

웨이블릿 이론을 이용한 주암호 TOC 자료의 분석 및 예측

Analysis and Prediction for TOC Data in the Juam-lake Using Wavelet Theory

오 창 열*, 진 영 훈**, 곽 필 정***, 박 성 천****

Chang Ryol Oh, Young Hoon Jin, Pil Jeong Gwak, Sung Chun Park

요 지

본 연구에서는 수질자료에 내재되어 있는 주기성 및 경향성 등을 파악하기 위해 웨이블릿 변환을 적용하였으며 비선형 시계열자료에 대한 예측력이 우수한 인공신경망을 적용하여 예측모형을 개발하였다. 대상자료는 섬진강 유역의 주암호 수질자동측정망 지점에서 측정되고 있는 수질자료 중 2002년 1월 1일 ~ 2004년 12월 31일까지의 일 TOC 수질자료를 이용하였다. 웨이블릿 변환을 위해 사용한 기저함수로는 Daubechies의 10번 웨이블릿 함수('db10')를 사용하였으며, 각 스케일링 및 웨이블릿 함수를 이용하여 5단계까지 변환하였다. 최종 변환된 근사성분과 D5, D4, D3, D2의 상세성분 자료를 이용하여 1시간후 TOC 예측 모형을 구성하였으며 그 결과 은닉층의 노드의 수가 17개인 모형인 Model_5_17 모형이 가장 우수한 예측력을 보였다.

핵심용어 : 웨이블릿, 인공신경망, TOC

1. 서 론

우리나라는 1960년대 이후 산업화에 따른 인구증가, 도시화 및 공단의 집단화 등으로 인하여 오염원의 양적증가와 지역적 집중화가 이루어지고 있다. 이러한 문제를 해결하기 위하여 정부에서는 물관리종합대책을 수립하여 시행중에 있으며 수질의 상시감시 기능과 조기경보체계의 구축을 위해 전국 4대강 유역 주요 20개 지점에서 수질자동 측정망을 설치 운영 중에 있다. 이러한 수질자동측정망의 설치 목적을 달성하기 위해서는 수질자료에 대한 특성분석 및 예측에 대한 심화연구가 필요하다.

따라서 본 연구에서는 수질자료에 내재되어 있는 주기성 및 경향성 등을 파악하기 위해 웨이블릿 변환을 적용하였다. 또한 최종 변환된 상세성분 및 근사성분을 이용하여 비선형 시계열자료에 대한 예측력이 우수한 인공신경망을 적용하여 예측모형을 개발하였다.

2. 이론적 배경

2.1 Wavelet 이론

웨이블릿 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 a 는 스케일링을 결정하는 값이고, b 는 함수를 얼마나 이동시킬 것인가를 결정하는 값이다.

* 정회원·동신대학교 대학원 토목공학과 박사과정·E-Mail : new2020@dsu.ac.kr

** 정회원·동신대학교 토목공학과 연구원 ·E-Mail : yhj@dsu.ac.kr

*** 정회원·동신대학교 대학원 토목공학과 박사수료·E-Mail : pjgwak@dsu.ac.kr

**** 정회원·동신대학교 토목공학과 부교수 ·E-Mail : psc@dsu.ac.kr

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad a, b \in R \quad \text{-----} \quad (1)$$

웨이블렛 변환은 식(1)에서 모함수를 b 만큼 이동하고 a 에 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.

$$f(t) = \sum_{k=-\infty}^{\infty} a_{J,k} \Phi_{J,k}(t) + \sum_{j=1}^J \sum_{k=-\infty}^{\infty} d_{j,k} \psi_{j,k}(t) \quad \text{-----} \quad (2)$$

2.2 인공신경망 이론

Haykin (1999)은 신경망이란 "자연적으로 경험적인 지식을 저장하고 유용하게 만드는 경향을 지닌 간단한 처리 단위로 구성된 거대하며 병렬적으로 분포된 연산 처리 장치"라고 정의하였다. 이러한 신경망의 여러 장점들 중에서 본 연구에 부합될 수 있는 면들은 다음과 같다.

첫 번째로 신경망은 선형 또는 비선형적 일 수 있으나 뉴런들간의 상호 연결에 의한 신경망은 비선형적이며, 이 비선형성은 신경망 전체에 걸쳐 분포되어진다는 것이다. 두 번째로 신경망은 입력과 출력간의 사상에 학습을 한다는 것이며, 세 번째로 신경망은 주위 환경에 적응되어지도록 연결강도를 조절할 수 있는 능력을 가지고 있다는 것이다.

본 연구에 사용된 인공 신경망의 종류는 다층(multi-layer) 신경망이며, 입력층과 은닉층, 그리고 출력층으로 구성하였다. 또한 매개변수 최적화를 위하여 사용된 학습 알고리즘은 모멘텀 상수와 적응식 학습율이 적용된 최급 강하법을 이용한 오차 역전파 학습 알고리즘이며, 이에 포함된 활성화 함수는 은닉층에서는 탄젠트 시그모이드 함수와 출력층에서 선형 함수를 사용하였다.

3. 대상지점 및 자료

본 연구의 대상지점인 주암호는 유역면적 1,010km², 저수량 4억 5700만t으로 보성군·순천시·화순군의 3개 시군에 걸쳐 있으며, 광주광역시·나주시·목포시·화순군 등 전라남도 서부권에 하루 80만t의 생활용수를 공급하고 있다.

주암호1 수질관측망은 전라남도 순천시 대광면 주암리에 위치해 있으며, 본 연구에서는 2002년 1월 1일 ~ 2004년 12월 31일까지 총 1,095개의 일평균 TOC 수질농도 자료를 이용하였다.

TOC 수질자료에 대한 기술통계 분석 결과 평균농도는 1.42 mg/L로 비교적 양호한 수질농도로 나타내었으며 0.98~2.27 mg/L의 분포를 보였다.



그림 1. 대상지점의 유역도

4. 모의결과 및 고찰

4.1 웨이블렛 변환 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 분해 가능한 함수 그리고, 원자료와 최종분해단계 근사성분과의 상관계수가 가장 높은 기

준으로 최종 웨이블릿 함수를 선택하였으며 그 결과 ‘db10’함수를 적용하였다.

원자료에서의 계절성 및 연주기성이 파악이 불명확한 TOC 자료에 대하여 웨이블릿 변환 함수(db10)를 적용한 결과 45일의 강한 주기성을 갖는 것으로 나타났으며 단계별 주기성 및 상세:근사성분의 에너지 값을 표 1에 제시하였다. 또한 TOC의 원자료 및 각 단계별 웨이블릿 변환 결과는 그림 2에 나타내었다.

표 1에 나타난 바와 같이 전체에너지 중 근사성분에서는 99.37%의 에너지 값에 따라 근사성분이 원자료에 대하여 99.37% 설명해 주고 있으며 나머지 상세성분에서 0.63%의 에너지 값을 나타내었다.

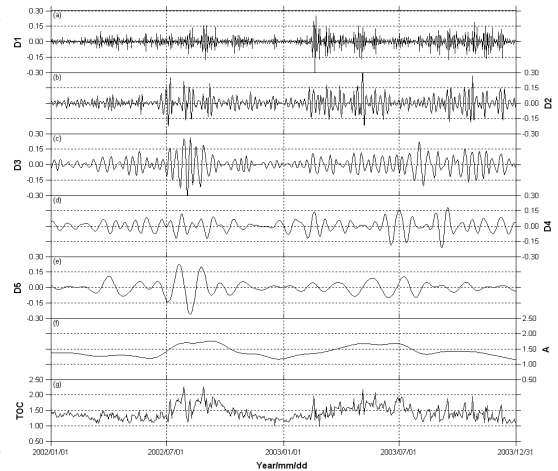


그림 2. TOC 자료에 대한 웨이블릿 변환 결과

4.2 TOC 수질농도 예측모형의 개발

본 연구에서는 주암호1지점에 대한 1일 후 TOC 수질농도를 예측하기 위하여 웨이블릿 함수 ‘db10’에 의하여 각 단계별로 최종 변환된 자료를 이용하였으며 다음의 식(4)~(8)까지 인공신경망의 예측모형을 구성하였다. 최적의 인공신경망 모형을 탐색하기 위하여 인공신경망 모형의 매개변수인 모멘텀 상수와 초기 학습률은 모든 모형에서 각각 0.1과 0.7를 일괄적으로 사용하여 모형을 선별하였으며, 그 결과 표 2와 같이 Model_5가 가장 우수한 예측력을 보였다.

표 1. TOC자료에 대한 웨이블릿 변환 후 각 단계별 주기 및 에너지 값

	Level (j)	Scale(a=2 ^j)	Frequency (1/Days)	Period (Days)	Energy(%) of details	Energy(%) of approximation
TOC	1	2	0.3529	2.8333	0.08	
	2	4	0.1765	5.6667	0.12	
	3	8	0.0882	11.3333	0.16	
	4	16	0.0441	22.6667	0.12	
	5	32	0.0221	45.3333	0.15	99.37

$$TOC_{M1}(t) = f \left[App_5(t-1, 2) \right] \dots\dots\dots \text{식 (4)}$$

$$TOC_{M2}(t) = f \left[App_5(t-1, 2), Det_5(t-1, 2) \right] \dots\dots\dots \text{식 (4)}$$

$$TOC_{M3}(t) = f \left[App_5(t-1, 2), Det_5(t-1, 2), Det_4(t-1, 2) \right] \dots\dots\dots \text{식 (5)}$$

$$TOC_{M4}(t) = f \left[App_5(t-1, 2), Det_5(t-1, 2), Det_4(t-1, 2), Det_3(t-1, 2) \right] \dots\dots\dots \text{식 (6)}$$

$$TOC_{M5}(t) = f \left[App_5(t-1, 2), Det_5(t-1, 2), Det_4(t-1, 2), Det_3(t-1, 2), Det_2(t-1, 2) \right] \dots\dots\dots \text{식 (7)}$$

$$TOC_{M6}(t) = f \left[App_5(t-1, 2), Det_5(t-1, 2), Det_4(t-1, 2), Det_3(t-1, 2), Det_2(t-1, 2), Det_1(t-1, 2) \right] \dots\dots\dots \text{식 (8)}$$

여기서, App₅는 웨이블릿 함수에 의하여 변환된 5단계 근사성분이며, Det₅₋₁은 5단계~1단계까지의 상세성분이다.

표 2. 모형의 구성에 따른 통계적 특성치

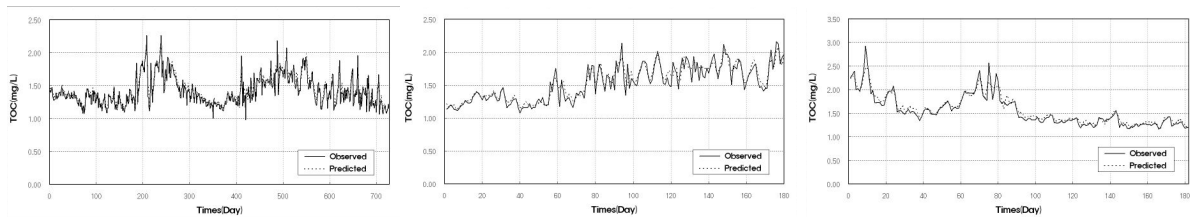
Model	Correlation Coefficient			RMSE			Model	Correlation Coefficient			RMSE		
	훈련	검증 I	검증 II	훈련	검증 I	검증 II		훈련	검증 I	검증 II	훈련	검증 I	검증 II
Model_1	0.760	0.750	0.785	0.149	0.223	0.176	Model_4	0.870	0.915	0.883	0.112	0.146	0.128
Model_2	0.817	0.899	0.832	0.134	0.163	0.159	Model_5	0.932	0.950	0.932	0.083	0.117	0.107
Model_3	0.854	0.909	0.865	0.115	0.148	0.139	Model_6	0.784	0.831	0.810	0.227	0.397	0.330

표 3은 Model_5에 대하여 은닉층 노드의 수에 따른 TOC 수질농도 예측 결과를 상관계수 및 RMSE의 통계적 특성치를 나타낸 표로서 최적 TOC 수질농도 예측모형은 은닉층의 노드의 수가 17개인 모형으로 나타났다. Model_5_17 모형은 훈련 및 검증 I, 검증 II 과정에서의 상관계수는 0.953, 0.968, 0.943으로 가장 우수한 예측력을 보였으며 RMSE 값 역시 0.067, 0.086, 0.090으로 가장 작은 오차값을 보임으로써 Model_5_17 모형이 TOC 수질농도에 예측력의 우수한 검증된 것으로 판단된다.

그림 3은 Model_5_17 모형의 훈련 및 검증 I, 검증 II 과정에 대한 관측값과 모의값을 도시한 그림으로써 TOC 농도의 최고값을 갖는 2004년 7월 11일의 2.93mg/L에 대한 예측농도는 2.45mg/L으로 관측값에 저평가된 것으로 나타났다. 이는 훈련과정에서의 최대값은 2.26mg/L이나 검증 II 과정에서의 최대값은 2.93mg/L임에 따라 최대 TOC 수질농도를 갖는 입력패턴에 대한 적절한 훈련이 이루어지지 않음에 기인한 것으로 판단된다.

표 3. 은닉층의 수에 따른 통계적 특성치

구분	Correlation Coefficient			RMSE			구분	Correlation Coefficient			RMSE		
	훈련	검증 I	검증 II	훈련	검증 I	검증 II		훈련	검증 I	검증 II	훈련	검증 I	검증 II
Model_5_6	0.945	0.960	0.940	0.074	0.098	0.097	Model_5_14	0.856	0.859	0.874	0.118	0.202	0.147
Model_5_7	0.950	0.969	0.939	0.073	0.089	0.095	Model_5_15	0.900	0.933	0.895	0.098	0.139	0.128
Model_5_8	0.945	0.958	0.942	0.076	0.101	0.098	Model_5_16	0.881	0.905	0.880	0.108	0.156	0.129
Model_5_9	0.913	0.938	0.921	0.095	0.122	0.108	Model_5_17	0.953	0.968	0.943	0.067	0.086	0.090
Model_5_10	0.932	0.950	0.932	0.083	0.117	0.107	Model_5_18	0.896	0.923	0.905	0.097	0.137	0.115
Model_5_11	0.926	0.932	0.930	0.089	0.139	0.110	Model_5_19	0.935	0.956	0.929	0.078	0.107	0.101
Model_5_12	0.844	0.839	0.858	0.118	0.194	0.140	Model_5_20	0.920	0.936	0.927	0.086	0.124	0.102
Model_5_13	0.938	0.959	0.933	0.079	0.105	0.104							



(a) 훈련과정의 관측값과 모의값 (b) 검증 I 과정의 관측값과 모의값 (c) 검증 II 과정의 관측값과 모의값

그림 3. Model_5_17 모형의 훈련 및 검증 I, 검증 II 과정에 대한 관측값과 모의값

5. 결론

본 연구에서는 수질자료에 내재되어 있는 주기성 및 경향성 등을 파악하기 위해 웨이블릿 변환을 적용하였다. 또한 최종 변환된 상세성분 및 근사성분을 이용하여 비선형 시계열자료에 대한 예측력이 우수한 인공신경망을 적용하여 예측모형을 개발하였다.

최적의 TOC 수질농도 예측 모형은 웨이블릿 변환에 의한 5단계 근사성분과 2·3·4·5단계의 상세성분을 입력자료를 갖으며 은닉층의 노드의 수가 17개인 모형인 Model_5_17 모형으로 나타났다. 본 모형은 훈련 및 검증 I, 검증II 과정에서의 상관계수는 0.953, 0.968, 0.943으로 가장 우수한 예측력을 보였으며 RMSE 값 역시 0.067, 0.086, 0.090으로 가장 작은 오차값을 보임으로써 Model_5_17 모형이 TOC 수질농도에 예측력의 우수성이 검증된 것으로 판단된다.

참고문헌

1. 안상진, 연인성(2004), 실시간 자동측정망 자료를 이용한 수질관리, 대한토목학회 논문집, 제24권 제3B호, pp. 221 ~ 228.
2. Daubechies, I.(1994), Ten lectures on wavelets, SIAM, **61**, pp. 258 ~ 261.