

# \*일기엔트로피 및 정보비에 의한 한국 장마기의 일기대표성 설정 ( 4대강 유역을 중심으로 )

박 현 옥

(조선대학교 강사)

## I. 서 론

### 1) 연구목적

한반도는 아시아 대륙의 북동쪽에 위치하여 지형이 복잡하여 강수현상의 출현빈도와 그 계절변화는 한반도의 수리적, 지리적 요인에 의해 강한 국지성을 보인다. 이러한 특정 일기의 특징은 일기출현율과 연변화 상에 잘 반영되고 있으나 각 지역의 기상 및 기후현상의 예측, 예보 및 기후대표성 설정은 매우 어렵다. 그런데 우리나라의 기후특성을 잘 반영하는 강수현상의 공간적 분포나 시간적 변동은 복수지점의 여러 기후요소와 상호 관련하여 발생하며, 강수의 특정한 process가 시, 공간상에 편중할 때 강수현상의 지역성 즉 대표성은 나타난다. 이러한 각 지역의 강수현상(일기)의 대표성은 정보이론을 이용하여 얻은 일기의 평균 정보량이며 통계학적인 parameter인 일기엔트로피와 정보비 개념을 기초로 하면 수리적, 지리적 요인 및 계절변화 등을 포함하여 보다 항상적으로 그 모양을 밝힐 수 있다.

본 연구에서는 각종 기후 인자의 영향을 명확히 반영하여 한반도의 장마기 날씨 및 기후특성을 잘 나타내는 탁월일기 및 강수현상의 장마기 출현빈도와 순별변화에 대해, 정보이론을 이용하여 일기엔트로피와 정보비를 추출하고 응용하여 그 공간스케일의 시간적 변동을 살폈고, 4대강 유역(서울, 대구, 광주, 충주, 부여)의 일기대표성을 정보비의 거리에 따른 감쇠특성과 정보비의 차에 의해 밝히고 그에 따라 일기대표성(예보구역)을 설정하고자 한다.

### 2) 자 료

본 연구에서는 기상청에서 발행한 1990-1999년까지의 기상월보에서 남한의 73개 관측지점의 매일의 기상자료 중 사계절을 대표하는 1,4,7,10월의 운량과 강수자료를 맑음(강수 없음), 강수(강수량 1mm 이상/일)의 두 계급으로 분류하여 분석하였다. 본 연구에 이용한 연구지역의 지리적 범위와 기상관측소는 그림 1에 나타냈다.

### 3) 연구방법

(1) 일기엔트로피에 관하여

Shannon(1948)에 의하여 제안된 엔트로피는 정보이론에서 중요시되는 개념으로써

열역학 제 2법칙, 「볼츠만의 통계역학」에 있어서의 엔트로피와 똑같은 형태이며, 정보원에서의 각 요소가 갖는 정보량의 평균값을 나타내는 확률적 개념의 측도(測度)이다. 이 엔트로피의 기상·기후학에서의 응용은 일기의 평균정보량인 일기엔트로피 및 정보비 개념에 의해 가능하다.

본 논문에서의 일기엔트로피는 다음 식에 의해 계산한다.

$$H = - \sum_{i=1}^m p_i \log_2 p_i \text{(단위 bits)}$$

여기에서 H는 엔트로피,  $p_i$ 는  $i$ 로 되는 날씨의 출현확률이다.  $i$ 는 1일 24시간 중에서 강수가 있는가 없는가의 2계급으로 나뉘므로  $i=1$ (맑음),  $i=2$ (강수)이다. 엔트로피는 어느 일기도 똑같이 나타날 때 ( $p_1=p_2=\dots=p_m$ ), 최대로 되며 이때 불확정성이 제일 크고, 출현확률이 특정의 일기에 집중하면 작기 때문에, 이와 같은 의미에서 엔트로피는 일기상태를 아는 지표의 1개로 있다.

## (2) 정보비에 관하여

일반적으로 각 지역이 갖는 일기엔트로피는 어느 기간중 매일 완전한 예측이 그 지역에 부여될 때의 1일 평균 정보량이다. 그런데 예보와 실황과의 일치정도를 비교하기에는 평균 정보량 만으로서는 충분하지 않다. 이는 어느 예보자가 어느 지점에 대해 매일 예보를 냈던 경우, 예보가 완전하게 적중한다면 그 지점 고유의 엔트로피에 똑같은 정보량을 준 것으로 되나 그 지점 고유의 엔트로피의 대소에 의해 그 의미가 달라지기 때문이다.

또한 예보의 일기가 맞지 않으면 평균으로 하여 정보량의 손실이 있게 되고 그 결과의 평균정보량  $\bar{I}$ 는 다음 식으로 계산된다.

$$\bar{I} = \sum_i \sum_j P(x_i, y_j) \log_2 \frac{P_{y_j}(x_i)}{P(x_i)}$$

여기에서  $x$ 는 실황,  $y$ 는 예보를 나타내며  $P(x_i, y_j)$ 는  $x_i$ 라는 실황과  $y_j$ 라는 예보가 동시에 일어나는 확률(동시확률),  $P_{y_j}(x_i)$ 는 한개 지점에 있어서의 일기  $y_j$ 에 대해 타 지점의 일기( $x_i$ )가 일어나는 확률(조건부 확률,  $y_j$ 가 조건),  $P(x_i)$ 는  $x_i$ 라는 실황의 출현확률이다.

위의 식은

$$\bar{I} = \sum_i P(x_i) \log_2 \frac{1}{P(x_i)} - \sum_i \sum_j P(y_j) P_{y_j}(x_i) \log_2 \frac{1}{P_{y_j}(x_i)}$$

여기에서 제1항은 일기가 완전히 일치하는 경우의 정보량이며, 제2항은 지역성에 의한 일기의 차이에서 오는 정보량의 손실이다.

결국 평균정보량  $\bar{I}$ 를 처음의 실황 엔트로피에 대한 비로서 표시한 것 ( $IR = \frac{\bar{I}}{H}$ )

$$IR = \frac{\sum_i \sum_j P(x_i, y_j) \log_2 [P_{y_j}(x_i) / P(x_i)]}{\sum_i P(x_i) \log_2 \frac{1}{P(x_i)}}$$

가 정보비이다.

이러한 정보비는 두 지점의 일기별 빈도가 어느 정도 관계가 있는가를 표시하는 지수이므로 일기 정보비를 계산하여 인접하는 2개 혹은 n개 지역의 예보중심의 대표범위를 논할 수 있다. 또한 정보비는 일기의 대표성의 거리에 따른 감쇠를 나타내는 측도(測度)로 유효하며 정보비의 거리에 따른 감쇠는 실험식으로 표시될 수 있고, 정보비의 반감거리도 구할 수 있어 기상관측지점의 입지선정과 예보적중률 향상에 합리적이고 과학적인 타당한 근거를 제시할 수 있다.

## II. 장마기의 강수, 일기엔트로피 및 정보비의 분포 특성

남한의 73개 관측지점 중 중,남부지방의 인문 및 자연지리적인 중심지이고, 특히 한국의 대부분의 인문, 경제활동이 집중적으로 이루어지고 있는 곳이면서 4대 강 유역의 중심지인 서울, 충주, 대구, 광주, 부여를 우선 연구대상 지점으로 선정하였다. 1990년~1999년의 매일의 일기자료에서 특히 각종 자연 및 기상재해가 두드러지게 나타나는 하계 중 장마기의 강수 분포, 일기엔트로피 및 정보비의 분포특성을 살폈다.

### 1) 장마기 강수의 분포 특성

남한 장마기의 순별 강수일수 및 일기엔트로피의 전국적 및 4대 강 유역의 중심지(5대 기준지점)의 분포상황을 표 1과 그림 2에 나타냈다.

표 1. 5개 기준지점의 장마기(순별) 강수일수 및 일기엔트로피 (1990-1999)  
(단위 : 일, bits)

순 지 점	6월하순		7월초순		7월중순		8월하순		9월초순		9월중순	
	강수일수	엔트로피	강수일수	엔트로피	강수일수	엔트로피	강수일수	엔트로피	강수일수	엔트로피	강수일수	엔트로피
서울	37	0.9507	43	0.9858	49	0.9997	38	0.9300	29	0.8687	16	0.6343
대구	33	0.9149	33	0.9149	45	0.9928	38	0.9300	29	0.8687	19	0.7015
광주	35	0.9341	29	0.8687	41	0.9765	40	0.9456	28	0.8555	16	0.6343
충주	27	0.8415	37	0.9507	53	0.9974	43	0.9654	29	0.8687	20	0.7219
부여	36	0.9427	34	0.9248	44	0.9896	38	0.9300	29	0.8687	15	0.6098
전국평균	33.0	0.930	30.7	0.905	40.4	0.969	36.0	0.925	24.8	0.829	16.1	0.656

## 2) 일기엔트로피 및 정보비의 분포 특성

1990-1999의 매일의 기상자료에서 운량과 강수량(1mm이상/일)에 주목하고 강수 유무의 2계급으로 분류한 일기자료를 이용한 장마기의 일기의 출현율과 엔트로피의 계산 결과를 이용하여 광주(이하,대구,충주,부여,서울 순)를 중심으로 남한의 72개 지점 각각에 대해 완전히 맞는 예보를 그대로 적용한 경우의 정보비를 구하고 그를 이용한 값에 의해 광주의 일기대표성을 논한다.

먼저 전 관측지점의 일기엔트로피의 장마기 분포상황 일부를 그림 3에 나타냈다. 그런데 일기엔트로피와 탁월날씨와는 역상관으로 되어있고, 일기엔트로피가 큰(작은) 곳일수록 base의 날씨가 변형되기 쉽고(어렵고) 예측이 어려운(쉬운) 지역이라 할 수 있으므로, 날씨는 지점별로는 서울과 충주의 7월중순이 가장 예측하기 힘들며, 부여의 9월중순의 날씨가 가장 한 모양으로 안정적이라 할 수 있다. 따라서 일기대표성은 가을장마기>여름장마기로 크고, 순별로는 9월중순에 부여를 중심으로 한 금강유역이 가장 크고 7월 중순에는 서울과 충주를 중심으로 한 한강유역이 가장 일기대표성이 작다.

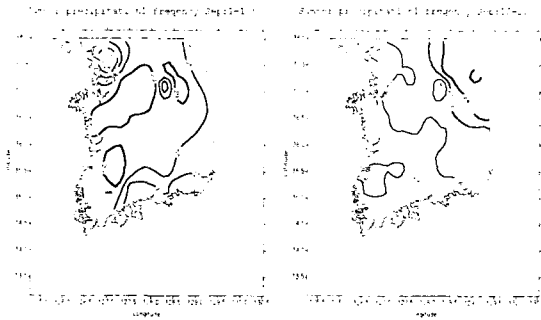


그림 2. 장마기 강수일수의 분포  
(단위 : 1mm이상/일)



그림 1. 관측지점의 분포와 지형

일반적으로 일기의 정보비는 거리,지형 등의 영향에 따른 일기대표성의 감쇠를 나타내는 測度로서 매우 유효하므로 일기의 정보비를 계산하면 2개의 예보중심지의 대표범위를 논할 수 있다. 광주기준지점에 대해 완전히 맞는 예보를 정보로 한 경우의 남한 내 정보비의 장마기 분포의 일부를 그림 4에 나타냈다. 여기서 기준지점 자신(원점)의 정보비 값은 1.00이다.

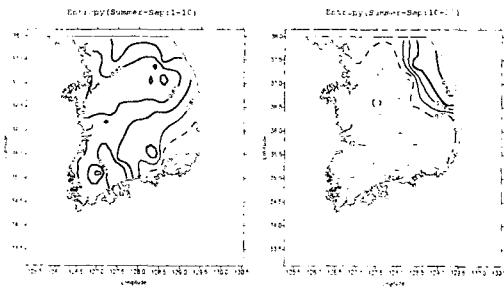


그림 3. 장마기 일기엔트로피의 분포  
(단위 : bits)

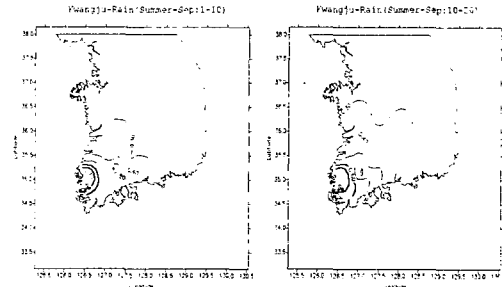


그림 4. 광주를 중심으로 한 장마기 정보비의 분포

광주 기준의 6월하순의 정보비 분포만 살펴보면, 무안(정보비 : 0.80, 기준점과의 거리:40.1km), 흑산도(0.61, 108.3km), 거창(0.57, 87.7km), 남해(0.55, 77.3km) 순으로 나타난다. 7월초순은 무안(0.55), 합천(0.50, 95.2km), 밀양(0.46, 127.4km), 마산(0.44, 110.4km) 순으로 나타나며, 7월중순은 거제(0.42, 116.9km), 순천(0.41, 26.5km), 무안(0.40), 부산(0.35, 141.2km)로 나타난다. 8월하순의 정보비 분포는 남해(0.72), 남원(0.60, 39.3km), 임실·문경(0.59, 56.6km, 181.5km) 순으로 나타나며, 9월초순은 남해(0.86), 남원(0.65), 무안(0.59), 목포(0.58, 51.6km) 순으로 나타나고, 9월중순은 해남·장흥·순천(0.55, 71.7km, 53.9km), 정읍(0.53, 44.5km) 순으로 나타난다.

전체적으로 보면 기준점별 정보비의 분포는 광주는 9월초순이 가장 크고 7월중순은 가장 작다. 대구는 6월하순이 크고, 8월하순은 작다. 충주는 9월중순이 크고 8월하순이 작으며, 부여는 9월중순이 크고 7월중순은 작다. 서울은 9월중순이 가장 크고 7월중순이 가장 작다.

정보비의 분포 값이 크다는 것은 기준지점의 일기 국지성이 약해 일기대표범위의 영역이 넓다는 것을 의미하므로, 대략 정보비 분포로 살펴본 일기대표성은 가을장마기 > 여름장마기이며, 광주·부여·서울의 7월중순은 일기대표성이 낮고, 서울·충주·부여의 9월중순은 일기대표성이 크다.

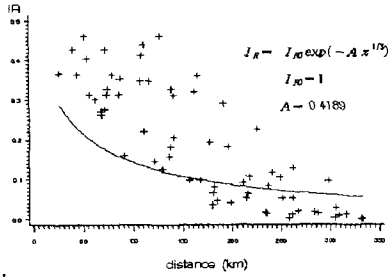
### Ⅲ. 일기대표성의 설정

#### 1) 정보비의 거리에 따른 감쇠특성

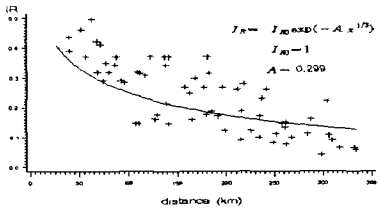
##### (1) 실험식의 적용

광주 기준의 장마기 정보비의 지리적 분포(그림 4)는 대체로 정보비의 분포형이 매우 단순하다. 즉 시기(순별)에 따라 지리적, 지형적 변동은 보이나 1차적인 요인이라 생각되는 거리에 따른 변동에 비교한다면 지리적, 지형적 변동은 2차적인 것을 알 수 있다.

그런데 정보비의 거리에 따른 감쇠는 실험식으로 표시될 수 있고 이는 일기대표성이 거리에 따라 감소하는 상황을 잘 나타내는 척도로서 유효하다. 따라서 광주 기준지점 자신과 그 외 남한의 72개 기상측후소와의 거리와 정보비의 관계를 실험식에 적용하여 얻은 결과의 일부를 그림 5에 나타냈다. 그림 5에서 점의 흩어짐은 주로 지형조건에 의한 변동을 표현한 것이라고 생각되나 거리에 의한 감소에 비교하면 문제가 안되고 있음을 보여주고 있다. 각 기준지점별로 장마기의 정보비와 거리와의 관계 즉, 반감거리와 감쇠계수 A를 살펴보면 표 3과 같다. 그런데 감쇠계수 A 값이 크면(작으면) 일기대표성이 작다(크다)라고 할 수 있으므로 전체적으로 보면 여름장마기, 9월중순과 부여의 일기대표성이 크고 가을장마기, 7월초순과 서울의 일기대표성이 가장 작다고 할 수 있다.



<광주중심 2상태 7월 초순>



<광주중심 2상태 9월 중순>

그림 5. 광주 기준지점의 정보비의 거리에 따른 감쇠경향 및 감쇠계수 A

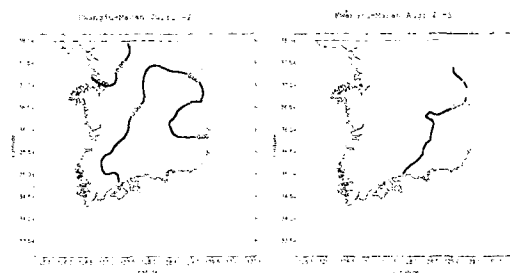
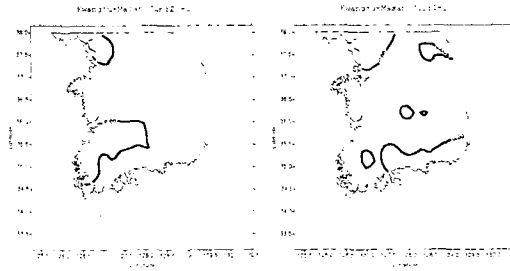


그림 6. 정보비차(광주-마산)에 의한 장마기의 일기대표 범위

## 2) 정보비의 반감거리

정보비의 반감거리는 지리·지형적 변동보다는 거리에 따른 변동이 우선적이거나 일기 대표성의 직관적인 척도로서 유효하게 활용할 수 있다.

전체로 보면 장마기의 반감거리는 모두 가을장마기>여름장마기로 나타나며, 순별로는 9월중순>6월하순>8월하순>9월초순>7월중순>7월초순 순이고, 지점별로는 부여>충주>광주>대구>서울 순이다.

따라서 일기대표성은 가을장마기>여름장마기이며, 순 및 지점별로는 각각 9월 중순과 부여를 중심으로 한 금강유역에서 일기대표성이 가장 크게 나타나며, 7월초순과 서울을 중심으로 한 한강유역은 가장 낮게 나타난다.

표 3. 5개 기준지점의 장마기(순별) 정보비의 반감거리 및 감쇠계수 A (단위 : km)

지점 \ 순	6월하순		7월초순		7월중순		8월하순		9월초순		9월중순	
	반감거리	감쇠계수	반감거리	감쇠계수	반감거리	감쇠계수	반감거리	감쇠계수	반감거리	감쇠계수	반감거리	감쇠계수
서울	101.29	0.3423	31.41	0.5058	56.81	0.4151	99.44	0.3444	103.20	0.3402	167.65	0.2894
대구	163.79	0.2917	77.96	0.3736	103.35	0.3401	118.72	0.3247	211.64	0.2678	191.40	0.2769
광주	143.16	0.3050	55.31	0.4188	82.61	0.3660	367.74	0.2227	157.51	0.2955	152.07	0.2990
충주	347.12	0.2271	65.72	0.3954	81.90	0.3675	132.04	0.3134	176.75	0.2843	296.99	0.2392
부여	182.73	0.2812	94.14	0.3508	89.13	0.3573	213.63	0.2669	185.78	0.2797	491.73	0.2022

## 3) 일기 대표성(예보구역)의 설정

광주 기준지점의 장마기 일기 정보비의 지역분포 변동은 예보중심지에서의 거리에 의한 효과가 일차적이며, 지리적인 또는 지형적인 효과는 2차적이라 하여도 인접하는 두 개의 관측지점에 관하여 일기의 대표성을 고려하여 그 경계를 논하는 경우에는 지리적, 지형적 효과가 의미있게 된다. 따라서 경계는 양 대표 관측지점에서 부여된 정보비 그것의 대소(정보비차)에 의해 정하는 것이 합리적인 방법이 된다.

정보비차에 의한 남한 내 5개 지점 기상측후소의 장마기의 순별 일기 대표범위(예보구역) 설정은 지리적위치, 지형효과 및 거리등을 고려하여 서울-부여, 서울-충주:충주-대구, 충주-부여:광주-부여, 광주-대구, 광주-마산:대구-마산, 대구-부여 등을 기준으로 한 경우 등 수많은 경우를 여러모로 나누어 보아야 한다. 여기서는 광주-마산의 것만을 살펴 보았다(그림 6).

광주-마산을 기준으로 한 경우의 정보비차에 의한 장마기의 일기대표 범위는 광주영역이 9월초순·중순과 8월하순일 때 마산영역보다 크게 나타난다. 6월하순과 7월하순을 제외하면 일기대표 범위는 소백산맥의 주향에 따른 지리, 지형적 효과가 의미있게 반영되고 있다.

#### IV. 결 론

본 연구에서는 각종 기후인자의 영향을 명확히 반영하여 한반도의 장마기 날씨 및 기 후특성을 잘 나타내는 탁월일기 및 강수현상의 장마기 출현다소와 순별변화에 대해 정보이론을 이용하여 일기엔트로피와 정보비를 구해 그의 시, 공간적 특성을 살피고, 4대강 유역(서울, 대구, 광주, 충주, 부여)의 일기 대표성을 정보비의 거리에 따른 감쇠특성과 정보비의 차에 의해 규명하고 그에 따른 예보구역 설정을 하였다.

본 연구의 결과를 요약하면 다음과 같다.

(1)장마기의 순별추이에 따른 강수일수는 여름장마기>가을장마기로 순별·지역차가 크며, 순별, 지점별로는 각각 7월중순과 서울이 강수일수가 가장 많고, 9월중순과 광주가 적다.

(2)강수강도는 서울을 제외하면 여름장마기>가을장마기이며, 순별·지점별로는 6월하순과 서울이 강수일수가 각각 가장 크며 9월초순과 대구는 가장 작다.

(3)일기엔트로피는 여름장마기>가을장마기로 나타나며 순별, 지점별로는 7월중순과 서울이 각각 일기엔트로피가 가장 크며, 9월중순과 광주는 가장 작다. 따라서 일기엔트로피의 특성에 따라 일기대표성은 가을장마기>여름장마기로 크며, 9월중순과 부여중심의 금강유역이 가장 크고, 7월중순의 서울과 충주중심의 한강유역에서 가장 작다.

(4)정보비의 분포는 가을장마기>여름장마기이며, 서울·충주·부여의 9월중순의 정보비의 분포가 가장 크고, 광주·부여·서울의 7월중순은 가장 작다. 따라서 전자는 일기의 국지성이 약해 일기대표성이 크고, 후자는 일기대표성이 작다.

(5)정보비의 감쇠계수는 가을장마기>여름장마기이며, 9월중순과 부여의 그 값이 가장 커서 일기대표성이 크고, 7월초순과 서울의 감쇠계수는 그 값이 가장 작아 일기대표성도 작다.

(6)정보비의 반감거리는 여름장마기>가을장마기이며, 7월초순과 서울의 반감거리가 가장 커서 일기대표성도 크며, 9월중순과 부여의 그 값이 가장 작아 일기대표성도 작다.

도시 녹지가 주변 시가지 기온에 미치는 영향