節為理論基礎。第三節為本論文所使用之資料，合成資料及實測資料之研究結果與討論列於第四節，最後，第五節為結論。

二、理論基礎

本論文主要之研究目的，在比較轉換前後序列之模式及特性，本節將針對研究中所使用方法作一理論背景說明，主要分成以下幾項①Box-Cox所提之轉換參數推估方法②本研究所提之轉換參數推估方法③SAR模式④模式適合度之比較及⑤模式預測能力比較。

2-1 Box-Cox指數轉換：

Box-Cox(1964)提出對因變數(dependent variable) $\{Z_t\}$ 轉換成 $\{Z_t^{(\lambda_1, \lambda_2)}\}$ 之轉換參數 λ_1 及 λ_2 的轉換公式

$$Z_t^{(\lambda_1, \lambda_2)} = \frac{(Z_t + \lambda_2)^{\lambda_1 - 1}}{\lambda_1} \quad (\lambda_1 \neq 0)$$
$$= \text{Ln}(Z_t) \quad (\lambda_1 = 0) \dots\dots(1)$$

其中參數 λ_2 為使序列 $\{Z_t\}$ 轉換為正值後再取指數轉換，但所分析之實測資料其值均為正，故本研究取式②之轉換公式加以轉換及分析。

$$Z_t^{(\lambda)} = \frac{Z_t^\lambda - 1}{\lambda} \quad (\lambda \neq 0) \dots\dots(2)$$
$$= \text{Ln}(Z_t) \quad (\lambda = 0)$$

其中時間序列 $\{Z_t\} = \{z_1, i=1, 2, \dots, n\}$ 式(2)必須滿足 $\{Z_t > 0 \mid t=1, 2, \dots, n\}$ ， λ 為欲轉換之參數，原始序列 $\{Z_t\} = \{Z_1, Z_2, \dots, Z_n\}$ 經式(2)轉換後可得 $\{Z_t^{(\lambda)}\} = \{Z_1^{(\lambda)}, Z_2^{(\lambda)}, \dots, Z_n^{(\lambda)}\}$ ，再經過選取適當線性模式之後可得 $E\{Z_t^{(\lambda)}\} = a\theta$

其中 a 是已知矩陣，而 θ 是未知參數，假設在轉換後 $\{Z_t\} = \{Z_1^{(\lambda)}, Z_2^{(\lambda)}, \dots, Z_n^{(\lambda)}\}$ ，滿足獨立之常態分布，沒有固定之變異數 (σ^2) 及期望值，則其轉換後參數 λ 之概似函數 (Likelihood function) 為

$$\frac{1}{(2\pi)^{n/2} \sigma^n} \text{Exp}\left\{-\frac{(Z_t^{(\lambda)} - a\theta)'(Z_t^{(\lambda)} - a\theta)}{2\sigma^2}\right\} J(\lambda : Z_t) \quad \dots\dots(3)$$

其中， $J(\lambda : Z_t) = \left| \frac{dz_i(\lambda)}{dz_1} \right|$ ，為函數變換式之 Jacobian 再利用最小二乘方法 (Least-squares method) 推估參數後可得

$$\hat{\sigma}^2(\lambda) = Z^{(\lambda)'} a r Z_t^{(\lambda)} \quad n = S(\lambda)/n \quad \dots\dots(4)$$

其中 $a r = I - a(a'a)^{-1}a'$ ， $S(\lambda)$ 為 Z_t 經模式分析後殘差之平方和，將式(3)取對數並予以簡化後可得

$$L_{\max}(\lambda) = -\frac{1}{2} n \ln \hat{\sigma}^2(\lambda) + \ln J(\lambda : Z_t) \quad \dots\dots(5)$$

由於 $\ln J(\lambda : Z_t) = (\lambda - 1) \sum_{i=1}^n \ln Z_i$ 式(5)可改寫為

$$L_{\max}(\lambda) = -\frac{1}{2} n \ln S(\lambda) + (\lambda - 1) \sum_{i=1}^n \ln Z_i \quad \dots\dots(6)$$

而 Hopwood, Mckeown 及 Newbold (1984) 提出式(7)

$$g(\lambda) = -\frac{1}{2} n \hat{\sigma}_{e,\lambda}^2 + (\lambda - 1) \sum_{i=1}^n \ln Z_i \quad \dots\dots(7)$$

其中 $\hat{\sigma}_{e,\lambda}^2$ 為轉換後資料擾動項之變異數。

本研究在實際推估時間序列轉換參數 λ 時，取 $\lambda = -1(0.05)1$ (表由 1 至 -1 每隔 0.05 取一 λ 推估值)，以 λ 代入式(2)將序列轉換，分析轉換後序列之模式並求出其 $\hat{\sigma}_{e,\lambda}^2$ ，將推估參數 λ 及 $\hat{\sigma}_{e,\lambda}^2$ ，代入式(7)求 $g(\lambda)$ ，發生 $g(\lambda)$ 最大值時之 λ 值即為所選取之轉換參數。

如將 Box-Cox 指數轉換在不同轉換參數 λ 下式(2)的轉換量與傳統上之轉換量之關係可得如表 1 所示之結果。

表 1 Box-Cox 指數轉換之轉換參數 λ 與傳統轉換式之比較

λ 值	傳統轉換式
-1.0	$1/Z_t$
-0.5	$1/\sqrt{Z_t}$
0.0	$\frac{\ln Z_t}{Z_t}$
0.5	$\sqrt{Z_t}$
1.0	Z_t (未轉換)

2-2 本研究所提轉換參數 λ 之推估方法：

若序列經過轉換後具常態分布之特性，其偏態係數 (coefficient of Skewness)， C_s 趨近理論值為 0，且峰度係數 (coefficient of Kurtosis)， K_u 趨近理論值為 3，基於此一常態分布特性，利用式(2)將原始序列做轉換後，求不同 λ 值轉換後序列之 C_s 及 K_u ，代入式(8)，發生最小值時 $g'(\lambda)$ 之 λ 值即為當經過此 λ 值轉換後，資料最接近常態分布，所以本研究提出式(8)推估轉換參數之判斷式。

$$g'(\lambda) = |C_s| + (k_u - 3) \lambda \quad \dots\dots(8)$$

2-3 部份自迴歸(SAR)模式：

部份自迴歸模式乃 AR(P) 模式之一特例，即 AR(P) 中有部份係數為零，又因研究之實測資料為台灣月流量資料，而由 Yu and Lin (1991) 所提之部份自迴歸模式 (簡稱 SAR 模式) 研究中發現絕大部份台灣實測月流量序列可由此選取模式來模擬，故本研究於轉換前及轉換後臺灣實測月流量序列，採用 SAR 模式來模擬。(因篇幅所限，有關自迴歸模式建立之詳細理論請參閱參考文獻 (11))

2-4 模式之比較：

本研究分別就轉換前後序列作模式之適合度 (Goodness-of-Fit) 及模式統計特性保有性 (Statistical Characteristic Perservation) 做深入分析，其詳細理論基礎簡述如下：

(1) 模式之適合度：

(A) 模式擾動項 $\{a_t\}$ 獨立性檢驗、常態測試及模式相關係數比較

(a) Q Test (或稱 Protomateau Test):

因為基本假設上殘值序列 $\{a_t\}$ 應該是 IID (Identical Independent Distribution) $\sim N(0, \sigma_a^2)$ ，所以如果模式正確，則 $\{a_t\}$ 應該是白噪音 (White Noise)。因此，若模式正確，則統計量 Q 之分佈近似於 $\chi^2(k-p-q)$ 卡方分佈。

$$Q = N \sum_{\tau=1}^k \gamma_a^2(\tau) \quad \dots\dots(9)$$

其中

γ_a = 模式殘值之自相關函數

K = 自相關函數之時間稽延 (Time Lag)

N = 資料個數

(b) 累積週期測試 (Cumulative Periodogram Test (C-P Test)):

分析週期性序列，必須檢驗其殘值是否依然存在著週期性，本研究使用累積週期測試加以檢定，其理論基礎簡述如下：

(i)週期圖 (periodogram)

時間序列 $\{a_t\}, t=1, 2, \dots, n$ 之週期圖為

$$I(F_1) = 2 / [(\sum_{t=1}^n a_t \cos 2\pi f_1 t)^2 + (\sum_{t=1}^n a_t \sin 2\pi f_1 t)^2] \dots\dots(10)$$

其中 $f_1 = i/n$

(ii)正規化累積週期圖 (normalized cumulative periodogram)

$$C(f_1) = \sum_{k=1}^i I(f_k) / (n \hat{\sigma}_a^2) \dots\dots(11)$$

(iii)週期檢定方法

本研究以 Kolmogorov-Smirnov 統計量 D_n 來檢定序列是否具週期性，取顯著水準為 0.05，得 $D_n = 1.36 \sqrt{N/2-2}$ ，當序列之 $C(f_1)$ 對應於 F_1 繪於 (0,0) 與 (0.5,1) 區間時，應趨近一直線，即若 $C(f_1)$ 均落於 $(2f_1 - D_n, 2f_1 + D_n)$ 間，則表序列不具週期性，反之則序列仍具週期性。

(C)模式擾動項 $\{a_t\}$ 常態分布檢定：

依樣本數 n ，將 $\{a_t\}$ 依等機率分成 k 組，再以依卡方檢定 (chi-squared test) 於顯著水準 α 檢定其是否為常態分布其中

$$k = [1 + 3.2 * \text{Log}_{10}(n)]$$

$[x] = x$ 之整數值

$$\text{統計量 } T = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi^2 (k-1-p) \dots\dots(12)$$

其中， E_i 表分割相等機率區間之理論資料個數。

O_i 實際落於區間內之資料個數。

P 參數個數 (常態分布 $P=2$)。

(B)模式相關係數 (R^2 : coefficient of determination)

$$R^2 = 1 - \frac{\sigma_a^2}{\sigma_z^2} \dots\dots(13)$$

其中

σ_a^2 = 資料殘差之變異數

σ_z^2 = 原始資料之變異數

②模式統計特性保有性

模式之保有性係指模式人造資料 (synthetic data) 是否保有原始資料之統計特性。本研究分別計算資料之平均值、變異數、偏態係數、峰度係數、組體圖 (Histogram)、自相關函數 (Autocorrelation Function (ACF)、部份自相關函數 (Partial Autocorrelation Function (PACF) 及波譜 (Spectral Density) 等做一比較。

模式之人造資料，本研究分別利用常態分布擾動項及 Bootstrap 重新抽樣擾動項產生，Bootstrap 重新抽樣法敘述如下：

由 Efron (1979) 所提之 Bootstrap resampling 理論為設 W_1, W_2, \dots, W_n 為彼此獨立，且為同一未知分布 F 之隨機變數，令 W 為由該組隨機變數 (Random variable) 所構成之隨機樣本，即 $W = (W_1, W_2, \dots, W_n)$ ；再令 w 為 W 之一組觀測資料，即 $w = (w_1, w_2, \dots, w_n)$ 。定義一隨機變數 $\alpha(W, F)$ 與 W 及未知分布 F 有關，則可利用觀測資料 w 推估 α 之抽樣分布 (Sampling distribution)。方法敘述如下：

(A) 利用觀測資料 w 建立 F 之樣本機率分布 (Sample probability distribution) F ，即令 F 之樣本空間 (Samples space) 為 $\{w_1, w_2, \dots, w_n\}$ ，而且使每一數值 w_1, w_2, \dots, w_n 均具有 $1/n$ 之機率分布。

(B) 依 F 之機率分布隨機替換抽樣 (Sampling with-replacement)，取得樣本長度為 n 之 Bootstrap 樣本 $W^* = (W_1^*, W_2^*, \dots, W_n^*)$

(C) 實際抽樣時，例如 u_1^* ，可利用蒙地卡羅模擬術 (Monte Carlo simulation) 由均勻分布 (0,1) 中產生亂數 u_1 ，令 k_1 為 nxu_1 之整數部分，則 $w_1^* = w_{k_1}$ ，再重複以上步驟，由原樣本空間 $\{w_1, w_2, \dots, w_n\}$ 抽樣得 $w_i^* = w_{k_i}, i=1, 2, \dots, n$ 。

(D) 實際產生擾動項時，利用原模式之擾動項為抽樣之樣本，利用上述方法重覆抽樣出人造合成序列之擾動項。

2-5 模式預測能力比較：

本研究主要針對實測資料於轉換前及轉換後之 SAR 模式作單期之預測 (One-Step-Ahead Forecast)，即

$$\hat{Z}_t(1) = E(Z_{t+1} | Z_t, Z_{t-1}, \dots)$$

本研究採用之預測能力比較指標為均方誤差 (MSE)，其定義如式(14)所示。

$$MSE = \frac{1}{K} \sum_{t=T+1}^{T+K} (Z_t(1) - Z_{t+1})^2 \quad \dots\dots(14)$$

其中

Z_{t+1} ≡ 第t+1期資料之觀測值

$Z_t(1)$ ≡ 單期預測值，即 Z_{t+1} 之預測值

T ≡ 資料長度

K ≡ 預測值個數

從MSE值大小判定預測值與觀測值之離散程度，其數值愈小，顯示其預測能力愈佳。

三、本研究所使用之資料

本研究所使用之資料分為合成資料與實測資料兩種，分述如下：

3-1 合成資料：

(1) 利用已知模式及參數產生具常態擾動項之合成原始時間序列，再利用式(1)之逆轉換式求得欲分析之原始時間序列。所使用之模式如下：

Model 1 : $Z_t = 0.5Z_{t-1} + a_t$

Model 2 : $Z_t = 0.8Z_{t-1} - 0.4Z_{t-2} + a_t$

Model 3 : $Z_t = 0.5Z_{t-1} + 0.4Z_{t-2} - 0.6Z_{t-3} + a_t$

Model 4 : $Z_t = 0.7Z_{t-1} - 0.2Z_{t-2} + a_t$

表2實測月流量資料之統計特性

區域	資料編號	站名	流域名稱	起訖年	平均值	變異數	偏態係數	峰度係數
北區	F030001	秀巒	淡水河	1957-1988	0.58	.50	4.09	31.42
	F030002	玉峰	淡水河	1957-1988	1.76	3.80	3.02	15.50
	F030004	菱角	淡水河	1957-1988	.71	.42	2.82	15.73
	F030005	高義	淡水河	1957-1988	3.02	11.69	3.67	23.44
	F030021	三峽	淡水河	1957-1988	.87	0.58	2.45	12.23
	F030026	福山	淡水河	1953-1988	1.80	2.41	2.72	13.25
	F100013	蘭陽大橋	蘭陽溪	1950-1988	6.30	41.29	2.33	9.34
	F130014	上溪	頭前溪	1951-1979	1.97	5.96	2.56	11.57
	F180001	山腳	南澳溪	1959-1988	.60	.54	4.31	31.08
中區	F250003	南湖	大甲溪	1960-1988	.67	.49	2.25	8.78
	F250004	環山合流點	大甲溪	1959-1988	1.40	1.69	2.10	9.23
	F250006	四季朗	大甲溪	1959-1988	.87	.78	3.56	21.70
	F270014	柑子林	烏溪	1956-1986	5.45	49.28	2.81	12.25
	F290009	武界	濁水溪	1932-1983	3.59	13.15	2.96	19.70
	F290042	桶頭	濁水溪	1957-1988	1.93	7.92	2.37	9.44
東區	F220013	綠水	立霧溪	1961-1988	3.32	9.87	2.21	7.81
	F300012	平林	花蓮溪	1960-1987	1.76	3.85	3.21	15.81
	F300037	仁壽橋	花蓮溪	1960-1987	3.37	8.73	2.96	14.33
	F300037	仁壽橋主流	花蓮溪	1960-1988	1.99	7.96	3.23	16.08
	F340011	立山	秀姑巒溪	1959-1988	2.02	4.40	2.83	12.69
	F400016	延平	卑南溪	1957-1988	3.69	20.21	2.37	9.34
	F400020	臺東大橋	卑南溪	1949-1988	9.90	168.92	2.38	9.75
南區	F390011	新營	急水溪	1961-1988	.85	1.84	2.64	11.53
	F51001	新發大橋	高屏溪	1959-1988	7.13	80.98	2.28	8.55
	F510031	月眉	高屏溪	1949-1986	4.07	39.57	2.50	9.99
	F510036	高屏大橋	高屏溪	1961-1988	24.00	1063.46	2.14	7.84

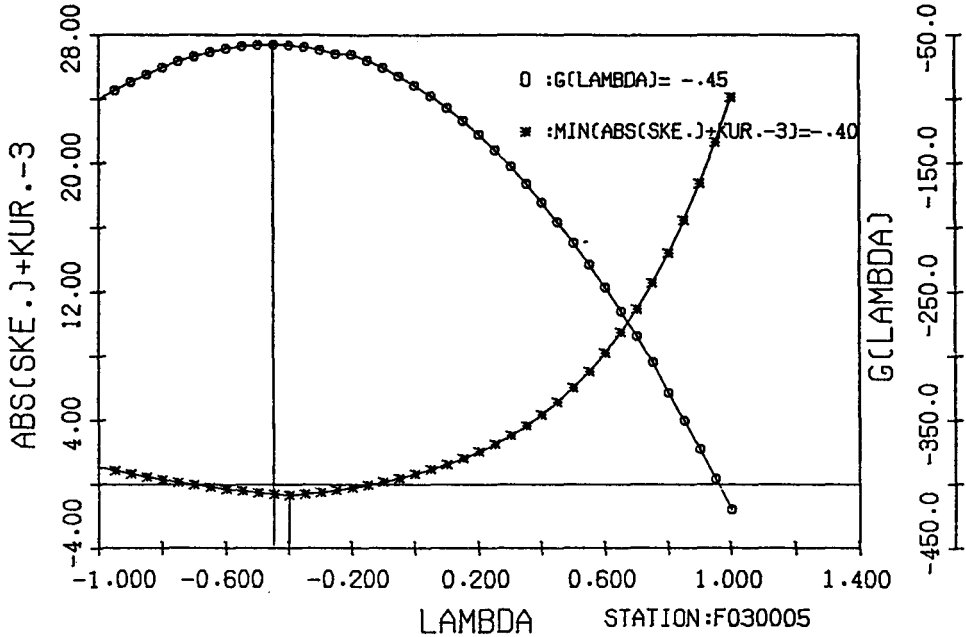


圖 1 F030005 (高義站, 淡水河) Box-Cox及本研究所提方法推估轉換參數之結果比較

其中 $a_t \sim N(0,1)$, 樣本數為120, 各模式分別產生100組, 轉換參數 λ 分別取-0.2、-0.4、-0.6、-0.8、0.0、0.2、0.4、0.6及0.8。

(2)利用已知模式及參數產生具非常態擾動項之合成時間序列。所使用之模式如下：

- Model 5 : $Z_t = 0.7Z_{t-1} + a_t$ $a_t \sim LN2$
- Model 6 : $Z_t = 0.7Z_{t-2} + a_t$ $a_t \sim Gumbel$
- Model 7 : $Z_t = 0.7Z_{t-12} + a_t$ $a_t \sim LN2$
- Model 8 : $Z_t = 0.7Z_{t-12} + a_t$ $a_t \sim Gumbel$
- Model 9 : $Z_t = 0.3Z_{t-1} + 0.2Z_{t-12} + a_t$ $a_t \sim LN2$

其中樣本數分別為60、120、240及360, 各模式分別產生100組。

3-2 實測資料：

本研究共計分析26站台灣實測月流量資料, 其基本統計特性如表2所示。

四、結果與討論

本節就合成資料及26組台灣實測月流量資料之研究結果分別討論如下(因篇幅所限, 僅列出部份結果, 詳細請參閱參考文獻⁽²⁾)。

4-1 合成資料：

探討Box-Cox方法與本研究所提方法於轉換參數

推估值之精確度, 同時亦以SAR模式探討Bootstrap重新抽樣法與傳統方法於產生合成資料時對統計特性保有性之比較。其研究結果分述如下：

(1)表3所示為模式1至4於不同真值 λ 時, Box-Cox方法100組轉換參數推估之結果。結果顯示, 逆轉換序列情況下, λ 之推估值相當靠近真值, 然存在高估現象。同時無論模式結構為簡單之AR(1)模式或較複雜之AR(3)、SAR(12)模式, 並不影響 λ 值的推估, 反而較複雜之模式所推估之結果更精確。另, 結果亦顯示逆轉換序列轉換後模式模擬之擾動項大都能通過常態分布檢定(於顯著水準 $\alpha=0.05$)。

(2)表4所示為模式4至8於不同樣本數時100組轉換參數推估之結果。結果顯示, λ 推估值之變異數隨樣本數增加而減少, 顯著當樣本數增加時 λ 之推估值越穩定, 為一致性推估值(Consistent Estimators), 另, 結果亦顯示於大樣本時轉換後模式之擾動項大都無法通過常態分布檢定。

(3)由模式1至模式8序列之研究結果顯示, 無論是逆轉換序列或未逆轉換之序列, 其推估 λ 後模式模擬之AR階數, 大多能保持原始序列之AR階數。

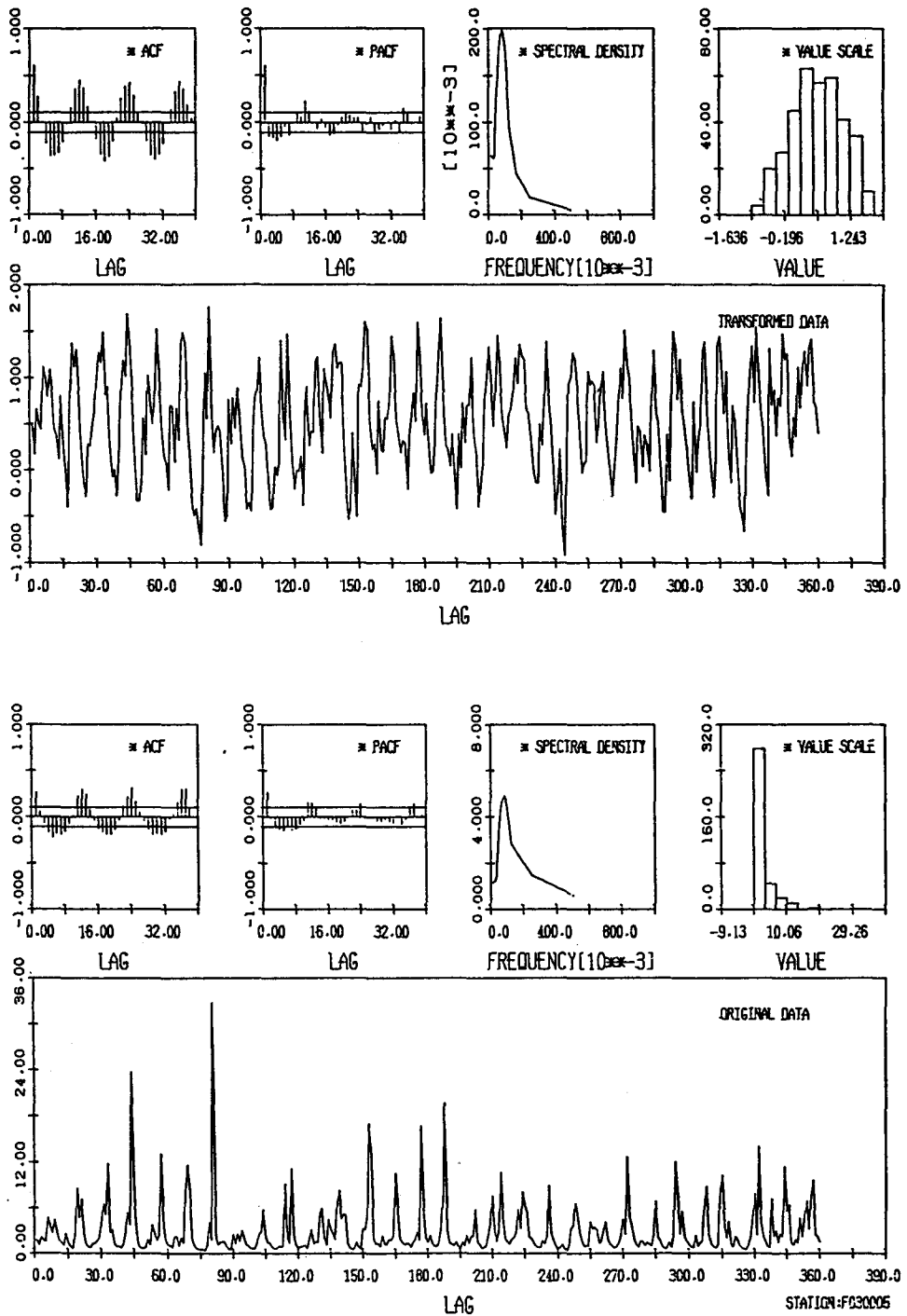


圖2 F030005 (高義站, 淡水河) 月流量基本統計特性((上): 轉換後, (下): 轉換前))

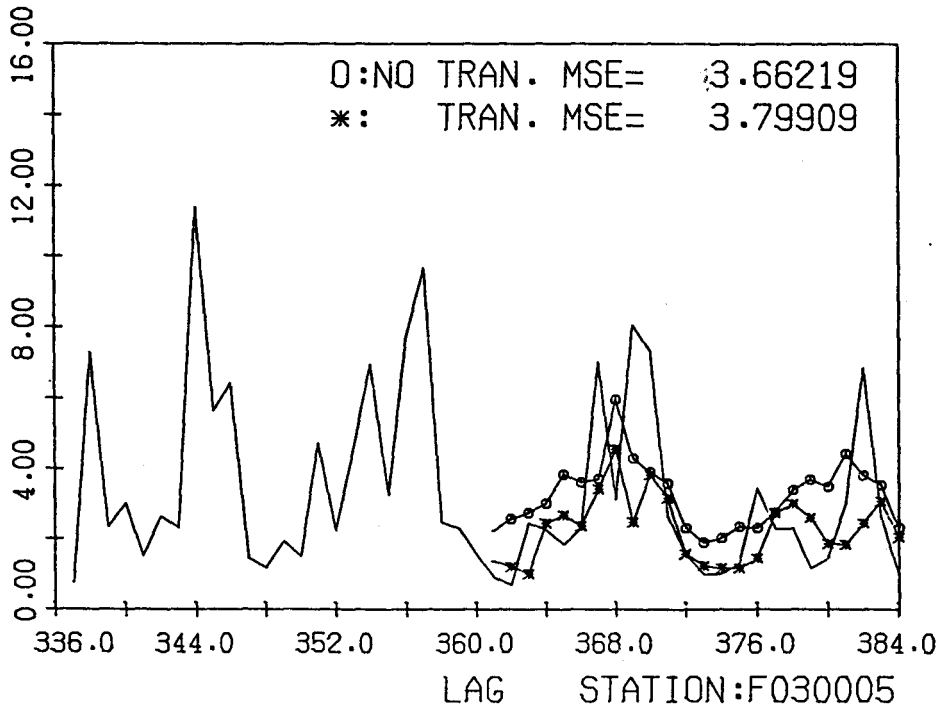


圖3 F030005 (高義站, 淡水河) 轉換前後SAR模式預測結果

(4)表5所示為模式9當 $n=360$ 情形下, 100組合成資料分析之結果, Box-Cox方法所推估100組轉換參數之平均值為0.138, 變異數為0.030, 而本研究所提方法推估100組轉換參數之平均值為0.1535, 變異數為0.029, 此結果顯示兩種方法所推估之結果大致上相同。同時, 由表5研究結果可知, 無論轉換前後合成序列之產生以Bootstrap較能保持原始資料之統計特性, 以三階及四階動差之保有性而言, 轉換後以Bootstrap產生合成序列較轉換前以Bootstrap產生合成序列為佳, 然其變異數較大。

4-2 實測月流量資料：

本研究共計分析26站實測月流量資料, 為了解預測能力, 以前N-24個資料擬合模式, 以最後24個資料驗證預測能力。轉換參數 λ 值分別以本研究所提之方法及Box-Cox方法加以推估, 並比較其差異。同時, 並對傳統方法與Bootstrap重新抽樣法產生合成序列於保有實測資料統計特性作一探討, 其結果與討論分述如下：

(1)圖1所示為淡水河高義站以Box-Cox方法及本研究所提轉換參數推估方法之比較, 表6所示為以Box-Cox方法及本研究所提方法所推估各月流量資料

之轉換參數 λ 值, 其值均為負值, 介於0.0~-0.75之間, 結果顯示其間並無明顯差異, 是故本研究所提之推估方法實為一有效快速之方法。圖2所示為淡水河高義站實測月流量及經 $\lambda = -0.35$ 轉換後之資料圖、ACF、PACF、波譜密度函數及組體圖, 由ACF、PACF及波譜密度函數可知月流量時序經Box-Cox轉換後其週期性較原始序列更強烈, 顯示轉換後SAR模式之模擬較原始資料更容易且精確。另, 由資料組體圖顯示, 轉換後資料較原始資料更趨近對稱分布。

(2)表7所示為部份實測月流量資料於轉換前後之SAR模式及模式適合度結果, 結果顯示轉換前後之SAR模式並無明顯差異, 此結果與圖2所示相同。然由於轉換後週期性更加明顯, 故轉換後模式之 R^2 值較轉換前模式為大。另, 轉換前後之擾動項大都能通過Q-TEST及C-PTEST。同時, 轉換前之擾動項均無法通過常態檢定, 顯示序列為非常態, 然轉換後之擾動項站中計有18站能通過常態檢定, 此結果顯示雖然大部份實測月流量序列為非非常態, 但仍具備常態分布擾動項。

(3)表8結果及圖3顯示, 轉換前之序列於水文年 (11

表3 模式1至模式4於不同真值 λ 時之轉換參數推估結果 (100組)

模式	λ 真 值	λ 推 估 值			未 通 過 常 態 檢 定 之 組 數
		平 均 值	變 異 數	MSE	
1	-0.8	-.71600	.04883	.05540	7
	-0.6	-.54900	.03424	.03650	8
	-0.4	-.37450	.01533	.01583	8
	-0.2	-.17400	.00932	.00990	7
	0.0	.01750	.01148	.01168	7
	0.2	.20550	.00944	.00938	7
	0.4	.41250	.01573	.01573	8
	0.6	.60900	.03310	.03285	8
	0.8	.78450	.04079	.04062	8
2	-0.8	-.72750	.03598	.04087	8
	-0.6	-.57400	.03424	.03650	8
	-0.4	-.37450	.01533	.01583	7
	-0.2	-.18800	.00556	.00565	8
	0.0	.00900	.00573	.00575	7
	0.2	.20150	.00482	.00478	9
	0.4	.39950	.01275	.01263	4
	0.6	.59850	.02911	.02882	6
	0.8	.77700	.03957	.03970	3
3	-0.8	-.75300	.03176	.03365	5
	-0.6	-.54900	.02103	.02150	5
	-0.4	-.38350	.00844	.00863	3
	-0.2	-.18900	.00276	.00285	4
	0.0	.01000	.00283	.00290	4
	0.2	.20300	.00221	.00220	6
	0.4	.40800	.00696	.00695	5
	0.6	.60400	.01726	.01710	5
	0.8	.80100	.02151	.02130	5
4	-0.8	-.75200	.02596	.28000	7
	-0.6	-.57250	.01815	.01873	7
	-0.4	-.38550	.00769	.00783	8
	-0.2	-.18650	.00343	.00358	5
	0.0	.00900	.00411	.00415	7
	0.2	.20200	.00333	.00330	6
	0.4	.41900	.00762	.00790	7
	0.6	.62050	.01910	.01933	8
	0.8	.81300	.02478	.02470	6

表4 模式5至模式8於不同樣本數時 λ 值推估結果(100組)

模式	樣本數	λ 推 估 值		通 過 常 態 檢 定 之 組 數
		平 均 值	變 異 數	
5	60	.39850	.02916	17
	120	.36750	.01153	78
	240	.30100	.00444	100
	360	.27300	.00330	100
6	60	.70550	.02005	5
	120	.67950	.00859	17
	240	.64700	.00615	40
	360	.62800	.00729	58
7	60	.38600	.07819	39
	120	.35900	.03361	87
	240	.30900	.00446	99
	360	.26750	.00542	100
8	60	.64000	.13838	10
	120	.67000	.03152	23
	240	.65850	.00824	46
	360	.64150	.00414	63

表5 模式9資料於轉換前後以傳統方法及Bootstrap重新抽樣法產生合成資料

方 法	平 均 值		變 異 數		偏 態 係 數		峰 度 係 數				
	平均值	變異數	平均值	變異數	平均值	變異數	平均值	變異數			
原 始 資 料	1.166	.022	1.148	.002	3.643	2.485	24.150	608.180			
轉 換 前	傳 統 方 法	1.531	.024	1.205	.005	-.234	.001	.008	.004		
	Bootstrap 重新抽樣法	1.167	.027	1.124	.116	3.252	1.923	18.528	327.050		
轉 換 中	Box-Cox 轉換參數 推估方法	1.202	傳 統 方 法	1.577	.034	1.389	.066	1.689	.052	4.297	1.629
	傳 統 方 法		1.567	.413	3.832	2.201	24.344	441.681			
轉 換 後	本 研 究 之 轉 換 參 數 推 估 方 法	1.198	傳 統 方 法	1.571	.033	1.342	.056	1.634	.064	4.014	1.875
	Bootstrap 重新抽樣法		1.511	.366	3.252	2.054	23.358	399.561			

表6 各實測月流量資料 λ 之推估值

區域	資料編號	站名	Box-Cox方法	本研究所之方法
北區	F020001	秀巒	-.40	-.20
	F030002	玉峰	-.45	-.35
	F030004	菱角	-.50	-.35
	F030005	高義	-.45	-.40
	F030021	三峽	-.05	.00
	F030026	福山	-.25	-.20
	F100013	蘭陽大橋	-.15	-.15
	F130014	上溪	-.05	.05
	F180001	山腳	-.40	-.40
中區	F250003	南湖	-.45	-.30
	F250004	環山合流點	-.35	-.15
	F250006	四季湖	-.45	-.30
	F270014	柑子林	-.45	-.45
	F290009	武界	-.20	-.15
	F290042	桶頭	-.15	-.25
東區	F220013	綠水	-.60	-.50
	F300012	平林	-.75	-.60
	F300037	仁壽橋	-.60	-.50
	F300937	仁壽橋主流	-.40	-.50
	F340011	立山	-.55	-.70
	F400016	延平	-.55	-.50
	F400020	臺東大橋	-.10	.00
南區	F390011	新營	.00	.00
	F510010	新發大橋	-.40	-.35
	F510031	月眉	.00	.05
	F510036	高屏大橋	-.10	-.15

表7 實測月流量資料之分析結果

站名	模 式		模 式 適 合 度				
			αa^2	是否通過 Q Test	是否通過 C-P Test	R^2 R^2	擾動項是否通過 常態檢定
淡水河 菱角 F030004	轉換前	$(1-.197B+.149B^5-.178B^{12}-.162B^{13})$ $(Zt-0.704)=at$.336	YES	YES	.180	NO
	轉換後	$(1-.580B+.212B^4-.178B^{11})$ $(Zt-0.906)=at$.458	NO	YES	.469	YES
蘭陽溪 蘭陽大橋 F100013	轉換前	$(1-.345B-.159B^{11}.289B^{12})$ $(Zt-6.640)=at$	26.388	YES	YES	.335	NO
	轉換後	$(1-.549B+.129B^5-.268B^{12}-.109B^{14})$ $(Zt-1.281)=at$.247	YES	YES	.457	YES
烏溪 柑子林 F270014	轉換前	$(1-.194B+.145B^2-.128B^4-.112B^5)$ $-.131B^{11}-.188B^{12})(Zt-5.468)=at$	33.885	YES	YES	.330	NO
	轉換後	$(1-.610B+.232B^4-.153B-.125B)$ $(Zt-0.786)=at$	0.087	YES	YES	.697	NO
濁水溪 桶頭 F290042	轉換前	$(1-.196B^2+.140B^4-.199B^{11}-.451B^{12})$ $(Zt-1.959)=at$	4.454	YES	YES	.438	NO
	轉換後	$(1-.445B+.229B^5-.163B^{11}-.224B^{12})$ $(Zt+0.420)=at$.426	NO	YES	.785	YES
花蓮溪 仁壽橋 F300037	轉換前	$(1-.255B-.214B^{12})(Zt-3.368)=at$	7.858	YES	YES	.112	NO
	轉換後	$(1-.525B-.153B^5-.100B^{12})$ $(Zt-0.673)=at$	0.071	YES	YES	.346	YES
卑南溪 臺東大橋 F400020	轉換前	$(1-.340B+.149B^5-.255B^{12})$ $(Zt-9.960)=at$	121.793	YES	YES	.281	NO
	轉換後	$(1-.587B+.176B^5-.253B^{11})$ $(Zt-1.355)=at$.468	YES	YES	.620	YES
高屏溪 高屏大橋 F510036	轉換前	$(1-.293^2+.167B^4-.281B^{11}-.322B^{12})$ $(Zt-24.37)=at$	586.889	YES	NO	.450	NO
	轉換後	$(1-.576B+.204B^5-.298B^{11})$ $(Zt-1.980)=at$.282	NO	YES	.761	NO

表8 月流量資料轉換前後SAR模式於濕季、乾季及水文年預測誤差之MSE

站名	濕季(5-10月)		乾季(11-4月)		水文年(11-10月)		
	轉換前	轉換後	轉換前	轉換後	轉換前	轉換後	
						Box-Cox方法	本研究方法
F030001	.181	.213	.054	.047	.117	.130	.117
F030002	2.123	2.437	.406	.310	1.265	1.374	1.361
F030004	.346	.551	.039	.052	.192	.301	.293
F030005	6.149	6.912	1.175	.686	3.662	3.799	3.787
F030021	.497	.610	.158	.119	.327	.365	.361
F030026	6.198	7.450	.591	.300	3.395	3.875	3.844
F100013	26.391	27.001	5.464	4.154	15.928	15.577	15.577
F130014	2.683	2.938	2.517	2.351	2.600	2.644	2.612
F180001	.310	.478	.054	.001	.182	.244	.244
F220013	7.082	7.983	.829	.419	3.953	4.201	4.155
F250003	.117	.097	.067	.027	.092	0.62	.061
F250004	.655	.747	.233	.196	.444	.471	.484
F250006	.099	.091	.118	.076	.109	.083	.083
F270014	18.157	28.092	2.541	2.148	10.349	15.120	15.120
F290009	17.618	17.287	5.275	4.707	11.446	10.997	10.985
F280042	8.928	5.911	.289	.054	4.609	2.982	4.130
F300012	1.123	1.677	.400	.097	.761	.887	.874
F300037	3.168	4.625	1.135	.369	2.152	2.497	2.444
F300937	3.015	4.267	.856	.129	1.935	2.198	2.256
F340011	5.644	7.173	1.073	.146	3.359	3.659	3.403
F390011	1.638	1.642	.044	.009	.841	.826	.826
F400016	21.008	34.864	2.179	.916	11.594	17.890	10.322
F400020	109.879	108.161	6.326	4.385	63.103	56.273	55.867
F510010	38.390	50.411	3.127	.399	20.757	25.405	25.110
F510031	9.878	15.370	2.314	.263	6.096	7.817	7.781
F510036	910.158	991.211	4.831	1.697	477.494	496.454	496.548

-10月)及豐水季節(5-10月)之MSE較轉換後為小,然於乾旱季節(11-4月)轉換後明顯較轉換前之MSE為小,此結果顯示原始資料之SAR模式於豐水季節能獲得較佳之預測結果,而轉換後之SAR模式於乾旱季節能得較佳之預測結果。

- (4)研究結果顯示,轉換前後分別以傳統方法及Bootstrap重新抽樣法產生之合成序列做統計特性保有性比較時發現,轉換前之合成序列以Bootstrap所產生者較能保持實測資料之統計特性。轉換後之資料無論以何種方法產生均無法保持實測資料之統計特性,其乃因為 $Z_t^{(\lambda)}$ $\lambda + 1$ 趨近於零及 λ 值為負時, $(Z_t^{(\lambda)})^{\lambda + 1}$ 趨避 ∞ ,將有特異值出現,影響統計特性之計算,是故,轉換後序之SAR模式無法用以產生合成序列。

五、結 論

根據本研究分析合成資料及實測資料之結果,歸納可得以下之結論:

- (1)以非常態分布擾動項產生合成序列時,序列經過轉換後模式之擾動項幾乎無法通過常態分布檢定。
- (2)模式合成序列統計特性比較上,轉換後模式之合成序列較轉換前為佳,而Bootstrap重新抽樣法又較傳統方法為佳。
- (3)實測月流量序列大都具備常態分布擾動項,轉換後之SAR模式擬合能力較佳。
- (4)轉換後序列之SAR模式於乾旱季節(11-4月)預測結果甚佳,而原始序列之SAR模式於豐水季節(5-10月)預測結果較佳。
- (5)轉換後序列之SAR模式無法用以產生合成序列。
- (6)本研究所提無須經過模式選取的轉換參數推估方法,能夠有效推估轉換參數,且能節省較多之計算時間。

謝 誌

本研究承蒙行政院國科會NSC-82-0115-E-032-039研究計劃之經費補助,特此致謝。

參考文獻

- 1.虞國興、陳惠龍、「合成資料之繁衍—Bootstrap重新抽樣法」,農業工程學報、第38卷、第3期、pp.14-26、民國81年12月。
- 2.歐立中,「Box-Cox指數轉換之研究及其應用於月流量時序之分析」,淡江大學水資源及環境工程研究所碩士論文,民國82年6月,PP.74
- 3.Box, G.E.P. and Cox, D.R. (1964), "An analysis of transformation", Journal of Royal Statistical Society B, 26, 211-243.
- 4.Box, G.E.P. and Jenkins, G.M. (1976), Time Series Analysis forecasting and control. Holden-Day, San Francisco.
- 5.Draper, N.R. and Cox, D.R. (1969), "On Distribution and Their Transformation to Normality", Journal of Royal Statistical Society B, 31, 472-476.
- 6.Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife", Ann.Statist. 7, 1-26.
- 7.Hopwood, W.S., Mckeown, J.C. and Newbold, P. (1984), "Time Series Forecasting Models Involving Power Transformations", Journal of Forecasting, Vol. 3, 57-61.
- 8.Nelson, C.R. (1973), Applied Time Series Analysis, Holden-Day, San Francisco.
- 9.Wei, William W.S. (1990), Time Series Analysis, Addison-Wesley Publishing Company. pp. 82-84
- 10.Yevjevich, V. (1972), Stochastic Processes in Hydrology. Water Resources Publications, Fort Collins, Colorado, U.S.A.
- 11.Yu, G.H. and Lin, Y.C. (1991), "A Methodology for Selecting Subset Autoregressive Time Series Models", Journal of Time Series Analysis, Vol. 12, No.4, pp.363-373.

收稿日期:民國82年6月22日

接受日期:民國82年7月21日