

실시간 TOC 자료의 장·단기 성분의 검출을 위한 이산형 웨이블릿 변환의 적용

진영훈·박성천
동신대학교 토목공학과
(2005년 11월 3일 접수; 2006년 9월 14일 채택)

Application of Discrete Wavelet Transform for Detection of Long- and Short-Term Components in Real-Time TOC Data

Young-Hoon Jin and Sung-Chun Park

Department of Civil Engineering, Dongshin University, Naju 520-714, Korea
(Manuscript received 3 November, 2005; accepted 14 September, 2006)

Recently, Total Organic Carbon (TOC) which can be measured instantly can be used as an organic pollutant index instead of BOD or COD due to the diversity of pollutants and non-degradable problem. The primary purpose of the present study is to reveal the properties of time series data for TOC which have been measured by real-time monitoring in Juam Lake and, in particularly, to understand the long- and short-term characteristics with the extraction of the respective components based on the different return periods. For the purpose, we proposed Discrete Wavelet Transform (DWT) as the methodology. The results from the DWT showed that the different components according to the respective periodicities could be extracted from the time series data for TOC and the variation of each component with respect to time could emerge from the return periods and the respective energy ratios of the decomposed components against the raw data.

Key Words : TOC(Total Organic Carbon), DWT(Discrete Wavelet Transform), Periodicity, Return period

1. 서론

현재까지 유기성 오염물질의 정도를 나타내기 위해 BOD나 COD를 사용하고 있다. 그러나 이러한 자료를 오염물질의 지표로 사용하기에는 측정에 소요되는 시간이 길다는 단점과 오염물질의 다양성 및 난분해성 문제, 또한 농도가 극히 낮은 경우 재현성에 어려움이 있다는 점이 지적되었다. 그에 대한 대안으로서 미량의 시료에 대해 고온 하에서 촉매를 이용·연소시켜 시료 중의 유기성탄소를 이산화탄소로 산화시킨 후 적외선 분석기를 이용하여 그 양이 정량적으로 측정될 수 있는 총유기탄소(Total Organic Carbon: TOC)가 제안되고 있다.

BOD나 COD는 시료에 포함되어 있는 물질의 산

화가능 정도와 산소요구량을 나타내지만 전체 유기물질의 양을 나타내지는 않는다. 이에 반해 TOC 측정은 산소요구량에 의한 유기물 농도의 추정치보다는 훨씬 객관적인 수치를 나타내므로 점차 중요한 수질 지표로 활용되고 있다¹⁾. 결과적으로 TOC는 수질오염의 정도를 보다 직접적으로 표현할 수 있다는 점에서 BOD나 COD를 대체할 수 있는 지표로서 사용될 수 있다. 이러한 TOC 자료는 1994년 낙동강에서의 수질오염 사고로 인해 4대강 유역에 수질자동측정망을 설치·운영하여 그 연속적인 측정이 이루어지고 있다²⁾.

따라서 오염물질의 지표로서의 TOC의 시계열에 내재된 경향성(trend) 및 주기성(periodicity)과 같은 확장론적 성분의 특성을 파악하는 것은 현재의 수질 정도를 정확히 파악하여 그에 대해 적절한 대처를 목적으로 하는 수질자료의 예측을 위해 필요할 뿐만 아니라, 미래의 수질관리를 위한 방향설정

Corresponding Author : Young-Hoon Jin, Department of Civil Engineering, Dongshin University, Naju 520-714, Korea
Phone: +82-61-330-2918
E-mail: yhj@dsu.ac.kr

기초적인 자료를 제공하기 위해서도 그 필요성이 절실하다. 따라서 그러한 목적을 달성하기 위해 TOC 시계열 자료의 해석은 필수적이다.

일반적으로 시계열 자료는 상기의 경향성과 주기성 및 무작위 추계성분(random stochastic component)의 선형조합으로 가정될 수 있으며, 특히 경향성과 주기성은 시계열모형을 위해 제거되어야 할 결정론적 성분으로 취급된다³⁾. 즉 대상자료에 내재된 경향성이나 주기성과 같이 비정상성을 나타내는 결정론적 성분들은 시계열의 모형화를 용이하게 하기 위해 제거되어야 하며, 그러한 비정상적 성분들의 추출을 위해 자기상관도와 스펙트럼분석이 주로 이용된다⁴⁾.

대개 자기상관도는 연주기성의 파악을 위해 쉽게 이용될 수 있는 반면, 그 외의 지배적인 주기성분의 파악을 위해서는 스펙트럼분석이 널리 사용되고 있다. 특히 스펙트럼분석은 시계열도에서 확연하게 나타나는 연주기 이외의 다양한 주기를 갖는 재현기간이 다른 지배적인 성분을 추출해 낼 수 있다. 그러나 이러한 재현기간별 특성을 파악하기 위해 이용되는 스펙트럼분석에서의 푸리에 변환(Fourier Transform)은 시간 영역에서의 자료를 주파수 영역으로 변환하면서 그 자료에 대한 시간정보를 잃는다는 점에서 그 한계성이 지적되고 있다⁵⁾.

그러므로 본 연구에서는 시계열 자료의 변환 후에 주파수정보는 물론 시간정보까지 동시에 유지할 수 있는 방법인 웨이블릿 변환(Wavelet Transform)의 적용을 제안하고자 한다. 이러한 웨이블릿 변환은 최근 다양한 분야에서 활발히 적용되고 있으며, 대상자료 별로 살펴보면 강우 및 유출^{6~9)}, 해수면 및 대기압^{10,11)}, 태양의 흑점^{12,13)}, 지진파¹⁴⁾ 자료의 특성을 파악하는 데 이용되었다.

따라서 본 연구에서는 주암호를 대상으로 하여, 해당 호소의 수질 자동측정망을 통해 연속적으로 측정된 TOC 자료에 대해 시계열 자료에 내재되어

있는 재현기간별 성분을 분리·추출하여 각 특성을 파악하고자 한다. 이를 위해 아직 우리나라의 수질 자료에 대해 그 적용이 아직까지 이루어지지 않은 웨이블릿 변환을 이용하여 대상자료의 재현기간별 특성을 이해하고 시계열 자료에 내재되어 있는 장·단기 성분을 파악하고자 한다.

2. 대상지점 및 측정자료

본 연구의 대상지점인 주암호는 유역면적 1,010 km², 저수량 4억 5,700만 t으로 보성군, 순천시, 화순군의 3개 시·군에 걸쳐 있으며, 광주광역시, 나주시, 목포시, 그리고 화순군 등 전라남도 서부권에 하루 64만 t의 생활용수를 공급하고 있다.

본 연구에서 사용한 수질 자동측정망 자료는 전라남도 순천시 대광면 주암리에 위치해 있는 주암호1 지점이며, 사용기간은 2002년 1월 1일 ~ 2004년 12월 31일까지 총 1096개의 일 TOC자료를 이용하였다. 대상자료에 대한 기술통계 분석결과를 시계열 자료(Fig. 1(a))와 함께 박스플롯(Fig. 1(b))에 나타내었다. 박스플롯은 월별 평균값, 평균±표준편차, 최대값 및 최소값을 나타내고 있다. 최대 TOC 농도는 2004년 7월 11일에 검출되었으며, 그 값은 2.928 mg/L이며, 본 연구를 위해 사용된 자료기간에 걸친 평균은 1.460 mg/L로 나타나 전반적으로 양호한 수질농도를 나타내고 있다.

3. 웨이블릿 변환

일반적으로 시계열내에 포함되어 있는 각 주파수별 성분들의 파악을 위해 푸리에 변환(Fourier transform)이 이용되어 왔으나, 이는 앞서 기술한 바와 같이 시간영역에서의 정보를 검출할 수 없다는 단점을 가지고 있다. 이러한 단점을 보완하기 위해 단시간 푸리에 변환(Short Time Fourier Transform: STFT)이 개발되었으나, 이 방법을 위해 사용되는 시간영역에서의 고정된 창함수(window function)는

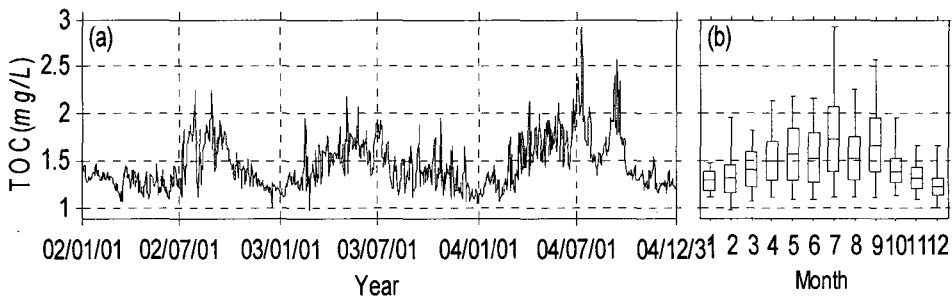


Fig. 1. Time series plot of daily TOC data and box-whisker plots with mean, mean±standard deviation, and maximum/minimum on a monthly basis.

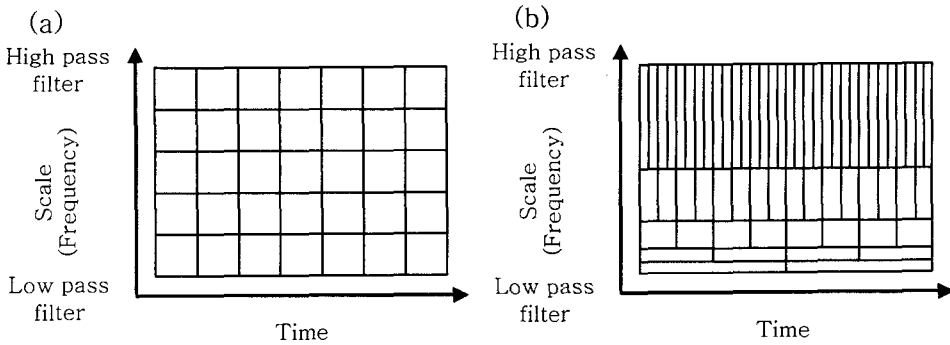


Fig. 2. Resolution in time-frequency domain: (a) STFT, (b) Discrete wavelet transform.

Fig. 2(a)에서 보여지듯이 시간-주파수의 전영역에서 동일한 분해능을 사용한다¹⁵⁾. 이러한 단시간 푸리에 변환의 경직된 분해능은 시계열자료를 시간-주파수 영역으로 변환할때 고정된 범위에 포함되지 않는 성분에 대해서는 해석이 곤란하며¹⁰⁾, 창 크기가 항상 일정하기 때문에 비정상 자료에 대해서도 그 적용이 곤란하다.

반면에 단시간 푸리에 변환의 단점을 보완하기 위해 개발된 웨이블릿 변환은 대상자료의 시계열을 각기 다른 주파수 성분으로 분리하는 도구이며, 각 스케일(주파수)에 따른 유연한 분해능에 의해 자료에 포함되어 있는 다양한 성분들을 조사할 수 있는 도구이다.

웨이블릿 변환은 대상자료와 웨이블릿 함수 사이의 회전적분으로 정의되며, 스케일 및 시간에 따라 달라진다. 이러한 과정을 위해 적용되는 웨이블릿 함수는 스케일과 전이항에 의해 그 크기와 위치가 변하여 Fig. 2(b)에 보여지듯이 시간-주파수 영역에서 그 분해능의 유연성을 나타낼 수 있다¹⁶⁾.

Heisenberg의 불확정성 원리에 의하면 대상자료의 시간-주파수영역으로의 변환에 있어 정보를 동시에 완벽하게 국소화시킬 수 있는 변환은 존재할 수 없으나¹⁷⁾, 웨이블릿 변환은 Fig. 2(b)에 나타난 바와 같이 저주파수영역에서는 주파수 분해능을 증가시키고 고주파수 영역에서는 시간영역의 분해능을 증가시키는 보완성을 지니고 있다¹⁸⁾. 따라서 웨이블릿 변환은 지금까지 개발된 시간-주파수영역에서의 변환방법들 중에서 가장 유연하며, 비정상성이나 급격한 변화를 보이는 자료에 적합하다.

Fig. 3은 본 연구에서 적용된 웨이블릿 변환의 최종단계까지의 반복 적용과정을 나타내고 있다. 즉 TOC 원시계열 자료는 제 1단계에서의 분해를 통해 근사(approximation)성분과 상세(detail)성분으로 나뉘어지며, 근사성분은 제 2단계의 분해를 위해 사용된다. 이와 같은 분해를 위해 사용되는 기저 웨이

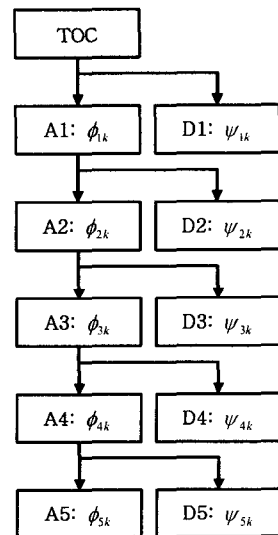


Fig. 3. Discrete wavelet transform at each decomposition level.

렛 함수를 Fig. 4에 나타내었다. Fig. 4(a)는 스케일링 함수($\phi_{j,k}(t)$)를 도시하고 있으며 이에 의해 근사(approximation)성분이 추출되며, Fig. 4(b)에 보이는 웨이블릿 함수($\psi_{j,k}(t)$)는 상세(detail)성분을 분리한다. 여기서 j 와 k 는 각각 스케일과 전이항을 나타내며, 선택된 기저 웨이블릿을 시간-주파수 영역에서 다른 분해능과 위치를 갖도록 변형시키는 역할을 한다. 각 함수는 단계별 스케일항과 전이항에 따라 변화하는 웨이블릿군(wavelet family)을 구성하며, 그 수학적 표현으로 Eq. (1)에 스케일링 함수를, 그리고 Eq. (2)에 웨이블릿 함수를 각각 나타내었다.

$$\phi_{j,k}(t) = 2^{-j/2} \phi(2^{-j}t - k) \quad (1)$$

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \quad (2)$$

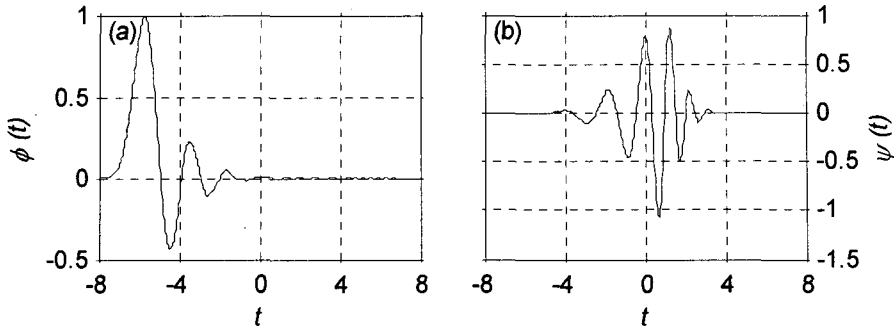


Fig. 4. (a) Scaling function and (b) wavelet function for the present study.

스케일링 함수에 의해 분리·추출된 근사성분은 다음 단계에서 다시 근사성분과 상세성분으로 분리되며, 이러한 과정은 최종분해단계까지 이산적으로 반복하여 진행된다 (Fig. 3 참조). 최종분해단계의 선택은 일반적으로 적용되는 웨이블릿 함수와 대상 자료의 길이에 따라 결정되며, 최종단계에서 적어도 하나 이상의 웨이블릿 계수가 유효해야 한다는 규칙에 바탕을 둔다.

본 연구에서는 Daubechies에 의해 제안된 함수들을 적용하여 각 웨이블릿 함수와 사용 자료의 길이에 따라 결정되는 최종분해단계 및 원자료와 최종분해단계에서의 근사성분 사이의 상관계수를 구하여 웨이블릿 함수를 선택하였다. 그 결과 원자료와 근사성분간의 상관계수값이 0.82로 비교적 높은 값을 나타내며, 본 연구에서 목적으로 하는 장기 성분의 추출을 위해 1096개의 TOC 자료를 5단계까지 분해할 수 있는 함수인 Daubechies의 12번 함수를 선택하여 웨이블릿 변환을 최종분해단계까지 반복 수행하였다.

4. 결과 및 고찰

본 연구에서는 주암호1 지점의 수질자동측정망에 의해 측정되어 축적된 자료인 TOC 시계열 자료에 이산형 웨이블릿 변환(DWT)을 적용하였으며, 최종분해 가능 단계 및 원시계열 자료와 근사성분간의 상관계수를 고려하여 웨이블릿 함수를 선택하였으며, 그 결과 Daubechies의 12번 웨이블릿 함수('db12')가 높은 상관성을 나타낼 뿐만 아니라 본 연구에서 사용한 1096개의 자료를 5단계까지 분해할 수 있는 것으로 나타났다.

따라서 원자료에 'db12'를 적용하였으며, 각 단계에 따라 그 크기를 달리하는 스케일항과 전이항을 이용하여 최종분해단계인 5단계까지 이산적으로 반복 적용하여 매 단계에서의 상세성분과 최종단계에서의 근사성분을 분리·추출하였다. 분리·추출된 성

분들을 Fig. 5에 도시하였으며, 그림에서 보이는 바와 같이 1단계에서 분리된 상세성분의 경우 그 변동의 빈도가 높은 고주파수 특성들을 포함하고 있으며, 단계가 증가할수록 그 빈도가 낮은 저주파수에 해당하는 특징들을 내포하고 있다. 특히 근사성분의 경우 대상자료에 대해 분해가능한 최대단계 내에서 가장 낮은 주파수 성분을 나타내고 있으며, 이는 본 연구에서 이용한 일 TOC 시계열 자료가 포함하고 있는 가장 긴 주기를 갖는 장기성분의 특성을 반영하고 있다.

또한 단기 및 장기성분에 해당하는 각각의 상세 및 근사성분에 대한 특성을 살펴보기 위해 각 단계별 스케일과 그에 따른 주파수 및 재현기간을 Table 1에 나타내었다. 재현기간에 따른 각 성분들의 주기성은 1단계에서의 약 3일부터 최종단계인 5단계에서의 약 46일까지 다양하게 적용되었다. 각 단계에서 분리·추출된 상세 및 근사성분들이 원자료를 설명하고 있는 정도를 나타내는 정량적인 지표인 에너지를 표시하였다.

표에서 보이는 바와 같이 모든 상세성분들이 차지하는 에너지 비율의 총합이 1% 이하로 나타났으며, 약 46일 이상의 주기성을 갖는 근사성분의 경우 원자료를 99% 이상 설명할 수 있는 것으로 나타났다. 이는 단주기성을 갖는 상세성분들이 원자료에 대해 극히 낮은 설명력을 보이는 반면, 장주기성을 갖는 근사성분이 원자료의 본질적인 거동을 나타내고 있는 것으로 판단된다.

5. 결론

본 연구에서는 주암호1 지점에서 측정된 일 TOC 시계열 자료의 장·단기 성분의 검출을 위해 이산형 웨이블릿 변환을 적용하였다. 그 결과 분리·추출된 성분들 중 상세성분이 차지하는 설명력이 1% 이하로 낮은 반면, 최종분해 단계의 근사성분의 설명력이 99% 이상을 보여주고 있었다. 이는 대상자

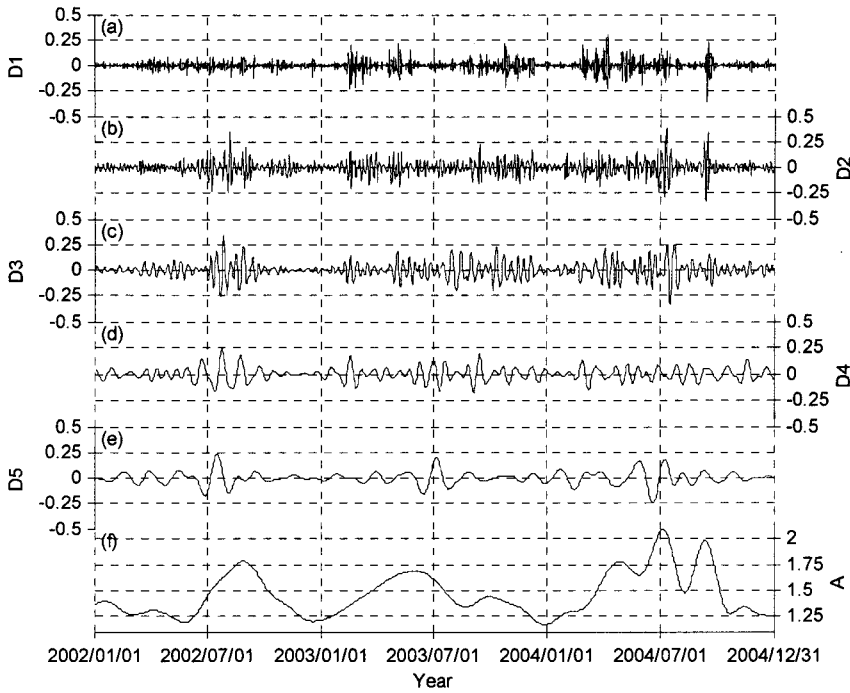


Fig. 5. Results from Discrete wavelet transform: ((a)~(e)) detail and (f) approximation components.

Table 1. Scale, frequency, return period, and energy corresponding to the respective levels

Level (j)	Scale ($a=2^j$)	Frequency (1/Day)	Period (Day)	Energy (%) of details	Energy (%) of approximation
1	2	0.3478	2.875	0.0928	
2	4	0.1739	5.750	0.1417	
3	8	0.0870	11.500	0.1809	
4	16	0.0435	23.000	0.1308	
5	32	0.0217	46.000	0.2152	99.2388

료로부터 이산형 웨이블릿 변환에 의해 분리된 46 일 이상의 장기성분이 원자료를 유의하게 표현해 낼 수 있음을 의미한다.

또한 그 주기가 약 3일부터 46일까지로 근사성분에 비해 상대적으로 단주기성을 나타내고 있으며, 낮은 비율의 설명력을 가지고 있는 상세성분들은 원자료의 측정시 발생할 수 있는 측정 및 기록 오차에 의한 단순 잡음 및 다른 자연현상의 간섭으로 발생할 수 있는 동역학적 잡음성분으로 간주될 수 있는 것으로 판단된다.

결론적으로, 본 연구에서 사용된 원자료는 그 대부분이 이산형 웨이블릿 변환에서 분리·추출된 근사성분에 의해 설명되고 있으며, 따라서 추출된 근사성분이 원자료의 본질적인 거동을 반영하고 있다고

판단된다. 이러한 결론은 향후 대상자료의 예측 모형의 개발을 위한 기초자료로 활용될 수 있을 것으로 기대된다. 또한 단주기성을 나타내는 상세성분의 극히 낮은 비율의 설명력을 고려할 경우, 대상자료의 예측모형 개발시 잡음성분으로 간주할 수 있는 것으로 판단되며, 본 연구에서 적용한 이산형 웨이블릿 변환방법이 잡음제거 또는 저감 방안으로 이용될 수 있을 것으로 기대된다.

참고 문헌

- 1) 정명규, 이동석, 김현기, 황갑수, 신현상, 김정호, 박홍재, 김정배, 2001, 환경분석화학, 동화기술, 406pp.
- 2) 안상진, 연인성, 2004, 실시간 자동측정망 자료

- 를 이용한 수질관리, 대한토목학회논문집, 24 (3B), 221-228.
- 3) 조용준, 김종문, 1998, Wavelet transform을 이용한 물수요량의 특성분석 및 다원 ARMA모형을 통한 물수요량 예측, 한국수자원학회논문집, 31(3), 317-326.
 - 4) 윤태훈, 1997, 응용수문학, 청문각, 716pp.
 - 5) Smith, L. C., D. L. Turcotte and B. L. Isacks, 1998, Stream flow characterization and feature detection using a discrete wavelet transform, Hydrological Processes, 12, 233-249.
 - 6) Kumar, P. and E. Foufoula-Georgiou, 1993a, A multicomponent decomposition of spatial rainfall fields 1. Segregation of large- and small-scale features using wavelet transform, Water Resources Research, 29(8), 2515-2532.
 - 7) Kumar, P. and E. Foufoula-Georgiou, 1993b, A multicomponent decomposition of spatial rainfall fields 2. Self-similarity in fluctuations, Water Resources Research, 29(8), 2533-2544.
 - 8) Cahill, A. T., 2002, Determination of changes in streamflow variance by means of a wavelet-based test, Water Resources Research, 38(6), 1-1-1-14.
 - 9) Kim, S., 2004, Wavelet analysis of precipitation variability in Northern California, U.S.A., KSCE Journal of Civil Engineering, 8(4), 471-477.
 - 10) Torrence, C. and G. P. Compo, 1998, A practical guide to wavelet analysis, Bulletin of American Meteorological Society, 79(1), 61-78.
 - 11) Drago, A. F. and S. R. Boxall, 2002, Use of the wavelet transform on hydro-meteorological data, Physics and Chemistry of the Earth, 27, 1387-1399.
 - 12) Oh, H. S., C. M. Ammann, P. Naveau, D. Nychka and B. L. Otto-Bliesner, 2003, Multi-resolution time series analysis applied to solar irradiance and climate reconstructions, Journal of Atmospheric and Solar-Terrestrial Physics, 65, 191-201.
 - 13) Polygiannakis, J., P. Preka-Papadema and X. Moussas, 2003, On signal-noise decomposition of time-series using the continuous wavelet transform: application to sunspot index, Monthly Notices of the Royal Astronomical Society, 343, 725-734.
 - 14) Pazos, A., M. J. González and G. Alguacil, 2003, Non-linear filter, using the wavelet transform, applied to seismological records, Journal of Seismology, 7, 413-429.
 - 15) Shin, T. and I. Han, 2000, Optimal signal multi-resolution by genetic algorithm to support artificial neural networks for exchange-rate forecasting, Expert Systems with Applications, 18, 257-269.
 - 16) Daubechies, I., 1992, Ten lectures on wavelets, SIAM, Philadelphia, PA, 1-352.
 - 17) Mallat, S., 1998, A wavelet tour of signal processing, Academic, San Diego, 2-590.
 - 18) 박형기, 박종열, 2003, 웨이블릿변환을 이용한 구조물의 동적 파라미터 추정, 대한토목학회논문집, 23(4A), 733-742.