

남방진동지수와 서울 강수량과의 Mutual Information

○ 김진영*, 문영일**, 황성환***, 박대형****

1. 서론

엘니뇨 사상이 기상이변의 주된 원인으로 밝혀지면서 국내에선 엘니뇨와 우리나라 강수량간의 상관관계를 분석하기 위한 많은 연구들이 진행되었다(강인식, 1994; 신현식, 1999; 차은정, 2000; Moon 등, 2000). 전통적인 통계적 기법을 사용하여 엘니뇨와 강수량간의 선형적인 상관관계를 분석한 것이 주를 이루고 있지만, 대부분의 수문자료들이 매우 불규칙한 비선형적인 성향을 가짐을 고려해 볼 때, 단순히 선형상관성에 의존하여 엘니뇨 사상이 우리나라에 미치는 영향을 분석하는 것은 바람직하지 않을 것이다.

본 연구에서는 우리나라의 강수량과 엘니뇨 사상의 정량적인 기준으로 많이 사용되어지는 남방진동지수(Southern Oscillation Index; SOI)를 가지고 자료의 비선형적인 분포에서도 상관성을 규명할 수 있는 기법을 사용하였다. 사용된 기법은 좌표공간에서 자료의 발생확률에 의해서 자료의 독립성을 찾아내는 원리를 이용한 Mutual Informaion(MI) 기법(Fraser와 Swinney, 1986; Moon 등, 1996)이다. 즉, 강수량자료와 남방진동지수 간의 대응하는 각 자료점에 핵함수를 위치시켜 결합확률밀도를 도출하여 엘니뇨 사상이 서울 강수량에 미치는 영향을 분석하였다.

2002년 초 현재 NOAA(National Oceanic and Atmospheric Administration) 발표에 따르면 엘니뇨사상이 서서히 발생되고 있으며 연말 정도에는 강하게 발생될 것으로 예측되고 있다. 이러한 시기에 본 연구의 결과가 엘니뇨 발생에 따른 피해 경감 대책의 수립에 도움이 되었으면 한다.

* 연구원, 서울시정개발연구원
** 부교수, 서울시립대학교 토목공학과
*** 박사과정, 서울시립대학교 토목공학과
**** 박사과정, 서울시립대학교 토목공학과

2. Mutual Information

관측된 임의의 두 변수의 시계열을 $s_1, s_2, s_3, \dots, s_i, q_1, q_2, q_3, \dots, q_n$ 으로 표현하자. 여기서 n 은 관측기간이고, 관측치 s_i 와 q_i 간의 Mutual Information은 다음 식(1)과 같은 조건 하에 정의된다(Fraser와 Swinney, 1986).

$$MI_{s,q}(s_i, q_j) = \log_2 \left(\frac{P_{s,q}(s_i, q_j)}{P_s(s_i)P_q(q_j)} \right) \quad (1)$$

여기서, $P_{s,q}(s_i, q_j)$ 는 (s_i, q_j) 에서 계산된 s 와 q 의 결합확률밀도함수(joint probability density function)이고, $P_s(s_i)$ 와 $P_q(q_j)$ 는 각각의 s_i 와 q_j 에서 계산된 주변확률밀도함수(marginal probability density function)이다. 따라서 두 관측자료 s 와 q 가 독립이면 $P_{s,q}(s_i, q_j) = P_s(s_i)P_q(q_j)$ 이므로, $MI = 0$ 이다. 그러므로, MI 값이 0에 가까울수록 독립의 관계로 볼 수 있다.

우리가 관심을 가지는 부분은 두 계열간의 전반적인 상관성이며, 평균 Mutual Information $I_{s,q}$ 를 다음과 같이 정의 할 수 있다.

$$I_{s,q} = \sum_{i,j} P_{s,q}(s_i, q_j) \log_2 \left(\frac{P_{s,q}(s_i, q_j)}{P_s(s_i)P_q(q_j)} \right) \quad (2)$$

식(2)는 종속과 독립여부를 떠나 다변량 sampling에서 구성요소를 구해내는데 유용하다. 특별히 최근에는 경험적 시계열로부터 상태공간을 구성할 때 적절한 지체시간 τ 의 선택(Martinerie 등, 1992; Abarbanel, 1994; Gao, 1994)에 사용되었다.

Fraser와 Swinney(1986)은 MI 의 추정과 τ 의 선택을 위해 다변량 히스토그램의 사용을 제안하였고, 본 논문에선 Moon 등(1996)이 사용한 비매개변수적 핵밀도함수를 이용하여 남방진동지수와 우리나라 강수량간의 상관성 분석에 응용하였다.

3. Mutual Information을 이용한 서울 강수량과 SOI의 상관분석

(1) 자료구성

본 논문에서 사용된 강수량자료는 자료축적 년수가 길고 비교적 신뢰성이 높은 서울강수량 자료로서 1932년 9월부터 2001년 7월까지의 총 827개의 월강수량 자료와 같은 기간의 SOI 자료를 사용하였다.

그림 1은 엘니뇨가 발생된 해의 여름강수량을 나타낸 그림으로 엘니뇨가 발생된 해의 여름철 강수량은 평균강수량을 기준으로 증가 또는 감소하는 일정한 경

향을 가지고 있지 않지만, 큰 규모의 강수가 대부분 엘니뇨 해에 발생함을 보여주는 사실은 주목할 만 하다.

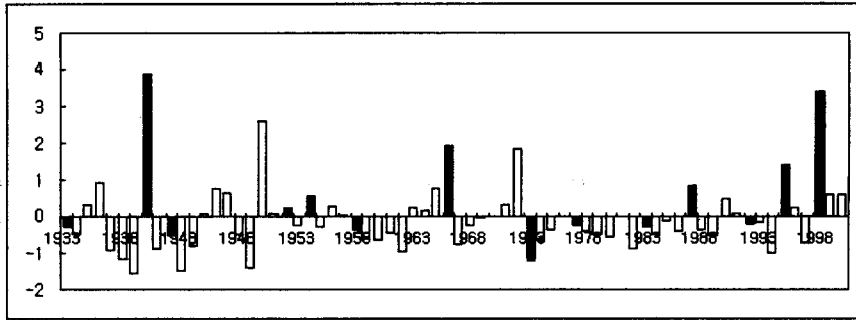


그림 1 엘니뇨가 발생된 해의 여름강수량 추이
(□ 정상시 ■ 엘니뇨발생)

서울 관측소의 강수량 자료는 년 주기 성분이 강하고 SOI 자료는 강수량 자료에 비하여 불규칙한 경향을 보인다. 본 논문에선 이렇게 특성이 다른 두 자료의 공통의 성분을 뽑아 그 상관성을 파악하기 위해서 강수량 자료에 대하여 년주기를 제거하기 위하여 강수량의 월평균과 표준편차를 이용하여 정상화시켰고, 이들의 공통인자를 추출하는 방법으로 다변량스펙트럼분석(Multi-channel Singular Spectral Analysis, M-SSA)을 이용하여, 자료를 재구성(Reconstructed Component, RC)하였으며, 이렇게 구성된 자료를 이용하여 MI 기법에 적용하였다.

(2) MI 기법을 이용한 상관분석 결과

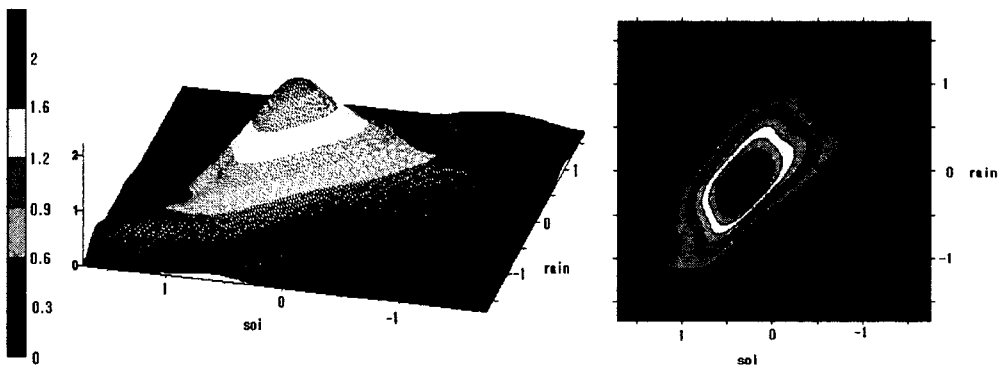


그림 2 강수량과 SOI의 3-D 결합확률밀도함수

그림 2는 두 자료의 상관관계를 핵함수를 이용한 결합확률밀도함수로 표현한 것이다. SOI와 서울강수량은 음의 상관을 보이고 있으며, SOI가 -4.6~2.9에 걸쳐 분포하고 SOI 값의 평균이 -0.123이며 최빈값이 양의 값에 있다. 그림 2에서 결합 확률밀도함수를 보면 서울관측소의 강수량자료는 평균에 비하여 낮은 경우가 많으며, 증가 될 때는 그 값이 크다고 할 수 있다.

이상의 결과에서 서울관측소의 강수량자료와 SOI 자료의 공통요소는 서로 음의 상관성으로 SOI 값이 감소하면, 서울관측소의 강수량은 커질 확률이 큰 요소를 내재하고 있다.

(3) 각 구간별 SOI에 따른 서울 강수량과의 관계분석

SOI의 총 발생 구간(-4.6~2.9)을 그림 3과 같이 나누어서 각 구간에서의 월별 평균강수량과 표준편차로 표준화된 서울 강수량의 평균값들과 증감횟수를 그림 4와 그림 5에 도시하였다.

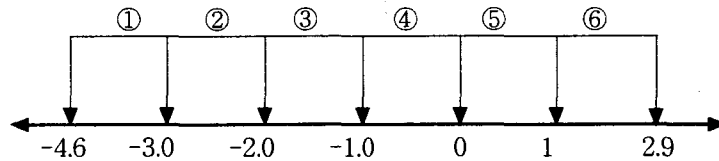


그림 3 SOI 각 구간의 모식도

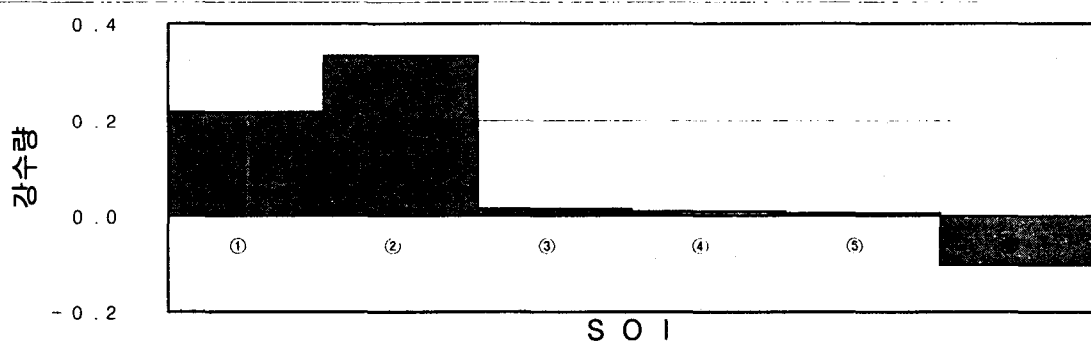


그림 4 구간별 SOI 값에 대한 강수량의 월별 강수량 표준화된 값의 평균

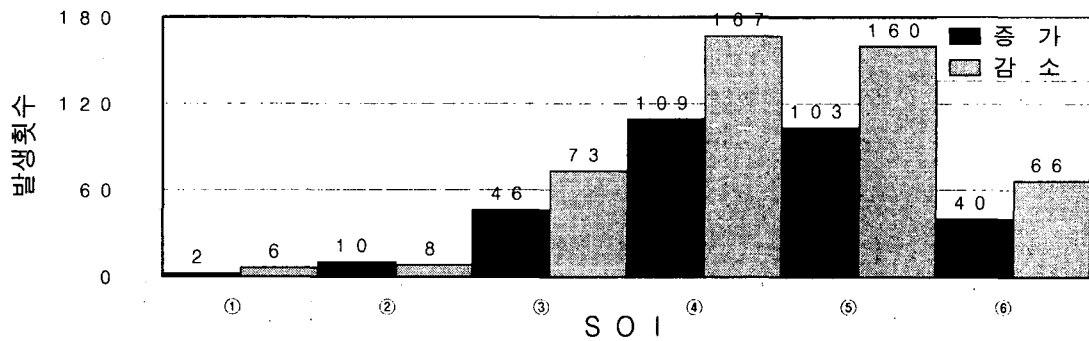


그림 5 SOI 각 구간별 강수량의 증감횟수

그림 4에서 SOI 값이 음수인 엘니뇨 사상이 나타날 때, 월강수량은 증가되는 경향임을 알 수 있다. 또한 그림 4와 그림 5를 통해 서울 강수량과 SOI 값의 관계를 분석해 보면, 강한 엘니뇨 사상을 나타내는 SOI 값이 음수인 ① 구간에선 월강수량이 감소된 빈도는 다소 많으나, 월평균값이 증가하였다. 이것은 엘니뇨 발생시 큰 사상의 강우가 발생할 확률이 높음을 의미한다. ② 구간에선 대체적으로 강수량의 증가할 확률도 높고, 평균 강수량도 큰 값을 보인다. ③~⑤ 구간에선 강수량이 평균이하일 때가 많고, 큰 홍수 사상이 잦은 우리나라 강수의 일반적인 형태를 보여주고 있다. 마지막으로 라니냐 사상쪽인 ⑥ 구간에선 강수량이 감소할 확률이 높고, 강수량 역시 감소함을 보이고 있다.

4. 결론

본 논문에서는 서울관측소의 강수량 자료와 SOI 자료간의 상관관계를 MI 기법을 통하여 알아보려고 하였다. 히스토그램을 개량한 핵함수를 이용한 MI 기법을 이용하였고, 3차원의 결합확률밀도함수를 도출하여 강수량과 SOI 자료간의 상관관계를 분석하였다. 또한 SOI 자료를 구간별로 나누어 구간에 따른 강수량의 증감과 표준화된 강수량의 구간에 따른 평균값을 통하여 MI 기법을 통한 분석결과를 확인하였다.

구간별로 분석한 결과, -4.6~-3.0의 극치의 SOI 값일 때 이때의 강수의 증가 확률은 낮지만, 큰 강수를 유발할 확률이 크다. 또한 SOI 값이 -3.0~-2.0일 때 서울관측소의 강수량은 평균보다 크며, 또한 증가할 확률이 더 높음을 보여 주고 있다. 이는 MI 기법의 3차원 결합확률밀도함수를 분석한 결과와 유사하다. 특히, 본 논문에 사용된 MI 기법은 핵함수를 이용하여 비선형적인 상관성 및 지체시간을 확인할 수 있다는 면에서 다른 기법에 비하여 실용성이 있다고 할 수 있다.

강수의 발생원인은 단순히 바람, 온도, 습도들만의 조합에 의한 발생이 아니라,

그 외에도 많은 인자들의 복합적인 조합으로 이루어졌다. 특히, 우리나라는 중위도에 위치하고 기단과 몬순 등 많은 기상적 요인들에 의하여 영향을 받고 있어, 서울 강수량과 엘니뇨 사상이 음의 상관성이 있다고 해서 반드시 엘니뇨 사상이 발생하면 서울 강수량은 증가한다고 확정적으로 말할 수 없다. 단지 결합확률밀도 함수가 의미하는 것처럼 SOI 값에 대한 서울 강수량은 음의 상관성을 보이지만, 항상 엘니뇨 사상에 대하여 강수량이 증가하는 것은 아니고, 엘니뇨 사상은 서울 강수량의 불규칙성이 증가시켜 큰 호우를 초래할 확률이 높다고 할 수 있다.

감사의 글

본 연구는 건설교통부의 2001년 국제수문개발계획(IHP)의 연구비지원에 의해 수행되었습니다.

참고문헌

- 강인식, 이양미, 안순일, 1994, 서태평양 지역에서 태풍활동의 경년변화와 엘니뇨, 한국기상학회지, 15-26.
- 신현식, 안재현, 윤용남, 1999, 엘니뇨와 우리나라 강수량의 시·공간적 상관관계 분석, 대한토목학회지, 1-12.
- 차은정, 최영진, 오재호, 2000, 1997/1998년 엘니뇨와 1998년 여름철 이상기상분석, 한국수자원학회지, 269-284.
- Fraser, A. M. and Swinney, H. L., 1986, Independent coordinates for strange attractors from mutual information., *Physical Review A*, 33, pp. 1134-1140.
- Gao, J. and Zheng, Z., 1994, Direct dynamical test for deterministic chaos and optimal embedding of a chaotic time series, *Physical Review E*, 49, pp. 3807-3814.
- Martinerie, J. M., Albano, A. M., Mees, A. I., and Rapp, P. E., 1992, Mutual information, strange attractors, and the optimal estimation of dimension, *Physical Review A*, 45, pp. 7058-7064.
- Moon, Y. I. and Park D. H., 2000, Low frequency teleconnections between seoul precipitation and southern oscillation index, International conference of hydraulic engineering.
- Moon, Y. I. and Lall, U., 1996, Atmospheric flow indices and interannual Great Salt Lake Variability, *Journal of Hydrologic Engineering*, April, 55-61.
- Silverman, B. W., 1986, Density estimation for statistics and data analysis, Chapman and Hall, New York.