

Wavelet 변환을 이용한 정상 시계열 데이터 해석에 관한 연구

*이준탁 · **최우진 · ***김태홍^o

*동아대학교 전기공학과 교수, **삼양무선(주), *** 동아대학교 대학원 전기공학과

Analysis of Stationary Time Series Using Wavelet Transform

Joon-Tark Lee* · Woo-jin Choi** · Tae-Hong Kim*

* Dong-A University, **SARACO

Abstract - Wavelet analysis is applying to many fields such as the time-frequency localization of a time series and a time varying data.

In this paper, a statistical testing based Wavelet power spectrum analysis for the stationary Nino3 Sea Surface Temperature(SST) data was executed.

Specially, the 95% confidence level for SST was effective in searching the periods of El-Nino using various wavelet basis functions.

1. 서론

Wavelet 변환은 1982년 프랑스의 Morlet에 의해 고안된 인공 지진파의 해석을 도구로 처음 사용되었으며, 이후, 수많은 수학자와 물리학자의 연구에 의해 이론의 수학적 기초를 확립하고, 음향 및 화상처리 분야 등에 응용되고 있다. Wavelet 변환과 푸리에 변환은 서로 다른 기저 함수를 사용하게 된다. 즉, 푸리에 변환은 $\exp(-j\omega t)$ 와 같이 연속이며, 무한히 미분 가능한 함수인 반면, Wavelet 변환은 기본 Wavelet에 시간적으로도 주파수적으로도 국재화시키는 함수를 사용하는 점에서 차이가 있다. 이처럼 Wavelet 변환은 고주파 성분에 대해서는 시간 분해능이 높고, 저주파 성분에 대해서는 주파수 분해능이 높아지므로, 주어진 신호에 대한 지식이 없이도 시간-주파수 해석이 가능하고, 고주파 성분을 갖는 특이점의 검출에도 적합하다. 즉, 국재화 가능한 Wavelet과 해석 대상 신호의 내적을 취하면, 추이 변환된 점 근방에서 패턴의 변화가 발생하므로, 시계열의 변동 주기 및 이상 현상 등을 추정할 수 있다.

따라서, 본 논문에서는 적도 근방의 NiNo3 해수표면 온도 정상 시계열 데이터를 적절히 가공하여, 발생 주기를 탐색하기 위하여 Wavelet 해석하였다. 주기 탐색의 정확성을 예측하기 위하여 여러 종의 Wavelet 함수가 채택되어졌으며, 유의 수준 및 신뢰구간 등의 통계학적 검정법이 이용되었다.

2. Wavelet 변환

Wavelet 변환은 많은 다른 주파수에서 비정상적 전력을 포함하는 시계열을 해석하는 데 사용될 수 있다. 시간 등간격 Δt 및 $n=0, 1, 2, \dots, N-1$ 인 시계열 x_n 에 대해 무차원의 시간 파라미터 η 에 종속되는 Wavelet 함수 $\Psi_0(\eta)$ 이 존재한다고 가정한다. Wavelet 변환이 허용적이 되려면, 평균치가 영이고 시간과 주파수 공간에서 국재화될 수 있어야 한다. 또한, 작고 Wavelet은 이산 Wavelet 변환을 시행하는 경우에 해당되고, 비직교 Wavelet은 이산 및 연속 Wavelet 변환에 통용될 수 있다.

본 연구에서는 통계학적 검정에 대한 해석을 용이하게 하기 위하여 연속 Wavelet 변환을 사용한다.

따라서, 이산 시계열 x_n 의 연속 Wavelet 변환은

$$W_n(s) = \sum_{n=0}^{N-1} x_n \Psi^* \left[\frac{(\eta-n)\Delta t}{s} \right] \tag{1}$$

와 같이 식 (1)과 같이 시계열 x_n 과 스케일 변환되고 친이된 Wavelet 함수 $\Psi_0(\eta)$ 의 합성적분으로 정의된다.

첨자*는 복소 공액을 나타낸다. Wavelet 스케일 s 를 변화시키고 국재화된 시간 지수 n 을 따라 천이시킴으로써, 스케일에 대한 어떤 특성의 진폭은 물론, 진폭이 시간과 변화에 대한 그림을 구성할 수 있다. 식 (1)을 사용해서 Wavelet 변환을 계산할 수 있지만, 푸리에 공간에서 계산하는 것이 훨씬 빠르다. 연속 Wavelet 변환의 근사는 매 스케일 s 에 대한 식 (1)의 합성적분을 N 회 수행하여야 한다. 여기서, N 은 시계열에서의 데이터 수를 나타낸다. 전체 N 회의 합성적분을 수행하는 것은 선택적이며 이보다 작게 취할 수도 있다. 전체 데이터 수가 N 이면, 이산 푸리에 변환을 사용해서 푸리에 공간에서 전체 N 회의 합성적분을 취하는 것이 원칙이다. 즉, 시계열 x_n 에 대한 이산 푸리에 변환은

$$\hat{x}_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N} \tag{2}$$

와 같으며, $k=0, 1, 2, \dots, N-1$ 은 주파수 지수이다. 한편, 스케일 변환된 함수 $\Psi(t/s)$ 의 푸리에 변환은 $\hat{\Psi}(s\omega)$ 로 주어진다. 따라서, 합성정리에 의하여, Wavelet 변환 $W_n(s)$ 은 시계열 x_n 의 이산 푸리에 변환 \hat{x}_k 과 Wavelet 함수의 푸리에 변환 $\hat{\Psi}(s\omega)$ 의 곱을 역푸리에 변환한

$$W_n(s) = \sum_{k=0}^{N-1} \hat{x}_k \hat{\Psi}(s\omega_k) e^{j\omega_k n \Delta t} \tag{3}$$

으로도 정의될 수 있다. 단, 각 주파수는

$$\omega_k = \begin{cases} \frac{2\pi k}{N\Delta t} & k \leq N/2 \\ -\frac{2\pi k}{N\Delta t} & k > N/2 \end{cases} \tag{4}$$

로 정의된다. 각 스케일 s 마다 식(4)와 표준 푸리에 변환 루틴을 사용해서 모든 n 에 대해 효율적이고도 동시에 연속 Wavelet 변환을 계산할 수 있다.

1) Wavelet 함수의 선택

Wavelet 변환은 시계열에 존재하는 특성의 형태를 반영해야 한다. 점프성이거나 계단상의 시계열에 대해서는 Harr와 같은 boxcar 타입의 함수가 선호되며, 완만하게 변화하는 시계열에 대해서는 감쇠 여현파와 같은 smooth한 함수를 선택해야 한다. Wavelet 전력 스펙트럼을 중요시한다면, 어떠한 Wavelet을 택해도 유사한 결과를 나타내므로 Wavelet 변환의 선택은 임의적이다.

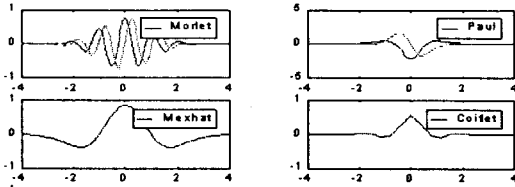


Fig.1 Wavelet Basis Function

직교 Wavelet에서는 각 스케일에 대한 Wavelet basis의 수는 그 스케일에서의 Wavelet basis의 폭과 같으므로, 이산 Wavelet 전력 블록을 포함하는 Wavelet 스펙트럼이 생기고, 신호를 가장 완벽하게 표현할 수 있기 때문에 신호 처리용에 적합하다. 그러나, 본 연구에서 사용된 비직교 Wavelet 함수는 시계열 해석에 유리하고, Wavelet 진폭을 smooth하고 연속적으로 변환시킬 수 있다.

Morlet과 Paul Wavelet은 복소수이나, Mexican Hat 및 Coiflet Wavelet은 실수치이다. 시간 영역에서 이들 Wavelet을 도시하면 그림 1과 같다. 복소 Wavelet 함수는 진폭과 위상에 관한 정보를 제공하고, 진동적인 특성을 포착하는 데 이용될 수 있다. 반면, 실수 Wavelet 변환은 하나의 성분만을 돌려주므로 peak점을 검출하거나 불연속성을 찾는 데 사용되어 질 수 있다.

따라서, Mexican hat Wavelet은 Wavelet 전력의 peak를 분리해내서 시계열의 음과 양의 진동을 동시에 포착할 수 있으며, 복소수 Morlet Wavelet은 Mexican hat의 경우보다 많은 진동을 포함하고 있는 데, Wavelet 전력은 높은 양과 음의 peak가 하나로 결합되어, 시간 공간은 넓게, 주파수 스펙트럼 공간은 좁게 나타난다.

Wavelet 함수가 선정되면, 스케일 s 를 선택해야 한다. 직교 Wavelet 변환에서는 1992년 Farge가 제안한 바와 같이 이산적인 스케일을 사용해야 한다. 그러나, 비직교적인 Wavelet 변환에서는 완전한 형상을 얻기 위해 스케일을 임의로 선택할 수 있다. 즉, 스케일을 다음과 같이 2의 멱승으로 표현하면 편리하다.

$$s_j = s_0 2^{j\Delta j}, \quad j = 0, 1, \dots, J$$

$$J = \Delta j^{-1} \log_2(N\Delta t/s_0) \quad (5)$$

여기서, s_0 는 최소의 분해능 스케일이며 J 는 최대의 스케일을 결정한다. s_0 는 등가 푸리에 주기가 $2\Delta t$ 와 거의 같아지도록 결정하고, Wavelet 함수의 스펙트럼 공간 폭에 의존해서 Δj 를 충분히 작게 선택한다. Morlet Wavelet 변환의 경우, $\Delta j = 0.5$ 는 스케일을 적절하게 샘플링할 수 있는 최대의 값이 된다. 그러나, 다른 Wavelet 변환 함수의 경우는 보다 더 큰 값도 선택해도 무방하다.

2) Wavelet 변환의 정규화

각 스케일 s 에서의 Wavelet 변환을 의미하는 식 (3)의 결과와 직접 비교해 볼 수 있도록, 이들에 대한 Wavelet 변환 $\Psi(s\omega_k)$ 가 단위 에너지를 갖도록 정규화시켜 $\Psi_0(s\omega_k)$ 로 표현한다.

$$\Psi(s\omega_k) = \left(\frac{2\pi s}{\Delta t} \right) \Psi_0(s\omega_k) \quad (6)$$

따라서, 무차원인 각 Ψ_0 는

$$\int_{-\infty}^{\infty} |\Psi_0(\omega)|^2 d\omega = 1 \quad (7)$$

로 된다. 이들을 정규화시키면 매 스케일마다

$$\sum_{k=0}^{N-1} |\Psi(s\omega_k)|^2 = N \quad (8)$$

를 가진다. 따라서, Wavelet 변환은 푸리에 계수 \hat{x}_k 의 진폭치로만 가중되어지며, Wavelet 함수에는 영향을 받지 않는다. 합성적분 공식 (1)을 사용하기 위해, 정규화시키면

$$\psi \left[\frac{(\eta-n)\Delta t}{s} \right] = \left(\frac{\Delta t}{s} \right)^{1/2} \psi_0 \left[\frac{(\eta-n)\Delta t}{s} \right] \quad (9)$$

로 된다

3) Wavelet 전력 스펙트럼

Wavelet 함수 $\Psi(\eta)$ 는 일반적으로 복소수이며, Wavelet 변환 $W_n(s)$ 또한 복소수이다. 변환은 실수부 $Re[W_n(s)]$ 와 $Im[W_n(s)]$ 로 나누어지며, 진폭 $|W_n(s)|$ 및 위상 $\tan^{-1} \frac{Re[W_n(s)]}{Im[W_n(s)]}$ 으로 표현된다.

3. 시계열 해석에의 응용

1) 시계열 데이터

본 논문에서 이용될 시계열 데이터는 ENSO (ElNino-Southern Oscillation)의 측도로서 사용되는 Nino3 해수 표면 온도이다. Nino3 데이터는 5°S-5°N 및 90°-150°W에 걸친 태평양 중앙부의 SST 평균치를 계절별로 나타낸 것이다. 1871년부터 1996년의 데이터는 영국 기상국 GISST2.3의 평균치이며, 1997년 1월에서 6월까지의 데이터는 기상예보센터(CPC)가 Nino3 데이터를 최적으로 보간한 것이다. 계절의 평균치는 이상 시계열을 얻기 위해 제거하였다. Nino3 데이터는 그림 2와 같다.

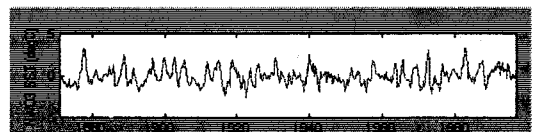
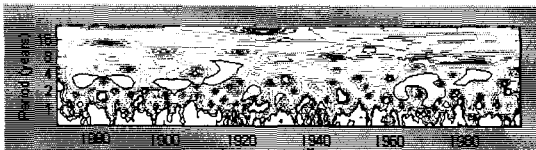


Fig. 2 Time Series Data

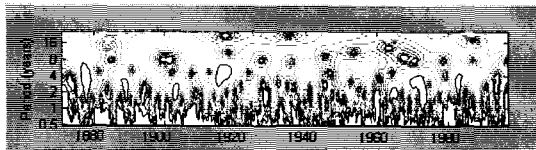
2) Wavelet 해석

그림 3(a)는 Nino3 SST 시계열에 대한 정규화된 Wavelet 전력 스펙트럼 $|W_n(s)|^2/\sigma^2$ 의 값을 나타내며, 0.5년에서 20년까지를 40 스케일로 구분하고 있다. $N=506$, $\Delta t=1/4$ 년, $s_0=2\Delta t$, $\Delta j=0.125$ 및 $J=56$ 으로 백하였다. Δj 의 값은 Wavelet 변환 전력의 그림을 매끈하게 할 정도로 충분하다.

$1/\sigma^2$ 으로 정규화시키면 백색 잡음에 대한 전력의 비를 알 수 있게 된다. 그림 3(a)는 Morlet Wavelet을 사용한 결과로써 긴 주기에서도 다소 큰 전력이 관찰되지만 대부분의 전력이 2-8년의 ENSO 대역 내에 집중되고 있다. 그림 3(b)는 Mexican Hat Wavelet을 사용한 결과로서 주기가 이산적으로 길게 분포되고 있어 해석이 용이하지 않으나, 2-4년의 주기가 관찰되었다. 이러한 연구 결과는 1976년의 Trenberth의 연구결과(Trenberth 1976)와도 일치한다. Wavelet 해석으로부터 ElNino(warm)와 Lanina(cold)의 빈도와 진폭이 변화하고 있음을 알 수 있다. 그림 3(d)에서는 95% 신뢰구간을 사용하지 않고 Coiflet Wavelet을 사용한 결과로서, 1920년대 전까지는 1875-1920년 및 1960-90년대에서는 진폭이 큰 ElNino 현상과 Lanina 현상들이 발생하고 있다. 반면, 1920-1955에는 그런 현상이 거의 없다. 1875-1910에서는, 4년이던 주기가 2년 정도로 조금씩 짧아지고 있으며, 1960-1990에서는 주기가 짧아지다가 다시 길어지는 것을 볼 수 있다.



(a) Morlet



(b) Paul



(c) Mexican Hat



(d) Coiflet

Fig.3 Results of Wavelet Analysis

3) 이론적 스펙트럼과 신뢰도

그림 3(a)~(c)는 국부적 Wavelet 전력 스펙트럼이다. 여기서, 두꺼운 contour로 둘러 쌓인 영역은 95% 이상의 신뢰도를 갖는다. 시계열에 대한 95% 신뢰구간은 표본 평균 \bar{X} 및 시계열 전체 평균 μ , 표준편차 $\sigma_{\bar{X}}$ 에 대하여

$$\bar{X} - 1.96\sigma_{\bar{X}} \leq \mu \leq \bar{X} + 1.96\sigma_{\bar{X}} \quad (10)$$

로 표현된다. Nino3 SST에 대한 95%의 신뢰구간의 경계는 그림 3(a)와 3(b)에서 진한 contour로 도시되었다. 1875-1910 및 1960-1990년은 2-8년 대역에서의 분산이 적색 잡음에 대해 95%이상의 신뢰도를 갖고 있다. 1920-1960년 동안 데이터의 약 5%가 95% 신뢰도의 contour 내에 포함됨을 알 수 있다. 약 2년 및 통상의 2-8년 ENSO 대역의 단부에서는 고립된 신뢰 영역을 갖는다.

4) 재현

역합성적이나 역 필터를 이용해서 원래의 시계열 신호를 재현시킬 수도 있고, 완전히 다른 Wavelet 변환을 사용할 수도 있다. 이 경우, 재현될 시계열은 모든 스케일에 대해

$$x_n = \frac{\Delta j \Delta t^{1/2}}{C_s \psi_0(0)} \sum_{j=0}^J \frac{\text{Re}[W_n(s_j)]}{s_j^{1/2}} \quad (11)$$

와 같이 Wavelet 변환의 실수부를 합한 결과로 된다. Nino3 SST에서는 시계열을 재현한 결과 1.4% 또는 0.087°C의 평균자승 오차를 발생시켰다.

4. 결론

Wavelet 해석은 분산이 변화되거나 다양한 시간 스케일을 갖는 시계열을 해석하는 유용한 도구가 될 수 있다. 해수표면 온도 데이터의 해석시, 본 논문에서 사용된 여러 종의 Wavelet 함수를 이용한 통계학적 검증법은 고차원의 Wavelet 해석에도 적합한 것으로 볼 수 있으나, cross-wavelet등에 관련된 세부적인 검증은 향후의 과제로 남아 있다.

참고문헌

1. Michet Misiti et alii, "MATLAB Wavelet Toolbox ver.1", MathWorks, 1997
2. M. Holschneider; "Wavelets: An Analysis Tool", Clarendon Press, 1995
3. Jacques Lewalle; "Tutorial on Continuous Wavelet Analysis of Experimental Data", 1995
4. M.J.McPhaden; "TOGA-TAO and the 1991-1993 El-Nino-Southern Oscillation Event", NOAA Report, 1994
5. 박정식, 윤영선; "현대통계학", 다산출판사, 1998
6. 김원경; "시계열 분석", 경문사, 1998