

유출량, 수질 농도 및 부하량의 상호관계를 이용한 자료보간 방법에 관한 연구

Study on the Method for Data Interpolation using the Correlation
among Runoff, Water Quality Concentration and Load

오창열*, 정우철**, 진영훈***, 박성천****

Chang Ryeol Oh, Woo Cheol Jung, Young Hoon Jin, Sung Chun Park

요 지

수문 및 수질자료는 일정한 기준에 의한 관측치를 시계열 자료로 기록하거나 전송할 때 다양한 형태의 오차가 발생하게 되며 또한 수문 및 수질자료를 관측하는 측정기기의 고장과 유지관리 등의 어려움으로 다양한 형태의 결측 자료가 발생하고 있다. 이와 더불어 수문 및 수질자료는 시공간적 변동성이 크며 비선형성이 강한 특성을 갖고 있다. 이러한 수문 및 수질 자료를 이용하여 모형을 구축할 경우 다양한 형태의 잡음에 대한 겹침 및 잡음저감이 필수적 요건이라 할 수 있다.

따라서 본 연구에서는 영산강 유역의 본류부를 대표하는 나주지점에 대한 유출량과 총유기탄소(TOC) 농도 및 TOC 부하량 예측모형을 개발하였으며. 이를 위한 방법으로는 잡음저감을 위하여 웨이블렛 변환과 인공신경망을 적용하였다.

TOC 부하량 자료는 유출량과 TOC 자료간의 함수로서 표현이 가능함에 따라 유출량 및 TOC 자료가 결측되었을 경우 역함수에 의한 계산으로 결측 자료에 대한 보간이 가능하다. 따라서 본 연구의 주안점을 잡음 저감 및 인공신경망에 의해 최적화된 예측 모형이 결측된 유출량과 TOC 자료에 대한 역함수로 정도있는 유출량과 TOC 자료 생성 가능성을 검토하고자 한다.

본 연구의 적용 결과, 유출량 자료가 결측되었을 경우 TOC 및 TOC 부하량 예측으로 유출량 자료에 대한 간접추정 및 결측 자료에 대한 보간의 정도를 평가한 결과 R^2 는 0.99 이상의 값을 보였다. 또한, TOC 자료가 결측되었을 경우 역시 R^2 는 비교적 우수한 0.97 이상의 값을 보였다. 따라서 본 연구에서 개발한 유출량 및 TOC, TOC 부하량 예측모형의 개발은 정도있는 유출량 및 TOC 수질 자료의 생성이 가능할 것으로 기대된다.

핵심용어 : 유출량, 총유기탄소(TOC), 부하량, 잡음저감, 인공신경망, 자료보간

1. 서 론

자연현상에서 발생하는 실제 신호는 적절한 형식으로 기록하거나 전송할 때 다양한 형태의 오차가 발생하게 되는데 다음과 같이 3가지 형태(측정오차, 계통오차, 동역학적 잡음)로 분류할 수 있다(Kaplan, 1997). 이와 더불어 수문 및 환경자료 측정기기의 고장과 유지관리 등의 어려움으로 결측 자료가 발생하고 있으며 이러한 오차들로 인하여 원시계열의 특성이 왜곡되는 현상으로 귀결된다.

유출량 및 수질 자료에 포함된 잡음을 저감하고 이에 따른 본질적인 거동 특성 분석과 예측모형의 개발을 통하여 정도있는 유출량 및 수질자료의 생성은 수자원의 효율적인 관리 측면에서 볼 때 중요한 연구과제이다.

이러한 오차를 저감하고 예측모형(유출량과 TOC 농도, TOC 부하량 예측모형)을 개발하기 위하여 시간과

* 정회원-유량조사 사업단 E-mail : croh@kict.re.kr
** 정회원-동신대학교 토목공학과 석사과정 E-mail : j01194337086@nate.com
*** 정회원-호남대학교 산업기술연구소 연구원 E-mail : nmdrjin@gmail.com
**** 정회원-동신대학교 토목공학과 교수 E-mail : psc@dsu.ac.kr

주파수 영역에서의 유연한 분해능에 의해 국소화가 가능하며 잡음저감에 있어서 효율적인 수학적 도구로 알려진 웨이블렛 변환(강현배 등, 2001; 권현한 등, 2005; 조용준 등 1998)과 비선형동역학 이론(김형수 1996a, 김형수 1996b), 인공신경망 이론을 적용하였으며 이에 따른 본 연구의 주요 내용 및 방법에 관한 흐름도는 Fig. 1과 같다.

따라서 본 연구에서는 유출량, TOC, TOC 부하량의 1시간 후 예측 모형을 개발을 통하여 측정기기의 고장이나 기타 원인에 의하여 유출량 및 TOC 자료가 결측되었을 경우 간접추정과 결측자료 보간의 정도를 평가하고자 하였으며 다만 본 고에서는 지면관계상 결측자료의 간접추정과 보간에 대한 평가 결과를 중심으로 기술하였다.

2. 대상지점 및 자료의 특성

2.1 대상지점

수자원에 대한 올바른 치수 및 이수 계획 수립을 위해서는 대상 유역에 대하여 홍수 예경보와 수질의 조기경보체계 구축을 위해서는 수문 및 환경부문의 정도있는 유량 및 수질자료가 담보되어야 한다.

따라서 본 연구에서는 이러한 목적을 달성하기 위하여 영산강의 본류부를 대표하는 나주지점을 대상으로 2003년 1월 1일 00시부터 2004년 12월 31일 23시까지의 유출량자료 및 TOC, TOC 부하량 자료를 이용하였다. 유출량 자료는 해당 년의 수위-유량관계 곡선식을 이용하여 산출하였으며 TOC 부하량 자료는 유출량 및 TOC 농도의 곱으로써 산출하였다.

본 연구의 대상인 나주 지점은 전라남도 나주시 남산동의 나주대교 (동경 $126^{\circ} 44' 03''$, 북위 $35^{\circ} 01' 52''$) 상류 우안에 위치한다. 유역 면적은 $2,058.72\text{km}^2$ 로 전체 영산강 유역의 약 60%에 해당되며 유로연장이 65.5 km 인 영산강 본류의 본류부를 대표하는 지점이다. 나주지점 상·하류의 하천 형태는 복단면의 형상이고, 하신은 만곡에 의하여 저수로부가 좌안 측으로 편기되어 있다(건설교통부, 1998; 2004)

2.2 자료의 특성

본 연구의 대상 자료는 유출량과 TOC 그리고 TOC 부하량 자료이다. 유출량은 대상 지점의 유역 내에서 발생하는 강우에 의하여 하천의 한 횡단면을 통과하는 단위시간에 따른 물의 양으로서 나주수위관측소의 2003년 ~ 2004년 시수위 자료에 대하여 수위-유량관계 곡선식을 이용하여 유출량으로 환산하였다.

TOC는 시료 중에 용존되어 있는 유기물을 촉매산화 또는 UV산화 등의 산화과정을 통하여 CO_2 가스를 발생시키고, 그 발생량을 CO_2 검출기를 이용하여 전체 유기물의 양을 나타내는 수질지표이다. 이는 기존의 BOD 및 COD 수질농도 측정에 비하여 적은 오차요인과 빠른 시간에 결과 값을 얻을 수 있으며 난분해성 물질에 대한 해석이 가능하고 재현성이 탁월하다는 장점이 있어 수질지표로 사용되는 데 적합하다 하겠다(정명규 등, 2001; 류동경 등, 2006).

TOC 부하량은 식 (1)과 같이 TOC 농도와 유출량의 곱과 단위 환산계수 1,000으로 나누어 산출한 자료로서 단위시간(sec)에서의 부하량을 의미한다. 일반적으로 부하량의 단위는 하루에 발생하는 오염물질의 총량으로서

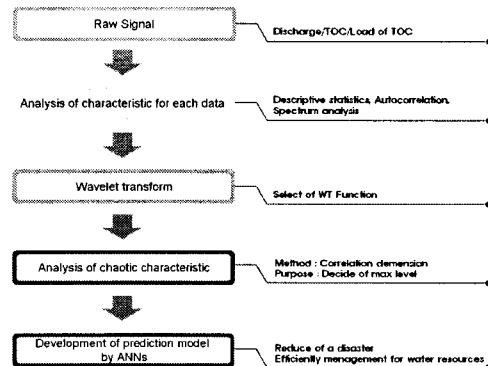


Fig. 1 Flow chart in this study

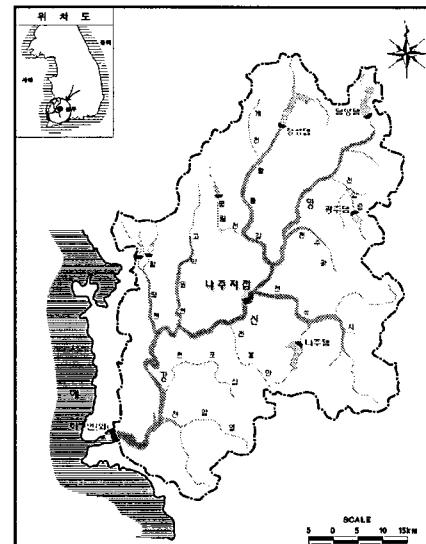


Fig. 2 Geographical location of the Naju station and the Young-San river basin

의 Kg/day를 이용한다. 그러나 본 연구에서는 부하량 예측모형의 개발과 동시에 유출량 및 TOC 결측 자료의 보간을 위하여 단위시간인 초(sec) 동안의 부하량 자료를 이용하였다.

$$TOC\ Load(Kg/sec) = \frac{Discharge(m^3/sec) \times TOC\ Concentration(mg/L)}{1,000} \quad - (1)$$

TOC 부하량은 식 (1)과 같이 TOC 농도와 유출량의 곱과 단위 환산계수 1,000으로 나누어 산출한 자료로서 단위시간(sec)에서의 부하량을 의미한다. 일반적으로 부하량의 단위는 하루에 발생하는 오염물질의 총량으로서의 Kg/day를 이용한다. 그러나 본 연구에서는 부하량 예측모형의 개발과 동시에 유출량 및 TOC 결측 자료의 보간을 위하여 단위시간인 초(sec) 동안의 부하량 자료를 이용하였다.

본 연구에서 이용한 유출량, TOC 농도, TOC 부하량 자료에 대한 기술통계 분석 결과는 Table 1과 같다.

Table 1. Descriptive statistics for Discharge, TOC, TOC Load data of Naju

	Max	Min	Mean	Standard deviation	Skewness	Kurtosis
Discharge	7,738.005	8.182	92.640	304.741	12.230	229.583
TOC	7.891	1.108	3.570	0.974	0.795	0.657
TOC Load	30.295	0.019	0.306	1.111	14.323	296.240

3. 결측 자료에 따른 간접추정 및 보간에 대한 평가

본 연구에서는 원자료에 포함된 잡음저감을 위해 웨이블렛 변환(곽필정 2006; 진영훈, 2006a; 진영훈, 2006b; Daubechies, 1992)을 적용하였으며 웨이블렛 변환시 최적파형 분해단계 결정의 도구로서 비선형 동역학적 이론을 적용하였다. 이러한 일련의 과정을 통하여 최종 분해된 최종파형 분해단계의 균사성분과 각 단계별 상세성분을 인공신경망 모형의 입력 자료로 활용하여 예측모형을 개발하였다.

본 연구에서 개발한 최적의 유출량 예측모형은 Model VI_23(Dis.) 모형으로서 훈련 및 검증과정에서의 R^2 에 대한 통계적 결과치는 각각 0.994, 0.993, RMSE는 각각 29.618, 20.454, PEP는 각각 -5.215, -0.647로 나타났으며 일반적인 신경망 모형에 나타나는 Persistence 현상을 발견할 수 없었다.

최적의 TOC 예측모형은 Model VII_29(TOC) 모형으로 훈련 및 검증과정에서의 R^2 에 대한 통계적 결과치는 각각 0.998, 0.998, RMSE는 각각 0.065, 0.056, PEP는 각각 -5.130, 0.800로 나타났으며 본 모형 역시 일반적인 신경망 모형에 나타나는 Persistence 현상을 발견할 수 없었다.

최적의 TOC 부하량 예측모형은 Model VI_17(T.L.) 모형으로 R^2 에 대한 통계적 결과치는 각각 0.989, 0.993, RMSE는 각각 0.212, 0.088, PEP는 각각 -7.608, -5.839로 나타났으나 본 모형에서는 시계반대 방향으로 전이되는 Persistence 현상이 발견되었다. 이는 대상 자료의 최대값과 최소값의 비율이 1,500배 해당하는 큰 범위를 갖는 것에 기인한 것으로 판단된다.

이상의 연구 결과는 1시간 후에 대한 예측모형으로서 최종 선정된 예측모형의 적용을 위해서는 1시간 전, 그리고 현재의 유출량과 TOC 자료가 필요하다. 그러나 측정기기의 고장이나 기타 원인에 의하여 유출량 및 TOC 자료가 결측되었을 경우 다음과 같은 방법에 의해 간접추정과 결측 자료의 보간이 가능하다.

1시간 전 또는 현재 시간의 유출량 자료가 결측된 경우는 1시간 후 TOC 및 TOC 부하량 예측으로 식 (2)에 의하여 간접추정이 가능하다. 또한 현재시간의 유출량 자료가 결측되었을 경우는 현재 시간의 TOC 부하량 예측을 통하여 식 (3)에 의하여 결측 자료에 대한 보간이 가능할 것으로 판단된다. TOC 자료가 결측된 경우 역시 식 (3)에 의하여 간접추정 및 결측 자료가 가능하다.

$$\text{Discharge}(\text{m}^3/\text{sec}) = \text{TOC Load}(\text{Kg/sec}) / \text{TOC Concentration}(\text{mg/L}) \times 1000 \quad (2)$$

$$\text{TOC Concentration}(\text{mg/L}) = \text{TOC Load}(\text{Kg/sec}) / \text{Discharge}(\text{m}^3/\text{sec}) \times 1000 \quad (3)$$

간접추정 및 결측 자료 보간에 대한 평가기간은 본 연구에서 검증과정으로 활용한 2003년 자료를 하였다. 유출량 자료가 결측되었을 경우 간접추정 및 결측 자료에 보간의 정도를 평가한 결과 R^2 는 0.990 이상 값을 보임에 따라 정도있는 수문자료의 생성이 가능할 것으로 판단되며 이에 따른 간접 추정에 의한 시계열은 Fig. 3, 결측자료 보간은 Fig. 4에 제시하였다.

TOC 자료가 결측되었을 경우 역시 R^2 는 비교적 우수한 0.970 이상의 값을 보임에 따라 정도있는 수질자료의 생성이 가능할 것으로 판단되며 이에 따른 간접추정에 의한 시계열은 Fig. 5, 결측 자료 보간은 Fig. 6에 제시하였다.

이와 같은 결과로 본 연구에서 개발한 유출량, TOC 및 TOC 부하량 예측 모형의 적용성은 우수한 것으로 판단되며 또한 정도있는 수문 및 수질자료의 구축이 가능할 것이다.

Table 2. Statistical analysis according to missing data

	R^2	
	간접추정	결측자료 보간
유출량	0.994	0.994
TOC	0.975	0.972

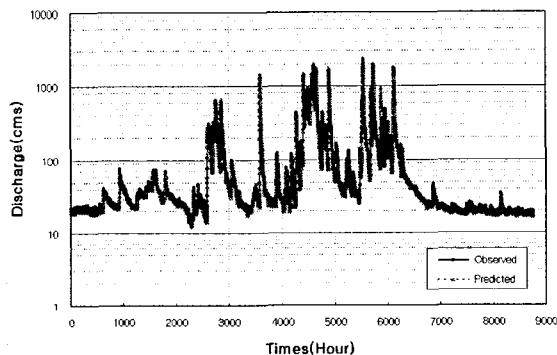


Fig. 3. Indirect estimation for missing data of discharge

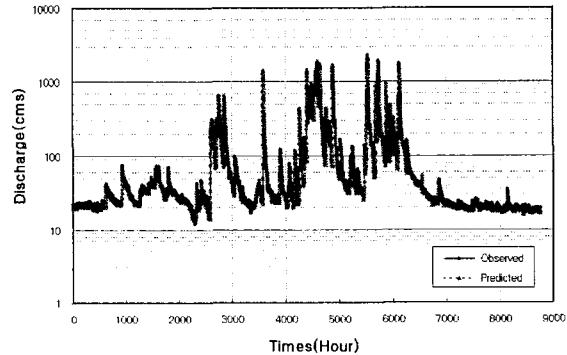


Fig. 4. Interpolation for missing data of discharge

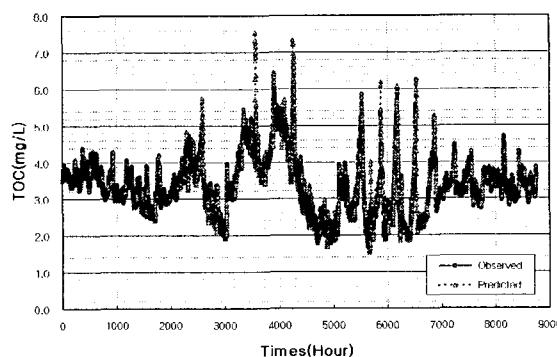


Fig. 5. Indirect estimation for missing data of TOC

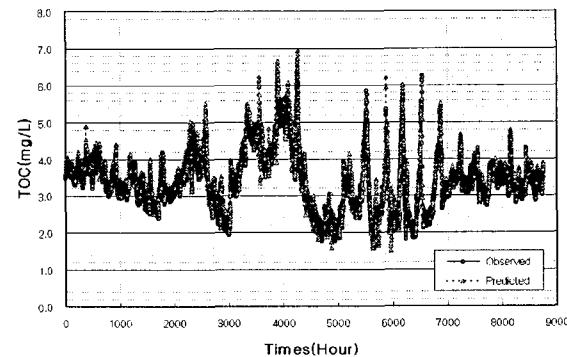


Fig. 6. Interpolation for missing data of TOC

4. 결 론

본 연구에서는 영산강 유역의 본류부를 대표하는 나주지점을 대상으로 인공신경망을 이용하여 유출량 예측 모형과 총유기탄소(Total Organic Carbon: 이하 TOC) 예측모형, TOC 부하량 예측모형을 개발하였으며 측정기

기의 고장이나 기타 원인에 의하여 유출량 및 TOC 자료가 결측되었을 경우 TOC 부하량 예측값을 이용한 유출량 및 TOC 자료의 간접추정과 결측자료 보간의 정도를 평가하였다.

유출량 예측모형을 개발하기 위하여 웨이블렛 변환함수 'db10'에 의해 최종 분해된 근사성분과 상세성분을 인공신경망의 입력 자료로 구축한 결과 최적의 유출량 예측모형은 Model VI_23(Dis.)이다. 본 모형의 훈련 및 검증과정에서의 R^2 에 대한 통계적 결과치는 각각 0.994, 0.993, RMSE는 각각 29.618, 20.454, PEP는 각각 -5.215, -0.647로 나타났으며 일반적인 신경망 모형에 나타나는 Persistence 현상을 발견할 수 없었다.

TOC 예측모형을 개발하기 위하여 웨이블렛 변환함수 'db10'에 의해 최종 분해된 근사성분과 상세성분을 인공신경망의 입력 자료로 구축한 결과 최적의 TOC 예측모형은 Model VI_27(TOC)이다. 본 모형의 훈련 및 검증과정에서의 R^2 에 대한 통계적 결과치는 각각 0.995, 0.996, RMSE는 각각 0.077, 0.058, PEP는 각각 -3.957, -0.162로 나타났으며 본 모형 역시 일반적인 신경망 모형에 나타나는 Persistence 현상을 발견할 수 없었다.

TOC 부하량 예측모형을 개발하기 위하여 웨이블렛 변환함수 'db10'에 의해 최종 분해된 근사성분과 상세성분을 인공신경망의 입력 자료로 구축한 결과 최적의 TOC 예측모형은 Model VI_17(T.L.)이다. 본 모형의 훈련 및 검증과정에서의 R^2 에 대한 통계적 결과치는 각각 0.989, 0.993, RMSE는 각각 0.212, 0.088, PEP는 각각 -7.608, -5.839로 나타났으나 본 모형에서는 시계반대 방향으로 전이되는 Persistence 현상이 발견되었는데 이는 대상 자료의 최대값과 최소값의 비율이 1,500배 해당하는 큰 범위를 갖는 것에 기인한 것으로 판단된다.

최적의 예측모형 개발은 측정기기의 고장이나 기타 원인에 의해 발생되는 결측 자료의 간접추정 및 결측 자료 보간의 정도를 평가한 결과 유출량은 R^2 값이 0.990 이상, TOC는 0.970 이상의 값을 보임에 따라 정도있는 수문 및 수질자료의 생성이 가능할 것으로 기대된다.

참 고 문 헌

1. 강현배·김대경·서진근(2001), “웨이블렛 이론과 응용”, 아카넷, pp. 219 ~ 238.
2. 건설교통부(1998), “영산강 하천정비기본계획(보완)”.
3. 건설교통부 영산강홍수통제소(2004), “영산강 수계 유량측정보고서”, (주)한국종합기술개발공사.
4. 곽필정·오창열·진영훈·박성천(2006), “웨이블렛 변환과 인공신경망을 이용한 일 TOC 자료의 예측에 관한 연구”, 한국물환경학회지, 22(5), pp. 952 ~ 957.
5. 권현한·문영일(2005), “Wavelet Transform을 이용한 수문시계열 분석”, 한국수자원학회논문집, 38(6), pp. 439 ~ 448.
6. 김형수·윤용남(1996a), “카오스의 위상학적 견지:(I) 일유량자료 계열 해석에의 응용”, 대한토목학회 논문집, 16(II-5), pp. 445 ~ 452.
7. 김형수·윤용남(1996b), “카오스의 위상학적 견지:(II) 독립성과 잡음저감”, 대한토목학회 논문집, 16(II-5), pp. 453 ~ 459.
8. 류동경·배상득·장재용·박재철·류재근(2006), “낙동강수계 수질오염총량관리 대상물질에 관한 연구 –TOC 유기물 중심으로–”, 한국물환경학회·대한상하수도학회 공동춘계학술발표회 논문집, pp. 1181 ~ 1188.
9. 정명규·외 7인 (2001), “환경분석화학”, 동화기술, pp. 406 ~ 407.
10. 조용준·김종문(1998), “Wavelet Transform을 이용한 물수요량 특성분석 및 다원 ARMA 모형을 통한 물수요량 예측”, 한국수자원학회 논문집, 31(3), pp. 317 ~ 326.
11. 진영훈·박성천(2006a), “수문기상자료의 웨이블렛 변환에 의한 비선형 동역학적 성분의 추출”, 대한토목학회, 26(5B), pp. 439 ~ 446.
12. 진영훈·오창열·박성천(2006b), “수질 및 강우자료의 상관분석을 위한 웨이블렛 변환의 적용”, 한국물환경학회지, 22(5), pp. 831 ~ 837.
13. Daubechies, I.(1992), "Ten lectures on wavelets", Soc. for Ind. and Applied Math., Philadelphia, PA.
14. Kaplan, D., & Glass, L.(1997), "Understanding Nonlinear Dynamics", Springer, pp. 280 ~ 286.