Transfer function 에 의한 수문시계열 모형

인하대학교 토목공학과 교수 강 관 원 *
인하대학교 대학원 토목공학과 김 주 환 **

1. 서 론

복잡한 자연현상의 하나인 수문과정을 모형화하는 것은 실제로 어려운 일이나 많은 수문학자들은 이러한 과정을 여러가지 수학적 방법을 이용하여 모형화하고 수문학에 있어서의 연구의 대상으로 삼아왔다. 수문학의 모형화 방법에는 결정론적 모형화와 추계학적 모형화 방법으로 구분된다.

결정론적 모형은 많은 시간과 방대한 자료를 요구한다. 또한 모형의 정확도는 필요한 자료가 많아야만 유지될 수 있는데 실제로 우리가 구할수 있는 자료는 한정되어 있는 것이 현실이다. 따라서 입력자료가 비교적 간단한 추계모형은 수문과정을 모형화하는데 더욱 광범위하게 사용되고 있다.

자료분석을 위한 많은 방법들은 대부분의 관찰치들이 서로 독립이라 가정한다. 그러나 관찰치들이 시간에 관하여 서로 종속적으로 연관된 경우 복잡하고 어렵게 여겨져 왔으나 실제의 관찰치들은 서로 종속적인 형태로 발생한다. 이러한 종속된 계열의 분석 기술이 시계열분석이며 이는 일련된 시간의 변동에 따른 자료를 현시점에서 과거의 관측된 자료를 이용하여 미래의 값을 예측하는 자료분석 방법이다. 이와 같이 우리가 시계열을 분석하는 목적은 크게 두가지이다. 첫째는 시계열 자료를 관찰하고 분석하므로써 주어진 자료를 발생시키는 확률적 체계를 이해하고 모형화한다. 둘째는 과거의 자료를 가지고 미래의 값을 예측하는데 있다.

따라서 본 연구에서는 수문시계열 과정의 모형화와 예측을 위하여 입력(input)과 출력(output)간의 상관관계를 모형화하는 이산형 linear transfer function model 을 사용하여 수문과정에 적용하고 이러한 모형에 대한 등정 (identification), 추정(estimation), 검진(diagnostic checking)에 대하여 고찰하는 것이다. 즉 시간에 따라 변화하는 수문자료에 대하여 이에 적합한 모형을 추정하였으며 추정된 모형에 대한 타당성을 통계학적 방법으로 조사, 검토하였다.

2. Transfer function model

2.1 Transfer function model 의 개요

X_t 를 어떤 시스템의 입력치라하고 Y_t 를 입력치의 영향을 받아 나타나는 출력치 이라 하자. 그리고 어떤 시스템에 입력치 X_t가 가해졌을 때 즉각적인 효과는 나타나지 않으나 동등(equilibrium) 하게 되기위한 출력치 Y_t가 있는 반응을 동적반응(dynamic response)이라하고 이는 현재의 출력치계열(output series)의 값 Y_t가 많은 과거의 입력치계열(input series)의 값 X_t와 관계가 있다고 가정한다면 이들의 관계를 다음식으로 나타낼 수 있다.

$$
\begin{aligned}
Y_t &= v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \cdots\cdots \\
&= (v_0 + v_1 B^1 + v_2 B^2 + \cdots) X_t \\
&= v(B) X_t \qquad (1)
\end{aligned}
$$

여기서, v_0 , v_1 , v_2 ,$\cdots$: 충격반응 함수(impulse response function),
$v(B) = (v_0 + v_1 B^1 + v_2 B^2 + \cdots)$: transfer function,
B : 후향연산자(backwardshift operator)

(1)식을 그림으로 나타낸 것이 그림 1 이다. 동적시스템(dynamic system)은 입력치계열 X_t를 transfer function 을 통하여 X_t값을 통과시키므로써 출력치계열 Y_t로 변환시킨다. transfer function 의 특성은 충격반응함수(impulse response function)인 가중치(weight) v_0, v_1, v_2, v_3, $\cdots\cdot$에 의하여 결정된다. (1)식은 이론적으로 무한히 많은 수의 계수들을 포함하는 v(B)는 추정하기 곤란하므로 다음과 같은 유리함수로 표현하여 해결할 수 있다.

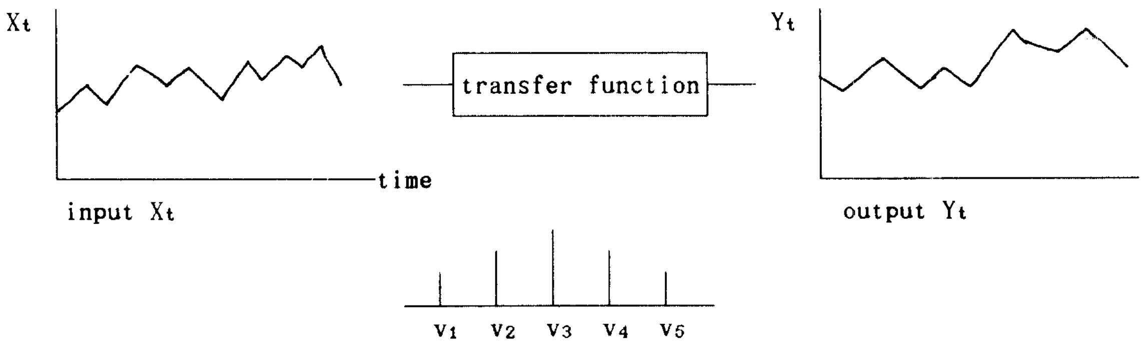

Fig. 1 Schematic representation of transfer function model.

$$v(B) = \delta^{-1}(B)\ \omega(B)\ B^{b}$$
$$= \delta^{-1}(B)\ \Omega(B) \qquad (2)$$

단, $\delta(B) = (\ 1 - \delta_1\ B^1 - \delta_2\ B^2 - \cdots - \delta_r\ B^r\ )$

$\omega(B) = (\ \omega_0 - \omega_1\ B^1 - \ \omega_2 B^2 - \cdots \omega_s\ B^s\ )$

(1)식과 (2)식을 합성한다면,

$$Y_t = \delta^{-1}(B)\ \omega(B)\ B^{b}\ X_t$$
$$= \delta^{-1}(B)\ \omega(B)\ X_{t-b} \qquad (3)$$

여기서, b는 지연변수(delayed parameter)로서 반응이 즉각적이지 못할 경우 시간이 얼마나 지연되는가를 알아보는 母數이다. (3)식이 아무리 정확하다 하더라도 출력치 Y_t 가 이 모델에 의해 결정된 일정한 형태를 반드시 따르리라고는 예상할수 없다. 즉, 예측하지 못했던 여러가지 오차(disturbance or noise)가 생기는데, 이 오차 N_t와 input X_t 는 독립이라 가정하고 N_t 자신은 자기상관 되어 있다고 가정한다면 다음과 같은 새로운 모델이 생긴다.

$$Y_t = \delta^{-1}(B)\ \omega(B)\ X_{t-b} + N_t$$
$$= \delta^{-1}(b)\ \omega(B)\ X_{t-b} + \varphi^{-1}(B)\ \theta(B)\ a_t \qquad (4)$$

단, $N_t = \varphi^{-1}(B)\ \theta(B)\ a_t \sim$ ARiMA(p, d, q)

$\varphi(B) = (\ 1 - \varphi_1 B - \varphi_2 B^2 - \cdots\cdots \varphi_{p+d}\ B^{p+d})$

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

a_t : 정규분포를 갖는 residual

여기서, N_t는 ARiMA(p,d,q) 과정으로 본다면 N_t 자신의 process를 나타낼수 있는 새로운 모델이 생긴다. 이러한 모형을 transfer function noise model 이라하며 다음과 같은 그림으로 나타낼수 있다.

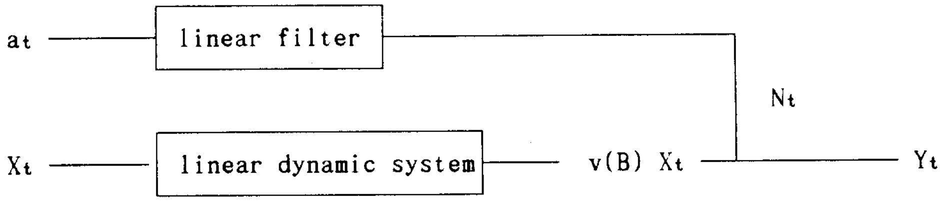

Fig.2 Transfer function model with added noise model.

만약 X_t와 Y_t가 비정상시계열(nonstationary time series)라면 적절히 계차(differencing)시켜 정상시계열(stationary time series)로 만들면 (4)식은 다음과 같다. 즉, $x_t=(1-B)^d X_t$, $y_t=(1-B)^d Y_t$, $n_t=(1-B)^d N_t$ 이다.

$$\begin{aligned} y_t &= \delta^{-1}(B)\ \omega(B)\ x_{t-b} + n_t \\ &= \delta^{-1}(b)\ \omega(B)\ x_{t-b} + \varphi^{-1}(B)\ \theta(B)\ a_t \qquad (5) \end{aligned}$$

2.2 Transfer function model 의 설정

2.2.1 transfer function model 의 설정

transfer function model의 자료분석에서 입력치와 출력치사이의 고차상관함수(cross correlation function)는 중요한 역할을 한다.

(X_t, Y_t)의 이변량추계과정(bivariate stochastic process)에서 x_t와 lag-k에서 y_t와의 고차공분산(cross covariance)은

$$\gamma_{xy}(k) = E[(x_t - \mu_x)(y_{t+k} - \mu_y)], \qquad k=1,2,3,\cdots$$

이고, y_t와 lag-k에서 x_t와의 고차공분산은

$$\gamma_{yx}(k) = E[(y_t - \mu_y)(x_{t+k} - \mu_x)], \qquad k=1,2,3,\cdots$$

이 된다. 이를 이용하여 시차(lag) k에서의 x와 y사이의 고차상관함수는 다음과 같이 정의 된다.

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \, \sigma_y} \quad , \quad k= 0,\ \pm1,\ \pm2,\ \cdots \tag{6}$$

σ_x : 계열 x_t의 표준편차

σ_y : 계열 y_t의 표준편차

교차상관함수는 자기상관함수와는 달리 lag k=0 에서 대칭이 아니므로 $\rho(k) \neq \rho(-k)$가 된다.

transfer function 모형을 설정하기 위한 첫단계는 input x_t가 추계학적 구조(stochastic structure)를 가진경우 백색잡음(white noise)으로 변환시키기 위하여 사전백색화(prewhitening)시켜야 한다. 이는 입력계열이 백색잡음일 경우 등정(identification)과정이 상당히 단순해지기 때문이다. 따라서 입력치 X_t를 계차시켜 정상상태로 만들면 이 때의 x_t는 ARIMA model을 갖게된다.

$$\phi(B)\ x_t = \theta(B)\ \alpha_t$$

$$\phi(B)\ \theta^{-1}(B)\ x_t = \alpha_t \tag{8}$$

α_t : 사전백색화된 입력치

즉, 입력치 x_t를 $\phi(B)\ \theta^{-1}(B)$ 라는 여과기능(filter)를 이용하여 사전백색화시켰다.

여기서 구한 여과기능을 이용하여 y_t도 사전백색화 시킨다.

$$\phi(B)\ \theta^{-1}(B)\ y_t = \beta_t \tag{9}$$

(8)식과 (9)식을 이용하여 다음과 같은 변형(transformation)을 할수 있다.

$$y_t = v(B)\ x_t + n_t$$

$$\beta_t = v(B)\ \alpha_t + \varepsilon_t, \qquad \varepsilon_t = \phi(B)\ \theta^{-1}(B)\ n_t \tag{10}$$

여기서 ε_t는 변형된 백색잡음 과정이다.

(10)식은 입력치 x_t 와 출력치 y_t 사이의 transfer function은 사전백색화된 α_t와 β_t 사이의 transfer function 과 같다는 것을 보여준다.

충격반응함수 v_k는 다음식에 의하여 구한다.

$$v_k = \frac{\gamma_{\alpha\beta}(k)}{\sigma_\alpha^2} = \frac{\rho_{\alpha\beta}(k)\ \sigma_\beta}{\sigma_\alpha}\ ,\ k=1,2,\cdots\cdots \tag{11}$$

2.2.2 잡음모형(Noise model)의 설정

transfer function model이 결정되면 잡음항(noise term)을 결정하는 일만 남게 된다. 잡음항은 초기 추정치 ω와 δ를 이용하여 아래와 같이 얻을 수 있다.

$$y_t = v(B)\ x_t + n_t$$
$$n_t = y_t - v(B)\ x_t \tag{12}$$

n_t는 식(13)을 이용하여 구하고 n_t의 표본자기상관함수(sample autocorrelation function : SACF)와 표본부분자기상관함수(sample partial autocorrelation function : SPACF)을 보고 ARiMA 모형을 설정한다.

이렇게 구한 transfer function model과 noise model을 결합하여 완성된 모형은 다음과 같다.

$$y_t = \frac{\omega(B)}{\delta(B)}\ x_{t-b} + \frac{\theta(B)}{\phi(B)}\ a_t \tag{13}$$

3. 분석 및 고찰

강우에 의하여 나타나는 영향은 유출에 일정한 충격을 줄 것이며 이러한 관계를 transfer function 과 비교하여 월강우량자료를 input 계열로, 월유출량은 output 계열로 가정하여 강우와 유출관계를 분석해 보기로 한다.

이들 시계열자료가 등분산성을 갖기위한 방법으로 대수변환(log transformation)한 후 사용하였다. 이는 시계열분석에서 잔차들이 이질성(heteroscedasticity)을 가질경우 오차항이 동일한 분포를 갖는다는 가정에 부합되지 않는다. 이러한 분산의 이질성문제를 해결하기 위한 방법으로 대수변환

한 자료를 사용하게 된다.(Box-Cox transformation)

3.1 모델의 설정

그림 3과 4는 강우량과 유출량의 시계열 분포도이다. 그림3에서 알수 있듯이 강우량자료가 비정상성(nonstationarity)를 가지므로 계차조작(differencing operation)시키면 이때의 x_t는 ARIMA과정을 갖게 된다. 따라서 강우자료에 대한 SACF와 SPACF를 구하여 적당한 모형을 찾아낸다.

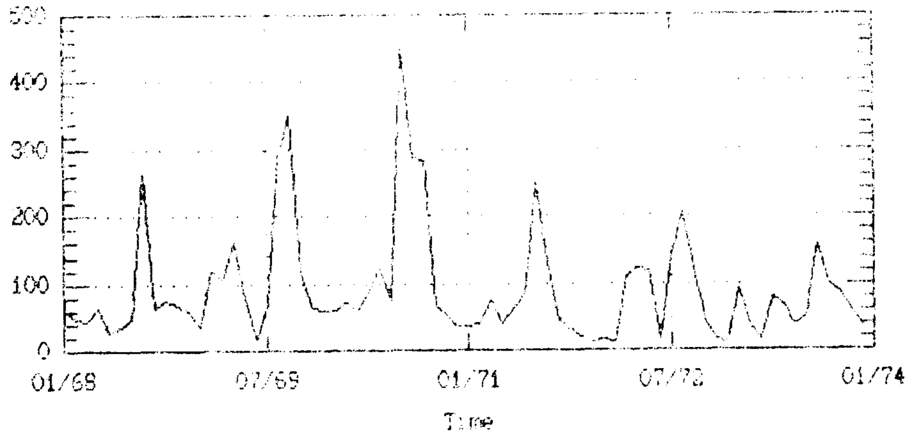

Fig.3 The input series (rainfall) shown graphically.

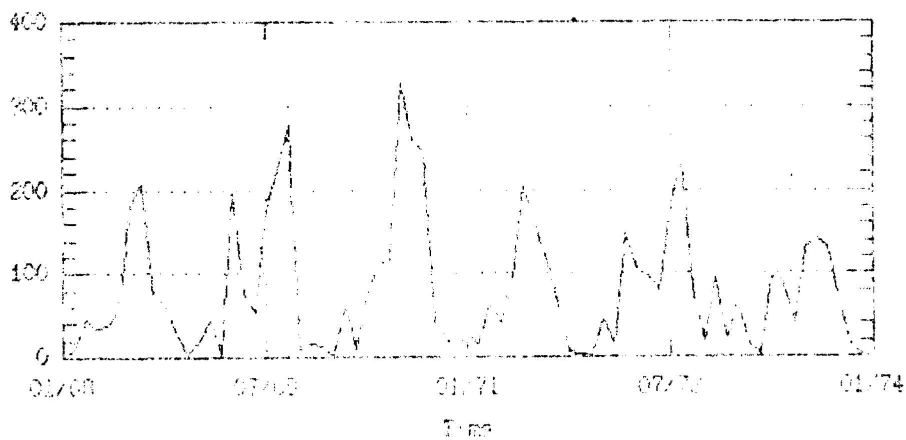

Fig.4 The output series (runoff) shown graphically.

위의 결과를 고려하여 잠정적 모형(preliminary model)은 다음과 같이 가정할수 있다.

$$y_t = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1 - \delta_1 B - \delta_2 B^2)} x_t + n_t \qquad (14)$$

이제 잡음항에 대한 모형을 설정하기 위하여 잡음항에 대한 잠정모형을 다음과 같이 가정하였다.

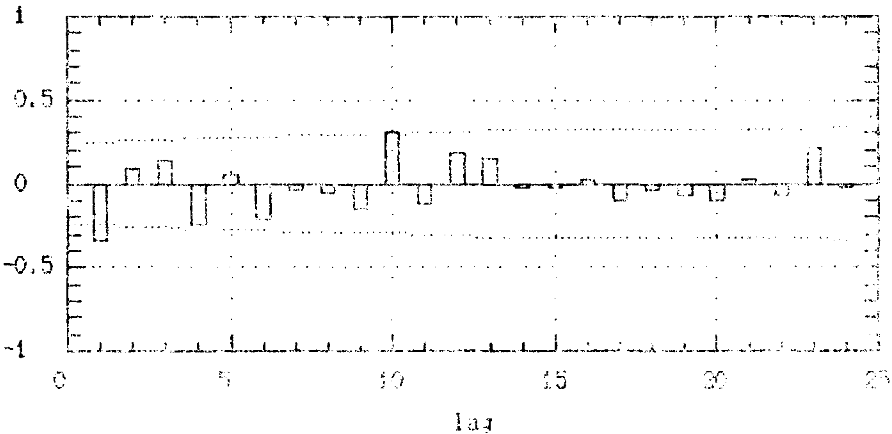

Fig.5 The estimated autocorrelation coefficients for differenced input series.

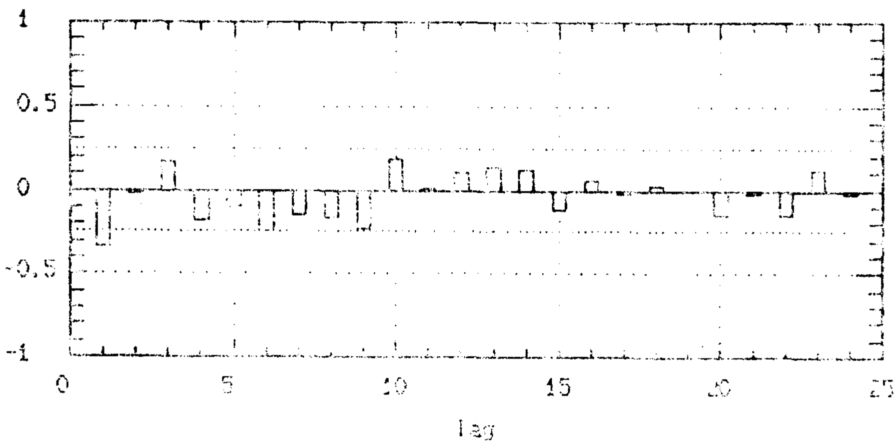

Fig.6 The estimated partial autocorrelation coefficients for differenced input series

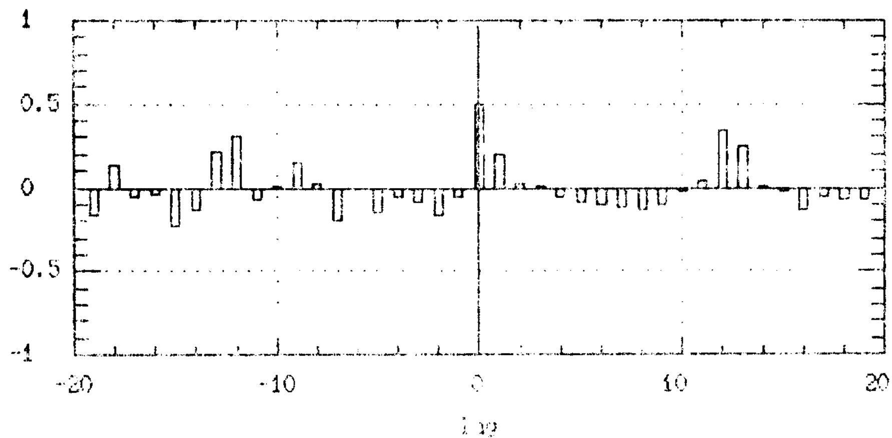

Fig.7 The estimated cross correlation coefficients between the prewhitened input series(α_t) and the prewhitened output series(β_t).

$$n_t = \frac{(1 - \theta_1 B - \theta_2 B^2)}{(1 - \phi_1 B)} a_t \qquad (15)$$

이렇게 하여 구한 transfer function noise model은 식(14) 과 식(15)을 합하여 다음과 같이된다.

$$y_t = (\omega_0 - \omega_1 B - \omega_2 B^2)\, x_t + \frac{(1 - \theta_1 B - \theta_2 B^2)}{(1 - \phi_1 B)} a_t \qquad (16)$$

3.2 모형의 검진(model diagnostic checking)

모형의 검진은 잔차분석(residual analysis)을 통하여 이루어지며 추정된 모형이 타당한 것이라면 다음의 가정을 만족하여야 한다.

첫째, 잔차들의 평균은 0 이고 분산은 σ_a^2인 백색잡음과 비슷한 성질을 가져야만 하며 이를 등분산성의 가정이라한다.

둘째, 잔차는 서로 독립이다. 즉, 잔차와 잔차간에 서로 자기상관되어 있지 않아야한다. 만약 잔차의 ACF가 어떤 추계학적 구조를 갖는다면 잠정적 모형의 등정은 잘못되었음을 의미한다.

셋째, 잔차는 정규분포를 따른다.

이러한 기준을 가지고 잔차(residual : a_t)를 분석하므로써 추정된 모형의 적합성을 검진해보기로 한다. 모형의 검진 방법으로 많이 사용되고 있는 방법으로 normal probability plot와 Box-Pierce에 의하여 제안된 Q-statistic 이다.

본 연구에서는 모형의 타당성을 검토하기 위하여 잔차들의 자기상관, 그리고 잔차와 α_t와의 고차상관을 조사하고 잔차를 정규확률지에 도시하므로써 모형의 적합성에 대하여 검진하였다.

3.2.1 Box-Pierce Chisquare test

ARIMA(p,d,q) 의 과정을 갖는 정상시계열에서 잔차 a_t가 여전히 자기상관되어 있는지의 여부를 알아보기 위한 방법이다. 즉,ARIMA 모형의 기본 가정을 만족하는지 여부를 알아보기 위하여 사용되는 통계량으로 "잔차는 백색잡음(white noise)이다" 라는 귀무가설을 검정하는 통계량이다.

만일 모형이 적합하다면

$$Q = m \sum_{k=1}^{K} r^2(k) \quad < \quad X^2{}_{(\alpha;k-p-q)} \tag{17}$$

r(k) : 잔차의 ACF

K : 충분히 큰값 (20 또는 30)

$m = n - u - p$

$u = \max.\{r, s+b\}$

n = N - d (N : 전체자료수, d : differencing 횟수)

와 같이 된다.

3.2.2 Crosscorrelation test

transfer function noise model 은 입력치와 오차사이에 독립성(independence)을 유지하여야 한다. 따라서 사전백색화된 α_t와 잔차 a_t가 서로 자기상관 되어 있는지를 알아본다. 즉, α_t가 a_t와 독립인가를 알아보기위

한 방법으로 다음의 식으로 검진될수 있다. 만일 모형이 적합하다면 다음과 같이 된다.

$$S = m \sum_{k=0}^{K} r^2(k) \quad < \quad X^2{}_{(\alpha;\, k+1-(r+s+1))} \qquad (18)$$

$r(k)$: α_t와 a_t의 교차상관함수

K : 충분히 큰값 (20 또는 30)

$m = n - u - p$

$u = \max.\{r, s+b\}$

$n = N - d$ (N : 전체자료수, d : differencing 횟수)

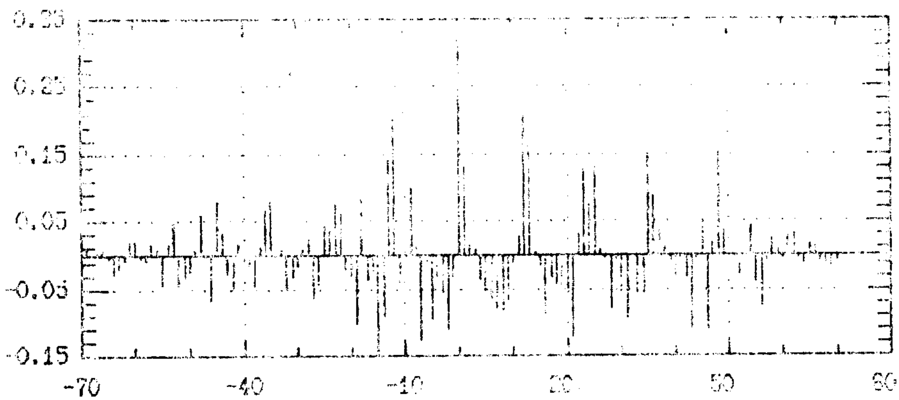

Fig.8 The estimated impulse response function.

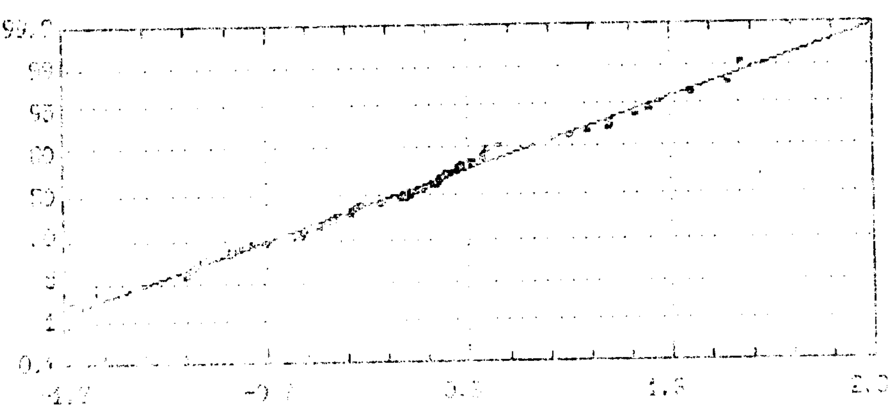

Fig.9 The normal probability plot of residual(a_t).

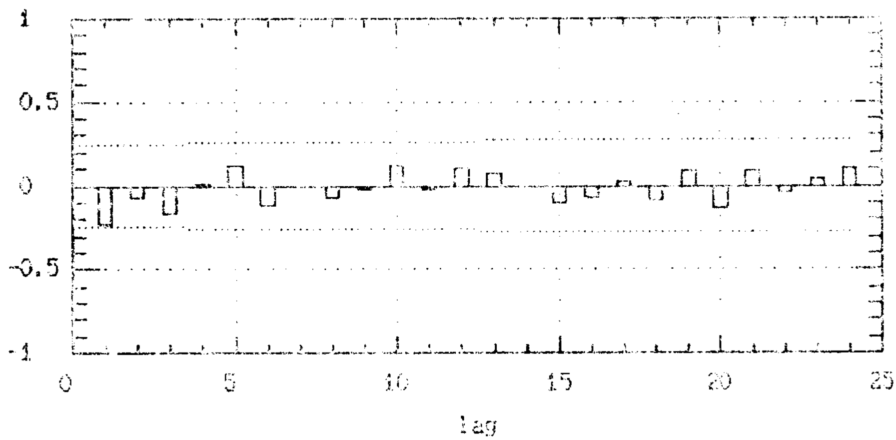

Fig. 10 The estimated autocorrelation coefficients for residual(at).

4. 결 론

Tansfer function의 오차항을 백색잡음이 아닌 어떤 시계열, 즉 ARiMA모형을 갖는다고 가정하고 월강우와 월유출량자료에 대하여 분석한 결과 입력치 X_t의 값은 서로 시간에 종속되어 있고 시간이 과거로 갈수록 Y_t에 주는 영향은 감소한다.

임의로 설정된 낙동강 유역의 용곡지점에 대한 월강우모형은 ARIMA(1,1,0)로 나타났으며 월유출량에 대하여 transfer function model을 적용하여 다음과 같은 모형을 찾아낼수 있었다.

$$y_t = (0.3253-0.1290B-0.0149B^2)\ x_t + \frac{(1-0.77266B-0.11183B^2)}{(\ 1\ -\ 0.5810\ B\ )}\ a_t$$

또한 충격반응함수 v_0, v_1, v_2, v_3에서 유의한 반응을 나타내고 있으나 v_3의 값은 아주 작은 값을 나타낸다. 따라서 월자료인 경우에는 강우가 유출에 약 2개월가량 영향을 미친다고 할수 있다. 그리고 충격반응함수의 반응은 lag-0에서부터 나타나므로 지연변수(delayed parameter) b는 0이 되는데 여기서 강우가 유출에 앞서 움직인다는 기본 성격은 나타나지 않지만 이는 비교적 시간간격이 긴 월자료를 사용했기때문으로 생각되며 시간간격이 짧은 자료를 사용한다면 지연변수의 값은 모형에 크게 영향을 미칠것이라 생각된다.

통계적 기법에 의한 수문시계열의 모형화는 여러가지 function에 의하여 나타난다. transfer function이라는 대체함수에 의하여 수문사상을 모형화하는데는 실제로 강우 이외의 여러가지 다른 변수를 포함할 것이다. 본 연구에서는 강우만을 유출에 영향을 미치는 인자로 가정하였으나 다른 수문학적 관련 인자를 고려한다면 더욱 정확한 모형화가 이루어지리라 확신한다. 따라서 single input에 의한 single output만이 아니라 여러가지 다른 인자를 고려한 multiple input에 의한 single output의 모형화와 multiple input에 의한 multiple output과 같은 결정론적 인자와 추계학적 인자를 함께 고려한 모형에 대한 적극적인 연구가 이루어진다면 좀더 정확한 수문사상을 예측할수 있을 것이며 이에대한 연구가 기대된다.