Journal of Korean Society on Water Quality, Vol. 22, No. 5, pp.831-837 (2006)

수질 및 강우자료의 상관분석을 위한 웨이블렛 변환의 적용

진영훈† · 오창열 · 박성천

동신대학교 토목공학과

Application of Wavelet Transform for Correlation Analysis between Water Quality and Rainfall Data

Jin, Young Hoon† · Oh, Chang Ryol · Park Sung Chun

Department of Civil Engineering Dongshin University

(Received 13 February 2006, Accepted 24 July 2006)

Abstract

The present study applies wavelet transform for the extraction of various periodicities which are included in TOC and pH time series of water quality and rainfall data. The primary objective of the present study is to detect the relationships between the respective data through the correlation analysis using the approximation components which are decomposed by wavelet transform. The results reveal the approximation components of TOC and pH in the 5th level of wavelet transform can explain more than 99% of the whole energy for the raw data respectively and there are considerably high correlation between the approximation components of the respective data used for the study even through no significant correlation between the raw data has been detected.

keywords : Approximation component, Detail component, Periodicity, pH, Rainfall data, Total organic carbon (TOC), Wavelet transform

1. 서 론

1991년 낙동강 페놀유출사고, 1994년 낙동강 휘발성유기용제류 유출사고 등을 계기로 수질의 상시 감시기능과 조기경보체계의 구축을 위한 수질자동측정망 설치의 필요성이 대두되어 현재 전국 4대강 유역에 수질자동 측정망을 설치 운영 중에 있다. 이러한 수질자동측정망의 설치의 목적을 달성하기 위해서는 수질자료에 적합한 분석 및 예측모형의 개발과 현장적용이 이루어져야 한다.

수질자동 측정망을 통해 관측되고 있는 자료들 중 총유기탄소(Total Organic Carbon: TOC)는 시료 중에 용존되어 있는 유기물을 산화과정을 통하여 CO_2가스를 발생시키고, 그 발생량을 CO_2검출기를 이용하여 측정한 수질지표이다. 이는 기존의 BOD 및 COD 수질농도 측정에 비하여 적은 오차요인과 빠른 시간에 결과 값을 얻을 수 있다는 장점이 있어 수질지표로 사용되는 데 적합하다 하겠다.

이러한 장점들로 인하여 정부에서는 오염부하량을 할당받은 자는 수질환경보전법시행규칙 제27조, 제45조, 제47조 및 하수도법시행규칙 제6조, 오수 · 분뇨및축산폐수에관한법률시행규칙 제30조, 제43조 및 제67조의 규정에 의하여 BOD 환산식에 따라 BOD 측정값 산출이 적정하다고 판단되는 TOC 및 COD 농도 자동측정을 하여야 한다고 규정하고 있다.

또한 수질자동 측정망을 통해 관측되고 있는 자료들 중 용액의 산도를 나타내는 수질 지표인 pH는 오염물질에 대한 생물학적 처리시에 pH 변화에 따라 미생물이 민감하므로 수처리공정에서의 pH값 관측은 중요하다.

TOC 자료를 적용한 연구를 살펴보면, 안 등(2004)은 평창강, 달천, 여주지점의 실시간 TOC 수질자료를 위한 예측모형을 개발하였으며, 이와 더불어 예측모형에 판단모형의 접합을 통해 기존의 인공신경망에 의한 수질예측모형보다 우수한 결과를 나타내었다. 그러나 TOC 시계열 자료 자체에 대한 분석에 관한 연구는 아직까지 깊이 있게 이루어지지 않았다. 따라서 본 연구에서는 TOC 시계열 자료에 내재되어 있는 특성들 중에서 주기성분에 관한 특징을 알아보고자 하며, 그를 위해 시계열 자료의 변환 후 시간 및 주파수 영역에서의 정보를 구할 수 있는 웨이블렛 변환(wavelet transform)을 적용하고자 한다.

자료의 특성을 검출하기 위한 웨이블렛 변환의 적용은 다양한 연구 분야에서 적용하였으며 국내의 사례를 살펴보면 다음과 같다. 물소비특성을 분석하기 위하여 조 등(1998)은 웨이블렛 변환을 위한 기저함수로 Coiflets5를 사용하였으며 박 등(2004)은 서울지역의 오존농도를 상시 관측하는 오존자료를 자료에 대하여 웨이블렛 변환함수 Coiflets5를 적용한 결과 23.98 hr의 강한 주기성을 확인하였다. 또한 권 등(2005)은 국내 강우량 자료 및 SOI, SST

† To whom correspondence should be addressed.
yhjin@dsu.ac.kr

의 수문기상자료에 대한 웨이블렛의 적용성 평가에서 통계적으로 유의한 결과를 제시하였다. 또한 외국에서의 수문 및 기상자료에 대한 웨이블렛 변환에 관한 연구를 살펴보면 해수위 및 대기압의 시계열 자료에 적용하여 대기압의 변화에 따른 상반된 해수의 변화에 대한 연구를 수행하였다(Drago, 2002).

따라서 본 연구에서는 섬진강 유역의 주암호 수질자동측정망 지점에서 측정되고 있는 수질자료 중 TOC 및 pH 시계열 자료와 주암호 인근의 순천 강우관측소의 일 강우량 자료에 내재되어 있는 특성을 파악하기 위해 웨이블렛 변환을 적용하여 각 변수들이 포함하고 있는 장 · 단기 성분의 추출을 통해 각각의 변화양상을 파악하고자 한다. 또한 각 변수들 간의 상관관계를 원자료가 아닌 웨이블렛에 의해 변환된 시계열 자료를 이용하여 그 정도를 파악하고자 하며, 이를 바탕으로 호소수 수질관리 및 환경정책입안의 기초자료를 제공하고자 한다.

2. 이론적 배경

웨이블렛 변환은 1900년대 초 Haar에 의해서 처음 제안되었으며, 지금까지 신호 처리 분야에서 가장 널리 사용된 푸리에 변환(Fourier Transform)과 같은 직교 변환 방법으로써 자료나 함수 등을 각기 다른 여러 주파수 성분으로 분해할 수 있으며, 각 스케일에 해당하는 해상도에 따라 분리된 각 성분들을 해석하는데 사용하였다. 웨이블렛 변환에 대한 연구는 1980년대에 들어서면서부터 본격적으로 진행되었으며 1990년 전까지 수학 분야에서 Morlet과 Grossman 등에 의해서 실증적으로 연구되었으며 공학 분야에서는 Vetterli과 Smith 등에 의해서 부대역 부호화(Filter Banks, Subband Coding)가 연구되어 왔다. 1990년 후에 Daubechies 등에 의해서 이 두 분야가 하나의 이론이라는 것이 수학적으로 유도되어 웨이블렛 이론으로 결합되었다(강 등, 2001).

이러한 웨이블렛 변환은 전통적으로 신호처리에 사용되어온 푸리에 변환에 비해 다음과 같은 장점을 가지고 있다. 첫째, 웨이블렛 변환은 푸리에 변환에 비하여 계산속도가 빠르며, 둘째 웨이블렛 변환은 주어진 신호에 대한 시간과 주파수 영역에서의 정보를 균형적으로 국소화시킬 수 있다. 마지막으로 불규칙한 데이터를 평탄하게 해줌으로써 데이터에 대한 예측이 가능하게 해주는 평탄성(smoothness) 분석이 가능하다는 장점이 있다(Daubechies, 1994).

웨이블렛 변환의 기저 함수로 사용되는 $\Psi(t)$를 모함수(mother wavelet)라고 하며 그 수학적 표현은 식 (1)과 같다. 식 (1)은 기저 함수의 스케일링과 전이를 나타내며 여기서 j는 스케일링을 결정하는 값이고, k는 함수를 얼마나 이동시킬 것인가를 결정하는 값이며, 정수범위에서 정의된다.

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) \tag{1}$$

웨이블렛 변환은 식 (1)에서 모함수를 k만큼 이동하고 j에 의해 크기가 변화하는 웨이블렛 함수를 사용한다. 이는 고주파로 갈수록 웨이블렛은 함수의 폭이 좁아지고, 저주파로 갈수록 함수의 폭이 넓어지는 것을 나타낸다.

웨이블렛 변환은 웨이블렛 기본 함수들의 중첩으로 임의의 함수를 표현하는 것이며, 이러한 웨이블렛 기본 함수들의 중첩은 각각 다른 스케일 레벨을 가지고 임의의 함수를 만들어 내며, 각 레벨은 그 레벨에 맞는 해상도를 가지게 된다. 결국 이산형 웨이블렛 변환은 각각의 스케일과 전이항에 의해 다음 식 (2)와 같다.

$$f(t) = \sum_{k=-\infty}^{\infty} a_{J,k}\phi_{J,k}(t) + \sum_{j=1}^{J}\sum_{k=-\infty}^{\infty} d_{jk}\,\psi_{j,k}(t) \tag{2}$$

여기서 J는 최대 분해단계를 나타내며 일반적으로 자료의 수에 따라 그 값이 결정된다. 최대 분해단계에 의해 도출된 계수인 $a_{J,k}$는 가장 낮은 주파수 성분 즉 가장 넓은 스케일 영역을 나타내며 대상자료($f(t)$)와 스케일링 함수($\phi_{j,k}(t)$)의 내적에 의해 산정된다. 반면에 각각의 이산형 스케일 j와 전이항 k에 의해 산정되는 $d_{j,k}$는 최대 분해단계 이전의 각 단계에서의 주파수 성분들에 대한 계수들이며, 원자료($f(t)$)와 웨이블렛 함수($\psi_{j,k}(t)$)의 내적값이며 본 연구에서 적용한 웨이블렛 함수는 Fig. 1과 같다.

3. 대상지점 및 자료

본 연구의 대상지점인 주암호는 유역면적 1,010 km^2, 저수량 4억 5700만 ton으로 보성군 · 순천시 · 화순군의 3개 시군에 걸쳐 있으며, 광주광역시 · 나주시 · 목포시 · 화순군 등 전

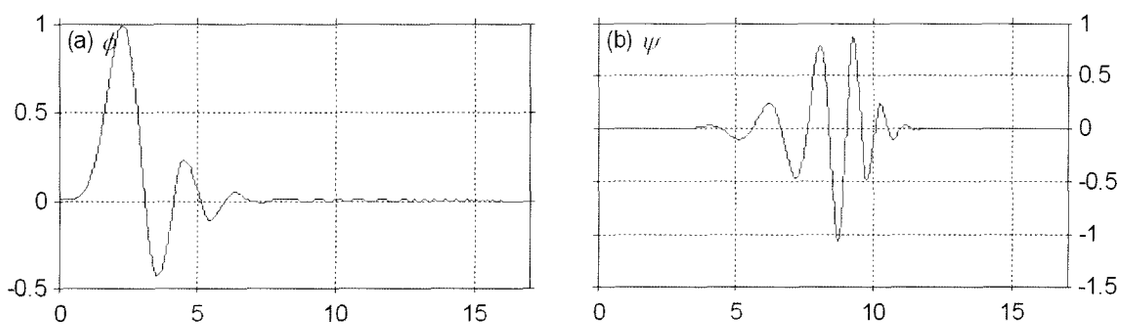

Fig. 1. (a) Scaling function and (b) wavelet function for the present study.

라남도 서부권에 하루 64만t의 생활용수를 공급하고 있다.

본 연구에서는 주암호에 위치한 수질관측소 중 주암호1 수질관측망 자료 중 TOC, pH의 일평균 수질자료와 주암호 1 수질관측망 인근에 위치한 순천강우관측소의 일강우량을 자료에 대하여 2002년 1월 1일 ~ 2003년 12월 31일까지 총 730개의 일평균 수질 및 강우량자료를 이용하였으며 대상 지점의 유역도는 Fig. 2와 같다.

본 연구의 대상 자료에 대한 기술통계 분석 결과는 Table 1에 나타낸 바와 같이 TOC 농도는 0.98~2.27 mg/L의 분포를 보였으며 평균농도는 1.42 mg/L로 비교적 양호한 수질농도로 나타났다. pH 수질자료는 계절적 요인에 의하여 겨울철에는 약산성의 특성을 보이다가 봄·여름·가을에는 약염기성의 특성을 갖는 것으로 나타났다. 순천지방의 대상 일강우량 자료 중 최대일강우량은 254.5 mm로 나타났으며 일평균강우량은 5.23 mm이다. 또한 대상 자료에 대한 시계열 및 월별 박스플롯은 Fig. 3 ~ 5와 같으며 순천지방 일별 강우량 자료의 강한 계절성은 월별 박스플롯 그래프에서 잘 보여주고 있다.

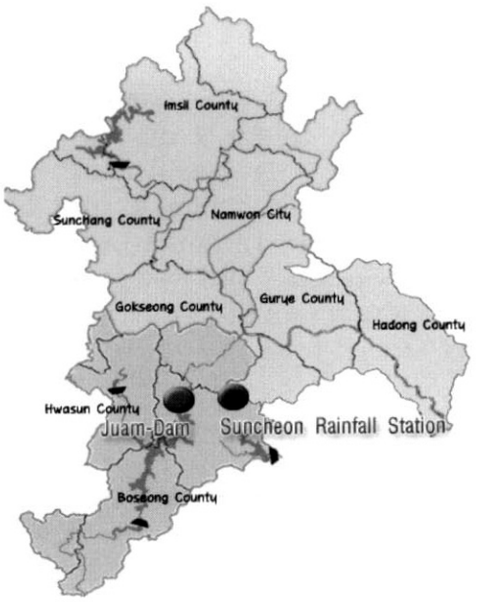

Fig. 2. Location of the study area.

4. 자료의 변환 및 결과

본 연구에서는 Daubechies에 의해 제안된 다양한 웨이블렛 함수 중 다음과 같은 두 가지의 기준에 의해 최적의 함수를 선택하였다. 그 기준으로는 대상자료에 대한 웨이블렛 변환 후 재현기간 45일 이상 해석이 가능한 5단계 이상의

Table 1. Descriptive statistics of data used

	TOC	pH	Rainfall
Max	2.27	9.30	254.50
Min	0.98	6.60	0.00
Mean	1.42	7.77	5.23
Stdev	0.22	0.82	17.90

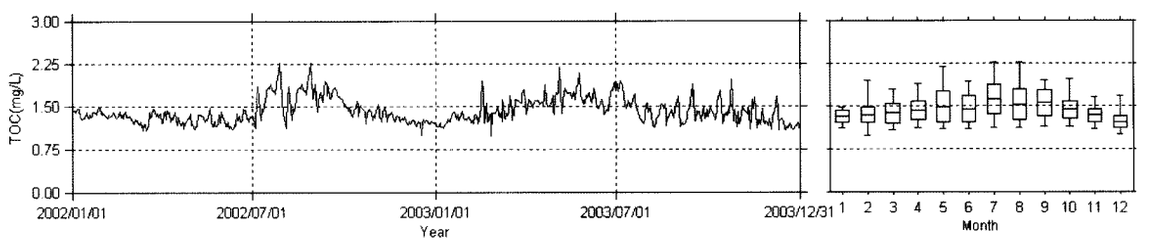

Fig. 3. Time series data and monthly box plot of daily TOC.

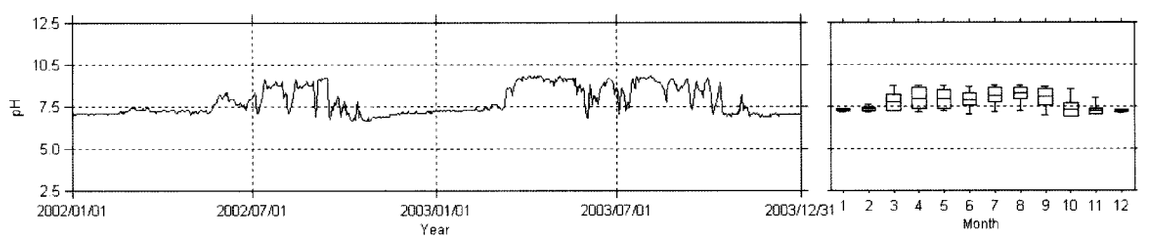

Fig. 4. Time series data and monthly box plot of daily pH.

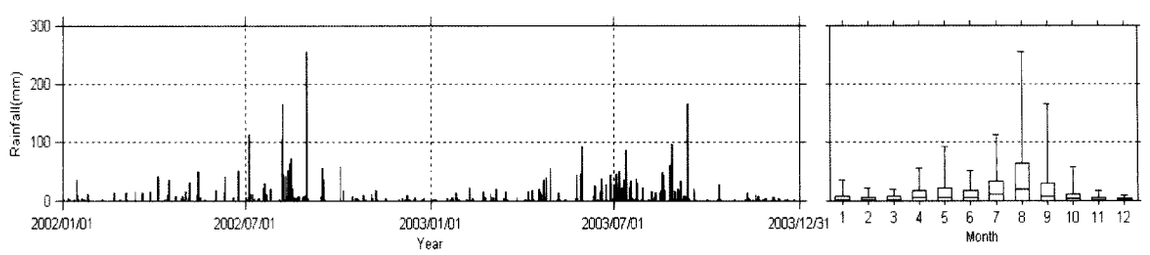

Fig. 5. Time series data and monthly box plot of daily rainfall.

분해 가능한 함수 그리고, 원자료와 최종분해단계 근사성분과의 상관계수가 가장 높은 웨이블렛 함수를 선택하였다. 그 결과 pH의 웨이블렛 함수는 Daubechies의 9번 함수('db9'), 강우량 및 TOC 자료에 대해서는 10번 함수('db10')가 가장 높은 상관계수 및 45일 이상 해석이 가능한 5단계까지 분해할 수 있는 것으로 나타났다.

Fig. 6~8은 TOC, pH의 수질자료 및 강우자료에 대한 웨이블렛 함수를 적용하여 1단계부터 5단계까지 추출된 상세성분(Detail component : D)과 근사성분(Approximation component : A)의 결과값을 도시하였으며 웨이블렛 변환의 각 단계별 스케일, 주파수, 재현기간 및 에너지 정보를 Table 2~4에 나타내었다. 여기서 에너지 정보는 웨이블렛 변환에 의한 각 단계별로 추출된 상세 및 근사성분이 원자료를 설명하고 있는 정도를 나타내는 정량적인 지표로서 다음의 식과 같다.

$$Energy(P(A,\ D_j)\ ;\ \%) = \frac{\sum_{n=1}^{k} P_n^2}{\sum_{n=1}^{k} R_n^2} \times 100 \qquad (3)$$

여기서, A는 근사성분, D_j는 각 단계별 상세성분, R은 원자료, P는 웨이블렛 변환된 근사 또는 상세성분 자료, k는 자료의 수이다.

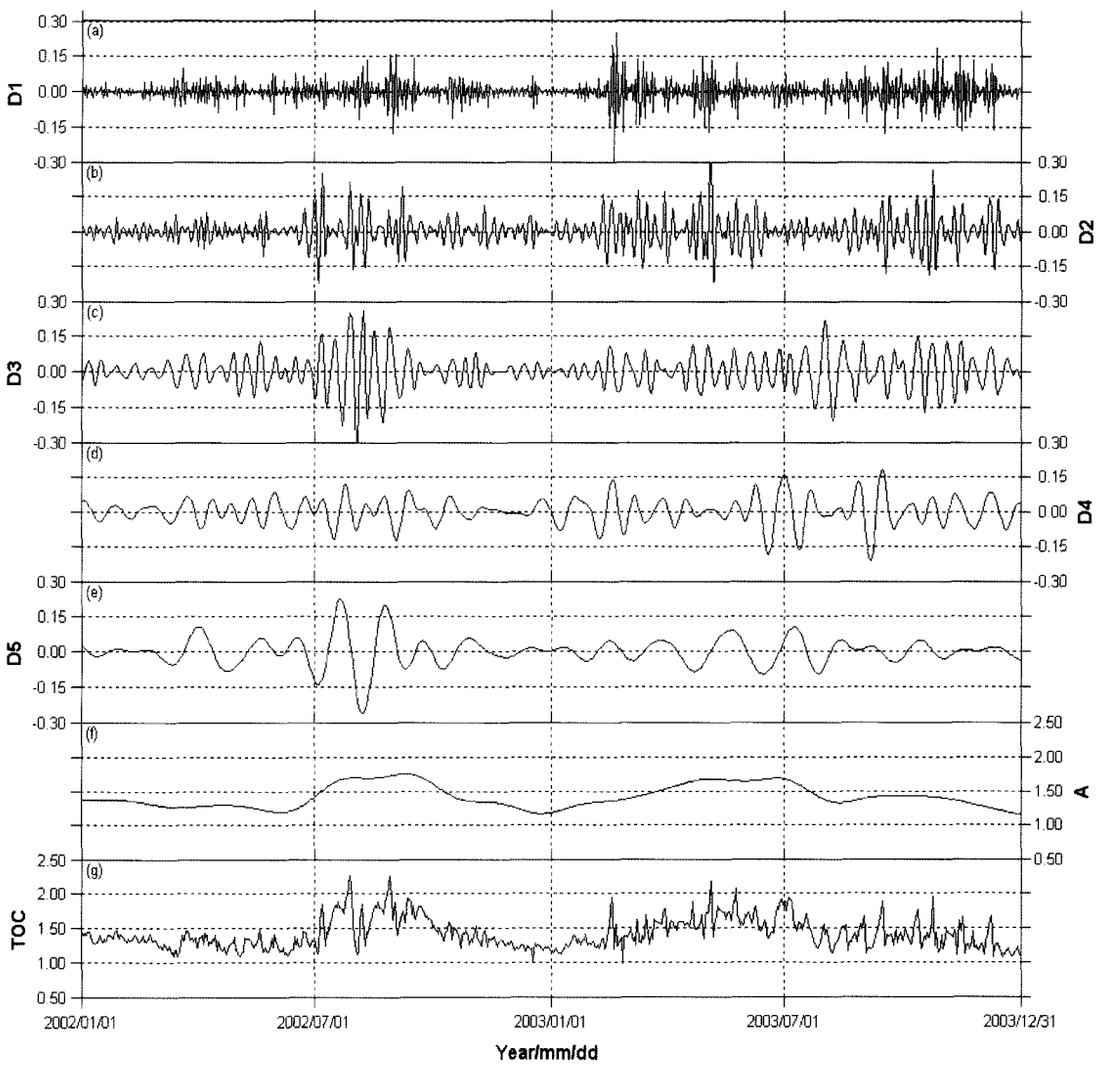

Fig. 6. Discrete wavelet transform and its approximation and detail components about daily TOC data.

Table 2. Each step scale, frequency, energy values result of discrete wavelet transform about daily TOC data

	Level (j)	Scale ($a=2^j$)	Frequency (1/Days)	Period (Days)	Energy(%) of details	Energy(%) of approximation
TOC	1	2	0.3421	2.9230	0.08	
	2	4	0.1712	5.8462	0.12	
	3	8	0.0855	11.6918	0.16	
	4	16	0.0428	23.3863	0.12	
	5	32	0.0214	46.7727	0.15	99.37

원자료에서의 계절성 및 연주기성이 파악이 불명확한 TOC 자료에 대하여 웨이블렛 변환 함수('db10')을 적용한 결과 약 47일의 강한 주기성을 갖는 것으로 나타났다. 이는 Table 3에 나타낸 바와 같이 원자료의 에너지를 100%로 할때 그 중 근사성분이 99.37%의 에너지를 차지하고 있으며, 이는 근사성분이 원자료에 대한 99.37%의 설명력을 갖는다는 것을 의미한다. 반면에 상세성분에서의 에너지 비율 총합은 1% 미만을 보임에 따라 어떠한 유의적인 해석은 불가능하며 TOC 원자료에 포함되어 있는 측정오차와 계통오차 및 변량간의 상호간섭에 따른 동역학적 잡음성분으로 판단된다(Kaplan et al., 1997).

pH 자료에 대하여 웨이블렛 변환 함수('db9') 적용한 결과 TOC 적용결과와 유사하게 45일 이상의 주기성을 갖는 근사성분에서의 전체에너지는 99.84% 이상을 설명하고 있는 것으로 나타났다. 또한 상세성분의 에너지 비율 역시 TOC 적용결과와 유사한 에너지 비율을 보이고 있다.

순천지방 강우관측소의 강우 자료에 대한 웨이블렛 변환을 적용한 결과 약 3일~11일의 주기성을 갖는 1단계 및 2·3단계의 상세성분과 5단계의 근사성분이 약 89%를 설명할 수 있는 것으로 나타났다.

일반적으로 상세성분 1·2단계는 원자료의 측정 및 기록시 단순 잡음 및 자연현상에서 발생할 수 있는 동역학적 잡음성분으로 볼 수 있으나 1·2단계의 상세성분은 여름철의 집중호우에 의해 기록된 큰 강우량 값에 기인하여 비교

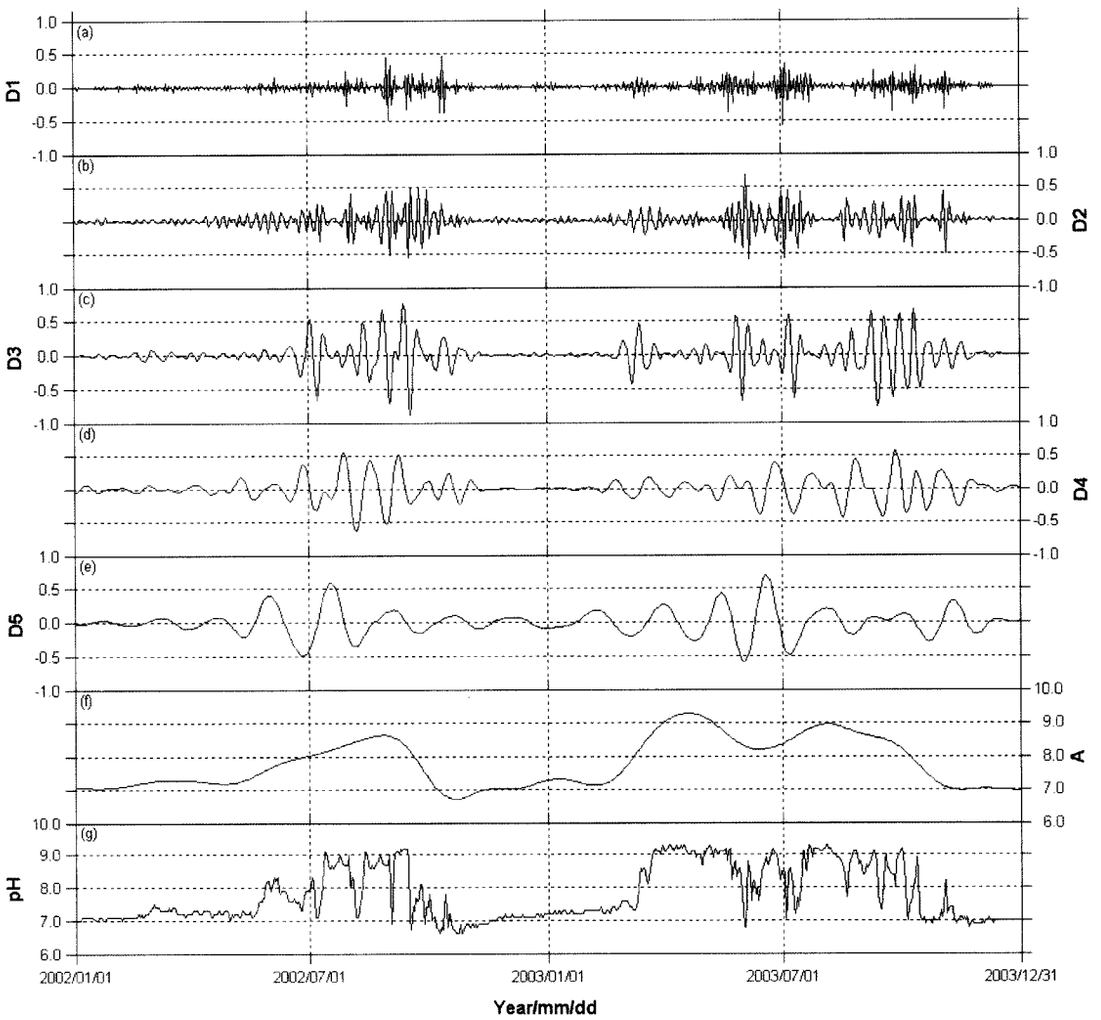

Fig. 7. Discrete wavelet transform and its approximation and detail components about daily pH data.

Table 3. Each step scale, frequency, energy values result of discrete wavelet transform about daily pH data

	Level (j)	Scale ($a=2^j$)	Frequency (1/Days)	Period (Days)	Energy(%) of details	Energy(%) of approximation
pH	1	2	0.3529	2.8333	0.01	
	2	4	0.1765	5.6667	0.02	
	3	8	0.0882	11.3333	0.04	
	4	16	0.0441	22.6667	0.04	
	5	32	0.0221	45.3333	0.05	**99.84**

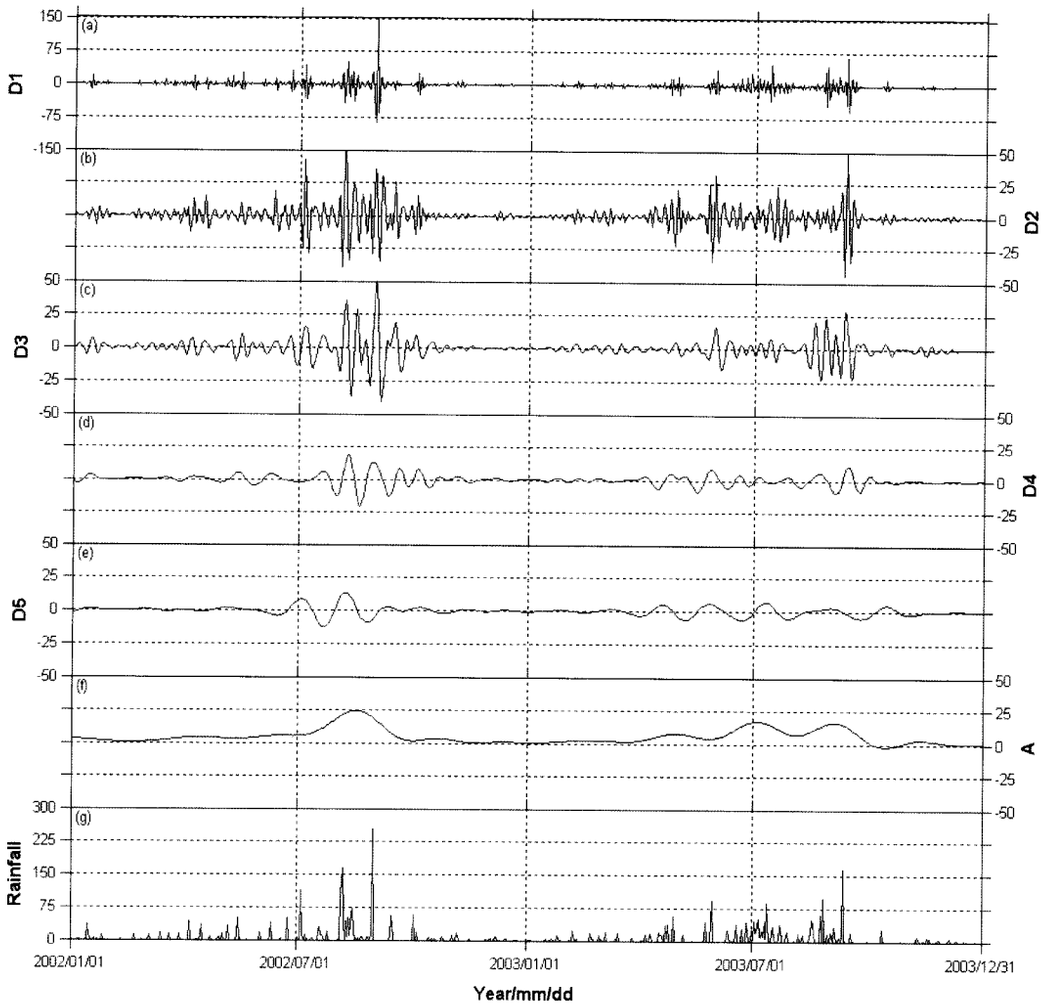

Fig. 8. Discrete wavelet transform and its approximation and detail components about daily rainfall data.

Table 4. Each step scale, frequency, energy values result of discrete wavelet transform about daily rainfall data

	Level (j)	Scale($a=2^j$)	Frequency (1/Days)	Period (Days)	Energy(%) of details	Energy(%) of approximation
Rainfall	1	2	0.3529	2.8333	**36.10**	
	2	4	0.1765	5.6667	**18.87**	
	3	8	0.0882	11.3333	**16.03**	
	4	16	0.0441	22.6667	5.72	
	5	32	0.0221	45.3333	5.27	**18.01**

Table 5. Correlation analysis result about rainfall and water quality data

		Raw data					Approximation component		
		Rainfall	TOC	pH			Rainfall	TOC	pH
Raw data	Rainfall	1.000			Approximation component	Rainfall	1.000		
	TOC	0.120	1.000			TOC	0.627	1.000	
	pH	0.103	0.469	1.000		pH	0.645	0.671	1.000

적 높은 에너지 비율을 나타내고 있다. 또한 Fig. 8(a)에 나타난 바와 같이 상세성분 1단계에서 여름철의 집중호우에 의하여 100 mm의 가장 큰 거동을 보이고 있다.

본 연구의 대상자료인 수질 및 기상자료와 웨이블렛 변환 후 근사성분 간의 상관분석 결과는 Table 3과 같다. Table 3에 나타낸 바와 같이 원자료 간의 상관분석에서는 통계학적으로 유의한 상관계수 값을 얻을 수 없었으나 웨이블렛 변환 후 근사성분간의 상관계수는 원자료 간의 상관계수에 비하여 0.62 이상의 비교적 높은 상관성을 확인할 수 있었다.

웨이블렛 변환 후 근사성분에 대한 강우자료 및 TOC 자료에 대한 상관분석 결과 0.627의 통계적으로 유의한 상관계수 값을 얻을 수 있었으며 이는 강우시 수계내로 유입하는 비점오염물질 유출로 인하여 TOC 농도가 높아진 것으

로 판단된다. 또한 강우자료 및 pH 자료의 근사성분에 대한 상관분석 결과 역시 원자료간의 상관계수가 0.103에 비하여 높은 상관성을 나타내었다. TOC와 pH 자료에 대한 상관분석 결과 역시 본 연구의 대상 자료 중 가장 큰 0.671의 상관계수 값을 얻을 수 있었다.

이러한 상관분석 결과를 이용하여 향후 강우량에 따른 TOC, pH 수질농도 예측에 관한 심화연구가 필요할 것으로 판단된다.

5. 결 론

본 연구에서는 섬진강 유역의 주암호 수질자동측정망 지점에서 측정되고 있는 수질자료 중 일 평균 TOC 및 pH 자료와 주암호 인근의 순천 강우관측소의 일강우량 자료에 대하여 대상 자료가 포함하고 있는 장·단기 특성 파악과 웨이블렛 변환 후 근사성분간의 상관분석을 하였다. 이와 같은 분석을 위하여 TOC 및 강우량 자료에 대하여는 웨이블렛 변환함수 Daubechies의 10번 함수('db10'), pH 자료는 웨이블렛 변환 함수 Daubechies의 9번 함수('db9')를 적용하였다.

TOC 및 pH에 대한 웨이블렛 변환을 적용한 결과 45일 이상의 주기성을 갖는 근사성분에서 전체에너지의 99% 이상을 설명하고 있는 것으로 나타났다. 순천지방 강우관측소의 일강우량 자료에 대한 웨이블렛 변환을 적용한 결과 약 3~11일의 주기성을 갖는 상세성분 1단계 및 2·3단계와 근사성분에 약 89%를 설명할 수 있는 것으로 나타났다. 일반적으로 상세성분 1·2단계의 큰 에너지비는 무강우 및 여름철의 집중호우에 의해 기록된 큰 강우량 값의 영향으로 판단된다.

웨이블렛 변환 후 강우자료와 TOC, 강우자료와 pH 그리고 TOC와 pH 자료의 근사성분 간의 상관분석 결과 0.627, 0.645와 0.671의 통계적으로 유의한 상관계수 값을 얻을 수 있었다. 강우자료 및 TOC와의 관계에서는 강우시 수계 내로 비점오염물질 유입의 영향으로 판단된다.

본 연구의 결과를 담보로 수질자료에 대한 예측모형의 개발시 원자료가 가지고 있는 잡음제거 및 전처리 과정으로서의 웨이블렛 변환을 적용한 수질농도 예측모형의 개발이 가능할 것으로 기대되며 다양한 수질자료에 대한 웨이블렛 변환의 적용과 수질예측 모형 개발에 대한 심화 연구가 필요할 것으로 판단된다.

사 사

이 논문은 2005년도 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구입니다(KRF-2005-037-D00023).

참고문헌

강현배, 김대경, 서진근, *웨이블릿 이론과 응용*, 아카넷, pp. 1-8 (2001).

권현한, 문영일, Wavelet Transform을 이용한 수문시계열 분석, *한국수자원학회논문집*, **38**(6), pp. 439-448 (2005).

박진수, 김신도, 웨이브릿 기법을 이용한 오존 특성분석 및 AR모델을 이용한 예측, *대한환경공학회지*, **26**(8), pp. 844-852 (2004).

안상진, 연인성, 실시간 자동측정망 자료를 이용한 수질관리, *대한토목학회 논문집*, **24**(3B), pp. 221-228 (2004).

조용준, 김종문, Wavelet Transform을 이용한 물수요량 특성분석 및 다원 ARMA 모형을 통한 물수요량 예측, *한국수자원학회 논문집*, **31**(3), pp. 317-326 (1998).

Daubechies, I., *Ten lectures on wavelets*, SIAM, **61**, pp. 258-261 (1994).

Drago, A. F. and Boxall, S. R., Use of the Wavelet Transform on Hydro-meteorological Data, *Physics and Chemistry of the Earth*, **27**, pp. 1387-1399 (2002).

Kaplan, D. and Glass, L., Understanding Nonlinear Dynamics, *Springer*, pp. 280-286 (1997).