학술기사

Discrete ARMA 모형과 일 강수사상의 모의

김 형 수 (선문대학교 건설공학부 조교수)
김 태 웅 (고려대학교 방재과학기술연구센터 연구원)

1. 서 론

복잡한 자연현상과 함께 발생하는 수문현상을 수학적으로 서술함으로써 모형화한다는 것은 쉬운 일이 아닐 것이다. 수문과정에서 일어나는 여러 기상학적인 오차, 강수의 시 · 공간적 불확실성, 토양수분과 증발산량 산정등에 대한 어려움은 수문현상의 모형화를 어렵게 하는 주요인이라 할 수 있을 것이다. 특히, 강수는 수문학적 입장에서 보면 강우-유출 해석에 있어 매우 중요한 입력자료이지만 강수가 가지는 특성들, 즉, 간헐성(intermittency), 주기성(periodicity), 그리고 추계학적 특성(stochastic characteristic)등 때문에 모형화에 많은 어려움이 따른다.

강수는 인간이 생존하는데 필수적인 담수의 유일한 원천이지만, 홍수나 가뭄과 같은 재해를 가져오는 원인을 제공하기도 한다. 즉, 같은 양의 강수일지라도 짧은 기간에 발생하면 홍수가 될 것이고, 긴 기간에 발생하면 가뭄이 될 것이다. 이처럼 강수의 양과 발생기간은 그 분석에 있어 중요한 인자라 할 수 있을 것이며, 통상 하천유량을 통해 정의되는 홍수나 가뭄을 해석하는데 있어서도 강수사상에 대한 정보는 매우 중요하다 할 수 있을 것이다. 그러나 강수의 양과 발생기간 그리고 홍수나 가뭄과 같은 수문사상들은 매우 불확실하여 확률론적으로 해석되어져야하며, 이를 위하여 대상유역에 대한 장기간의 수문학적 자료가 요구 된다. 수문학적 자료들은 대부분 이산적(discrete)으로 관측되어 시계열을 형성하거나 이산적 시간구간에 대해 평균값을 취함으로써 연속적(continuous)으로 기록되기도 한다.

그동안 수문학자들은 확률법칙을 근거로 수문시계열을 수학적으로 모형화한 추계학적 모형들을 연구하여 왔다. 그러나 추계학적 모형을 일단위 수문시계열에 적용하는데는 여러 가지 단점들이 있었다. 첫째로 정규분포특성은 ARMA와 같은 긴 시간단위을 갖는 수문 시계열에는 적절하였으나, 비교적 짧은 시간단위인 일단위 수문과정에는 유효하지 않았다. 두 번째로 강수와 같이 0의 값을 가지고 지속되는 수문 과정은 ARMA 모형에 적합하지가 않다. 세 번째는 단일모형에 의해서 홍수나 가뭄의 특성을 완전히 해석하기는 힘들다. 마지막으로 공학적으로 적용하는데 간단하지가 않다는 것이다.

이러한 단점을 해결하기 위하여 일단위와 같이 짧은 지속기간을 갖는 강수계열의 모의를 위하여 Discrete Autoregressive Moving Average (DARMA) 모형이 개발되었으며, 또한 일단위 강수모형으로부터 일단위 하천유량을 모의할 수 있도록 하였다.

2. DARMA 모형의 구조

DARMA(1, q), DARMA(p, 0), 그리고 DARMA(0, q)등은 1978년 Jacobs와 Lewis에 의해 수학적으로 개발되었다. DARMA(p, q) 모형은 $\{Y_n\}$의 확률론적 결합에 의해 형성된$\{X_n\}$의 과정이다. 여기에서, $\{Y_n\}$은 iid(identically independent distributed)이고, p는 자기회귀과정의 차수이고 q는

이동평균과정의 차수이다. $\{Y_n\}$가 확률분포 π를 갖는다면 π는 다음과 같이 정의될수 있다.

$$\pi_i = P(Y_n = i) \; ; i = 0,1,2,... \tag{1}$$

그리고 $\{U_n\}$과 $\{V_n\}$을 {0, 1}사이의 값을 가지는 독립적인 무작위 변량의 계열이라면,

$$P(U_n = 1) = \beta = 1\text{-}P(U_n = 0),\ 0 \le \beta \le 1 \tag{2}$$
$$P(V_n = 1) = \rho = 1\text{-}P(V_n = 0),\ 0 \le \rho \le 1 \tag{3}$$

을 정의할 수 있다.

*m*1은 {0, 1, ⋯, q-1}의 iid 무작위 변량 계열로 β_0, β_1, ⋯, β_{q-1}의 확률을 가지고, *m*2는 {1, 2, ⋯, p}의 iid 무작위 변량 계열로 α_1, α_2, ⋯, α_p의 확률을 갖는다고 하면, $\beta_0 + \beta_1 + \cdots + \beta_{q-1} = 1$이고, $\alpha_1 + \alpha_2 + \cdots + \alpha_p = 1$ 이다. 그러면 DARMA(p, q)는 다음과 같이 정의된다.

$$X_n = U_n Y_{n-m1} + (1\text{-}U_n)A_{n-q} \tag{4}$$

여기서, Y_{n-m1}은 이동평균항이고 A_{n-q}는 자기회귀 꼬리(autoregressive tail)로 다음과 같이 정의된다.

$$A_n = V_n A_{n-m2} + (1\text{-}V_n)Y_n \tag{5}$$

A_n은 $\{Y_n\}$과 같은 확률분포를 가지며 A_{1q}, A_q, ⋯, A_{1qp}중에서 한 무작위 변량을 가지고 시작을 하지만 $\{Y_n\}$과는 상호 독립적이다. 다음에 $\{X_n\}$는 주변 분포(marginal distribution)가 $\{Y_n\}$과 같고, 종속적인 이산 무작위 변량의 정상 계열처럼 거동한다. 즉, $\{X_n\}$과정은 주변 분포와 상관구조를 분리하여 생각할 수 있다는 장점이 있다.

3. DARMA 모형에 의한 일 강수량의 모델링

일 강수량을 모의하기 위한 DARMA모형에는 두 가지 형태가 있는데 그 첫 번째는 Binary-Discrete Autoregressive Moving Average Mixed with Exponential Model(B-DARMA-E(p, q))이고 두 번째는 Multistate DARMA Model(M-DARMA(p, q:r))이다. 각각의 모형을 살펴보면 다음과 같다.

3.1 B-DARMA-E(p, q)

B-DARMA-E(p, q) 모형에서 p는 자기회귀성분의 차수이고, q는 이동평균성분의 차수로써 식(6)과 같이 모형을 나타낼수 있다.

$$W_i = X_i Z_i \tag{6}$$

여기서, i는 시간상태(time state)이고, X_i는 다음과 같이 정의되는 일 강수 시계열의 이산상태(discrete state)계열이다.

$$X_i = \begin{cases} 1, \text{ if } & R_i \ge 0.01 \\ 0, \text{ if } & R_i < 0.01 \end{cases} \tag{7}$$

R_i는 관측된 강수량으로서 0.01 inch 이하는 무강우를 의미한다. 단지 0과 1만을 포함하는 $\{X_i\}$는 이진 이산 시계열(binary discrete time series)로서, 1은 습윤일을 나타내고, 0은 건조일을 나타낸다. $\{Z_i\}$는 습윤일일 때, 즉 $\{X_i\}$가 1의 값을 가질 때의 강수량 계열로 지수분포를 갖고, $\{W_i\}$는 일단위 강수계열로 $\{X_i\}$와 $\{Z_i\}$의 결합으로 구성된다. 물론 이 모형에서 $\{X_i\}$와 $\{Z_i\}$는 서로 독립적이라고 가정한다.

B-DARMA-E 모형은 습윤일과 건조일 계열을 위한 B-DARMA와 0이 아닌 강수계열을 위한 독립적 지수분포를 갖는 하위 모형으로 구성되어 있고, 미국 인디애나 주의 일 강수 시계열에 적용된 바 있다. B-DARMA-E 모형을 구성하기 위해서는 모형인식과정(identification), 매개변수추정(estimation), 그리고 검정(diagnostic checking) 등 3가지 절차가 있는데, 일강수 시계열에 적용하기에 매우 효과적이다.

3.2 M-DARMA(p, q: r)

B-DARMA-E 모형은 습윤일과 건조일에 의해서

강수량을 구별할 수가 없기 때문에 지수분포를 이용하여 0이 아닌 강수계열을 모의하여야만 한다. 만약 강수계열을 이산 강수 상태(discrete precipitation state)로 구성된 계열로 변환하여 일 강수량이 여러상태로 구별될 수만 있다면 B-DARMA-E 모형에서의 지수분포 가정을 피할 수가 있다. 이를 위하여 M-DARMA(p, q: r) 모형이 개발되었으며, p는 자기회귀성분의 차수, q는 이동평균성분의 차수, 그리고 r은 상태의 수를 나타내는데 강수의 양에 따라 분류된다.

표본 자기상관함수의 비교를 통하여 다음과 같은 3가지 상태의 일 강수계열이 적절하다는 것이 밝혀졌으며 M-DARMA 과정의 모의에 적용되었다.

$$X_i = \begin{cases} 0, & R_i < \hat{R}_j \\ 1, & \hat{R}_j \le R_i < \hat{R}_j + \hat{\sigma}_j \\ 2, & R_i > \hat{R}_j + \hat{\sigma}_j \end{cases} \tag{8}$$

R_i는 i번째 날에 관측된 일 강수량, $\hat{R}_j$는 j번째 계절의 일 강수량의 평균이고 $\hat{\sigma}_j$는 j번째 계절의 일강수량의 표준편차이다.

M-DARMA 모형의 검정을 위하여 다상태 런 길이(multistate run length)에 대한 개념을 도입하였다. 이는 일 강수 시계열의 지속성을 보존하는데 중요한 개념이다. 즉, 여러상태의 런 길이에 대한 확률분포를 얼마나 잘 재현하는지를 판단하여 최적의 모형을 선택하는 것이다.

4. B-DARMA와 M-DARMA 모형의 런 길이

4.1 이진 런 길이(Binary run length)

B-DARMA 모형에서 $\{Y_n\}$는 확률 π_0와 π_1을 가지고 $\{0, 1\}$을 취하며 식(9)의 관계를 갖는다.

$$\pi_0 + \pi_1 = 1 \tag{9}$$

그러므로, B-DARMA에서는 T_0로 표현되는 0의 런 길이와 T_1로 표현되는 1의 런 길이를 포함하며 식(10)과 (11)처럼 나타낼수 있다.

$$\{T_0 = n\} = \{X_0 = 1, X_1 = 0, \cdots, X_n = 0, X_{n+1} = 1\} \tag{10}$$

$$\{T_1 = n\} = \{X_0 = 0, X_1 = 1, \cdots, X_n = 1, X_{n+1} = 0\} \tag{11}$$

따라서, 식(10)과 (11)에 대한 확률은 다음과 같이 정의할 수 있다.

$$\begin{aligned} P(T_0 = n) &= P\{X_0 = 1, X_1 = 0, \cdots, X_n = 0, \\ &\quad X_{n+1} = 1 \mid X_0 = 1, X_1 = 0\} = P\{X_0 = 1, X_1 \\ &\quad = 0, \cdots, X_n = 0, X_{n+1} = 1\} \,/\, P\{X_0 = 1, X_1 \\ &\quad = 0\} \end{aligned} \tag{12}$$

$$\begin{aligned} P(T_1 = n) &= P\{X_0 = 0, X_1 = 1, \cdots, X_n = 1, \\ &\quad X_{n+1} = 0\} / P\{X_0 = 0, X_1 = 1\} \end{aligned} \tag{13}$$

4. 2 다상태 런 길이(multistate run length)

$\{Y_n\}$이 확률 $\pi_0, \pi_1, \pi_2, \cdots, \pi_{m-1}$을 가지며, $\{0, 1, 2, \ldots, m-1\}$을 취한다고 하면,

$$\pi_0 + \pi_1 + \pi_2 + \cdots + \pi_{m-1} = 1 \tag{14}$$

따라서, $\{X_n\}$은 강수량에 따라 m개의 이산 상태(discrete state)를 가지게 되고 M-DARMA모형에 의해 모의할 수가 있다. M-DARMA 모형에서는 i의 다상태 런 길이를 T_i 라 하고, 이는 식(15)와 같이 정의한다.

$$\begin{aligned} \{T_i = n\} &= \{X_0 \ne i, X_1 = i, \cdots, X_n = i, \\ &\quad X_{n+1} \ne i \mid X_0 \ne i, X_1 = i\} \end{aligned} \tag{15}$$

그러면 T_i의 확률은 다음과 같다.

$$\begin{aligned} P(T_i = n) &= \mathrm{P}\{X_0 \ne i, X_1 = i, \cdots, X_n = i, \\ &\quad X_{n+1} \ne i \mid X_0 \ne i, X_1 = i\} \\ &= \mathrm{P}\{X_0 \ne i, X_1 = i, \cdots, X_n = i, \\ &\quad X_{n+1} \ne i\} \,/\, P\{X_0 \ne i, X_1 = i\} \end{aligned} \tag{16}$$

5. 전이모형(Transfer Model)

입력으로서 강수모형이 구성되면 Transfer DARMA Model(T-DARMA(p,q,m,n,1))에 의해서 일단위 하천유량을 모의할 수 있다. T-DARMA(p, q,m,n,1) 모형에서 p는 자기회귀성분의 차수, q는 이동평균성분의 차수, m은 전이과정에서의 자기회귀성분의 차수, 그리고 n은 전이과정에서의 이동평균성분의 차수를 나타낸다. 그리고 모형은 식(17)과 같이 표현된다.

$$\Phi(B)\ Q(k) = \theta(\mathrm{B})R(k) - L \tag{17}$$

여기서, $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \cdots - \phi_m B^m$,

$\Phi(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \cdots - \theta_m B^m$,

$Q(k)$는 시간 k에서의 유출, $R(k)$는 시간 k에서의 강수입력이고, L_j는 j계절에서의 손실량이다. 강수모형, B-DARMA-E나 M-DARMA는 T-DARMA 모형에서의 입력 $R(k)$로 사용된다.

6. 요약 및 결론

지금까지 본 논고에서 일 강수량 모의와 일 강우-유출 전이를 위하여 이용할 수 있는 DARMA 모형에 대하여 살펴보았으며, 여기에서 DARMA모형의 응용에 있어 그 효율성을 요약해보면 다음과 같다.

(1) B-DARMA-E는 습윤일과 건조일 과정을 모의하기 위한 B-DARMA와 0이 아닌 강수과정을 모의하기 위한 독립적인 지수분포로 구성되어 있어 기존에 ARIMA모형을 적용하였을 때 가지고 있던 어려움들을 해결하였다.

(2) DARMA 모형의 간단한 매개변수추정 방법은 여러지점의 자료에 모형을 적합시키는 데 효과적이다. 또한 모형을 구성하는데 있어 간단한 절차는 여러 모형을 구성하여 최적의 모형을 보다 쉽게 선정할 수 있다.

(3) B-DARMA-E 모형에서 습윤일과 건조일의 지속성과 M-DARMA 모형에서 다중 런(multiple run)을 보존할 수가 있어 이론적인 런 길이의 분포를 여러 모형들이 쉽게 간직할 수가 있다.

(4) 홍수와 가뭄의 특성들이 이론적 런 길이분포와 추정된 런 길이 분포와의 비교를 통하여 개념적으로 해석될 수 있다.

(5) B-DARMA-E 모형과 M-DARMA 모형에 의해서 모의된 일 강수 계열은 일 하천유량을 위한 T-DARMA 모형의 입력자료로 이용될 수가 있다.

〈참고문헌〉

이재준(1996), 강수의 발생이론과 강수사상의 수문학적 모델링, 한국수자원학회지 제29권 제 1호, 한국수자원학회, pp. 114-116

Chang, T.J., Delleur, J.W. and Kavvas M.L.(1987), Application of Discrete Autoregressive Moving Average Models for Estimation of Daily Runoff, Journal of Hydrology, 91(1987), pp. 119-135.

Chang, T.J., Kavvas M.L., and Delleur, J.W.(1984), Daily Precipitation Modeling by Discrete Autoregressive Moving Average Processes, Water Resources Research, Vol. 20, No. 5, pp. 565-580.

Chang, T.J., Kavvas M.L., and Delleur, J.W.(1982), Stochastic Daily Precipitation Modeling and Daily Streamflow Transfer Processes, Purdue University, Technical Report No. 146.

Jacobs, P.A., and Lewis P.A.W.(1978a), Discrete Time Series Generated by Mixtures, 1, Correlation and Run Properties, J.R. Stat. Soc., Ser. B, 40(1), pp. 94-105.

Jacobs, P.A., and Lewis P.A.W.(1978b), Discrete Time Series Generated by Mixtures, 2, Asymptotic Properties, J.R. Stat. Soc., Ser. B, 40(2), pp. 222-228.

Jacobs, P.A., and Lewis P.A.W.(1983), Stationary Discrete Autoregressive Moving Average Time Series Generated by Mixtures, J. Time Ser. Anal., 4(1) pp. 19-36.