

통계-물리학적 모형을 이용한 강우예측모형의 개발

Development of Precipitation Forecasting Model by Using Stochastic and Physical Model

정창삼*, 허준행**

Jeong, Chang-Sam, Heo, Jun-Haeng

1. 서론

일반적으로 강우예측모형은 통계적 모형(statistical model)과 물리·동력학적 모형(physical dynamical model)으로 구분될 수 있다. 물리적 모형은 강우 발생의 원리를 질량 보존식, 운동량 방정식 및 에너지 방정식으로 표현하여 계산하는 모형으로 각종 지상기상 관측, 해양기상 관측, 고공기상 관측, 기상레이더 관측 및 기상위성 자료를 활용하고 있으나 강우발생의 지역적 다양성, 강우형성에 관한 정확한 구름과정 및 지표과정에 대한 지식의 한계, 계산의 어려움, 계산 시간의 과다 등으로 인해 향후 보다 많은 연구가 필요한 실정이다. 이에 비해 통계적 모형은 지표에 도달한 강우와 주변의 물리적 특성 등을 규명하여 미래에 발생할 강우를 예측하는 모형으로 이는 물리적 모형의 어려운 점을 극복하기 위한 방안의 일환으로 시도되었으며 주로 과거의 지역적 관측자료를 근거로 모형이 개발되었다. 이러한 모형은 대개 강우의 시간적 특성이 경향성을 띠는 지역에 일반적으로 사용할 수 있으며, 과거로부터 널리 적용되고 있다. Grace와 Eagleson(1966)은 alternating renewal 모형을 이용하여 확률적인 강우예측모형을 제안하였고, Duckstein 등(1972)은 Poisson 모형을, Smith와 Schreber(1973)은 Markov chains을 이용하였으며, Chang 등(1984)은 discrete autoregressive moving average 기법을, Kavvas와 Delleur(1981)은 point process 모형을, Saleh(1990)은 Hybrid모형을 이용하여 Dacca시에 적용한 바 있다.

본 연구에서는 통계적 모형의 하나인 Hybrid 모형을 이용하여 강우예측모형을 구성하여 보았다. Hybrid라는 원래의 의미처럼 본 모형은 두 가지 요소에 의해 강우를 예측하게 된다. 하나는 seasonal한 요소이고, 다른 하나는 perturbation 요소이며, 아래의 식(1)과 같이 표현할 수 있다.

$$(i_d - \bar{i}_d \rightarrow O_d - \bar{O}_d) \quad (1)$$

여기서 \bar{i}_d 는 입력값으로 d일에 대한 seasonal 평균치를 의미하며, \bar{O}_d 는 출력값으로 d일에 대한 seasonal 평균치를 의미한다.

위의 개념을 바탕으로 seasonal 값을 구하기 위해 식(2)와 같이 Fourier series 모형을 구성하였다.

* 연세대학교 사회환경시스템공학부 토목공학과 박사과정

** 연세대학교 사회환경시스템공학부 토목공학과 부교수

$$x_k = \bar{x} + \sum_{m=1}^A [a_m \cos(2\pi mk/365) + b_m \sin(2\pi mk/365)] \quad (2)$$

여기서

$$\bar{x} = (1/365) \sum_{k=1}^{365} x_k \quad (3)$$

$$a_m = (2/365) \sum_{k=1}^{365} x_k \cos(2\pi mk/365) \quad (4)$$

$$b_m = (2/365) \sum_{k=1}^{365} x_k \sin(2\pi mk/365) \quad (5)$$

위의 식을 이용하여 모형의 입력자료를 선정하여 평활화된 seasonal 값을 구하게 된다.

Geogakakos와 Bras(1984)는 강우와 기온, 습도, 기압, 등과 같은 기상 관측자료의 상호 연관성을 해석한바 있다. 본 연구에서는 강우 예측값을 얻기 위해 기압, 습도, 증발량, 온도, 습도, 동을 입력자료로 사용하였으며, 이를 식으로 나타내면 식(6)과 같다.

$$(R_{t+1} - \bar{R}_{t+1}) = c_1(P_t - \bar{P}_t) + c_2(H_t - \bar{H}_t) + c_3(U_t - \bar{U}_t) + c_4(D_t - \bar{D}_t) + c_5(E_t - \bar{E}_t) + e_d$$

여기서 R은 강우, p는 기압, h는 습도, u는 일 최고기온, d는 일 최저기온, E는 증발량, e_d는 오차를 의미한다. 이를 다른 형태로 나타내 보면 다음과 같다.

$$\Delta R_{t+1} = c[\Delta P_t, \Delta H_t, \Delta U_t, \Delta D_t, \Delta E_t]' + e_d \quad (7)$$

일반적으로 기압과 습도, 기온 등은 교차상관관계(cross-correlation)를 가지게 된다. 만약 이러한 교차상관관계가 존재할 경우, 아래의 식(8)을 이용하여 회귀분석시 발생하는 multicollinearity문제를 해결할 수 있다.

$$W = ZA \quad (8)$$

$$Z = [z_{tj}] = [(x_{tj} - \bar{x}_j) / S_j] \quad t = 1, 2, \dots, n, \quad j = 1, 2, \dots, p \quad (9)$$

여기서 W는 변형된 값이며, Z는 입력자료이며, A는 입력자료의 cross-correlation행렬로부터 구해지는 고유행렬을 의미한다. \bar{x}_j 는 j번째 변수의 평균값을 의미하고, S_j는 j번째 변수의 표준편차를 의미하며, x_{tj}는 $\Delta P_t, \Delta H_t, \Delta U_t, \Delta D_t, \Delta E_t$ 를 의미한다.

회귀분석시 각 독립변수들은 noncommensurate문제를 제거하기 위해 식(8)의 방법과 같이 일반적으로 표준화(standardize)되어진다. 이러한 과정을 통해 회귀식은 다음과 같은 형태로 나타낼 수 있다.

$$R = WB + E \quad (10)$$

여기서 R은 계산결과 나오는 예측치를 의미하고 B는 회귀행렬을 의미한다.

다음으로는 이렇게 구성된 모형이 적합한가에 대한 적합도 분석을 실시하여야 한다. 본 연구에서는 일반적으로 적합도 분석시 널리 사용되는 T-test와 F-test를 실시하였다. 우선 식(10)의 B에서 최소자승법을 이용한 임의의 sample estimate b를 구하여 보면

$$b = [W'W]^{-1} W'R \quad (11)$$

이 되며,

$$\text{Cov}(b) = \sigma^2 (W'W)^{-1} \quad (12)$$

여기서 σ 는 회귀식의 오차다.

만약 모형이 바르게 구성되었다면, b_j/S_{b_j} 는 $n-p$ 의 자유도를 가지는 t -분포형이 되어야 한다. 여기서 S_{b_j} 는 $\text{Var}(b_j)$ 에 대한 추정값이다. 가설검증을 실시하여보면 아래 식(13)의 검정 통계량 t 가 $|t| > t_{1-\alpha/2, n-p}$ 인 경우 가정 H_0 는 기각되는데, 이는 j 번째 독립변수가 전체에 대해 상당히 큰 변동성을 가지고 있다는 것이다. 반대인 경우는 일반적으로 그 변수는 모형에서 제거하여도 무방하다고 할 수 있다.

$$t = b_j / S_{b_j} \quad (13)$$

다시말해 모든 회귀식이 각 종속변수의 충분히 변동성을 나타내고 있지 못하다는 가설을 만족하기 위해서는 $H_0 : B_1 = B_2 = B_p = \dots = 0$ 되어야 하고, 반대인 경우에는 최소 1개 이상의 B_s 가 0이 아닌 값을 가지게 된다.

2. 본론

본 연구에서는 Hybrid 모형을 적용하여 소양강댐 유역의 기상자료에 적용하여 1일 강우예측모형을 구성하였다. 일반적으로 본 모형과 같은 회귀모형의 경우 강우형태의 경향성이 강한 지역에 적합하다고 알려져 있다(Saleh, 1990).

기본적인 자료로는 취득이 용이한 강우량, 습도, 기압, 증발량, 최고기온, 최저기온의 6가지 자료를 사용하였고, 적용 기간은 1974~1996년의 자료 중 부정확한 기록을 가진 1980, 1982, 1983년의 세 연도에 대한 기록을 제외한 20년간의 일간 자료를 이용하여 seasonal한 특징을 구하였으며, 1997년과 1998년의 자료를 이용하여 회귀분석을 통해 오차(deviation)를 구하였다.

우선 20년간의 자료를 바탕으로 Fourier series를 이용하여 seasonal 값을 구하고, 1997년과 1998년의 자료를 이용하여 오차(deviation)를 구하였다.

식(2)를 이용하여 seasonal 특성값을 구하기 위해서는 harmonic을 결정하여야 한다. β 로 표시되는 이 값은 seasonal 특성치를 나타내는 중요한 값으로, 일반적으로 그림 1이나 2에서와 같이 $p(i)$ 를 구하여 critical harmonic을 구하거나, variance analysis를 이용하여 harmonic을 결정한다. 본 연구에서는 $\beta = 3$ 으로 harmonic을 선정하였으며, 각각의 요소에 대해 harmonic과 이에 따른 계산결과를 표 1~6에 나타내었다. 또한 이를 통해 정해진 모형에서의 각 요소별 상관계수를 나타내면 표 7과 같다. 표 7의 값에 대해 식(8)을 이용하여 eigen vector를 계산하면 아래와 같으며, 각각의 eigen value는 2.363, 1.231, 0.360, 0.045이 된다.

$$\begin{bmatrix} 0.0606 & 0.8896 & 0.1568 & -0.4249 \\ 0.5410 & -0.2965 & 0.7386 & -0.2710 \\ 0.5859 & -0.1431 & -0.6542 & -0.4568 \\ 0.6003 & 0.3166 & -0.0429 & 0.7331 \end{bmatrix}$$

이를 바탕으로 회귀방정식을 구성하면 아래 (13)식과 같다.

$$R_t = \bar{R}_t - 2.459 - 0.238AG + 0.056AS - 3.526AE - 0.301AU + 0.201AD \quad (13)$$

이를 바탕으로 '97년과 '98년의 예측치와 관측치를 나타내면, 그림 8, 그림 9와 같다.

표 1 강우

Number of Harmonics	Fourier Coefficients		Variance Accounted for by m th Harmonics
	a_m	b_m	
1	-3.44	-1.29	6.79
2	0.94	1.56	1.67
3	0.16	-1.21	0.75
4	-0.44	0.29	0.14
5	-0.35	0.09	0.06

표 2 기압

Number of Harmonics	Fourier Coefficients		Variance Accounted for by m th Harmonics
	a_m	b_m	
1	16.87	1.81	143.88
2	-1.47	-1.64	2.43
3	-0.19	3.84	7.38
4	0.16	1.03	.54
5	1.30	-0.10	.85

표 3 습도

Number of Harmonics	Fourier Coefficients		Variance Accounted for by m th Harmonics
	a_m	b_m	
1	-1.97	-6.82	25.16
2	2.30	2.34	5.39
3	-0.44	-0.02	0.10
4	0.14	-0.44	0.11
5	0.27	-0.54	0.18

표 4 증발량

Number of Harmonics	Fourier Coefficients		Variance Accounted for by m th Harmonics
	a_m	b_m	
1	-1.90	0.25	1.83
2	-0.09	-0.20	0.02
3	0.20	0.08	0.02
4	-0.08	0.01	0
5	0.06	0.05	0

표 5 최고기온

Number of Harmonics	Fourier Coefficients		Variance Accounted for by m th Harmonics
	a_m	b_m	
1	-13.44	-4.23	99.29
2	-1.86	-0.31	1.78
3	0.14	-0.18	0.03
4	-0.19	0.32	0.07
5	0.37	0.09	0.07

표 6 최저기온

Number of Harmonics	Fourier Coefficients		Variance Accounted for by m th Harmonics
	a_m	b_m	
1	-13.42	-5.33	104.24
2	-0.59	0.59	0.34
3	-0.04	-1.00	0.50
4	-0.24	-0.19	0.05
5	0.27	-0.13	0.04

표 7 상관계수표

	습도	기압	최고기온	최저기온	증발량
습도	1	-0.2001	-0.1007	0.4168	-0.7637
기압	-0.2001	1	0.6326	0.6314	0.3460
최고기온	-0.1007	0.6326	1	0.7707	0.3744
최저기온	0.4168	0.6314	0.7707	1	-0.1166
증발량	-0.7637	0.3460	0.3744	-0.1166	1

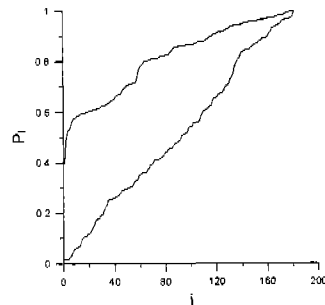


그림 1 Periodogram(강우)

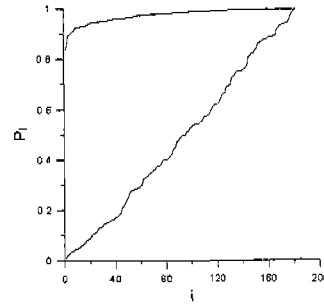


그림 2 Periodogram(기압)

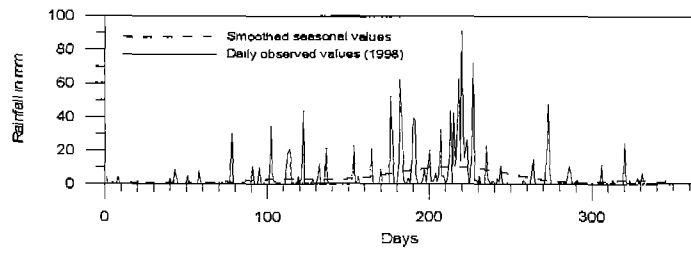


그림 3 소양강댐 유역의 강우 관측자료 (1998, 일단위) & seasonal values

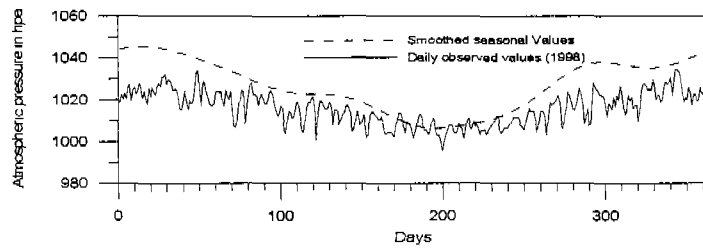


그림 4 소양강댐 유역의 기압 관측자료 (1998, 일단위) & seasonal values

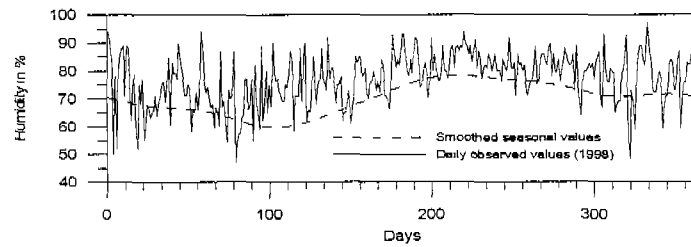


그림 5 소양강댐 유역의 습도 관측자료 (1998, 일단위) & seasonal values

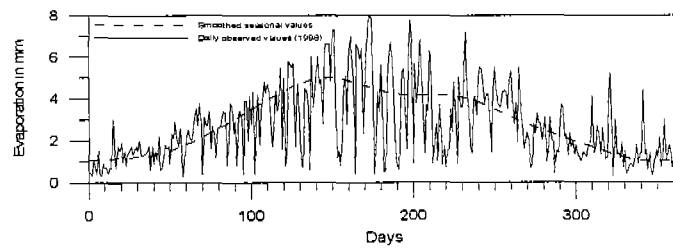


그림 5 소양강댐 유역의 증발량 관측자료 (1998, 일단위) & seasonal values

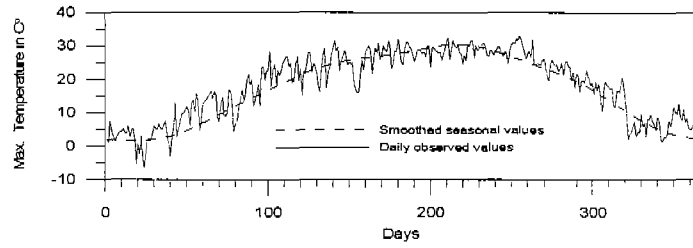


그림 6 소양강댐 유역의 최고기온 관측자료 (1998, 일단위) & seasonal values

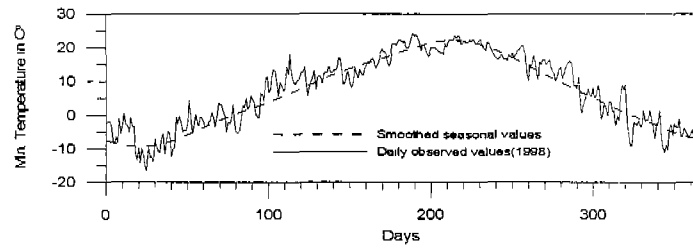


그림 7 소양강댐 유역의 최저기온 관측자료 (1998, 일단위) & seasonal values

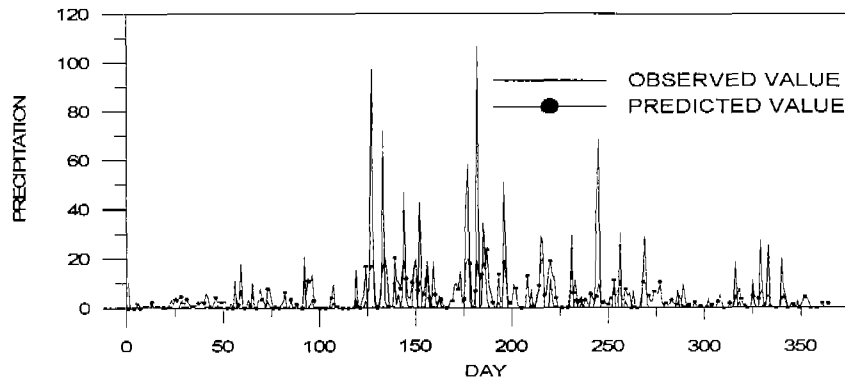


그림 8 소양강댐 유역의 1997년 관측강우량과 예측강우량

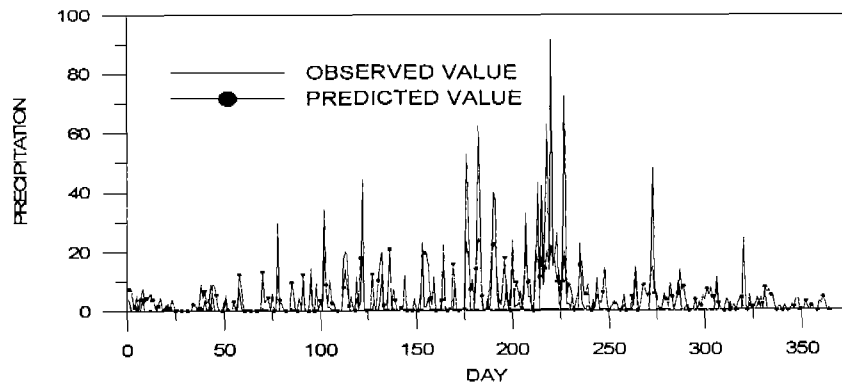


그림 9 소양강댐 유역의 1998년 관측강우량과 예측강우량

3. 결론

수공학에 있어 일일 강우예측은 중요한 요소이며, 그 활용분야도 광범위하다고 할 수 있다. 현재 일일 강우예측의 경우 대부분 위성사진의 관측과 기상레이더 등을 이용하여 실시하고 있으나, 미개발지역이나 지역적인 연구가 미비한 지역에 대해서는 아직까지 적용이 어려운 상황이다. 본 연구에서 적용한 모형의 경우 기상관측자료와 계절별 주기성이 강한 지역에 대해 쉽게 적용할 수 있으며, 비교적 용이한 결과를 얻을 수 있었다. 하지만 첨두값이 큰 강우사상에 대해서는 어느 정도 한계를 가지고 있음을 결과를 통해 알 수 있었다.

현재와 같이 이상기후가 빈번한 경우 기상레이더와 같은 물리적인 강우 예측이 우선화되어야 하겠지만, 이러한 예보시스템이 미비한 지역에 대해서는 본 모형이 어느 정도의 대안이 되리라 판단된다.

4. 참고문헌

Chang, T. J., Kavvas, M. L., and Delleur, J. W., Daily precipitation modeling by discrete autoregressive moving average process, *Water Resour. Res.*, 20(5), 565~580, 1984.

Draper, N. R., and Smith, H., *Applied Regression Analysis*, 2nd ed., pp.70~71, John Wiley, New York, 1981.

Duckstein, L., Fogel, M., and Kisiel, C. C., A stochastic model of runoff producing rainfall for summer type storms, *Water Resour. Res.*, 8(2), 410~421, 1972.

Georgakakos, K. P., and Bras, R. L., A hydrologically useful station precipitation model, *Water Resour. Res.*, 20(11), 1585~1610, 1984.

Grace, R. A., and Easleson, P. S., The synthesis of short time increment rainfall sequences, *Hydrodyn. Lab. Rep. 91*, Dep. of Civ. Eng., Mass Inst. of Technol., Cambridge, 1966.

Kavvas, M. L., and Deller, J. W., A stochastic cluster model of daily rainfall sequences, *Water Resour. Res.*, 17(4), 1151~1160, 1981.

Saleh, A. W., A hybrid model for forecasting daily rainfall, *Water Resour. Res.*, 26(11), 2741~2746, 1990.

Smith, R. E., and Schreiber, H. A., Point process of seasonal thunderstorm rainfall, I, Distribution of rainfall events, *Water Resour. Res.*, 9(4), 871~884, 1973