

月 流 量 序 列 補 遺 之 研 究

Studies on the Estimation of Missing Data in Monthly Streamflow Series

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

魏 柏 村
Po-Tsun Wei

摘 要

水文資料常因儀器、人爲、天災等因素發生遺失或錯誤之現象，致使水文分析上發生不少困擾，故遺失資料之補遺爲一十分重要之研究課題，本研究以定常性合成資料驗證 Ljung (1989)、Pourahmadi (1989) 及 Abraham (1981) 所提方法之精確度，並探討其於實測資料之適用性。

研究結果顯示，Ljung、Abraham 及 Pourahmadi 三種方法之精確度大致相同，然 Abraham 方法偶有發散現象。配合 Yu 及 Lin (1991) 之部份自迴歸模式應用於台灣實測月流量資料之補遺，以本研究所提之修正 Abraham 方法精確度最高。另，研究中亦探討具遺失值時序模式判定之正確性，研究結果顯示，模式之選取準則以 BIC 較佳。

關鍵詞：遺失值，時間序列模式。

ABSTRACT

Due to the instruments malfunction, human factor or disaster, hydrological data are often missing. It causes lots of problem in the analysis. Therefore, the estimation of missing data plays an important role in the hydrological analysis. In the present study, stationary synthetic data are used to investigate the accuracy of estimating the missing proposed by Ljung (1989), Pourahmadi (1989) and Abraham (1981). Meanwhile, its aptness for analysis the real data is also investigated.

The results indicate that Ljung, Abraham and Pourahmadi are with the same accuracy. However, there is divergent situation for Abraham method. For the estimation of the missing data of monthly streamflow, the SAR model proposed by Yu and Lin (1991) is used for modelling. The modified Abraham method proposed in the present study gives the highest accuracy for the monthly streamflow in Taiwan. Besides, the accuracies of model selection criteria are also investigated for the time series with some missing data. The results indicate the

BIC serves the better performance of model selection.

Keywords : Missing data, Time series model.

一、前 言

完整且正確紀錄之水文資料，為水資源規劃及設計之重要依據，然而，水文資料常因儀器、人為、天災等因素發生遺失或錯誤紀錄之現象，致使水文分析上發生不少困擾。倘若一水文時序中某些資料發生遺失，導致破壞資料之連續性，造成僅能以具連續性之部份資料分析的遺憾，例如：乾旱頻率分析中之聯程 (Run Length) 分析需要具連續性之資料。又，時間序列之自迴歸移動平均模式 (ARMA 模式)，其理論以等間距之時間序列為基礎，若序列中發生遺失值，造成資料之間距並不完全相等，使模式之構建發生困擾。另，若資料發生明顯之錯誤紀錄，可將此紀錄視為遺失值，並以適當之推估值取以，以減低使用此組資料所造成之錯誤推論。

近十多年來，時間序列遺失值補遺問題引起十分廣泛之研究與探討，諸如 Ljung(1989)、Pourahmadi (1989) 及 Abraham(1981)，其研究探討大都侷限於理論方法之研析，並未對實際例子作一完整之研究。本研究之主要目的為：

- (1)以合成資料驗證上述方法之精確度。
- (2)探討上述方法於實測資料之適用性。

本文大綱如下：第二節為理論基礎，第三節為本研究所使用之資料，第四節為研究結果與討論，最後為結論。

二、理論基礎

2-1 時間模式：

自迴歸移動平均模式 (ARMA Model) 之定義如下：

$$\Phi(B)Z_t = \theta(B)a_t \dots\dots\dots (1)$$

其中 $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

B 為後移運算子， $B^k X_t = X_{t-k}$

$\{a_t\}$ 為常態分布白噪音過程 (White Noise Process)

，具平均值 0 及變異數為 σ_a^2 。

因為模式為可逆，(1)式可轉成 AR(∞) 模式，表示如下：

$$\pi(B)Z_t = a_t \dots\dots\dots (2)$$

其中 $\pi(B) = \Phi(B) / \theta(B)$

若 $\Phi(B)=0$ 特徵方程式之根 B_i 皆落在單位圓外，則 $\{Z_t\}$ 為定常性過程 (Stationary Process)，若 $\Phi(B)=0$ 之根 B_i 皆落在單位圓外，則 $\{Z_t\}$ 為可逆過程 (Invertible Process)。

部份自迴歸模式 (Subset Autoregressive Model) 為 AR(P) 模式之特例，即其 P 個參數中有 "部份" 為零，故稱為部份自迴歸模式。本研究採用 Yu 及 Lin(1991) 所提方法建立，部份自迴歸模式，以利實測月流量資料之分析，其優點為：

- (1)於線性時間序列模式中，自迴歸模式物理意義較清楚，理論發展較完整，且其參數推估簡單。
- (2)「部份模式」(Subset Model) 能以較少之參數表現較多或相當之訊息，為吾人所期望，尤其當資料具週期時，若擬合全階 (Full Order) 模式，常會有某些參數近乎或等於零，此時以部份模式來表示將更合適。
- (3) Yu 及 Lin(1991) 所提方法可以自動建立最佳模式，不須經繁瑣之人工判斷。

2-2 具遺失值時序之模式判定：

為探討具遺失值之時間序列模式判定之正確性，本研究以 AIC 及 BIC 為選取之準，其定義如下：

$$AIC = n \ln \hat{\sigma}_a^2 + 2(p+q) \dots\dots\dots (3)$$

$$BIC = n \ln \hat{\sigma}_a^2 + (p+q) \dots\dots\dots (4)$$

其中 n 為資料長度， $\hat{\sigma}_a^2$ 為模式殘差變異數 σ_a^2 之推估值，(p+q) 為 ARMA(p,q) 模式之參數個和。

於研究中先將遺失資料段以其他觀測值之平均值取代，藉以消弭遺失值於推估自相關函數之影響，再分別將資料擬合 AR(1) ~ AR(P) 模式，參數及模式殘差變異數之推估本研究採用動差法。將資料擬合各階模式其對應最小 AIC 或 BIC 值之模

式階數即是最適模式。

2-3 預測

(1) 最小均方誤差預測：

由預測原點 t 向前 1 時間對 Z_{t+1} 之最小均方誤差預測 $Z_{t(1)}$ ，定義如下：

$$\hat{Z}(1) = E[Z_{t+1} | H_t]$$

其中， $H_t = \{Z_t, Z_{t-1}, Z_{t-2}, \dots, \Phi_1, \theta_1, \dots, \theta_q, \sigma_a^2\}$

最小均方誤差預測之通式可表示為：

$$\hat{Z}(l) = \sum_{j=1}^l \pi_j^{(l)} Z_{t-j+1} \dots \dots \dots (5)$$

其中 $\pi_j^{(l)} = \pi_{j+l-1} + \sum_{k=1}^{l-1} \pi_k \pi_j^{(l-k)}, l > 0$

$$\pi_j^{(0)} = \pi_j, j = 1, 2, \dots$$

(2) 反向預測：

考慮(4)式之反向表示式如：

$$\pi(F)Z_t = C_t \dots \dots \dots (6)$$

其中 $\{C_t\}$ 為 $N(0, \sigma_c^2)$ 白噪音過程，且 $\sigma_c^2 = \sigma_a^2$ 。因(6)式與(2)式之共變異數結構(Covariance structure)相同，因此可證明由 t 反向 1 時間對 Z_{t-1} 之最小均方誤差預測 $Z_{t(-1)}$ 之通式可表示為：

$$\tilde{Z}(l) = \sum_{j=1}^l \pi_j^{(l)} Z_{t-j+1} \dots \dots \dots (7)$$

其中 $\pi_j^{(l)} = \pi_{j+l-1} + \sum_{k=1}^{l-1} \pi_k \pi_j^{(l-k)}, l > 0$

$$\pi_j^{(0)} = \pi_j, j = 1, 2, \dots$$

2-4 定常性時序遺失值之推估理論：

本節將就研究中所選定之 Abraham (1981)、Ljung(1989) 及 Pourahmadi(1989) 方法作一簡要之說明。

(1) Abraham 方法：

Abraham 方法為利用最小均方誤差預測理論，分別自遺失資料段前後之觀測資料預測遺失值，再以其線性組合推估遺失值。若有一時間序列 $Z = (Z_1, Z_2, \dots, Z_n)$ ，滿足定常性及可逆性之 ARMA(p,q) 模式，其中 $Z_m = (Z_{q+1}, Z_{q+2}, \dots, Z_{q+m})$ ，資料段遺失，Abraham 假設遺失值之推估值形式如：

$$\hat{Z}_{q+l} = W_{1l} \hat{Z}_q(l) + W_{2l} \tilde{Z}_{q+m+1}(m+l-1) \dots \dots \dots (8)$$

其中 W_{1l} ， W_{2l} 為未定常數， $l = 1, 2, \dots, m$ 。

m 為遺失值個數。

$\hat{Z}_q(l)$ 為由預測原點 q 向前 l 時間對 Z_{q+1} 之最小均方誤差預測值。

$Z_{q+m+1}(m+l-1)$ 為由預測原點 $q+m+1$ 反向 $(m+l-1)$ 時間對 Z_{q+1} 之最小均方誤差預測值。

因此應用 Abraham 方法推估遺失值可分成兩個步驟：

(a) 前向及後向預測值之推估：

$$\hat{Z}_q(l) = \sum_{j=1}^l \pi_j^{(l)} Z_{q-j+1}, l = 1, 2, \dots, m \dots \dots \dots (9)$$

$$\tilde{Z}_{q+m+1}(m+l-1) = \sum_{j=1}^{m+l-1} \pi_j^{(m+l-1)} Z_{q+m+j} \dots \dots \dots (10)$$

其中 $\pi_j^{(l)} = \pi_{j+l-1} + \sum_{k=1}^{l-1} \pi_k \pi_j^{(l-k)}, l > 0$

$$\pi_j^{(0)} = \pi_j, j = 1, 2, \dots$$

$$s = q + m + 1$$

(b) 係數 W_{1l} 及 W_{2l} 之決：

$$K = E[Z_{q+l} - \hat{Z}_{q+l}]^2 \\ = E[Z_{q+l} - W_{1l} \hat{Z}_q(l) - W_{2l} \tilde{Z}_{q+m+1}(m+l-1)]^2$$

則利用

$$\frac{\partial K}{\partial W_{1l}} = 0$$

$$\text{及 } \frac{\partial K}{\partial W_{2l}} = 0$$

可求得 W_{1l} 與 W_{2l} ，最小值其計算如(11)及(12)式所示。

$$W_{1l} = L_{22}(L_{11} - L_{12}) / H \dots \dots \dots (11)$$

$$W_{2l} = L_{11}(L_{22} - L_{12}) / H \dots \dots \dots (12)$$

其中， $H = L_{11}L_{22} - L_{12}^2$

$$L_{11} = \sigma_a^2 - \sigma_b^2$$

$$L_{22} = \sigma_a^2 - \sigma_b^2$$

$$L_{12} = \sigma_a^2 + \sigma_b^2 - \sigma_a^2 - \sigma_b^2$$

$$\sigma_a^2 = \left(\sum_{j=0}^{\infty} \Psi_j^2 \right) \sigma_a^2$$

$$\sigma_b^2 = \left(\sum_{j=0}^{l-1} \Psi_j^2 \right) \sigma_a^2$$

$$\sigma_c^2 = \left(\sum_{j=0}^{m-l} \Psi_j^2 \right) \sigma_a^2$$

$$\sigma_{12} = A_1' D A_2'$$

$$A_1' = (1 \Psi_1 \dots \Psi_{l-1})$$

$$A_2' = (1 \Psi_1 \dots \Psi_{m-l})$$

$D = (d_{ij})$ 為 $1 \times (m+l-1)$ 矩陣， d_{ij} 定義為：

$$d_{ij} = \left(\sum_{k=0}^{\infty} -\pi_k \Psi_{h+i+j-2} \right) \sigma_a^2, \text{ 令 } \pi_0 = -1$$

$$i = 1, 2, \dots, l; j = 1, 2, \dots, m+1-l, l = 1, 2, \dots, m$$

(2) Ljung 方法：

Ljung 方法以具遺失值時間序列之最大概似函

數推估遺失值。若有一時間序列 $Z=(Z_1, Z_2, \dots, Z_n)$ ，滿足定常性及可逆性之 ARMA(p,q) 模式，其概似函數為：

$$\rho(Z | \beta, \sigma^2) = (2\pi \sigma^2)^{-n/2} | \Sigma |^{-1/2} \exp \left\{ \frac{-Z' \Sigma^{-1} Z}{2\sigma^2} \right\} \dots \dots \dots (13)$$

若 Z 中 $Z_m=(Z_q, Z_{q+1}, Z_{q+2}, \dots, Z_{q+m-1})$ ，資料段遺失，則具遺失值時間序列之概似函數為：

$$\rho(Z | \beta, \sigma^2) = \rho(Z_m | \beta, \sigma^2) \rho(Z | Z_m, \beta, \sigma^2) \dots \dots \dots (14)$$

其中 Z_m 為 $n-m$ 向量，表示 Z 中除了 Z_m 之外的 $n-m$ 個觀測值。欲分解 (13) 式，先將 Z 中 Z_m 以零取代，並以 $n \times 1$ 向量 Z_n 表示之。令 $X=(x_{ij})$ 為 $n \times m$ 矩陣；其中當 $i=q+j-1, x_{ij}=1$ ，否則 $x_{ij}=0$ 。將 Z 分解成 $Z=Z_n + XZ_m$ ，故 (12) 式中指數部份之分子可改寫成：

$$\begin{aligned} Z' \Sigma^{-1} Z &= (Z_n + XZ_m)' \Sigma^{-1} (Z_n + XZ_m) \\ &= (Z_n + X\hat{Z}_m)' \Sigma^{-1} (Z_n + X\hat{Z}_m) + \\ &\quad (Z_m - \hat{Z}_m)' X' \Sigma^{-1} X (Z_m - \hat{Z}_m) \end{aligned}$$

其中 \hat{Z}_m 為 $m \times 1$ 向量，表示遺失值 Z_m 之推估值。將上式對 Z_m 作偏微分，求 $Z' \Sigma^{-1}$ 之極小值，得遺失值 Z_m 之推估值，與 (13) 式中之條件機率分佈函數如：

$$\begin{aligned} \hat{Z}_m &= -(X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} Z_n \\ p(Z_m | \beta, \sigma^2) &= (2\pi \sigma^2)^{-(m-n)/2} | \Sigma |^{-1/2} | X' \Sigma^{-1} X |^{-1/2} \\ &\quad \exp \left\{ \frac{-1}{2\sigma^2} (Z_n + X\hat{Z}_m)' \Sigma^{-1} (Z_n + X\hat{Z}_m) \right\} \\ p(Z_m | Z_m, \beta, \sigma^2) &= (2\pi \sigma^2)^{-m/2} | X' \Sigma^{-1} X |^{1/2} \\ &\quad \exp \left\{ \frac{-1}{2\sigma^2} (Z_m - \hat{Z}_m)' X' \Sigma^{-1} X (Z_m - \hat{Z}_m) \right\} \end{aligned}$$

若資料滿足 AR(p) 模式，且遺失值發生時間大於 p 時，其推估值可以 (14) 式求得。

$$\hat{Z}_m = -(\tilde{X}' \tilde{X})^{-1} \tilde{X}' \tilde{a} \dots \dots \dots (15)$$

其中 $\tilde{X}' \tilde{a} = \Phi(F) \Phi(B) Z_t, t=q, q+1, \dots, q+m-1, FZ_t = Z_{t+p}$ 。

$\tilde{X}' \tilde{X}$ 為 $m \times m$ 矩陣，其第 i 列 j 行元素為：

$$\sum_{k=|i-j|}^p \phi_k \phi_k - |i-j|, \text{ 令 } \phi_0 = -1$$

(3) Pourahmadi 方法：

Pourahmadi 方法以遺失資料段 (Missing Data Block) 之預測值，與其後觀測值之預測誤差之線性組合推估遺失值。若有一時間序列 $Z=(Z_1, Z_2, \dots, Z_n)$ ，滿足定常性及可逆性之 ARMA(p,q) 模式，其中 Z_q 為遺失值。 H_q 表示用以推估遺失值之 Hilbert 向量空間 (Hilbert Space) 即 $H_q = \text{SPAN} \{ Z_t ; t = 1, 2, \dots, n, t \neq q \}$ ，Pourahmadi(1989) 證明 H_q 可分解為兩正交子空間 (Orthogonal Subspace) H_{q-1} 與 S 之和：

$$H_q = H_{q-1} \oplus S \dots \dots \dots (16)$$

其中 $H_{q-1} = \text{SPAN} \{ Z_{q-1}, Z_{q-2}, \dots, Z_1 \}$

$$S = \text{SPAN} \{ Z_{q+k} - Z_{q-1}(k+1) ; k=1, 2, \dots, r=n-q \}$$

$\hat{Z}_{q-1}(k+1)$ 為由預測原點 $q-1$ ，向前 $(k+1)$ 步對 Z_{q+k} 之最小均方誤差預測值。

Pourahmadi 將 Z_q 遺失值之推估值表示為 Z_q 於 H_q 上之正交投影：

$$\hat{Z}_q = \hat{Z}_{q-1}(1) + P_S^Z \dots \dots \dots (17)$$

其中 \hat{Z}_q 為遺失值 Z_q 之推估值。

$\hat{Z}_{q-1}(1)$ 為由預測原點 $q-1$ ，向前 1 步對 Z_q 之最小均方誤差預測值。

P_S^Z 為 Z_q 在向量空間 S 上之投影，可以 (17) 式表示：

$$P_S^Z = \sum_{k=1}^r C_k (Z_{q+k} - \hat{Z}_{q-1}(k+1)) \dots \dots \dots (18)$$

其中，

$$C_k = (1 + \sum_{i=1}^k \phi_i^2)^{-1} (\phi_k - \sum_{i=1}^{k-1} \phi_i \phi_{k+i}), k=1, 2, \dots, r$$

將上述理論推廣至具 m 個遺失值之時間序列之情形，並設遺失資料段為 $Z_m=(Z_q, Z_{q+1}, \dots, Z_{q+m-1})$ ；則 Z_m 可由 (18) 式推估。

$$\hat{Z}_m = \hat{Z}_m + C'(Z_t - \hat{Z}_t) \dots \dots \dots (19)$$

其中 \hat{Z}_m 為 $m \times 1$ 向量，為遺失值 Z_m 之推估值。

\hat{Z}_m 為 $m \times 1$ 向量，即由 H_{q-1} 對 Z_m 之預測值。

\hat{Z}_t 為 $r \times 1$ 向量，為遺失資料段之後之觀測值。

\hat{Z}_k 為 $r \times 1$ 向量，表示由 H_{q-1} 對 Z_k 之預測值。

C 為 $r \times m$ 矩陣， $C = D(I_m + A'A)^{-1} T_m$ 。

$r = n - (q + m) + 1$ 為遺失資料段之後的觀測值個數。

I_m 為 m 階單位矩陣。

T_m 為 $m \times m$ 矩陣； $T_m = (\Psi_{j-i})$ ， $i = 0, m-1$
 $j = 0, m-1$

A 為 $r \times m$ 矩陣； $A = (a_{ij})$ ， $i = 1, r$
 $j = 1, m$

A 矩陣中第 m 行為 $\pi^{(i)}$ ， $i = 1/r$

第 $m-1$ 行為 $\pi^{(i)}$ ， $i = 1/r$

第 1 行為 $\pi^{(i)}$ ， $i = 1/r$

D 為 $r \times m$ 矩陣； $D = (d_{kj})$ ， $k = 1, r$ ；
 $j = 1, m$

其中，

$$d_{kj} = a_{kj} - \sum_{i=1}^{r-k} a_{ki} a_{i+j}$$

2-5 本研究所提之修正方法：

因實測月流量資料具週期性，且約 78% 之年雨量集中於 5 ~ 10 月，且受颱風雨量影響甚大，因而非常不利於線性時間序列之模擬，故本研究利用 Ljung (1989) 方法分別對 Abraham 方法加以修正，茲將修正步驟說明如下：

(1) 以 Yu 及 Lin (1991) 所提之 SAR 模式配合 Abraham 方法補遺，並將補遺後之序列視為 "完整" 之資料。

(2) 利用 Abrahamz 方法求得 Z_t 之推估值 \hat{Z}_t ， $t=1, 3, \dots, n=12$ ，

並計算一比例常數 $\frac{\hat{Z}_t}{Z_t} = C_t$ 。若 C_t 小於 1 表示高估 Z_t 。

(3) 計算 1 ~ 12 個月份中 \bar{C}_j 之平均值 c_{nj} ， $j=1, 2, \dots, 12$ 。 c_{nj} 表示平均而言，吾人高或低估觀測值之比例常數。

(4) 將步驟(1)所估計之遺失值，分別依遺失值發生之月份乘上該月份之 \bar{C}_j ，即完成修正工作。

2-6 補遺精確度之判定：

就合成資料而言，吾人常就同一模式所產生之資料分別以不同方法加以分析，此時，理論上，因資料具相同之變異數；故本研究採用均方差 MSE 判定補遺精確度。然而，由於實測資料間之變異情形彼此不同，為能了解各種方法之精確度，本研究採用判定係數 R^2 。

(1) 合成資料之分析：

本研究首先利用模式產生完整之合成資料，

然後再設定遺失資料，假設遺失值發生時間為 $n/2$ ， $(n/2)+1, \dots, (n/2)+m-1$ ；共連續 m 個資料遺失（ n 為資料長度）。例如：先產生資料長度為 60 之 AR (1) 序列，若欲使之發生連續兩個遺失值，則將 30 及 31 兩點之觀測值當成遺失值。在合成資料之研究中，均方差定義為：

$$MSE = \frac{1}{nset} \sum_{k=1}^{nset} (Z_k - \hat{Z}_k)^2 \dots\dots\dots (10)$$

其中 $nset$ 為合成資料之組數。

k 為發生遺失值之時間。

Z_k 為實際資料，但假設其為遺失值。

\hat{Z}_k 為遺失值之推估值。

(2) 實測資料分析：

本研究設定遺失值由第 13 個資料開始，若探討連續 m 個遺失值時，則令資料由第 13 至 $(13+m-1)$ 共 m 個資料發生遺失，稱為一個遺失資料段，並進行補遺工作，再令資料由第 $(13+m)$ 至 $(13+2m-1)$ 間之 m 個資料遺失，並進行補遺工作，重覆上述步驟直至最後一次設定遺失資料段其最後一個遺失值發生時間小於或等於 $n-12$ 。

實測資料分析以中判定係數為判定補遺精確度，其定義為：

$$R^2 = 1 - \frac{\sum_{k=T_1}^{T_2} (Z_k - \hat{Z}_k)^2}{\sum_{k=T_1}^{T_2} (Z_k - \bar{Z}_k)^2} \dots\dots\dots (11)$$

$$\bar{Z}_k = \frac{1}{(T_2 - T_1 + 1)} \sum_{k=T_1}^{T_2} Z_k \dots\dots\dots (12)$$

其中 T_1 為第一個遺失資料之時間。

T_2 為最後一個遺失資料之時間。

Z_k 為實際資料，但假設其為遺失值。

\hat{Z}_k 為遺失值之推估值。

三、本研究所使用之資料

本研究使用之資料分合成資料與實測資料兩種。為能有效探討 Abraham、Ljung 及 Pourahmadi 三種方法之適用性，研究中藉合成資料加以驗證。另，本研究亦分析九組實測資料及十五組台灣地區之實測月流量資料，藉以探討上述方法於實測資料之適用性，茲分述如下：

(1) 合成資料：

本研究分別利用 AR(1) 及 AR(2) 模式繁衍合成

表 1. AR(2) 模式數表

模式 編號	參 數	
	Φ_1	Φ_2
1	-.40	-.04
2	-.80	-.16
3	-1.20	-.36
4	-1.60	-.64
5	.40	-.04
6	.80	-.16
7	1.20	-.36
8	1.60	-.64
9	.00	.04
10	.20	.08
11	.40	.12
12	.60	.16

模式 編號	參 數	
	Φ_1	Φ_2
13	-.20	.08
14	.00	.16
15	.20	.24
16	.40	.32
17	-.40	.12
18	-.20	.24
19	.00	.36
20	.20	.48
21	-.60	.16
22	-.40	.32
23	-.20	.48
24	.00	.64

表 2. 九組實測資料之基本統計特性表

代號	資 料 說 明	來 源	資料長度	平均 值	變 異 數	偏態係數
C1	U.S. Quarterly Unemployment Rates	C	121	5.11	1.98	0.43
C4	Monthly AA Railroad Bond Yields	C	102	787.92	4530.70	-0.72
B_C	Chemical Process Temperature	B	226	22.97	4.22	-0.23
B_D	Chemical Process Viscosity	B	310	9.13	0.36	-0.46
B_E	Wolfers Sunspot Number	B	100	46.93	1382.19	0.85
W3	Blowfly Data	W	82	4138.65	1824775.00	0.58
W4	Monthly Unemployed Females	W	300	599.34	30046.32	0.00
W5	Yearly Accidental Death Rate	W	35	44.26	29.42	0.12
W6	Yearly U.S. Tobacco Production	W	114	1321.99	334666.10	0.01

註： C 表示 Jonathon D. Cryer (1986) ,Time Series Analysis.
 B 表示 Box, G. E. P. and Jenkins, G. M. (1976) ,Time Series Analysis, Forecasting and Control.
 W 表示 Wai W.W.S. (1989) ,Time Series Analysis.

資料；其資料長度各為 60,120,240。其中 AR(1) 模式之參數 ϕ_1 由 -0.99,-0.9,-0.8,-0.7,.....,0.7,0.8,0.9,0.99 共二十一種變化。AR(2) 模式之參數如表 1 所示，共計變化二十四種。模式中隨機干擾項皆為常態

分佈具平均值 0 及變異數 1，故本研究共計分析 AR(1) 模式合成資 56700 組及 AR(2) 模式合成資料 68400 組。

(2)實測資料：

表 3. 九組實測資料以 BIC 準則選取之模式

資料	模 式	σ_a^2
C1	$(1-1.5B+0.61B^2)(\varepsilon_t-5.11)=a_t$	0.15
C4	$(1-0.96B)(\varepsilon_t-787.92)=a_t$	325.98
B_C	$(1-1.19B+0.07B^2+0.16B^3)(\varepsilon_t-22.97)=a_t$	0.17
B_D	$(1-0.86B)(\varepsilon_t-9.13)=a_t$	0.09
B_E	$(1-1.32B+0.63B^2)(\varepsilon_t-46.93)=a_t$	289.21
W3	$(1-0.73B)(\varepsilon_t-6138.65)=a_t$	839413.50
W4	$(1-0.61B-0.37B^2)(\varepsilon_t-599.34)=a_t$	1646.08
W5	$(1-0.87B)(\varepsilon_t-64.26)=a_t$	7.11
W6	$(1-0.63B-0.3B^2)(\varepsilon_t-1321.99)=a_t$	52012.78

本研究使用之九組實測資料來源、資料之基本統計特性及如表 2 所示。鑑於本研究之主要目的在於比較 Abraham, Ljung, Pourahmadi 三種方法之優劣，故直接採用完整資料以 BIC 準則擬合模式，結果如表 3 所示，其主要目的在於避免因具遺失值時發生模式誤判之困擾。然而，為避免過度理想化，於分析時模式之參數為於發生遺失值之情形下所推估的。

另，本研究分析之十五組台灣地區實測月流量資料因其皆呈現正偏態，且具週期性，研究中先以自然對數轉換後（轉換後資料之偏態係數趨近於零），再以 Yu 及 Lin(1991) 所提之方法選取 SAR 模式，其結果及資料之基本統計特性如表 4 所示。

四、結果與討論

本研究採用 IMSL 套裝軟體產生合成資料與推估其參數。另，因篇幅所限，本文僅列部份成果，其他請參閱 [1]。

(1) 合成資料：

(a) 具遺失值時序模式判定之正確性：

研究中以 AIC 與 BIC 準則探討具遺失值時模式判定之正確性，表 5 所示為以 AIC 及 BIC 準則於 10 0 組具遺失值之 AR(2) 序列中選出正確模式之組數 \oplus ，結果顯示當資料長度為 240，模式之參數為 0.4 及 0.32，遺失值個數為 1，AIC 準則於 100 組中選得正確模式之組數為 69 組，而 BIC 準則選得正確模式之組數為 96 組。當 AR(2) 模式及 $n=60$ 時，

AIC 及 BIC 準則於遺失值個數為 1,2 及 3 個時之平均精確度分別為 60、64,65、68,65、68 組，當 $n=240$ 時，AIC 及 BIC 準則於遺失值個數為 1,2 及 3 個時之平均精確度分別為 64、88,67、94,64 及 92 組。另，本研究亦探討 (1) 模式其結果顯是當 $n=60$ 時，AIC 及 BIC 準則於遺失值個數為 1,2 及 3 個時之平均精確度分別為 77、96,80 及 96 組，當 $n=120$ 時，AIC 及 BIC 準則於遺失值個數為 1,2 及 3 個時之平均精確度分別為 73、94,73、95,74 及 95 組，當 $n=240$ 時，AIC 及 BIC 準則於遺失值個數為 1,2 及 3 個時之平均精確度分別為 75、96,75、97,75 及 97 組。整體而言，具遺失值時序之模式選取準則，以 BIC 為佳。

(b) 補遺之精確度：

當 AR(1) 模式時，研究發現無論時序具 1,2 或 3 個遺失值時，三種方法補遺之精確度幾乎完全相同。另，當參數越接近 ± 1 時，補遺之 MSE 越小，而當資料接近白噪音時，Abraham 方法偶有發散現象，如表 6 所示，然此現象將隨資料長度增加而減少。

當 AR(2) 模式時，研究發現當資料僅具一個遺失值發生時，Ljung 方法與 Pourahmadi 方法補遺之精確度完全相同。然而，Abraham 方法之精確度與上述二種方法亦十分接近。當資料具一個以上之遺失值時，Pourahmadi 方法呈現不穩定之現象，其補遺之 MSE 較其他方法為。整體而言，Ljung 方法之精確度與穩定性最高，其次為 Abraham 方法，Pourahmadi 方法最差。

(2) 實測資料：

(a) 表 7 所示分別為利用 Abraham、Ljung 及 Pourahmadi 方法於九組實測資料補遺之 R^2 之平均值結果，由其結果可知，當遺失值個數 $m=1$ 及 2 時，三種方法之精確度十分接近，然 $m=3$ 時，其精確度之高低依次為 Ljung, Abraham, Pourahmadi。

(b) 表 8 所示分別為利用 Abraham、Ljung 及 Pourahmadi、修正 Ljung 及修正 Abraham 方法於台灣地區實測月流量資料補遺之 R^2 之平均值結果，由其結果可知，當遺失值個數 $m=1$ 時，其精確度之高低依次為修正 Abraham，修正 Ljung, Abraham, Ljung, Pourahmadi 方法，當遺失值個數 $m=2$ 時，其精確度之高低依次為修正 Abraham，修正 Ljung,

表 4. 台灣地區實測月流量資料之 SAR 模式表

資料編號	站名	資料長度	平均值	變異數	偏態係數	SAR 模式
F030002	玉峰	384	2.48	0.69	0.58	SAR(1,5,11,12)
F030004	菱角	384	1.68	0.50	0.60	SAR(1,4,5,12)
F030005	高義	384	3.04	0.64	0.67	SAR(1,5,11,12)
F030009	石門(3)	408	3.48	0.57	0.33	SAR(1,4,6,12)
F100013	蘭陽大橋	468	3.75	0.75	0.26	SAR(1,5,12)
F180001	山腳	372	1.42	0.62	0.80	SAR(1,5,12)
F250004	環山合流點	372	2.30	0.70	0.19	SAR(1,5,11)
F270014	柑子林	372	3.47	0.92	0.67	SAR(1,4,11,12)
F290042	桶頭	408	2.02	1.93	0.34	SAR(1,4,8,11)
F340011	立山	360	2.68	0.53	0.93	SAR(1,5,12)
F400016	延平	384	3.08	0.94	0.64	SAR(1,5,11)
F400020	霧鹿	480	3.84	1.66	-0.25	SAR(1,5,11)
F510010	新發大橋	360	3.64	1.20	0.45	SAR(1,5,11,12)
F510031	月眉	336	2.46	3.12	-0.06	SAR(1,5,11)
F510036	高屏大橋	360	4.60	1.91	0.21	SAR(1,5,11)

註：資料經取自然對數。

Abraham, Pourahmadi, Ljung 方法，當遺失值個數 $m=3$ 時，其精確度之高低依次為修正 Abraham，修正 Ljung, Abraha, Ljung, Pourahmadi 方法。整體而言，以 SAR 模式分析臺灣地區實測月流量資料時，本研究提修正 Abraham 方法之精確度最高。圖 1 為高屏溪－高屏大橋站 (F510036) 之月流量紀錄假設其發生一個遺失值時，利用修正 Abraham 方法補遺之值與實測值之比較圖。

五、結 論

根據本研究採用之合成資料與實測資料研究結果，可得下列結論：

- (1) 具遺失值時序之模式選取準則以 BIC 較佳。
- (2) Ljung、Abraham 及 Pourahmadi 三種方法之精確度大致相同，然 Abraham 方法偶有發散現象。
- (3) 以 SAR 模式分析臺灣地區實測月流量資料時，本研究提修正 Abraham 方法之精確度最高。

表 5. AIC 與 BIC 準則選取正確模式之組數 (AR(2),100 組)

資料長度		60						240					
遺失資料時間		30 - 30		30 - 31		30 - 32		120 - 120		120 - 121		120 - 122	
遺失資料個數		1		2		3		1		2		3	
準則		AIC		BIC		AIC		BIC		AIC		BIC	
AR2 參數 φ1 φ2		AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
-1.6 -0.64		39	37	58	60	59	58	39	64	46	91	40	81
0.4 0.32		58	39	53	37	54	39	69	96	73	96	73	97
0.20 0.48		73	86	75	83	72	82	74	94	75	94	73	95
0.00 0.64		70	92	75	91	74	91	75	96	74	96	70	96
平均精確度		60	64	65	68	65	68	64	88	67	94	64	92

表 6. Abraham 方法補遺之 MSE (AR(1),100 組)

資料長度		60						240					
遺失值個數		1		2		3		1		2		3	
遺失點 φ1		30	30	31	30	31	32	120	120	121	120	121	122
-0.99	.50	.51	.64	.74	1.00	.73	.58	.70	.60	1.00	1.16	.96	
-0.90	.55	.55	.72	.75	1.11	.78	.63	.83	.66	1.08	1.24	1.01	
-0.80	.59	.61	.81	.80	1.20	.83	.70	.92	.73	1.16	1.31	1.04	
-0.70	.64	.67	.91	.84	1.27	.85	.78	1.01	.80	1.22	1.32	1.04	
-0.60	.68	.73	1.00	.86	1.29	.85	.87	1.10	.86	1.25	1.28	1.01	
-0.50	.73	.78	1.09	.86	1.28	.83	.96	1.17	.90	1.26	1.19	.95	
-0.40	.77	.82	1.15	.86	1.25	.81	1.05	1.22	.92	1.26	1.09	.89	
-0.30	.81	.84	1.19	.85	1.21	.80	1.14	1.24	.93	1.25	.99	.84	
-0.20	.84	.85	1.20	.87	1.18	.79	1.20	1.25	.92	1.25	.92	.81	
-0.10	.86	*****	*****	*****	*****	*****	1.24	1.25	.91	1.24	.88	.79	
0.00	.87	.85	1.18	*****	*****	*****	1.24	1.24	.90	*****	*****	*****	
0.10	.87	.85	1.15	*****	1.19	*****	1.21	1.24	.90	1.24	.89	.82	
0.20	.85	.84	1.10	.85	1.21	.91	1.16	1.23	.91	1.23	.94	.84	
0.30	.81	.82	1.04	.85	1.23	.98	1.08	1.21	.91	1.22	1.00	.87	
0.40	.77	.80	.96	.83	1.23	.96	.99	1.19	.90	1.20	1.06	.91	
0.50	.72	.78	.88	.81	1.20	.96	.90	1.14	.87	1.17	1.12	.93	
0.60	.67	.75	.79	.78	1.14	.94	.80	1.07	.83	1.12	1.16	.95	
0.70	.63	.71	.70	.75	1.06	.88	.71	.99	.77	1.06	1.15	.95	
0.80	.58	.67	.62	.71	.94	.80	.64	.90	.71	.97	1.11	.93	
0.90	.53	.62	.54	.65	.82	.71	.57	.81	.64	.88	1.02	.88	
0.99	.49	.58	.49	.60	.74	.66	.51	.72	.57	.78	.82	.82	

註：***** 表示發散。

謝 誌

本研究承蒙行政院國科會 NSC84-2211-E-032-017 研究計畫之經費補助，特此致謝。同時，論文審查者所提供之寶貴意見，作者由衷感謝。

參考文獻

1. 魏柏村，"月流量序列補遺之研究"，淡江大學水資源及環境工程研究所碩士論文，pp.77，民國 83 年 6 月。

表 7. Abraham、Ljung 及 Pourahmadi 方法於九組實測資料補遺之整體精確度 R^2

方法	m=1	m=2	m=3
Abraham	0.899	0.827	0.831
Ljung	0.900	0.829	0.836
Pourahmadi	0.900	0.828	0.812

表 8. Abraham、Ljung、Pourahmadi、修正 Ljung 及修正 Abraham 方法於台灣地區實測月流量資料補遺之整體精確度 R^2 值

方法	m=1	m=2	m=3
Abraham	0.318	0.277	0.306
Ljung	0.306	0.274	0.297
Pourahmadi	0.306	0.276	0.286
修正 Ljung	0.333 ^(*)	0.300 ^(*)	0.323 ^(*)
修正 Abraham	0.346 ^(*)	0.302 ^(*)	0.333 ^(*)

註：(*)中數字為精確度之排名次序。

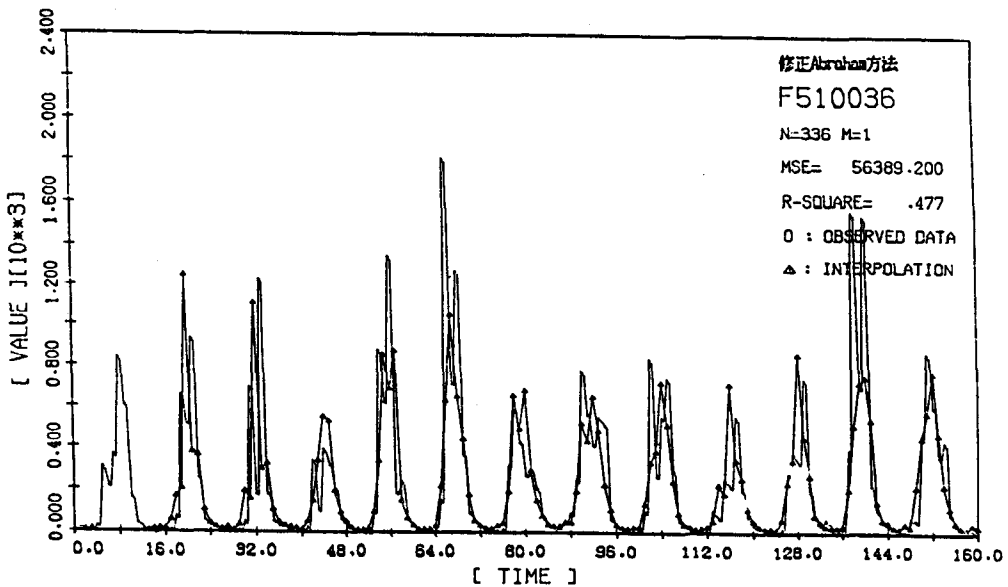


圖 1. 高屏溪—高屏大橋站月流量資料補遺之結果 (m = 1, 修正 Abraham 方法)

2. Brubacher S.R. and Tunnicliffe Wilson G., " Interpolating Time Series with Application to the Estimation of Holiday Effects on Electricity", Appl. Statist. Vol.25, No.2, pp.107-116, January, 1976.
3. Ljung G.M., "A Note on The Estimation of Missing Values in Time Series", Commun. Statist. -Simula., 18 (2), pp.459-465, 1989.
4. Pourahmadi M., "Estimation and Interpolation of Missing Values of a Stationary Time Series", Journal of Time Series Analysis, Vol.10, No.2, pp. 149-169, 1989.
5. 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 83 年 12 月 10 日
 修正日期：民國 84 年 8 月 20 日
 接受日期：民國 84 年 9 月 7 日