

엘니뇨현상에 관한 통계적분석

김 해 경
연세대학교 이과대학 수학과

A Statistical Analysis for El Niño Phenomenon

HAE-KYUNG KIM
Department of Mathematics, College of Sciences, Yonsei University

본 논문의 목적은 엘니뇨현상의 예측을 위한 확률모델의 개발과 그 응용에 있다. 이를 위해, 먼저 태평양 적도지역의 월평균 해면수온의 편차시계열을 기초로 하여 엘니뇨현상의 지속기간, 강도의 결정방법과 이 현상의 출현에 대한 판별방법을 제안하였다. 다음으로 과거 40년(1951-1990) 자료의 편차시계열에 나타난 엘니뇨의 연변동성, 주기성, 종속성 등 확률구조 및 통계적 특성을 파악하였고, 이 결과를 엘니뇨현상의 예측을 위한 시계열 비선형확률모델을 유도하였는데 이용하였다. 마지막으로, 유도된 확률모델의 실제적용을 위한 통계절차를 제안하였다.

This paper is concerned with the development and application of a stochastic model for predicting El Niño phenomenon. For this, first a general criterion for determining El Niño phenomenon, including period and strength, which is based on partial sum of monthly sea surface temperatures (SST) anomalies, is proposed. Secondly, the annual fluctuations, periodicity and dependence of monthly mean of equatorial Pacific SST during the period 1951-1990 are analyzed. Based on these, time series nonlinear regression model for the prediction of SST have been derived. A statistical procedure for using the model to predict the SST level is also proposed.

서 론

남아메리카 서해안의 페루해류에 적도쪽으로부터 침입하는 異常暖流로 인하여 해수온이 일정기간 급상승하는 엘니뇨현상은 여러가지로 해양 및 기상현상에 영향을 미치는 징후가 있다. 이 현상의 발생 요인이나 메카니즘은 아직 확실하게 밝혀지지 않았으나 적도지역의 海水와 大氣와의 상호작용에 기인하는 현상으로 보는 견해가 지배적이다. 즉, 남태평양에서 고기압이 강하게 발달하면 남동무역풍이 서태평양으로 불어 들어와 그 기간에는 이 지역의 수온과 수위를 높이고, 또 이 고기압이 다시 약해지면서 서태평양에 집적된 고온의 해수가 赤道反流에 의해 동쪽으로 이동하여 적도 동부해역의 수온과

수위를 높이고 이것이 페루연안까지 침입하는 현상을 말한다.

근래는 물론이고 역사적으로도 엘니뇨현상에 의하여 초래된 기상이변을 포함한 해양 그리고 생태학적인 변화가 많이 나타났다. Quinn and Neal (1987)은 15세기이후 현재까지 4세기여동안 남아메리카 북부의 서쪽연안과 여기에 인접된 태평양상에서 발생한 엘니뇨의 해양생태학적, 기상학적인 근거와 그 영향들을 史料를 통하여 정리분석하였다. 뿐만 아니라 엘니뇨출현과 영향에 대한 지질학적인 근거들도 DeVries (1987)에 의하여 확인된 바 있다. 이들의 분석에 의하면 엘니뇨현상은 대부분 해양 및 인접연안의 생태계에 대한 크고 작은 변화와 豪雨, 洪水, 海溢 등을 포함한 氣象異變을 동반하고 있다.

18세기초 이후 최근까지(1803-1987년)의 자료만을 근거로 할 때 이러한 엘니뇨현상은 평균적으로 대략 3.8년을 週期로 하여 발생한 것으로 분석되었다.

생태계의 변화 그리고 기상이변 등을 포함하여 우리생활에 많은 영향을 주고 있는 엘니뇨현상의 출현에 대한 事前豫測 그리고 이 현상이 주변 해양생태계에 미치는 영향의 분석은 재앙의 경보나 예보의 측면에서 매우 중요하다.

본 연구의 목적은 엘니뇨현상의 統計 및 確率의 특성을 파악하고, 이 현상의 數值豫測을 위한 確率 모델을 확립하는데 있다. 이를 위해, 먼저 엘니뇨현상을 설명하는 적도지역 해수면온도의 週周期性, 從屬性 등 확률 및 통계적 특성을 파악하고, 다음으로 이 결과들을 기초로 하여 엘니뇨의 出現, 強度, 持續期間 등에 대한 수학적 정의를 확립하였다. 마지막으로, 엘니뇨현상의 예측을 위한 確率모델을 유도하고 이 모델의 적용한계를 오차분석을 통하여 논하였다. 또, 유도된 모델을 이용한 엘니뇨예측의 統計的절차를 제안하였다.

본 분석에서 사용된 자료는 일본기상청에 의하여 과거 40년(1951~1990)간의 赤道地域 東太平洋(4°N-4°S, 150°W-90°W)에서 관측된 해수면 온도의 월평균값들이다.

統計的分析

1. 定義와 強度

個個의 엘니뇨현상은 규모나 지속시간이 일정하지 않다. 같은 현상일지라도 관측의 위치나 시점에 따라 서로 다른 자료를 얻을 수 있다. 따라서, 엘니뇨현상의 정의나 규정은 분석자에 따라 다소 다를 수 있으나, 통상적으로는 해수면의 년평균온도가 평년보다 0.5°C 이상 상승할 때를 엘니뇨현상으로 정의한다. 한편, 해양연구과학위원회(Scientific Committee on Oceanic Research, SCOR)는 엘니뇨를 에콰도르와 페루를 따라 리마(Lima: 12°S)에 이르는 연안의 해수면온도가 적어도 3개월이상 연속적으로 정상온도의 표준편차를 초과상승하는 현상으로 이 지역 5개연안지점(Talara, Puerto, Chicama, Chimbote, Isla Don Martin, Callao)중 3지점 이상에서 확인되는 상태로 정하고 있다(Quinn and Neal, 1987).

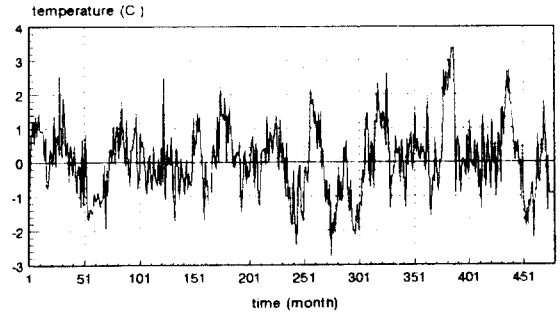


Fig. 1. Time series of original monthly sea surface temperature(SST) anomalies: 1951-1991.

엘니뇨의 강도는 기본적으로 일정기간 이상을 정상온도로 부터 이탈된 해수면온도의 크기와 이 현상에 동반하거나 후속되는 기상이변이나 해양생태계가 받는 충격 그리고 연안환경의 변화 등의 정도에 따라 결정된다. 그러나, 이러한 외형적인 변화의 크기는 결국 해수면온도의 상승수준에 비례하므로 Quinn and Neal(1987)에서 언급한 바와 같이 엘니뇨의 강도는 남반부 적도지역에서 해수면온도의 상승수준에 따라 2-3°C, 3-5°C, 7-12°C를 근방으로 하여 각각 普通, 强 그리고 超强으로 구분하기도 한다.

그러나, 이와같은 정의들은 특정해역의 수온을 전제로 하기 때문에 광범위한 영역에서 관찰되는 일반적인 자료에 적용하기 어렵다. 특히, 년평균을 기초로하는 엘니뇨의 판별방법은 이 현상의 初期와 末期의 시점을 구체적으로 설명할 수 없을 뿐 아니라 지속기간이 12개월보다 크게 길거나 짧을 때에는 強度와 持續期間의 판정에 오류를 범할 수 있다.

본 분석에서는 엘니뇨현상의 始點과 終點의 판별에 月平均海面水溫의 偏差時計列과 그 部分合曲線을 이용한다. 그리고 強度의 판정은 평균의 표본분포를 기초로 한다. 여기서, 偏差時系列이란 Fig. 1에서와 같이 長期平均이 제거된 月平均海面水溫의 시계열 $\{X_t\}$ 을 말하고, 部分合時系列 $\{S_t\}$ 는 $S_t = \sum X_t$ 로 정의된다. 部分合曲線이란 부분합시계열을 근사적으로 맞추어 주고 또 시간을 독립변수로 갖는 解析的인 곡선을 의미한다. 부분합곡선이 어떤 구간에서 零의 기울기를 가지면 이 구간에서 (월평균)해면수온이 長期平均값을 유지함을 의미한다. 또, 주어진 구간에서 이 곡선의 陽(陰)의 기울기란 그 구간에서 수

온이 長期平均을 上廻(下廻)하는 상태를 의미한다. 특히, 이웃하는 두 변곡점사이에서 이 근사곡선이 양(음)의 기울기를 가지면 본래의 편차시계열은 이 구간에서 위(아래)로 오목한 상태가 된다. 따라서, 엘니뇨의 시점과 종점은 근사곡선상에 대응되는 변곡점 또는 그 안쪽점으로 설명되고, 이 변곡점사이 존재하는 편차들의 평균은 여기서 발생된 월평균수온의 평균변화량으로써 엘니뇨의 강도를 나타낸다.

본 분석에서는 엘니뇨의 출현과 그 강도를 다음과 같이 정의한다: 부분합곡선이 두 변곡점 (t, C_t), (s, C_s) 사이에서 陽의 기울기를 갖고 그 사이의 평균값 (C_s-C_t)/(s-t)이 z(α)[σ_T/√(s-t)] + μ_T(단, μ_T, σ_T²는 全偏差時計列의 平均과 分散, 그리고 z(α)는 표준정규분포의 우측꼬리확률이 α되는 값)를 초과하면 이 구간에서는 有意水準 α에서 엘니뇨출현으로 규정한다. 한편, 이 방법은 對立假說 H₁: “엘니뇨의 存在”에 대한 歸無假說 H₀: “엘니뇨의 非存在”의 檢定統計量

$$Z^* = \frac{M_{st} - \mu_T}{\frac{\sigma_T}{\sqrt{s-t}}} \quad (\text{단, } M_{st} = (C_s - C_t) / (s - t))$$

의 歸無分布인 표준정규분포의 p-값이 α보다 작으면 귀무가설을 棄却하여 엘니뇨의 출현을 인정하는 통계적절차와 同值이다. 또, 이웃하는 두 변곡점사이에서 편차시계열의 값들이 평균 (C_s-C_t)/(s-t)를 초과하는 期間 그리고 그 始點과 終點을 각각 그 엘니뇨의 지속기간과 시점, 종점으로 정의한다.

이 정의는 편차시계열이 (弱)定常이고 평균과 분산이 각각 μ_T, σ_T²인 정규분포를 갖는다는 假定을 기초로 한다. 따라서, 이 모집단에서 抽出된 크기 n의 確率標本平均들은 평균과 분산이 각각 μ_T, σ_T²/n인 정규분포를 갖기 때문에 한 구간에서 엘니뇨의 출현이란 그 구간에 나타난 편차시계열이 평균적으로 보아 (有意水準 α에서) 純粹確率過程의 實現値로 볼 수 없는 異常상태를 의미한다.

엘니뇨의 強度도 같은 방법으로 정의한다. 엘니뇨기간의 편차시계열이 평균과 분산이 각각 μ_E, σ_E²인 正規分布를 따른다는 假定과 그 平均分布의 近似正規性を 이용하여, 統計量

$$S^* = \frac{M_{st} - \mu_E}{\frac{\sigma_E}{\sqrt{s-t}}}$$

Table 1. Moments of monthly SST anomalies.

Moments			
N	480	Sum Wgts	480
Mean	0.0655	Sum	31.44
Std Dev	1.054538	Variance	1.11205
Skewness	0.27745	Kurtosis	0.156973
USS	534.7314	CSS	532.6721
CV	1609.98	Std Mean	0.048133
T:Mean=0	1.36082	Prob> T	0.1742
Sgn Rank	2804.5	Prob> S	0.3523
Num=0	477	Q3-Q1	1.385
Range	6.16	Mode	0.91

의 표준정규분포 p-값이 0.60(미만), 0.60(이상)-0.90(미만), 0.90-0.99에 따라 각각 弱, 普通, 强으로 구분하고, 또 0.99를 넘으면 초강으로 판정한다. 여기서 엘니뇨의 약, 보통, 강의 수준이란 과거의 자료에 기초를 둔 強度分布에서 아래로부터 각각 60%, 60-90% 그리고 90-99%를 포함하는 규모를 의미하고, 초강이란 이 強度分布의 범위를 넘어선 異常크기를 말한다.

본 분석에 있어서, 全偏差時計列의 近似的正規性 ((μ_T, σ_T²)=(0.0655, (1.0545)²))은 Table 1에서 쉽게 확인된다. 또, (μ_E, σ_E²)=(1.0199, (0.9097)²))를 母數로 갖는 엘니뇨편차시계열의 경우도 같은 방법으로 확인된다. 분석결과는 Table 2와 같다. 월평균 편차시계열의 부분합곡선(Fig. 2)으로 부터 먼저 변곡점들을 구하고, 이 변곡점사이의 기간과 그 평균을 계산하였다. 두 변곡점 (t, C_t)와 (s, C_s)의 사이에서 평균값(C_s-C_t)/(s-t)을 Manhattan diagram으로 Fig. 3에 표시하였다. 열한가지의 경우중에서 Table 2에서 보는 바와 같이 여섯경우만이 有意水準 0.01에서 엘니뇨현상으로 판정된다. 그리고, 나머지 4개는 비교적 큰 p-값을 보이고 있어 엘니뇨현상으로 보기 어렵다. 이 결과에 의하면, 과거 40년동안 6회: 1951.4-1952.2, 1965.4-1966.2, 1972.5-1973.2, 1977.5-1978.2, 1982.6-1983.4, 1986.11-1987.9에 엘니뇨가 발생하였고, 평균지속기간은 10.8개월, 그리고 월별로 보아 보통 3-6월에 發生하여 11-1월에서 成熟期를 거치고 다음해 3-4월頃に 終熄된다. 強度에 있어서는 통계량 S*의 p-값크기에 따라 처음 세 경우는 微弱수준으로 그리고 1972.5-1973.2, 1986.11-1987.9의 두 경우는 普通수준으로 평가된다. 특히, 1982.6-19

Table 2. Summary of statistical analysis for monthly SST anomalies, where the hypothesis is Ho: No El niño vs. H1: El niño, and n=sample size.

Time (initial-terminal)	Duration (month)	Z*(n)	p-value (of Z*)	Ho:	S*(n)	p-value (of S*)	Strength
Apr. 1951-Feb. 1952	11	2.625(16)	0.0043	reject	-1.154(16)	0.8749	weak
Apr. 1965-Jan. 1966	10	3.523(19)	0.0002	reject	-0.491(19)	0.6517	weak
May 1972-Feb. 1973	10	3.423(13)	0.0003	reject	0.183(13)	0.4286	moderate
Mar. 1977-Feb. 1978	11	4.097(23)	0.0000	reject	-0.283(23)	0.6103	weak
Jun. 1982-Apr. 1983	12	6.797(17)	0.0000	reject	3.549(17)	0.0002	(very)strong
Nov. 1986-Sep. 1987	11	4.565(18)	0.0000	reject	0.837(18)	0.2033	moderate
Dec.1952-Sep.1953	10	2.303(16)	0.0107	accept	(No El niño)		
Jul. 1957-May.1959	10	2.021(32)	0.0217	accept	(No El niño)		
Jun.1963-Dec.1964	7	1.990(9)	0.0223	accept	(No El niño)		
Jan.1968-Feb.1970	14	2.255(21)	0.0122	accept	(No El niño)		
Jul. 1976-Sep.1976	3	1.825(4)	0.0344	accept	(No El niño)		

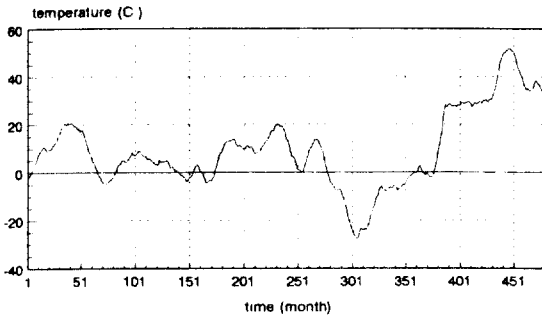


Fig. 2. Plot of partial sum of monthly SST anomalies.

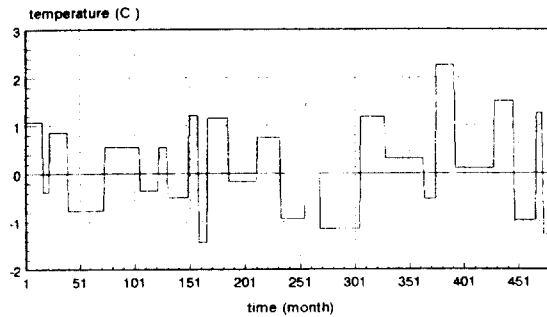


Fig. 3. Manhattan diagram of monthly SST anomalies.

83.4의 경우는 정상적인 강도를 넘어선 超強의 수준으로 평가된다. 그러나, 분석자의 관점에 따라서는 이 판정에서 제외된 다섯경우중 한 둘이 p-값의 크기에 따라 엘니뇨의 현상으로 설명될 수도 있다.

이 판정결과는 Quinn and Neal 또는 SCOR의 판정결과와 대부분 일치하는 것으로 통상적인 판정

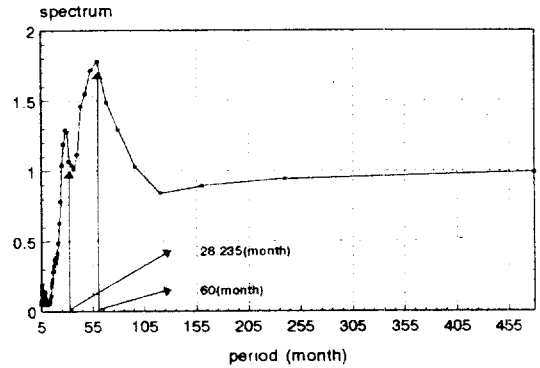


Fig. 4. Estimated power spectrum for the monthly SST anomalies.

기준과 큰 차이를 보이지는 않는다. 그러나, 지속기간의 다양성을 고려하면 년평균을 기초로 하는 기간과 강도에 대한 종래의 판정방법은 바람직하지 않다.

2. 週期性

기온 등 기상자료에서와 같이, 엘니뇨현상이 발생하는 적도지역의 월평균 해수면온도에도 한 回歸直線의 주위를 상하진동하며 어떤 기간을 두고 반복되는 週期性의 존재가 가정된다. 이러한 결과는 Fig.1에 주어진 해수면온도의 월평균 편차시계열의 외관상의 특징에서도 관찰된다. 즉, 시간의 변화에도 일정한 分散을 유지하며 원점을 지나는 직선주위를 상하진동하고 있는 定常狀態의 특징과 특정한 회귀

추세는 없지만 어느정도 같은 기간을 두고 같은 수준이나 패턴이 반복되는 週期性的의 특징이 나타난다. 이러한 결과는 이 시계열의 標本 自己相關函數에서도 쉽게 확인될 수 있다.

특히, 週期性的의 분석은 위 자료를 기초로한 periodogram 또는 이것을 적당한 가중치를 사용하여 平滑시킨 spectrum의 추정값으로 이루어진다. 이 때의 평활은 periodogram이 漸近的 不偏性은 갖지만 충분한 표본크기에도 分散이 감소하지 않는 一致性을 갖지 못하는 결점을 보완할 수 있다. Fig. 4는 40년간 (480월)의 자료를 기초로 하여 얻어진 periodogram을 三角加重函數의 移動平均에 의하여 평활된 spectrum의 추정값을 나타낸 것이다. 대략 60.0월과 28.2월이 각각 主副週기로 나타나고 있다. 이 두 주기의 존재는 여러가지 다른 형태의 濾過函數를 통해서도 평활되지 않는 사실에서도 설명되지만, 다음에 주어지는 主副週期的의 有意性檢定에 의하여 그 존재가 더욱 확실해 진다. 즉, 主週期 60.0월의 존재에 대한 검정은 periodogram을 기초로 하고, 근사분포가 $P\{T_F > g\} \cong N(1-g)^{N-1}$, $N = [n/2]$, 으로 주어지는 Fisher(1929)의 검정통계량

$$T_F = \frac{I^{(1)}(w_{(1)})}{\sum_{k=1}^{[n/2]} I(w_k)}$$

를 이용한다. 주파수 $w(1)$ 에서 나타난 최대 periodogram $I^{(1)}(w(1)) = I^{(1)}(w_9) = 40.9729$ 과 $N = 240$ 을 이용하면 $T_F = 18.3935$ 로 주어지고, 유의수준 $\alpha = 0.05$ 에서 $N(1-g)^{N-1} = 240(1-g)^{239} = \alpha$ 를 만족하는 $g = 0.0348$ 이 $T_F = 18.3935$ 에 비하여 너무 작기 때문에 유의수준 0.05에서 주파수 $w_9 = 0.10472$ (주기 = 60.00월)의 주기성분의 존재가 확인된다. 부주기 28.2월의 존재에는 통계량 T_F 를 부주기까지 확장시킨 Whittle(1952)통계량

$$T_W = \frac{I^{(2)}(w_{(2)})}{\left[\sum_{k=1}^{[n/2]} I(w_k) \right] - I^{(1)}(w_{(1)})}$$

이 이용된다. T_F 의 근사분포에서 N 이 $N-1$ 로 바뀐 T_W 의 근사분포를 이용하면, 유의수준 0.05에서 $T_W = 0.0377 > 0.0348 = g$ 으로 주파수 $w_{18} = 0.22253$ (주기 =

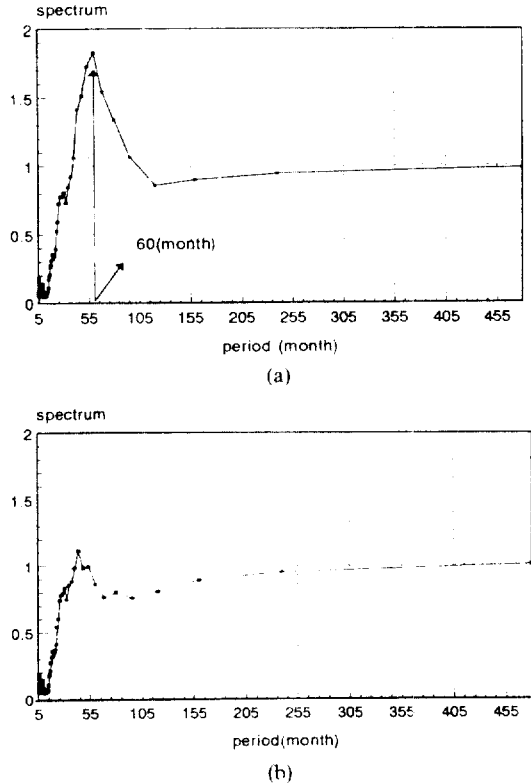


Fig. 5. Estimated power spectrum for the monthly SST anomalies: (a) after deleting the sub-period. (b) after deleting both the sub- and main periods.

28.235월)에서 존재하는 부주기를 부정할 근거는 없다.

여기서 主週期 60.0월은 副週期 28.2월주기의 근사적 倍數로 調和現狀의 결과일 수도 있다. 그러나, 본래의 자료에서 이 부주기성분을 母數의 표현인 방정식

$$\beta_{10} \sin(\beta_{11} + \beta_{12} t), \quad \beta_{12} = \frac{2\pi}{28.2}$$

을 통하여 제거시키면, 즉 최소제곱법을 이용하여 위 방정식을 자료에 맞추고 여기서 얻어지는 殘差의 週期성을 분석하면, Fig. 5(a)에서와 같이 60.0월의 週期는 계속 강하게 남는다. 이것은 두 주기가 서로 독립적으로 발생하고 있음을 의미한다.

같은 방법으로, 主副週期的의 모수방정식

$$\beta_{10} \sin(\beta_{11} + \beta_{12} t) + \beta_{20} \sin(\beta_{21} + \beta_{22} t) \quad (2.1)$$

$$(\beta_{12} = \frac{2\pi}{60.0}, \beta_{22} = \frac{2\pi}{28.2})$$

을 자료에 적용하여 얻어진 최소제곱곡선의 殘差는 Fig. 5(b)에서 보는바와 같이 主副週期가 모두 제거된 시계열이 된다.

월평균온도의 시계열자료로부터 얻을 수 있는 최소주기가 두달(Nyquist frequency = π)이고, 그리고 Fig. 4에서 보는 바와 같이 이 최소주기에서는 spectrum이 근사적으로 零이므로 주어진 월평균자료는 60.0월, 28.2월의 두 주기의 추정에 충분한 정보를 주고 있으며, 또 aliasing문제도 발생하지 않는다. 또 이 두 주기의 산술평균(3.68년)은 위에서 언급한 Quinn and Neal의 결과와 평균적으로는 일치하지만 이 평균값이 그대로 엘니뇨의 주기로 설명될 수는 없다. 왜냐하면 3.68년이란 다만 과거 약 1세기동안에 발생된 강도 보통이상인 엘니뇨사이의 평균기간이기 때문이다. 그 밖에 高,低돛수의 주요 주기성은 더 이상 관찰되지 않는다.

豫測모델

앞에서 언급한 바와 같이 엘니뇨는 적도지역의 海水와 大氣와의 상호작용에 기인하는 현상으로, 남태평양상의 일정기간 강했던 고기압이 다시 약화되는 상태에서 발생하는 현상으로 설명되고 있다. 다른 한편으로, 엘니뇨는 지구의 自轉速度가 가속에서 감속으로 전환하는 상태와 유관하다는 분석(Zheng *et al.*, 1990)도 있다. 이러한 결과들이 어느 정도 타당성을 갖는다면, 엘니뇨현상의 정확한 예측을 위해서는 기본적으로 해당지역의 대기상태나 엘니뇨현상과 同伴現象이라고 평가되는 熱帶大氣의 南方振動(Southern Oscillation : Deser and Wallace, 1987 ; Gage and Reid, 1987)이나 그 熱帶海洋의 鹽分濃度의 변화 또는 지구자전률의 연중변화 등과 같은 엘니뇨의 前兆 또는 同伴요인에 대한 분석이 병행되어야 한다. 이러한 관점에서 볼 때, 엘니뇨의 예측에는 두가지 형태의 확률모델의 개발이 가능하다. 엘니뇨현상에 관련된 주요인자 그리고 그 관련된 決定的 또는 確率的 형태를 밝히고 그 관계를 모델개발에 이용하는 傳達函數確率모델의 개발이 그 하나이다. 이에 반하여, 시간에 따라 내

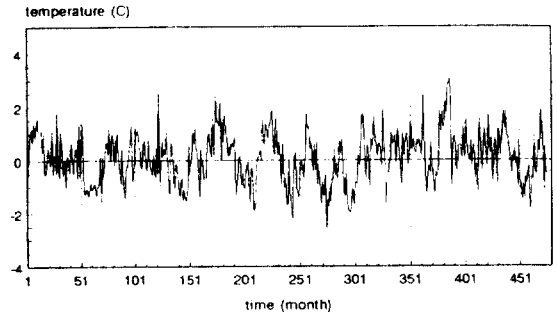


Fig. 6. Plot of residual series $\{R_t^{(1)}\}$.

존하는 해수표면온도의 확률적 종속관계를 파악하고 이것을 모델에 이용하여 수온의 예측에는 과거와 현재의 해수면온도만을 기초로하는 單變數確率모델이 생각될 수 있다.

본 분석에서는, 다음과 같은 이유로 後者の 모델 개발에 역점을 둔다. 前者의 경우는 독립인자를 파악하는데 대규모의 실험 그리고 대기와 해양의 상호작용에 관한 역학구조의 광범위하고 완벽한 분석이 요구될 뿐 아니라, 대부분의 이 인자들은 기상인자 또는 물리현상의 지배인자로서 각각 고유의 回歸趨勢, 週期性 그리고 從屬性등 복잡한 確率構造를 가지며 어떤 시점에서 엘니뇨현상의 예측에는 적어도 같은 시점에서 이 변수들의 정확한 예측값을 필요로 한다. 이것은 이 방법이 後者보다 확률적으로 非效率일 수도 있음을 의미한다. 둘째, 前者의 개발은 後者の 개발과정에서 필수적으로 요구된다. 왜냐하면, 전달함수모델의 독립변수선별과 효율평가의 기준은 단변수모델을 이용한 殘差의 분석으로 이루어지고, 또 전달함수의 수정과 보완을 필요로하는 독립인자의 異常상태나 자료의 왜곡에 대한 파악도 단변수확률모델의 殘差를 통해서 확인되기 때문이다.

앞 절에서 밝혀진 바와같이, 해수면온도의 편차 시계열이 갖는 주기성분은 방정식(2.1)의 母數의 모델로 표현된다. 또, 수온의 변화는 순간도약이 없는 연속변화로써 인접된 온도상호간에 종속성의 존재가 가정될 수 있다. 이와같은 종속성의 존재는 주기성분의 함수식(2.1)을 주어진 자료에 적용하여 얻어지는 Fig. 6의 잔차시계열에서도 엿 볼 수 있다. 이와 같은 관점에서 볼 때, 해수면온도의 편차 $X_t(t)$ 는 月을

Name of variable = R_t⁽¹⁾
 Mean = -0.07014. Standard deviation = 0.962483, Number of observations = 480

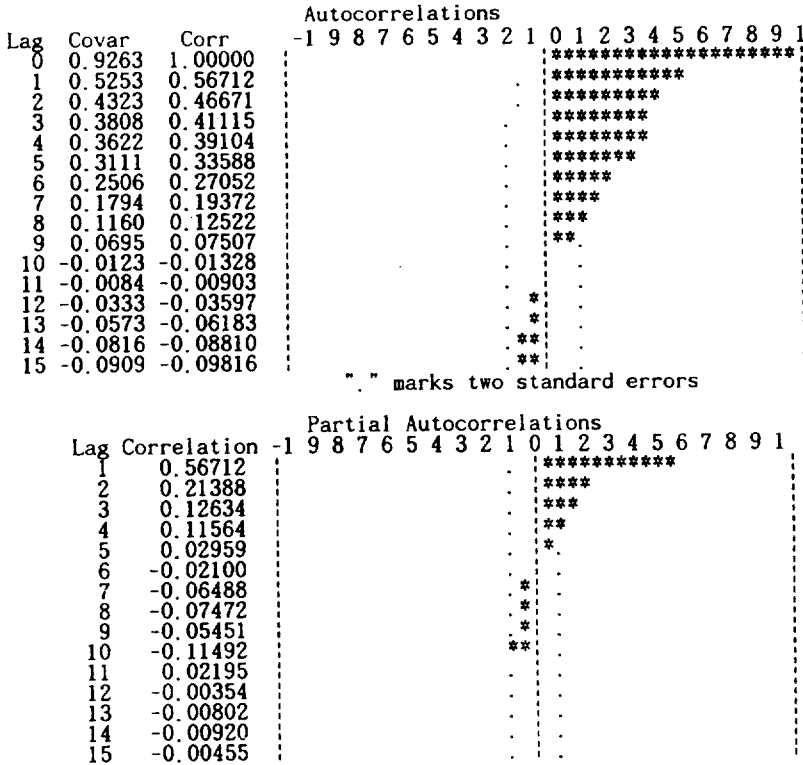


Fig. 7. Sample autocorrelation function and sample partial autocorrelation function, of residual series {R_t⁽¹⁾}.

나타내는 시간)는 기본식

$$X_t = C_t + S_t + E_t$$

으로 표현된다. 여기에서, C_t는 (2.1)로 표시되는 주기성분, S_t는 종속성을 나타내는 확률성분 그리고 E_t는 측정오차를 포함하는 확률오차성분이다. 또, 확률성분 S_t와 측정오차 E_t의 합, Z_t=S_t+E_t,은 보통 定常 또는 弱定常 ARMA(p, q)모델

$$\sum_{j=0}^p a_j Z_{t-j} = \sum_{k=0}^q b_k e_{t-k} \quad (p, q \text{는 양의 정수}) \quad (2.2)$$

로 설명된다. 이 때 a_j와 b_k는 상수(a₀=b₀=1) 그리고 e_t는(正規)白色誤差이다. 따라서, 위 기본식은 (2.2)의 Z_t를 사용하여

$$X_t = C_t + Z_t \quad (2.3)$$

로 표시된다.

모델(2.2)에서 確率成分 Z_t는 시계열 X_t에 내존하는 從屬性을 설명한다. 확률성분 Z_t의 확률모델은, X_t를 (2.1)로 정의된 함수식 C_t로 濾過하여 X_t에서 週期成分이 제거된 잔차 R_t⁽¹⁾, R_t⁽¹⁾=X_t-C_t,의 분석으로 결정된다(C_t는 C_t의 추정값). Fig. 6에서와 같이 시간에 따라 나타나는 일정한 평균과 분산 등은 시계열 {R_t⁽¹⁾}의 定常 또는 弱定常상태를 의미하며, 이 사실은 時差에 따른 標本 自己相關函數와 標本 部分 自己相關函數로도 확인된다. Fig. 7에서 보는 바와 같이, 標本 自己相關函數는 時差가 증가할수록 零에 가까워지고, 標本 部分 自己相關函數는 時差 3이후에서 급히 감소하고 있다. 이것은 Z_t가 3계의 自己 回歸過程(AR(3))

$$Z_t = \mu + a_1 Z_{t-1} + a_2 Z_{t-2} + a_3 Z_{t-3} + e_t \quad (2.5)$$

로 설명됨을 의미한다. 이 결과는 Akaike's AIC (Akaike, 1974) 또는 Schwartz's SBC (Schwartz,

1978)판정에서도 확인된다.

앞 분석들로 얻어진 성분 C_t, Z_t 등의 確率모델들을 (2.2)에 대입하여 시계열 X_t 의 確率모델이 완성된다. 즉, X_t 의 確率모델은

$$X_t = \beta_{10} \sin(\beta_{11} + \beta_{12}t) + \beta_{20} \sin(\beta_{21} + \beta_{22}t) + \mu + a_1 Z_{t-1} + a_2 Z_{t-2} + a_3 Z_{t-3} + e_t \quad (2.6)$$

이다. 여기서 Z_t 는 殘差 $R_t^{(2)}$ 로 近似化될 수 있는 確率變數이고, e_t 는 正規($0, \sigma^2$)確率變數이다.

모델(2.6)의 적합성확인인 이 모델을 자료에 적용한 후 얻어지는 殘差 $\{R_t^{(2)}\}$, $R_t^{(2)} = X_t - (C_t + \hat{Z}_t)$ 의 분석으로 이루어진다. 즉, (2.6)의 週期成分函數를 最小제곱법을 이용하여 먼저 Fig. 1의 자료에 적용하고, 그 殘差에 (2.6)의 從屬성분을 같은방법(조건 最小제곱법)으로 적용하여 추정값을 얻는다. 잔차제곱합이 前後者 각각 447.02, 283.08인 추정값과 표준오차(SE)는

$$\begin{aligned} \hat{\beta}_{10} &= 0.4004(0.0626) & \hat{\beta}_{11} &= -5.4649(0.3176) \\ & & \hat{\beta}_{12} &= 0.2169(0.0011) \\ \hat{\beta}_{20} &= -0.4644(0.0627) & \hat{\beta}_{21} &= -4.6119(0.2702) \\ & & \hat{\beta}_{22} &= 0.1074(0.0009) \\ \hat{a}_1 &= 0.4194(0.0454) & \hat{a}_2 &= 0.1585(0.0488) \\ & & \hat{a}_3 &= 0.1245(0.0454) \\ \hat{\mu} &= 0.0597(0.1173) \end{aligned}$$

이다. 추정의 표준오차는 대부분 작은 값을 가지며, 이것은 서로 다른 標本에 의한 추정값의 변화를 적게 만들고 따라서 추정값의 신뢰도를 높일 수 있음을 의미한다. 殘差 $R_t^{(2)}$ 의 標本 部分自己相關函數의 관찰로 결정된 종속성분의 階數는 해수면온도의 지속 상태를 설명할 수 있어 유의성의 검정이 특히 요구된다.

歸無假說 $H_0: a_i = 0 (i=1,2,3)$ 의 t-검정통계량

$$t^* = \frac{\hat{a}_i - 0}{SE(\hat{a}_i)} \quad (i=1,2,3)$$

의 값은 모두 2보다 큰 9.24, 3.25, 2.74이므로 有意水準 5%에서 세 가설은 모두 기각된다. 이 결과는 세 係數가 모두 零과 有意的인 差를 갖고 따라서 종속성분의 階數 3 그리고 종속성분의 예측모델

$$Z_t = 0.0597 + 0.4195 Z_{t-1} + 0.1585 Z_{t-2}$$

$$+ 0.1245 Z_{t-3} + e_t \quad (2.7)$$

들이 다시 확인된다.

종속성분의 AR(3)는 월평균 해수면온도의 “3個月持續性”을 의미한다. 즉, 평균온도에 있어서 이달의 해수면온도는 내달과 그리고 그 다음달의 해수면온도에만 직접 영향을 주고 있음을 말한다.

한편, 殘差시계열 $\{R_t^{(2)}\}$ 는 확률오차 $\{e_t\}$ 들의 추정값으로 이것들의 (正規)白色過程의 성립은 모델의 적합함을 의미한다. 殘差 $\{R_t^{(2)}\}$ 의 spectrum은 물론이고 Fig. 8에서 주어진 自己相關係數와 部分自己相關係數는 모두 白場時系列의 특성을 나타내고 있다. 더욱이, Fig. 8에서 時差 30까지 사용하여 계산되는 Box-Pierce 카이-제곱검정량 의 값이

$$Q(r) = (N-k-p) \sum_{i=1}^{k=30} r_i^2 = 38.13$$

(단, r_i =時差 i 의 自己相關係數, N =자료의수, k =自己相關係數가 零이 아닌 最大時差 그리고 p =推定된 母數數)로 주어지고, 有意水準 0.05에서 같은 自由度(d.f.= $k-p$)27의 이론적 카이-제곱값이 40.11임을 참작할 때 確率誤差의 白場性を 부정할 근거가 없다.

엘니뇨현상의 발생예측은 모델(2.6)를 이용한 적도지역 해수면온도의 상승의 예측으로 가능해진다. 과거의 해수면온도의 월평균편차시계열 $\{X_t\}$, $t=0, 1, \dots, T-1$,을 사용하여 모델(2.6)로 부터 미래 해수면온도의 豫測방정식을 결정한 위의 통계적 과정을 요약하면 다음과 같다: 먼저, $\{X_t\}$ 의 主,副週期를 平滑週期圖를 통해 구한다. 그리고 이것을 初期값으로 하여 週期성분

$$C_t = \sum_{i=1}^p \beta_{i0} \sin(\beta_{i1} + \beta_{i2}t)$$

의 未知係數 $\beta_{i0}, \beta_{i1}, \beta_{i2}(i=1, \dots, p)$ 에 대한 가중최소제곱 추정값을 구한다. 이때 p 는 主,副週期の 갯수로 2 (또는 3)이다. 다음으로, 殘差 $\{R_t^{(1)}\}$, $R_t^{(1)} = X_t - \hat{C}_t$ 의 標本 自己相關 및 部分自己相關함수를 관찰하여 自己回歸過程의 階數 $q(=3)$ 를 결정하고, 또 이 殘差를 이용하여 식 (2.7)에서와 같은 방법으로 未知係數의 最小제곱 추정값을 구한다. 마지막으로, 위에서 결정된 각 성분들로 부터 豫測방정식을 완성한다. 즉,

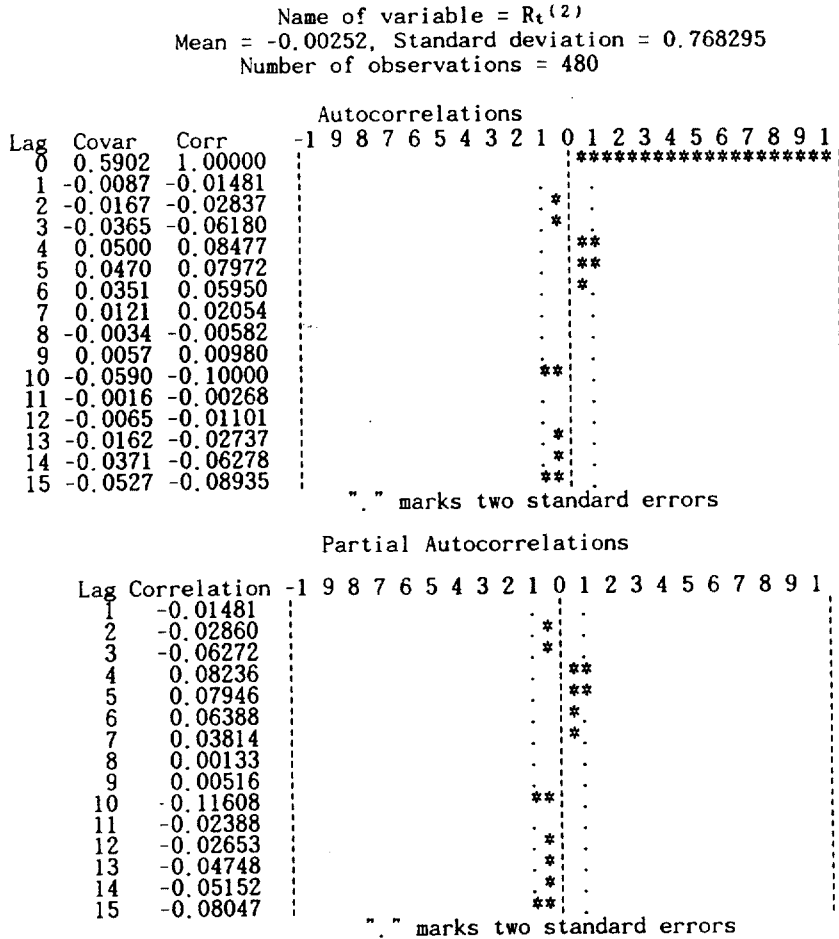


Fig. 8. Sample autocorrelation function and sample partial autocorrelation function, of residual series $\{R_t^{(2)}\}$.

$t \geq T+1$ 에 대하여

$$X_t = \sum_{i=1}^p \hat{\beta}_{i0} \sin(\hat{\beta}_{i1} + \hat{\beta}_{i2}t) + \hat{\mu} + \sum_{\mu=1}^q \hat{a}_\mu Z_{t-\mu} \quad (3.2)$$

이다. 여기서, Z_t , $t = -q, \dots, -1, 0, 1, 2, \dots$ 는 다음식

$$Z_t = \begin{cases} 0 & t = -q, -q+1, \dots, -1 \\ R_t^{(2)} & t = 1, 2, \dots, T \\ \sum_{\mu=1}^q \hat{a}_\mu Z_{t-\mu} & t = T+1, T+2, \dots \end{cases}$$

으로 계산한다.

앞의 결과를 이용한 해수면온도의 월평균편차들에 관한 예측결과는 Fig. 9과 같다. 전반부 480개월(1951-1990)자료를 기초로 하여 후반부 36개월(1991-

1993)을 예측하였다. 전반부(480개월)에서는 실측값과 예측값이 근사적으로 일치하고 있음을 볼 수 있고, 이것은 모델의 適合과 推定의 最適을 의미한다. 또, 이 결과는 前方豫測에 대한 이 豫測方程式의 受容과 適用의 限界를 설명하기도 한다. 즉, 전방예측에 있어서 약 3개월까지는 비교적 정확한 예측이 이루어 진다. 구체적으로 말하면, 이 예측에서는 (絶對)誤差가 平均적으로 0.61°C 정도이고 또 예측 오차가 1°C 이상 발생될 확률은 0.20정도이다. 이에 반하여, 이 간격(3개월)을 넘어선 예측은 週期成分만을 이용한 예측으로 한정되며, 이 때 발생하는 예측오차는 평균적으로 0.78°C 그리고 1°C 이상의 오차가 허용될 확률은 0.32로 각각 증가된다. 이러한 평가는 예측방정식에서 만들어지는 殘差들의 分析과

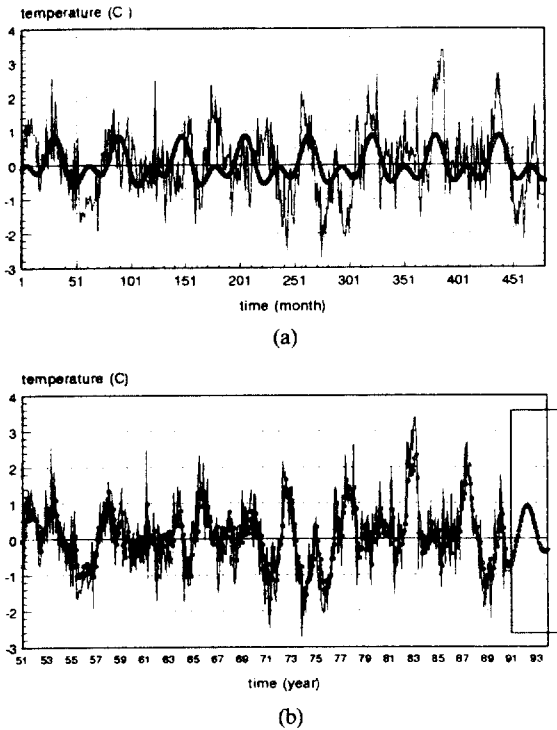


Fig. 9. Plot of original series (-) and predicted values (-) using (a) only cyclic component, (b) both cyclic and stochastic components.

推定된 (階數3의)自己回歸過程이 갖는 기본성질에서 설명된다. 다시 말하여, 週期成分만을 이용한 絶對殘差의 평균이 0.78°C 또 絶對殘差가 $1.0(^{\circ}\text{C})$ 를 초과할 확률이 0.32인데 반하여, 週期成分과 從屬確率成分이 동시에 사용될 때는 평균과 확률이 각각 0.61, 0.20으로 감소되기 때문이다. 더욱이, 階數3의 自己回歸過程에 있어서 예측간격이 3(개월)을 넘을 때는 平均(여기에서는 0.0597)이 예측의 최적값이기 때문이다. 이러한 성질은 定常的인 自己回歸過程의 속성이기도 하다.

위와 같은 이유에서, 해수면온도의 예측에 있어서 종속성분까지 고려된 單變數모델의 적절한 예측간격은 보통 3, 4개월이며, 따라서 이 모델은 長期보다는 6개월이하의 短期豫測에 보다 적절하다고 평가된다. 엘니뇨의 평균지속기간이 10.8개월임을 감안하면 이 단변수모델은 엘니뇨의 정확한 強度豫測에는 다소의 결점을 지니고 있다. 그러나, 엘니뇨의 持續期間과 出現의 예측과 확인을 위한 推檢定에서

는 이러한 결점은 더 이상 제기되지 않는다. 그것은 여기에 이용되는 통계량과 검정량은 從屬成分의 平均에 의존하지 않기 때문이다.

36개월 전방예측값(Fig. 9참조)을 기초로 한 엘니뇨출현과 지속기간에 대한 판정결과는 다음과 같다. 즉, 長期平均水溫을 초과하는 해수면온도의 상승현상이 1991.7-1993.2의 기간에 나타나며, 특히 1991.11-1992.9의 약 10개월 기간에는 엘니뇨현상의 발생이 예측된다.

短期뿐 아니라 長期에서도 효과적인 強度判定에 적용될 수 있는 확률모델을 위해서는 엘니뇨현상의 先導因子를 독립변수로 하는 傳達函數確率모델의 개발을 필요로 한다.

結 論

적도지역 해수면온도의 자료를 통하여 얻은 엘니뇨현상의 統計 및 確率의 특징은 다음과 같다.

첫째, 2節에서 제안된 정의 및 통계적방법의 판정에 의하면, 과거 40년(1951-1990)동안 6회의 엘니뇨가 발생하였고, 평균지속기간은 10.8개월이다. 보통 3-6월에 發生하여 11-1월의 成熟期를 거쳐 다음해 3-4월頃に 끝난다. 6회의 엘니뇨 현상중 3회(1951.4-1952.2, 1965.4-1966.2, 1977.5-1978.2)는 弱, 2회(1972.5-1973.2, 1986.11-1987. 9)는 普通(또는 强) 그리고 1회(1982.6-1983.4)의 경우는 超强의 수준이다.

둘째, 엘니뇨는 60.0월(5년)과 28.2월(2.35년)를 각각 主, 副週期로 하여 발생한다.

셋째, 엘니뇨는 階數 3의 自己回歸過程으로 설명되는 從屬性을 가지며, 따라서 새로운 상승요인이 발생하지 않는 한 한번 상승된 해수면온도는 3개월간 영향이 지속된다.

넷째, 엘니뇨현상을 설명하는 해수면온도의 변동은 確率모델(3.2)로 주어지며, 이 모델을 이용한 예측의 統計의 절차는 3節에서와 같다. 이 때, 週期成分과 從屬成分이 모두 고려된 적정한 전방예측간격은 3, 4개월이고, 이 간격을 넘어선 예측은 週期成分만으로 설명된다. 從屬成分까지 고려된 예측에서는 絶對誤差가 平均적으로 0.61°C 정도 이고, 예측오차가 1°C 이상 발생될 확률은 0.20이다. 또, 週期成分만이 이용되는 長期豫測(6개월이상)에는 평균적으로 0.78°C 정도의 절대오차가 발생하고 또 1°C 이상의

오차가 허용될 확률이 0.32로 증가된다.

다섯째, 3節의 예측방법에 의하면, 平均水溫을 초과하는 해수면온도의 상승현상이 1991.7-1993.2의 기간에 나타나며, 특히 1991.11-1992.9의 약 10개월 기간에는 엘니뇨현상의 발생이 예측된다.

참고문헌

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19: 716-723.
- Deser, C. and J.M. Wallace, 1987. El Niño events and their relation to the Southern Oscillation: 1925-1986. *J. of Geophys. Res.*, **92**: 14,189-14,196.
- DeVries, T.J., 1987. A review of geological evidence for ancient El niño activity in Peru. *J. of Geophys. Res.*, **92**: 14,471-14,479.
- Fisher, R.A., 1929. Tests of significance in harmonic analysis. *Proc. Roy. Soc., Ser. A*, **125**: 54-59.
- Gage, K.S. and G.C. Reid, 1987. Longitudinal variations in tropical tropopause properties in relation to tropical convection and El Niño-Southern Oscillation Events. *J. of Geophys. Res.*, **92**: 14,191-14,203.
- Quinn, W.H. and V.T. Neal, 1987. El Niño occurrences over the past four and a half centuries. *J. of Geophys. Res.*, **92**: 14,449-14,461.
- Schwartz, G., 1978. Estimating the dimension of a model. *Ann. Statist.*, **6**: 461-464.
- Zheng, D.W., G.X. Song and S.F. Luo, 1990. EL Niño prediction. *Nature*, **348**: 8-119.
- Whittle, P., 1952. The simultaneous estimation of a time series' harmonic covariance structure. *Trabajos Estadist.*, **3**: 43-57.

Accepted March 9, 1992