

ARMAX모형을 이용한 금강수계 월별유출량 예측

○ 권지혜¹⁾, 허준행²⁾, 맹승진³⁾, 신은우⁴⁾

1. 서론

이·치수면에서 불리한 조건을 가진 우리나라 수계를 고려할 때 대체수자원을 적극적으로 개발·활용하는 방안이나 댐 건설 등의 근본적인 대책이 절실히 요구되고 있다. 그러나 이런 방안들이 지닌 사회적·정책적인 한계로 인하여 신뢰할 수 있는 유출량의 해석과 예측의 중요성이 더욱 부각되고 있다. 본 연구에서는 유출량을 정확히 예측하기 위한 방안의 하나로 해당 지역의 시계열자료에 적합한 추계학적 모형을 구축하고, Kalman filter 기법을 적용함으로써 예측결과를 개선하고자 하였다.

본 논문에서는 유출량 예측을 위하여 금강의 대청댐 하류에 위치한 석화지점이 속해 있는 소유역과 하구둑이 속해 있는 소유역을 선정하였으며 ARMAX(AutoRegressive Moving Average eXogenous) 모형을 구성하여 예측을 수행하였다. 구성된 모형을 이용하여 해당 소유역의 2000년 월별 유출량을 예측하여 실제 자료와의 비교를 통하여 모형을 검증하였는데, 예측시 새로운 관측치가 입력될 때마다 모형의 매개변수들을 adaptive하게 재산정함으로써 선행시간에 따른 예측치를 실측치와 비교하였다. 또한, 구성된 ARMAX 모형에 대하여 기준시간에 따라 재산정되는 매개변수들을 Kalman filter에 시간대별 입력치로 고려하여 해당 매개변수들을 시불변으로 보고 수행했던 filtering 기법을 개선하고자 하였다.

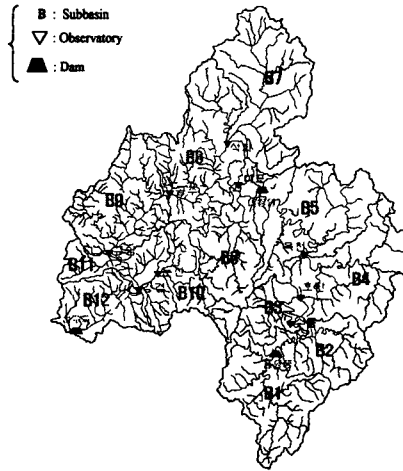
2. 연구 개요

유출량 예측에 사용되는 시계열 모형으로는 과거의 유출자료만을 이용하여 구성하는 AR(AutoRegressive), ARMA(AutoRegressive Moving Average), ARIMA (AutoRegressive Integrated Moving Average) 모형 등이 널리 사용되고 있으나, 본 연구에서는 이 외에도 유출에 영향을 미치는 강우, 기온, 선행토양함수, 상류부의 댐의 방류량에 의해 제어되는 유입량, 각종 기후자료 등의 외래유입성분에 의한 인과관계를 고려할 수 있는 전이함수계(Transfer Function)의 일종이라 볼 수 있는 ARMAX 모형을 선정하였다.

또한 기본적인 외래유입요소로는 자료구축면에서 유리하고 유출량에 미치는 영향이 가시적으로 표현되기에 가장 적합하다고 판단되는 강우량을 선정하였다. 유출량 예측 대상유역을 선정할 때에는 외래유입요소들을 조건별로 다양하게 고려하여 선정하게 되나, 본 연구에서는 필요로 하는 자료들의 신뢰성을 고려하여 상류부에서의 제약 없이 유역 내의 강우만이 유출에 영향을 미치는 외래유입요소로서 고려되는 경우에 대한 모형만을 구축하였다.

유출량 예측을 위한 모형은 시계열 모형을 사용하였으며, 적용대상 소유역으로는 <그림 1>에서 소유역 B7과 소유역 B12를 선정하였다. 대상유역에 대해 적합한 시계열 모형을 구축하기 위해 B7 소유역과 소유역 B12의 1981년~1999년까지 강우량 자료와 유출량 자료를 사용하였는데, 이들 소유역은 상류의 영향을 받지 않으므로 소유역 B7과 소유역 B12의 자연유량 이외에 입력치로서 동일 유역에 대한 강우량자료만을 사용하였다.

- 1) 정희원, 한국시설안전기술공단 진단2본부 댐항만실, 공학석사
- 2) 정희원, 연세대학교 사회환경시스템공학부 토목·환경공학전공 교수, 공학박사
- 3) 정희원, 한국수자원공사 수자원연구소 책임연구원, 농학박사
- 4) 정희원, 한국시설안전기술공단 진단2본부 본부장, 기술사



<그림 1> 금강유역의 소유역도

3. 모형의 구성

본 연구에서는 ARMAX 모형을 구성하여 유출량 예측을 수행하였다. ARMAX 모형은 기존의 ARMA 모형에 외래유입요소인 exogenous 항, X가 추가된 형태이다. 이 모형은 유출을 예측하기 위해 과거 유출자료만을 input으로 사용하는 기존의 AR, ARMA 모형 등에 비해 강우, 기후자료, 선행토양조건, 상부로부터 조절되는 유입량 자료, 특히 상류부에 댐이 있어서 대상 소유역으로의 유입이 조절되는 양 등의 외래유입요소인 exogenous term 중 적절한 사항을 고려할 수 있다는 장점이 있다. Hipel과 McLeod(1994)와 Salas 등(1980)에 의한 ARMAX 모형의 일반적인 수식은 다음과 같다.

$$y_t = \frac{\theta(B)}{\phi(B)} B^b x_t + \frac{\theta(B)}{\phi(B)} a_t = \nu(B)x_t + N_t \quad (1)$$

여기서 $\frac{\theta(B)}{\phi(B)} = \nu$: 전이함수 ϕ_j : AR 계수(매개변수), ϵ : 백색잡음, B : 후진연산자 [$B y_t = y_{t-1}$]

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$: 차수가 p인 AR operator, θ_j : j번째 MA의 매개변수

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$: 차수가 q인 MA operator, a_t : 백색잡음 ($N_t = \frac{\theta(B)}{\phi(B)} a_t$)

이를 위해 해당 소유역의 유출량 자료 외에 유출에 미치는 영향이 가시적으로 표현되기에 가장 적합하다고 판단되는 강우량 자료를 기본적인 외래유입요소로서 고려하였다.

3.1 시계열 자료

소유역 B7과 소유역 12에 대한 시계열 모형을 구성하기 위한 자료로 각 소유역의 1981년부터 1999년까지 19년간의 월간 유출량 자료와 같은 기간의 강우량 자료를 사용하였다. 여기서 유출량 자료는 해당 소유역만의 자연유량을 산정한 자료이다.

3.2 시계열 자료의 정규화 및 표준화

실제 자료에 대한 통계치를 구한 결과 대부분의 월자료가 정규분포를 따르지 않으므로 정규화 및 표준화 과정이 필요하다. 여기에서는 다음 식 (2)로 정의되는 Box-Cox 변환공식(Box and Cox, 1964)를 이용하여 자료를 정규화하였다.

$$z_t^{(\lambda)} = \begin{cases} \lambda^{-1}[(z_t + c)^\lambda - 1] & , \lambda \neq 0 \\ \ln(z_t + c) & , \lambda = 0 \end{cases} \quad (2)$$

여기서 λ : Box-Cox 매개변수, c : 위치 매개변수

3.3 차수의 결정

차수 p, q의 결정은 시계열의 상관계수를 지체시간 k에 대해 도시한 sample correlogram나 partial correlogram을 분석하여 판별하게 된다. 또한, PACF(Partial Autocorrelation Function), $\phi_k(k)$ 는 ACF (Autocorrelation Function)와 함께 모형을 판별하기 위해 사용된다.

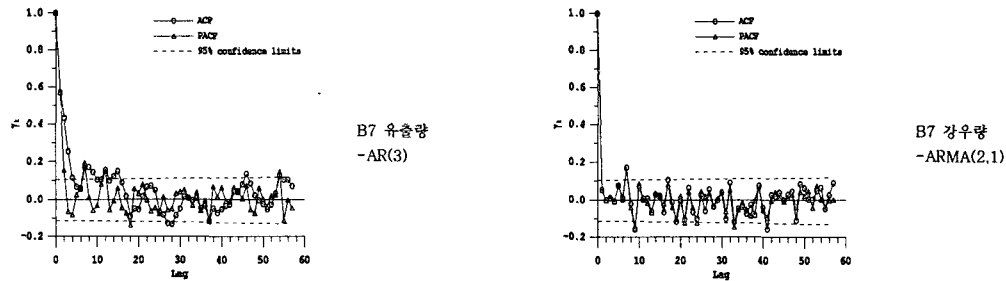
ACF와 PACF를 구하는 식은 각각 Salas 등(1980)과 Durbin(1960)의 관계에 의한 다음의 식 (3), 식 (4)과 같고, 이를 계산하여 도시하고 차수를 판별한 결과에 대하여 B7 소유역의 사항만을 <그림 2>에 나타내었다.

$$\gamma_k = \frac{\sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2} \quad (3)$$

$$\phi_1(k) = \rho_1, \quad \phi_1(2) = \frac{\rho_1(1 - \rho_2)}{(1 - \rho_1^2)}, \quad \phi_2(2) = \frac{\rho_2 - \rho_1^2}{(1 - \rho_1^2)} \quad (4a)$$

$$\phi_k(k) = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_j(k-1)\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_j(k-1)\rho_j} \quad (4b)$$

$$\phi_j(k) = \phi_j(k-1) - \phi_k(k)\phi_{k-j}(k-1) \quad (4c)$$



<그림 2> 소유역 B7과 강우량, 유출량 자료에 대한 상관계수도와 차수의 결정

3.4 매개변수의 산정

식 (5)와 식 (6)~식 (9)에 의하여 매개변수들을 구하고 그 결과 구축된 모형을 식 (10)~식 (13)에 나타내었다.

$$\begin{aligned} c_{q+1} &= \theta_1 c_q + \theta_2 c_{q-1} + \dots + \theta_p c_{q+1-p} \\ c_{q+2} &= \theta_1 c_{q+1} + \theta_2 c_q + \dots + \theta_p c_{q+2-p} \\ &\vdots \\ c_{q+p} &= \theta_1 c_{q+p-1} + \theta_2 c_{q+p-2} + \dots + \theta_p c_q \end{aligned} \quad (5)$$

여기서 c_j 는 자기공분산의 추정치이다.

$$z_t' = z_t - \theta_1 z_{t-1} - \theta_2 z_{t-2} - \dots - \theta_p z_{t-p} \quad (6)$$

$$c_j' = \sum_{i=1}^p \theta_i^2 c_j + \sum_{i=1}^p (\theta_0 \theta_i + \theta_1 \theta_{i+1} + \dots + \theta_{p-i} \theta_p) d_j \quad (7)$$

$$d_j = c_{j+1} + c_{j-1}; \quad j=0, 1, \dots, q; \quad \theta_0 = -1$$

매개변수 θ 와 잔차의 분산 σ_ϵ^2 는 식 (8)과 (9)에 의하여 반복적으로 계산된다.

$$\hat{\sigma}_\epsilon^2 = \frac{c_q'}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} \quad (8)$$

$$\theta_j = - \left(\frac{c_j'}{\hat{\sigma}_\epsilon^2} - \theta_1 \theta_{j+1} - \theta_2 \theta_{j+2} - \dots - \theta_{q-j} \theta_q \right) \quad (9)$$

$$Y_t = 0.4920 Y_{t-1} + 0.1881 Y_{t-2} - 0.0653 Y_{t-3} + \epsilon_t \quad (10)$$

$$X_t = -0.0139 X_{t-1} + 0.4554 X_{t-2} + \epsilon_t + 0.0355 \epsilon_{t-1} \quad (11)$$

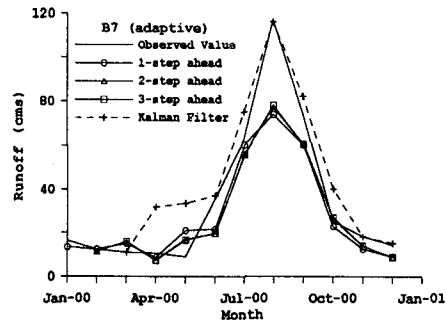
$$Y_t = 0.4888 Y_{t-1} + 0.1458 Y_{t-2} + \epsilon_t \quad (12)$$

$$X_t = 0.0493 X_{t-1} + \epsilon_t + 0.1062 \epsilon_{t-1} \quad (13)$$

4. 모형의 적용

앞에서 구축된 각 모형별로 Kalman Filter 기법을 적용하여 소유역 B7에 대한 결과만을 <그림 3>에 나타내었다.

original time	B7 leading time											
	1	2	3	4	5	6	7	8	9	10	11	12
99/12	13.37	12.17	14.83	8.58	20.79	20.92	60.22	73.29	61.11	23.04	12.61	9.04
00/01	11.64	15.75	7.40	16.79	19.55	56.12	77.43	60.96	26.65	13.80	8.87	
00/02	16.03	7.28	16.58	19.72	55.82	78.13	60.74	26.99	13.84	8.81		
00/03	7.15	16.14	19.47	55.38	78.43	60.80	27.39	13.99	8.82			
00/04	16.98	19.65	55.84	77.96	60.81	26.94	13.85	8.83				
00/05	21.65	60.92	72.84	60.90	22.43	12.34	9.03					
00/06	60.26	73.54	60.91	23.05	12.57	9.01						
00/07	72.93	60.87	22.18	12.24	9.06							
00/08	56.74	41.48	17.47	8.06								
00/09	40.48	17.45	8.06									
00/10	17.11	8.06										
00/11	7.84											



<그림 3> B7 소유역의 월간 유출량 예측 (2000년 1월에서 12월)

위의 예측 결과에 대한 절대오차(abs), 상대오차(rel), 평균제곱오차(MSE), 평균제곱근오차(RMSE)들을 계산하였다.

5. 결론

본 연구에서는 금강유역의 B7, B12 소유역에 대한 월간 강우, 유출량 자료를 이용하여 ARMAX 모형을 구축함으로써 유출량을 예측하였으며 Kalman filtering 기법을 이용하여 예측치들을 시간단계별로 갱신하였으며, 이에 의해 얻은 결론은 다음과 같다.

ARMAX 모형을 이용하여 구한 예측치는 각 소유역의 실측값의 경향을 잘 표현하였으며, 1~3의 선행시간에서 집중호우기를 제외하면 예측이 잘 되는 것으로 나타났다. 여기에서 집중호우기에 대한 예측치는 Kalman filtering 기법에 의하여 개선되었으며, 특히 침투치에 대하여 상당한 예측치의 개선이 이루어졌다.

소유역 B7의 선행시간 2의 평균제곱오차를 제외하고는 선행시간이 증가할수록 일반적으로 예측오차도 증가하는 경향을 보였다. ARMAX 모형을 Kalman filtering 기법을 사용하여 보완한 경우 단순히 ARMAX 모형만을 사용하여 예측하는 것에 비하여 예측치에 대한 오차가 작아져, 대상 유역에 있어서 ARMAX 모형과 Kalman filtering 기법의 적용이 유용한 것으로 나타났다.

감사의 글

본 연구는 21세기 프론티어연구개발사업인 수자원의 지속적 확보기술개발사업단의 연구비지원(과제번호:1-6-1)에 의해 수행되었습니다. 이와 관련하여 관계자 분들에게 감사드립니다.

6. 참고문헌

- Hipel, K.W. and McLeod, A.I. (1994). Time Series Modeling of Water Resources and Environmental Systems, Elsevier.
- Salas, J.D., Delleur, J.W., Yevjevich, V., and Lane, W.L. (1980). Applied Modeling of Hydrologic Time Series, Water Resources Publications, Littleton, Colorado.
- Box, G.E.P. and Cox, D.R. (1964). "An analysis of transformations", Journal of the Royal Statistical Society, Series B, 26, pp. 211-252.
- Durbin, J. (1960). "The fitting of time series model.", Rev. Int. Inst. Stat., 28, p. 233.