

점 강우량 결측시 보정방법에 관한 비교 연구

김응석 *, 김 형수**, 김중훈***

1. 서론

임의의 대상유역에 존재하는 강우관측소의 결측치에 대한 보정을 위하여 여러가지 방법들이 이용되고 있다. 이러한 방법들은 대부분 실측 강우관측소 지점과 결측 강우관측소 지점간의 거리, 고도차, 강우관측치 등을 이용하여 결측 강우지점의 가중치를 산정, 실측지점에 이를 적용하여 보간하는 방법이다. 본 연구에서는 결측치 보간을 위하여 널리 이용되고 있는 산술평균법(Arithmetic Average method, AA), 년정상강우량법(Normal Ratio method, NR), 수정년정상강우량법(Modified Normal Ratio method, MNR), 역거리법(Inverse-Distance method, ID), 거리고도비율법(Ratio of Distance and Elevation method, RDE), 선형계획법(Linear Programing method, LP), 크리깅방법(Simple Kriging method, KM) 등을 적용, 비교 검토하고자 한다. 즉, 한강수계의 평창강 유역을 대상으로 연 중 발생하는 큰 호우사상에 대해 임의의 강우관측소를 결측지점으로 가정하고 주변의 강우관측소로부터 각각의 방법을 이용해 가중치들을 산정하여 결측지점의 강우량 값을 보정하고자 한다. 또한 각각의 방법을 이용하여 얻어진 결과에 대해 실측값과 보정값의 오차정도를 평균절대오차법(Mean Absolute Error, MAE)과, 제곱평균자승오차법(Root Mean Squared Error, RMSE)에 의해 산정하였다.

2. 대상유역선정과 기본자료

본 연구에서는 태백산맥 지역에 위치한 한강수계의 평창강 유역에 있는 11개 관측소 지점을 선택하였고 (국제수문개발계획(IHP) 대상지역), 관측소 지점 대부분이 해발고도가 높은 관계로 산악지형 특성에 따른 영향을 많이 받아 국지적으로 큰 호우 사상을 갖는다 (표 1 및 그림 1) (김응석 등, 1995). 평창강 유역은 한강홍수 통제소에서 이를 관할하며, 11개 지점의 관측은 자기우량계를 이용하여 강우량을 관측하고 있다. 본 연구에서 사용된 자료는 평창강 유역에서 발생한 호우사상으로 1982년부터 1989년 까지 총 43개의 호우사상을 선택하였다.

본 연구에서는 그림 1.에서 등매지점을 결측지점으로 가정하고 관측지점들로부터 선택된 호우사상과 각 보정방법들을 이용하여 관측지점과 결측지점간의 가중치를 산정한 후 관측치와 가중치로부터 등매지점에 대한 결측치를 구하고, 산정된 결측치와 실제 관측된 값들과의 차이로부터 그 오차정도를 비교한다.

* 고려대학교 대학원 토목환경공학과 박사과정

** 선문대학교 건설공학부 전임강사

*** 고려대학교 토목환경공학과 부교수

표 1. 평창강 유역내 강우관측소 지점의 현황

관측지점	위도	경도	해발고(m)	10년 평균 년 강우량 (mm)
No. 1 (방림)	N 37-25-25	E 128-23-50	480	1322.02
No. 2 (가평)	N 37-26-40	E 128-29-05	470	1234.29
No. 3 (수동)	N 37-27-05	E 128-20-50	440	1302.45
No. 4 (계촌)	N 37-27-00	E 128-18-25	620	1399.24
No. 5 (대화)	N 37-29-40	E 128-27-40	400	1151.13
No. 6 (신리)	N 37-32-05	E 128-27-35	500	1358.95
No. 7 (유천)	N 37-31-40	E 128-19-40	490	1775.94
No. 8 (등매)	N 37-32-55	E 128-23-45	500	1421.78
No. 9 (용전)	N 37-36-05	E 128-26-40	560	1282.47
No. 10 (계방)	N 37-40-15	E 128-27-55	700	1390.65
No. 11 (홍정)	N 37-38-50	E 128-19-45	640	1627.75

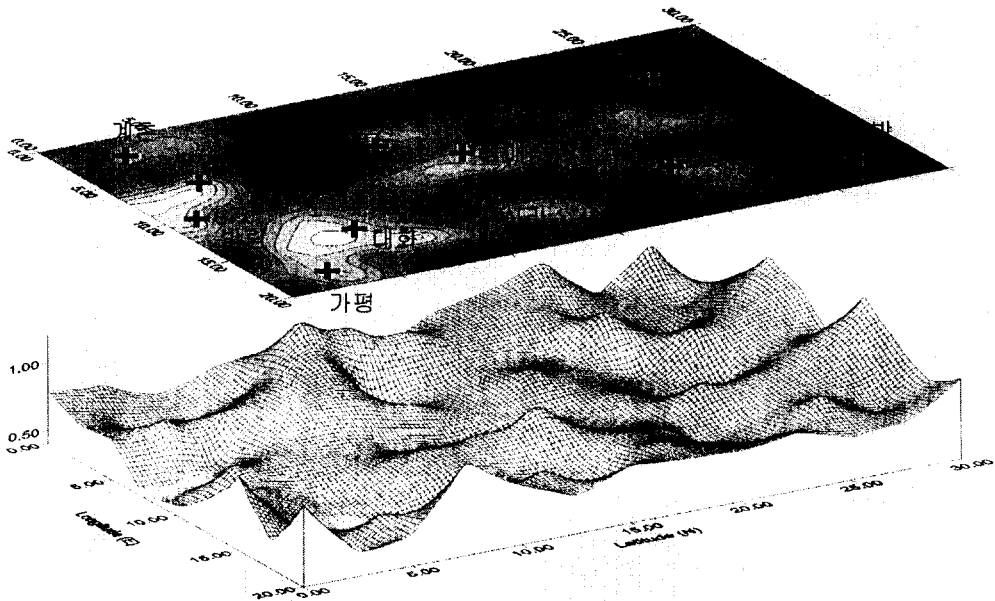


그림 1. 평창강 유역의 지형도 및 등고선

3. 결측 강우량 보정 방법

3.1 기본공식

각각의 방법을 간략하게 설명하면, AA 방법(식 1.)은 기본적인 산술평균값으로 실측지점의 관측소 개수를 평균하여 가중치를 구하며, NR 방법(식 2.)은 10년 동안 년 평균강우량 값

을 이용하여 구하였으며, MNR^a 와 MNR^b 방법(식 3.)(Tung, 1983)은 NR 방법을 수정하여 연평균강우량과 거리의 함수로 가중치를 구할 수 있다. 여기서 사용된 지수 a 또는 b 값은 적용지역에 따라서 조금씩 다르고 대략 1.5 에서 2.0 사이의 값을 시행착오방법으로 구하여야 하나 본 연구에서는 a 값은 1.5 로 b 값은 2.0 값으로 사용하였다. ID^a 과 ID^b 방법(식 4.)은 결측지점에서 각각의 관측지점에 대한 거리의 반비례로 가중치를 구하였으며, RDE 방법(식 5.)은 결측지점과 관측지점과의 거리에 비례하며, 고도차에 반비례하는 함수 관계로 가중치를 구할 수 있다

(1) 산술평균법(AA)

$$R = \frac{1}{N} \sum_{i=1}^n P_i \quad \dots(1)$$

(2) 년정상강우량법 (NR)

$$R = \frac{N_x}{n} \sum_{i=1}^n \left(\frac{P_i}{N_i} \right) \quad \dots (2)$$

(3) 수정년정상강우량법 (MNR)

$$R = \frac{\sum_{i=1}^n D_i^{(1/b)} \left(\frac{N_R}{N_i} \right)}{\sum_{i=1}^n D_i^{(1/b)}} \quad \dots(3)$$

(4) 역거리법 (ID)

$$R = \frac{\sum_{i=1}^n \left(\frac{P_i}{D_i} \right)}{\sum_{i=1}^n \left(\frac{1}{D_i} \right)} \quad \dots(4)$$

(5) 거리고도 비율법 (RDE)

$$R = \frac{\sum_{i=1}^n \left(\frac{P_i}{D_i^b} \right)}{\sum_{i=1}^n \left(\frac{1}{D_i^b} \right)} \quad \dots(5)$$

여기서, P_i 는 관측지점의 시 강우량, N 는 관측지점의 개수, ΔE_i 는 결측지와 관측지점 사이의 고도차, D_i 는 등매의 관측지점에서 기지의 관측지점까지의 거리, N_R 는 결측지점의 년 평균강우량(등매 관측지점) 값, N_i 는 관측지점의 년 평균강우량 값, n 는 관측지점의 총 관측 시간 개수이다.

3.2 선형계획법 (LP)

$$\min \sum_{j=1}^n (U_j + V_j) \quad \dots(6)$$

$$\text{subject to } \sum_{i=1}^n a_j P_{ij} - U_j + V_j = P_{bj}, \quad j = 1, \dots, K \quad \dots(7)$$

$$\sum_{i=1}^n a_i = 1.0 \quad \dots(8)$$

여기서, $a_i \geq 0, i = 1, \dots, n$; $U_j \geq 0, V_j \geq 0, j = 1, \dots, K$, U_j 와 V_j 는 각각의 결측관측지점의 Slack variables 이며, P_{ij} 는 각각의 관측지점에서의 시간 강우량, P_{bj} 는 결측지점에서의

시간 강우량, a_i 는 각각의 관측지점의 가중치, K 는 관측하고자 하는 강우 시간의 개수이다. 이방법의 목적은 구하고자 하는 가중치 값의 편차를 최소화하기 위한 것이다. 본 연구에서는 GAMS (Brook et. al, 1992) 프로그램을 이용하여 가중치 값을 계산하였다.

3.3 크리깅 방법 (KM)

크리깅 방법은 분산이 최소가 되도록 불편의 된 가중치들을 구하는 선형기법이다 (Kitanidis, 1983; Delhomme, 1978). 어떤 지점들 x_1, x_2, \dots, x_n 에 있어서 변수 z 에 대한 n 개의 관측치가 있을 때, 이의 선형조합에 대한 추정치는 식(9)와 같이 표현할 수 있다.

$$\hat{Z}_0 = \sum_{i=1}^n \lambda_i Z(x_i) \quad \dots(9)$$

여기서, \hat{Z}_0 는 참값 Z_0 의 추정치이며 λ_i 는 지점 x_i 의 관측치에 대한 가중치이다. 불편의를 나타내는 식은 참값과 추정치의 차이에 대한 기대치가 0인 경우이며 식(10)과 같이 쓸수 있다.

$$E[\hat{Z}_0 - Z_0] = 0 \quad \dots(10)$$

여기에서 E 는 기대치를 나타낸다. 그리고 분산이 최소가 되도록 하는 것은 식(11)과 같이 나타낼수 있다.

$$E[(\sum_{i=1}^n \lambda_i Z(x_i) - Z_0)^2] = \text{minimum} \quad \dots(11)$$

또한 불편의 조건을 만족하기 위하여 가중치들은 식(12)의 조건을 만족해야 한다.

$$\sum_{i=1}^n \lambda_i - 1 = 0 \quad \dots(12)$$

추정치와 참값의 차이로부터 발생하는 추정오차는 식(13)과 같이 나타낼수 있다.

$$e = \sum_{i=1}^n \lambda_i Z(x_i) - Z_0 \quad \dots(13)$$

크리깅 기법은 식(12)를 제한조건으로 식(11)을 최소화 하는 것이다. 최소화를 위한 필요조건은 식(14)와 같은 미지 값 들과 함께 $n+1$ 개의 식을 갖는다.

$$\sum_{j=1}^n \lambda_j \gamma(x_i - x_j) = \gamma(x_i - x_0) + \mu, \quad i = 1, 2, \dots, n \quad \dots(14)$$

$$\sum_{i=1}^n \lambda_i = 1 \quad \dots(15)$$

여기서 $\gamma = E[Z(x_i) - Z(x_j)]^2$ 는 semivariogram 을 나타내며 μ 는 Lagrange multiplier 이다. 식(14)와 (15)를 풀어 최적 가중치들을 얻을수 있으며, 추정치에 대한 분산은 식(16)과 같이 주어진다.

$$\sigma^2 = -\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i - x_j) + 2 \sum_{i=1}^n \lambda_i \gamma(x_i - x_0) \quad \dots(16)$$

x_0 가 x_i 와 일치할 때 시스템의 해는 λ_i 가 1이 되며, 이는 x_i 에 있어서 $\hat{Z}_0 = Z(x_i)$ 이고 $\sigma^2=0$ 이 된다.

4. 각 방법에 의한 가중치 산정

등매 관측지점을 결측지점으로 가정하고, 주변의 관측지점으로부터 각각의 방법에 의하여 가중치를 구한다. 우선 주변의 5개(방림, 수동, 신리, 계방, 흥정) 지역을 중심으로 가중치 값을 구하고, 계속해서 결측지점과 가까운 거리의 순서로 1개씩 관측지점을 증가하여 가중치 값을 구하는 방법으로 가중치 값을 산정하였다. 표 2는 각각의 관측 지점간의 거리를 1:50000 지도에서 구한 값이며, 표 3은 5개 지점을 고려했을 때 각 방법에 의해 산정된 가중치들이다.

표 2. 각 관측지점간의 거리 (Km)

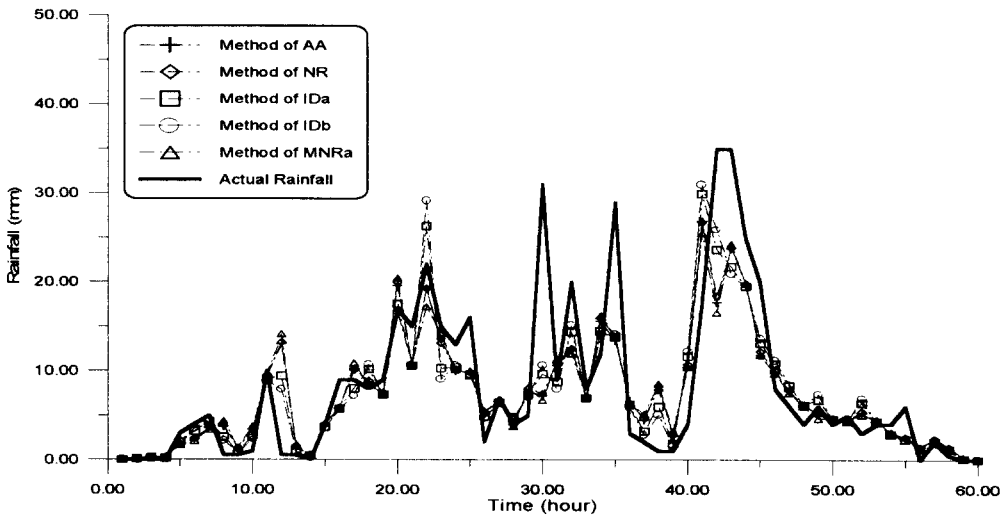
구분	방림	가평	수동	계촌	대화	신리	유천	등매	용전	계방	흥정
방림	0.0	8.3	5.05	7.45	9.9	13.4	12.8	13.7	20.05	27.9	25.35
가평	-	0.0	12.1	15.75	6.0	10.2	16.7	14.05	17.8	25.1	26.55
수동	-	-	0.0	3.7	11.1	13.45	8.65	11.65	18.75	26.45	21.8
계촌	-	-	-	0.0	14.45	16.4	8.8	13.5	20.8	28.2	21.95
대화	-	-	-	-	0.0	4.6	12.3	8.3	11.95	19.55	20.7
신리	-	-	-	-	-	0.0	11.7	5.9	7.6	15.1	17.75
유천	-	-	-	-	-	-	0.0	6.5	13.25	20.0	13.25
등매	-	-	-	-	-	-	-	0.0	7.3	14.8	12.55
용전	-	-	-	-	-	-	-	-	0.0	7.85	11.7
계방	-	-	-	-	-	-	-	-	-	0.0	12.6
흥정	-	-	-	-	-	-	-	-	-	-	0.0

표.3 각 방법에 의해 산정된 가중치

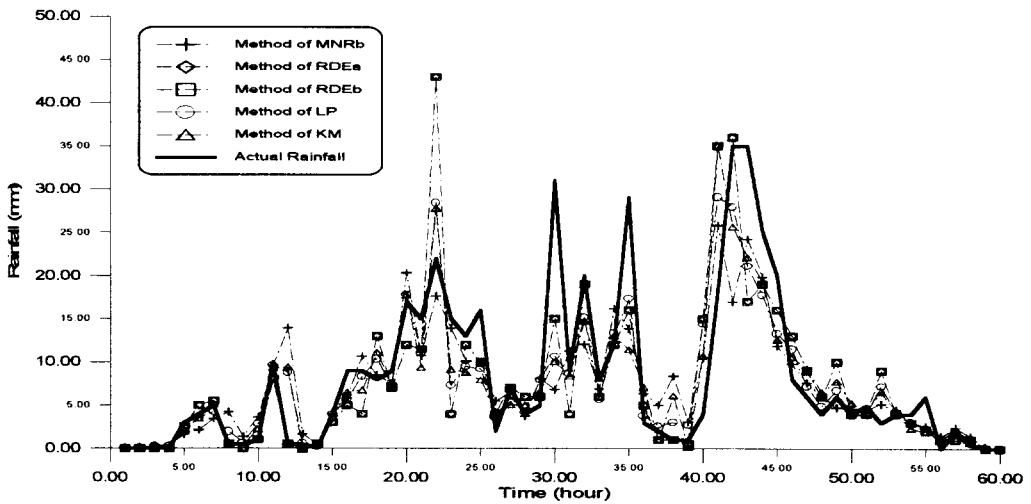
방법 지역	AA	NR	ID ^a	ID ^b	MNR ^a	MNR ^b	RDE ^a	RDE ^b	LP	KM
방림	0.2	0.2151	0.1147	0.0856	0.2438	0.2375	0.0171	0.0045	0.086	-0.018
수동	0.2	0.2032	0.1462	0.1184	0.2068	0.2069	0.0026	0.0004	0.160	0.2140
신리	0.2	0.2092	0.5063	0.6205	0.1226	0.1408	0.9789	0.9950	0.542	0.5335
계방	0.2	0.2045	0.1021	0.0734	0.2441	0.2347	0.0006	0.0001	0.212	0.0277
흥정	0.2	0.1747	0.1308	0.1021	0.1868	0.1846	0.0008	0.0001	0.000	0.2432

5. 결측 호우사상의 추정 및 오차

43 개 호우사상에 대하여 각각의 방법으로 구한 가중치를 이용해 결측지점에 대한 호우사상을 각각의 보간방법들에 의해 산정하였고, 이들 산정된 값과 관측치와의 오차를 평균절대오차법(MAE)과 제곱평균자승오차법(RMSE)으로 실제의 값과 비교해 보았다. 그림 2(a)와 (b)는 각 방법들에 의해 산정된 가중치를 이용해 등매지점에 대한 결측치를 구한후 이 지점의 실제 관측치와 산정된 결측치를 비교한 것이다. 표 4와 5는 등매지점에 대해 산정된 결측치와 실제 관측치간의 오차를 이용한 관측지점의 수에 따라 구한 것이다.



(a)



(b)

그림 2. 등매지점에 대해 각 방법으로부터 산정된 결측치와 관측치의 비교

표 4. 등매지점에 대해 산정된 결측치와 관측치의 평균절대오차 (mm)

방법 지역	AA	NR	ID ^a	ID ^b	MNR ^a	MNR ^b	RDE ^a	RDE ^b	LP	KM
5 개	1.684	1.677	1.540	1.516	1.736	1.722	1.613	1.613	1.485	1.556
6 개	1.764	1.601	1.503	1.491	1.666	1.65	1.613	1.613	1.436	1.506
7 개	1.630	1.541	1.433	1.415	1.61	1.593	1.613	1.613	1.371	1.465
8 개	1.555	1.528	1.439	1.424	1.583	1.568	1.613	1.613	1.369	1.452
9 개	1.547	1.555	1.451	1.432	1.623	1.607	1.613	1.613	1.369	1.469
10 개	1.561	1.569	1.456	1.436	1.628	1.614	1.565	1.613	1.369	1.484

표 5. 등매지점에 대해 산정된 결측치와 관측치의 제곱평균자승오차 (mm)

방법 지역	AA	NR	ID ^a	ID ^b	MNR ^a	MNR ^b	RDE ^a	RDE ^b	LP	KM
5 개	3.289	3.263	3.082	3.066	3.361	3.338	3.473	3.473	3.016	3.191
6 개	3.681	3.116	2.977	2.991	3.226	3.199	3.473	3.473	2.906	3.073
7 개	3.400	3.007	2.889	2.892	3.113	3.085	3.473	3.473	2.896	3.057
8 개	3.241	3.02	2.921	2.916	3.102	3.08	3.473	3.473	2.903	3.063
9 개	3.172	3.084	2.939	2.924	3.186	3.161	3.473	3.473	2.905	3.076
10 개	3.128	3.125	2.957	2.933	3.227	3.202	3.402	3.473	2.905	3.083

6. 토론 및 결론

본 연구에서는 평창강 유역의 등매지점을 강우 결측지점으로 가정하고 3 절에서 설명한 각 방법들을 이용하여 관측지점의 개수에 따른 가중치를 산정하였다. 또한 산정된 가중치를 근거로 등매지점의 결측치를 구하여 이 지점의 실제 관측치와 비교하고, 각 방법에 따른 오차의 정도를 MAE 와 RMSE 방법으로 검토하였다.

가중치 값의 산정에서 ID, MNR, RDE, KM 방법은 신리와 유천 관측지점에 큰 가중치 값을 보이고 있다. 이는 이들 방법들이 거리의 비례하여 가중치를 산정되는 함수로 구성되어 있기 때문에 큰 값을 가지는 것으로 나타났다. 특히 RDE 방법에서 구한 가중치 값은 신리 지점이 가장 큰 값인 1에 가까운 값을 보여 주고 있다. 또한 KM 방법에서 구한 가중치 중에는 음의 값을 가지는 지점도 볼 수 있었다. 이는 등매 관측지점으로부터 상대적으로 다른 지점에 비해 거리가 멀리 떨어져 있는 지점에서 볼 수 있었으며, 이는 이들 지점들이 결측지점과 어느 정도 이상 멀리 떨어져 있을 경우 아무런 영향을 미치지 않음을 알 수 있다.

관측지점별 개수에 따른 오차산정에서는 LP의 방법이 다른 방법에 비해서 보다 더 작은 오차 값을 보여 주었다. 그러나, 이 LP 방법은 결측지점의 자료가 있다고 가정하였기에 결측지점의 강우량 값과 주변 지역에서 관측된 값과의 편차를 줄이면서 계산한 가중치 값이므로 실제 실무에서 적용은 다소 어려운 점이 있다고 생각된다. 그 외 다른 방법들이 보여준 오차 값에서 ID, KM, NR, MNR, RDE 방법 순서로 작은 오차 값을 보여주고 있다.

본 연구에서 기존의 강우량 보간에 관한 방법들을 살펴 보았으며, 특히 많이 사용되는

ID 방법은 다른 방법들에 비해서 손쉽게 결측지점의 강우량을 보정할 수 있으며, 또한 기존에 많이 사용되는 방법들과 비교해 상대적으로 작은 오차를 가지는 것을 알 수 있었다. 또한 이들 방법 이외에 신경망 이론이나 유전자 알고리즘과 같은 새로운 이론을 적용해 본다면 기존의 간단한 방법과 새로운 방법들간의 차이를 비교해 볼 수 있을 것으로 사료된다.

참고문헌

1. 김응석, 김중훈, 김태균 (1995). “산악지역을 고려한 점강우량 결측시 보정방법”, 대한토목학회 학술발표회 논문집.
2. Brook, A., Kendrick, D., and Meeraus, A. (1992). “Release 2.25 GAMS user guide”, The Scientific Press.
3. Delhomme, J. P. (1978). “Kriging in Hydrosciences”, Advances in Water Resources, Vol. 1, No. 5, pp. 251-266.
4. Kitanidis, J. (1983). “Statistical Estimation of Polynomial Generalized Covariance Functions and Hydrological Applications”, Water Resources Research, Vol. 19, No. 4, pp. 909-921.
5. Tung, Y. K. (1983). “Point Rainfall Estimation for a Mountainous Region”, Journal of Hydraulic Engineering, Vol. 109, No. 10, pp.1386-1398.