

# 주기성을 갖는 탁도자료의 결측치 보완 기법

## Filling Method for Missing Turbidity Data having Periodicity

백경오\*, 조흥연\*\*, 이삼희\*\*\*

Kyong Oh Baek, Hong Yeon Cho, Sam Hee Lee

### 요 지

한강 하구부의 3개 지점에서 수중 계류방식으로 약 5개월에 걸쳐 탁도를 관측하였다. 이 과정에서 관측 기기의 한계로 인해 탁도 자료의 결측치가 발생하였고, 이를 효율적으로 보완하기 위해 새로운 결측치 보완 기법을 개발하였다. 개발된 기법, 일명 면적비법은 시계열 자료가 단일주기와 상이한 진폭을 갖는다는 가정 하에, 각 사이클의 면적비율을 통해 결측치를 보완하는 방법이다. 면적비법과 기존의 최소제곱법을 검증하기 위해 결측치가 없는 정상적인 자료에 적용해 보면, 두 방법 모두 첨두치를 약간 과소 산정하는 경향이 있었다. 하지만 면적비법의 경우, 원자료의 총 면적과 보완자료의 총 면적간의 차이가 거의 없었다. 이 방법들을 한강 하구부에서 관측된 탁도자료에 적용해 본 결과, 면적비법은 합리적으로 결측치를 보완하는 반면, 최소제곱법은 보완자료의 총면적이 원자료에 비해 작아지는 오류가 발생하였다. 따라서 최소제곱법에 비해 면적비법이 결측치 보완에 더 우수한 결과를 제공함을 알 수 있었다. 본 연구에서 개발한 면적비법은 주기성이 뚜렷한 시계열자료의 결측치 보완에 유용하게 쓰일 수 있으리라 기대된다.

**핵심용어** : 탁도, 하구, 시계열 자료, 결측치, 보완기법

### 1. 서론

하구부에서 탁도(turbidity)나 부유사농도(suspended sediment concentration; SSC)의 장기적인 관측시, 측정기기의 관리소홀이나 오작동으로 인해 시계열자료의 결측치(missing data)가 발생할 수 있다. 특히 조석간만의 영향으로 탁도의 변동폭이 큰 하구부인 경우, 관측 기기의 최대 관측 한계를 넘어서는 경우가 많아, 관측 자료의 첨두치가 결측되곤 한다(해양연구원, 2005). 이러한 경우 결측된 첨두치를 합리적인 방법을 통해 정확히 보완(filling)해 줄 필요가 있다. 중·장기간의 수문시계열 자료의 경우, 경험적 방법(empirical method), 추계적 방법(stochastic method), 신경망 기법(neural network)등을 이용하여 결측치를 보완할 수 있다고 알려져 있다(수자원의 지속적인 확보기술 개발 사업단, 2004). 반면 하구부에서 관측된 수리자료들은 수문자료와 달리 단기간의 변동량 및 변동폭이 매우 심하여, 기존의 수문시계열 보완을 위한 기법들로 결측자료를 보완하기엔 어려움이 많다. 따라서 조석의 영향으로 강한 주기성을 내포한 수리자료의 결측치를 보완할 수 있는 새로운 기법의 제안이 필요한 실정이다. 본 연구에서는 수중 계류방식으로 약 5개월에 걸쳐 한강 하구부의 3개 지점에서 탁도를 관측한 바, 이 과정에서 관측기기의 한계로 인해 탁도 자료의 결측치가 발생하였다. 이를 효율적으로 보완하기 위해 새로운 결측치 보완기법을 개발하였다. 그리고 결측되지 않은 정상적인 자료 일부를 추출하여 개발된 기법의 적용성을 검토해

\* 정회원 · 한국건설기술연구원 박사 후 과정 · E-mail : kopaek1@kict.re.kr

\*\* 정회원 · 한국해양연구원 책임연구원 · E-mail : hycho@kordi.re.kr

\*\*\* 정회원 · 한국건설기술연구원 수석연구원 · E-mail : samhee.lee@kict.re.kr

보았다. 검증과정을 통해 개발된 기법의 합리성을 확인 후 이를 관측된 탁도자료에 적용하여 결측치를 보완하였다.

## 2. 결측치 보완법

본 연구에서는 시계열 자료가 단일주기(unique period)와 상이한 진폭(different amplitude)을 갖는다는 가정하에 선 사이클(preceding cycle)과 후 사이클(following cycle)의 면적비율을 이용하여 결측치를 보완할 수 있는 새로운 기법을 개발하였다. 먼저 진폭이 다르며 주기  $T$ 를 갖는 시계열자료를 단순 도식화하면 그림 1과 같다. 이 그림에서 후 사이클( $C_2$ ) 구간  $a$ 에서  $T-b$ 사이의 관측값이 측정범위를 넘어서 결측되었다고 가정한다. 이를 보완하기 위해 선 사이클( $C_1$ )에서  $C_2$ 의 결측구간과 대응하는 구간( $a \sim T-b$ )을 제외한 나머지 구간의 면적( $S_1 + S_2$ )을 계산한다. 그리고  $C_2$ 에서도 결측구간을 제외한 면적( $S_3 + S_4$ )을 구한다. 이 둘의 면적비(area ratio)를  $\alpha$ 라 하면 다음과 같이 표현된다.

$$\alpha = \frac{S_3 + S_4}{S_1 + S_2} = \frac{aC_2(a) + bC_2(T-b)}{aC_1(a) + bC_1(T-b)} \quad (1)$$

$C_2$ 의 결측구간에  $C_1$ 의  $\alpha$ 배한 자료를 대치(overlap)시킴으로써 결측치를 보완할 수 있다. 이를 수식으로 표현하면 다음과 같다.

$$[C_2(a), C_2(a+1), \dots, C_2(T-b)] = \alpha [C_1(a), C_1(a+1), \dots, C_1(T-b)] \quad (2)$$

이상의 결측치 보완법을 면적비법(area ratio method; ARM)이라 명한다. 이 방법은 시계열자료의 주기성을 이용하여 간단한 원리로 비교적 손쉽게 결측치를 보완할 수 있는 장점이 있다.

본 연구에서 개발한 면적비법은 Hamilton(1999)이 제안한 시계열자료 예측법의 기본 원리인 선형 투영법(linear projection)에 바탕을 두고 있다. Hamilton(1999)은 과거 관측자료가 미래 예측자료와 선형적인 함수관계임을 제안한 바 있다. 이를 수식으로 표현하면 다음과 같다.

$$C_{i+1|t} = \alpha' C_t \quad (3)$$

여기서  $C_{i+1|t}$ 는 과거 관측 자료들에 바탕을 둔 예측치이고,  $\alpha'$ 는 투영계수(projection coefficient)이며,  $C_t$ 는 과거 관측자료들로서  $\{C_t, C_{t-1}, C_{t-2}, \dots\}$ 이다. 식(3)을 응용하여, 결측되지 않은 선 사이클 자료를 과거 관측자료로 하고, 결측된 후 사이클 자료를 미래예측자료로 삼아 결측치를 보완할 수 있다. 본 연구에서 개발한 면적비법 또한 투영계수를 면적비로 대체하는 원리이다. 반면 Hamilton(1999)은 최소제곱법(least square regression; LSR)을 통해 투영계수를 산정함을 제안한 바 있다.

## 3. 결측치 보완법의 검증 및 적용

본 연구에서는 한강 하류부의 신곡 수중보 직상류(C3 지점)와 직하류(C2 지점), 그리고 전류 수위표(C1 지점) 근처에서 5개월 이상 장기간 관측장비를 수중에 계류하여 유속 및 탁도를 관측하였다. 여기서 탁도계를 통해 수집된 탁도 자료는 조홍연과 김백운(2003)이 제안한 방법을 따라 부유사 농도자료로 환산되었다. 그 결과를 그림 2에 도시하였는데, 고농도의 부유사 자료가 관측 기기의 한계로 인해 결측되었다. 결측된 자료를 새로 개발된 보완법으로 보완하는 절차는 다음과 같다.

### 3.1 주기성 검토

본 연구에서 제안한 결측치 보완법은 주기성을 갖는 시계열 자료에만 적용 가능하므로, 이를 적용하기 위해서 우선 한강하구부에서 관측한 부유사 자료가 주기성을 가지고 있느냐를 판별해야 한다. 정상자료(stationary random data)인 경우 자기 스펙트럼 밀도 함수(autospectral density function)를 통해 주기를 파악할 수 있다(Bendat 와 Piersol, 2000). 시간  $t(-\infty < t < \infty)$ 와 지수(Index)  $k$  를 갖는 단일 시계열자료  $\{x_k(t)\}$ 에 대해 자기 상관함수(autocorrelation function)  $R_{xx}$ 는 다음과 같이 표현할 수 있다.

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2) dx_1 dx_2 \quad (4)$$

여기서  $\tau$ 는 지체시간(time lag),  $x_1 = x_k(t)$ ,  $x_2 = x_k(t + \tau)$ , 그리고  $p(x_1, x_2)$ 은 결합 확률밀도함수(joint probability density function)이다. 자기 스펙트럼 밀도함수(autospectral density function)  $G_{xx}$ 는 Fourier 변환에 의해 자기상관함수의 적분형태로 다음과 같이 표현된다.

$$G_{xx}(f) = 4 \int_0^{\infty} R_{xx}(\tau) \cos 2\pi f \tau d\tau \quad 0 < f < \infty \quad (5)$$

여기서  $f$ 는 진동수이다. 식(5)를 통해 시계열 자료의 자기 상관성이 가장 큰 진동수를 파악할 수 있다. 관측점 C1의 부유사 자료 중 결측되지 않은 자료(약 6 일간 자료)를 추출하여 진동수와 자기 스펙트럼 밀도 함수를 도시하면 그림 3과 같다. 이 그림에서 보는 바와 같이 자기상관성이 가장 큰 경우, 즉 자기 스펙트럼 밀도 함수가 가장 큰 값을 갖는 진동수는 0.013514이고 주기는 74이다. 부유사 농도자료가 10분 간격으로 독취 되었으므로 12시간 20분이 부유사 농도자료의 주기임을 확인할 수 있다. 이는 조석의 반일주조 주기인 12시간 25분과 거의 일치한다.

### 3.2 보완법의 검증

개발된 결측치 보완법의 적용성을 검증하기 위해 C1 지점의 결측되지 않은 부유사 농도자료 일부를 추출하여 적용하였다. 이 자료는 또한 주기성 검토가 행해진 자료와 동일한 자료이다. 추출된 원자료에서 부유사 농도가 400 mg/l 이상인 경우에 자료가 결측되었다는 가정하에 면적비법(ARM) 및 최소제곱법(LSR)을 적용하였다. 그 결과를 그림 4에 도시하였다. 이 그림에서 보듯이 각 주기마다 과대 산정이 되는 부분도 있고, 과소 산정되기도 하지만 전체적으로 결측치를 잘 보완하고 있다. 면적비법과 최소제곱법에 의한 보완치와 원자료와의 rms(root mean square) 오차를 보면, 면적비법이 약 98.58 mg/l, 최소제곱법이 약 101.1 mg/l로서 면적비법이 최소제곱법에 비해 더 나은 결과를 제공한다.

### 3.3 보완법의 적용 및 검토

개발된 결측치 보완법을 사용하여 한강 하구부에서 관측된 부유사 농도자료의 결측치를 보완하였고, C1지점의 보완결과를 그림 2에 도시하였다. 이 그림에서 면적비법은 합리적인 보완치를 산출하고 있는 반면, 최소제곱법은 다수 결측구간에서 결측 하한치 이하로 예측하거나 낮은 침투치를 산출하였다. 이로 인해 최소제곱법으로 보완된 총부유사량이 오히려 결측된 원자료의 총부유사량 보다 적어지는 오류가 발생하였다. 이러한 원인은 불규칙한 주기자료에 기인한 것이라 사료된다. 최소제곱법은 선 사이클 자료와 후 사이클 자료의 일대일 대응관계가 선형성을 벗어날수록 오차는 커질 수 밖에 없다. 또한 C1 지점의 자료처럼 각각의 사이클에서 결측된 구간이 넓을수록 일대일 대응할 자료수가 줄어들기 때문에 오차 발생 가능성은 커진다. 반면 면적비법은 합

을 통한 면적비를 고려하므로 일대일 대응관계에 덜 민감하게 반응할 수 있다. 따라서 이러한 경우, 최소제곱법에 비해 면적비법이 결측치 보완에 더 우수한 결과를 제공함을 알 수 있다.

#### 4. 결론

면적비법은 시계열 자료가 단일주기와 상이한 진폭을 갖는다는 가정하에, 각 사이클의 면적비를 이용하여 결측치를 보완하는 방법이다. 면적비법과 최소제곱을 검증하기 위해 결측치가 없는 정상적인 자료에 적용해 보면, 두 방법 모두 첨두치를 약간 과소산정하는 경향이 있다. 하지만 면적비법의 경우, 원자료의 총 면적과 보완자료의 총 면적간의 차이가 거의 없었다. 이 방법들을 한강 하구부에서 관측된 탁도자료에 적용해 본 결과, 면적비법은 합리적으로 결측치를 보완하는 반면, 최소제곱법은 보완자료의 총면적이 원자료에 비해 작아지는 오류가 발생하였다. 따라서 최소제곱법에 비해 면적비법이 결측치 보완에 더 우수한 결과를 제공함을 알 수 있었다. 본 연구에서 개발한 면적비법은 주기성이 뚜렷한 시계열자료의 결측치 보완에 유용하게 쓰일 수 있으리라 기대된다.

#### 참고문헌

1. 수자원의 지속적인 확보기술 개발 사업단 (2004). 시공간 자료의 통계적 검·보정 기술의 개발, 1단계보고서, 과학기술부.
2. 조홍연, 김백운, (2003), “포획된 부유물질의 입도분포를 고려한 탁도계 검교정.” 한국해양·해양공학회지, 한국해양·해양공학회, 제15권, 제1호, pp. 33-38.
3. 한국해양연구원 (2005). 한강 하구부 수리특성 측정조사 연구, 한국해양연구원.
4. Bendat, J.S., and Piersol, A.G. (2000). *Random data*, 3rd ed., John Wiley & Sons Inc., NY.
5. Hamilton, J.D. (1999). *Time series analysis*, Princeton Univ. Press, NJ.

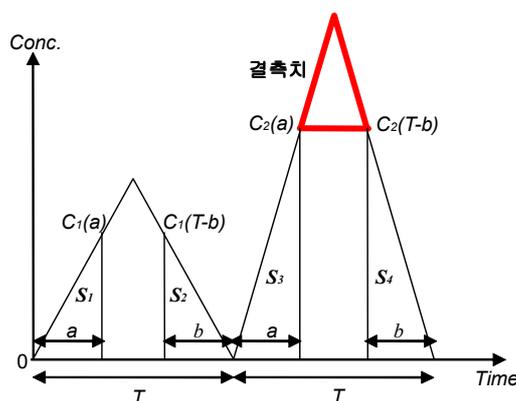


그림 1. 결측치 보완법(ARM)의 개념도

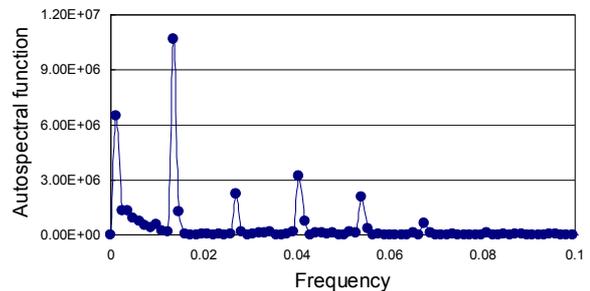


그림 3. 주기성 파악

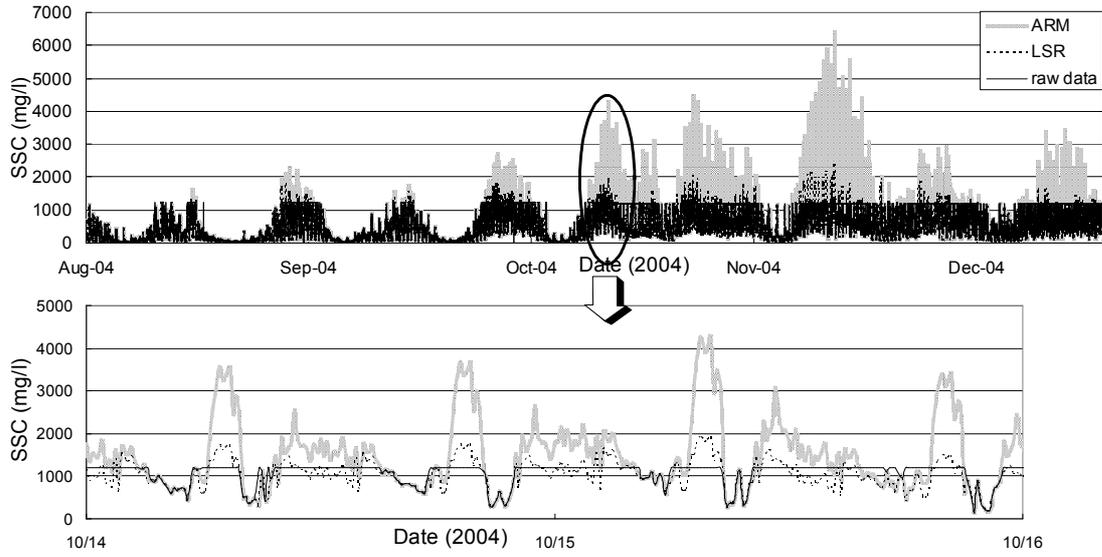


그림 2. 결측치 보완법의 적용

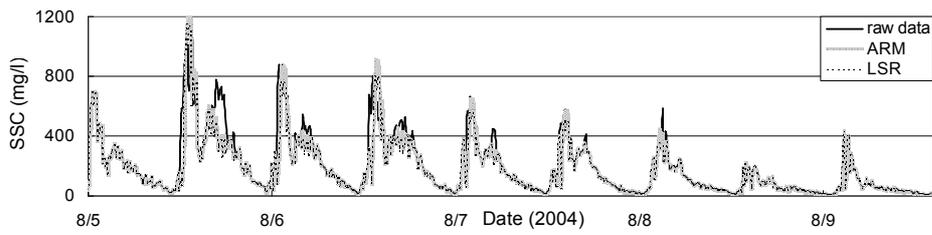


그림 4. 결측치 보완법의 검증