合成資料之繁衍—Bootstrap重新抽樣法

Generation of Synthetic Data—Bootstrap Resampling Technique

淡江大學水資源及環境
工程研究所教授
虞　國　興
Gwo-Hsing Yu

淡江大學水資源及環境
工程研究所研究助理
陳　惠　龍
Huey-Long Chen

摘　　要

傳統上，非常態水文序列須經過轉換爲常態水文序列，或經由判定水文序列擾動值之分布，方能以時間序列模式繁衍水文合成資料。然而由於資料轉換失眞或擾動值之分布誤判，同時，傳統上合成資料之產生以能保存資料之樣本統計特性爲依據，然樣本統計特性與理論統計特性可能存在差異，故所繁衍之合成資料將不具資料之理論特性，因此本論文提出一種方法，利用 Bootstrap resampling 之抽樣方法，配合ARIMA模式可直接繁衍未知分布之水文合成資料，而資料不須轉換爲常態。研究結果顯示，本論文所提方法可保存資料之理論統計特性，而非僅保存該組資料之樣本統計特性。

ABSTRACT

One of the conventional approaches to analyze the non-Gaussian time series is to transform it to be a Gaussian time series or to detect the distribution of the noise. The synthetic data are then generated based on the time series model. However, the statistical characteristics of time series may be distorted by using the transformation approach and by drawing a wrong inference of the distribution of noise from the data. Meanwhile, the synthetic data were generated to preserve the statistical characteristics of the samples. The statistical characteristics of samples and population are usually different. Therefore, the statistical characteristics of population will not be preserved by the synthetic data. A method for generating the synthetic data by employing Bootstrap resampling technique and ARIMA models was proposed in this study. The results indicate that the statistical chasacteristic of population instead of samples were preserved by this proposed method.

一、緒　　論

所有水文歷程都屬於序率歷程 (Stochastic process)，而水文序列可視爲以時間爲變數之序列，故可利用時間序列分析 (Time series analysis) 方法，根據過去的水文資料，繁衍過去或預測未來可能發生之水文歷程。

時間序列可分爲常態過程(Normal process

)及非常態過程 (Non-normal process)，二者又可再分爲定常性時間序列(Stationary time series) 及非定常性時間序列 (Nonstationary time series)。一般水文時間序列多屬於非定常性非常態時間序列，所以傳統上在繁衍合成資料 (Synthetic data) 時，常須藉由轉換 (Transformation) 方法將非定常性非常態水文時間序列轉換爲常態定常性時間序列，以自廻歸—移動平均模式 (Autoregressive moving average model，簡稱ARMA模式) 分析轉換後之序列模式並推求參數，再將轉換後之序列模式所繁衍之合成資料反轉換爲具原始資料統計特性之合成資料。然而模式建立時，當資料因轉換而造成失眞，則模式建立後，經反轉換所得之合成資料將不具有原始資料之眞正統計特性。

一般常態時間序列可分爲定常性時間序列和非定常性時間序列。定常性時間序列可利用 ARMA 模式，予以分析。非定常性時間序列又可分爲平均值不固定之非定常性序列，二階動差不固定之非定常性序列，及兩種性質皆有之非定常性序列。Box and Jenkins(1970) 將平均值不固定之非定常性序列定義爲均齊性非定常性序列(Homogeneous nonstationary time series) 並且提出自廻歸—整合—移動平均模式(Autoregressive Integrated Moving Average model，簡稱ARIMA模式)，以差分方法 (Difference)將均齊性非定常性序列轉換成定常性序列，再利用 ARMA 模式予以分析。

對於非常態時間序列，傳統上利用轉換方式如除勢模式 (Detrended model) 或 Box-Cox 轉換 (Box and Cox, 1964) 將資料轉換成常態。由於水文資料常被視爲對數常態分布 (Lognormal distribution) 或加瑪分布 (Gamma distribution)，水文學者 Yevjevich (1972) 直接以邊際率分布 (Marginal probability distribution) 爲加瑪分布及對數常態分布之 AR(1) 模式模擬水文資料，而不須將資料轉換爲常態。O'Connel and Jones(1979) 利用以對數常態分布爲擾動值 (Noise 或 Disturbance) 之 AR 模式模擬河流量資料。Li and McLeod (1988) 以最大概似法 (Maximum likelihood estimation) 推估非常態分布擾動值之 ARMA模式之參數。上述不須轉換資料之方法雖可模擬非常態時間序列，然而其模式參數之推求，及合成資料之繁衍，必須在序列分布或擾動值分布已知之條件下。由於實際水文資料之分布未知，其擾動值之分布亦未知，因此上述方法於實際應用時，必須先判斷序列或擾動值之分布，然而當資料分布判斷錯誤時，則所繁衍之合成資料將不具有原始資料之眞正分布特性。

由於以傳統方法繁衍合成資料容易因資料轉換失眞或擾動值分布判定錯誤，同時，傳統上合成資料之產生以能保存資料之樣本統計特性爲依據，然樣本統計特性與理論統計特性可能存在差異，而造成所繁衍之合成資料不具有原始資料之理論統計特性，故本論文提出一種不須轉換原始資料分布的方法繁衍合成資料，即利用ARIMA模式模擬非定常性非常態水文時間序列之相關性，再根據 Bootstrap resampling 抽樣方法繁衍具原始資料分布特性之合成資料。

本論文利用合成資料及臺灣水文實測資料比較傳統繁衍合成資料方法與本研究所提繁衍合成資料方法之優劣。

以下爲本論文之研究主題：

1. 比較以均勻分布 (Uniform distribution)、常態分布 (Normal distribution)、對數常態分布 (2-parameter log-normal distribution) 及加瑪分布 (Gamma distribution) 爲擾動值之時間序列的平均值、變異數、偏態係數 (Skewness)、峰度 (Kurtosis)、序列相關性及利用動差法 (Method of moments) 與最小二乘方法 (Least squares method) 所推估之模式參數。
2. 比較由理論分布抽樣之擾動值與 Bootstrap抽樣之擾動值所構成序列之相關性及統計特性。
3. 比較本論文所提繁衍資料方法與除勢模式 (Detrended model) 繁衍資料方法對不同擾動值分布之序列統計特性的保存能力。
4. 比較因擾動值分布誤判所繁衍合成資料之統計特性與原始資料眞正統計特性之差異。

本論文之大綱如下：第二節爲有關時間序列及 Bootstrap resampling 抽樣方法之理論基礎，第三節爲討論合成資料及實測資料之分析結果，最後，第四節爲結論。

二、理論基礎

本章將對時間序列模式，Bootstrap resampling 抽樣理論及合成資料繁衍方法分別說明如下：

2-1 離散線性序率過程 (Discrete linear-stochastic process)：

若時間序列$\{Z_t\}$屬於一離散線性序率過程，則Z_t可表示如下：

$$Z_t=\mu+a_t+\psi_1a_{t-1}+\psi_2a_{t-2}+\cdots\cdots$$

其中：

μ＝歷程Z_t之平均值

ψ_i＝模式參數

$\{a_t\}$＝平均值爲零，變異數爲σ_a^2之IID(Independent and identical distribution)分佈。

Z_t可視爲由過去及現在之擾動值a_t所構成之線性組合。Z_t之期望值如下：

$$E[Z_t]=\mu+E[a_t+\psi_1a_{t-1}+\psi_2a_{t-2}+\cdots\cdots]$$

若Z_t存在，則必須滿足條件如下：

$\sum_{i=0}^{\infty}\psi_i=K$，其中

$\psi_0=1$，$-\infty<K<\infty$

則$E[Z_t]=\mu$，且Z_t之變異數C_0及自相關變異數C_k爲：

$$C_0=Var[Z_t]=E[(Z_t-\mu)^2]=\sigma_a^2\sum_{i=0}^{\infty}\psi_i^2$$

$$C_k=Cov[Z_tZ_{t-k}]=E[(Z_t-\mu)(Z_{t-k}-\mu)]$$

$$=\sigma_a^2\sum_{i=0}^{\infty}\psi_i\psi_{i+k}$$

當$\sum_{i=0}^{\infty}\psi_i^2$及$\sum_{i=0}^{\infty}\psi_i\psi_{i+k}$存在，則$Z_t$之平均值、變異數及自相關變異數 (autocovariance)不隨時間t而改變，故$\{Z_t\}$滿足定常性。

由上述證明過程可知，只要擾動值a_t之平均值及變異數σ_a^2爲定值，則a_t之分布並不影響$\{Z_t\}$滿足定常性之條件。

2-2 ARI MA模式：

(1)非季節性模式 (Non-seasonal model)

非季節時間序列$\{X_t\}$之ARIMA(p, d, q)模式如下：

若$Z_t=(1-B)^d(X_t-\mu)$且$\{Z_t\}$滿足 ARMA(p, q)模式，

即 $Z_t=\Phi_1Z_{t-1}+\Phi_2Z_{t-2}\cdots\cdots+\Phi_pZ_{t-p}+a_t-\theta_1a_{t-1}-\theta_2a_{t-2}\cdots\cdots-\theta_qa_{t-q}$

其中：

d＝非季節性差分之階數

$\{\Phi_1, \Phi_2, \cdots\cdots, \Phi_p\}$＝非季節性自迴歸參數

$\{\theta_1, \theta_2, \cdots\cdots, \theta_q\}$＝非季節性移動平均參數

或寫成：

$\Phi(B)[(1-B)^d(X_t-\mu)]=\theta(B)a_t$，

故建立 ARIMA (P, d, q) 模式必須推估 p+q+3 個參數 (p個Φ，q個θ及d，μ，σ_a^2)。

(2)季節性模式 (Seasonal model)

季節性時間序列 $\{X_t\}$ 之 ARIMA(p, d, q)×(P, D, Q)$_s$ 模式定義爲：

$$\Phi(B)\Phi(B^s)[(1-B^s)^D(1-B)^d(X_t-\mu)]$$
$$=\theta(B)\theta(B^s)a_t$$

其中：

s＝季節長度，如：月資料 s＝12

D＝季節性差分之階數

μ＝歷程X_t之平均值

$\Phi(B^s)=1-\Phi_1B^s-\Phi_2B^{2s}-\cdots\cdots-\Phi_PB^{s}$ 階次爲P之季節自迴歸運算子

$\theta(B^s)=1-\theta_1B^s-\theta_2B_2^s-\cdots\cdots-\theta_QB^s$ 階次爲Q之季節移動平均運算子

所以，建立 ARIMA(p, d, q)× (P, D, Q)$_s$ 模式必須推估 p+q+P+Q+4 個參數 (p個Φ，q個θ，P個Φ，Q個θ 及 d，μ，σ_a^2，D)。例：模式 ARIMA(1, 1, 2)×(0, 1, 1)$_{12}$ 可寫爲 $(1-\Phi_1B)[(1-B)(1-B^{12})(X_t-\mu)]$
$=(1-\theta_1B-\theta_2B^2)\times(1-\theta_1B^{12})a_t$

2-3 Bootstrap resampliog 抽樣理論：

Efron (1979) 所提之 Bootstrap resampling 理論如下：

設 $W_1, W_2, \cdots\cdots, W_n$ 爲彼此獨立，且爲同一未知分布F之隨機變數，令w爲由該組隨機變數 (Random variable) 所構成之隨機樣本，即 $W=(W_1, W_2, \cdots\cdots, W_n)$；再令w爲W之一觀組測資料，即 $w=(w_1, w_2, \cdots\cdots, w_n)$ 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 distribwtion)。

Bootstrap resampling 之方法如下：

1.利用觀測資料w建立$\hat{F}$之樣本機率分布 (Sample probability distribution) $\hat{F}$，即令

F之樣本空間 (Sample space) 爲 $\{w_1, w_2, \cdots\cdots, w_n\}$，而且使每一數值 $w_1, w_2, \cdots\cdots, w_n$ 均具有 1/n 之機率分布。

2.依F之機率分布隨機替換抽樣 (Sampling with replacement)，取得樣本長度爲n之 Bootstrap 樣本 $W^*=(W_1^*, W_2^*, \cdots\cdots, W_n^*)$。

3.則 Bootstrap 分布 $\alpha^*(W^*, \hat{F})$ 近似於 $\alpha(W, F)$ 之抽樣分布。

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 爲 $n \times u_1$ 之整數部分，則 $w_1^* = w_{k1}$，再重複以上步驟，由原樣本空間 $\{w_1, w_2, \cdots\cdots, w_n\}$ 抽樣得 $w_i^* = w_{ki}$，$i=1, 2, \cdots\cdots, n$。

2-4 本研究所提合成資料繁衍方法：

本論文利用 Bootstrap resampling 繁衍未知分布時間序列之合成資料，其方法如下：

1.判斷模式及推估參數：由於實際水文序列其擾動值 a_t 之分布未知，因此模式之參數無法以最大概似法推估，但可利用最小二乘方法或動差法推估。

2.將實際水文序列 $\{X_t\}$ 代入模式可求得原始序列之擾動值 $(a_1\ a_2, \cdots\cdots, a_n)$。

3.利用 Bootstrap resampling 由原始序列之擾動值 $(a_1, a_2, \cdots\cdots, a_n)$ 中隨機抽取所需繁衍資料個數n之擾動值 $(a_1', a_2', \cdots\cdots, a_n')$。

4.再將所抽出之擾動值代入原來序列模式，即可繁衍合成資料 $(X_1', X_2', \cdots\cdots, X_n')$。

2-5 傳統上利用除勢模式繁衍合成資料之方法：

傳統上，對於非常態或非定常性之時間序列，可利用除勢模式將序列轉換爲常態定常性，再以 ARMA 模式分析。

利用除勢模式繁衍合成資料之步驟如下：

1.繪製時間序列資料分佈圖。

2.去除序列 $X_{v,h}$ 之趨勢 (Trend) 或週期性；即

$$Y_{v,h} = \frac{X_{v,h} - \mu_h}{\sigma_h}$$

μ_h ＝週期性平均值 (Periodic mean)

σ_h ＝週期性標準偏差 (Periodic standard deviation)

v ＝年份

h ＝年份內之時間距 (Time interval within the year)　如：日、月等

3.建立 $Y_{v,h}$ 之 ARMA(p, q) 模式。

4.利用所建立之 ARMA(p, q) 模式繁衍合成資料 $Y'_{v,h}$，再將 $Y'_{v,h}$ 反轉換爲 $X'_{v,h}$ (如下式)，即爲$X'_{v,h}$之合成資料。

$$X'_{v,h} = \mu_h + Y'_{v,h} \times \sigma_h$$

三、結果與討論

本論文所使用之資料包含合成資料、十組實測資料及臺灣月流量資料。本節將根據各研究主題所使用之資料及其分析結果 (由於篇幅所限，僅列出部份結果) 分別討論如下：

3-1 序列相關性、統計特性及參數推估與擾動值分布之關係：

本節之主要研究目的在於比較以均勻、常態、對數常態及加瑪分布爲擾動值之時間序列的平均值、變異數、偏態係數 (Skewness Coefficient)、峰度 (Kurtosis)、序列相關性，及利用動差法與最小二乘方法所推估之模式參數。

(1)所使用之合成資料：

本節所使用之合成資料模式如下：

1.1: $(1-\Phi B)X_t = a_t$，$\Phi = 0., 0.8, 1.0$

1.2: $(1-B)X_t = 0.5 + a_t$

1.3: $(1-\Phi B^{12})X_t = a_t$，$\Phi = 0.8, 1.0$

1.4: $(1-0.8B)(1-0.8B^{12})X_t = a_t$

1.5: $X_t = (1-\theta B)a_t$，$\theta = 0.5, 0.8$

1.6: $(1-0.8B)X_t = (1+0.4B)\ a_t$

所使用之合成資料樣本數爲200。模式 1.1 爲 AR(1) 模式其中 $\Phi = 1.0$ 時爲隨機步 (Random walk) 模式，模式1.2爲具趨勢之非定常性模式，模式1.3 爲具週期12之定常性及非定常性模式，模式 1.4爲 AR(13) 模式，模式1.5爲 MA(1) 模式，模式1.6爲 ARMA(1.1) 模式。

a_t (平均值爲 0，標準差爲 1) 由均勻分布、常態分布、對數常態分布及加瑪分布所產生，各分布之形式如下：

1.均勻分布：

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{其它} \end{cases}$$

平均值爲 (a+b)/2

變異數爲 $(b-a)^2/12$

本研究利用參數 $a=0$，$b=1$ 產生均勻分布之 a_t，再以理論平均值及標準差標準化。

2.常態分布：

$$f(x)=\frac{1}{\sigma(2\pi)^{1/2}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

平均值爲 μ

變異數爲 σ^2

本研究利用參數 $\mu=0$，$\sigma=1$ 產生標準常態之 a_t。

3.對數常態分布：

$$f(x)=\frac{1}{x\sigma_n(2\pi)^{1/2}}\exp\left[-\frac{1}{2}\left(\frac{\ln x-\mu_n}{\sigma_n}\right)^2\right]$$

平均值爲 $\exp(\mu_n+\frac{1}{2}\sigma_n^2)$

變異數爲 $\mu^2_n(\exp(\sigma^2_n)-1)$

本研究利用參數 $\mu_n=0$，$\sigma_n=1$ 產生對數常態之 a_t，再以理論平均值及標準差標準化。

4.加瑪分布：

$$f(x)=\frac{1}{\Gamma(a)}x^{a-1}\exp(-x)$$，其中

$$\Gamma(a)=\int_0^{\infty}x^{a-1}\exp(-x)dx$$

其平均值及變異數均分別爲 a

本研究利用參數 $a=5$ 產生加瑪分布之 a_t，再以理論平均值及標準差標準化。

(2)結果：

將上述四種分布之擾動值代入上述各合成資料模式（AR、MA、ARMA及ARI），各產生100組序列資料，其100組資料平均之統計特性及相關性如表1所示，而利用動差法與最小二乘方法所推估之模式參數如表2所示。

表1 $(1-0.8B)(1-0.8B^{12})Z_t=a_t$ 100 組資料之平均統計特性表

a_t之分布	均勻	對數常態	加瑪	常態
平均值	-.245	.091	-.038	-.170
變異數	5.995	6.376	6.415	6.088
偏態係數	-.018	.707	.177	-.021
峰度	2.669	3.752	2.769	2.682
ACF				
1	.716	.731	.735	.718
2	.507	.528	.542	.511
3	.367	.383	.392	.372
4	.269	.284	.291	.277
5	.208	.215	.227	.216
6	.178	.187	.196	.185
7	.168	.176	.190	.176
8	.190	.201	.218	.199
9	.246	.256	.276	.254
10	.338	.346	.375	.346
11	.480	.481	.502	.486
12	.675	.667	.687	.681
13	.454	.463	.484	.462
14	.296	.312	.339	.307
15	.194	.206	.225	.208
16	.121	.135	.152	.137
17	.075	.082	.105	.090
18	.053	.060	.081	.067
19	.043	.048	.075	.056
20	.061	.064	.095	.074
PACF				
1	.716	.731	.735	.718
2	-.021	-.026	-.011	-.018
3	.009	.012	-.012	.010
4	-.003	.009	.005	-.002
5	.025	.007	.028	.028
6	.024	.044	.037	.024
7	.059	.041	.065	.063
8	.075	.089	.088	.082
9	.126	.122	.141	.130
10	.153	.165	.179	.153
11	.262	.257	.237	.264
12	.368	.365	.373	.373
13	-.540	-.544	-.544	-.538
14	-.008	-.002	-.007	-.002
15	-.019	-.010	-.005	-.014
16	-.024	-.006	-.007	-.024
17	-.012	-.011	-.014	-.018
18	-.008	-.022	-.016	-.007
19	-.017	-.004	-.005	-.015
20	-.001	-.015	-.016	-.001

結果顯示，在相同序列模式下，不同分布之擾動值所產生之序列，其偏態係數、峰度有明顯的差異，但相關係數之差異很小，而當序列爲定常性，序列平均值及變異數並無明顯之差異。由此可知，時間序列資料之偏態係數、峰度基受擾動動值之分布的影響，但資料之相關性只與序列模式影響，而不受擾動值之分布影響。利用動差法或最小二乘方法所推估之模式參數，在相同序列模式下，不同擾動值分布所造成的影響很小。

3-2 由理論分布抽樣之擾動值與 Bootstrap 抽樣之擾動值所構成序列之比較：

本節之主要研究目的在於比較由理論分布抽樣之擾動值與 bootstrap 抽樣之擾動值所構成序列之相關性及統計特性。

(1)所使用之合成資料：

2.1： $(1-\Phi B)X_t=a_t$, $\Phi=0.,0.8$

2.2： $(1-B)X_t=a_t$

2.3： $(1-B)X_t=0.5+a_t$

2.4： $(1-\Phi B^{12})X_t=a_t$,$\Phi=0.8,1,0$

2.5： $X_t=(1-\theta B_t)a_t$, $\theta=0.5,0.8$

2.6： $(1-0.8B)X_t=(1+0.4B)a_t$

所使用之合成資料樣本數爲200。 a_t 之分布如3-1節所示。

(2)結果：

將3-1 節所述四種分布之擾動值代入上述各合成資料模式 (AR、MA、ARMA及ARI) ，產

表2 動差法及最小二乘方法推估之模式參數 $(1-\phi B^{12})Z_t=a_t$

ϕ	方法	均勻	對數常態	加瑪	常態
1.000	動差法	.751	.759	.778	.745
	最小二乘方法	.994	.996	.996	.994
.950	動差法	.782	.793	.795	.778
	最小二乘方法	.938	.938	.940	.938
.800	動差法	.680	.683	.689	.682
	最小二乘方法	.784	.778	.792	.784
.500	動差法	.423	.422	.432	.422
	最小二乘方法	.481	.475	.496	.479
.000	動差法	-.014	-.018	-.002	-.018
	最小二乘方法	-.011	-.016	.001	-.016

表3 $(1-B)Z_t=a_t$ 原始資料與合成資料之平均統計特性表

a_t之分布	均勻	BOOTSTRAP	對數常態	BOOTSTRAP	加瑪	BOOTSTRAP	常態	BOOTSTRAP
平均值	-2.9591	-4.5046	1.3348	2.5039	-1.7851	5.3502	-2.4178	-2.1288
變異數	34.0541	52.2842	30.5936	48.1427	32.2381	51.4685	34.1817	53.5394
偏態係數	-.0668	.0304	-.0056	-.0505	.0270	-.0732	-.1041	.0595
峰度	2.2943	2.3504	2.4154	2.3825	2.3626	2.2673	2.2994	2.3509
ACF								
1	.9633	.9662	.9630	.9660	.9633	.9670	.9633	.9647
2	.9279	.9340	.9265	.9325	.9277	.9360	.9277	.9311
3	.8947	.9027	.8909	.9000	.8927	.9054	.8944	.8988
4	.8628	.8727	.8564	.8698	.8583	.8751	.8624	.8676
5	.8305	.8440	.8222	.8403	.8253	.8458	.8305	.8380
6	.7986	.8163	.7896	.8120	.7929	.8174	.7989	.8098
7	.7675	.7902	.7574	.7838	.7619	.7895	.7681	.7833
8	.7370	.7649	.7260	.7560	.7323	.7624	.7376	.7570
9	.7068	.7404	.6958	.7289	.7039	.7363	.7075	.7309
10	.6774	.7167	.6659	.7027	.6762	.7109	.6778	.7059
PACF								
1	.9633	.9662	.9630	.9660	.9633	.9670	.9633	.9647
2	-.0062	-.0005	-.0165	-.0149	-.0101	.0003	-.0087	-.0013
3	.0018	-.0069	-.0072	-.0055	-.0132	-.0114	.0050	-.0082
4	-.0004	-.0051	-.0117	.0006	-.0123	-.0162	-.0004	-.0086
5	-.0155	-.0032	-.0177	-.0102	.0029	-.0041	-.0094	.0005
6	-.0207	-.0070	-.0018	-.0061	-.0159	-.0131	-.0203	-.0065
7	-.0045	.0006	-.0113	-.0096	-.0054	-.0117	-.0066	-.0008
8	-.0099	-.0046	-.0082	-.0131	-.0045	-.0156	-.0103	-.0111
9	-.0191	-.0092	-.0066	-.0104	-.0019	-.0024	-.0175	-.0131
10	-.0157	-.0056	-.0158	-.0082	-.0145	-.0095	-.0200	-.0066

生1組原始序列資料，再由同一組擾動值經 Bootstrap resampling 所得之擾動值代入同一時間序列模式，產生1組合成序列資料。重複上述步驟100 次，其100 組原始資料及合成資料之平均統計特性及相關性如表3及表4所示。

結果顯示，除了模式2.2 之外，對於其他模式，不同分布之擾動值，由原始序列及由同一組擾動值經 Bootstrap resampling 所產生合成序列的平均值、變異數及相關係數之差異都很小；模式2.2爲隨機步模式，結果顯示該模式經 Bootstrap resampling 所繁衍合成序列的變異數明顯較原始序列爲大。同一結果顯示，當原始序列之擾動值爲對常態分布（具較大偏態）時，則 Bootstrap resampling 所繁衍合成序列在偏態係數及峰度上略有低估；而當原始序列之擾動值爲均勻分布、常態分布（不具偏態）及加瑪分布（具較小偏態）時，則 Bootstrap resampling 所繁衍合成序列在偏態係數及峰度上沒有明顯之差異。

由此可知，除了隨機步模式外，由 Bootstrap resampling 所繁衍相同模式之合成序列可保存原始序列之統計特性及相關性；當原始序列擾動值具較大偏態，則 Bootstrap resampling所繁衍合成序列其偏態係數及峰度略有低估。

3-3 本論文所提繁衍合成資料方法與除勢模式繁衍資料方法之比較：

本節之主要研究目的在於比較本論文所提繁衍資料方法與除勢模式繁衍資料方法對不同擾動值分布之序列統計特性的保存能力。

(1)所使用之合成資料及研究結果：

由於一般水文資料常具有週期，所以本節使用具週期12之 SAR (12) 合成資料模式如下：

$(1-\Phi B^{12})X_t=a_t,\ \Phi=0.8,1.0$

所使用之合成資料樣本數爲 200。a_t之分布如3-1節所示。

本研究將3-1 節四種分布擾動值代入上述合成資料模式（具週期性），產生100 組原始序列資料，求出每一組資料之樣本統計特性，再根據每一組原始序列經本論文所提方法及除勢模式各繁衍 100 組合成資料，分別求出其100 組平均統計特性，然後根據每一組原始序列樣本統計特性比較兩種方法

表4 $(1-0.8B)Z_t=(1+0.4B)a_t$ 原始資料與合成資料之平均統計特性表

a_t之分布	均勻	BOOTSTRAP	對數常態	BOOTSTRAP	加瑪	BOOTSTRAP	常態	BOOTSTRAP
平均值	-.0372	-.0536	.0133	.0413	.0409	.0262	-.0071	-.0143
變異數	4.5622	5.0304	4.8644	5.6119	4.7378	4.8076	4.5810	5.1472
偏態係數	.0553	-.0057	1.4887	1.4194	.2714	.2989	.0920	.0179
峰度	2.7190	2.7368	6.3727	5.9396	2.8347	2.9863	2.9110	2.9513
ACF								
1	.8592	.8669	.8635	.8622	.8641	.8639	.8582	.8682
2	.6609	.6782	.6683	.6644	.6689	.6688	.6574	.6814
3	.5098	.5301	.5143	.5064	.5120	.5138	.5044	.5345
4	.3942	.4127	.3913	.3843	.3882	.3879	.3886	.4178
5	.3013	.3212	.2922	.2900	.2921	.2880	.2978	.3258
6	.2242	.2475	.2157	.2178	.2168	.2113	.2232	.2512
7	.1626	.1864	.1551	.1622	.1582	.1518	.1630	.1889
8	.1122	.1383	.1069	.1170	.1158	.1054	.1132	.1400
9	.0695	.1002	.0688	.0800	.0854	.0685	.0706	.1023
10	.0337	.0665	.0388	.0501	.0622	.0400	.0334	.0702
PACF								
1	.8592	.8669	.8635	.8622	.8641	.8639	.8582	.8682
2	-.2982	-.3021	-.3084	-.3140	-.3113	-.3093	-.3030	-.3010
3	.1153	.1084	.1079	.0967	.0967	.1016	.1175	.1072
4	-.0455	-.0515	-.0613	-.0408	-.0476	-.0678	-.0441	-.0499
5	.0005	.0189	.0021	.0075	.0128	.0137	.0072	.0155
6	-.0230	-.0259	-.0085	-.0115	-.0215	-.0166	-.0237	-.0247
7	.0015	-.0055	-.0130	-.0065	.0006	-.0034	.0014	-.0093
8	-.0214	-.0047	-.0091	-.0247	.0036	-.0158	-.0220	-.0017
9	-.0159	-.0127	-.0105	-.0089	-.0002	-.0106	-.0168	-.0077
10	-.0230	-.0244	-.0100	-.0112	-.0070	-.0058	-.0275	-.0215

所繁衍合成資料對每一組原始序列樣本統計特性的保存能力，結果如圖 1 所示；再根據原始序列 100 組之平均統計特性比較兩種方法對原始序列之平均統計特性的保存能力，結果如圖 2 所示。

由圖 1 中可以看出，除勢模式所繁衍之合成資料其平均統計特性與每一組原始序列之樣本統計特性，皆對應成一條45° 的直線，顯示除勢模式能保存每一組原始序列之樣本統計特性；本論文所提方法繁衍之合成資料其平均統計特性之平均值及變異數亦能保存每一組原始序列之樣本統計特性，然而偏態係數及峰度則較不受每一組原始序列之樣本統計特性影響，且由圖 2 可以看出，本論文所提方法繁衍之合成資料其平均統計特性之偏態係數及峰度明顯較趨近原始資料100 組之平均統計特性（即其理論特性），尤其當資料爲定常性（即 $\Phi=0.8$）。

由此可知，除勢模式所繁衍之合成資料只能保存每一組原始序列之樣本統計特性，而非原始序列之理論特性。本論文所提之繁衍方法能反應原始序列其偏態係數及峰度之平均統計特性，即資料之理論特性，但其合成資料之平均值及變異數則受每一組原始序列之樣本統計特性影響。

(2)所使用之實測資料及研究結果：

本節利用七個臺灣水文站月流量實測資料，比較上述兩種方法對於原始資料統計特性的保存能力。根據每一組實測資料合繁衍100 組合成資料，求出其平均統計特性，結果如表 5 所示。

結果顯示，本論文所提之繁衍方法在乾溝、菱角、義里、桶頭四站對原始資料統計特性之保存能力，明顯較除勢模式爲佳；對於其他三站則兩種方法對原始資料統計特性之保存能力互有優劣。

綜合本節合成資料及實測資料之研究結果，本論文所提繁衍資料方法能反應原始資料其偏態係數及峰度之理論特性，而除勢模式則只能反應每一組原始資料之樣本統計特性。由於一般水文資料其理論統計特性未知，而觀測資料之樣文統計特性與理論統計特性存在差異，因此利用本論文所提繁衍資料方法可繁衍比較接近理論統計特性之合成資料。

3-4 誤判擾動值分布繁衍合成資料之結果：

合成資料之繁衍，亦可根據原始資料擾動值分布之判定，再由判定之分布抽樣獲得所需擾動值繁衍合成資料。然而，當擾動值之分布判定錯誤，則合成資料將不具有原始資料之特性。本節之主要研究目的在於比較因擾動值分布誤判所繁衍合成資料之統計特性與原始資料眞正統計特性之差異。

表5 台灣月流量資料之統計特性及合成資料之平均統計特性

流量資料	樣本數	平均值	變異數	偏態係數	峰度
淡水河乾溝站		.783	.325	2.021	9.659
Bootstrap	120	.764	.330	1.995	9.408
Detrended		.813	.336	.665	4.588
淡水河菱角站		.210	.048	3.783	23.544
Bootstrap	120	.198	.043	3.737	22.878
Detrended		.204	.045	1.707	9.143
東港溪潮州站		.590	.432	1.824	6.658
Bootstrap	120	.595	.399	1.077	5.303
Detrended		.634	.499	1.336	4.045
高屏溪九曲堂站		7.367	97.072	2.121	8.119
Bootstrap	120	7.287	80.759	2.151	9.201
Detrended		7.243	96.120	1.337	5.206
烏溪柑子林站		1.593	4.162	2.961	13.135
Bootstrap	120	1.449	2.885	2.347	9.939
Detrended		1.770	5.523	2.084	8.198
大安溪義里站		.786	1.291	2.523	10.152
Bootstrap	120	.805	1.247	2.136	8.437
Detrended		.799	1.412	1.487	5.706
濁水溪桶頭站		.560	.679	2.255	8.388
Bootstrap	120	.561	.677	2.338	8.982
Detrended		.569	.705	1.726	5.496

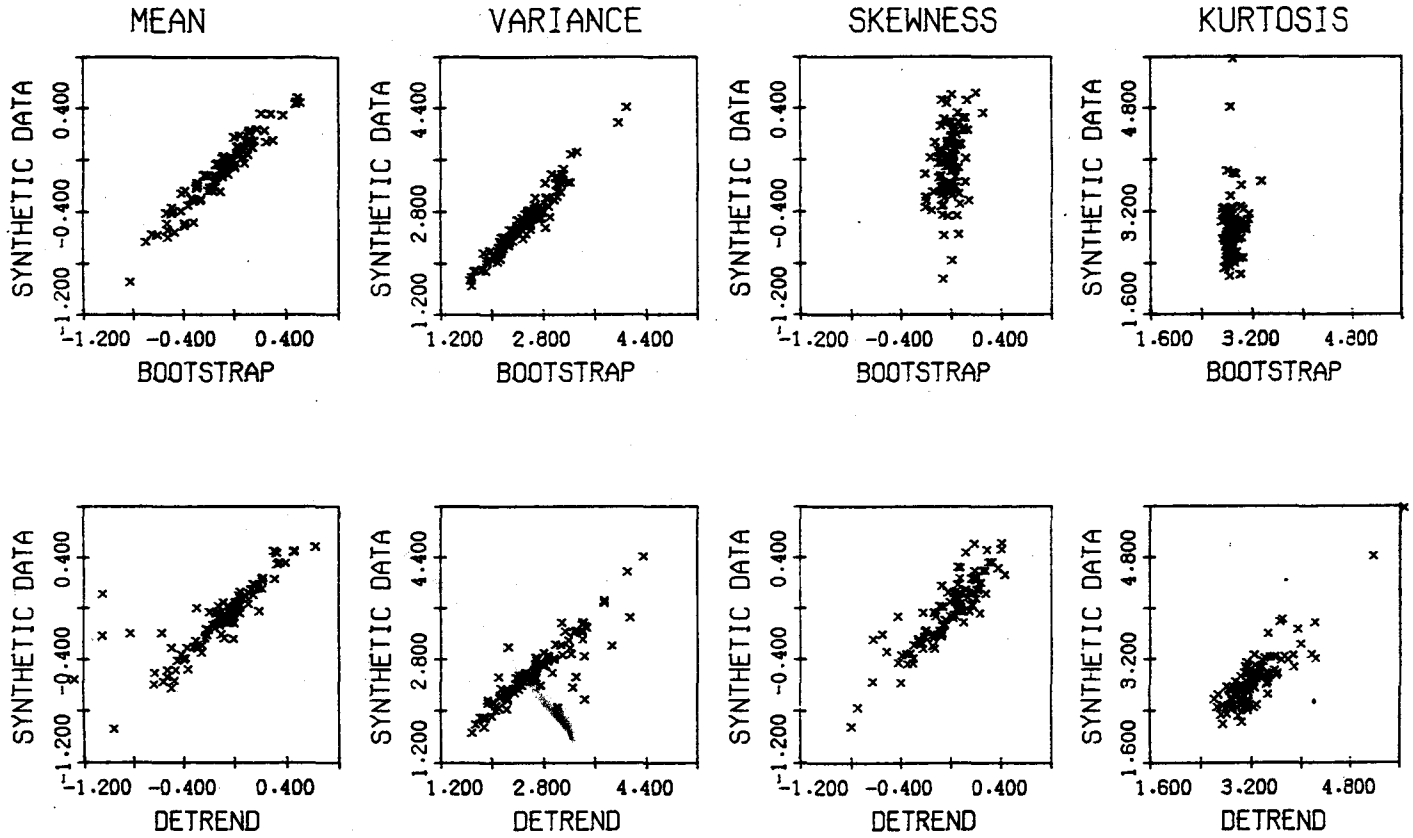


圖1　單組原始資料統計特性與合成資料統計特性對應圖
模式：$(1-0.8B^{12})Z_t=a_t$
a_t爲常態分布

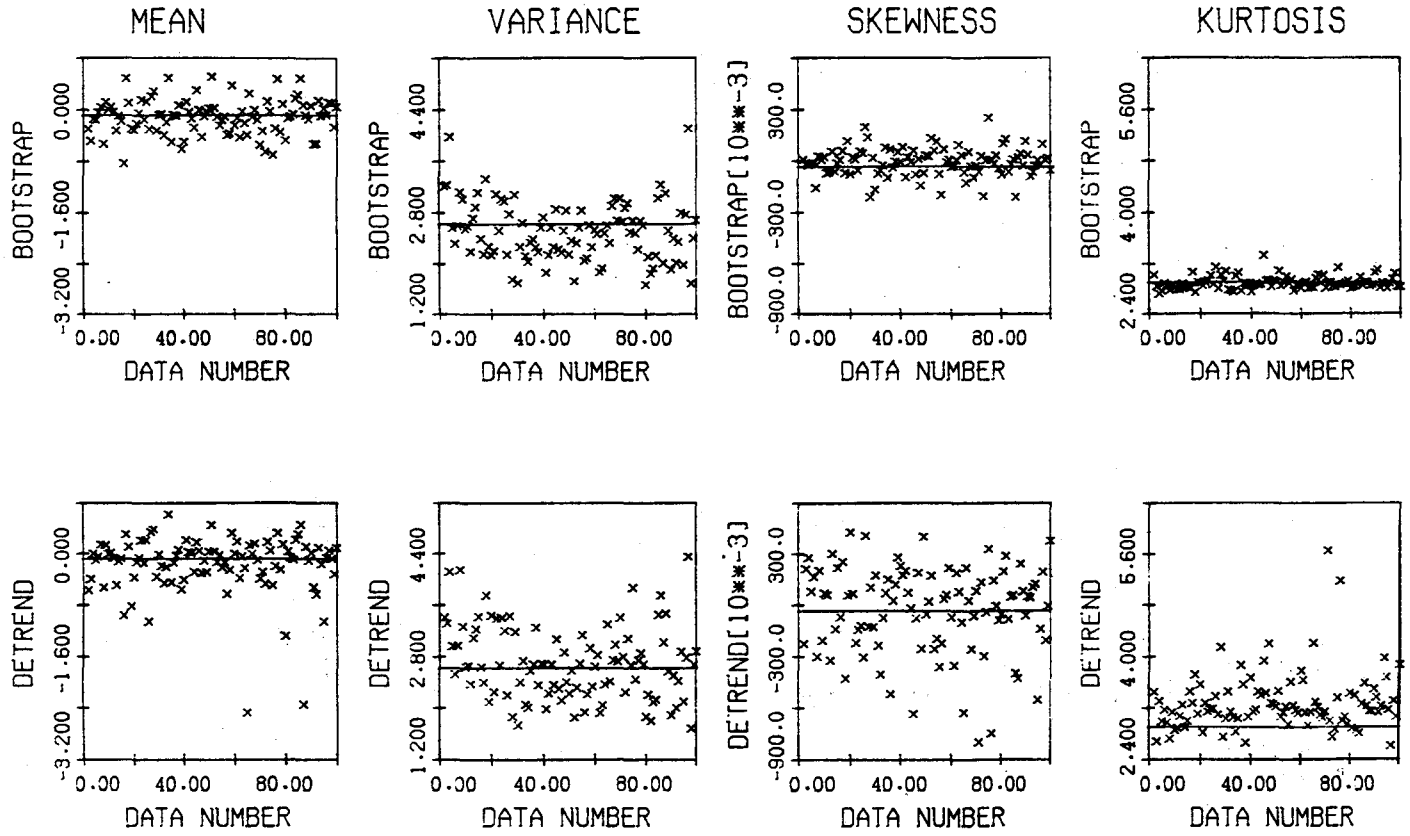
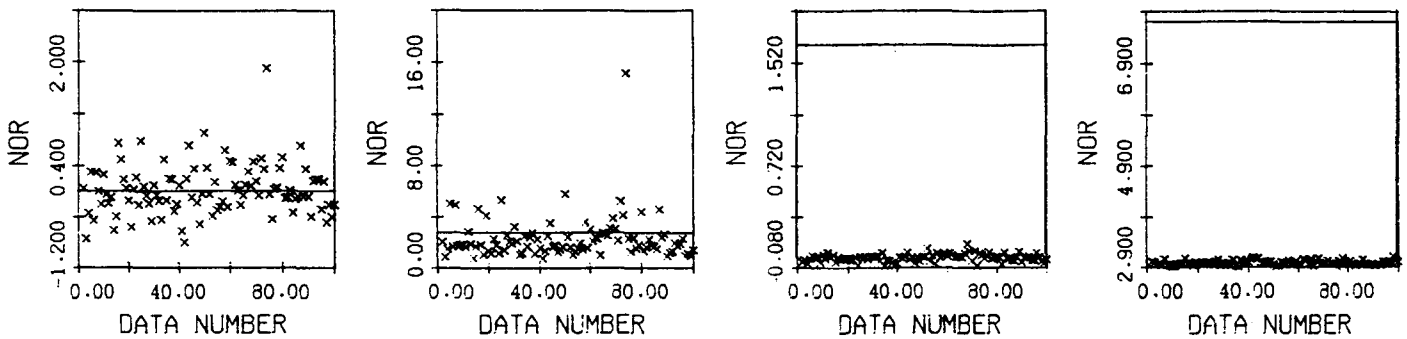


圖2　原始資料平均統計特性與合成資料統計特性對應圖
模式：$(1-0.8B^{12})Z_t=a_t$
a_t爲常態分布



附圖3-4-5 原始資料(a_t爲對數常態分布)平均統計特性與合成資料(a_t爲常態分布)統計特性對應圖

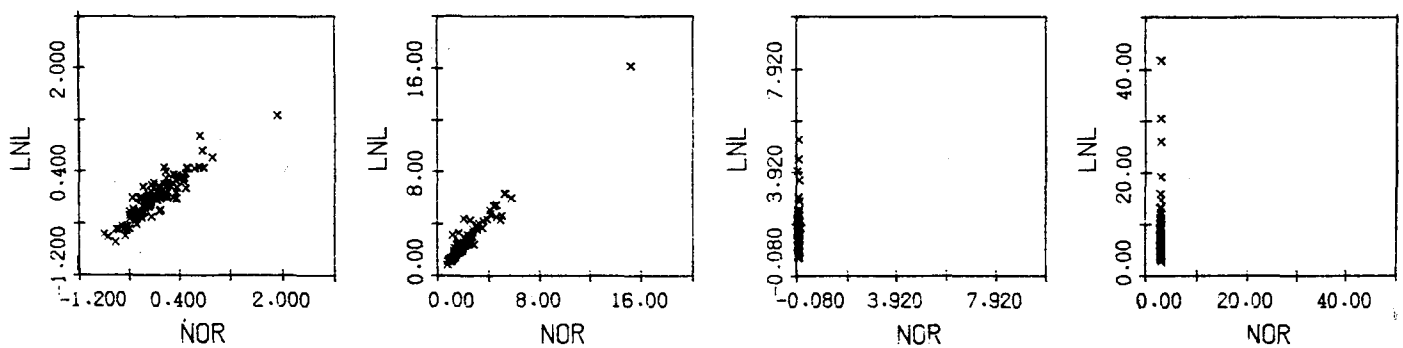


圖3 原始資料(a_t爲對數常態分布)單組統計特性與合成資料(a_t爲常態分布)統計特性對應圖

所使用之合成資料及研究結果：

以對數常態分布（如3-1 節所示）爲擾動值之原始序列，其模式爲 $(1-0.8B^{12})X_t=a_t$，樣本數爲 200，經模式判定後，分別以常態分布、加瑪分布及均勻分布爲擾動值繁衍合成資料，比較每一組原始資料之樣本統計特性和該組資料所繁衍100 組合成資料之平均統計特性，同時亦比較原始資料理論統計特性和每一組資料所繁衍100 組合成資料之平均統計特性，結果如圖 3 所示。

由圖 3 中可以看出，當擾動值誤判，則利用誤判之擾動值所繁衍之合成資料，其平均值與變異數仍可反應原始序列的樣本統計特性，但偏態係數及峰度則與原始序列的樣本統計特性及理論統計特性存在明顯之差異。由此可知，當原始資料擾動值分布誤判，則所繁衍之合成資料將不具有原始資料之眞正統計特性。

四、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.在相同序列模式下，不同分布之擾動值所產生之序列，其序列之偏態係數、峰度有明顯的差異，但相關性並不受擾動值分布之影響。當擾動值之分布未知時，雖無法利用最大概似法推估參數，但仍可使用動差法與最小二乘方法推估參數。

2.根據 AR、MA、ARMA 及ARI 模式研究結果顯示，除了隨機步模式外，由 Bootstrap 抽樣之擾動值所繁衍之合成序列，其平均值、變異數及序列相關性皆與理論分布抽樣之擾動值所繁衍相同模式合成序列之統計特性一致，而當序列擾動值具較大偏態時（如對數常態分布），Bootstrap 抽樣之擾動值所繁衍序列其偏態係數及峰度則較由理論分布抽樣之擾動值所繁衍序列略有低估。

3.根據利用本論文所提繁衍合成資料方法及除勢模式繁衍合成資料方法分別繁衍 SAR (12) 模式之合成資料的研究結果顯示，本論文所提繁衍合成資料方法可反應資料偏態係數及峰度之理論統計特性，而除勢模式只能保存資料之樣本統計特性。

4.經由判定序列擾動值之分布而由所判定之理論分布抽樣繁衍合成資料時，當擾動值分布之判定錯誤，則繁衍之合成資料將不具有原始資料之統計特性。本論文所提繁衍資料方法可避免因資料轉換失眞或擾動值分布誤判而繁衍錯誤資料。

謝　　誌

本研究承蒙行政院農業委員會『81農建—12.2—林—05(10)研究計畫』之經費補助及臺灣省水利局提供資料，特此致謝。最後，論文審查者所提供之寶貴意見，對提高本文可讀性及完整性之幫助甚大，作者由衷感謝。

參 考 文 獻

1.虞國興、施國肱、林佑昌（1987）、『部份自迴歸時間序模式自動化之研究』，臺灣水利、第35卷，第四期，pp. 36-45。

2.虞國興、莊明德（1989），『長記憶水文時序之分析——新方法』，臺灣水利，第37卷，第三期，pp. 23-32 。

3.虞國興、莊明德（1989），『長記憶水文時間序列模式之探討』，臺灣水利，第37卷，第四期，pp. 37-49。

4.Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis forecastig and control.* Holden-Day, San Francisco.

5.Box, G. E. P. and Cox, D. R., "An Analysis of Transformation", Jnurnal Royal Stat. Soc., B26, P. 211, 1964.

6.Davies, N., Spedding, T. and Watson. W. (1980), "Autoregressive Moving Average Process with Non-normal Residuals", J. Time Series Anal. 2, 103-9.

7.Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife", Ann. Statist. 7, 1-26.

8.Li, W. K. and Mcleod, A. I. (1988), "ARMA Modelling with Non-gaussian Innovations", J. Time Series Anal. 9, 155-68.

9.Nelson, C. R. (1973), *Applied Time Series Analysis.* Holden-Day, San Francisco.

10.O' Connell, P. and Jones, D. A. (1979), *Some Experience with the Development of Models for the Stochastic Simulation of*

Daily Flows, Inputs for Risk Analysis in Water Systems. Water Resources Publications, Colorado, 281-314.

11. Wei, William W. S. (1990), *Time Series Analysis.* Addison-Wesley Publishing Company.

12. Yevjevich, V. (1972), *Stochastic Processes in Hydrology.* Water Resources Publications, Fort Collins, Colorado, U.S. A.

13. Yu, G. H. and Lin, Y. C., "A Methodology for Selecting Subset Autoregressive Time Series Models", Journal of Time Series Analysis, Vol. 12, No. 4, pp. 363-373, 1991.

收稿日期：民國81年8月14日
修正日期：民國81年9月1日
接受日期：民國81年10月2日