

Nino3.4지역 SST 및 여름강수량의 독립성분분석

Independent Component Analysis of Nino3.4 Sea Surface Temperature and Summer Seasonal Rainfall

권 현 한* / 문 영 일**

Kwon, Hyun-Han / Moon, Young-Il

Abstract

We examined problems of the principal component analysis(PCA), which is able to analyze at the low dimensionality as a methodology to assess hydrologic time series, and introduced the theory and characteristics of independent component analysis(ICA) that can supplement problems of principal component analysis. We also applied the global sea surface temperature(SST) of the Nino region and assessed the correlation between El Niño-Southern Oscillation(ENSO) and SST. The results of examining separation-ability of principal components using mixed signals indicate that the independent component analysis is statistically superior compared to that of the principal component analysis. Finally, we assessed correlation between ENSO and global anomaly SST. The independent component analysis was applied to the $5^{\circ} \times 5^{\circ}$ (latitude and longitude) global anomaly SST in the Nino+3.4 region that is the El Niño observation section. We assessed the correlation with the ENSO years. These results of the analysis show that only one independent component(86%) was able to represent the entire behavior and was consistent with the main ENSO years. Finally, we carried out independent component analysis for summer seasonal rainfalls at nine stations and could extract ICs to reflect geographical characteristics. The increasing trend has been shown at IC-1 and IC-2 since 1970s.

keywords : Independent component analysis, principal component analysis, global sea surface temperature, ENSO

요 지

수문시계열을 분석하기 위한 방법으로 낮은 차원에서 해석이 가능한 주성분분석 방법의 문제점을 검토하고 이를 보완할 수 있는 독립성분분석의 이론과 특성을 검토하였고 수문기상자료인 Nino지역의 해수면온도에 적용하여 El Niño-Southern Oscillation(ENSO) 사상과의 상관성을 평가하였다. 혼합자료를 사용하여 독립성분분석 방법의 주성분 분리 능력을 검토한 결과 독립성분분석이 기존 주성분분석에 비해 통계적으로 우수한 결과를 나타내었다. El Niño의 감시구역인 Nino+3.4지역의 $5^{\circ} \times 5^{\circ}$ (위도와 경도)의 총 20개 Global Anomaly SST 격자자료를 대상으로 분석을 실시하여 El Niño-Southern Oscillation(ENSO)사상의 발생시기와의 상관관계를 평가하였으며 한 개의 독립성분(86%)만으로 SST의 전체적인 거동을 표현할 수 있었으며 주요 ENSO 발생시기와 일치하는 결과를 나타내었다. 또한 국내 주요지점의 여름 강수량을 대상으로 독립성분분석을 적용한 결과 지역적인 특성을 고려하여 비교적 합리적으로 독립성분을 추출할 수 있었으며 IC-1과 IC-2에서 1970년대 이후로 여름강수량의 증가를 확인할 수 있었다.

핵심용어 : 독립성분분석, 주성분분석, Global SST, ENSO

* 정희원, Columbia University, Postdoctoral Research Associate, NY, USA (e-mail: hk2273@columbia.edu)

** 정희원, 서울시립대학교 토목공학과 부교수 · 공학박사 (e-mail: ymoon@uos.ac.kr)

1. 서 론

수자원의 효율적인 관리를 위해서 강우, 유출 및 기상자료 등의 상관관계를 규명하는 것은 수문학자들의 주요 관심사가 되고 있으며 특히 최근 이상기후의 영향으로 빈발하는 홍수와 가뭄 등의 수자원관리에서의 문제점을 극복하려는 수단으로서 기상자료와 수문자료와의 상관성을 이용하여 이를 예측하고자 하는 많은 연구가 진행되고 있다. 그러나 자연계에서 발생하는 수문현상은 매우 복잡한 Mechanism을 통한 결과로서 이러한 복잡성을 보다 작은 차원에서 해석하기 위한 주요 수단으로 주성분분석(principal component analysis)이 주로 이용되고 있다. 여러 개의 반응변수(response variable)에 대하여 얻어진 다변량 자료(multivariate data)를 분석의 대상으로 하는 주성분분석은 다차원적인 변수들을 축소, 요약하는 차원의 단순화와 더불어 일반적으로 서로 상관되어있는 반응변수들 상호간의 복잡한 구조를 분석하는데 그 목적을 두고 있다. 이를 위하여 주성분분석은 반응변수들을 일반적으로 선형 변환시켜, 주성분(principal component)이라고 부르는 서로 상관되어 있지 않은(uncorrelated), 혹은 독립적인 새로운 인공변수들을 유도한다. 이때 각 주성분이 보유하는 변이의 크기를 기준으로 그 중요도의 순서를 생각할 수 있는데, 그들 중 몇 개의 주성분에 의해 원래자료에 내재하는 전체변동 성분 중 가능한 한 많은 부분이 보유되도록 변환시킴으로서 정보의 손실을 최소화하는 차원의 축소(dimension reduction)가 주성분분석에 기본 개념이라 할 수 있다. 이러한 주성분분석 방법으로 박대형 등(2000)은 Multi-Channel Singular Spectrum Analysis(MSSA)를 이용하여 국내 주요강우와 SOI의 지면도 상관관계를 추정하였으며 김진영 등(2002)은 상호정보(mutual information, MI)를 이용하여 서울강수량과 SOI의 비선형상관관계에 대한 연구를 수행하였고 박상우 등(2003)은 주성분분석을 수문분석에서 지점강우의 권역화 등에 응용한 사례가 있다. 그러나 주성분분석에서 비상관성을 가지는 새로운 주성분을 유도할 수 있지만 이들 주성분이 최종적으로 독립성을 확보하기에는 미흡하다는 지적이 있다(Common, 1994; Bell and Sejnowski, 1995; Lee et al., 1999).

최근에 이러한 단점을 보완할 수 있는 방법론으로서 자료의 상관관계이외에 더 높은 차수의 상관관계를 제거하여 최종적으로 차원들의 관계를 독립으로 변환시켜 독립성분을 추정할 수 있는 독립성분분석(independent component analysis)이 제안되어 신호처리 및 다변량 시계열 분석에서 응용이 이루어지고 있다(Common,

1994; Bell and Sejnowski, 1995; Lee et al., 1999). 이 방법은 기존의 주성분분석이 2차 통계치인 분산을 최대화하기 위해서 직교좌표(orthogonal basis)와 정규성(gaussianity)을 이용하는 반면에 독립성분분석은 비직교좌표(non-orthogonal basis)와 비정규성(non-gaussianity)을 토대로 고차의 통계치를 이용하는 방법론이다.

따라서 본 연구에서는 기존 주성분분석과 독립성분분석을 비교하여 수문기상자료에 대한 적합성을 분석하고자 한다. 이를 위하여 2가지 방법론의 이론적인 배경 및 특성을 검토하고 모의자료에 대해서 각 방법론의 타당성을 평가하고 El Niño의 감시구역인 Nino+3.4지역의 위·경도 $5^{\circ} \times 5^{\circ}$ 의 총 20개 Global Anomaly SST 격자자료를 대상으로 분석을 실시하여 El Niño-Southern Oscillation(ENSO)사상의 발생시기와 상관관계를 평가하였다. 또한 국내 주요 9개 지점의 여름 강수량(6월-9월)을 대상으로 독립성분분석을 실시하여 전체적인 경향성 및 각 지점과 추정된 독립성분간의 상관관계를 검토하였다.

2. 주성분분석과 독립성분분석

2.1 주성분분석

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 기법(second order statistics)이다. 자료를 설명하기 위해 투영(projection)시켜 분산이 최대가 되는 기본벡터(basis vector)를 찾게 되며 이들 가운데 서로 직교하는 성질을 만족하는 기본벡터를 찾게 되는데 이 때 기본벡터를 주성분이라고 하며 이 벡터를 자료로 취급해 분석하게 된다. 이는 입력 자료의 가장 중요한 축들을 찾아 효율적으로 자료의 차원을 줄일 수 있는 장점을 갖게 됨을 의미한다. 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.

$$X = TS^T + E \quad (1)$$

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 Eigenvalue 중에서 큰 순서대로 p 만큼 택하여 이에 해당하는 Eigenvector로 구성하면 된다. 여기서, p 는 주성분분석에서 나타내는 Eigenvector의 수가 된다. Eigenvalue는 PC로부터 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. (1)은 다음 Eq. (2)와 같은 형태가 된다.

$$\bar{T} = \bar{X} S \quad (2)$$

여기서, 벡터 \bar{X} 는 행렬 X 의 열(row)성분이며 단일 자료의 벡터이며, 벡터 \bar{T} 는 행렬 T 혹은 특성 공간(feature space) \bar{X} 에서 해당하는 열성분을 의미한다. 행렬 S 는 선형 변환의 계수를 의미한다.

2.2 독립성분분석

독립성분분석은 혼합된 신호로부터 숨어있는 신호 분리(blind source separation, BSS)의 문제를 해결하는 방법으로 가장 널리 사용되고 있다. BSS란 아무런 사전 지식 없이 선형 혼합된 신호로부터 알려지지 않은 근원신호 또는 혼합 행렬을 구해내는 것이 목적인 통계적인 방법이다. 독립성분분석은 오직 기본 시계열 성분들이 서로 통계적으로 독립이라는 사실만 가지고 자료의 고차원 통계 성분을 이용하여 분리해내는 방법으로 신호처리나 인공 신경망의 여러 분야에서 다양하게 사용되고 있다(Comon, 1994).

독립성분분석은 주성분분석과 유사하게 자료를 투영시키는 방법이라는 공통점을 갖는 반면 가장 큰 차이점은 비직교변환과 비정규성을 고려할 수 있다는 것이다. 주성분분석은 2차 통계치(variance)를 최대화해 주는 반면 독립성분분석은 비정규성을 규정할 수 있는 왜곡도(skewness), 첨예도(kurtosis)와 같은 고차원 통계치를 고려하며 비직교 함수를 찾아주게 된다. 이는 비정규성을 최대화 시키거나 상호정보를 최소화시켜주는 방향의 주성분을 찾는 것을 의미하며 Table 1은 독립성분분석

과 주성분분석의 해석상의 개념적인 차이를 비교하여 나타내었다. Fig. 1은 독립성분분석과 주성분분석의 해석과정을 나타낸다.

Fig. 1에서 2개의 독립적인 신호를 혼합시켜서 독립성분분석과 주성분분석의 해석상의 차이점을 설명하였으며 여기서 주성분분석은 새로운 좌표축을 통해 자료의 분산을 최대화하기 위해서 직교좌표를 이용하여 투영하며 이러한 특성으로 인해 원신호로 재구성 시킬 때 왜곡되는 문제점이 발생하게 된다. 반면 독립성분분석은 혼합신호에 대해서 비직교좌표를 구성할 수 있으므로 자료를 재구성할 때 왜곡되지 않고 구성이 가능하다.

독립성분분석은 근원신호들의 선형조합으로부터 얻어지는 신호로부터 독립된 근원신호를 회복하는데 이용된다(Comon, 1994). 독립성분분석의 해석단계는 선형이라는 것을 제외하고 혼합과정에 관한 가정은 존재하지 않는다. 독립성분분석에 대한 해석과정을 쉽게 이해하기 위해서 Sine 함수(sinusoid)의 혼합된 성분을 예를 들어 생각해 보기로 한다. 만약 일정한 주기를 갖는 3개의 서로 독립적인 Sine 함수가 있다고 가정하고 이를 $o_1(t)$, $o_2(t)$, $o_3(t)$ 로 표현할 수 있다. 이들 3개의 함수를 서로 다른 가중치를 가지고 선형적으로 혼합시킬 수 있으며 각각 $u_1(t)$, $u_2(t)$, $u_3(t)$ 로 나타낼 수 있다. 따라서 3개의 독립적인 Sine 함수 $o_1(t)$, $o_2(t)$, $o_3(t)$ 로 혼합성분을 함수로 나타내면 다음 Eq. (3)과 같다.

$$\begin{aligned} u_1(t) &= a_{11}o_1 + a_{12}o_2 + a_{13}o_3 \\ u_2(t) &= a_{21}o_1 + a_{22}o_2 + a_{23}o_3 \\ u_3(t) &= a_{31}o_1 + a_{32}o_2 + a_{33}o_3 \end{aligned} \quad (3)$$

따라서 이들 혼합된 함수로부터 근원적인 독립성분을 분리해 낼 수 있다면 다변량(multivariate) 자료 분석에 도움이 될 수 있을 것이다. 우리는 $o_1(t)$, $o_2(t)$, $o_3(t)$ 가 통계적으로 독립이라면 독립성분분석을 이용하여 $a_{ij}(t)$ 를 추정할 수 있다. 만약 n 개의 선형적으로 혼합된 성분을 관측했다고 가정하면 다음 Eq. (4)로 나타낼 수 있다.

Table 1. Comparison between independent component analysis and principal component analysis

Principal Component Analysis	Independent Component Analysis
Orthogonal transform Two order statistics Gaussianity	Non-orthogonal transform Higher order statistics Non-gaussianity

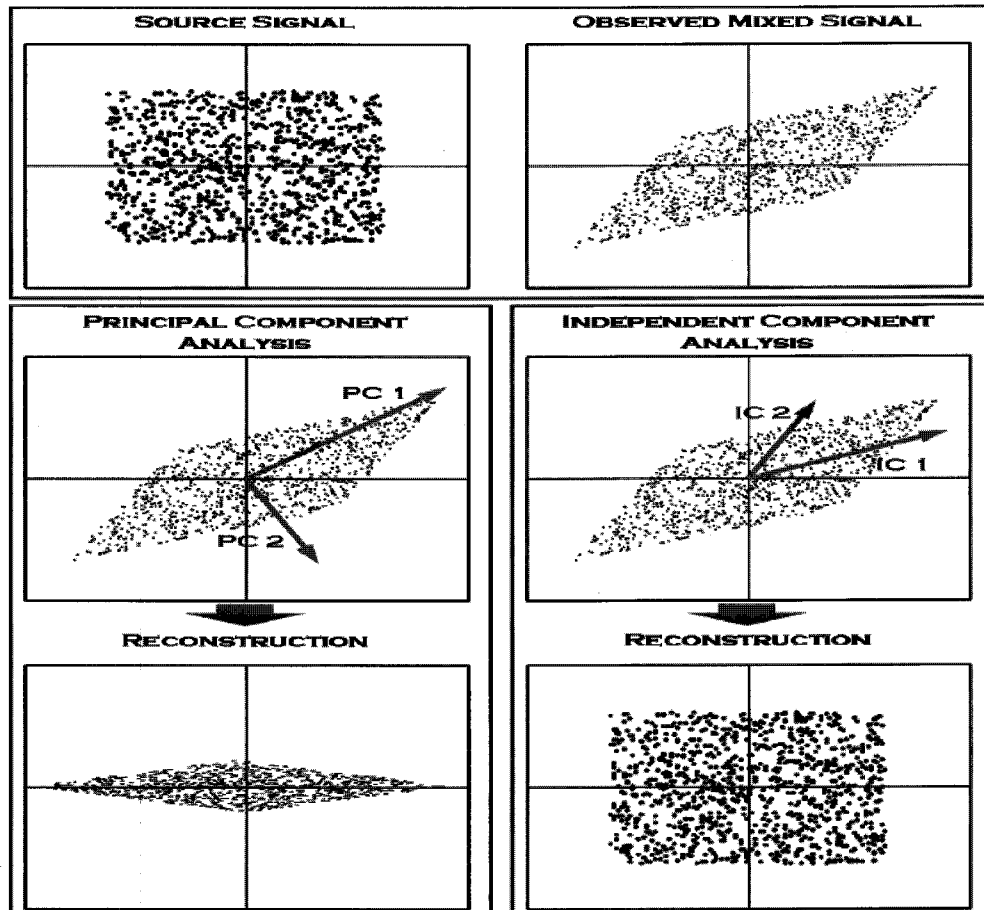


Fig. 1. Graphical comparison between independent component analysis and principal component analysis

$$u_i(t) = a_{i1}o_1 + a_{i2}o_2 + \dots + a_{in}o_n, \text{ for all } i \quad (4)$$

수학적으로 관측된 성분들인 $u = \{u_1(t), u_2(t), \dots, u_N(t)\}$ = Ao 을 나타내기 위해서 행렬 A 를 곱하면 근원성분들을 고립시키기 위한 알고리즘으로 표현할 수 있다. 여기서 행렬 A 를 곱한다는 것은 선형혼합과정(linear mixing process)을 나타내고 알고리즘은 행렬 A 또는 각 구성 성분들의 기본적인 특징을 가능한 최대로 유지시킬 수 있도록 가정한다. 편의성을 위해서 혼합행렬을 정방행렬로 가정하자. 독립성분분석을 이용하여 행렬 A 를 추정 한 후에, A 의 역행렬 W 를 계산할 수 있으며 다음 Eq. (5)에 의해서 독립성분을 추정하게 된다.

$$o = Wu \quad (5)$$

W 를 추정하기 위해서 독립성분분석은 다음 3가지를 가정한다. 1) 근원성분은 통계적으로 독립이다. 2) 대부분의 성분들은 비정규분포를 따른다. 3) 성분들은

Stationary하다. 일반적인주성분 분석과 같은 비상관(decorrelation) 기법과 대조적으로 독립성분분석은 출력성분들이 비상관성을 유지할 수 있을 뿐만 아니라 근원성분들의 독립성을 최대한 유지시킬 수 있도록 결합 확률밀도함수가 주변확률밀도함수의 곱이 되도록 하는 기준을 바탕으로 하는 방법론이다. Eq. (6)은 독립성분 분석의 독립성을 유지하기 위한 기준을 나타낸다(Bell and Sejnowski, 1995; Lee, Girolami and Sejnowski, 1999).

$$f_o(o_1, \dots, o_n) = \prod_{i=1}^n f_{o_i}(o_i) \quad (6)$$

독립성분분석에서 추정 한 주성분은 일반적인 주성분 분석과 같이 자료를 설명해 주는 분산의 정도로서 주성분의 순위를 추정할 수 없으며 단지 총 Eigenvalue에서 각 성분이 나타내는 Eigenvalue의 정도로서 표현이 가능하다. Fig. 2는 독립성분분석의 해석과정을 그림으로 나타내었다. 원자료 s 를 혼합시킨 후 새로운 비직교좌

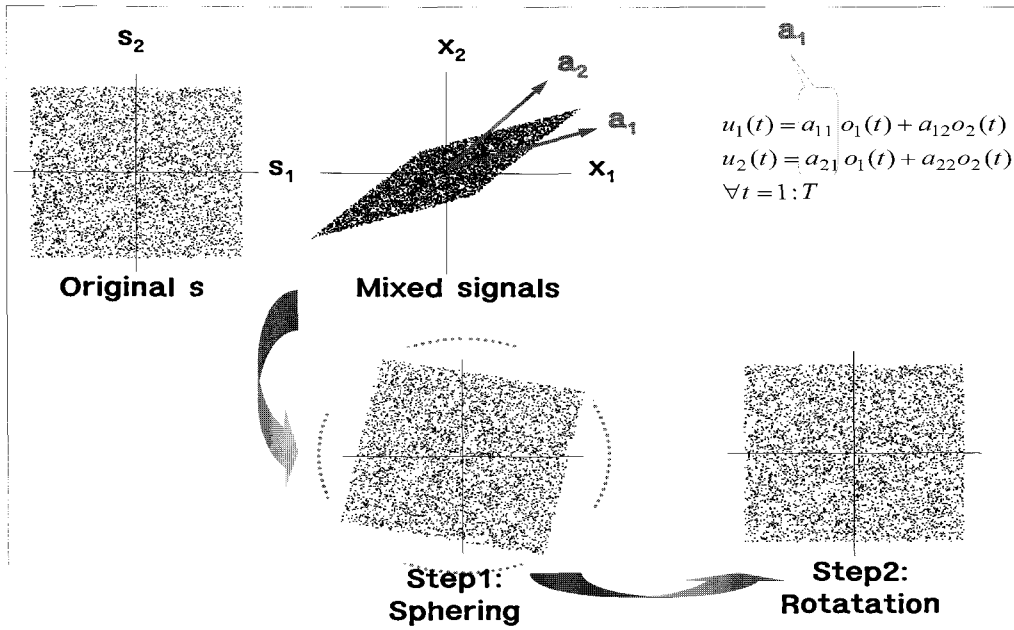


Fig. 2. Concept of independent component analysis

표 a_1, a_2 를 추정하고 이를 회전시켜 최종적으로 원자료와 유사한 특성으로 자료를 추출할 수 있다.

3. 혼합자료에 대한 주성분분석과 독립성분분석

본 연구에서는 앞서 언급된 독립성분분석과 주성분분석의 특성을 비교함과 동시에 모의 자료를 구성해 2가지 방법론을 평가하고자 한다. 모의자료는 두 개의 Sine 함수 $SIN(12X+30)$, $SIN(3X+14)$ 를 조합하여 자료크기 $N=500$ 개의 주기성을 갖는 2개의 시계열로 구성하였다. Fig. 3은 모의된 자료계열을 나타낸다.

Fig. 3에서 실선은 $Y=0.1[SIN(12X+30)]+0.9[SIN(3X+14)]$ 를 점선은 $Y=0.1[SIN(3X+14)]+0.9[SIN(12X+30)]$

$+0.9[SIN(12X+30)]$ 를 나타낸다. 본 연구에서는 2개의 혼합된 신호를 대상으로 독립성분분석과 주성분분석을 각각 실시하였다. 각 방법론에서 추정된 주성분과 모의된 자료의 주성분인 $SIN(12X+30)$ 과 $SIN(3X+14)$ 를 비교하여 Fig. 4~5에 나타내었다.

Fig. 4는 주성분 $SIN(3X+14)$ 의 추정결과를 나타내며 굵은 실선은 원시계열을 나타내며 얇은 실선과 점선은 각각 주성분분석과 독립성분분석을 통한 주요성분을 나타낸다. Fig. 5에서 주성분분석은 실제자료의 거동과 유사한 형태를 보이지만 잡음이 섞인 불규칙한 성분으로 추정이 되고 있다. 반면 독립성분분석은 실제자료와 거의 같은 거동을 나타내고 있으며 일관된 결과를 제시하여 주고 있고 자료를 재구성시 왜곡되지 않는다.

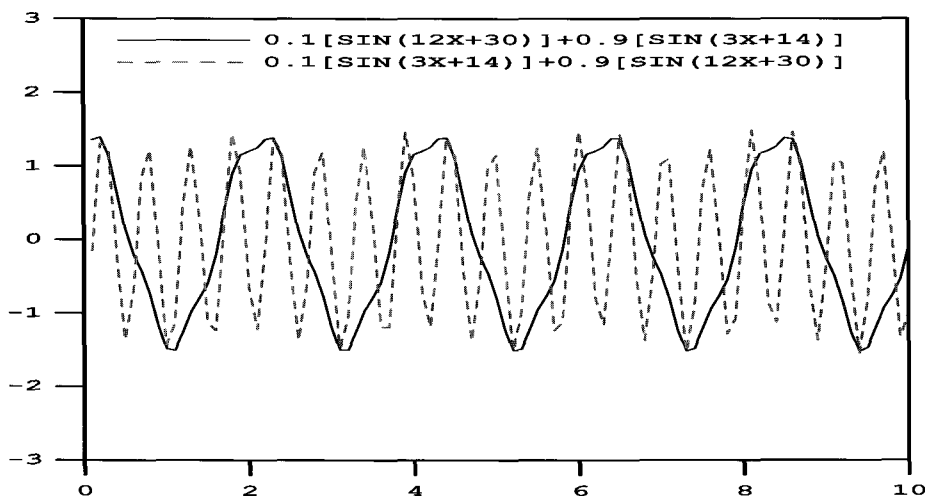


Fig. 3. Synthesize mixing data for experiment

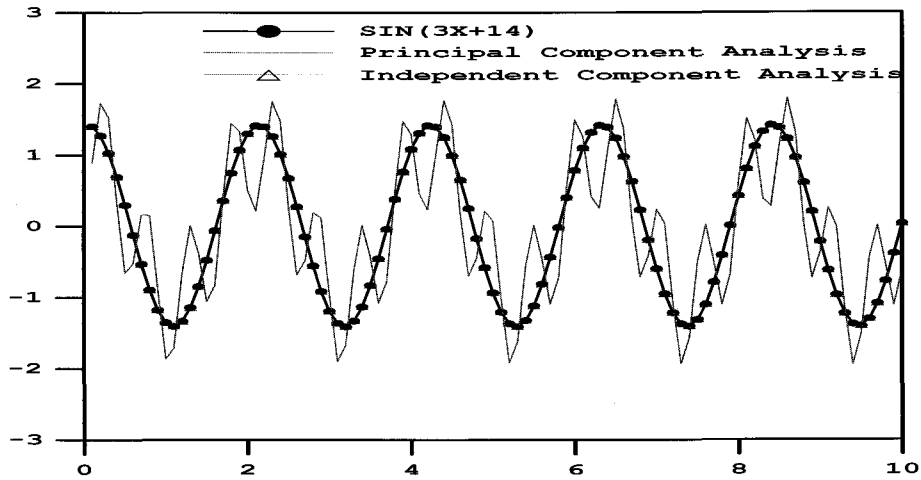


Fig. 4. Comparison of blind source separation of mixing data using PCA and ICA

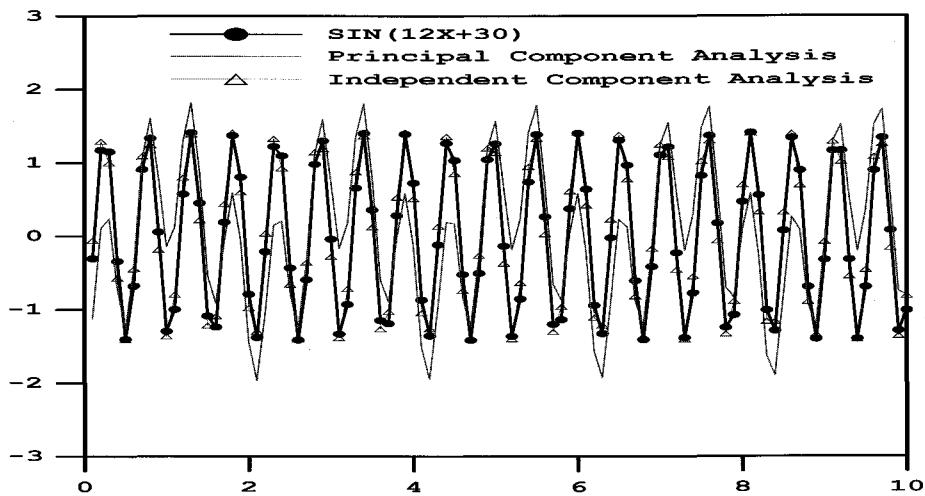


Fig. 5. Comparison of blind source separation of mixing data using PCA and ICA

자료간의 거동 특성을 평가할 수 있는 통계적지표인 상관계수를 가지고 판단해보면 주성분분석은 원자료와 0.77의 상관관계를 나타내고 있는 반면에 독립성분분석은 0.99의 상관계수를 갖는 거의 동일한 거동을 하는 성분으로 추출이 가능하였다.

Fig. 5는 주성분 $SIN(12X+30)$ 의 추정결과를 나타내며 앞선 예와 동일하게 주성분분석은 실제자료의 거동과 유사한 형태를 보이지만 잡음이 섞인 불규칙한 성분으로 추정이 되고 있으나 독립성분분석은 실제자료와 거의 같은 거동을 나타내고 있으며 일관된 결과를 제시하여 주고 있다. 상관계수의 경우에도 주성분분석은 원자료와 0.78의 상관관계를 나타내고 있는 반면에 독립성분분석은 0.99의 상관계수를 갖는 거의 동일한 거동을 하는 성분으로 추출이 가능하였다.

4. 독립성분분석의 수문시계열 적용

본 연구에서는 상대적으로 주기성이 강한 Nino3.4지

역 SST와 지역적인 편차가 존재하는 국내 9개 지점의 여름강수량을 대상으로 주성분을 분석하였다.

4.1 Nino3.4 지역 SST에 적용

본 연구에서는 El nino의 발생여부를 판단할 수 있는 대표적인 실제 수문기상자료인 Global Sea Surface Temperature(SST)의 주성분을 추정하여 El nino 발생과의 상관관계를 평가하였다. El nino의 해양성분에 대한 기준이 되는 SST를 측정하는 기준이 되는 지점들을 살펴보면, Nino12($0^{\circ}\sim 10^{\circ}\text{S}$, $90^{\circ}\text{W}\sim 80^{\circ}\text{W}$), Nino3($5^{\circ}\text{N}\sim 5^{\circ}\text{S}$, $150^{\circ}\text{W}\sim 90^{\circ}\text{W}$), Nino4($5^{\circ}\text{N}\sim 5^{\circ}\text{S}$, $160^{\circ}\text{E}\sim 150^{\circ}\text{W}$), Nino.4($5^{\circ}\text{N}\sim 5^{\circ}\text{S}$, $170^{\circ}\text{W}\sim 120^{\circ}\text{W}$)등이 있으며 본 논문에서는 Nino3.4지역(1856년~2002년)의 위·경도 $5^{\circ}\times 5^{\circ}$ 의 총 20개 Global Anomaly SST 격자자료를 대상으로 분석을 실시하였다.

El nino는 적도 동태평양의 해수면 온도의 공간평균값(Nino+3 지수 또는 Nino+3.4 지수)이 평상시보다 0.5

도 이상 따뜻한 상태가 5개월 이상 지속될 때를 El Niño 시기로 정의하며 La Niña 시기는 El Niño와 반대로 해수면 온도의 평균값이 0.5도 낮은 상태가 5개월 이상 지속되는 때로 정의된다.

비교방법은 총 20개의 격자를 갖는 SST자료를 독립 성분분석을 통하여 통계적으로 유의한 주성분을 추출한 후 미국 NOAA(National Oceanic and Atmospheric Administration)에서 정기적으로 분석하고 있는 Niño+3.4 Anomaly SST와 비교하여 분석하였다. Niño+3.4의 Anomaly SST 자료는 ENSO 사상을 판단하는 대표적인 지표로서 El Niño와 관련된 많은 연구에서 사용되는 기본 자료라 할 수 있으며 이 지표는 Niño+3.4(5°N~5°S, 170°W~120°W)지역의 SST를 공간적으로 평균하여 사용하고 있다.

본 연구에서는 20개의 차원을 갖는 자료계열에서 주

성분을 선택하기 위해서 Eigenvalue를 가지고 차원을 감소시켰다. Fig. 6은 자료계열에 대한 Eigenvalue를 도시하여 나타내었다. Fig. 6에서 1개의 Eigenvalue가 전체의 86%를 설명해주고 있으며 나머지 19개의 주성분은 14%를 설명하고 있다. 이러한 결과는 적도지역에 해당하는 20개 격자의 SST가 상대적으로 변동성이 적고 유사한 거동을 나타내는데 기인하는 것으로 사료된다. 따라서 본 연구에서는 1개의 독립성분만으로 충분히 자료의 변동성을 평가할 수 있을 것으로 판단되었으며 이를 토대로 ENSO 사상과의 상관관계를 평가하였다.

Fig. 7은 본 연구에서 추정된 독립성분과 NOAA에서 제공하는 Niño+3.4지역 SST 지수를 비교하여 나타내었다. 추정된 독립성분과 NOAA의 지표와 같은 거동을 보여주고 있으며 상관계수가 0.99로서 거의 같은 시계열 특성을 나타내고 있다.

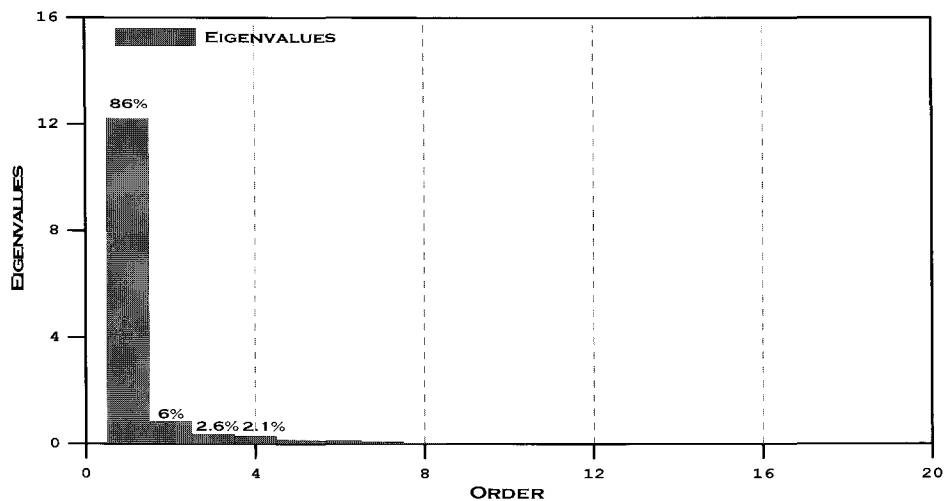


Fig. 6. Eigenvalue of each independent component of Niño3.4 grids data

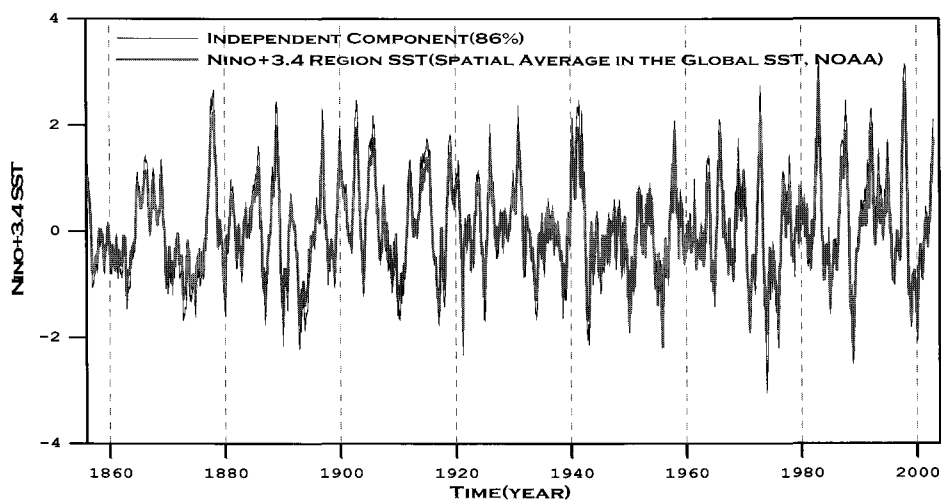


Fig. 7. WIC-1 of Niño3.4 SST grids data and Niño3.4 SST anomaly provided by NOAA

Fig. 8은 추정된 독립성분과 주요 ENSO 사상과의 적합성을 평가하기 위한 그림이며 삼각형과 역삼각형은 대표적인 El Niño와 La Niña 사상의 발생년도를 나타낸다. 앞서 언급했듯이 일반적으로 El Niño는 Niño+3 지수 또는 Niño+3.4 지수가 평상시보다 0.5도 이상 따뜻한 상태가 5개월 이상 지속될 때를 El Niño 시기로 정의하며 La Niña 시기는 El Niño와 반대로 해수면 온도의 평균값이 0.5도 낮은 상태가 5개월 이상 지속되는 때로 정의된다. 본 연구에서는 Kiladis and Diaz(1989), Meyers et al. (1997)이 제시한 ENSO 사상을 토대로 독립성분과 비교하였으며 Fig. 8에서 평상시보다 높은 해수면온도를 나타내는 시점과 주요 El Niño 발생시점이 일치하는 결과를 보이고 있으며 동일하게 평상시보다 해수면 온도가 낮은 시점과 La Niña 발생시점이 상

관관계를 가지고 발생함을 확인할 수 있다.

4.2 국내 주요지점 여름 강수량의 적용

독립성분분석을 국내 주요 9개 지점 여름 강수량에 적용하여 주성분을 검토하였다. 각 관측지점은 1954년 이후부터 2003년까지 결측 자료를 갖지 않는 서울, 인천, 부산, 대구, 목포, 강릉, 제주, 광주, 울산 등을 고려하였으며 6월부터 9월까지의 강수량을 합하여 총 50년의 자료로 구성하였다. 9개 지점에 대해서 독립성분분석을 적용한 결과 3개의 주성분이 전체 분산의 86%를 설명해주고 있으며 3개의 독립성분과 각 지점간의 상관계수를 추정하여 Table 2에 제시하였다. Fig 9부터 11은 각 독립성분과 이들 성분을 대표하는 강수지점을 나타내었다.

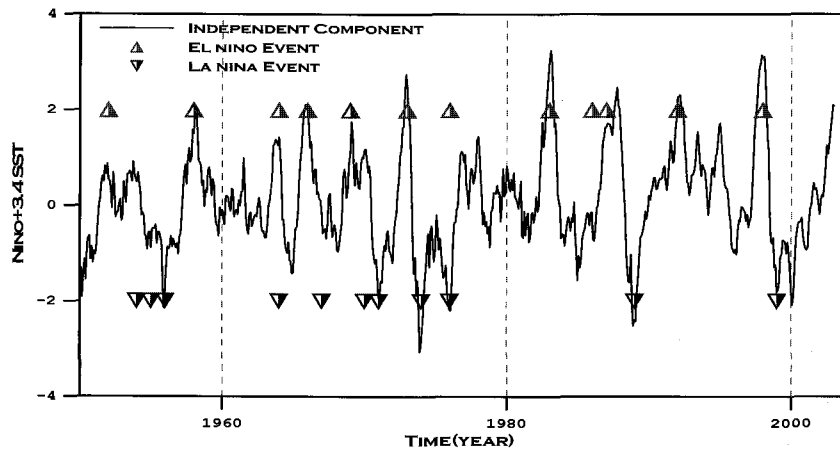


Fig. 8. Main ENSO events and WIC-1 of Niño3.4

Table 2. Correlation analysis between summer rainfall of main stations and WIC-1 using independent component analysis

	Seoul	Incheon	Busan	Mokpo	Daegu	Gwangju	Gangreng	Jeju	Ulsan
IC-1	57%	59%	90%	84%	84%	73%	59%	65%	85%
IC-2	78%	70%	-27%	-23%	-17%	-20%	29%	10%	-31%
IC-3	-11%	-11%	3%	-10%	-22%	-42%	-21%	72%	2%

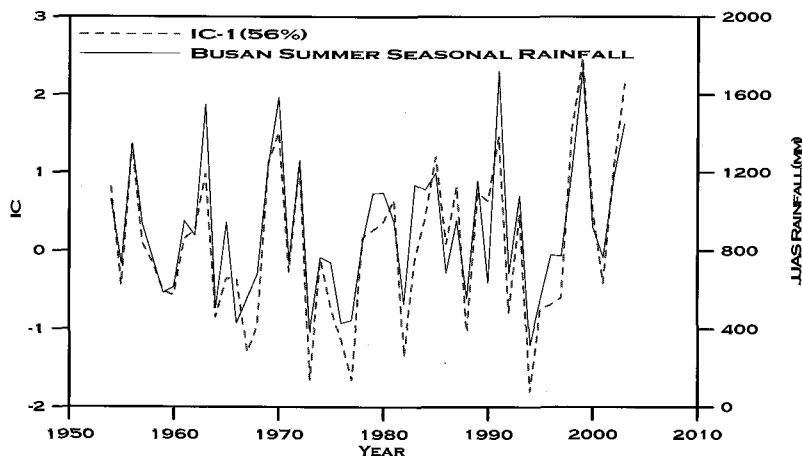


Fig. 9. Comparison between WIC-1 and Busan summer seasonal rainfall

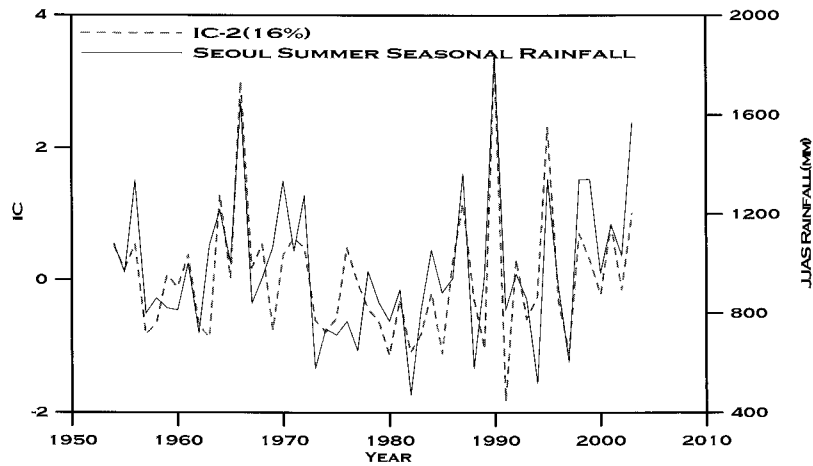


Fig. 10. Comparison between WIC-1 and Seoul summer seasonal rainfall

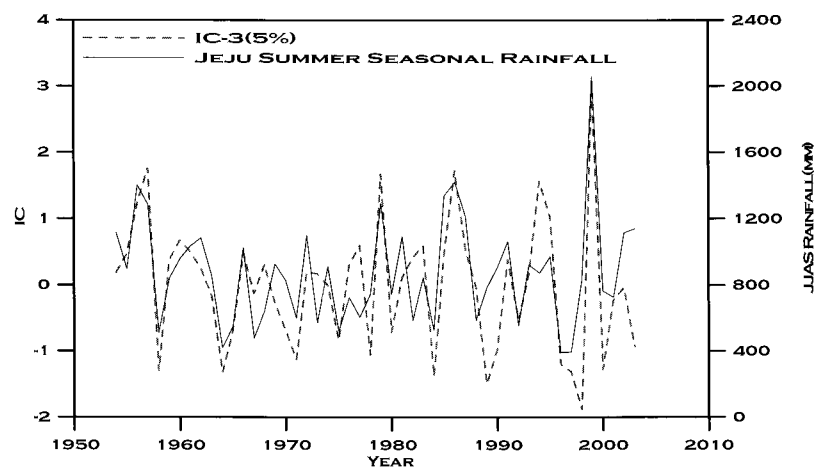


Fig. 11. Comparison between WIC-1 and Jeju summer seasonal rainfall

독립성분 1은 전체 분산의 56%를 차지하고 있으며 대부분 남부지방인 부산, 목포, 대구, 광주, 울산 등과 70%이상의 상관성을 나타내 주고 있다. 강릉의 경우 독립성분 1과 가장 큰 상관관계를 나타내지만 다른 지역에 비해 상대적으로 작은 상관성을 보이고 있으며 이는 지형적인 영향으로 인해 강릉만이 갖는 여름 강수량의 특성으로 사료된다. 두 번째 독립성분은 전체 분산의 16%를 차지하며 중부지방인 서울, 인천과 70%이상의 상관성을 보이고 있으며 이들 2지역을 대표하는 성분으로 사료된다. 마지막으로 독립성분 3은 전체 분산의 5%를 설명하고 있으며 제주지방만이 72%의 상관성을 보이고 있다. 이는 제주도의 여름 강수 특성을 고려한 결과로 판단된다. Fig. 9~11에서 IC-3을 제외하고 IC-1과 IC-2는 1970년 이후로 증가하는 경향을 보이고 있다.

5. 결 론

독립성분분석은 자료의 상관관계이외에 비정규성을

갖는 성분에 대해서도 비직교좌표를 이용하여 더 높은 차수의 상관관계를 제거할 수 있으며 최종적으로 차원들의 관계를 독립으로 변환시켜 독립성분을 추정할 수 있으며 주성분분석에 비해 근원성분에 가까운 주성분분리가 가능하다. 따라서 본 연구에서는 이러한 중요한 해석수단이 주성분분석의 특징 및 문제점을 검토하고 이를 보완할 수 있는 독립성분분석의 이론과 타당성을 검토하였으며 수문기상자료인 SST에 적용하여 ENSO 사상과의 상관성을 평가하였다. 본 연구를 통해 도출된 결론은 다음과 같다.

2가지 방법론의 이론적인 배경 및 특성을 검토하고 혼합자료를 이용하여 2가지 방법론의 주성분분리 능력을 검토한 결과 독립성분분석이 기존 주성분분석에 비해 통계적으로 우수한 결과를 나타내었다. 최종적으로 대표적인 수문기상자료이면서 El Niño의 감시구역인 Niño+3.4지역의 위·경도 $5^{\circ} \times 5^{\circ}$ 의 총 20개 Global Anomaly SST 격자자료를 대상으로 분석을 실시하여 El Niño-Southern Oscillation(ENSO)사상의 발생시기와

의 상관관계를 평가하였으며 한 개의 독립성분만으로 전체 SST의 전체적인 거동을 표현할 수 있었으며 주요 ENSO 발생년도와 거의 일치하는 결과를 나타내었다. 또한 국내 주요지점의 여름 강수량을 대상으로 독립성분분석을 적용한 결과 지역적인 특성을 고려하여 비교적 합리적으로 독립성분을 추출할 수 있었으며 IC-1과 IC-2에서 1970년대 이후로 여름강수량의 증가를 확인할 수 있었다. 따라서 독립성분분석을 수문자료의 다변량분석에서 이용한다면 주성분분석의 보조수단으로서 보다 통계적으로 유의한 분석이 가능할 것으로 사료되며 또한 수문자료와 기상자료간의 상관성 분석에서도 유용한 해석수단으로 적용이 가능할 것으로 판단된다.

참 고 문 헌

- 김진영, 문영일, 황성환 (2002). “남방진동지수와 서울 강수량과의 Mutual Information”, **한국수자원학회 학술발표회논문집**, pp. 1205-1210.
- 박대형, 황성환, 문영일 (2000). “서울 강수량과 남방진동지수간의 저빈도 관계 분석”, **한국수자원학회 학술발표회논문집**, pp. 341-346.
- 박상우, 전병호, 장석환 (2003). “다변량 분석기법에 의한 지점강우의 권역화 연구”, **한국수자원학회 논문집**, **한국수자원학회**, 제 36권, 5호, pp. 879-892.
- Bell, A.J., and Sejnowski, T.J. (1995). “An information-maximization approach to blind separation and blind deconvolution.” *Neural Computation*, vol. 7, No. 6, pp. 1129-1159.
- Common, P. (1994). “Independent Component Analysis, A New Concept?”, *Signal Processing*, Vol. 36, pp. 287-314.
- Hyvarinen, A. (1999). “Fast and Robust Fixed-Point Algorithms for Independent Component Analysis”, *IEEE Transaction on Neural Networks*, Vol. 10, No. 3, pp. 626-634.
- 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.
- Lee, T.W., Girolami, M., and Sejnowski, T.J. (1999). “Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources”, *Neural Computation*, vol. 11, No. 2, pp. 417-441.
- Meyers, S.D., O'Brien, J.J., and Thelin E. (1997). “Reconstruction of monthly SST in the Tropical Pacific Ocean during 1868-1993 using adaptive climate basis functions,” *Monthly Weather Review*, vol. 127, No. 7, pp. 1599-1612.
- Moon, Y.I., and Lall, U. (1996). “Atmospheric Flow Indices and Interannual Great Salt Lake Variability”, *Journal of Hydrologic Engineering*, vol. 1, pp. 55-62.
- Vautard, R., Yiou, P., and Ghil, M. (1992). “Singular Spectrum Analysis: a toolkit for short, noisy chaotic signals”, *Physica D*, 58, pp 95-126.

(논문번호:05-50/접수:2005.03.23/심사완료:2005.10.06)