해수면온도와 우리나라 월강우량과의 관계분석에 관한 연구

A Study on the Analysis of the Relationship between Sea Surface Temperature and Monthly Rainfall

오 태 석* / 문 영 일**

Oh, Tae Suk / Moon, Young-Il

Abstract

Rainfall events in the hydrologic circulation are closely related with various meteorological factors. Therefore, in this research, correlation relationship was analyzed between sea surface temperature of typical meteorological factor and monthly rainfall on Korean peninsula. The cluster analysis was performed monthly average rainfall data, longitude and latitude observed by rainfall observatory in Korea. Results from cluster analysis using monthly rainfall data in South Korea were divided into 4 regions. The principal components of monthly rainfall data were extracted from rainfall stations separated cluster regions. A correlation analysis was performed with extracted principal components and sea surface temperatures. At the results of correlation analysis, positive correlation coefficients were larger than negative correlation coefficients. In addition, The 3 month of principal components on monthly rainfall predicted by locally weighted polynomial regression using observed data of sea surface temperature where biggest correlation coefficients have. The result of forecasting through the locally weighted polynomial regression was revealed differences in accuracy. But, this methods in the research can be analyzed for forecasting about monthly rainfall data. Therefore, continuous research need through hydrological meteorological factors like a sea surface temperature about forecasting of the rainfall events.

Keywords : correlation analysis, monthly rainfall, sea surface temperature, rainfall forecast

요 지

수문학적 물순환과정에서 강우는 여러 기상학적 인자들과 밀접한 관련을 갖고며 발생한다. 따라서 본 연구에서는 대표적인 수문기상인자인 해수면온도와 한반도에 발생하는 월강우량 사이의 관계에 대하여 분석하였다. 우리나라의 61개 지점의 월평균 강우량과 위도 및 경도 자료를 이용하여 군집분석을 수행하였다. 군집 분석 결과에서 우리나라의 월강우자료를 이용하여 크게 4개의 군집으로 구분할 수 있었다. 군집별로 구분된 강우관측소의 월강우량 자료들을 주성분을 추출하였다. 추출된 주성분과 해수면온도와의 상관성 분석을 수행하였다. 상관성 분석 결과에서 양(+)의 상관관계가 음(-)의 상관관계보다 더 크게 나타났다. 또한, 상관관계가 가장 큰 지점의 해수면온도를 이용하여 3개월의 월강우량을 지역가중다항식을 통해 예측하였다. 지역가중 다항식을 통한 예측 결과는 군집에 따라 정확성에 차이는 있으나, 정량적인 예측이 가능한 것으로 판단되었다. 따라서 해수면온도와 같은 수문기상인자를 통한 강우량의 예측에 대한 지속적인 연구가 필요하다.

핵심용어 : 상관분석, 월강우량, 해수면온도, 강우 예측

* 서울시립대학교 공과대학 토목공학과 박사후과정

The University of Seoul, Siripdae-gil 13, Dongdaemun-gu, Seoul 130-743, Korea (e-mail: waterboy@uos.ac.kr)

** 교신저자, 서울시립대학교 공과대학 토목공학과 교수

Corresponding Author, Department of Civil Eng., Professsor, The University of Seoul, Siripdae-gil 13, Dongdaemun-gu, Seoul 130-743, Korea (e-mail: ymoon@uos.ac.kr)

1. 서론

수문학적 물순환과정에서 강우는 수문해석의 기본적인 인자 중의 하나로 해석된다. 물순환 과정에서 강우는 물순환을 유발시키는 인자로서 여러 수문기상인자와 밀접한 관련을 갖으며 발생하는 것으로 알려져 있다. 그러나 강우의 발생형태와 빈도는 현재로서 명확히 규명하기는 매우 어려운 실정이다. 이는 강우의 발생형태와 발생빈도가 여러 기상인자들과 매우 복잡한 관계를 맺으며 비정상적인 형태를 보이면서 발생하고 있기 때문이다. 여러 기상인자 중에서 해수면온도(Sea Surface Temperature, SST)는 강우와 밀접한 관련이 있는 매우 중요한 요소로 인식되고 있으며 최근에는 강우의 빈도 및 강우량과 저빈도(low frequency) 등을 밝히기 위한 상관성 분석을 통해서 시·공간적인 분석을 시도하고 있는 추세이다.

Klein and Bloom(1987), Kiladis and Diaz(1989), Cayan and Peterson(1989) 및 Kim et al. (2006)는 북미의 저빈도 분석을 수행하여 해양과 대기 성분이 강우사상에 많은 영향을 끼치고 있음을 밝혔다. Moon and Lall(1996)은 기후변동과 수문학적 요인의 동역학적 관계를 분석하기 위하여 미국의 Great Salt Lake와 남방진동지수(Southern Oscillation Index, SOI)를 포함한 여러 대기순환지수를 이용한 저빈도 분석을 수행하였다. Uvo et al. (1998), Harshburger et al. (2002), Rautenbach and Smith(2001), 및 Misra (2003)은 태평양 등의 해수면온도 변화와 강우량과의 상관관계에 대한 분석을 수행하였다. Higgins et al. (2004) 및 Barnston and Smith(1996)은 해수면온도 변화에 따른 계절 강우량에 대한 예측을 여러 통계적 기법을 통하여 수행하였다. 우리나라의 강우사상과 수문기상인자에 대한 연구는 유철상 등(2000)은 한반도 근해의 해수면온도와 월강수량 및 월평균 기온에 대한 상관관계를 분석하였다. 문영일 등(2005)은 우리나라의 계절강우량을 Wavelet Transform 분석을 통해 저빈도 주기성분을 추출하여 독립성분과 주성분 분석을 통해 추출한 자료와 태평양 해수면온도와의 상관성 분석을 수행하였다. 오태석 등(2007, 2008)은 월강수량과 해수면온도

와의 상관성 분석을 여러 통계기법을 이용하여 수행하였다. 추현재 등(2007)은 월해수면온도의 표준화를 통해 연주기를 제거하여 5개월 이동평균한 자료와 계절강우량과의 상관성 분석을 수행하였다.

이와 같이, 강우사상은 여러 수문인자 중에서 특히, 해수면온도와 밀접한 관련을 갖고 있는 것으로 나타났다. 선행연구사례에서 해수면온도와 강우량을 이용한 분석은 대부분 원격지체상관분석이 주를 이루고 있는 것으로 나타났다. 또한, 일부 연구결과에서는 해수면온도를 통해 계절강우량의 예측을 수행하였다. 예측방법은 선형적인 회귀분석을 이용하는 방법을 이용하고 있는 것으로 나타났다. 따라서 본 연구에서는 비모수적(nonparametric) 회귀분석의 일종으로 비선형자료의 예측에 많이 이용되고 있는 지역가중다항식(Locally weighted polynomial regression)을 이용하여 월강우량의 주성분을 예측하였다.

본 연구는 월강우량과 해수면온도의 상관성 분석을 수행하였다. 월강우량은 우리나라의 기상청에서 관할하는 강우자료를 이용하여 30년 이상 관측된 자료를 이용하였으며, 해수면온도는 ERSST(Extended Reconstructed SST) Ver.3의 범지구적 자료를 이용하였다. 분석방법은 월강우량의 주성분과 해수면온도의 원격지체상관계수를 산정하여 우리나라의 강우특성과 상관성이 큰 지역을 설정하였으며, 해수면온도를 이용하여 월강우량의 주성분의 예측을 수행하였다.

2. 분석대상 자료

분석대상 자료로써 강우자료는 우리나라의 기상청에서 관할하는 강우관측소의 일강우 자료로부터 월강우량을 추출하였으며, 해수면온도는 NOAA(National Oceanic and Atmospheric Administration)에서 제공하고 있는 ERSST Ver. 3을 이용하였다.

2.1 강우 자료

기상청에서 관할하고 있는 76개 강우관측소 중에서 30년 이상 장기간의 관측자료를 보유하고 있는 61개 지점에

Fig. 1. Monthly Rainfall Data at Seoul Station

서 관측된 월강우량 자료를 이용하였다. 월강우자료를 이용하여 월강우량 자료를 구성하였으며, 관측기간은 분석 대상자료가 모두 동일하게 1973년부터 2008년까지 36개년의 자료를 이용하였다. Fig. 1은 서울지점에서 관측된 월강우량 자료를 1973년부터 2008년까지 도시한 결과이다.

2.2 해수면온도 자료

해수면 온도는 NOAA의 Satellite and Information Service에서 관측한 ERSST의 자료 중에서 60°S~60°N, 180°W~180°E 범위 내에서 관측된 월해수면 온도 자료를 이용하였다. Smith and Reynolds (2003, 2004)에 따르면, 해수면온도는 1854년 1월부터의 관측자료를 제시하고 있으나 1880년 이전의 관측자료는 매우 부족하여 전지구적 변동을 거의 반영하지 못한 것으로 알려져 있으며, 해수면온도의 재구성 자료는 ERSST Ver. 2가 사용가능하도록 제공되고 있다. ERSST Ver. 1에서 재구성한 자료에 비하여 ERSST Ver. 2는 공간적인 모형을 구성하여 해수면온도 관측 자료를 재구성한 것으로 관측 자료의 변동에 취약한 지역의 자료를 실제에 보다 근사하게 향상시킨 것으로 평가받고 있다. 최근에는 ERSST Ver. 3가 제공되었다 (Smith et al., 2008). ERSST Ver. 3는 ERSST Ver. 2의 모의 결과를 검증을 통해 향상시켰다. 가장 큰 차이점은 1930년대 이전 자료에 대하여 최적화된 매개변수를 통해 해수면온도의 아노말리 (anomaly) 자료가 갖는 진폭을 감소시켰다. 또한, 15년 주기의 저빈도를 갖는 자료의 필터링을 수행하여 현재 자료에 맞게 재구성을 시도하였다. 따라서 본 연구에서는 가장 최근에 재구성되어 제공되어지는 ERSST Ver. 3 자료를 활용하였다. Fig. 2는 해수면온도의 관측범위를 도시한 결과이다. Fig. 2에서 나타낸 ERSST 자료에서 해수면 온도의 관측 격자는 2°×2°이므로 총 격자의 개수는 10,980개이며, 이중에서 육지 부분인 2,699개 격자를 제외한 8,281개 지점에서 관측한 1973년부터 2008년까지의 월해수면 온도 자료를 이용하였다.

3. 월강우량을 이용한 군집분석과 주성분 분석

분석대상자료인 61개 지점에서 관측된 월강우량 자료를 이용하여 군집분석을 수행하였다. 군집분석은 지점별로 월평균강우량을 산정하여 Ward's minimum variance를 통해 개체간의 거리를 산정하여 수행하였다. 군집분석을 통해 강우발생특성이 비슷한 지역으로 구분하여, 지역별로 주성분을 추출하였다.

3.1 월강우자료를 이용한 군집분석

모집단에 소속된 많은 개체들의 특성인자를 통한 분류를 수행하기 위해서 사용되는 다변량 분석 기법이 군집분석이다. 군집분석을 통해 개체를 분류하기 위해서는 Euclidean 거리를 계산하고 거리가 가까운 개체끼리 묶어야 한다. Euclidean 거리는 두 개체 사이의 유사 정도를 거리로 표현할 수 있다. 다음 Eq. (1)은 r 번째 개체와 s 번째 개체의 Euclidean 거리를 나타내며, $\underline{x}_r$, $\underline{x}_s$ 는 각 개체의 위치를 나타내는 변수이다.

$$d_{rs} = \left[(\underline{x}_r - \underline{x}_s)' (\underline{x}_r - \underline{x}_s) \right]^{\frac{1}{2}} \quad (1)$$

군집분석의 절차는 다음과 같다. 첫 번째로 처음에는 개체의 수만큼에 해당하는 군집이 존재하게 된다. 두 번째로 유사성이 가장 가까운 개체를 군집으로 묶게 된다. 세 번째로 개체가 군집으로 묶이게 되면 개체와 새로 만들어진 군집과의 유사성을 계산하여 새로운 군집을 형성하게 된다. 이 때, 군집과 군집 (혹은 개체)의 유사성 (거리)을 측정하는 방법은 Nearest neighbor, Furthest neighbor, Centroid neighbor, Average neighbor, Ward's minimum variance로 총 5가지 기법이 있다. Nearest neighbor 기법은 두 군집의 각 개체 중 가장 가까이 있는 개체의 거리를 이용하며, Furthest neighbor 기법은 두 군집의 각 개체 중 가장 멀리 있는 개체의 거리를 이용해 군집을 구분한다. Centroid neighbor 기법은 군집간의 평균거리를,

Fig. 2. Observed Range of Sea Surface Temperatures

Average neighbor 기법은 한 군집의 개체와 다른 군집 개체들의 각 평균거리를, Ward's minimum variance 기법은 군집의 평균간 거리를 각 군집의 개체 개수의 역의 합으로 나눈 제곱근을 구한 거리를 통해 구분하게 된다. 따라서 본 연구에서는 Ward's minimum variance를 이용하여 군집을 구분하였다. Ward's minimum variance를 이용한 방법은 우리나라의 지역빈도해석을 수행하기 위해서 군집 분석을 적용할 때에 많이 이용된 방법으로 우리나라의 강우자료를 통한 지역구분에 많이 활용되고 있는 기법이다.

군집의 개수를 결정하기 위해서는 그래프적 방법으로 계층적 나무 다이어그램 (tree diagram)과 검정 통계량을 이용하는 방법으로는 Hotelling's T^2 검정이 있다. Hotelling's T^2 검정 통계량은 두 집단 다변량 평균의 차이를 보는 통계량이다. Hotelling's T^2 는 "변수들의 평균이 같다"라는 귀무가설을 세우고 가설 검정에서 포함된 변수들의 평균이 다른지에 대한 검정을 수행하는 방법이다 (Hotelling, 1931). 따라서 개체의 군집간 평균 차이가 유의하지 않으면 두 군집을 합치고 유의하면 군집 그대로 유지하는 방법이다.

(a) Tree Diagram

(b) Result of Hotelling's T^2 Test

Fig. 3. Decision of the Cluster Number by Cluster Analysis using Ward's Minimum Variance

Table 1. Result of Cluster Analysis

NCL	Clusters Joined		Ferq.	Distance	PST2
60	143	281	2	0.0002	.
59	114	221	2	0.0003	.
58	146	238	2	0.0003	.
⋮	⋮	⋮	⋮	⋮	⋮
5	CL12	115	4	0.074	9.5
4	CL8	CL9	14	0.0751	9.5
3	CL16	CL6	43	0.0752	16.5
2	CL5	CL4	18	0.1668	10.3
1	CL2	CL3	61	0.3114	26.7

는 적정 군집수를 4개로 선정하였다. Table 2는 각 군집에 포함되는 관측소를 정리한 결과이다. 군집 1에 포함되는 지역은 대부분 한반도의 중부와 남부 지역에서 관측되는 강우관측소가 포함되는 것으로 나타났으며, 군집 2는 중북부 지역, 군집 3은 남해안 지역, 군집 4는 강원 북부 지역의 관측소가 포함되는 것으로 분석되었다.

3.2 군집별 주성분 분석

주성분분석은 차원축소를 통하여 저차원상에서 변수의 관계를 규명하는 다변량 자료 분석기법이라 할 수 있다. 주성분분석은 다변량기법에서 폭넓게 사용되는 기법중 하나이며 주성분분석은 평균과 분산까지의 통계적 성질을 이용한 2차 통계적 기법이다. 주성분분석에 해석과정을 간단히 설명하면 다음과 같다.

다변량 X 가 $n \times m$ 의 행렬이라고 생각해보자(여기서, n 은 관측된 자료의 수, m 은 변수의 수를 의미한다). 주성분분석은 X 를 $n \times p$ 차원의 행렬 T 와 $m \times p$ 차원의 행렬 S 와 $n \times m$ 의 잔차(residual) 행렬로 분해할 수 있는 방법이다.

$$X = TS^T + E \quad (2)$$

여기서, p 는 새로 구성된 차원의 수이며, $p < m$ 의 크기를 갖는다. 가장 최적으로 분할하는 조건은 주어진 요소에 대하여 잔차 행렬의 Euclidean 거리를 최소화하는 것이다. 이 기준을 만족시키려면 S 행렬의 행(column)은

y 의 공분산(covariance) 행렬의 고유값(eigenvalue) 중에서 큰 순서대로 p 만큼 택하여 이에 해당하는 고유벡터(eigenvector)로 구성하면 된다. 여기서, p 는 주성분분석에서 나타내는 고유벡터의 수가 된다. 고유값은 주성분으로부터 데이터를 다시 복원할 때 해당하는 자료에 대한 가중치의 역할을 하게 된다. 이러한 관점에서 주성분분석은 차원 m 을 차원 p 로 줄이는 동시에 선형으로 재구성하는 것이므로 $S^T S = I$ 가 되고 Eq. (2)는 다음 Eq. (3)과 같은 형태가 된다.

$$\bar{T} = \bar{X} S \quad (3)$$

여기서, 벡터 $\bar{X}$ 는 행렬 X 의 열성분이며 단일 자료의 벡터이며, 벡터 $\bar{T}$ 는 행렬 T 혹은 특성 공간 $\bar{X}$ 에서 해당하는 열성분을 의미한다. 행렬 S 는 선형 변환의 계수를 의미한다.

군집분석을 통해 구분된 4개 지역에 포함된 월강우량의 시계열 자료를 이용하여 주성분 분석을 수행하였다. 주성분 분석은 n 개의 원변수를 선형결합의 주성분변수를 이용하여 원변수의 공분산 구조를 설명하는 방법이다. 따라서 4개 군집에 대하여 각각의 공분산을 계산하여 군집별로 주성분을 추출하였다. Table 3은 각 군집별로 계산된 고유치와 총변동설명 비율을 정리하여 나타낸 결과이다. Table 3에서 C. P.는 누적비(Cumulative Proportion)으로 총변동에 대한 주성분의 누가설명비율을 의미한다.

Table 2. Result of Clustering Analysis

Cluster	Rainfall observatory
1 (32 stations)	Chungju, Seosan, Uljin, Cheongju, Daejeon, Chupungryeong, Pohang, Gunsan, Daegu, Jeonju, Ulsan, Gwangju, Mokpo, Boeun, Cheonan, Boryeong, Buyeo, Geumsan, Buan, Imsil, Jeongeup, Namwon, Haenam, Yeongju, Munkyeong, Yeongdeok, Uiseong, Gumi, Yeongcheon, Geochang, Habcheon, Milyang
2 (11 stations)	Chuncheon, Seoul, Incheon, Wonju, Suwon, Gangwha, Yangpyeong, Icheon, Inje, Hongcheon, Jecheon
3 (14 stations)	Busan, Tongyeong, Yeosu, Wando, Jeju, Seongsan, Seogwipo, Jinju, Suncheon, Jangheung, Goheong, Sancheong, Geoje, Namhae
4 (4 stations)	Sokcho, Daegwanryeong, Gangreung, Ulreungdo

Table 3. Eigenvalue and Cumulative Proportion of Covariance Matrix by Principal Component Analysis

PCN	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	Eigenvalue	C. P.	Eigenvalue	C. P.	Eigenvalue	C. P.	Eigenvalue	C. P.
1	333181.1	0.834	192214.5	0.920	210681.0	0.838	53490.9	0.876
2	19773.0	0.883	5791.4	0.948	14416.4	0.895	3760.9	0.937
3	10419.9	0.909	3127.1	0.963	9683.2	0.934	2520.8	0.979
4	7064.4	0.927	2245.8	0.973	4125.1	0.950	1304.3	1.000
5	4395.1	0.938	1539.9	0.981	2747.7	0.961	-	-

PCN은 주성분의 번호이다.

Table 3에서 나타난 바와 같이, 주성분 분석 결과에서 4개의 군집 모두 첫 번째 주성분만으로 80% 이상의 변동을 설명하고 있는 것으로 나타났다. 이는 첫 번째 주성분의 설명력이 대상 자료의 전체 분산의 80% 이상을 포함하고 있음을 의미한다. 즉, 첫 번째 주성분만으로도 군집별 강우자료가 갖는 특성을 표현하기에 충분한 것으로 판단된다. 따라서 각 군집별로 추출된 첫 번째 주성분만을 이용하여 향후 분석을 수행하였다. 계산된 첫 번째 주성분 자료는 우리나라의 각 군집별 월강우량 자료의 경년별 변화를 표현하고 있는 것으로 나타났다. 두 번째 주성분은 변수들의 월별 변화를 보이고 있으며, 세 번째 이상의 주성분은 앞의 주성분에서 설명되지 못하는 분산의 조합인 잡음에 가까운 특성을 보여주고 있다.

4. 강우자료와 해수면온도의 원격지체상관분석

군집분석과 주성분분석을 이용하여 우리나라의 61개 지점에서 관측되는 월강우량을 4개의 지역으로 구분하여 4개의 주성분을 추출하였다. 추출된 4개의 주성분 자료는 1973년부터 2008년까지의 월강우량 자료를 내포하고 있다. 따라서 추출된 주성분 자료와 해수면온도와 원격지체상관분석을 수행하였다. 상관계수의 산정을 위하여 선형적인 상관관계를 나타낼 수 있는 Pearson 상관계수를 범지구적 해수면온도 자료와 월강우량의 주성분을 이용하여 산정하였다. 또한, 월강우량 자료에 0 (zero)개월에서

11개월까지 지체를 주어 상관계수를 산정하였다. 산정된 상관계수는 유의수준 5%에서 t-분포를 통해 검정하였다. 본 연구에서 분석한 대상자료는 시계열별로 432개이며, 일반적으로 0.1보다 큰 상관계수를 갖는 경우에 통계적으로 유의한 의미를 갖는 것으로 나타났다(김우철 등, 2008).

상관성 분석 결과를 Table 4에 나타내었다. Table 4에서 지체시간별로 산정된 상관계수를 정리하였다. 분석대상이 되는 해수면온도의 관측지점이 약 8,000개이므로 5%에 해당하는 400개 지점의 상관계수를 이용하여 통계량을 나타내었다. 상관분석을 통해 월강우량의 주성분과 상관관계가 큰 해수면온도 관측지역을 설정하기 위해서는 다음과 같은 조건을 만족하여야 한다. 첫 번째로 산정된 상관계수의 크기가 통계적으로 유의한 상관계수의 값보다 큰 값을 가져야 한다는 것이다. 이는 모든 자료에서 가장 큰 상관계수가 통계적으로 유의한 값을 계산하였다. 두 번째로 적절한 지체시간(월)을 계산하기 위해서 지체시간별로 다르게 산정되는 상관계수가 보다 큰 값이 계산되어야 한다. Table 4에서 나타난 바와 같이 양(+)의 상관계수가 음(-)의 상관계수보다 더 크게 산정되었으며, 최대값을 갖는 상관계수의 지체시간은 약간씩 상이한 것으로 나타났다. 지체시간이 5개월 이상인 경우에는 지체시간이 다소 길고 상관성이 크게 나타나는 지역의 편차가 심한 것으로 나타나 별도로 표에 제시하지는 아니하였다. 따라서 Table 4에 가장 큰 상관계수의 평균값과 최대값을 갖는 지체시간에 음영을 주어 표기하였다. 세 번째로 산정된 상관계수가 크게 산정된 지점들이 어느 정도 같은

Table 4. Results of Correlation Analysis

Classification			0	1	2	3	4
P01	Positive C.C.	Mean	0.65	0.47	0.53	0.64	0.68
		Max.	0.71	0.55	0.64	0.71	0.70
	Negative C.C.	Mean	-0.60	-0.37	-0.29	-0.52	-0.65
		Min.	-0.72	-0.54	-0.43	-0.60	-0.68
P02	Positive C.C.	Mean	0.63	0.45	0.51	0.61	0.64
		Max.	0.69	0.54	0.63	0.70	0.66
	Negative C.C.	Mean	-0.58	-0.37	-0.25	-0.47	-0.60
		Min.	-0.68	-0.52	-0.38	-0.54	-0.63
P03	Positive C.C.	Mean	0.57	0.43	0.52	0.63	0.67
		Max.	0.64	0.55	0.60	0.69	0.69
	Negative C.C.	Mean	-0.53	-0.27	-0.38	-0.57	-0.66
		Min.	-0.66	-0.45	-0.49	-0.63	-0.69
P04	Positive C.C.	Mean	0.63	0.49	0.43	0.53	0.58
		Max.	0.65	0.57	0.56	0.59	0.62
	Negative C.C.	Mean	-0.57	-0.44	-0.20	-0.35	-0.50
		Min.	-0.62	-0.57	-0.33	-0.43	-0.53

(a) Lag Time Zero (0) Month

(b) Lag Time 3 Months

Fig. 4. Results of C.C. Analysis Compare to the Lag Time using P1 Data

(a) Lag Time 3 Months

(b) Lag Time 4 Months

Fig. 5. Results of C.C. Analysis Compare to the Lag Time using P3 Data

Fig. 6. Correlation Analysis Result of P1 (Lag 3 Month)

Fig. 7. Correlation Analysis Result of P2 (Lag 3 Month)

Fig. 8. Correlation Analysis Result of P4 (Lag 3 Month)

Fig. 9. Correlation Analysis Result of P3 (Lag 3 Month)

동일한 범위의 해수면온도 관측 지역에 포함되어야 한다는 점이다. Figs. 4~5는 상관계수가 가장 크게 나타난 400개 지점을 도상에 표현한 결과이다. 군집 1은 지체시간 0개월과 3개월을 비교하였으며, 군집 3은 지체시간 3개월과 4개월을 비교하여 나타내었다. 분석 결과에서 지체시간이 3개월인 경우에 큰 값을 갖는 상관계수의 해수면온도 지점이 보다 잘 모여 있는 것을 확인할 수 있다.

Figs. 6~9는 월강우량의 주성분과 해수면온도의 상관성 분석을 수행한 결과를 도시하여 나타낸 것이다. 상관성 분석 결과에서 지체시간 3개월일 때에 산정된 양(+) 상관계수가 가장 유의한 것으로 나타났다. Figs. 6~9의 P1, P2, P3 및 P4는 군집 분석 결과에서 각각의 군집번호를 의미한다.

Figs. 6~9에서 나타난 바와 같이, 각 지역별로 3개월 지체시간을 보일 때에 산정된 상관계수가 가장 유의한 것

으로 나타났다. 상관계수가 가장 크게 나타난 지역을 군집에 따라서 약간씩 차이를 보이는 것으로 계산되었다. 군집 1의 주성분은 북위 10°~남위 5°, 동경 55°~동경 80°에서 관측된 해수면온도와 가장 유의한 상관관계를 보이는 것으로 나타났다. 군집 2는 북위 20°~0°, 동경 60°~동경 90°, 군집 3은 남위 5°~남위 20°, 서경 5°~동경 35°, 군집 4는 북위 15°~남위 5°, 동경 60°~동경 90° 지역에서 관측된 해수면온도와 가장 상관관계가 큰 것으로 나타났다. 군집 1, 2, 4의 주성분은 아라비아해와 뱅골만 주변의 해수면온도와 원격지체상관성이 큰 것으로 나타났으며, 군집 3의 주성분은 기니만 인근의 해수면온도와 원격지체상관성이 크게 나타났다. 강우량은 여러 수문 기상·기후학적 인자들과 복잡하고 밀접한 관련을 갖으며 발생하게 된다. 따라서 해수면온도와 월강우량의 주성분 사이의 원격지체상관성 분석 결과에서 한반도와 다소 먼 지역의 해수면온도가 보다 밀접한 상관성이 있는 것으로 나타났다. 이는 전지구적인 해류와 대기의 순환과정에서 강우량과의 상호간에 주고받는 복잡한 과정에 기인한 것으로 판단되며, 몬순기후의 영향으로 인해 인도양의 해수면온도가 높이면 한반도의 강우량이 증가하는 것으로 판단된다. 그러나 현재까지는 명확한 인과관계에 대한 파악이 이루어지지 않는 것이었다. 따라서 이러한 과정에 대한 보다 정밀한 분석이 필요한 것으로 판단된다.

5. 지역가중 다항식을 이용한 월강우량의 예측

상관성이 큰 해수면온도를 이용하여 우리나라의 월강우량의 단기예측을 수행하였다. 일반적으로 강우예측은 관측된 월강우량 자료를 이용하여 ARIMA와 같은 모형을 통해 예측을 수행하기도 한다. 본 연구에서는 해수면온도를 이용하여 월강우량을 예측하기 위해서 비모수적(nonparametric) 회귀관계를 이용하였다. 일반적인 회귀분석에서 선형적인 회귀관계를 통해 해수면온도에 따라 월강우량을 예측할 수 있으나, 해수면온도와 월강우량 사이의 관계가 선형적이지 않으므로 지역가중 다항식(Locally weighted polynomial regression)을 이용하였다.

지역가중다항식을 이용한 회귀분석은 독립변수 x 와 종속변수 y 사이에 비선형적인 관계가 있을 경우에는 변수 변환(prior transform)을 통해 원변수를 다항식 형태로 바꾸어 회귀식을 유도하게 된다. 지역가중다항식은 Loader (1999)에 의해 가중이동평균을 이용한 비모수적(nonparametric) 회귀분석(kernel regression)의 일반화가 이루어졌다. 비모수적(nonparametric) 회귀분석은 Cleveland and Devlin (1988)에 의해 제시된 방법을 이용하였으며, 일반적인 이론은 Wand and Jones (1995)에 설명되어 있

다. 일반적으로 지역가중다항식을 이용한 회귀분석은 k-nearest neighbor 기법에 의해 추정된다.

기본원리는 함수 $f(\cdot)$ 를 추정에서부터 Eq. (4)의 일반적인 매개변수적 회귀분석과 차이점을 가지고 있다. k-nearest neighborhood에 의해 $y_i = x_i^*$ ($i = 1, 2, \dots, n$)에 따라 x 를 추정할 수 있다. x_i^* 의 $i = 1, 2, \dots, n$ 는 원자료 x^* 를 오름차순으로 정렬한 것이다. 여기서 추정된 함수 f 는 newton-taylor 함수식으로 정리할 수 있다.

$$\begin{aligned} f(x) &= f[x_1^*] + f[x_1^*, x_2^*](x - x_1^*) + \dots \\ &+ f[x_1^*, x_2^*, \dots, x_k^*](x - x_1^*) \cdot (x - x_2^*) \dots (x - x_{k-1}^*) \\ &+ f^{(k)}(\zeta)(x - x_1^*) \dots (x - x_{k-1}^*) \cdot (x - x_k^*)/k! \\ &= p_k(x) + f^{(k)}(\zeta)(x - x_1^*) \dots (x - x_k^*)/k! \end{aligned} \quad (4)$$

여기서, x 는 추정하고자 하는 추정치이며 ζ 는 $(x_1^*, x_2^*, \dots, x_k^*)$ 사이에서 내삽한 값을 의미한다. $f[x_1^*, x_2^*, \dots, x_k^*]$ 는 관측 자료인 $(x_1^*, x_2^*, \dots, x_k^*)$ 에 맞는 $f(\cdot)$ 인 $f(x_1)$, $f(x_2)$, ..., $f(x_j)$ 의 $j-1$ 차원의 다항식을 의미한다. ζ 는 일반적으로 미지 값이나, 평활화 과정을 통해 추정할 수 있다. 또한, 관측된 개체의 개수가 적다면, $f(\cdot)$ 는 평활화 과정을 통한 국부적인 C^k 평활자(smooth)로 추정할 수 있으며, 관측치의 최댓값보다 큰 값이 요구될 때에는 k-차원의 다항식을 통해 $p_k(x)$ 는 $f(x)$ 를 통해 추정할 수 있다.

지역가중다항식(locally weighted polynomial regression)을 이용하여 행렬 $(n \times M)$ 을 갖는 X 로 설명할 수 있는 x_l^* , $l = 1, 2, \dots, m$ 을 통해 행렬 $(n \times 1)$ 을 갖는 응답 벡터인 y 를 추정하기 위해서는 아래의 Eq. (5)를 만족하여야 한다.

$$\min_{\beta_l} (y_l - Z_l \beta_l)^T W_l (y_l - Z_l \beta_l) \quad (5)$$

여기서, 아래첨자 l 은 평가된 x_l^* 로부터 결합된 요소를 의미한다. β_l 은 Z_l 로부터 유도된 계수들(coefficients)을 나타내며, Z_l 은 p -차원을 갖는 행렬 X 에 의한 다항식의 확장을 통해 유도된다. W_l 은 $k \times k$ 차원을 갖는 대각항의 가중 행렬로써 다음과 같은 요소를 내재하고 있다. $w_{k,l} = K(u_{k,l}) / \sum_{j=1}^k K(u_{j,l})$ 이며, $u_{k,l} = d_{k,l} / d_{k,l}$ 이다. $d_{k,l}$ 은 행렬에서 사용되는 x_i^* 로부터 x_j^* 까지의 거리를 나타낸다. 그리고 $K(\cdot)$ 는 가중함수(weight function)이다.

지역가중다항식을 통해 해수면온도를 이용하여 월강우량의 예측을 수행하였다. 군집 1의 주성분을 예측하기 위해서 남위 2°, 동경 62°에서 관측된 해수면온도를 이용하였으며, 군집 2는 적도 0°, 동경 62°, 군집 3은 남위 8°, 서경 4°, 군집 4는 적도 0°, 동경 62° 지점에서 관측된 해수면온도를 이용하였다. Table 5는 지역가중다항식의 입력 자료로 활용된 값을 정리하여 나타난 결과이다. Table 5

Table 5. Input Data for Locally Weighted Polynomial Regression

Year	Month	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
		Rainfall	SST	Rainfall	SST	Rainfall	SST	Rainfall	SST
1973	4	137.07	28.55	7.05	28.55	280.54	26.21	-107.14	28.51
1973	5	113.84	28.73	-56.61	28.73	568.57	27.04	0.69	28.64
1973	6	-52.33	29.28	61.96	29.28	-220.12	27.58	61.40	29.25
1973	7	142.45	29.92	77.83	29.92	82.40	27.97	-63.22	29.97
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2007	10	-312.52	28.86	-267.18	28.86	-67.56	25.03	-105.69	29.15
2007	11	-525.32	28.74	-243.94	28.74	-489.41	24.41	-153.55	28.99
2007	12	-387.92	28.84	-318.43	28.84	-279.39	24.33	-207.48	28.99

Fig. 10. Comparison Analysis between Observed and Simulated Monthly Rainfall Data

에서 연도와 월은 강우자료를 기준으로 나타내었다. 분석 방법은 1973년부터 2007년까지 관측된 해수면온도와 월 강우량 자료를 이용하여 회귀관계를 설정하고 2008년의 월강우량을 예측하여 비교하였다. 해수면온도는 상관성 분석에서 가장 큰 양(+)의 상관계수를 갖는 지점의 자료를 이용하였으며, 해수면온도를 3개월 지체시켜 분석에

이용하였다. Fig. 10은 각 군집별로 추출된 주성분의 관측 값과 해수면온도를 이용하여 예측한 값을 비교하여 도식한 결과이다.

Fig. 10에서 나타낸 결과에 대해서 예측치와 관측치에 대한 통계적 분석을 수행하였다. 예측 결과를 비교분석하기 위해서 상관계수(C.C.), R-Bias (Relative Bias), RMSE

Table 6. Accuracy Measures of the Simulation

Classification	P1	P2	P3	P4
RMSE	160.80	260.47	376.79	72.68
C.C.	0.94	0.86	0.78	0.93
R-Bias	-0.57	2.01	-1.91	-0.91
R-RMSE	-0.08	-1.06	-0.23	-0.16

(Root Mean Square Error) 및 R(Relative)-RMSE 오차를 비교하였다. RMSE는 추정치와 관측치의 차이를 제곱근하여 평균을 취한 값으로 zero (0)에 가까울수록 검정 결과가 좋은 것으로 알려져 있다. 상관계수는 예측값과 관측값의 정확성을 측정하며, 1에 가까울수록 검정결과가 우수하다.

Fig. 10에서 나타난 것과 같이, 월강우량의 주성분에 따라서 해수면온도를 이용한 예측값이 관측값과 비슷한 결과를 보이거나, 차이가 발생하고 있음을 확인할 수 있다. 군집 2의 주성분처럼 매우 큰 값을 갖는 경우나, 군집 3의 주성분과 같이 변동이 심한 경우에는 예측값이 좋은 결과를 보이고 있지 못하는 점을 확인할 수 있다. 특히 군집 3의 지역은 남해안 지역에서 관측되는 월강우량의 주성분 자료로써 이 지역은 전통적으로 태풍의 영향을 많이 받는 지역으로 태풍의 내습영향에 따라 강우량의 발생에 많은 영향을 받는 지역이다. 따라서 이 지역은 강우량은 월별로 변화가 매우 극심한 지역이기 때문에 예측결과가 다소 미흡한 것으로 판단된다. 그러나 군집 1과 4의 주성분은 예측결과가 관측값과 매우 유사한 크기를 갖는 것으로 나타났다. 이는 해수면온도와 같은 수문기상인자와 강우사상을 결합하여 분석하는 연구가 지속적으로 수행될 필요성이 있음을 의미한다.

6. 결 론

본 연구에서는 해수면온도와 월강우량의 주성분과의 관계에 대하여 분석을 수행하였다. 분석방법은 월강우량 자료를 이용하여 군집분석을 통해 4개 지역으로 구분하여 각각의 주성분을 추출하였다. 추출된 주성분과 해수면온도와의 상관성 분석을 수행하였으며, 이를 바탕으로 해수면온도를 이용하여 월강우량 자료의 단기예측을 수행하였다.

1. 군집분석을 통하여 우리나라의 월강우량 자료를 이용해 월강우특성이 비슷한 4개의 지역으로 구분하였다. 구분된 지역별로 주성분 분석을 통해 추출된 주성분은 모두 첫 번째 주성분이 80% 이상의 분산을 설명하는 것으로 나타났다. 이는 한반도에 발생

하는 월강우량의 특성이 4개 군집의 지점별로 큰 차이를 보이지 않음을 의미한다. 따라서 추출된 4개의 주성분을 이용하여 해수면온도와 상관성 분석을 수행하고, 해수면온도를 이용한 월강우량의 단기예측에 활용하였다.

2. Pearson 상관계수를 이용하여 월강우량의 주성분들과 범지구적 해수면온도와의 상관성 분석을 수행하였다. 상관성 분석 결과에서 우리나라의 월강우량의 특성을 갖는 주성분들과 해수면온도는 통계적으로 유의한 상관관계를 갖고 있는 것으로 나타났다. 군집 1, 2, 4는 아라비아해와 뱅골만 주변의 해수면온도와 상관성이 높으며, 군집 3은 기니만 인근과의 상관성이 높게 나타났다. 각 군집별로 군집 1은 북위 10°~남위 5°, 동경 55°~동경 80°, 군집 2는 북위 20°~0°, 동경 60°~동경 90°, 군집 3은 남위 5°~남위 20°, 서경 5°~동경 35°, 군집 4는 북위 15°~남위 5°, 동경 60°~동경 90° 지역의 해수면온도와 상관관계가 큰 것으로 계산되었다.
3. 지역가중다항식을 통하여 해수면온도를 이용해 월강우량의 주성분의 예측을 수행하였다. 분석 결과에서 분석 결과의 정확성을 대상 자료에 따라서 차이를 보이는 것으로 나타났다. 군집 1과 4는 관측값과 예측값의 편차가 크지 않은 것으로 나타났으며, 군집 2는 경향은 어느 정도 일치하나, 7월의 큰 월강우주성분을 예측하지 못하였다. 군집 3의 주성분은 변동이 심해 예측값과 관측값의 차이가 다소 큰 것으로 나타났다. 이는 해수면온도를 반영하여 월강우량을 예측하는데 현재까지는 한계가 있음을 의미한다. 그러나 절반이상의 강우관측소를 포함하고 있는 군집 1에 포함된 월강우량의 예측이 적절히 수행되었으므로 수문기상인자를 반영한 강우사상의 예측에 대한 연구는 지속적으로 수행할 필요성이 있다.

분석 결과에서 해수면온도를 이용하여 월강우량을 예측한 결과를 지역에 따라 정확성에서 차이가 나는 것으로 나타났다. 그러나 일부 지역에서는 해수면온도를 이용하여 월강우량을 예측한 결과가 관측값과 매우 비슷하게 나

타나고 있다. 따라서 해수면온도와 같은 수문기상인자는 강우사상과 밀접한 관련을 맺으며 복잡한 과정을 통해 발생하므로, 이에 대한 연구를 지속할 필요성이 있다.

감사의 글

이 논문은 2009년도 서울시립대학교 교내학술연구비에 의하여 연구되었음.

참 고 문 헌

- 김우철, 김재주, 박병욱, 박성현, 송문섭, 이상열, 이영조, 전종우, 조진섭 (공저) (2008). 일반통계학. 영지문화사, pp. 338-339.
- 문영일, 권현한, 김동권 (2005). “해수면온도와 우리나라 강수량과의 상관성 분석.” **한국수자원학회논문집**, 한국수자원학회, 제38권, 제12호, pp. 995-1008.
- 오태석, 문영일 (2008). “우리나라 월강수량과 범지구적 해수면온도의 상관성 분석.” **대한토목학회논문집**, 대한토목학회, 제28권, 제2B호, pp. 237-248.
- 오태석, 안재현, 문영일, 권현한 (2007) “우리나라 강수량과 한반도 근해의 수온 및 기온의 상관성 분석.” **대한토목학회논문집**, 대한토목학회, 제27권, 제2B호, pp. 1-15.
- 유철상, 정건희, 김중훈 (2000). “우리나라 근해 해수면 온도와 기온 및 강수량과의 상관관계에 관한 연구.” **한국수자원학회 학술대회지**, 한국수자원학회, pp. 145-150.
- 추현재, 김태웅, 이정규 (2007). “우리나라 강수량과 태평양 해수면온도의 원격상관관계 분석.” **한국수자원학회 2007년도 학술발표회 논문집**, 한국수자원학회, pp. 1953-1957.
- Barnston, A.G., and Smith, T.M. (1996). “Specification and Prediction of Global Surface Temperature and Precipitation from Global SST Using CCA.” *Journal of Climate*, Vol. 9, pp. 2660-2697.
- Cayan, D.R., and Peterson, D.H. (1989). “The Influence of North Pacific Atmospheric Circulation on Streamflow in the West.” *American Geophysical Union*, Washington, D.C. pp. 375-397.
- Cleveland, W.S., and Devlin, S.J. (1988). “Locally weighted regression: An approach to regression analysis by local fitting.” *Journal of the American Statistical Association*, Vol. 83, No. 403, pp. 596-610.
- Harshburger, B., Ye, H., and Dzialoski, J. (2002). “Observational evidence of the influence of Pacific SSTs on winter precipitation and spring stream discharge in Idaho.” *Journal of Hydrology*, Vol. 264, pp. 157-169.
- Higgins, R.W., Kim, H.-K., and Unger, D. (2004). “Long-Lead Seasonal Temperature and Precipitation Prediction Using Tropical Pacific SST Consolidation Forecasts.” *Journal of Climate*, Vol. 17, pp. 3398-3414.
- Hottelling, H. (1931). “The generalization of student’s ratio.” *Annals of Mathematical Statistics*, Vol. 2, No. 3, pp. 360-378.
- Klein, W.H., and Bloom, H.J. (1987). “Specification of monthly precipitation over the United States from the surrounding 700 mb height field.” *Month Weather Review*, Vol. 115, pp. 2118-2132.
- Kiladis, G.N., and Diaz, H.F. (1989). “Global climatic anomalies associated with extremes in the Southern Oscillation.” *Journal of Climate*, Vol. 2, pp. 1069-1090.
- Kim, T.-W., Valdes, J.B., Nijssen, B., and Roncayolo, D. (2006). “Quantification of linkages between large-scale climatic patterns and precipitation in the colorado River Basin.” *Journal of Hydrology*, Vol. 321, pp. 173-186.
- Loader, C. (1999). *Local regression and likelihood*. Springer, USA.
- Misra, V. (2003). “The Influence of Pacific SST Variability on the Precipitation over Southern Africa.” *Journal of Climate*, Vol. 16, pp. 2408-2418.
- Moon, Y.I., and Lall, U. (1996). “Atmospheric flow indices and interannual Great Salt Lake variability.” *Journal of Hydrologic Engineering*, Vol. 1, No. 2, pp. 55-62.
- Rautenbach, C.J., deW., and Smith, I.N. (2001). “Teleconnections between global sea-surface temperatures and the interannual variability of observed and model simulated rainfall over southern Africa.” *Journal of Hydrology*, Vol. 254, pp. 1-15.
- Uvo, C.B., Repelli, C.A., Zebiak, S.E., and Kushnir, Y. (1998). “The Relationships between Tropical Pacific and Atlantic SST and Northeast Brazil Monthly Precipitation.” *Journal of Climate*, Vol. 11, pp. 551-562.
- Smith, T.M., and Reynolds, R.W. (2003). “Extended Reconstruction of Global Sea Surface Temperatures Based on COADS Data (1854-1997).” *Journal of Climate*, Vol. 16, pp. 1495-1510.
- Smith, T.M., and Reynolds, R.W. (2004). “Improved Extended Reconstruction of SST (1854-1997).”

Journal of Climate, Vol. 17, pp. 2466–2477.

Smith, T.M., and Reynolds, R.W., Peterson, T.C., and Lawrimore, J. (2008). “Improvements to NOAA’s Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006).” *Journal of Climate*, Vol. 21, pp. 2283–2296.

Wand, M.P., and Jones, M.C. (1995). *Kernel smoothing*. Chapman & Hall/CRC, USA.

논문번호: 09-131	접수: 2009.11.25
수정일자: 2010.02.26/04.20	심사완료: 2010.04.20