추계학적 모형을 이용한 금강수계 유출량 예측

권지혜*, 허준행**

1. 서론

유출량 예측에 있어서 강우-유출모형을 사용하기에는 유량관측망 구축에 대한 비용문제, 매개변수들의 검증문제, 유역의 제반 물리조건들의 변화에 관한 문제 등에 의해 제약이 있음을 고려할 때, 이에 대한 대안으로서 다양한 형태의 추계학적 모형들을 이용해왔다. 그러나 일반적인 추계학적 모형에서는 유출량을 예측하기 위해서는 과거의 유출량 자료만이 사용되므로 기후조건, 토양조건, 강우, 상류로부터의 유입량 또는 상류에 댐이 있을 경우 댐의 방류량에 의한 영향 등을 고려할 수 없다는 한계가 있다. 따라서 본 논문에서는 유출량 외에 앞에서 언급한 다른 조건들을 입력치로 고려할 수 있는 ARMAX(AutoRegressive Moving Average eXogenous) 모형을 이용하여 유출량을 예측함으로써 기존의 일반적인 추계학적 모형에 의한 유출량 예측에서 나타나는 단점을 보완하고자 한다.

2. 기본이론

2.1 ARMAX 모형

본 연구에서 적용된 모형인 ARMAX 모형은 전이함수계(Transfer Function) 모형과 유사한 모형으로서 시계열 자료에 의해 형성된 모형은 동적구성요소(dynamic component)와 잡음항(noise)로 구성되어 있다. 동적구성요소는 다음의 관계로 모형화될 수 있다.

$$Y_t = \nu_0 X_t + \nu_1 X_{t-1} + \nu_2 X_{t-2} + \cdots = \nu(B) X_t \tag{1}$$

여기서, ν_0, ν_1, ν_2, $\cdots$는 충격반응함수(impulse response function or impulse response weights), $\nu(B) = \nu_0 + \nu_1 B + \nu_2 B^2 + \cdots$로서 전이함수, 그리고 B는 후진연산자(backward-shift operator;

---

* 정회원 · 연세대학교 토목공학과 석사과정 · 02-393-1597 (E-mail: wisdomdada@yonsei.ac.kr)
** 정회원 · 연세대학교 공과대학 사회환경 · 건축공학부 토목전공 교수 · 02-2123-2805 (E-mail: jhheo@yonsei.ac.kr)

$B^nP_t = P_{t-n}$)이다. 또한, X_t와 Y_t는 각각 입력치와 출력치이고 각각의 평균이 0이 아닌 경우에는 평균을 제외한 시계열을 의미한다.

또한 식 (2)와 같은 ARMA 모형에서 무한 MA 연산자를 식 (3)과 같이 표현할 수 있는 것과 유사하게 TF에 대해서도 $\nu(B)$에 대해 경제적으로 표현하면 식 (4)와 같이 된다.

$$Z_t - \mu = \phi(B)^{-1}\theta(B)a_t = \Psi(B)a_t \tag{2}$$

$$\Psi(B) = \frac{\theta(B)}{\phi(B)} = 1 + \psi_1 B + \psi_2 B^2 + \cdots \tag{3}$$

$$\nu(B) = \frac{\omega(B)}{\delta(B)} = \frac{\omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_m B^m}{1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_r B^r} \tag{4}$$

여기서, $\Psi(B)$는 무한 MA 연산자, $\theta(B)$는 q차의 MA연산자, 그리고 $\phi(B)$는 p차의 AR 연산자이다. 또 $\omega(B)$는 TF의 분자에서 차수가 m인 연산자, $\delta(B)$는 TF의 분모에서 차수가 r인 연산자이고, ω_0, $\omega_1, \cdots$, ω_m과 δ_1, $\delta_2, \cdots$, δ_r들은 각각 $\omega(B)$와 $\delta(B)$의 매개변수들이다.

그리고 지체시간(time delaying) b을 갖는 전이함수모형은 다음 식 (5)와 같은 형태를 갖게 된다.

$$Y_t = \nu(B)X_t = \frac{\omega(B)}{\delta(B)} B^b X_t \tag{5}$$

여기서, 지체시간 b는 입력치 X가 출력치 Y에 영향을 미칠 때까지 걸리는 시간을 의미한다. 또한 실제적으로 물리적 현상이 반영되는 계에서 식 (5)의 결정론적 전이함수모형만으로는 모형화될 수 없기 때문에 다음 식 (6)처럼 ARMA 모형을 갖는 것으로 가정되는 잡음항을 고려하게 된다.

$$\phi(B)N_t = \theta(B)a_t \text{ or } N_t = \frac{\theta(B)}{\phi(B)} a_t \tag{6}$$

여기서, $\phi(B)$, $\theta(B)$는 각각 차수가 p, q인 AR과 MA의 연산자이고 a_t는 $IID(0, \sigma_a^2)$인 백색잡음계열이다. 이 식 (6)을 식 (5)에 결합하면 동적구성요소와 잡음항이 모두 포함된 다음과 같은 TFN(Transfer Function-Noise) 모형을 구성된다.

$$\begin{aligned} Y_t &= \nu(B)X_t + N_t \\ &= \frac{\omega(B)}{\delta(B)} B^b X_t + \frac{\theta(B)}{\phi(B)} a_t \end{aligned} \tag{7}$$

또한 TFN 모형의 변형이라고 볼 수 있는 ARMAX 모형은 다음 식 (8)로 표현되며 그림 1은 TFN 모형과 ARMAX 모형의 개괄적인 흐름도이다.

$$\begin{aligned} Y_t &= \nu(B)X_t + N_t \\ &= \frac{\omega(B)}{\phi(B)} B^b X_t + \frac{\theta(B)}{\phi(B)} a_t \end{aligned} \tag{8}$$

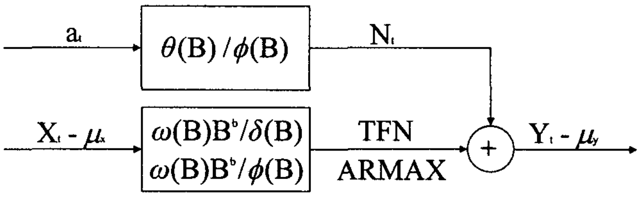

그림 1 TFN과 ARMAX 모형의 흐름도

위의 식 (8)과 그림 1에서 알 수 있듯 동적 구성요소를 나타내는 항의 분모가 TFN 모형의 경우는 새로운 매개변수 $\delta(B)$로 표현되는 반면, ARMAX 모형의 경우는 ARMA 모형의 AR 매개변수인 $\phi(B)$가 사용됨을 알 수 있다.

2.2 모형 적용 절차

단일 입력을 통한 ARMAX 모형의 설계를 위해서는 모형의 형태 결정 (Model Identification) →매개변수 추정 (Parameter Estimation)→검토의 과정을 따르게 된다. 모형의 형태를 결정하기 위해서는 주로 경험적 방법(Empirical Identification Approach)이 사용된다. 이 단계에서는 ARMAX 모형의 수학적 성질과 발생된 시계열의 물리적 성질에 대해 고려함으로써 전이함수 $\nu(B)$를 규명하게 된다. 즉, 대상 시계열의 상관계수를 구하여 지체시간 k에 대해 도시한 sample correlogram나 partial correlogram을 분석하여 입력치와 출력치 각각에 대한 모형의 차수를 먼저 판별하게 된다. 상관계수는 다음 식 (9)에 의해 계산된다(Salas 등, 1980).

$$\gamma_k = \frac{\sum_{t=1}^{N-k} (X_t - \overline{X})(X_{t+k} - \overline{X})}{\sum_{t=1}^{N} (X_t - \overline{X})^2} \tag{9}$$

여기서, X_t는 시간 t에서의 관측값, $\overline{X}$는 관측값의 평균, 그리고 N은 표본자료의 개수이다.

모형의 차수가 결정되면 역시 입력치와 출력치 각각에 대해 매개변수를 구하게 된다. 이 때 각각에 대해 매개변수 $\phi = (\phi_1, \phi_2, \cdots, \phi_r)$, $\theta = (\theta_1, \theta_2, \cdots, \theta_q)$를 먼저 구한 후 입력치와 출력치간의 관계를 파악하기 위해 사전백색화(prewhitening) 모형을 형성하게 된다. 즉, 식 (10)에 의해 시스템의 입력치계열 X_t을 백색잡음계열 α_t로 변환시키고 이 식을 출력치계열인 Y_t에 적용하면 식 (11)과 같이 α_t 계열에 대응하는 β_t 계열을 얻을 수 있다.

$$\alpha_t = \phi_x(B)\,\theta_x^{-1}(B)\,X_t \tag{10}$$

$$\beta_t = \phi_x(B)\,\theta_x^{-1}(B)\,Y_t \tag{11}$$

여기서, α_t, β_t 간의 교차공분산(cross-correlation function)을 구함으로써 입력치가 출력치에 영

향을 미치게 되는 시간 b를 결정하고 남아있는 매개변수인 $\omega = (\omega_0, \omega_1, \omega_2, \cdots, \omega_m)$들도 산정할수 있다.

마지막으로, 결정된 모형에 관한 검토는 각 시간대별로 동적성분을 구한 후 잡음항인 N_t와 잔차인 a_t를 차례로 구한 후, 잔차 a_t가 독립적으로 분포되어 있어야 하고, 경향성을 갖지 않으며 예측값과도 어떤 경향성이 없이 homoscedasticity를 가져야 한다는 기본 가정사항을 확인함으로써 가능하다.

3. 모형의 적용

본 연구에서는 금강유역의 규암지점에 대해 모형을 적용해보았다. 규암지점의 상류에는 대청댐이 있고 공주를 지나 규암으로 이르게 되는데 추계학적 모형을 적용하기 위해 규암지점의 유출량 자료로부터 공주지점의 유출량자료를 제외함으로써 규암유역의 순유출량에 대해서만 고려하였다. 또한 외래유입요소로는 유출량에 대해 가장 직접적인 영향을 미칠 것으로 보이는 규암유역의 강우량만을 고려하였다. 모형구축에는 1993년부터 1997년까지의 5년치 일간 유출량과 강우량 자료를 이용하였으며 이를 이용하여 1998년의 규암지점의 순유출량을 일별로 예측하였다.

앞에서 제시한 절차에 의해 모형을 구축한 결과 입력치인 강우량에 대해서는 ARMA(2,1)을, 출력치인 유출량에 대해서는 AR(2) 모형을 사용하게 되었다. 또한 매개변수 추정과정을 통하여 확정된 본 유역에 대한 모형은 다음 식 (12)와 같다.

$$\begin{aligned} Y_t &= 0.7409\,Y_{t-1} - 0.0013\,Y_{t-2} \\ &\quad + 0.0321\,X_{t-5} + 0.0560\,X_{t-6} + 0.0429\,X_{t-7} + \varepsilon_t \end{aligned} \qquad (12)$$

예측을 수행할 때 모형의 차수는 해당기간에 있어서 본 유역에 가장 적절한 모형으로 보고, 각 시간대별로 새로운 관측값이 들어오면 매개변수를 산정하는 과정부터 다시 수행하여 선행시간은 1일이 되게 하여 다음 시간대에 대해 예측하였다.

4. 모형적용 결과

다음 그림 2에 1998년에 대한 일별 예측유출량을 도시하였고 그림 3a~3c에는 각각 1일~160일, 161일~280일, 281~365일에 대한 예측값과 관측값을 확대하여 도시하였다. 이 그림들을 보면 전기간에 걸쳐 유출량이 증가하고 감소하는 경향과 시간은 잘 따라가고 있고, 특히 건기 동안에는 유출량을 비교적 정확히 예측하고 있음을 알 수 있다. 그러나, 과거자료들의 평균에 가깝게 예측하게 되는 것이 일반적이기 때문에 집중호우기에 유출량이 급격히 변하는 경우에는 그 경향은 어느정도 반영하고 있지만, 값은 제대로 예측하지 못하고 있음을 알 수 있다. 또한 상류부에 있는 공주의 유출량 값이 하류부에 있는 규암의 유출량 값보다 오히려 큰 날은 유출량을 0으로 처리하였는데 예측값은 0까지 떨어지지는 않고 있다.

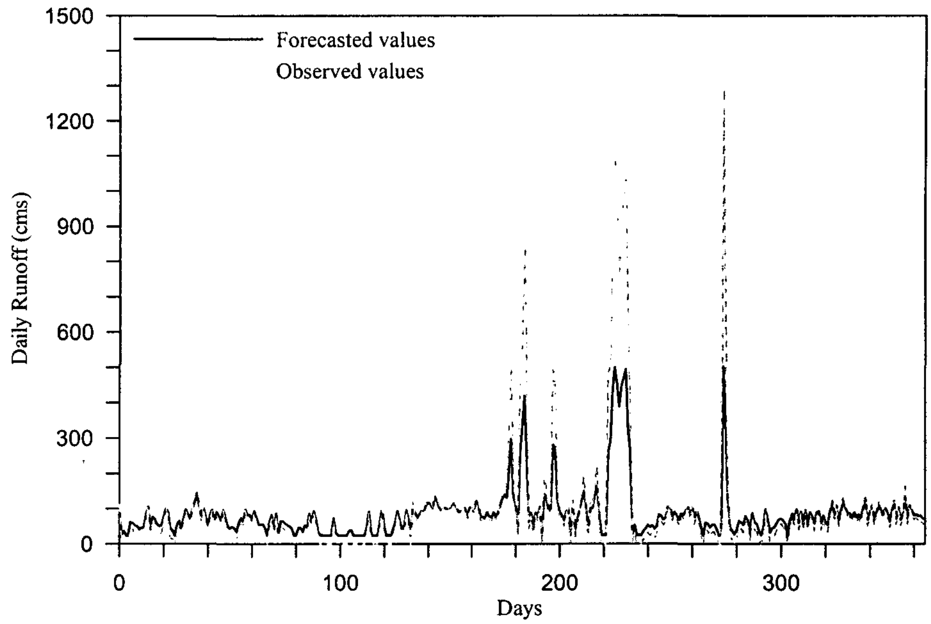

그림 2 일별 유출량의 관측값과 예측값

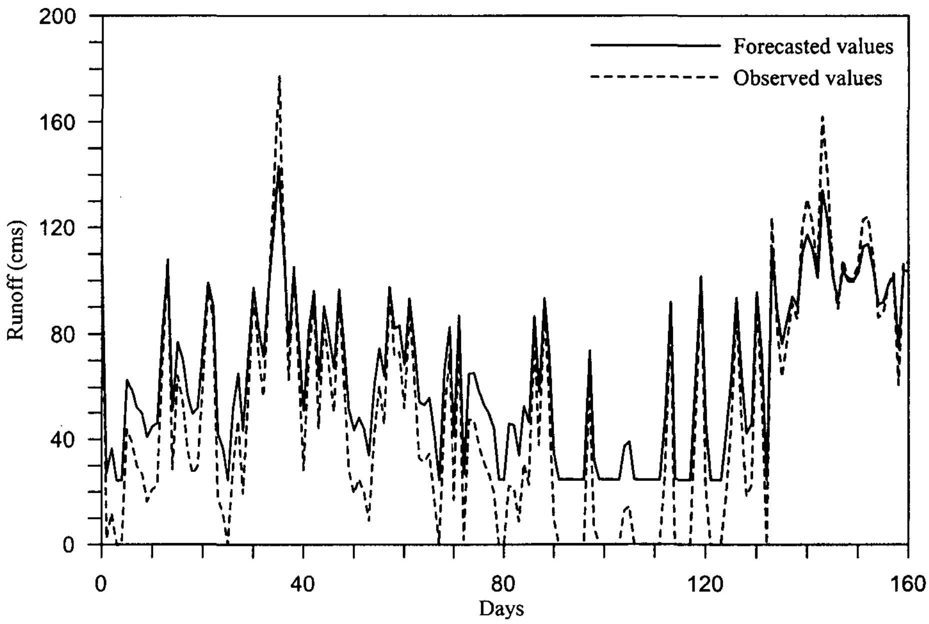

그림 3a 1일~160일에 대한 일별 유출량의 관측값과 예측값

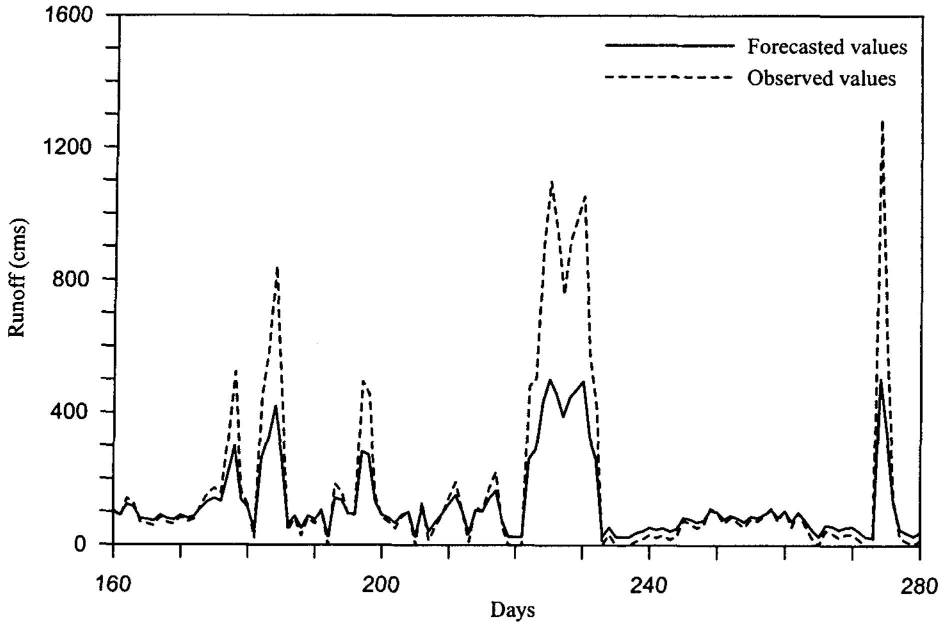

그림 3b 161일~280일에 대한 일별 유출량의 관측값과 예측값

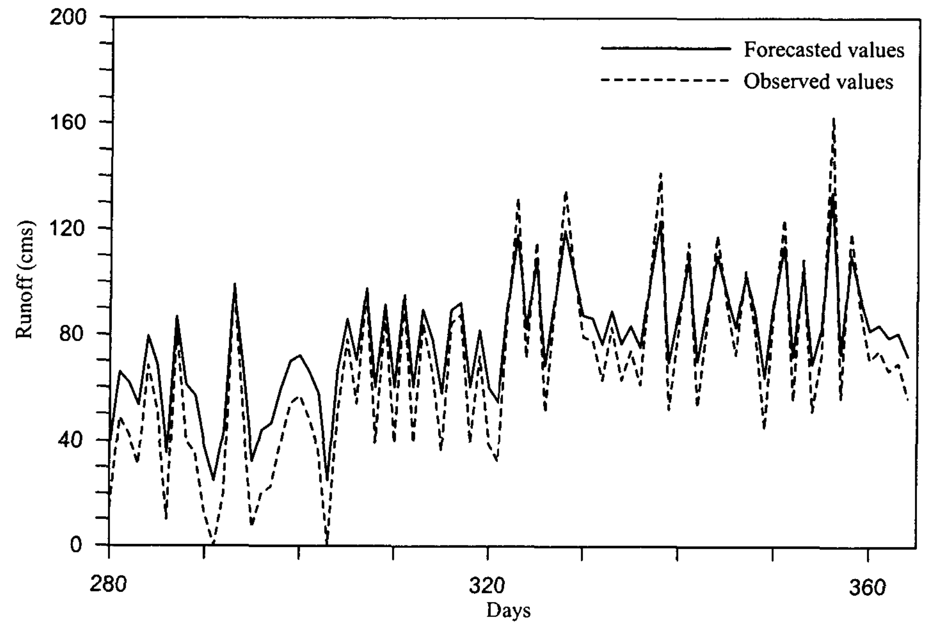

그림 3c 281일~365일에 대한 일별 유출량의 관측값과 예측값

이에 대한 RMSE(Root Mean Squared Error)를 월별과 전체에 대해 구한 결과를 표 1에 나타

내었다.

표 1 1998년에 대한 예측유출량에 대한 RMSE (전체기간 : 93.65)

기간	1월	2월	3월	4월	5월	6월
RMSE	19.72	16.44	20.36	23.09	15.00	47.24
기간	7월	8월	9월	10월	11월	12월
RMSE	113.71	251.23	16.37	150.60	13.26	12.64

5. 결론

본 연구는 기존의 추계학적 모형들이 예측대상자료계열 외의 유역의 다른 조건들은 직접적으로 고려하고 있지 않기 때문에 발생할 수 있는 단점을 보완하고자 강우량 자료를 고려한 ARMAX 모형을 적용하여 유출량을 예측해 보았다. 이에 의해 집중호우기에 상류부에 있는 대청댐의 방류에 의한 급격한 유출량의 변화나 공주의 유출량이 규암의 유출량에 비해 더 커서 이상치를 갖는 것으로 간주되는 경우를 제외하면, 일반적으로 경향이나 시간대에 있어서는 정확한 예측을 하고 있음을 알 수 있다. 특히 건기에 있어서는 매우 정확하게 예측하고 있음을 알 수 있다.

앞에서 언급한 예측상의 문제는 일차적으로는 이상자료의 보완을 통해 그리고, 추후에 적절한 Kalman Filtering 기법을 적용함으로써 개선해보고자 한다.

6. 참고문헌

George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel (1970). Time Series Analysis, Forecasting and Control. Holden-Day Inc., San Francisco, California.

Keith W. Hipel, A. Ian McLeod (1994). Time Series Modelling of Water Resources and Environmental Systems. Elsevier Science B. V., Amsterdam, Netherlands

J. D. Salas, J. W. Delleur, V. Yevjevich, W. L. Lane (1995). Applied Modeling of Hydrologic Time Series. Water Resources Publications, Littleton, Colorado.

J. P. Haltiner and J. D. Salas (1988). "Short-Term Forecasting of Snowmelt Runoff Using ARMAX Models" Water Resources Bulletin, 24(5), pp. 1083-1089.