엔트로피 이론을 이용한 강우관측망 평가; 혼합분포와 연속분포의 적용 비교

유철상[1] · 정광식[2]

1. 서론

엔트로피는 일반적으로 무질서도 또는 불확실성의 척도로 알려져 있으나, 정보이론(information theory)에서는 신호가 가지고 있는 정보용량으로 정의되고 있다(Shannon과 Weaver, 1949). 즉, 정보교환의 과정으로 신호(signal)가 보내질 때, 이 신호의 불확실성은 신호에 대한 정보가 불확실성을 제거할 정도로 많아지면 감소하게 되고, 따라서 불확실성의 감소정도로 신호에 대한 정보를 간접적으로 측정할 수 있는 것이다.

이러한 엔트로피의 개념은 많은 분야에 응용되고 있으며 특히 수문분야에서는 관측망을 평가하는데 주로 이용되어 왔다. 예를 들어, Caselton과 Husain(1980)은 수문관측망 설계에 정보전달의 개념을 적용한 예가 있고, Chapman(1986)은 엔트로피를 이용하여 수문학적 자료들이 가지고 있는 불확실성을 추정하고 모델에 적용시켜 불확실성의 감소정도를 평가한 바 있다. Al-Zahrani와 Husain(1998)은 조밀한 관측망에서의 최적 관측소 개수와 밀도가 낮은 관측망에서의 관측망 확장에 엔트로피 개념을 이용하였다. 그러나 현재까지 엔트로피 개념을 이용한 관측망의 평가에서는 강우에 대한 정보만이 고려되고 있으며, 무강우기간에 대해서는 고려되고 있지 않고 있다.

일반적으로 강우는 그 시-공간적 변동성이 크고 아울러 시-공간적으로 큰 간헐성(intermittency)을 보이기 때문에, 특히 큰 유역의 경우 강우가 유역 전체에 고르게 내리는 경우는 드물다. 이는 시간강우를 사용하는 경우뿐만 아니라 일 강우를 다루는 경우도 마찬가지이다. 특히, 시간 단위의 자료를 사용하는 경우는 일 단위의 자료를 사용하는 경우에 비해 강우의 시-공간적 변동정도가 크므로 강우의 특성 자체가 관측망의 평가에 중요한 요소로 작용하게 되나 일 자료를 사용하는 경우는 상대적으로 강우의 발생확률 및 각 지점사이의 동시발생 정도가 관측망의 평가에 중요한 요소로 작용하게 된다. 그러나 보다 엄밀히 살펴보면, 두 경우 모두 강우 자체의 특성뿐만 아니라 강우의 발생확률 및 각 지점 사이의 동시발생 확률 등이 모두 중요하게 고려되며 단지 이동평균의 정도에 따른 평활화(smoothing)의 정도만이 달라질 뿐이다. 물론 상대적으로 작은 유역에 대해 홍수기만을 대상으로 강우 관측망을 평가한다면 전 유역에 대해 강우가 발생이 일정하다고 가정하는 경우와 큰 차이가 없으리라 판단된다. 그러나, 대유역의 경우는 이러한 경우와는 다른 양상을 보일 것임이 분명하다.

본 연구에서는 무강우에 대한 정보와 강우에 대한 정보를 함께 고려하기 위해서 혼합 대수정규분포를 이용한 관측망 평가를 수행하였으며, 그 결과를 강우에 대한 정보만을 고려하였을 경우(연

1) 성균관대학교 토목환경공학과 부교수(cyool@skku.ac.kr)
2) 유일엔지니어링 상하수도부 사원(edjks@hanmail.net)

속분포인 대수정규분포를 적용한 경우)와 비교하였다. 본 연구에서는 일 강우량 자료를 이용하였고 충주댐 상류유역에 적용하였다. 사용된 자료가 일 강우량 자료이므로 평가된 관측망도 치수보다는 이수의 관점에서 이해되는 것이 바람직하다고 판단된다. 그러나 혼합분포 및 연속분포를 적용한 결과를 비교함으로서 강우 자체의 특성만을 이용한 경우와 강우의 공간적 발생특성을 함께 고려하는 경우의 차이를 비교할 수 있을 것이며, 이를 통해 치수 및 이수 측면에서의 관측망의 차이도 살펴볼 수 있을 것이다.

2 강우관측망 평가

2.1 엔트로피를 이용한 정보량 측정

Shannon과 Weaver(1949)에 의하면 이산무작위변량(discrete random variable) X에 대한 한계엔트로피는 다음과 같이 정의된다.

$$H(X) = -\sum_{n=1}^{N} P(x_n) \ln P(x_n), \quad n = 1, 2, 3, \ldots, N \tag{1}$$

여기서 $p(x_n)$은 x_n의 발생확률이며, 한계엔트로피 $H(X)$은 X가 가지고 있는 정보량 또는 불확실성을 의미한다.

만일 무작위 변량 x_n과 관계 있는 $y_m (m = 1, 2, \ldots, N)$이 존재할 경우, y_m으로부터 x_n을 추정한다면 x_n이 가지고 있는 불확실성을 감소시킬 수 있을 것이다. 이러한 원리에 의해 주어진 변량 Y에 의해 무작위 변량 X에 남아 있는 불확실성은 다음과 같이 추정할 수 있다.

$$H(X|Y) = -\sum_{n=1}^{N} \sum_{m=1}^{N} P(x_n, y_m) \ln P(x_n | y_m) \tag{2}$$

여기서 $P(x_n, y_m)$은 $X = (x_n)$와 $Y = (y_m)$의 결합확률을, $P(x_n | y_m)$은 주어진 Y에 대한 X의 조건확률을 의미한다. 따라서 $H(X|Y)$는 주어진 Y에 대한 X의 조건엔트로피이며, 이는 X와 Y 간의 정보전달 사이에 손실되는 정보량을 나타내기도 한다(Yang과 Burn, 1994).

주어진 Y에 의해서 X가 가지고 있는 불확실성의 감소정도 또는 X와 Y 사이의 정보 전달량(공유되는 정보 혹은 중복되는 정보)은 다음과 같다.

$$T(X, Y) = H(X) - H(X|Y) \tag{3}$$

2.2 엔트로피를 이용한 관측망 평가

유역 내 존재하는 관측망의 최적화는 관측소간의 중복되는 정보가 최소가 되도록 관측소의 수를 줄이고 아울러 유지되는 관측소들로부터 유역의 정보를 최대로 얻을 수 있어야 함을 의미한다. 즉, 최소의 관측소로부터 유역의 정보를 최대로 얻을 수 있어야 하며 따라서 최적화의 목적함수는 다음과 같이 나타낼 수 있다(Al-Zahrani와 Husain, 1998).

$$\text{MAX} \quad T(X_1, X_2, \ldots X_m; X_k, X_l, \ldots X_p) \tag{4}$$

여기서, m은 현재 유역 내에 있는 관측소의 총 개수이며, p는 유지되어야 할 관측소 개수를 의미한다.

따라서 $T(X_1, X_2, \ldots X_m; X_k, X_l, \ldots X_p)$는 p개의 관측소로부터 얻을 수 있는 유역의 정보를 의미한다. 식 (4)는 다음과 같이 나타낼 수도 있다.

$$\text{MAX} \sum_{i=1}^{m} T(X_i; X_k, X_l, \ldots\ldots X_p)$$

$$= \text{MAX}\left[H(X_k) + \ldots . H(X_p) + \sum_{i=1}^{m-p} \sum_{j=k}^{n} T(X_i, X_j)\right], i \neq j \qquad (5)$$

$H(X_k) + H(X_l) + \ldots\ldots H(X_p)$는 선택된 관측소 각각의 정보량을 합한 것이며, $\sum_{i=1}^{m-p} \sum_{j=k}^{n} T(X_i, X_j)$는 선택된 관측소와 선택되지 않은 관측소간의 정보 전달량, 또는 선택한 관측소로부터 얻을 수 있는 선택되지 않은 관측소의 정보량을 의미한다. 선택된 관측소의 수가 증가할수록 얻을 수 있는 정보량은 증가할 것이며 어느 시점 이후에는 관측소들 간의 중복되는 정보량에 의해 선택된 관측소로부터 얻을 수 있는 정보량이 감소하게 된다. 따라서 최적의 관측망이란 유역에 대한 정보량을 최대로 얻을 수 있는 관측소 조합을 의미한다.

3 적용 예

3.1 대상유역

본 연구에서는 충주댐 상류지역을 대상으로 강우관측망을 평가하였다. 최근 수자원 분야에서 활발히 이용되고 있는 수자원 단위지도에는 충주댐 상류지역을 3개의 중권역(1001, 1002, 1003 유역)으로 나누고 있다. 본 연구에서는 이들 중 1001 유역에 대한 분석을 수행하기로 한다.

1001 유역은 남한강 상류에서 평창강 합류지점이며 유역면적은 2,445.84 ㎢이다. 현재 1001 유역 내에는 20개의 관측소가 운영중이며 관측망 평가에는 1993~1999년의 일 강우량 자료가 사용되었다.

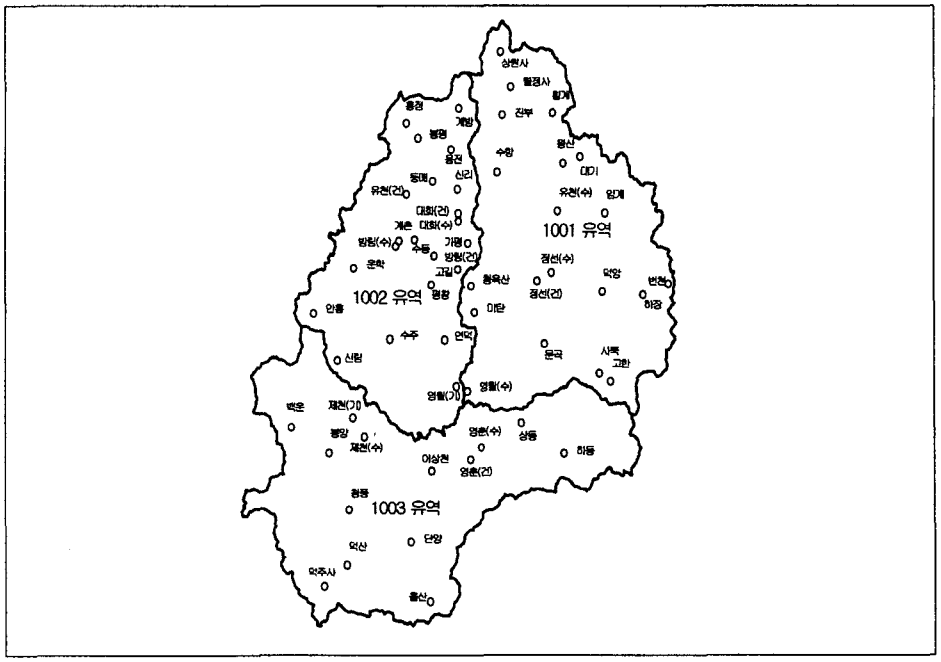

그림 1. 충주댐 상류지역 관측소 현황(수:수자원공사, 건:건설교통부, 기:기상청)

3.3 강우관측망 평가 결과

1001 유역에 대해 정보전달 행렬을 혼합분포와 연속분포를 적용하여 각각 구성하고, 유역내의 관측소로부터 유역 강우에 대한 정보를 최대로 얻을 수 있는 조합을 구성하였다. 이때 얻어지는 정보량을 도시한 것이 그림 2이다. 그림에서는 강우자료의 계급구간 수를 100, 200, 300, 400, 및 500개인 경우에 대해 총 정보량을 각각 도시하고 있으며, 계급구간 수가 증가함에 따라 정보전달량은 증가하지만 선택되는 관측소의 수나 위치에는 큰 영향을 미치지 않는 것으로 나타났다.

그러나 혼합분포를 사용하는 경우와 연속분포를 사용하는 경우의 최적 관측소의 수에는 크게 차이가 있는데, 이는 기본적으로 강수부분이 얼마나 민감하게 고려되느냐의 차이로 설명된다.

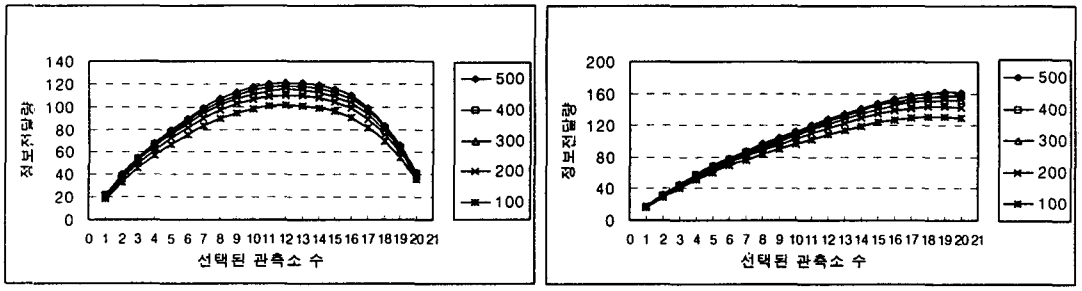

그림 2. 1001 유역 정보전달량 변화(좌:혼합분포, 우:연속분포)

표 1은 관측망 평가결과로서 유역별 최적 관측소 수 및 이때의 우량계 밀도를 나타내고 있다. 관측망 평가결과, 최적 관측소 수는 혼합분포를 적용한 경우가 연속분포를 적용한 경우보다 더 작게 나타났다. 이렇게 최적 관측소 수가 상이하게 나타나는 이유는 그림 2를 통해 쉽게 파악할 수 있다.

표 1. 유역별 최적 관측소 수 및 밀도

유역	현재	최적 관측망	
		혼합분포	연속분포
관측소 수	20	12	18~19
관측소 밀도(1/km^2)	122.3	203.8	135.9~128.7

우리나라에서 통용되고 있는 우량계 밀도의 기준은 세계기상기구(World Meteorological Organization; 이하 WMO라 칭함)에서 권장하는 우량계 밀도이다(윤용남, 2000). 우리나라가 포함된 온대기후 지역의 경우 평지에는 1개소/(600 ~ 900)㎢가, 산지에는 1개소/(100 ~ 250)㎢이 권장되고 있다. 표 1에서 살펴볼 수 있는 것처럼 혼합분포를 이용하여 최적화된 강우관측망이 WMO의 산지기준을 무리 없이 만족시키고 있는 것이 특이하다. 반면에 연속분포를 이용한 경우는 우량계의 수가 과다하게 산정되는 결과를 보여주고 있다.

4 결론

본 연구에서는 엔트로피의 개념을 이용하여 충주댐 상류지역의 관측망을 평가하였다. 강우에 대한 정보만을 고려하는 연속분포와 강우와 무강우에 대한 정보를 함께 고려하는 혼합분포를 각각

적용함으로서 두 경우가 비교될 수 있도록 하였다. 본 연구의 결과를 정리하면 다음과 같다.

(1) 엔트로피를 이용하여 관측소의 정보량을 추정한 결과 관측소 고유의 정보량은 연속분포를 적용하는 경우에 더 크게 추정되었으며, 선택된 관측소와 선택되지 않은 관측소들간의 중복되는 정보량의 총합은 혼합분포를 적용하는 경우에 더 크게 추정되었다.

(2) 현재 운영중인 관측망의 최적관측소 수는 혼합분포의 경우가 더 적게 추정되었으며, 혼합분포를 이용하여 최적화된 강우관측망도 WMO의 산지기준을 무리 없이 만족시키고 있는 것으로 나타났다. 반면에 연속분포를 이용한 경우는 우량계의 수가 과다하게 산정되는 결과를 보여주고 있다. 따라서, 본 연구의 결과는 강우관측망의 평가에 혼합분포를 이용해야 한다는 점을 간접적으로 시사해 주는 것이라고 말할 수 있다.

서론에서 언급한 바와 같이 본 연구의 결과는 일 강우자료를 이용한 것으로 치수보다는 이수측면에서 이해되어야 할 것이다. 그러나 연속분포와 혼합분포를 적용하는 경우의 차이가 매우 크게 나타났듯이 치수를 목적으로 하는 경우와 이수를 목적으로 하는 경우의 관측망이 크게 다를 수 있음도 아울러 살펴볼 수 있다. 따라서 추후의 연구에서는 시간 자료를 이용한 강우관측망의 평가와 함께, 우기만을 대상으로 한 평가 및 강우사상 별 평가가 동시에 수행되어 비교되어야 할 것으로 판단된다.

5. 참고문헌

윤용남 (1998). 공업수문학, 청문각.

Al-Zahrani, M. and Husain, T. (1998). "An algorithm for designing a precipitation network in the south-western region of Saudi Arabia" *J. Hydrology*, Vol. 205, pp. 205-216.

Caselton, W. F. and Husain, T. (1980). "Hydologic networks: Information transmission" *J. Water Resources Planning and Management Division*, ASCE, 106, WR2, pp. 503-529.

Chapman, T. G. (1986). "Entropy as a measure of hydrologic data uncertainty and model performance" *J. Hydrology*, Vol. 85, pp. 111-126.

Shannon, C. E., and Weaver, W. (1963). *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, Ill.

Yang, Y. and Burn, D. H. (1994). "An entropy approach to data collection network design" *J. Hydrology*, Vol. 157, pp. 307-324.