

論 文

非定常 月流量時系列의 解析과 豫測

Analysis and Forecast of Non-Stationary

Monthly Steam Flow

李 在 烟*

Lee, Jae Hyung

鮮 于 仲 皓**

Sonu, Jung Ho

An attempt of synthesizing and forecasting of monthly river flow has been made by employing a linear stochastic difference equation model

As one of the linear stochastic difference equation model, an ARIMA Type is tested to find the suitability of the model to the monthly river flows. On the assumption of the stationary covariance of differenced monthly river flows the model is identified and is evaluated so that the residuals have the minimum variance. Finally a test is performed to find the residuals being White noise.

Monthly river flows at six stations in Han River Basin are applied for case studies. It was found that the difference operator is a good measure of forecasting the monthly river flow.

要 旨

比較的週期성이 강하고 편향성이 있는 경우에 있어서豫測 및 模擬發生을 위한 模型開發이試圖되었다. 原時系列로부터 求한 差分時系列(Differenced time series)이 定常共分散을 갖는다는 假定하에 模型의 同定化(Model Identification)가 實施되었으며, 定常假定을 正當化하기 위해 残差(Residual)의 統計的性質을 檢討하였다. 또한, 同定된 模型의豫測精度를 높이기 위하여豫測誤差의 分散이 最小가 되도록 推計的制御(Stochastic Control)된 模型을豫測에 使用하였다.

韓國主要河川流域의 流量資料에 對한 模型의 同定과豫測結果로부터, 差分演算子(Difference operator)는 偏向과 週期을 除去하는데 좋은 方法이 됨이 判斷되었다.

I. 序 論

一般的으로 流量時系列은 週期성이 강하고 경우에 따라서는 偏向성을 나타내고 있기 때문에 이들의 特性을 分析하고, 實際 水資源計劃에 應用하기에는 상당한 어려움이 있다. 이들의 將次舉動을豫測하기 위한手段으로서 많은 統計的인 方法이 利用되고 있으나, 制御 및 產業工學에서 利用되고 있는 예측이론의導入으로 流量時系列을 分析,豫測할 수 있다.

推計的差分方程式模型(Stochastic Difference-Equation Model)은 이러한 時系列의 分析方法中 한 方法으로 이는 몇段階後의 月流量을豫測하는데 有用하여 水資源計劃과 最適運營方案(Optimal Operating Policy)을樹立하는데 基礎가 되며 또한 記錄된 時系列과 같은 統計的인 特性을 갖는 時系列을 合成할 수 있어서 復雜한 水資源系 模擬實驗研究에 도움이 된다.

* 서울大學校 工科大學 土木工學科 助教

** 서울大學校 工科大學 助教授, 工學博士.

$$(1-L)Y(t) = Y(t) - Y(t-1)$$

이다. 式(6)을 간략하게 ARIMA $(p, d, q) \times (P, D, Q)$ 로 表記한다. 여기서 G, H 는 季節成分에 對한 加重值이다. Fig. 1은 式 (6)에 對한 filtering 過程을 “Block Diagram”으로 나타낸 것이다.

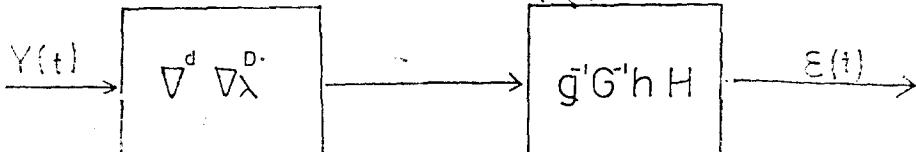


Fig. 1 Block Diagram to white Monthly River flow

parzen⁽⁹⁾은 이러한 模型이豫測을 爲해서는 적절하나
季節性이나 傾向을 說明하는데는 不足함을 指摘하고
있다. 또한 時系列이 季節性이나 傾向성이 支配的일
때는 ARIMA模型의豫測精度는 差分($\Delta^d \nabla^D$)에는 크게 영향을 받으나 推定變數에는 그 영향이 미미하다
는 것을 說明하고 있다. 그러나 parzen自身도 더 좋은
對案을 提示하지 못하고 있다.

III. 原時系列의 豫測

時系列 $Y(t)$ 는 豫測이 可能한 項 $Y^b(t)$ 와 豫測不可
能한 $Y^u(t)$ 의 합으로 나타낼 수 있다. 즉 未來에 實
現되는 時系列 $Y(t)$ 에 對하여 $Y^b(t)$ 는 예측치이며
 $Y^u(t)$ 는 $Y(t)$ 와 $Y^b(t)$ 의 差 즉 Model의豫測에 對한
誤差라고 생각할 수 있다. $Y^u(t)$ 는 “Innovations”이
라 불리워지며 이 時系列은 自由雜音⁽¹⁰⁾이다. 즉

로부터 次段階에 對한豫測誤差의 分散은

$$\sigma_k^2 = E\{|Y^b(t+k|t) - Y(t+k)|^2\} \quad \dots \dots \dots (9)$$

로 表示된다. $Y^u(t)$ 가 白色雜音인 $\varepsilon(t)$ 와 같다는 事實은 模型의 同定이나 變數推定에 非常 重要하며 다음과 같이 간단히 證明 (4) (11)를 수 있다. $Y(t)$ 를 k 段階後の MA模型으로 나타내면

$$Y(t+k) = C(k) \sum_{j=0}^{k-1} \psi(j) \varepsilon(t+k-j) \dots \quad (10)$$

이때, 最適豫測值가 다음式으로 表示된다고하면

$$Y^b(t+k, |t) = \psi*(k) \cdot \varepsilon(t) + \psi*(k+1) \varepsilon(t-1) \quad \dots (11)$$

k-段階後의豫測誤差의分散은

$$E[(Y(t+k) - Y(t+k|t))^2] = (1 + \phi(1)^2 + \dots + \phi(k-1)^2)\sigma_{\epsilon^2} + \sum_{j=1}^{\infty} \{\phi(k+j) - \phi^*(k+j)\}^2 \sigma_{\epsilon^2} \dots \quad (12)$$

이다. 이를 最小로 하기 為해서는 式(12)의 두번째項이 零이 되어야한다. 結果的으로 式(11)은 式(10)의

$$g'G'hH \xrightarrow{\varepsilon(t)}$$

$$C(k) \text{ 가 되며前述한바와같이豫測誤差, } Y^u(t+1), \text{ 은} \\ Y^u(t+1) = Y(t+1) - Y^b(t+1)$$

$$Y^u(t+1) = Y(t+1) - Y^b(t+1) \\ = \varepsilon(t+1)$$

이 된다.

豫測不可能한 $Y^u(t)$ 의 白色性은豫測誤差를 把握하는 觀點에서 뿐만아니라 Model을 推定하는데도 重要的한 열쇠가 된다. 따라서 式 (5)를 $\epsilon(t) = g(t)Z(t)$ 로 變換해서 생각하면 g 가 얼마나 真值에 가까운가 하는것이 Model의 成敗를 좌우한다.

IV. 모델의 同定과 評價

時系列 $Y(t)$ 를 $Z(t)$ 로 變換하기 為해서는 可能한
集合 $(d, D)_1$ 中에서 最適의 것을 選擇해야한다. 過去
여러學者⁽¹⁾⁽²⁾⁽³⁾들에 依하여 研究된 바에 依하면
 (d, D) 는 $(3, 3)_1$ 以上을 넘지 않는 것으로 알려져 있
다. 따라서 $Y(t)$ 的 自己共分散分析에 依하거나 經驗
點에 依해 λ 가 推定되었다고 하면 上記 $(d, D)_1$ 中에
서 九個의 集合에 對해서만 調査하면 充分하다. 未知
의 變數 $(\alpha, \beta, \phi, \theta)$ 的 數를 決定하는 方法으로서
는 偏自己相關函數를 使用하는 경우가 많다. 定常時
系列 $Z(t)$ 를 AR模型으로 나타내면

$$Z(t) = \phi_1 Z(t-1) + \phi_2 Z(t-2) + \dots + \phi_j Z(t-j) + \varepsilon(t)$$

로 되며 또한自己相關函數의 Generating Function은 다음과 같이定義된다.

式(18)에서 ϕ_{kk} 는 偏自己函數로 定義되어 $\phi_{jj} (j \neq k)$ 는 ϕ_{kk} 들 중에서 最後의 零이 아닌 係數를 나타내므로 이로부터 p, P 를 求할 수 있다. 따라서 偏自己相關函數는 AR의 次數를 決定해준다. AR의 次數와 MA의 次數를 같게한다면 $(p, d, q) \times (P, D, Q)$ 는 決定된다. 決局 ARIMA 模型은 다음과式에서 係數集合 θ 를 決定하는 問題로 考察되게 된다.

$$\varphi(t) = [-Z(t-1) \cdots Z(t-p-P), \varepsilon(t-1) \cdots \varepsilon(t-q-P)]$$

$$\theta = [\alpha(1), \dots, \alpha(p+P), \beta(1), \dots, \beta(q+Q)]^T$$

$$Z(t) = \theta^T \cdot \varphi(t) + \varepsilon(t) \dots \dots \dots (14)$$

式(14)의 解를 求하는 方法은 크게 두 가지로 分類될 수 있다. 첫째는 Brophy와 Salzar¹²⁾등에 依해 提案된 模擬技法과 前章에서 記述한 豫測誤差를 利用하는 方法이다. 그러나, 模擬技法은 一般的으로 豫測을 爲해서는 適合치 않은것으로 알려지고 있다.

前記 두個方法中 後者는 豫測誤差의 分散을 最小로 하는 方法으로서 이의 推定值 $\sigma_{\varepsilon}^2(\theta)$ 는 다음과 같이 表示될 수 있다. 즉

$$\sigma_{\varepsilon}^2(\theta) = \frac{1}{T} \sum_{t=1}^T \varepsilon^2(t) \dots \dots \dots (15)$$

$\sigma_{\varepsilon}^2(\theta)$ 를 最小로 하는 方法中 最尤推定方法(maximum likelihood method)의 적용을 $Z(t), \varepsilon(t)$ 들이 각각 獨立이라는 假定下에서는 非線形最小自乘推定法¹³⁾이 널리 쓰여지고 있다. 여기서 $\sigma_{\varepsilon}^2(\theta)$ 를 "Loss function"이라 부른다. 또한 이 方法은 初期值에 依해 式(15)의 수렴 成敗가 좌우되기 때문에 ARIMA $(p, d, q) \times (P, D, Q)$ 의 初期值은 ARIMA $(p, d, q) \times (P, D, Q)$ 에 對한 Yule Walker⁹⁾公式의 解에 依해 주어지는 경우가 많다. 즉

$$\left. \begin{array}{l} \sum_{j=0}^p \alpha(j) R(j-v) = 0 \\ \sum_{k=0}^q \beta(k) R_i(k-v) = 0 \\ \alpha(0) = \beta(0) = 1 \end{array} \right\} v = 1, 2 \dots \dots \dots n \quad (16)$$

으로 表示되는 Yule-walker公式의 解는 雜音 $\varepsilon(t)$ 의 推定分散을 求하는데 使用되고 Åström¹⁰⁾의 積分定理는 式(16)의 推定值들로 構成된 模型의 安定與否를 말해줄과 同時に 安定일 때는 다음式의 積分值가

$$\sigma_z^2 = \sigma_{\varepsilon}^2 \frac{h(z)h(z-1)}{g(z)g(z-1)} \frac{dz}{z} \dots \dots \dots (17)$$

"Recursive Integration"으로 주어짐을 뜻한다. 實際로 이들 初期值들은 解析的인 解를 求하기는 어려우므로 電算機를 利用한 數值解析에 依存하고 있다.

最小分散을 갖도록 制御된 模型이 最適의 模型인지의 與否는 初期模型의 残差가 白色雜音 $\varepsilon(t)$ 와 같은 特性을 가진다는 假定의 成立與否에 對한 試驗을 함으로서 實施할 수 있다. 이 實施方法으로서는 残差의 自己相關係數 $\rho_{\varepsilon}(k)$ 를 利用하여 $\rho_{\varepsilon}^2(k)$ 의 合 측

$$Q = T \sum_{k=1}^m \rho_{\varepsilon}^2(k) \dots \dots \dots (18)$$

Q 를 自由度 m 을 가지는 χ^2 分布로 假定, χ^2 檢正을 하는 方法과 Cumulative spectral density나 Cumulati-

ve Periodogram을 直線인가를 檢討하는 方法이다. 後者에는 kolmogorov-smirnov 檢正方法에 依하여 이를 檢正할 수 있다. 즉 여기서 使用되는 Spectral Density의 推定值은

$$P(f) = 2 \{1 + 2 \sum_k \rho_{\varepsilon}(k) \cos 2\pi f k\} \dots \dots \dots (19)$$

로서 계산될 수 있다. Fig. 2는 以上的 模型推定過程을 圖示한 流程圖이다.

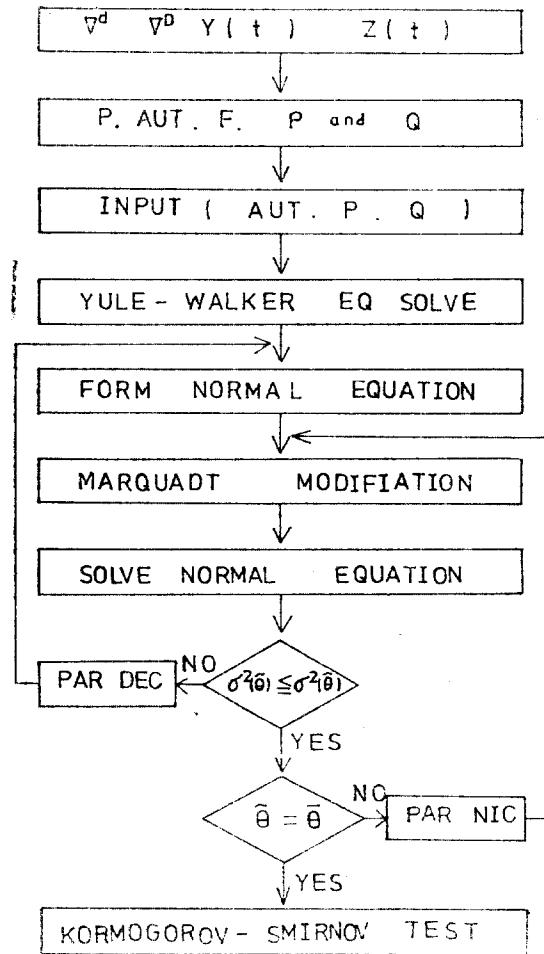


Fig. 2 Flowchart to identify
A Model (ARIMA)

V. 모델의 適用例

ARIMA模型의 適用可能與否를 檢討하기 위하여 漢江流域의 6個地點의 月流量時系列(1917~1930年)에 對한 資料를 適用하였다. 前章에서 記述한바와 같이 $g(L)G(L\lambda)\nabla^d\nabla^{D_1} Y(t) = h(L)H(L\lambda)\varepsilon(t)$ 을 一般的의 ARIMA型으로 同定하기 爲해서 (d, D) 를 決定하여 差分時系列을 얻고, 偏自己相關函數로 부터 變數의 數를 推定하여, 豫測誤差가 最小分散을 갖

도록 制御되어야 한다. 이러한 一聯의 過程에 對해 상세한 記述은 昭陽江 資料에 限하여 나머지 五個時系列에 對해서는 過程을 생략하고 結果만을 表 1에 수록하였다.

	ϕ^2	ϕ_1	ϕ_2	ϕ_3	β_1	β_2	β_3	β_4	θ	d , τ
HWA-	0.755	0.102			0.052					(1, 1)
CHON	0.448	0.376			0.912					
DAL-	0.453	0.453	-0.019		0.285	-0.034				(1, 1)
CHON	0.376	0.212	0.358		0.039	0.227			0.022, 0.845	
FAL-	0.582	-0.818	0.529	-0.04	0.022	0.657	0.108			(1, 1)
DANG	2.300	-0.059	-0.739	0.220	0.446	-0.452	0.681	0.045	0.737	
SOY-	0.642	-0.222	-0.053		0.196	0.125				(1, 1)
ANG	0.362	0.866	-0.368		0.608	0.366		0.042	0.848	
HONG-	0.719	-0.434	-0.001	-0.057	0.128	0.364	-0.037			(1, 1)
CHON	0.466	-0.284	0.034	0.002	0.102	0.295	0.148	-0.011	0.745	
KAN-	2.164	0.453	-0.018		0.284	-0.032				(1, 1)
HYON	368	0.220	0.385		0.846	0.251				

Table 1 Models of Monthly River Flows

Fig. 3은 昭陽江月流量時系列에 對한 原時系列과 差分時系列의 自己相關函數를 나타낸것인데, Fig. 3(a)에서 보인바와 같이 $Y(t)$ 에 對한 自己相關函數는 明確한 週期性을 나타내어 주고있으며 差分時系列 $\nabla_{12}Y(t)$ 에 對한 自己相關函數 Fig. 3(b)에서는 原時系列에서 1年週期는 除去된것으로 보여지나 $\sum|R(v)| < \infty$ 의 條件을 滿足하지는 않는다. 이에 比해 $\nabla\nabla_{12}Y(t)$ 는 無作爲性을 보여주고 있으며 (Fig. 3(c)) 特히 $V>15$ 以下 값들에 對해서는 自己相關係數는 零에 가깝다.

Fig. 4는 이들 두 差分時系列의 偏自己相關函數이며 $k>13$ 일 경우의 偏自己相關係數들의 值은 아주 적은 值들을 나타내어 주고있다. 따라서 $\nabla\nabla_{12}Y(t)$ 의 時系列은 傾向이나 週期가 거의 除去되었다고 判斷되며 $(3, 1, 3) \times (3, 1, 3)_{12}$ 以下의 낮은 次數로도 昭陽江의 月流量時系列를 나타낼 수 있음을 보여준다.

表 2는 Åström의 定理에 依해 安定度가 檢討된 Yule-Walker-EQ의 數值解이다.

表 2에서豫測模型의 初期值로서 分散이 제일 적은 $(3, 1, 3) \times (0, 1, 1)$ 를 指하여 主週期 12에 對한 加重值를 考慮해서 推計的 制御를 Fig. 2의 過程에 따라 實行

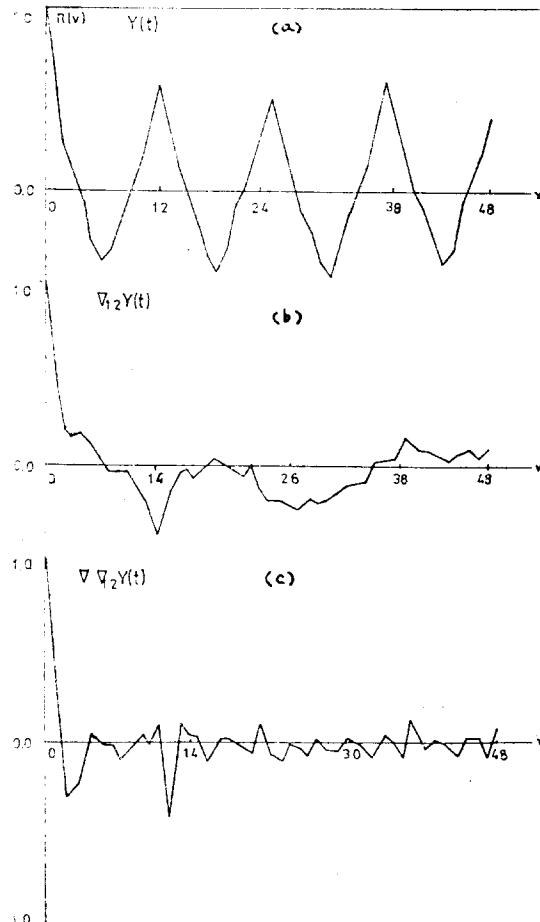


Fig. 3 A Family of Autocorrelations

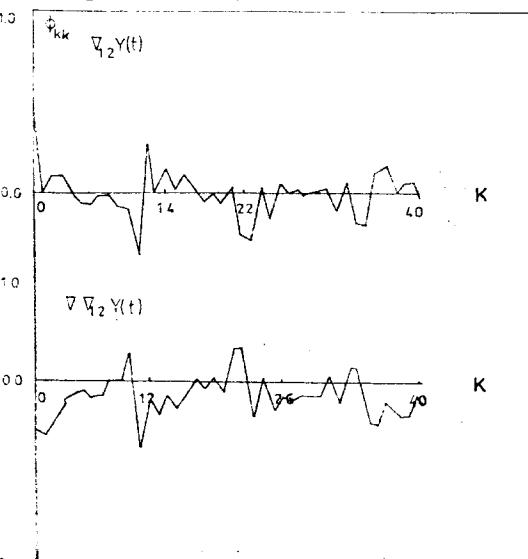


Fig. 4 Partial Autocorrelations of Monthly Stream Flow for Soyang River

$(p, d + q)$	$\hat{\alpha}_q^2$	α_1	α_2	α_3	β_1	β_2	β_3
(1, 1, 1)	0.72	-0.202			0.132		
(2, 1, 2)	0.64	-0.222	0.109		0.135	0.215	
(3, 1, 3)	0.58	-0.297	-0.035	-0.000	0.236	0.351	0.146

Table. 2 Solution of Yule-walker EQ

했다. 이로 부터 얻은 加重值들과 “Loss function”들을 Fig. 5에 나타내었는데 加重值의 變化에 對해 Loss Function이 線型的으로 變化는 角視點의 雜音差分時系列成分들이 서로 獨立이며 1次 moment가 零인 正規分布를 한다는 假定이 證明된 셈이다.

또한 分散이 0.58에서 0.38로 0.2程度 줄었음은 이 模型의 適合性을 나타내주고 있다. 前記한 模型에 평均分散 0.38에서의 變數들을 對入하면

$$(1+0.149L^1+0.191L^2-0.169L^3)(1-0.043L^{12})$$

$$\Delta \Delta_{12} Y(t) = (1-0.367L^1-0.230L^2-0.377L^3)(1-0.808L^{12})\varepsilon(t) \dots \dots (20)$$

이다.

Fig. 6은 残差의 Cumulative Spectral density가 kolmogorov-smirnov의 75%, 95% 限界内에서 有意함으로 残差는 白色雜音이라는 것을 證明하여 주고 있다. 變數의 共分散 matrix는

$$V(\theta) = Cov[\theta, \theta^T]$$

$$= 0.38 \left(\begin{array}{cccccccccc} 0.99E-04 & 0.274E-06 & 0.381E-06 & 0.294E-08 & -0.294E-08 & -0.693E08 & -0.590E-08 & 0.201E-08 \\ 0.990E-04 & 0.243E-06 & -0.569E-08 & 0.399E-09 & -0.947E-08 & 0.140E-06 & -0.881E-06 \\ 0.990E-04 & -0.377E-08 & -0.34E-08 & 0.240E-08 & 0.282E-08 & -0.862E-08 \\ 0.990E-04 & -0.976E-08 & -0.966E-08 & 0.295E-08 & -0.479E-08 \\ 0.990E-04 & -0.957E-08 & 0.245E-08 & -0.432E-08 \\ 0.990E-04 & 0.308E-08 & -0.473E-08 \\ 0.990E-04 & 0.774E-08 \\ 0.990-04 \end{array} \right)$$

이고 여기서 對角線값에 比해서 나머지 값들은 아주 적다. 따라서 各推定值가 獨立되어 있음을 알 수 있다.

表 3은 標準誤差를 考慮한 推定值의 값들이다.

Fig. 7은 1940年の豫測值와 實測值들의 比較인데 加重值가 많으면 끌을수록 좋은 模型이라는 것이 通念

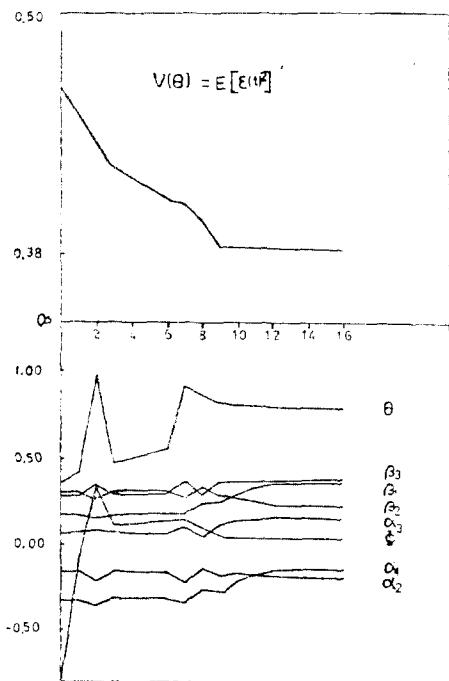


Fig. 5 Behavior of coefficients and Loss Function

이나 그림에서 알 수 있듯이 係數의 精確度가豫測을支配하는 데에 기여하는 듯한 결과를 보여준다.

이豫測值는 1939년의 値들이 比較的 많이 반영되었다고 볼 수 있는데 實測值가, 1940年最高值는 1939年最高值의 무려 10배나 되었다.

α_1	α_2	α_3	β_1	β_2	β_3	δ	θ
0.149 ± 0.003	0.191 ± 0.003	0.169 ± 0.003	0.367 ± 0.003	0.230 ± 0.003	0.377 ± 0.003	0.642 ± 0.030	0.01 ± 0.003

Table. 3 Weights and Standard Errors

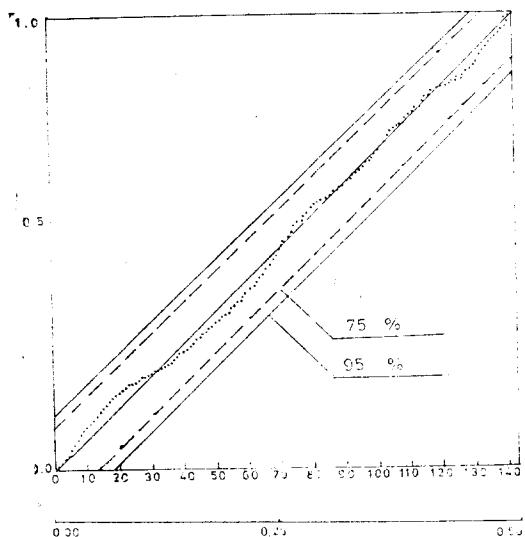


Fig. 6 Cumulative Spectral Density check on Residuals

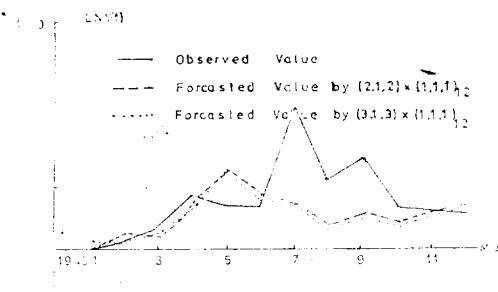


Fig. 7 1-year Forecasting

VI. 結 論

韓國主要河川의 月流量時系列의 ARIMA過程을 짜르는가를 檢討하기 為해 六個地點(1919~1939年)의 觀測資料를 使用하여 模型化를 試圖하였다. 그 結果 使用된 資料는 低次數의 差分時系列이 定當時系列로 取及될 수 있었고 ARIMA型에 依한 白色化는 보장되었다. 따라서 月流量時系列은 ARIMA模型에의 適合性을 보여주고 部分時系列에 依한 方法보다 簡單한 形態의 模型을 얻을 수 있었다.

豫測에 對해서도 實現值의 動向을 따르는 比較的 좋은豫測結果를 얻었다. 또한 變數의 精確度보다는

差分이 豫測에 미치는 영향은 절대적이었다.豫測技法에 바탕을 두어 推定된 ARIMA模型은 模擬發生模型으로도 有用하다. 이러한 技法이 土木計劃의 最適運營體系를 為해서 많이 利用될것이 期待된다.

參 考 文 獻

1. VUJICA YEVJEVICH; Stochastic Process in Hydrology p.8 Water Repources Publications Fort Collins, Cororado, U.S.A.
2. 長谷部正彦; Box & Jenkinsの理論による 非定常降水量時系列の解析と豫測について 日本土木學會論文報告集 第261號 5. 1977
3. B.M. METHA & R.C. AHLERT & S.L. Yu; "Stochastic Variation of water Quality of the Passaic River" Water Resources Research Vol. 11. No. 2 4. 1974
4. G.E.P. Box & G.M. JENKINS; Time Sejies, Forcasting, & Control, Sanfransisco, Calif. Holden Day 1970
5. RAMACHANDRA A. RAO & R.L. KASHYA P; "Stochastic Modeling of River Flows" IEEE Transactions on Auto. Cont. Vol. AC-19, Nos 6 12. 1974
6. Robert B. ASH & MELVIN F. GARDANER; Topics in Stochastic processes p. 65 Academic Press 1975
7. K.E. KALMAN & R.S. BUCY; "New Results in Linear Filtering & prediction Theory", J. of Basic Engineering 1961
8. LENNARI LJUNG, Torsten Södersröm & IVAR GUSTAVSSON; "Counter Examples to General Convergence of a Commonly used Recursive Identification Method", IEEE Transaction Aut. Cont. Vol. AC-20, No. 5 8, 1975
9. EMANUEL PARZEN; "Some Recent Advances in Time Series Modeling" IEEE Transuction Aut. Cont. Vol. AC-19 No.6, 12, 1974.
10. Åström; Introduction to Stochastic Control Theory p.227
11. EMANUEL PARZEN; "Some Solutions to the time Series Modeling and Prediction Problem", Stâtist. sci. Division, State Univ. New York, Buffalo, N.Y. 5, 1974
12. F. BROPHY & A.C. SALZAR; Synthesis

of Spectrum Shaping Digital Filters of Recursive Design, IEEE Transactions on Circuits and Systems, Vol. CAS-22, No. 3, 3, 1975

13. DONAL W. MARQUARDT; "An Algorithm

For Least-Squares Estimation of Nonlinear Parameters," J. Soc. Indust. APPL. Math. Vol. 11, No. 26, 1963.