

엘니뇨/남방진동(ENSO) 지수의 모형 및 예측

김형수 (선문대학교 건설공학부 전임강사)

안재현 (고려대학교 토목환경공학과 박사과정)

강두선 (세일종합기술공사 기술연구소 연구원)

1. 서론
 2. 남방진동지수
 3. 선형 추계학적 모형의 구축
 - 3.1 ARMA 모형
 - 3.2 잔차 분석
 4. BDS 통계
 5. Close Returns Test
 - 5.1 Close Returns Plot(CRP)
 - 5.2 SOI에 대한 CRP의 적용
 6. ARCH 모형의 적용
 7. ENSO 예측모형
 8. 결론
- 참고문헌

1. 서론

AR(autoregressive)이나 ARMA(autoregressive/moving-average)와 같은 선형 추계학적(linear stochastic) 모형들은 여러 분야에서 자연현상의 양상을 이해하고 해석하기 위하여 광범위하게 이용되어 왔으나 대부분의 실제 자연현상들은 비선형적인 특성을 가지고 있다. 따라서 좀더 정도 높은 자연현상의 이해를 위해 비선형 시계열(nonlinear time series)에 대한 연구가 진행되어 왔으며, 최근에 들어서는 통계학이나 동역학 분야에서 비선형 시계열 모형의 구축과 예측이 이루어지고 있다(Tong, 1990; Farmer와 Sidorowich, 1987).

본 연구에서는 전세계적인 기상재해를 발생시켜 갈수록 피해가 심각해져 가는 엘니뇨(El Niño) 현상의 심도를 나타내는 여러 가지 지수중에서 가장 일반적으로 사용되는 남방진동지수(SOI, Southern Oscillation Index) 자료에 대한 분석을 실시하였다. 즉, SOI 자료가 어떤 시계열적인 특성을 가지고 있는가를 분석하여 적절한 추계학적 모형을 구성하였다. 또한 모형의 적합성 여부를 판단해본 후에, 그 자료가 가지고 있는 선형 혹은 비선형적인 특성에 대한 파악을 통해 적절한 예측모형의 구축여부를 판단하였다.

이를 위해 SOI 시계열을 선형 모형인 ARMA 모형에 적합시킨 후 기존의 통계기법인 ACF와 Porte Manteau test를 이용하여 그 적합성 여부를 판단하였으며, 새로운 통계기법인 BDS 통계통계를 통해 그 적합성 여부를 검증하였다. 또한 SOI 시계열의 특성을 파악하기 위하여 스펙트럼 분석을 실시하였는바 카오스 특성의 가능성이 있는 것으로 판단되어 Close Returns Test 기법을 이용, SOI 시계열의 카오스 특성 여부를 검토하였다. 최종적으로 비선형 추계학적 모형의 적용을 위해 ARCH(autoregression conditional heteroscedasticity) 모형에 대한 SOI 시계열의 적합을 시도함으로써 적절한 모형 구축여부를 판단하였고, 현재 연구되고 있는 예측 모형에 대하여 간략하게 기술하였다.

2. 남방진동지수

스페인어로 '남자아이'라는 뜻을 지닌 엘니뇨는 1726년부터 관찰되고 기록되어진 현상(Cane, 1983)으로 원래는 매년 크리스마스 시즌에 남미 페루연안의 해수면 온도가 올라가는 계절적인 특성으로서 바닷물의 온도가 올라가면 연안의 바다에서 물고기떼가 다른 지역으로 이동하고 비가 많이 내리는 현상을 말한다. 그러나 최근 들어서는 그 개념이 바뀌어서 언제라도 동태평양 적도지방의 해수면 온도가 평년보다 0.5℃ 이상 높아지는 현상을 엘니뇨라 부르며, 반대로 0.5℃ 이상 낮은 경우에는 '여자아이'란 뜻의 라니냐(La Niña)로 부른다. 또한 엘니뇨가 나타나는 지역의 범위도 달라져서 최근에는 페루 연안에 국한된 것이 아니라 페루 연안에서 날짜변경선까지 약 1만 km에 달하여 발생하는 매우 큰 규모의 기상 이변 현상으로 알려지고 있다.

엘니뇨가 발생하게 되는 가장 큰 원인은 바람인 것으로 알려져 있다. 평년에는 열대 태평양에서 바람(무역풍)이 동쪽에서 서쪽으로 불게되며 그 결과로 서태평양에서는 따뜻한 바닷물이 쌓여서 서태평양의 인도네시아 연안은 동태평양의 남미 지역보다 바닷물의 높이 및 온도가 상승하게 된다. 따라서 서태평양은 전세계적으로 바닷물의 온도가 가장 높은 지역이 되며 반대로 남미 연안에서는 서쪽으로 쓸려간 바닷물을 보충하기 위해서 바다밑의 차가운 해수가 솟아오르는 용승(upwelling) 현상이 일어나면서 해수의 온도가 낮아지게 되는 라니냐 현상이 발생하게 된다. 그러나 엘니뇨 해에는 바람의 강도가 약해지며 반대의 현상이 발생한다. 즉, 서태평양으로 이동하는 따뜻한 바닷물의 움직임은 약해지고, 이에 따라 동태평양의 바닷물 용승현상이 약해지며 적도 부근에서 가열된 따뜻한 바닷물이 정체되어 물의 온도가 올라가는 엘니뇨현상이 일어나게 되는 것이다. 이러한 해수 온도의 변화는 평년에 서태평양 지역에 형성되는 강우지역을 동쪽으로 이동시켜서 동태평양 부근의 기압을 떨어뜨리고 반면에 서태평양 부근의 기압을 상승시키게 되어 여러 가지 이상기후를 야기하는 원인이 되는 것이다. 하지만 최근들어 엘니뇨가 좀더 복잡하게 지구의 기상 및 수문 현상과 연관되어 있다고 알려지고 있으며(원격관계, Teleconnection) 그에 대한 연구가 계속적으로 진행되고 있다.

영국의 기상학자인 Walker(1923)는 인도에서 아시아 몬순에 관한 연구를 위하여 기상 및 해양 자료들을 수집하여 분석하던 중에 남방진동(Southern Oscillation) 현상을 발견하였다. 남방진동이란 적도부근 남태평양 타히티(Tahiti)섬과 오스트레일리아 북부 다아윈(Darwin) 지역에서 발생하는 기압의 시소(Seesaw)현상으로 타히티의 기압이 올라가면 다아윈의 기압이 낮아지고, 반대로 타히티의 기압이 낮아지면 다아윈의 기압이 올라가는 현상을 말한다. Walker의 발견 이후 수십년이 지나서 Bjerknes(1966)에 의해 남방진동이 엘니뇨현상과 밀접한 관계가 있음이 처

음으로 알려졌다.

일반적으로 엘니뇨와 그때의 남방진동 현상을 ENSO(El Niño/Southern Oscillation)라고 하며 이것을 정량적으로 지수화한 것이 SOI로서 이는 월별로 각각 표준화(standardization)된 타히티의 기압에서 다아윈의 기압을 빼줌으로써 구할 수 있게 된다(신현석, 1997). 대부분 4년에서 7~8년 사이인 것으로 알려져 있는 엘니뇨의 발생주기에 대해 1900년 이후부터 1991년까지의 SOI 자료를 분석해보면 지금까지 총 21회의 엘니뇨가 발생한 것으로 조사되었다(Ropelewski와 Halpert, 1987; Chu, 1995).

본 연구에서는 1932년 9월부터 1998년 5월까지 총 789개의 월별 SOI를 이용하여 분석을 실시하였다. 그림 2.1은 분석대상기간의 월별 SOI를 도시한 것으로서 음(-)의 값으로 커질수록 엘니뇨의 정도가 심함을 의미하는데, 엘니뇨가 극심했던 1982~83년이나 1997~98년의 경우에 음(-)으로 큰값을 보여주고 있음을 알 수 있다.

또한 그림 2.2는 SOI 시계열에 대한 파워 스펙트럼(power spectrum)으로 $f=0.0127 \sim 0.0203(1/\text{months})$ 에서 첨두값을 가짐을 알 수 있는데 이를 주기로 환산하면 대략 49개월~79개월로서 일반적으로 알려진 엘니뇨의 주기인 4~7년을 잘 반영하고 있다. 그러나 첨두값의 발생이후에는 특별한 주기가 없이 거의 무작위(random)한 특성을 갖는 시계열의 양상을 보여주고 있음을 알 수 있다. Berge 등(1984)에 따르면 카오스의 정의가 “파워 스펙트럼이 첨두의 존재에도 불구하고 연속적으로 넓게 분포할 때 혹은 시계열의 자기상관함수가 유한시간에 0으로 될 때”이므로 그림 2에서 SOI의 파워 스펙트럼이 연속적인 형태를 보이는 특성이 나타남을 고려하여 본 연구에서는 Close Returns Test(Gilmore, 1993a; 1993b)를 이용하여 카오스 분석을 실시하였다.

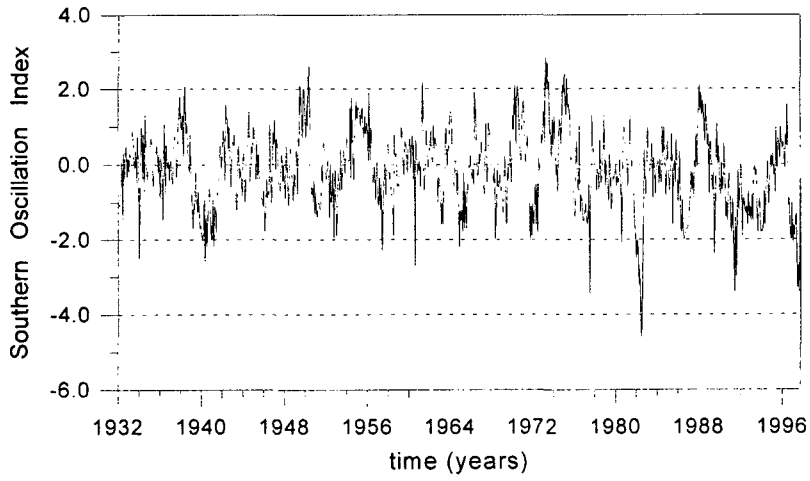


그림 2.1. SOI 시계열(Sep. 1935 ~ May 1998)

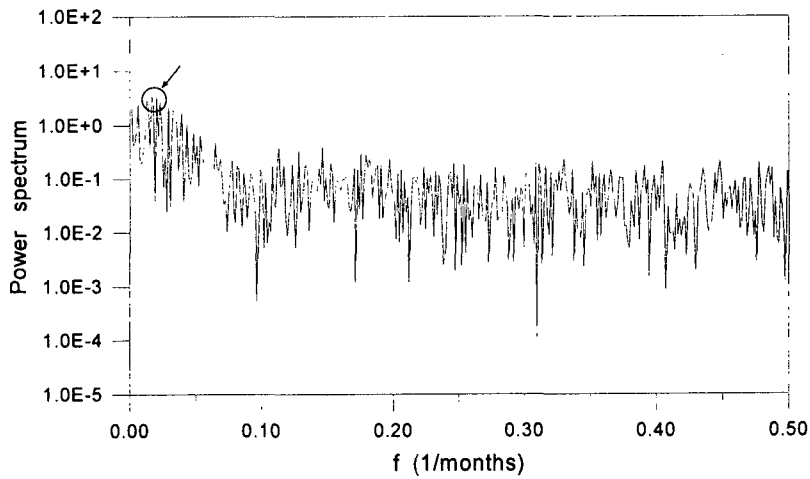


그림 2.2. SOI의 Power Spectrum(Sep. 1935 ~ May 1998)

3. 선형 추계학적 모형의 구축

3.1 ARMA 모형

SOI에 대한 추계학적 예측모형의 연구는 Chu와 Katz(1985)에 의해 되어진 바 있다. 그들은 1935년 1월부터 1983년 8월까지의 SOI 시계열 자료를 이용해서 월별 ARMA 모형과 계절별 ARMA 모형을 수립하여 적용가능성을 검토한 바 있다. 특히 월별 ARMA 모형의 경우는 특이하게도 ARMA(1,7;1)의 차수를 가지는 $X_t = \phi_1 X_{t-1} + \phi_7 X_{t-7} + a_t + \theta_1 a_{t-1}$ 모형이 적합한 모형으로 선정되었는데, 여기서 $\phi_1 = 1.011$, $\phi_7 = -0.115$, $\theta_1 = -0.680$ 이며, a_t 는 평균 0과 표준편차 σ_a 로서 $a_t \sim [0, \sigma_a^2]$ 인 백색잡음(white noise)이며 $\sigma_a^2 = 1.481$ 이다. 본 연구에서는 Chu와 Katz(1985)의 구와 같은 방법으로 SOI에 대한 ARMA 모형의 수립을 시도하였다.

먼저 적절한 모형의 차수를 결정하기 위하여 자기상관함수(autocorrelation function, ACF)와 편자기상관함수(partial autocorrelation function, PACF)를 그림 3.1 및 3.2에 도시하였다.

자기상관함수와 편자기상관함수로부터 판단해보면 시계열은 AR(p)과 MA(q)의 차수가 각각 10차와 3차 이내일 것으로 판단되었다. 따라서 각각의 항에서 적절한 차수를 결정하기 위하여 Schwarz(1978)가 제안하고 Chu와 Katz(1985)가 모형의 차수를 결정하기 위하여 사용하였던 BIC(Bayesian Information Criterion)를 본 연구에서도 모형 결정의 기준으로 이용하였다. 식 (3.1)과 같이 정의되는 BIC는 BIC(p,q)가 최소가 될 때의 p와 q를 선택해서 모형의 차수로 결정한다.

$$\text{BIC}(p ; q) = \ln [(\hat{\sigma}_a^2)^n] + r \frac{\ln(n)}{n} \quad (3.1)$$

여기서 $(\hat{\sigma}_a^2)^n$ 는 σ_a^2 의 최우도(maximum likelihood)이며, $r = p + q + 1$ 은 모형에서 사용되는 매개변수의 수를 나타낸다.

표 3.1과 같이 각각의 차수에 대해 BIC를 산정하였으며, 이를 통해 BIC가 가장 작은 값을 보이는 ARMA(1;1) 모형이 적합한 것으로 나타났다.

또한 추가적으로 좀더 정확한 모형의 수립을 위해 복잡한 ARMA 모형을 이용하였다. 즉, 위에서 선정된 ARMA(1;1) 모형의 특성을 거의 변화시키지 않는 한도내에서 하나의 고차 AR 항(지체 3, 4, ..., 9)을 추가하였다(Chu와 Katz, 1985).

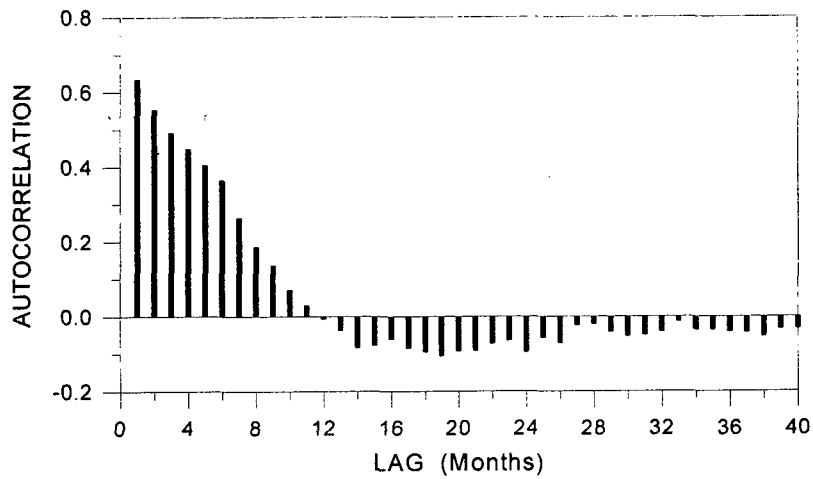


그림 3.1 SOI의 자기상관함수

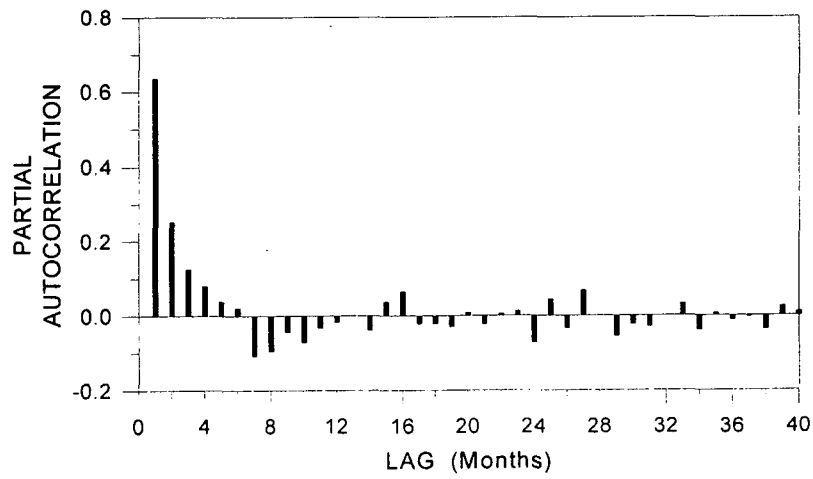


그림 3.2 SOI의 편자기상관함수

표 3.1 ARMA(p ; q) 모형의 차수 결정

p	q	BIC	WN(σ_a^2)	p	q	BIC	WN(σ_a^2)
0	1	2101.652	0.8331	6	0	1877.339	0.6052
	2	2009.568	0.7356		1	1882.419	0.6055
	3	1963.072	0.6885		2	1882.284	0.6019
1	0	1925.398	0.6657	3	1884.061	0.5998	7
	1	1859.443	0.6074	0	1872.951	0.5982	
	2	1865.001	0.6074	1	1865.862	0.5892	
2	3	1869.466	0.6068	2	1870.672	0.5895	8
	0	1878.038	0.6220	3	1874.454	0.5891	
	1	1865.246	0.6076	0	1870.336	0.5927	
3	2	1870.220	0.6073	1	1869.942	0.5889	9
	3	1875.146	0.6072	2	1873.964	0.5887	
	0	1870.135	0.6114	3	1878.342	0.5888	
4	1	1869.781	0.6070	0	1873.203	0.5914	10
	2	1875.166	0.6073	1	1877.794	0.5916	
	3	1879.992	0.6073	2	1878.172	0.5887	
5	0	1869.616	0.6069	3	1882.545	0.5889	10
	1	1873.546	0.6060	0	1873.782	0.5886	
	2	1878.710	0.6063	1	1878.533	0.5890	
5	3	1864.560	0.5917	2	1881.568	0.5881	10
	0	1872.907	0.6055	3	1885.915	0.5883	
	1	1877.931	0.6057				
	2	1881.364	0.6047				
	3	1869.164	0.5918				

표 3.2 ARMA(1,p ; 1) 모형의 차수결정

p	q	BIC	WN(σ_a^2)
3	1	1864.476	0.6070
4	1	1863.983	0.6066
5	1	1865.043	0.6074
6	1	1862.187	0.6052
7	1	1849.333	0.5954
8	1	1849.192	0.5952
9	1	1852.958	0.5981

이렇게 추가한 AR항에 따라 앞에서와 같이 BIC를 산정하여 표3.2에 나타내었는데 단순한 ARMA(1;1)모형에 비해 훨씬 작아지는 BIC의 값들을 발견할 수 있다. 특히 7차 혹은 8차에서 아주 좋은 결과를 나타내었고, 좀더 나은 값을 보이는 8차가 최종적으로 선정되었다. 이하 ARMA(1,8;1) 또는 일반적인 표기를 위하여 ARMA(1,p;1)로 표시하기로 한다.

따라서, 결정된 모형은 ARMA(1,8;1) 모형이며 식 (3.2)와 같이 나타낼 수 있다.

$$x_t = \phi_1 x_{t-1} + \phi_8 x_{t-8} + a_t + \theta_1 a_{t-1} \quad (3.2)$$

여기서 매개변수의 추정에는 Hannan 과 Rissanen(1982)의 방법을 이용하였으며, 추정된 매개변수는 $\phi_1=0.9613$, $\phi_8=-0.0768$, $\theta_1=-0.5852$, $a_t \approx N(0, \sigma_a^2)$ 이다. 이러한 ARMA(1,8;1) 모형을 분석해 보면 월별 SOI는 1개월전과 8개월전 SOI와 무작위항(random noise term)의 선형적인 조합으로 이루어짐을 알 수가 있다. 즉, SOI는 1개월과 8개월전의 SOI 및 당월의 무작위항과 1개월전의 무작위항에 의존한다는 것이다.

3.2 잔차 분석

선정된 모형의 적합성을 검토하는 방법은 잔차(residual)의 무작위성(randomness) 여부를 판단하는 것으로 잔차가 백색잡음(white noise)과 같이 완전하게 무작위하다면 그 모형은 적합하게 선정된 것이다. 이러한 잔차의 무작위성 여부를 판단하는 것은 크게 두가지 방법이 있다.

첫째는 잔차의 자기상관함수(ACF) 및 편자기상관함수(PACF)를 산정한후 모든 지체(lag)에서 ACF 및 PACF가 충분히 0에 가까운가를 확인해서 만약에 충분히 0에 가깝다면 잔차는 무작위한 것이며 선정된 모형은 적합한 것이된다.

둘째는 Porte Manteau test로서 검정통계량 Q를 식(3.3)과 같이 산정한다.

$$Q = n(n+2) \sum_{k=1}^K r_k^2 / (n-k) \quad (3.3)$$

여기서 n은 ARMA 모형의 선정에 쓰인 자료수이고, r_k 는 지체 k에서 잔차의 자기상관계수이며 K는 충분히 큰 지체로서 일반적으로 $K=n/5$ 를 사용한다. 만약 귀무가설 “적합한 ARMA 모형이 선정되었다”가 사실이라면 Q는 근사적으로 자유도가 (K-p-q)인 χ^2 (Chi-squares) 값 보다 작은 값을 가지게 된다.

선정된 ARMA(1,8;1) 모형의 적합성 여부를 판단하기 위하여 앞에서 언급한 두가지의 방법

을 모두 적용하였다. 먼저 잔차의 ACF 및 PACF를 그림 3.3 및 3.4와 같이 산정하여 도시하였는데 모두 충분히 0에 가까운 값을 나타내고 있음을 알 수 있다. 또한 Porte Manteau test를 위해서 Q와 χ^2 값을 산정하였으며 Q=17, $\chi^2=20$ 을 얻을 수 있었는데 $Q < \chi^2$ 이므로 적합한 것으로 판단되었다.

따라서 ACF 및 PACF와 Porte Manteau test를 통해 잔차의 무작위성을 확인할 수 있었으며, 이를 통해 선정된 ARMA(1,8;1) 모형이 적합함을 알 수 있었다.

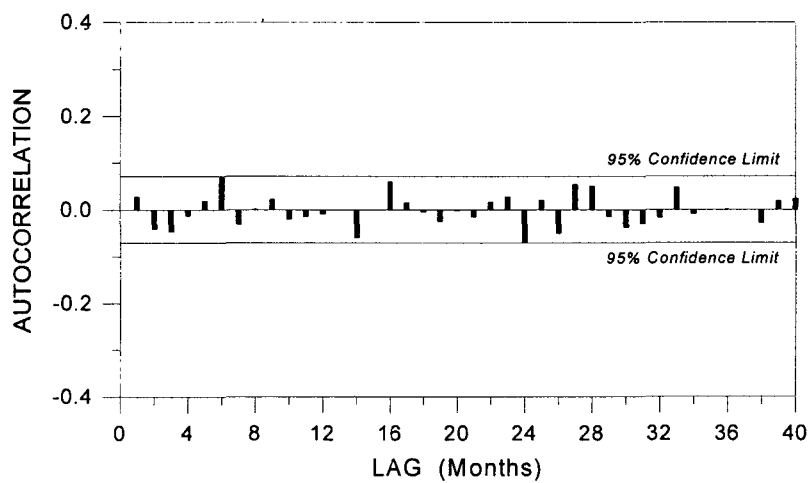


그림 3.3 잔차의 자기상관함수

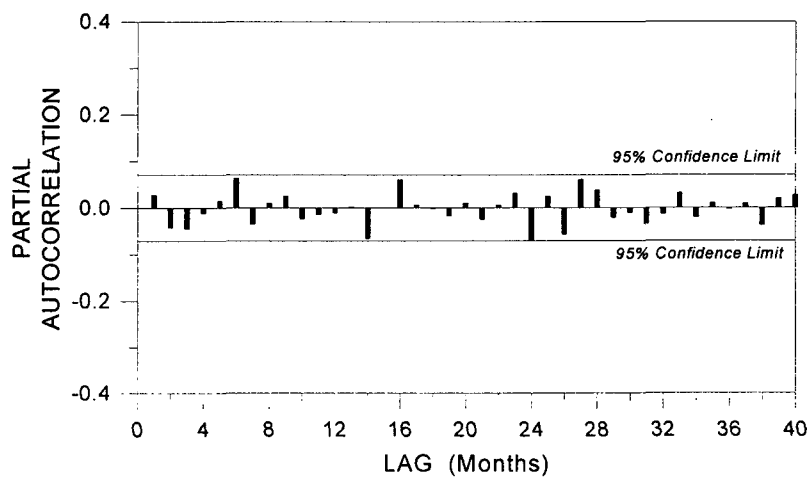


그림 3.4 잔차의 편자기상관함수

4. BDS 통계

기존의 통계기법들이 밝혀내지 못하는 시계열의 비선형성을 파악할 수 있는 BDS 통계 (Brock 등 1987)는 상관적분(correlation integral)에 기초한 검정방법으로서 시계열 자료가 iid(independently and identically distributed)를 따른다는 가설의 검증을 통해서 자료의 무작위 여부를 결정하는 방법이다. 특히, 이 방법은 무작위한 시계열자료와 카오스나 추계학적 비선형 계를 구분하는데 매우 유용한 통계기법이다.

Packard 등(1980)에 따르면 시간지체법(method of delays) (식 (4.1)를 참고)을 이용하여 스칼라(scalar) 시계열을 m차원 공간상의 상태벡터 점(state vector point)들로 나타낼 수 있다.

$$\vec{x}_i = (x_t, x_{t+\tau_d}, x_{t+2\tau_d}, \dots, x_{t+(m-1)\tau_d}), \quad x_i \in R^m \quad (4.1)$$

여기서 τ_d = 지체시간(delay time), m = m차원(embedding dimension)

무작위한 시계열은 어느 지체시간(delay time) τ_d 에서도 ACF는 0일 것이다. 그러므로 $\tau_d=1$ 일 경우의 ACF도 0일 것이며, 따라서 시계열의 무작위성 여부를 판단하기 위해서 위의 식 (4.1)을 이용할 때에 τ_d 는 1을 사용하여도 무방할 것이므로 식(4.1)은 다음과 같이 표시할 수 있게 된다.

$$\vec{x}_i = (x_t, x_{t+1}, x_{t+2}, \dots, x_{t+(m-1)}), \quad x_i \in R^m \quad (4.2)$$

BDS 통계를 위한 상관적분은 임의의 상태벡터 점을 중심으로 일정한 거리(r) 안에 있는 상태벡터 점들의 수를 계산한 값으로서 식(4.3)과 같이 나타낼 수 있다(Grassberger 와 Procaccia, 1983).

$$C(m, N, r) = \frac{2}{M(M-1)} \sum_{1 \leq i < j \leq M} \Theta(r - \|\vec{x}_i - \vec{x}_j\|), \quad r > 0 \quad (4.3)$$

$$\begin{aligned} \Theta(a) &= 0, \quad \text{if } a < 0 \\ \Theta(a) &= 1, \quad \text{if } a \geq 0 \end{aligned}$$

여기서 N=사용된 자료의 수, M=N-(m-1)은 m차원(embedding dimension) 공간상에서 상태벡터 점들의 수이다.

시계열 자료가 무작위한 분포를 따른다는 가설하에 $m > 1$ 인 경우의 BDS 통계는 다음과 같은 식으로 나타낼 수 있다.

$$BDS(m, M, r) = \frac{\sqrt{M}}{\sigma} [C(m, M, r) - C^m(1, M, r)] \quad (4.4)$$

여기서

$$\begin{aligned} \sigma^2(m, M, r) = & 4 m(m-1)C^{2(m-1)}(K - C^2) + K^m - C^{2m} \\ & + 2 \sum_{i=1}^{m-1} [C^{2i}(K^{m-i} - C^{2(m-i)}) - mC^{2(m-i)}(K - C^2)] \end{aligned} \quad (4.5)$$

$$K(m, M, r) = \frac{6}{M(M-1)(M-2)} \sum_{1 \leq i < j \leq M} [\Theta(r - \|\vec{x}_i - \vec{x}_j\|) \times \Theta(r - \|\vec{x}_j - \vec{x}_k\|)] \quad (4.6)$$

식 (4.4)에서 $\sqrt{M}[C(m, M, r) - C^m(1, M, r)]$ 은 $m \rightarrow \infty$ 일때 평균인 0인 정규분포(normal distribution)로 근접하게 되고 이때 정규분포의 분산이 식 (4.5)의 $\sigma^2(m, M, r)$ 이며, 식 (4.6)를 통해 계산에 필요한 K값을 산정하게 된다.

식(4.4)의 BDS 통계는 확정론적 비선형계(nonlinear deterministic system)와 추계학적 비선형계(nonlinear stochastic system)를 구분할 수는 없으나, 카오스나 추계학적 비선형계로부터 발생된 시계열과 무작위 시계열 자료를 구분할 수 있으며, 이것은 기존의 통계방법들이 밝히지 못하는 시계열의 무작위성 여부를 판단할 수 있는 아주 강력한 도구이다.

원래 확정론적 시계열로서 무작위하지 않지만 기존의 통계기법으로 분석할 때 무작위하다는 결과를 보여주는 Logistic map 자료를 BDS 통계로 판단해보면, 표 4.1에서와 같이 95% 유의수준(-1.96 ~ 1.96)에서 기각되는 값을 산정하며, 따라서 Logistic map이 무작위한 특성을 갖지 않은 시계열이라는 것을 BDS 통계만이 밝히고 있음을 알 수 있다(Brock 등, 1991). 또한 그 무작위성이 확인된 Gaussian noise의 경우는 BDS 통계에서도 무작위하다는 가설이 받아들여짐을 보여주고 있다. BDS 통계를 실시하기 위해서 필요한 m값은 $2 \leq m \leq 5$ 의 네개의 정수값이, r값은 $0.5\sigma \sim 2\sigma$ 사이의 값이 적절한 범위로 추천되며 (Brock 등, 1991, 1996), 여기서 σ 는 자료의 표준편차이다. 따라서 본 연구에서는 $m=2, 3, 4, 5$ 및 $r=0.5\sigma, 1.0\sigma, 1.5\sigma, 2.0\sigma$ 를 선정하여 각각의 경우에 대해 BDS 통계치를 산정하였다. SOI 시계열 자료와 잔차에 대해 산정된 BDS 통계치를 표 4.2에 수록하였으며 95% 유의수준으로 가설의 적정성 여부를 판단하였는데, 먼저 SOI의 경우에는 모든 경우에 있어서 가설에 위배되는 값이 산정되었고, 잔차의 경우도 $m=2, r=0.5\sigma$ 일때를 제외하고는 모든 경우에서 가설에 위배되는 값이 산정되었다.

표 4.1 BDS 통계의 비교 검증

r	Embedding dimension (m)	Logistic map		Gaussian noise	
		Value of statistic	Null of Hypothesis ($\alpha=0.05$)	Value of statistic	Null of Hypothesis ($\alpha=0.05$)
1.0 σ	2	199.66	Rejected	0.1879	Accepted
	3	190.73	Rejected	0.6013	Accepted
	4	182.45	Rejected	0.3538	Accepted

[95% 유의수준($\alpha=0.05$) : -1.96 ~ 1.96]

표 4.2 BDS 통계에 의한 검정결과

r	Embedding dimension (m)	SOI data		Residual data	
		Value of statistic	Null of Hypothesis ($\alpha=0.05$)	Value of statistic	Null of Hypothesis ($\alpha=0.05$)
0.5 σ	2	24.0512	Rejected	1.5745	Accepted
	3	26.7669	Rejected	2.3484	Rejected
	4	29.4167	Rejected	2.6349	Rejected
	5	31.7034	Rejected	4.1547	Rejected
1.0 σ	2	23.5987	Rejected	1.9746	Rejected
	3	25.8505	Rejected	2.5409	Rejected
	4	27.5752	Rejected	2.7861	Rejected
	5	29.1348	Rejected	2.6752	Rejected
1.5 σ	2	22.5198	Rejected	2.1932	Rejected
	3	24.1542	Rejected	2.6637	Rejected
	4	24.9882	Rejected	2.9192	Rejected
	5	25.5016	Rejected	2.8808	Rejected
2.0 σ	2	22.2353	Rejected	2.7024	Rejected
	3	23.6704	Rejected	2.9521	Rejected
	4	23.8717	Rejected	3.2617	Rejected
	5	23.5495	Rejected	3.1440	Rejected

[95% 유의수준($\alpha=0.05$) : -1.96 ~ 1.96]

즉, BDS 통계를 통해서 SOI 및 ARMA 모형의 잔차가 무작위한 특성을 가지지 않는 것으로 나타났는데, 이것은 SOI 자료가 가지는 비선형적인 특성에 기인한 것으로 보이며, 따라서 선형 추계학적 모형(linear stochastic model)인 ARMA 모형은 본 자료의 경우에는 적합하지 않는 것으로 판단되었다.

5. Close Returns Test

소멸하는 동력계 (dissipative dynamic system)가 카오스적인 거동을 보이면 그 시스템은 위상 공간에 이상한 어트랙터(strange attractor)를 가지게 된다(이재수, 1994). Gilmore(1993a, 1993b)에 의해 제안된 Close Returns Test는 비선형 동역학계에서 이상한 어트랙터를 찾기위한 간단한 위상학적인 접근방법으로서 지체시간의 산정을 통한 어트랙터의 도시를 하지 않고서도 시계열의 카오스적인 특성을 찾아낼 수가 있으며, 또한 상대적으로 다른 방법에 비해 적은 수의 자료를 이용해서 필요한 해석을 실시할 수 있다는 장점이 있다.

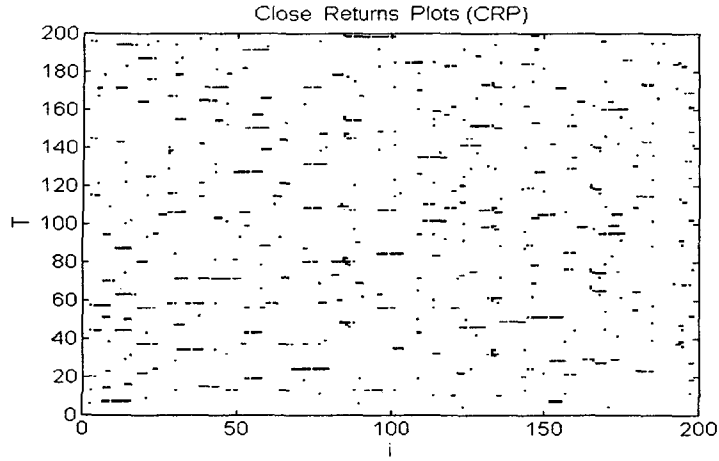
4장에서 SOI는 비선형적인 특성을 갖고 있음을 알 수 있었으나 추계학적(stochastic) 특성과 확정론적(deterministic) 특성중 어느 특성을 가지고 있는지는 판단할 수 없었다. 따라서 SOI가 가지는 추계학적 비선형성(stochastic nonlinearity) 혹은 확정론적 비선형성(deterministic nonlinearity), 즉 카오스(Chaos)의 존재여부를 판단하기 위하여 Close Returns Test를 실시하였다.

5.1 Close Returns Plot(CRP)

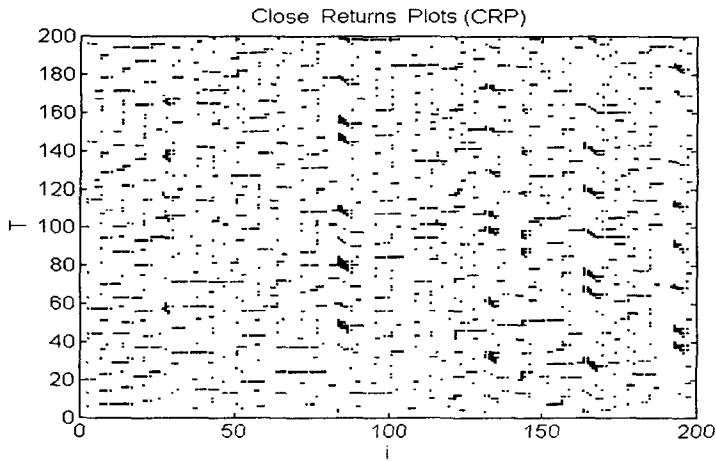
이상한 어트랙터의 중요한 특성은 불안정한 주기의 궤도(orbis)들로 채워져 있다는 것인데, 궤적(trajecory)이 이러한 궤도들 중의 하나를 아주 가깝게 지나가게 된다면, 짧은 시간동안 서로 매우 근접해서 머무르게 된다. 이것은 카오스 시계열에서 주기의 간격이 짧아지게 만드는데 추계학적 시계열에서는 이러한 특성을 보이지 않는다.

시계열의 어떤 점인 x_i 가 불안정한 주기의 궤도근처에 위치한다면, 연속하는 다른 점들도 지수적인 발산에 의해 궤적이 멀어지는 순간까지는 궤도근처에 머물게 된다. 만약 궤적이 불안정한 주기의 궤도근처에 충분히 오랫동안 머물수 있다면, 그것은 적당한 시간 T가 지난 후에 x_i 의 근처로 되돌아갈 것이다. 이런 경우에 두 점의 차인 $\delta = |x_i - x_{i+T}|$ 는 연속적인 값 i에서 매우 작은 값을 가질 것이며, 이것은 $\delta < r$ 일때의 점들을 i와 T축에 도시함으로써 보여질 수 있다. 이런 그림에서 주기성의 간격은 수평적으로 끊어지는 직선들을 그릴 것이며, 따라서 Close Returns Plot(CRP)에서 수평으로 끊어지는 직선의 존재유무를 통해 추계학적 특성과 카오스적 특성을 구분할 수 있는 것이다. 일반적으로 $r=0.01 \sim 0.1$ 의 값을 가지며 $r=0.05$ 라면 이는 $0.05(x_{max} - x_{min})$ 을 나타낸다.

그림 5.1은 카오스의 특성을 가지는 시계열인 Hennon map 자료에 대하여 $r=0.05$ 및 0.1일때의 CRP를 도시한 것으로서, 수평으로 끊어지는 직선이 불규칙하게 존재함을 확인할 수 있으며 이를 통해 CRP가 카오스 자료의 특성을 반영함을 알 수 있다.



(a) $r=0.05$

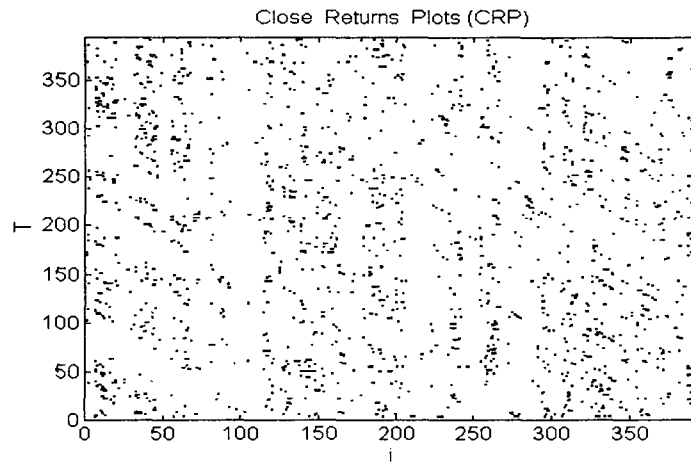


(b) $r=0.1$

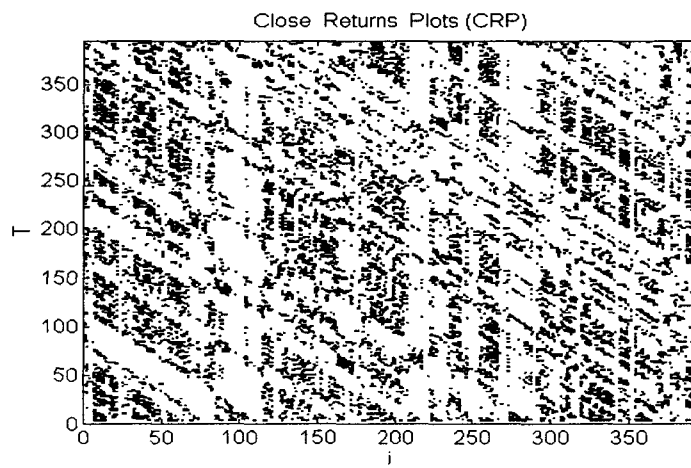
그림 5.1 Hennon map의 Close Returns Plot

5.2 SOI에 대한 CRP의 적용

SOI의 CRP는 그림 5.2와 같으며 각각 $r=0.05$ 와 0.1 의 결과에 대하여 도시하였다. 그림 5.2를 살펴보면 각각의 경우에 대하여 수평으로 끊어지는 직선이 존재하지 않는 것을 볼 수 있으며 따라서 CRP로 판단해볼 때 SOI는 카오스적인 특성을 가지고 있지 않는 것으로 판단된다.



(a) $r=0.05$



(b) $r=0.1$

그림 5.2 SOI의 Close Returns Plot

6. ARCH 모형의 적용

4장에서 BDS 통계를 이용한 분석 결과에 따르면 SOI 시계열은 비선형적인 특성을 가지고 있으며, ARMA와 같은 선형 추계학적 모형으로는 분석할 수 없음을 알 수 있었다. 따라서 비선형 추계학적 과정에 적용할 수 있는 모형인 ARCH류의 모형을 이용하여 SOI 시계열에 대한 분석을 시도하였다.

선형 추계학적 모형에 주어진 시계열 자료를 적합시키는 경우에 가장 중요한 가정중의 하나는 잔차들의 분산이 시간에 따라 항상 일정하다는 것이다. 즉, 등분산(homoscedastic)을 따른다는 것이다. 따라서 일반적으로 모형을 적합시킬 때 등분산을 가정하므로 이 가정이 성립하지 않는 이분산(heteroscedastic)의 경우에는 모형의 적합성이 떨어지게 된다(Engel, 1982). 또한 추정값의 분산 등도 영향을 받게되어 예측값과 예측값의 신뢰구간 등이 정확하지 않게 되므로 이분산을 갖는 경우에는 이에 맞는 추정법을 사용하여야 한다.

ARCH 류의 모형은 이러한 이분산을 고려하여 적합시킬 수 있는 모형으로서, 오차의 분산을 과거의 관측값들의 제곱의 선형합수 형태로 표현하는 ARCH 모형이 시초이며, 이를 일반화한 GARCH(Generalized ARCH) 모형 및 GARCH 회귀모형 등이 있다.

이러한 이분산성을 설명하기 위한 모형 중 일반적인 모형이 다음과 같은 GARCH 회귀모형이다.

$$y_t = x_t \beta + \nu_t \quad (6.1)$$

여기서

$$\nu_t = \varepsilon_t - \phi_1 \nu_{t-1} - \dots - \phi_m \nu_{t-m} \quad (6.2)$$

$$\varepsilon_t = \sqrt{h_t} \cdot e_t \quad (6.3)$$

$$h_t = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \gamma_j h_{t-j} \quad (6.4)$$

$$e_t \sim \text{IID}, N(0, 1) \quad (6.5)$$

식 (6.1)의 GARCH 회귀모형은 오차가 m차 자기회귀오차모형을 따르며 동시에 오차의 분산이 GARCH(p, q) 분산모형을 따르는 GARCH 모형 중에서 가장 일반적인 AR(m)-GARCH(p, q) 회귀모형이다.

이 모형 중에서 가장 간단한 형태의 모형이 ARCH 모형으로서 GARCH(p, q) 모형에서 p=0

인 경우를 ARCH(q) 모형이라고 한다. 또한 $p > 0$ 이면 적어도 하나 이상의 ARCH 모형의 매개 변수는 0이 아니어야 한다. 즉, $p > 0$ 이면 $q > 0$ 이다.

GARCH 모형은 모든 과거의 잔차들의 제곱합을 이용하여 분산을 설명하므로 주로 장기기억과정(long memory process)의 경우에 적합하며, $p=0$ 인 경우의 ARCH(q) 모형은 가장 최근의 q개의 잔차들의 제곱만을 이용하여 분산을 설명하므로 단기기억과정(short memory process)의 경우에 적합하다.

본 연구에서는 이러한 AR(m)-GARCH(p, q) 모형을 SOI 시계열에 적합을 시켜보았다. 이를 위해 $p=1, q=1$ 인 AR(m)-GARCH(1, 1) 모형과 $p=0, q=2$ 인 AR(m)-ARCH(2) 모형을 가정하였다.

또한 최종적으로 모형의 차수를 결정하기 위해서 AR항의 차수 m은 BDS 통계를 이용하였는데, m의 범위는 3장에서와 마찬가지로 1~9로 변화시키며 BDS 통계 결과에 의해 가장 작은 값을 산정하는 차수를 결정하였다.

표 6.1 BDS 통계를 이용한 AR(m)-GARCH(p, q) 모형의 차수 결정

r	m	AR(1)		AR(2)		AR(3)	
		ARCH(2)	GARCH(1,1)	ARCH(2)	GARCH(1,1)	ARCH(2)	GARCH(1,1)
0.5	2	0.6122	1.0429	2.2202	2.2332	2.6136	2.4931
	3	1.3888	1.9918	3.6605	3.5996	3.9721	3.9396
	4	1.4845	2.1462	4.5556	4.4586	3.9289	4.1141
	5	1.8078	2.4365	3.9140	3.8610	4.0032	4.6078
r	m	AR(4)		AR(5)		AR(6)	
		ARCH(2)	GARCH(1,1)	ARCH(2)	GARCH(1,1)	ARCH(2)	GARCH(1,1)
0.5	2	2.0758	2.1006	2.0091	2.0599	1.9478	1.8733
	3	2.9968	3.0340	2.7708	2.8461	2.8380	2.6556
	4	3.3025	3.4476	3.0796	3.1492	2.9712	2.8281
	5	3.5429	3.7454	3.6451	3.5505	3.2722	2.9843
r	m	AR(7)		AR(8)		AR(9)	
		ARCH(2)	GARCH(1,1)	ARCH(2)	GARCH(1,1)	ARCH(2)	GARCH(1,1)
0.5	2	1.7840	1.7765	1.3969	1.4322	1.1146	1.0944
	3	2.2116	2.2083	2.2025	2.1704	1.9059	1.8698
	4	1.9729	2.0034	2.2704	2.2731	2.2034	2.3085
	5	1.2869	1.1176	2.3114	2.3575	2.2404	2.5278

[95% 유의수준($\alpha=0.05$) : -1.96 ~ 1.96]

위와 같은 분석을 통해 얻어진 BDS 통계 결과를 표 6.1에 수록하였는데, 모든 경우에 대해 유의수준을 만족하는 결과를 산정하는 것은 AR(1)-ARCH(2) 모형임을 알 수 있다.

BDS 통계를 이용해서 결정된 AR(1)-ARCH(2) 모형의 매개변수를 결정하기 위하여 SAS/ETS 프로그램을 사용하였으며, 이를 이용하여 결정된 모형의 매개변수는 각각 $\phi_1=0.6411$, $\alpha_1=0.0583$, $\alpha_2=0.0625$ 로 결정되었다. 따라서 최종적으로 결정된 SOI 시계열에 대한 비선형 추계학적 모형은 식 (6.6)과 같다.

$$y_t = -0.1877 + \nu_t \quad (6.6)$$

여기서

$$\nu_t = \varepsilon_t - \phi_1 \nu_{t-1} \quad (6.7)$$

$$\varepsilon_t = \sqrt{h_t} \cdot e_t \quad (6.8)$$

$$h_t = 0.5829 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 \quad (6.9)$$

$$e_t \sim \text{IID}, N(0, 1) \quad (6.10)$$

이렇게 구축된 AR(1)-ARCH(2) 모형은 잔차항은 자기상관을 가지면서 독립적이지 못하며, 모형의 조건부 분산은 과거 2개의 시계열의 함수로 보여질 수 있다는 것을 의미한다.

7. ENSO 예측모형

본 장에서는 현재 세계적으로 연구되어지고 있는 ENSO 예측 모형을 간략하게 소개하고자 한다. 물론 엘리뇨현상을 추계학적 모형으로 구축할 수만 있다면 추계학적 모형에 의해 예측이 가능할 수도 있을 것이다. 추계학적 모형에 엘리뇨현상을 적합시키고자 하는 노력이 현재 진행중이나 아직 많은 학자들은 이에 대하여 의문을 제기하고 있는 실정이다. 본 연구에서도 비선형 추계학적 모형을 이용하여 엘리뇨현상을 분석하고자 하였으며, 비선형 모형을 구축하였다. 엘리뇨 현상에 대하여 비선형 추계학적 모형을 구축한 예는 아직 존재하고 있지 않아 이에 대한 신뢰성 여부 또한 논란의 여지가 있을것으로 판단되어 본 장에서는 현재 연구되어지고 있는 예측모형을 서술하고자 한다.

지금까지 ENSO현상을 예측하기 위하여 개발되고 있는 모형은 크게 세가지 형태로 대별할 수 있으며 그 기법들은 다음과 같다.

첫 번째는 통계학적 기법으로 CCA(canonical correlation analysis; Graham 등, 1987a, b; Xu와 von Storch, 1990; Barnston과 Ropelewski, 1992), POPs(principal oscillation patterns; Xu와 von Storch, 1990), linear inverse modeling (Penland와 Magorian, 1993), constructed analogue model (Van den Dool, 1994), singular spectrum analysis-maximum entropy method(Keppenne과 Ghil, 1992).

두 번째로는 기상통계학적 모형과 혼합한 물리적 모형으로 the hybrid coupled model이 있다 (Barnett 등, 1993; Balmaseda et al., 1994).

마지막으로 동역학적 모형을 포함한 물리적 모형(Cane 등, 1986; Zebiak과 Cane, 1987; Kleeman, 1993)과 the fully coupled ocean and atmosphere general circulation models(GCMs; Latif 등, 1991, 1993; Ji 등, 1994a, b; Kirtman 등, 1996, 1997) 등이 있다.

엘리뇨현상을 예측하기 위한 노력은 많은 학자들에 의해 연구 진행중이며 위에서 열거한 방법론들도 이러한 노력의 결과이다. 그러나 방법론들에 따라서 장점과 단점들이 존재하여 그들에 대한 비교분석과 새로운 기법들의 개발이 연구 중에 있다.

8. 결론

본 연구에서는 엘니뇨의 지수중 하나인 SOI 자료가 어떤 시계열적인 특성을 가지고 있는가를 분석하여 적절한 추계학적 모형을 구성하였다. 또한 모형의 적합성 여부를 판단해본 후에, 그 자료가 가지고 있는 선형 혹은 비선형적인 특성에 대한 파악을 통해 다음과 같은 결론을 얻을 수 있었다.

- (1) SOI 자료를 이용하여 선형 추계학적 모형의 적정성 여부를 판단한 결과 ARMA(1,8;1) 모형이 가장 적합한 모형으로 결정되었다.
- (2) 구축된 ARMA(1,8;1) 모형의 잔차에 대한 ACF, PACF 분석 및 Porte Manteau test 결과에서는 무작위성을 나타내었다.
- (3) 그러나, SOI 자료와 ARMA(1,8;1) 모형의 잔차에 대한 BDS 통계에 의하면 무작위하지 않다는 결과가 산정되었으며, 이는 SOI 자료가 가지는 비선형적인 특성에 의한 것으로 판단된다. 따라서 SOI 자료에 대한 ARMA 모형은 적절하지 않음을 알 수 있었다.
- (4) 시계열이 가지는 확정론적 카오스 특성의 유무를 판단할 수 있는 Close Returns Test를 통해서 SOI 자료가 카오스적인 특성을 가지고 있지 않음을 알 수 있었다.
- (5) 비선형 추계학적 모형인 ARCH 류의 모형에 SOI 시계열을 적합시켜 보았으며, AR(m)-ARCH(2) 모형이 적합한 모형으로 분석되었다.
- (6) 이상의 분석에 따르면 엘니뇨의 지수인 SOI 자료에 대한 예측모형을 구성할 때는 비선형 추계학적 모형(nonlinear stochastic model)을 적용하는 것이 바람직할 것으로 판단되며, 이에 대한 지속적인 연구가 수행되어야 할 것으로 사료된다.

참고문헌

- 신현석 (1997). “엘니뇨의 이해와 한국의 자연재해.” 한국수자원학회지, 제30권, 제6호, pp. 75-82.
- 이재수 (1994). “물수지 방정식의 카오스적 분석.” 한국수자원학회지, 제27권, 제3호, pp. 45-54.
- Balmaseda, M.A., Anderson, D.L.T., and Davey, M.K. (1994). “ENSO prediction using a dynamical ocean model coupled to statistical atmospheres.” *Tellus*, Vol. 46A, pp. 66-68.
- Barnett, T.P., Latif, M., Graham, N., Flugel, M., Pazan, S., and White, W. (1993). “ENSO and ENSO-related predictability. Part 1: Prediction of equatorial Pacific sea surface temperature with a hybrid coupled ocean-atmosphere model.” *J. Climate*, Vol. 6, pp. 1545-1566.
- Barnston, A.G., and Ropelewski, C.F. (1992). “Prediction of ENSO episodes using canonical correlation analysis.” *J. Climate*, Vol. 5, pp. 1316-1345.
- Berge, P., Pomeau, Y., and Vidal, C. (1984). *Order within Chaos: Toward a deterministic approach to turbulence*. John Wiley & Sons, New York.
- Bjerknes, J. (1966). “Survey of El Niño, 1957-58 in its relation to Tropical Pacific Meteorology.” *Bull. Int. American Tropical Tuna Commission*, Vol. 12, pp. 3-62.
- Brock, W.A., Dechert, W.D., Scheinkman, J.A., and LeBaron, B. (1996). “A test for independence based on the correlation dimension.” *Econometric Reviews*, Vol. 15, No. 3, pp. 197-235.
- Brock, W.A., Hsieh, D.A., and LeBaron, B. (1991). *Nonlinear dynamics, Chaos, and instability: Statistical theory and economic evidence*. The MIT Press.
- Brock, W.A., Dechert, W.D., and Scheinkman, J.A. (1987). *A test for independence based on the correlation dimension*. Univ. of Wisconsin Press, Madison.
- Cane, M.A. (1983). “Oceanographic events during El Niño.” *Science*, Vol. 222, pp. 1189-1195.
- Cane, M.A., Zebiak, S.E., and Dolan, S.C. (1986). “Experimental forecasts of El Niño.” *Nature*, Vol. 321, pp. 827-832.
- Chu, P.S. and Katz, R.W. (1985). “Modeling and forecasting the Southern Oscillation: A time-domain approach.” *Mon. Wea. Rev.*, Vol. 113, pp. 1876-1888.
- Chu, P.S. (1995). “Hawaii Rainfall Anomalies and El Niño.” *J. of Climate*, Vol. 8, No. 6, pp. 1697-1703.
- Engel, R.F. (1982). “Autoregressive conditional heteroskedsticity with estimates of the variance of United Kingdom inflation” *Econometrica*, Vol. 50, pp. 987-1008.
- Farmer, J.D. and Sidorowich, J.J. (1987). “Predicting chaotic time series.” *Physical Review Letters*, Vol. 59, No. 8, pp. 845-848.
- Gilmore, C.G. (1993a). “A new test for Chaos.” *Journal of Economic Behavior and Organization*,

Vol. 22, pp. 209-237.

- Gilmore, C.G. (1993b). "A new approach to testing for Chaos with applications in finance and economics." *International Journal of Bifurcation and Chaos*, Vol. 3, No. 3, pp. 583-587.
- Graham, N.E., J. Michaelsen, and T.P. Barnett (1987a). "An investigation of the El Nino: Southern oscillation cycle with statistical models. 1. Predictor field characteristics." *J. Geophys. Res.*, Vol. 92, pp. 14251-14270.
- Graham, N.E., Michaelsen, J., and Barnett, T.P. (1987b). "An investigation of the El Nino-Southern Oscillation cycle with statistical models. 2. Model results." *J. Geophys. Res.*, Vol. 92, pp. 14271-14289.
- Grassberger, P. and Procaccia, I. (1983). "Measuring the strangeness of Strange Attractors." *Physica D.*, Vol. 7, pp. 153-180.
- Hannan, E.J. and Rissanen, J. (1982). "Recursive estimation of mixed autoregressive moving average order." *Biometrika*, Vol. 69, pp. 81-94.
- Ji, M., Kumar, A., and Leetmaa, A. (1994a). "A multiseason climate forecast system at the National Meteorological Center." *Bull. Amer. Meteor. Soc.*, Vol. 75, No. 4, pp. 569-577.
- Ji, M., Kumar, A., and Leetmaa, A. (1994b). "An experimental coupled forecast system at the National Meteorological Center: Some early results." *Tellus*, Vol. 46A, pp. 398-418.
- Keppenne, C.L. and Ghil, M. (1992). "Adaptive filtering and prediction of the Southern Oscillation index." *J. Geophys. Res.*, Vol. 97, pp. 20449-20454.
- Kirtman, B.P., and Schneider, E.K. (1996). "Model based estimates of equatorial Pacific wind stress." *J. Climate*, Vol. 124, pp. 1077-1091.
- Kirtman, B.P., Shukla, J., Huang, B., Zhu, Z., and Schneider, E.K. (1997). "Multiseasonal predictions with a coupled tropical ocean global atmosphere system." *Mon. Wea. Rev.*, Vol. 125.
- Kleeman, R. (1993). "On the dependence of hindcast skill on ocean thermodynamics in a coupled ocean-atmosphere model." *J. Climate*, Vol. 6, pp. 2012-2033.
- Latif, M. and Flugel, M. (1991). "An investigation of short range climate predictability in the tropical Pacific." *J. Phys. Oceanogr.*, Vol. 96, pp. 2661-2673.
- Latif, M., A. Sterl, E. Maier-Reimer, and M.M. Junge (1993). "Climate variability in coupled GCM. Part I: The tropical Pacific." *J. Climate*, Vol. 6, pp. 5-21.
- Packard, N.H., Crutchfield, J.P., Farmer, J.D., and Shaw, R.S. (1980). "Geometry from a Time Series." *Phys. Rev. Lett.*, Vol. 45, pp. 712.
- Penland, C. and Magorian, T. (1993). "Prediction of Nino 3 sea surface temperatures using linear inverse-modelling." *J. Climate*, Vol. 6, pp. 1067-1076.
- Ropelewski, C.F. and Halpert, M.S. (1987). "Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation." *Mon. Wea. Rev.*, Vol. 115, pp.

1606-1626.

- Schwarz, G. (1978). "Estimating the dimension of a model." *Ann. Statist.*, Vol. 6, pp. 461~464.
- Tong, H. (1990). *Non-linear time series: A dynamical system approach*. Clarendon Press, Oxford.
- Van den Dool, H.M. (1994). "Searching for analogues, how long must we wait?" *Tellus*, Vol. 46A, pp. 314-324.
- Walker, G.T. (1923). "Correlation in seasonal variations of weather(VIII): A preliminary study of world weather." *Mem. Indian Meteorol. Dep.*, Vol. 24, pp. 75-131.
- Xu, J.-S. and von Storch, H. (1990). "Predicting the state of the Southern Oscillation using principal oscillation pattern analysis." *J. Climate*, Vol. 3, pp. 1013-1029.
- Zebiak, S.E. and Cane, M.A. (1987). "A model of El Nino-Southern Oscillation." *Mon. Wea. Rev.*, Vol. 115, pp. 2262-2278.