한국해안 · 해양공학회논문집/ISSN 1976-8192(Print), ISSN 2288-2227(Online)
Journal of Korean Society of Coastal and Ocean Engineers 26(4), pp. 207~216, Aug. 2014
http://dx.doi.org/10.9765/KSCOE.2014.26.4.207

화학적산소요구량의 총유기탄소 변환을 위한 이상자료의 탐지와 처리

Outlier Detection and Treatment for the Conversion of Chemical Oxygen Demand to Total Organic Carbon

조범준* · 조홍연** · 김 성***
Beom Jun Cho*, Hong Yeon Cho** and Sung Kim***

요 지 : 총유기탄소(TOC)는 해양의 탄소순환 연구분야에서 직접적인 생물학적 지표로 이용되는 중요한 인자다. 가용한 TOC 자료가 상대적으로 화학적산소요구량(COD) 자료 보다 부족하기 때문에 COD 자료를 활용하여 TOC 자료를 추정할 수 있다. COD를 TOC 로의 변환 시 TOC 추정에 직접적으로 영향을 미치는 COD 관측자료에 포함된 이상자료의 탐지와 적절한 처리는 합리적이고 객관적으로 수행되어야 한다. 본 연구에서는 국내 연안해역에서 관측된 염분, COD 및 TOC 자료에 대한 최적회귀모형을 제시하였다. 최적회귀모형은 이상자료와 영향자료를 여러 가지 탐색방법으로 진단하여 제거 전·후의 자료 개수 변화, 변동계수 및 RMS 오차를 비교 및 분석하여 선택하였다. 연구수행 결과, Cook의 진단방법과 SIQR의 boxplot 방법을 조합한 방법이 가장 적절한 것으로 파악되었다. 최적 회귀 함수는 TOC(mg/L) = 0.44·COD(mg/L)+1.53 이고, 결정계수는 0.47 정도로 나타났으며, RMS 오차는 0.85 mg/L이다. RMS 오차와 지레계수(leverage values)의 변동계수는 이상자료 제거 전에 비하여 각각 31%, 80%로 크게 감소되었다. 본 연구에서 제시된 방법을 통해 COD와 TOC 관측자료에 포함된 이상자료와 영향자료의 과도한 영향을 진단 및 제거하였기 때문에 보다 적절한 회귀곡선식을 제시할 수 있었다.

핵심용어 : 이상자료, 최적회귀모형, RMS 오차, 결정계수, SIQR boxplot과 Cook's 계수

Abstract : Total organic carbon (TOC) is an important indicator used as an direct biological index in the research field of the marine carbon cycle. It is possible to produce the sufficient TOC estimation data by using the Chemical Oxygen Demand(COD) data because the available TOC data is relatively poor than the COD data. The outlier detection and treatment (removal) should be carried out reasonably and objectively because the equation for a COD-TOC conversion is directly affected the TOC estimation. In this study, it aims to suggest the optimal regression model using the available salinity, COD, and TOC data observed in the Korean coastal zone. The optimal regression model is selected by the comparison and analysis on the changes of data numbers before and after removal, variation coefficients and root mean square (RMS) error of the diverse detection methods of the outlier and influential observations. According to research result, it is shown that a diagnostic case combining SIQR (Semi - Inter-Quartile Range) boxplot and Cook's distance method is most suitable for the outlier detection. The optimal regression function is estimated as the TOC(mg/L) = 0.44·COD(mg/L)+1.53, then determination coefficient is showed a value of 0.47 and RMS error is 0.85 mg/L. The RMS error and the variation coefficients of the leverage values are greatly reduced to the 31% and 80% of the value before the outlier removal condition. The method suggested in this study can provide more appropriate regression curve because the excessive impacts of the outlier frequently included in the COD and TOC monitoring data is removed.

Keywords : outlier, optimal regression model, RMS error, determination coefficient, SIQR boxplot and Cook's distance

1. 서 론

총유기탄소(TOC, Total Organic Carbon)는 해양생물이나 미생물의 먹이원일 뿐 아니라 탄소순환에서 저장고의 역할을 하는 중요한 인자이다(Chen and Bada, 1992; Hedges, 2002; Kim et al., 2006). 이 생물학적 인자는 해수 중 유기물의 함량을 파악할 수 있는 직접적인 지표로 매우 유용하다

*한국해양과학기술원 해양환경·보전연구부 (Corresponding author : Beom Jun Cho, Marine Environments & Conservation Research Division, Korea Institute of Ocean Science & Technology, Ansan, GyeongGi-Do, 426-744, Korea, Tel.:+82-31-400-7807, Fax:+82-31-400-7868, bjcho@kiost.ac)
**한국해양과학기술원 해양환경 · 보전연구부 (Marine Environments & Conservation Research Division, Korea Institute of Ocean Science & Technology, Ansan, GyeongGi-Do, 426-744, Korea)
***한국해양과학기술원 해양생태계연구부 (Marine Ecosystem Research Division, Korea Institute of Ocean Science & Technology, Ansan, GyeongGi-Do, 426-744, Korea)

(Doval and Hansell, 2000). 그러나, TOC는 화학적산소요구량(COD, Chemical Oxygen Demand)에 비해 분석비용이 2배 정도 높고(Ministry of Maritime Affairs and Fisheries, 2013a), 최근에서야 유기물질의 중요한 지표로 인식되었기 때문에 장기간 측정 자료가 상대적으로 부족하다.

COD는 해수내의 유기물을 과망간산칼륨으로 산화시켜 소비되는 산소의 양으로부터 측정된 유기물의 농도이다. COD의 분석 시간은 3시간에 불과(Tchobanoglous and Schroeder, 1985)하여 해양환경 관측항목에서 장기간의 측정 자료가 TOC보다 풍부하다. 해양의 탄소 순환 연구에서 직접 이용이 가능한 TOC 자료를 확보 또는 일반적 환경 모니터링을 위해서 COD와 TOC간 자료 변환 관계식은 매우 중요하다.

TOC 측정이 어려울 경우, COD 분석을 통한 TOC 추정은 유용성이 매우 높을 것이다. 해양에서 조사된 COD와 TOC 농도의 상관관계에 대한 연구는 Son et al. (2003)의 자료와 같이 신뢰성 높은 결과를 찾기가 매우 어렵다. 왜냐하면, 해수의 COD 측정은 유기물의 난분해성 정도와 염분 등의 간섭조건에 따라 산화정도의 영향을 받기 때문이다(Son et al., 2003; Ministry of Maritime Affairs and Fisheries, 2013ab). 즉, 분석 과정에서 측정 오류가 동반된 또는 평균적인 범위를 크게 벗어나는 이상자료가 발생할 가능성이 높아진다.

현장 관측에서 발생하는 빈번한 이상자료는 분석결과에 큰 영향을 미치기 때문에 관측자료를 분석하기 전에 이상자료를 진단하고 처리하는 과정은 매우 중요하다.

이상자료의 발생 요인은 관측장비의 검교정(calibration), 자료 입력의 실수, 센서관리 미흡, 안정적인 전원공급 제한, 센서의 오작동 등 대우 다양하다(Cho and Oh, 2012). 이러한 이상자료 진단과 처리는 의학, 보건, 수문, 교통, 통신, 대기 및 통계 등의 다양한 분야에서 폭넓게 활용되고 있다. 해양에서는 수온이나 음향자료 등 매우 제한 영역에서만 이상자료의 검출과 처리를 통해 자료의 품질을 높이고 있다(Cho and Oh, 2012; Lee et al., 2001). 이외에도 국내외 연구사례는 강이나 하천에서의 COD와 TOC의 상관성에 대한 연구는 있었으나 해양에서는 기존 Son et al. (2003)의 논문 이외에는 연구사례가 없고, COD와 TOC의 직접적인 상관관계에 대한 근거가 다소 미미하기 때문에 기존 논문결과를 활용하기에 매우 제한적 이었다. 현 상황에서 국내외 사례가 많지 않아서 이전의 연구결과와 연관시키기에는 객관적인 신뢰성을 판단하기가 어려움이 있다.

일반적으로 연구자가 각자의 경험과 자료특성을 감안하여 주관적 판단 하에 이상자료를 처리할 경우 분석결과의 차이가 발생할 수 있다(Cho and Oh, 2012). 이에 따라 관측자료의 개수가 적은 경우에 이상자료를 간단하게 처리하여 제거할 수 있으나, 관측자료의 개수가 많을 경우 이상자료를 인위적인 수작업으로 처리하는 데 한계가 있기 때문에 자동화된 또는 객관화된 처리기법이 필요하다.

본 연구에서는 해양환경인자인 염분, COD 및 TOC의 가용한 동시 관측 자료 전부를 이용하여 사전 회귀분석을 수행하였으며, 이 회귀분석 과정에서 발견된 이상자료와 영향자료를 진단하여 처리하였다. 처리 전과 후의 회귀분석 계수와 오차의 변화양상을 진단-처리기법 별로 비교 검토하여 객관적이며 효율적인 이상자료 및 영향자료 진단-처리기법을 선정하고, 선정된 방법을 이용하여 최종적으로 보다 개선되고, 안정된(robust) 환산공식을 제안하였다.

2. 자료 및 분석방법

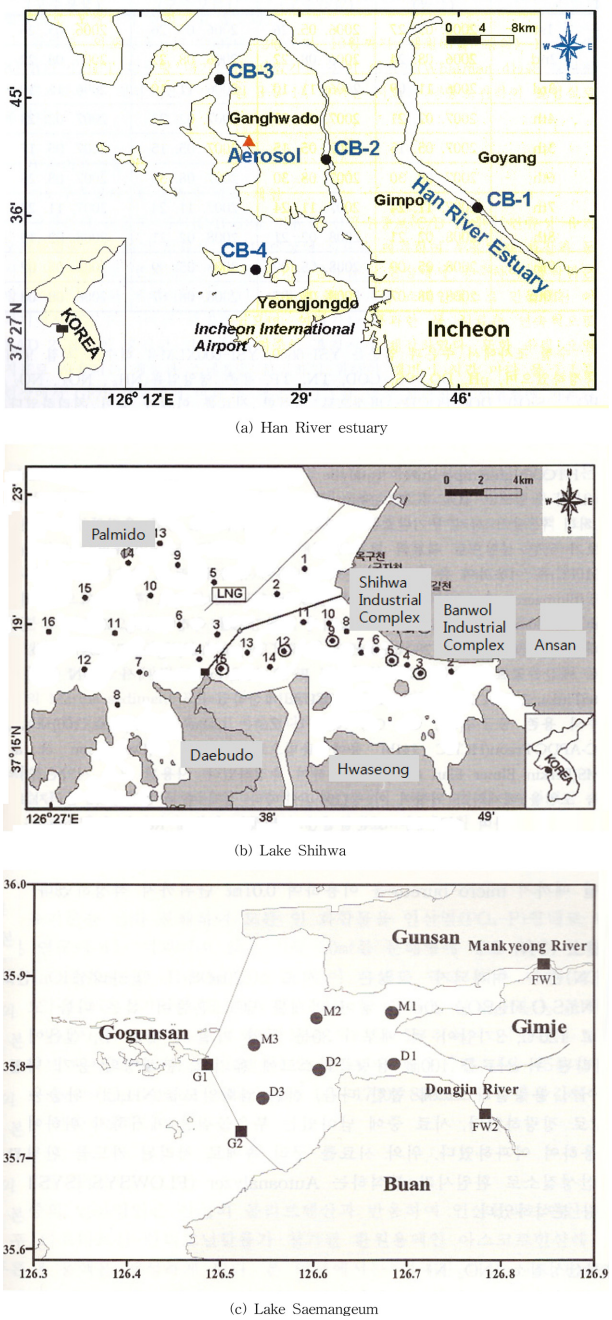
2.1 자료 현황

본 연구에서 사용된 관측자료는 경기만, 시화호, 새만금호에서 관측된 것과 Son et al. (2003)의 자료이다. 경기만 자료는 한강하구역의 신곡수중보, 강화도 서쪽의 염하수로, 강화도와 교동도 사이의 석모수로, 장봉도와 시도 사이에서 수집한 것이다(Fig. 1(a)). 관측기간은 2006년 5월부터 2008년 2월까지 계절별로 총 8회이다. 표층에서 8~12시간씩 시간별로 연속 관측한 자료이다(Korea Ocean Research & Development Institute, 2008). 시화호의 자료는 2006년 4월, 7월, 8월, 10월에 수집된 것이다. 표층 조사 정점은 14개, 수층별 조사 정점은 5개(정점 3, 5, 9, 12, 15) 이다(Fig. 1(b); Ministry of Maritime Affairs and Fisheries, 2006). 새만금호의 관측시기는 2010년 3월부터 10월, 관측 주기는 1개월 간격이다. 시료 채취는 새만금호의 6개 정점(M1-M3, D1-D3)으로 표층과 저층에서 이루어 졌다(Fig. 1(c); Ministry of Land, Transport and Maritime Affairs, Korea Institute of Marine Science & Technology, 2011). Son et al. (2003)의 자료는 경기만, 강화대교 부근, 세어도, 형산강, 부산해역 등 5개 해역에서 수집된 것이다. 경기만의 자료는 1999년 9월과 12월, 2000년 1월과 4월에 21개 정점의 표층에서 수집된 것이다. 형산강 자료는 2000년 3월 염분 구배에 따라 19개 정점의 표층에서 획득한 것이다. 부산해역의 것은 1999년 5월에 8개 정점의 표층과 2개 정점(정점 4와 8)의 수층별 자료이다. 강화대교의 시계열 자료는 1999년 12월에 9시 45분부터 17시 15분까지 30분 간격으로 관측된 자료이다. 인근 세어도 자료는 2000년 2월에 9시 30분부터 16시 30분까지 30분 간격의 측정된 자료이다(Son et al., 2003).

본 논문에 사용한 관측자료는 문헌자료(논문 및 보고서)에 공개된 원시자료를 이용하였기 때문에 관측된 원시자료에 대하여 객관적인 신뢰성을 설명하기가 어렵고 한계가 있다.

2.2 분석방법

일반적으로 이상자료는 “대부분의 자료와는 다른 특성을 가진 자료”로 정의되고 있으며, 매우 다양한 정량적인 이상자료 진단기법 등이 개발·제안되고 있다(Barnett and Lewis, 1978; Kottegoda and Renzo, 1997). 그러나 자료의 분포 및



(a) Han River estuary

(b) Lake Shihwa

(c) Lake Saemangeum

Fig. 1. Monitoring stations of the salinity, COD and TOC.

변동 특성이 매우 다양하기 때문에 어떤 하나의 절대적인 기준을 적용하여 진단하는 것은 무리가 있으며, 설사 이상자료로 진단되더라도 처리에 관한 선택문제가 대두된다. 이상자료 처리를 주관적으로 수행하는 경우, 동일한 분석을 답습할 수 없기 때문에 본 연구에서는 회귀분석과정에서의 이상자료를 다음과 같이 정의하고 객관적으로 진단·처리과정을 제시하고자 한다.

모집단의 관측자료에서 이상자료 제거에 관한 객관적인 기준(수치)은 명확하게 제시되어 있지 않다. 다만, 이상자료로 판단되는 관측자료의 탐지 및 제거에 관한 내용을 살펴보면, 일반적인 관측 분포현황에서 벗어나 너무 높거나 낮은 관측자료를 이상자료로 간주하여 이상자료 제거 및 존속을 위한 기초자료로 활용 가능하며, 변량에 따라 약간 차이가 있지만, 모집단의 관측자료에서 대략 5~10% 정도를 이상자료로 간주하여 제거하여도 무방하다고 제시하고 있다(Hair et al., 2010). 본 논문은 이러한 근거에 따라 관측자료를 이상자료 또는 영향자료로 판단함으로써 이상자료를 처리하여 분석하였다.

2.2.1 이상자료(outliers) 진단 및 분석

Robust 회귀분석 결과로 제시되는 각각의 자료에 대한 가중계수가 유난히 작은 경우에 해당하는 자료로, 가중계수의 SIQR boxplot 기법을 이용하여 진단한다. 회귀분석 양상에서 크게 벗어나는 자료로 이상자료로 진단된 자료는 전체 자료의 5% 범위 내에서 모두 제거한다. 이상자료가 모두 제거된 경우 Robust 회귀분석과 기본적인 회귀분석에 의한 매개변수 추정결과 및 RMS 오차, 결정계수(R^2, coefficient of determination) 등이 95% 유의수준에서 동일하게 되어야 한다.

2.2.1.1 Boxplot 기법을 활용한 이상자료 진단

Robust 회귀분석에서 산정된 가중계수를 boxplot 기법에 적용하여 가중계수의 분포가 각각 $Q_1 - 1.5 \times IQR$ 및 $Q_3 + 1.5 \times IQR$ 범위 이외일 경우 약한 이상자료(mild outlier), 분포값이 각각 $Q_1 - 3.0 \times IQR$ 및 $Q_3 + 3.0 \times IQR$ 범위 이외일 경우 강한 이상자료(extreme outlier) 라고 정의한다(Lyman and Longnecker, 2001). 여기서, 1.5, 3.0은 각각 Whisker 길이이며, Q_1, Q_3는 각각 제1사분위수, 제3사분위수(분위수, quartile : 크기 순서에 따라 늘어놓은 자료를 4등분하는 수)이다. IQR(interquartile range)은 $Q_3 - Q_1$이다. 본 연구에서는 이상자료 진단기준으로 강한 이상자료를 이용하였다. Robust 회귀분석의 가중계수(weight)를 사용하는 이유는 관측자료 중에서 이상자료의 영향을 줄일 수 있기 때문이다.

2.2.1.2 SIQR boxplot 기법을 활용한 이상자료 진단

본 연구에서는 이상자료의 판단기준으로 boxplot 통계기법 중에서 Whisker 길이 3.0의 하한 및 상한 이외의 범위와 IQR의 반(Semi-IQR)을 활용한 SIQR boxplot 방법(이하 S-boxplot)을 적용하였다(Kimber, 1990; Aucremanne et al., 2004; Hubert and Vandervieren, 2008). S-boxplot 방법은 하한과 상한을 각각 $SIQR_L = Q_2 - Q_1$, $SIQR_U = Q_3 - Q_2$ 로 정의하며, 관측자료 중에서 하한값이 관측자료의 최소값에서 Q_1까지, 상한값이 Q_3에서 관측자료의 최대값까지를 벗어난 자료에 대하여 이상자료로 판단한다. 다시 말해서 이상자료는 Whisker 길이가 3.0으로 강한(extreme) 이상자료 판단방법과 더불어 이상자료의 손실을 완화시킨 IQR의 반(Semi-IQR)을 활용한 S-boxplot 방법을 비교하여(Kimber, 1990; Aucremanne et al., 2004; Hubert and Vandervieren, 2008) 이상자료 판단 근거로 간주하였으며, 이상자료의 제거 비율은 상기에 제

시된 모집단의 관측자료 중 대략 5~10% 비율에 해당되는 이상자료를 판단 기준으로 설정하였다.

2.2.2 영향자료(influential observations)의 진단 분석 방법

회귀분석 결과(매개변수, 오차, 상관계수 등)에 큰 영향을 미치는 자료로, 이 자료의 유무에 따라 회귀분석 결과가 크게 달라진다. 영향자료도 일반적으로는 이상자료에 포함되지만 회귀분석에서는 자료의 상관양상에서는 벗어나지 않는다는 점에서 차이가 난다. 영향자료의 판단기준도 다양하게 제시되고 있으나 본 연구에서는 널리 이용되고 있는 지레계수(leverage values)와 Cook 계수(Cook's distance)를 이용하여 판단한다(Chatterjee and Hadi, 1986). 영향자료로 진단되는 경우에도 자료의 5% 범위 내에서 모두 제거하였으며, 영향자료가 모두 제거되는 경우, 남아있는 다른 자료를 무작위로 제거하여도 회귀분석 결과에 미치는 영향은 95% 유의수준에서 미미한 수준으로 판단되어야 한다. 그러나 제거 자료의 개수를 한정하기 때문에 미미한 수준을 넘어서는 자료가 남아있을 수도 있다. 한편, 관측자료에서 영향자료 제거에 대한 절대적인 기준은 없다. 다만, 관측된 원시자료에서 영향자료를 너무 적게 또는 과도하게 제거할 경우 영향자료 판단기준(연구자에 따라 절대적인 기준을 제시한 것보다 어느 정도의 이상의 범위를 제시)에 영향을 미칠 수 있으므로 관측자료의 일반적인 오차범위 5%를 감안하여 근거로 제시한 것이다.

2.2.2.1 지레계수(leverage values)

지레계수는 관측자료 중에서 영향자료가 있어 그것이 추정결과에 영향을 미치는지 파악하기 위한 것이다. 다른 관측자료와 비교하여 큰 지레계수를 가지는 관측개체는 영향자료이기 때문에 지레계수는 영향자료의 측도로 사용될 수 있다. Hoaglin and Welsch(1987)는 $h_i\left(=1/n+(x_i-\bar{x})^2/\sum_i(x_i-\bar{x})^2\right) \geq 2(p+1)/n$ 조건을 만족하는 관측자료를 영향자료로 판단하였다. 여기서 h_i는 지레계수, n은 관측자료의 개수, p는 추정된 회귀계수의 개수이고, $x_i, \bar{x}$는 각각 COD의 관측자료, COD 관측자료의 평균이다.

2.2.2.2 Cook 계수(Cook's distance)

영향자료 판단기준의 하나로 널리 이용되는 Cook 계수는 전체 데이터로부터 얻은 회귀계수들과 i번째 개체를 제거하고 얻은 회귀계수들의 차이를 측정하며, i번째 관측개체에 대한 영향력을 다음 식으로 진단한다. Cook(1977)은 C_i의 값이 자유도(degree of freedom)가 $(p+1,\ n-p-1)$인 F분포의 50% 보다 큰 경우 i번째 관측자료를 영향자료로 판단하여도 좋다고 제안하였다($C_i \geq F(p+1, n-p-1; 0.5)$). 그러나 Kim and Storer(1996)가 모의실험을 통해 Cook(1977)이 제안한 기준치가 적절하지 못함을 밝혔다. 대신 $C_i \geq 3.67/(n-p)$을 Cook 계수의 기준치로 제시하였다. 본 연구에서는 이 식을 영향자료의 기준치로 사용하였다. 여기서 $C_i = r_i^2/(p+1)\cdot h/(1-h_i)$이며, r_i는 표준화 잔차이다.

2.2.3 분석 절차

전술한 바와 같이, OLS(Ordinary Least Square) 방법과 Robust 방법을 이용하여 각각 회귀식을 먼저 추정하였다. 여기서 산정된 4 가지 경우(영향자료 판단기준 2가지, 이상자료 판단기준 2가지)의 판단기준을 활용하여 이상자료 및 영향자료를 진단하였다. 분석 절차는 다음과 같다.

제1단계 : 상관분석을 수행할 두 인자, 즉 COD 농도, TOC 농도 관측자료를 입력한다. 여기서 독립변수는 COD 농도이고, TOC 농도는 종속변수이다.

제2단계 : 관측자료에 대한 OLS 방법과 Robust 방법으로 추정 회귀식의 기울기, 절편, 결정계수 및 RMS 오차의 변화 정도로 이상자료와 영향자료의 제거 효과를 분석하였다. 여기에서 OLS 회귀분석은 Robust 회귀분석에서 가중치 함수(weight function)가 없음을 의미한다.

제3단계 : 영향자료 판단기준에 지레계수 $h_i \geq 2(p+1)/n$ 와 Cook 계수 $C_i \geq 3.67/(n-p)$ 를, 이상자료 판단기준에 boxplot과 S-boxplot을 활용하여 네 가지 경우의 판단방법을 적용하였다($2 \times 2 = 4$ Cases).

제4단계 : 제3단계에서 적용한 네 종류의 자료 판단방법에 따라 영향자료 판단기준인 Cook 계수와 지레계수의 기준치를 만족하는 관측자료에 일련번호(index)를 부여하였다. 이상자료 판단기준인 boxplot과 S-boxplot의 Robust 가중계수가 하한과 상한 범위 이외인 경우도 관측자료에 번호를 할당하였다. 두 가지 일련번호의 합집합에 해당하는 관측자료를 영향자료 및 이상자료로 간주하여 관측자료에서 제외시켰다.

제5단계 : 이상자료 및 영향자료를 제거 후 다시 네 가지 방법을 통해 OLS와 Robust 추정 회귀식의 기울기 및 절편, 결정계수 및 RMS 오차를 산정하였다.

제1단계에서 제5단계의 자료처리 과정에서 자료의 손실이 가장 적고 RMS 오차 및 변동계수 감소 기준으로 처리 효과가 가장 큰 것은 Cook 계수와 S-boxplot 기법을 조합한 Case-2 진단방법 이다. 자료의 처리효과를 파악하기 위해 이상자료 및 영향자료의 제거 전과 후의 RMS 오차와 영향계수의 변동계수를 분석하였다.

2.2.4 관측 자료의 일련번호

본 연구에서 사용한 자료의 일련번호(data index)는 Son et al. (2003)의 자료 뿐 아니라 한강하구, 시화호 및 새만금호의 COD 및 TOC 관측자료에서 이상자료와 영향자료를 일시적으로 명시하기 위하여 부여된 것이다. 일련번호 부여 순서

Table 1. Index numbering according to observed sampling stations

data index	sampling stations	Number of missing data
1-79	Incheon coastal area	
80-95	Kanghwa Island	5
96-110	Seo Island	Sts. 7, 13, 17, 18, 1999; St.4, 2000
111-130	Hyungsan River	in Incheon coastal area
131-150	Busan coastal area	
151-182	Han river estuary	-
183-300	Lake Shihwa (MOMAF 2006, 158-162 p)	2 St.6, 2006.4 ; St.9(4m), 2006.8
301-390	Lake Saemangeum (MOLTMA 2011, 206-213 p)	6 Sts. M1-M3, D1-D3, 2010.5

* data index of Son's et al. (2003) : 1-150

Table 2. Observed data in Han river estuary

date	Index	Salinity (PSU)	COD (mg/L)	TOC (mg/L)
Singok submerged weir (CB-1)				
2006. 05	151	0.10	3.80	4.75
2006. 08	152	0.09	3.15	2.99
2006. 11	153	0.28	5.51	5.98
2007. 02	154	0.49	6.56	6.20
2007. 05	155	0.17	6.16	4.95
2007. 08	156	0.11	3.04	3.17
2007. 11	157	0.17	4.67	3.89
2008. 02	158	0.23	11.93	9.41
Yeomha Channel (CB-2)				
2006. 05	159	14.61	3.91	3.91
2006. 08	160	9.67	4.37	3.92
2006. 11	161	23.40	8.28	8.55
2007. 02	162	25.50	9.03	11.28
2007. 05	163	20.40	4.69	8.05
2007. 08	164	7.38	5.13	6.22
2007. 11	165	20.08	7.53	7.97
2008. 02	166	24.23	10.35	13.70
Seokmo Channel (CB-3)				
2006. 05	167	19.84	3.75	3.08
2006. 08	168	15.08	3.21	3.60
2006. 11	169	25.83	6.35	7.74
2007. 02	170	26.36	6.94	10.57
2007. 05	171	23.13	3.06	3.84
2007. 08	172	10.58	3.72	4.54
2007. 11	173	22.69	5.07	5.97
2008. 02	174	25.65	6.67	7.96
Jangbong Channel (CB-4)				
2006. 05	175	28.49	2.23	1.78
2006. 08	176	26.77	2.94	2.88
2006. 11	177	30.14	2.25	3.00
2007. 02	178	30.09	2.30	3.93
2007. 05	179	30.37	1.97	3.30
2007. 08	180	24.97	2.71	2.98
2007. 11	181	28.97	1.32	2.05
2008. 02	182	30.50	1.59	2.18

는 결측된 관측자료를 제외한 후 조사 시기와 조사 정점 순으로 정렬하여 정하였다. 관측지점이 고정지점일 경우 시간과 수심별로 정렬한 후 순서를 부여하였다(Table 1 참조). 조사에 사용된 원시 관측자료가 수록된 Son et al. (2003), 새만금호(Ministry of Land, Transport and Maritime Affairs, Korea Institute of Marine Science & Technology, 2011), 시화호(Ministry of Maritime Affairs and Fisheries, 2006)를 제외한 한강하구의 관측자료는 Table 2에 제시하였다.

3. 결 과

3.1 COD와 TOC 관측자료 분석 결과

본 연구에서 활용된 COD와 TOC의 관측자료는 전체 390개 이다. 조사 해역별로 관측자료의 개수는 한강하구 32개, 시화호 118개, 새만금호 90개, Son et al. (2003)의 자료 150개 이다.

이들 자료를 한강하구, 시화호, 새만금호 및 Son et al. (2003)의 자료를 모두 포함한 전체 관측자료와 Son et al. (2003) 자료의 두 집단으로 구분하였다. COD와 TOC 관측자료에서 OLS 방법과 Robust 방법을 통해 회귀식으로 추정한 결과를 Table 3에 나타내었다(Case-0). 두 종류의 자료에서 추정된 OLS와 Robust 회귀분석의 기울기, 절편, 결정계수 및 RMS 오차 등의 차이는 OLS 회귀분석 결과를 기준으로 전체자료는 0.9~15.4%, Son et al. (2003)의 자료는 3.5~18.5% 였다(Table 3). 관측자료 중에서 일부 자료는 회귀직선 상에서 과도하게 벗어나고 있다(Fig. 2(a)). 이 자료는 이상자료 이거나 영향자료일 가능성이 높다.

3.2 이상자료와 영향자료 진단 및 처리 결과

네 가지 진단방법을 적용하여 COD와 TOC 관측자료 390개와 Son et al. (2003)의 자료 150개에서 이상자료와 영향자료를 진단하였다. 관측자료의 진단과 처리에서 자료의 손실이 가장 적고 처리효과가 가장 큰 Case-2로 확인된 이상자료와 영향자료는 각각 22개와 12개 였다(Table 3). 이러한 자료들을 제거한 후에 OLS와 Robust 회귀분석을 수행하였다(Fig. 2(b)).

OLS와 Robust 회귀분석으로 추정된 기울기, 절편, 결정계수 및 RMS 오차 등에 대한 두 종류의 자료 차이는 OLS 회귀분석 결과를 기준으로 0.1~4.2%로 동일하게 나타났다. 이 차이는 이상자료와 영향자료 제거 후에 1/3 이상 정도로 감소하였다.

3.3 이상자료와 영향자료 처리 효과

자료 처리를 통한 자료의 품질 향상 효과 분석에 이상자료와 영향자료의 제거 전과 후의 RMS 오차와 변동계수(표준편차/평균)의 변화를 이용하여 판단하였다. 네 종류의 방법으로 자료 처리 후 전체 관측자료의 OLS 회귀분석 방법에서 RMS 오차와 영향계수에 대한 변동계수의 저감비율은 각각 29~36%와 74~80%로 나타났다. Son et al. (2003) 자료의 분석결과 RMS 오차는 34~42%, 영향계수에 대한 변동계수 저감비율은 59~71%로 파악되었다(Table 4). RMS 오차 및 변동계수의 저감비율이 크면 클수록 관측자료에 포함된 이상자료와 영향자료가 많이 제거되었음을 의미한다. RMS 오차 저감비율의 의미를 보면, 관측된 원시자료는 이상자료와 영향자료가 모두 포함되어 있으며, 이러한 자료들을 4가지 판단방법에 따라 제거한 후 원시자료(Case 0)에 대한 제거 후 자료(Case 1~4)의 비율[(1 – 제거 후 비율/원시자료) × 100%]을 산정하였다. 이상자료 및 영향자료 제거 후 오차 저감은 당연히 있을 수 있겠지만, 실질적으로 어느 정도의 오차 저감이 있는지 수치로 산정해 보고자 하였다.

Robust 회귀분석 방법에 대한 RMS 오차와 변동계수 비율은 결과에서 별도로 제시하지 않았다. 제거 후 진단방법에 따라 OLS 방법과 Robust 회귀분석 방법의 RMS 오차가 거의 유사하게 파악되어서 상기에 제시한 비율과 유사한 것으로 판단되었기 때문이다(Table 3).

Table 3. Analysis results according to diagnostic methods of Son's et al and all data

DM (diagnostic method)		regression parameters		R^2	RMS error	n	outlier index	influential data index
		intercept	slope					
all data								
Case-0	BOR_OLS	1.77	0.40	0.46	1.23	390	-	-
	BOR_RLS	1.50	0.44	0.45	1.24			
Case-1	AOR_OLS	1.56	0.43	0.43	0.88	369	1, **17,** 27, 31, 36, **120,** 161, 162, 163, 165, **166**, 169, 170, 174, 257, 305	**17, 120,** 158, **166,** 183, 271, 307, 313
	AOR_RLS	1.50	0.44	0.43	0.88			
Case-2	AOR_OLS	1.53	0.44	0.47	0.85	368	**1, 17, 27, 31,** 36, **120, 161, 162, 163, 165, 166, 169, 170, 174,** 257, **305**	**1, 17, 27, 31, 120,** 158, **161, 162, 163, 165, 166, 169, 170, 174,** 256, **305,** 307, 309, 310, 313
	AOR_RLS	1.46	0.45	0.47	0.85			
Case-3	AOR_OLS	1.55	0.42	0.48	0.79	358	1, **17,** 27, 31, 36, **120, 158,** 161, 162, 163, 164, 165, **166,** 169, 170, 173, 174, 199, 207, 234, 256, 257, 259, 305, **307,** 309, 310, 312, 327	**17, 120, 158, 166,** 183, 271, **307,** 313
	AOR_RLS	1.50	0.44	0.47	0.79			
Case-4	AOR_OLS	1.52	0.44	0.51	0.79	360	**1, 17, 27, 31,** 36, **120, 158, 161, 162, 163,** 164, **165, 166, 169, 170,** 173, **174,** 199, 207, 234, **256,** 257, 259, **305, 307, 309, 310,** 312, 327	**1, 17, 27, 31, 120, 158, 161, 162, 163, 165, 166, 169, 170, 174, 256, 305, 307, 309, 310,** 313
	AOR_RLS	1.46	0.45	0.51	0.79			
Son's et al.								
Case-0	BOR_OLS	1.72	0.32	0.65	0.88	150	-	-
	BOR_RLS	1.40	0.36	0.63	0.91			
Case-1	AOR_OLS	1.32	0.40	0.67	0.58	141	1, **17,** 21, 27, 31, 36, 101, 115	**17**, 120
	AOR_RLS	1.27	0.39	0.67	0.59			
Case-2	AOR_OLS	1.42	0.36	0.60	0.56	138	**1, 17,** 21, **27, 31, 36,** 101, 115	**1, 17, 27, 31, 36,** 111, 112, 118, 120
	AOR_RLS	1.35	0.36	0.59	0.57			
Case-3	AOR_OLS	1.33	0.38	0.69	0.52	136	1, 11, **17,** 21, 27, 31, 36, 101, 102, 114, 115, 116, 118	**17**, 120
	AOR_RLS	1.29	0.37	0.69	0.53			
Case-4	AOR_OLS	1.40	0.35	0.64	0.51	134	**1,** 11, **17,** 21, **27, 31, 36,** 101, 102, 114, 115, 116, **118**	**1, 17, 27, 31, 36,** 111, 112, **118,** 120
	AOR_RLS	1.34	0.35	0.64	0.51			

* **Case-0** : raw data with before removal of outlier and influential data, **Case-1** : leverage and SIQR boxplot, **Case-2** : Cook's distance and SIQR boxplot, **Case-3** : leverage and boxplot, **Case-4** : Cook's distance and boxplot, **n** : data number, **BOR, AOR** : before and after outlier removal, respectively, **OLS, RLS** : ordinary and robust least square, **bold underline** : outlier index and influential data index

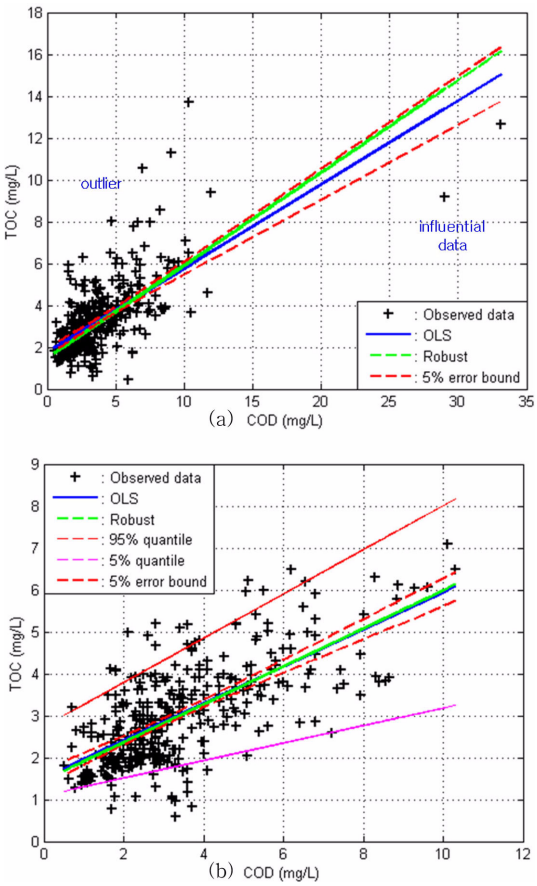


Fig. 2. Scatter plot and regression curves of COD and TOC (a) before removal in outlier and influential data. (b) after removal in outlier and influential data.

Table 4. Reduction ratio before and after removal of the outlier and influential data

(a) RMS error due to outlier

Son's et al.				Total data			
DM	OLS	RR(%)	DR	DM	OLS	RR(%)	DR
Case-0	0.88		BR	Case-0	1.23		BR
Case-1	0.58	34	AR	Case-1	0.88	29	AR
Case-2	0.56	36	AR	Case-2	0.85	31	AR
Case-3	0.52	40	AR	Case-3	0.79	36	AR
Case-4	0.51	42	AR	Case-4	0.79	36	AR

(b) coefficient of variation due to influential point

Son's et al.				Total data			
DM	Cook's	LV	RR(%)	DM	Cook's	LV	RR(%)
Case-0	7.38	3.05		Case-0	10.35	3.42	
Case-1	-	0.90	71	Case-1	-	0.87	75
Case-2	2.41	-	67	Case-2	2.11	-	80
Case-3	-	0.93	70	Case-3	-	0.90	74
Case-4	2.99	-	59	Case-4	2.17	-	79

* RR(reduction ratio, %) : reduction ratio by RMS error and variation coefficient based on the case-0, DM : diagnostic method, DR : before and after outlier removal, BR : before removal, AR : after removal, LV : leverage value

3.4 염분과 COD 자료의 진단 및 처리 효과

Son et al. (2003)의 자료를 포함한 염분과 COD 전체자료에 Case-2 진단방법을 적용하여 이상자료와 영향자료를 진단

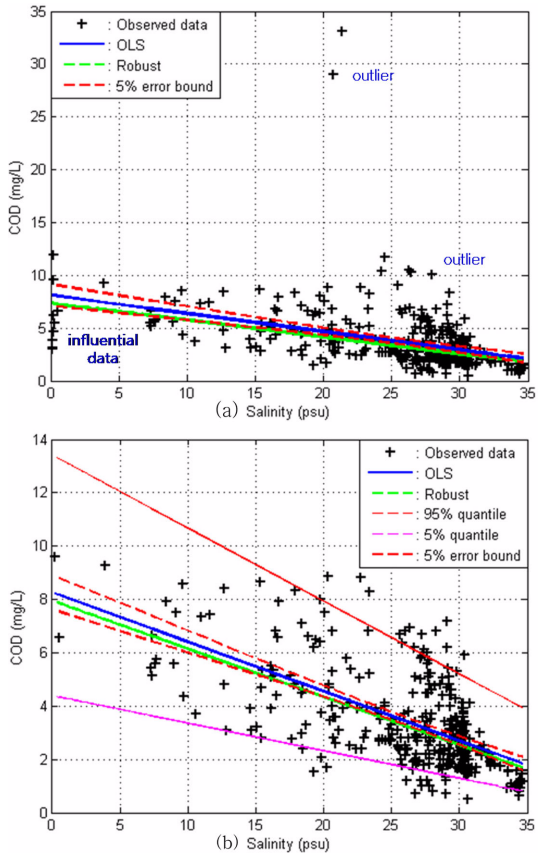


Fig. 3. Scatter plot and regression curves of Salinity and COD (a) before removal in outlier and influential data. (b) after removal in outlier and influential data

Table 5. RMS error and RR(%) of outlier and influential data due to variation coefficient in total salinity, COD and TOC

(a) salinity and COD

DM		correlation intercept	correlation slope	R^2	RMS error	RMS error RR (%)	Cook's	RR (%)
Case-0	BOR_OLS	8.12	-0.17	0.18	2.57	-	5.91	-
	BOR_RLS	7.39	-0.16	0.15	2.61			
Case-2	AOR_OLS	8.27	-0.19	0.39	1.41	45	2.04	65
	AOR_RLS	7.96	-0.18	0.38	1.42			

(b) salinity and TOC

DM		correlation intercept	correlation slope	R^2	RMS error	RMS error RR (%)	Cook's	RR (%)
Case-0	BOR_OLS	5.94	-0.10	0.18	1.52	-	3.65	-
	BOR_RLS	5.37	-0.09	0.15	1.54			
Case-2	AOR_OLS	5.71	-0.11	0.35	0.88	42	1.64	55
	AOR_RLS	5.68	-0.11	0.35	0.88			

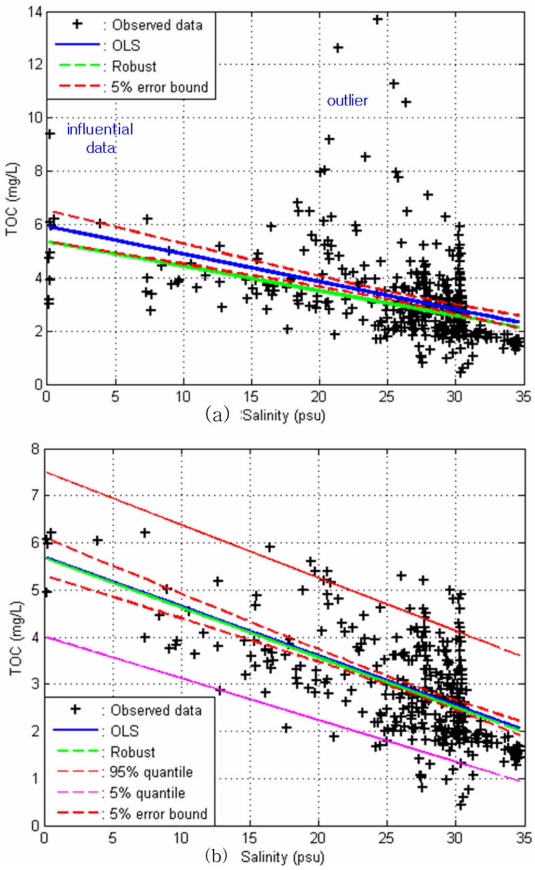


Fig. 4. Scatter plot and regression curves of Salinity and TOC (a) before removal in outlier and influential data. (b) after removal in outlier and influential data

한 후 처리하였다.

이상자료와 영향자료 제거 전 염분과 COD에 대해 OLS와 Robust 회귀분석으로 추정된 기울기, 절편, 결정계수 및 RMS 오차 등은 OLS 회귀분석 결과를 기준으로 두 분석의 차이가 1.5~13.7% 였다. 관측자료 중에서 일부 자료는 회귀직선 상에서 과도하게 벗어났다(Fig. 3a). 이 자료들은 이상자료 이거나 영향자료일 가능성이 높다.

Case-2 진단방법으로 이상자료와 영향자료를 탐색하여 16개 자료를 제거하였다(Fig. 3(b)). 자료 처리 후 이들 자료 제거 후 OLS와 Robust 회귀분석으로 추정된 기울기, 절편, 결정계수 및 RMS 오차 등은 OLS 회귀분석 결과를 기준으로 두 분석의 차이가 0.8~3.7% 였다. 제거 후 OLS 회귀분석 방법에 대한 RMS 오차의 저감비율이 45%, 영향계수의 변동계수 저감비율이 65%로 파악되었다(Table 5).

3.5 염분과 TOC 자료의 진단 및 처리 효과

Son et al. (2003)의 자료가 포함된 염분과 TOC 전체자료에서 이상자료와 영향자료를 진단하고 처리하기 위해 Case-2 진단방법을 적용하였다.

염분과 TOC의 이상자료 및 영향자료에 대한 Case-2 진단방법을 통해 제거 전 OLS와 Robust 회귀분석으로 추정된 기울기, 절편, 결정계수 및 RMS 오차 등은 OLS 회귀분석 결과를 기준으로 두 회귀분석의 차이는 1.8~16% 였다. 관측자료 중에서 일부 자료는 회귀직선 상에서 과도하게 벗어났다(Fig. 4(a)). 이 자료들은 이상자료 이거나 영향자료일 수 있다.

Case-2 진단방법으로 이상자료와 영향자료를 검출하여 33개 자료를 제거하였다(Fig. 4b). Case-2 진단방법 적용 후 염분과 TOC 자료에 대해 OLS와 Robust 회귀분석을 실시하였다. 두 회귀식에서 추정된 기울기, 절편, 결정계수 및 RMS 오차 등에 대한 두 분석결과의 차이는 OLS 회귀분석을 기준으로 0.1~0.4% 정도로 매우 미미하였으며, 이는 이상자료가 적절하게 제거되었음을 의미한다. 제거 후 RMS 오차의 저감비율이 42%, 영향계수의 변동계수 저감비율이 55%로 나타났다(Table 5).

4. 고 찰

분석결과의 정밀도를 현저히 떨어뜨리는 이상자료와 영향자료들(Lee et al., 2001)의 처리 유무에 따라 일반적으로 널리 쓰이는 OLS 회귀분석에서 결정계수, RMS 오차 및 추정 회귀식의 기울기와 절편의 변화가 심하다(Chatterjee and Hadi, 1986). 이상자료와 영향자료의 영향력을 줄이기 위해 OLS 회귀분석의 대안으로 Robust 방법이나 Quantile 회귀분석방법을 사용하기도 한다(Koenker and Bassett, 1978; Koenker and Hallock, 2001; So et al., 2012). 하지만 선형회귀분석의 정밀도와 안정도(robustness)를 높이기 위해서 이상자료와 영향자료의 진단과 처리는 매우 중요하다.

이상자료와 영향자료의 제거에 대한 정량적인 판단이 어려우며(Cho and Oh, 2012), 모집단의 관측자료에서 이상자료 제거에 관한 객관적인 기준(수치)은 명확하게 제시되어 있지 않다. 다만, 이상자료로 판단되는 관측자료의 탐지 및 제거에 관한 내용을 살펴보면, 모집단의 관측자료에서 대략 5~10% 정도를 이상자료로 간주하여 제거한다(Hair et al., 2010).

본 연구에서 TOC, COD, 염분 관측자료를 Cook 계수와 S-boxplot를 활용하여 이상자료와 영향자료 제거후에 OLS와 Robust 회귀분석에서 RMS 오차와 영향계수에 대한 변동계수 모두 큰 폭으로 감소하였다(Fig. 2b, 3b, 4b). 이러한 감소는 TOC, COD, 염분 관측자료에 포함된 이상자료와 영향자료의 진단과 처리에 Cook 계수와 S-boxplot 방법이 매우 효과적임을 시사한다.

본 연구에서 사용한 염분, COD, TOC 자료에서 이상자료와 영향자료의 진단 및 처리방법의 적절성을 확인하기 위해 Son et al. (2003)과 비교하였다. Son et al. (2003)이 분석한 TOC와 COD에 대한 회귀식은 COD(mg/L) = 1.63 × TOC(mg/L) – 0.88, 결정계수는 0.66 이었다. 이 회귀식의 측정단위는 기존 논문에서 원래 mol 농도로 제시되었으나, 본 연구결과와 비교하기 위하여 mg/L 단위로 환산하였다 (1 mmol O_2 [COD]/L = 32 mg/L, 1 mmol C[TOC]/L = 12 mg/L). 본 연구에서 Case-2 방법으로 Son et al. (2003)의 자료를 재분석하여 이상

Table 6. Comparisons of the regression models by Son et al. and this study in case of using the Son et al.'s data

items	this study		Son's study	
	regression model	R^2	regression model	R^2
COD-TOC	COD = 1.57·TOC – 0.98	0.69	COD = 1.63·TOC – 0.88	0.66
TOC-Salinity	TOC = –0.11·Salinity + 5.50	0.67	TOC = –0.10·Salinity + 5.28	0.66
COD-Salinity	COD = –0.21·Salinity + 8.49	0.71	COD = –0.22·Salinity + 8.96	0.62

* units : COD(mg/L), TOC(mg/L), salinity(psu)

자료와 영향자료의 진단 및 처리 결과, COD와 TOC에 대한 OLS 추정회귀식은 COD(mg/L) = 1.57 × TOC(mg/L) – 0.98, 결정계수는 0.69 였다. 두 결과에서 기울기와 절편이 유사하였고, 결정계수는 Son et al. (2003)의 결과 보다 약간 증가하였다(Table 6). 즉, 두 분석방법에서 이와 같은 결과의 유사성은 본 연구에서 염분, COD 및 TOC 등에 포함된 이상자료와 영향자료의 진단 및 처리에 사용한 Case-2 방법이 적절했음을 의미한다.

이상자료와 영향자료 진단 및 처리과정에서 자료 손실은 불가피하지만 가능하면 최소화 하는 것이 바람직하다. 염분, COD, TOC 자료에서 이상자료와 영향자료의 진단 및 처리를 위한 적정 방법을 찾기 위해 본 연구에서는 네 가지 분석방법을 비교하였다. 그 중에 Cook 계수와 S-boxplot 기법을 조합한 Case-2 진단방법을 최종적으로 결정하였다. Case-2를 결정한 이유는 S-boxplot은 SIQR을 활용하여 상한값이 boxplot 상한값보다 커지기 때문에 그만큼 이상자료가 줄어들어서 이상자료로 제거될 관측자료의 손실율이 낮아진다(Hubert and Vandervieren, 2008). 즉 S-boxplot를 활용한 이상자료 일련번호의 개수가 boxplot를 활용한 이상자료 일련번호의 개수 보다 훨씬 적음을 알 수 있었다(Table 3). 영향자료의 진단과 처리를 위해 Cook 계수 및 지레계수를 활용하였다. Cook 계수가 지레계수 보다 제거될 관측자료의 개수가 훨씬 많아서 Cook 계수가 지레계수를 활용한 방법 보다 훨씬 엄격한 진단기법이기 때문이다(Chatterjee and Hadi, 1986; Hoaglin and Welsch, 1987; Kim and Storer, 1996). 또한, Case-2 진단방법이 Case-1 진단방법 보다 RMS 오차가 보다 적다. 따라서 관측자료의 손실율이 높은 Case-3과 Case-4 방법 보다 손실율이 낮은 Case-1과 Case-2 판단방법 중 RMS 오차가 더 적은 Case-2 방법을 최종적으로 결정하였습니다.

본 연구에서 Case-2 진단방법을 이용하여 염분, COD 및 TOC 자료에서 이상자료와 영향자료를 처리한 후 OLS와 Robust 회귀분석의 결과(Table 3)는 경험과 자료 특성을 고려한 Son et al. (2003)의 분석결과와 큰 차이가 없었다. 자료 처리 효과는 회귀분석 시 이상자료와 영향자료에 대한 변동계수의 감소율이 각각 31%와 80%에 이를 정도로 매우 컸다. 따라서 주관적인 판단방법 이외 본 연구의 Case-2 진단방법도 매우 효율적이고 적절하다고 할 수 있다. 특히, 관측자료가 많아질수록 모든 자료를 하나하나 확인하면서 인위적인 수작업으로 이상자료 및 영향자료를 처리하는 것은 한계가 있기 때문에 본 연구의 진단방법과 같은 객관적이고, 자동화된 이상자료 처리기법이 매우 필요하다고 볼 수 있다. 그러나, 기존 연구사례가 없어서 과거 논문결과를 일반화하여 보편적으로 모든 해역에 적용하는데 제한적이며, 무리가 있다고 판단된다.

5. 결론 및 제언

본 연구를 수행한 결과, 얻어진 주요 결론으로 이상자료 및 영향자료 처리기법으로는 S-boxplot 기법과 Cook 계수를 이용한 기법을 조합한 Case-2 진단방법이 적절한 것으로 파악되었으며, 이 방법으로 진단된 이상자료 및 영향자료는 모두 22개(전체자료에서 차지하는 비율 = 5.6%)로, 본 연구에서는 과도한 자료 제거는 발생하지 않았다. 또한, 이 방법을 이용하여 최적 추정된 회귀식은 TOC(mg/L) = 0.44·COD(mg/L) + 1.53이며, RMS 오차는 0.85 mg/L, 결정계수는 0.47이다. 그리고, Case-2 진단방법으로 처리된 이상자료 및 영향자료의 개수가 전체 자료에서 차지하는 비율이 큰 경우에는(대략 5~10% 이상) 과도한 자료 제거를 방지하기 위하여 적절한 처리 및 제거 개수에 대한 제한이 필요할 것으로 판단된다.

감사의 글

본 논문은 2014년 해양수산부의 재원으로 한국해양과학기술진흥원의 지원을 받아 수행된 연구(운용해양(해양예보)시스템 연구 (2단계)-PM57701)입니다. 연구비 지원에 감사드립니다.

References

Aucremanne, L., Brys, G., Hubert, M., Rousseeuw, PJ. and Struyf, A. (2004). A study of belgian inflation, relative prices and nominal rigidities using new robust measures of skewness and tail weight. In: Hubert, M, Pison, G, Struyf, A, Van Aelst, S. (Eds.), Theory and Applications of Recent Robust Methods, Series: Statistics for Industry and Technology. Birkhauser, Basel, pp. 13-25.

Barnett, V. and Lewis, T. (1994). Outliers in Statistical Data, John Wiley & Sons, pp. 320-328.

Chatterjee, S. and Hadi, A.S. (1986). Influential observations, high

leverage points, and outliers in linear regression, Statistical Science, Vol. 1, No. 3, pp. 379-416.

Chen, R.F. and Bada, J.F. (1992). The fluorescence of dissolved organic matter in seawater, Marine Chemistry, Vol. 37, pp. 191-221.

Cho, H.Y. and Oh, J.H., (2012). Outlier Detection of the Coastal Water Temperature Monitoring Data Using the Approximate and Detail Components, Journal of the Korean Society for Marine Environmental Engineering, Vol. 15, No. 2, pp. 156-162.

Cook, R.D. (1977). Detection of Influential Observations in Linear Regression, Technometrics, 19, pp. 15-18.

Doval. M.D. and Hansell, D.A. (2000). Organic carbon and apparent oxygen utilization in the western south and the central Indian Ocean, Marine Chemistry, Vol. 68, pp. 249-264.

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010). Multivariate Data Analysis. Seventh Edition. Chapter 2. pp. 64-70.

Hedger, J.I. (2002). Why dissolved organic matter, In : Biogeochemistry of marine dissolved organic matter, edited by Hansell, D.A. and C.A. Carlson, Academic Press, Amsterdam, pp. 1-33.

Hoaglin, D.C. and Welsch, R.E. (1978). The Hat Matrix in Regression and ANOVA, The American Statistician, Vol. 32, pp. 17-22.

Hubert, M. and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions, Computational Statistics and Data Analysis, Vol. 52, pp. 5186-5201.

Kim, C. and Storer, B.E. (1996). Reference Values for Cook's Distance, Communications in Statistics Simulations and Computations, Vol. 25, pp. 691-708.

Kim, K.H., Son, S.K., Son, J.W. and Ju, S.J. (2006). Methodological comparison of the quantification of total carbon ad organic carbon in marine sediment, Journal of the Korean Society for Marine Environmental Engineering, Vol. 9, pp. 235-242.

Kimber, A.C. (1990). Exploratory data analysis for possibly censored data from skewed distributions, Applied Statistics, Vol. 39, pp. 21-30.

Koenker, R. and Bassett, J.G. (1978). Regression quantile, Econometrica : Journal of the Econometric Society, Vol. 46, No. 1, pp. 33-50.

Koenker, R. and Hallock, K.F. (2001). Quantile regression. Journal of Economic Perspectives, Vol. 15, No. 4, pp. 143-156.

Korea Ocean Research & Development Institute. (2008). Development of management and restoration technologies for estuaries with focus on Han River estuary region, BSPE98101-2028-7, pp. 349-371 (in Korean).

Kottegoda, N.T. and Renzo, R. (1997). Statistics, Probability, and Reliability for Civil and Environmental Engineers, pp. 375-380.

Lee, J.S., Kim, S.Y., Lee, Y.K., Shin, D.W., Kim, H.J. and Jou, H.T. (2001). A Study on Outlier Adjustment for Multibeam Echosounder Data, The SeaJournal of the Korean Society for Marine Environmental Engineering, Vol. 6, No. 1, pp. 35-39.

Lyman, O.R. and Longnecker, M. (2001). An Introduction to Statistical Methods and Data Analysis, pp. 96-101.

Ministry of Land, Transport and Maritime Affairs, Korea Institute of Marine Science & Technology. (2011). Saemangeum coastal system research for marine environmental conservation, Korea Ocean Research & Development Institute, BSPM55630-2269-2, pp. 206-213 (in Korean).

Ministry of Maritime Affairs and Fisheries. (2006). Research on Marine Environmental Improvement of Shihwa Lake, Korea Ocean Research & Development Institute, BSPM38800-1825-4, pp. 158-162 (in Korean).

Ministry of Maritime Affairs and Fisheries. (2013a). Marine Environment Process Test Standard, Notification No. 2013-230 of the Ministry of Maritime Affairs and Fisheries (in Korean).

Ministry of Maritime Affairs and Fisheries. (2013b). Marine Environment Management Act Enforcement Regulations, Act No. 63 of the Ministry of Maritime Affairs and Fisheries (in Korean).

So, B.J., Kwon, H.H. and An, J.H. (2012). Trend Analysis of Extreme Precipitation Using Quantile Regression, Journal of Korea Water Resources Association, Vol. 45, No. 8, pp. 815-826.

Son, J.W., Park, Y.C. and Lee, H.J. (2003). Characteristics of Total Organic Carbon and Chemical Oxygen Demand in the Coastal Waters of Korea. The SeaJournal of the Korean Society of Oceanography, Vol. 8, No. 3, pp. 317-326.

Tchobanoglous, G. and Schroeder, E.D. (1985). Water Quality, pp. 101-104.

원고접수일: 2014년 6월 26일
수정본채택: 2014년 7월 29일(1차)
2014년 8월 4일(2차)
2014년 8월 14일(3차)
게재확정일: 2014년 8월 22일