

Note

이상자료가 연안 환경자료의 통계 척도에 미치는 영향

조흥연<sup>1\*</sup> · 이기섭<sup>2</sup> · 안순모<sup>2</sup>

<sup>1</sup>한국해양과학기술원 연안공학연구본부  
(15627) 경기도 안산시 상록구 해안로 787

<sup>2</sup>부산대학교 자연과학대학 해양학과  
(46241) 부산광역시 금정구 부산대학로 63번길 2

Impact of Outliers on the Statistical Measures of the Environmental Monitoring Data in Busan Coastal Sea

Hong-Yeon Cho<sup>1\*</sup>, Ki-Seop Lee<sup>2</sup>, and Soon-Mo Ahn<sup>2</sup>

<sup>1</sup>*Division of the Coastal and Environmental Engineering Research, KIOST  
Ansan 15627, Korea*

<sup>2</sup>*Division of Earth Environmental System, College of Natural Sciences  
Pusan National University, Busan 46241, Korea*

**Abstract :** The statistical measures of the coastal environmental data are used in a variety of statistical inferences, hypothesis tests, and data-driven modeling. If the measures are biased, then the statistical estimations and models may also be biased and this potential for bias is great when data contain some outliers defined as extraordinary large or small data values. This study aims to suggest more robust statistical measures as alternatives to more commonly used measures and to assess the performance these robust measures through a quantitative evaluation of more typical measures, such as in terms of locations, spreads, and shapes, with regard to environmental monitoring data in the Busan coastal sea. The detection of outliers within the data was carried out on the basis of Rosner's test. About 5–10% of the nutrient data were found to contain outliers based on Rosner's test. After removal (zero-weighting) of the outliers in the data sets, the relative change ratios of the mean and standard deviation between before and after outlier-removal conditions revealed the figures 13 and 33%, respectively. The variation magnitudes of skewness and kurtosis are 1.36 and 8.11 in a decreasing trend, respectively. On the other hand, the change ratios for more robust measures regarding the mean and standard deviation are 3.7–10.5%, and the variation magnitudes of robust skewness and kurtosis are about only 2–4% of the magnitude of the non-robust measures. The robust measures can be regarded as outlier-resistant statistical measures based on the relatively small changes in the scenarios before and after outlier removal conditions.

**Key words :** statistical measures, robust measures, outlier, Rosner's test, Busan coastal sea

## 1. 서 론

통계 모수는 분산분석, 모수의 신뢰구간 추정, 회귀분석, 상관분석, 분포 추정 등 다양한 통계적인 추정·검정에서 필수적인 정보로 요구되고 있다. 통계 모수는 기술통계의 관점에서 다수의 자료를 축약된 소수의 수치로 표현하기 때문에 전체 자료를 대표할 수 있는 정보를 포함하고 있어야 한다. 그러나 이상자료(outliers, abnormal data)가 존재하는 경우에는 통계모수를 왜곡시키고, 더불어 통계적인 추론 및 검정도 왜곡될 수 있다. 따라서 이상 자료에 저항할 수 있는(resistant) 굳건한(robust) 통계모수 추정이 필요하다. 이상 자료는 다양한 원인에 의하여 발생할 수 있으며, 일반적으로 기본적인 통계모형에서 크게 벗어나는 자료 또는 대부분의 자료와는 다른 아주 크거나 작은 관측 자료로 정의된다(Barnett and Lewis 1994). 따라서 이상자료는 인간의 자료관리 오류, 관측 센서의 작동불량 등의 원인으로 발생하는 실제 존재하지 않는 자료일 수도 있고, 정상조건(normal condition)과는 다른 어떤 극한 상황(event, abnormal or extreme condition)에 해당하는 실제 존재하는 자료일 수도 있다. 따라서 실제 존재하지 않는 이상자료는 마땅히 제거되어야 하며, 실제 존재하는 자료는 유지되어야 하지만 그 판단이 곤란한 경우도 발생할 수 있다.

그러나 어떤 경우라 할지라도 이상자료는 기본적인 통계모수 추정에 직접적인 영향을 줄 수 있다. 표본으로부터 추정되는 통계모수에 해당하는 평균, 분산(또는 표준편차), 왜도, 첨도 등은 다양한 통계학적 추론 및 검정, 관측 자료의 분포형태 추정에 이용되기 때문에 이상자료의 영향을 받는 경우, 그 추정결과가 크게 왜곡되는 경우가 발생할 수 있다. 이러한 영향을 저감하기 위한 방법으로 다양한 로버스트(robust) 추정방법이 제시되고 있으며, 이상자료의 영향을 덜 받는 다양한 통계 모수 추정방법이 제안되고 있으나(Huber and Ronchetti 2009), 그 활용수준은 상대적으로 미흡한 수준이다(Erceg-Hurn and Miroseovich 2008).

이상자료는 현장 관측자료의 변동 특성 및 관측방법에 따라 매우 다양하게 발생되기 때문에 이상자료의 발생 양상 및 이러한 이상자료가 자료의 통계모수 추정에 미치는 정량적인 영향분석이 선행되어야 한다. 이러한 분석은 이상자료를 포함하는 현장자료 통계 모형의 적절한 모수를 추정하기 위하여 이상자료를 어떻게 처리할 것인가를 판단하는 기준이 된다. 이상자료에 대한 판단은 이상자료를 제외한 대부분의 정상자료에 대한 통계모형을 구성하고, 구성된 통계모형을 이용하여 이상자료로 간주되는 자료가 어느 정도로 대부분의 정상자료에서 벗어나고 있는가를 객관적이고, 정량적으로 평가하여야 한다. 또한 이러한 평

가자료를 바탕으로 이상자료가 ‘현실적으로 발생가능한 자료인가? 불가능한 자료인가?’를 판단할 수도 있다.

본 연구에서는 연안 환경자료를 이용하여 이상자료를 진단하고, 이러한 이상자료가 기본적인 통계모수 추정에 미치는 정량적인 영향 분석을 수행하였으며, 더불어 로버스트 추정방법으로 제시된 통계모수 추정방법과 비교하여 그 성능평가를 수행하였다.

## 2. 연안 환경자료 및 방법

### 부산연안 정점자료(2000–2014년, 연4회 측정자료)

부산연안 환경 모니터링 자료(국가해양환경정보통합시스템 2016; 해양환경관리공단 2016)는 해양환경관리공단에서 연 4회 정기적으로 측정·제공되고 있다. 측정 정점은 1997년 4개 정점을 시작으로, 2004년부터 2개 정점이 추가되고, 2013년부터 11개 정점이 추가되어 모두 17개이며, 각각 표층과 저층 자료를 제시하고 있다(Fig. 1 참고). 측정항목은 수온, 염분, pH, 용존산소(Dissolved Oxygen, DO), 화학적산소요구량(Chemical Oxygen Demand, COD), 질산성질소, 아질산성질소, 암모니아성질소, 용존무기질소, 총질소, 용존무기인(인산염인), 총인, 규산규소, 부유물질, 엽록소 농도이다. DO, COD, 부유물질 농도자료의 단위는 mg/L이며, 영양염류 및 엽록소 농도는 모두  $\mu\text{g/L}$ 이다. 관측항목에 따라 관측 기간이 차이를 보이고 있다. 정점에서의 자료 결측이 없는 경우, 하나의 환경인자 자료개수는 표층-저층 각각 464개(계산 : 4개 정점 18년 $\times$ 4회/연 $\times$ 4개 지점 = 288개, 2개 정점 11년 $\times$ 4회/연 $\times$ 2개 지점 = 88개, 11개 정점 2년 $\times$ 4회/연 $\times$ 11개 지점 = 88개)로 모두 464개 $\times$ 2 = 928개가 된다. 연안환경 관측자료는 연 4회 측정하는 자료로 연간 변화의 비교분석은 자료의 개수제한으로 한계가 있으며, 계절변화가 크기 때문에 계절에 따른 자료구분이 적절하다. 따라서 본 연구에서는 연 4회 측정 자료가 각각의 계절을 대표하는 자료로 간주하여 계절 자료에 대한 통계 모수를 추정하였다. 이 경우 각각의 계절에 대한 통계모수 추정에 사용되는 자료의 개수는 표층, 저층 각각 116개로 모두 232개이다.

관측 정점의 위치는 다음과 같다(Fig. 1 참조).

### 이상자료 진단

이상자료는 대부분의 자료와는 다른 아주 크거나 작은 자료를 의미하며, 통계 모수 추정에 큰 영향을 미치기 때문에 자료 대부분의 변화양상 및 분포 특성을 나타내는 통계모수 추정을 위해서는 이상자료 진단-제외 과정이 필요하다. 가장 일반적인 이상자료 진단방법은 Box-plot 방법이 제안되고 있으나, 편중된(skewed) 분포를 가지는 자료의 경우 진단에 한계가 있다. 또한 특이한 1-2개 자료

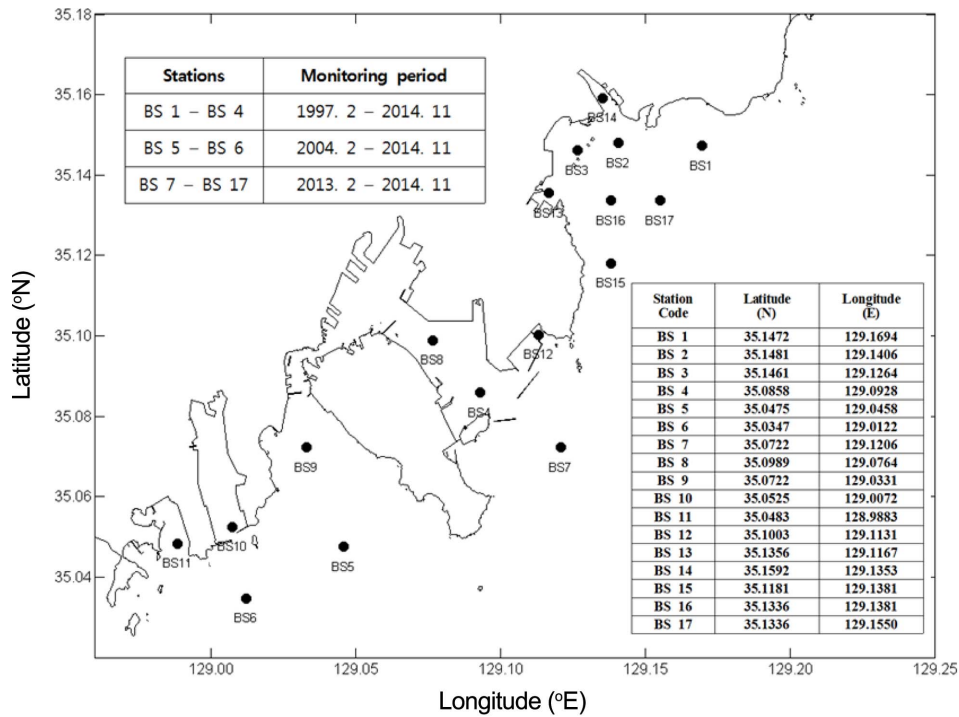


Fig. 1. Monitoring stations of the environmental data in Busan coastal sea

의 이상자료 여부만을 판단하는 Grubbs 또는 Dixon 검정 방법(Dixon 1951; Grubbs 1950) 등 다양한 방법이 제시되고 있으나, 실질적으로는 이상자료의 개수를 모를 때 때문에 본 연구에서는 전체 자료에서 이상자료(outliers)와 대부분의 정상자료(inliers, normal data)를 구분하는 개념으로 다수의 이상자료를 진단하는 Rosner 검정 방법(Rosner 1983)을 선정하였으며, 가능한 이상자료의 개수는 전체 자료의 1/2 이하로 간주하였다. Rosner 검정 방법은 이상자료의 개수를 지정하지 않고, 이상자료 개수의 상한의 범위를 지정하여 이 범위에서 이상자료 여부를 판단한다. 따라서 관측 자료를 이상자료와 정상자료로 양분하는 개념이기 때문에 이상자료 개수의 가능한 상한은 50% 미만( $< 50\% = 1/2$ )이 된다. 그러나 실질적으로는 전체 자료에서 이상자료가 차지하는 비중은 미미한 수준이다.

Rosner 방법은 어떤 자료가 정규분포를 따른다는 가정을 하고, 이상자료가 대부분의 자료의 중심에서 얼마나 벗어날지를 설정한 유의수준 조건에서  $t$ -분포를 이용하여 추정·판단하는 방법으로 세부 과정은 다음과 같다. Rosner 방법을 이용한 환경자료의 이상자료 검정은 R ‘EnvStats’ 패키지의 ‘rosnerTest’ 함수를 이용하여 간단하게 추정할 수도 있다(Millard 2013; R Core Team 2015).

#### 입력조건:

이상자료를 포함한 자료( $x_i, i = 1, 2, \dots, n$ ),

유의수준( $\alpha = 0.05$ ),

이상 자료 진단에 필요한 자료의 최대개수 비율( $r = n/2$ )

제1단계:  $n$ 개의 입력 자료를 이용하여 절대편차를 각각 계산하고, 그 편차 중에서 최대 편차(extreme Studentized deviate)를  $R_1$  통계정보로 지정한다.

$$R_1 = \max(|(x_i - \bar{x})/\sigma_x|, i = 1, 2, \dots, n)$$

제2단계: 제1단계에서 최대편차를 생성하는 자료를 제거하고, 나머지 자료 ( $n - 1$ )개를 이용하여 다시 절대편차를 계산하고, 그 편차 중에서 최대편차를  $R_2$  통계정보로 저장한다. 이 과정을  $r$ 번 반복하여 모두  $r$ 개의 통계정보(test statistics)  $R_i (i = 1, 2, \dots, r)$ 를 생성한다(참고:  $R_r$  최대편차 추정에 사용한 자료의 개수는  $n - r + 1$ ).

제3단계:  $t$ -분포를 이용하여 유의수준에 대한 한계 통계정보( $\lambda_i, i = 1, 2, \dots, r$ )를 계산한다.

$$\lambda_i = \frac{(n-i) \cdot t_{p, n-i-1}}{\sqrt{(n-i-1 + t_{p, n-i-1}^2)(n-i+1)}}$$

여기서,  $t_{p, df}$  = 자유도  $df$  조건, 초과확률  $100p$ 에 해당하며,  $p = 1 - [\alpha/(n-i+1)]$ 이다.

제4단계 (판단):  $R_i > \lambda_i$  조건을 만족하는 최대  $i (\leq r)$

가 이상 자료로 진단된 개수이며,  $R_i$  계산에서 단계적으로 제거된 모든 자료를 이상 자료로 진단한다.

**출력조건:**

이상자료의 개수( $n_o$ ),  
 이상자료로 진단된 자료( $[x_o], j = 1, 2, \dots, n_o$ )

**기본 통계모수 추정 방법**

본 연구에서 이상자료의 정량적인 영향분석을 평가하기 위하여 선정한 기본적인 통계모수는 평균, 표준편차, 왜도 및 첨도이다. 또한 기본적인 통계모수에 대한 로버스트 통계 모수로 제시되는 변수는 다음과 같이 분류할 수 있다 (Table 1). 본 연구에서는 이상자료 제외 전·후의 통계모수 변화 정도를 파악하기 위한 통계모수로 평균, Median, 표준편차, IQR, MAD, 전통적인 왜도 및 첨도와 로버스트 추정기법에 근거하는 팔분위수(octile) 왜도와 팔분위수 첨도를 선정하였으며(Brys et al. 2004; Bonato 2011; Kim and White 2004; Moors 1988; Rousseeuw and Croux 1993), 각각의 통계모수에 대한 추정 공식은 다음과 같다.

연안 환경자료를  $x$  ( $i = 1, 2, \dots, n =$  자료의 개수)라고 하면,  $Q_1, Q_2, Q_3 =$  각각 제1사분위수, 제2사분위수 (median), 제3사분위수이고,  $E_1, E_2, \dots, E_7 =$  각각 제1팔분위수, 제2팔분위수, ..., 제7팔분위수이며,  $E_2, E_4, E_6$  분위수는 각각 제1,2,3 사분위수에 해당한다. 사분위수는 오름차순으로 정렬된 자료(최소 = 0, 최대 = 1)에서 1/4, 2/4, 3/4 위치에 해당하는 자료이며, 팔분위수는 1/8, 2/8, ..., 7/8 위치에 해당하는 자료이며, 해당위치에 자료가 없는 경우, 인접 자료를 산출 평균한다.

평균(mean),

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

표준편차(SD, standard deviation),

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Median, median( $x_i$ ) =  $Q_2$ ,

IQR(inter-quartile range), IQR( $x_i$ ) =  $Q_3 - Q_1$

MAD(median absolute deviation),

$$MAD(x_i) = \text{median}(|x_i - \text{median}(x_i)|)$$

왜도(coefficient of skewness),

$$\gamma_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 / s_x^3$$

첨도(kurtosis),

$$\delta_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 / s_x^4 - 3.0$$

로버스트 사분위수 왜도(robust quartile skewness),

$$QSK_x = \frac{Q_3 + Q_1 - 2Q_2}{IQR}$$

Pearson 왜도계수(Pearson coefficient of skewness),

$$PSK_x = \frac{\bar{x} - Q_2}{s_x}$$

**Table 1. Summary on the measures of the statistical parameters**

Measures	Classical parameters and/or methods (Non-robust estimation)	Robust parameters. Robust estimation
Central tendency (location)	Mean	Trimmed mean, Windsorized mean, Median, mode
Variation, dispersion (spreading)	Standard deviation (or variance), range	IQR (Inter-Quartile Range) MAD (Median Absolute Deviation) Coefficient of variation, Trimmed-range, Windsorized variance
Shape of distribution	Asymmetry Skewness (coefficient of skewness) (normalized third central moment)	Pearson coefficient of skewness, MC (Med-Couple) Quartile-skewness Octile-skewness
	Peakedness Kurtosis (normalized fourth central moment)	Octile-kurtosis

로버스트 팔분위수 왜도(robust octile skewness),

$$OSK_x = \frac{(E_7 - E_4) - (E_4 - E_1)}{E_7 - E_1}$$

로버스트 팔분위수 첨도(robust octile kurtosis),

$$KR_x = \frac{(E_7 - E_5) + (E_3 - E_1)}{E_6 - E_2} - 1.23$$

이상자료 제외 전·후에 대한 통계모수 변화 정도는 환경자료의 단위와 같은 평균, 표준편차는 다음과 같은 상대 변화비율(Relative Change, RC)로 제시하였으며, 무차원 모수에 해당하는 왜도 및 첨도는 절대적인 차이(Absolute Change, AC)로 제시하였다.

$$RC = \frac{(m_{BEO} - m_{AEO})}{m_{BEO}} \times 100\%$$

$$AC = m_{BEO} - m_{AEO}$$

여기서,  $m$  = 자료의 기본 통계 척도(모수)이며, BEO, AEO = 각각 이상자료 제외 전·후 조건(before and after

excluding outliers' conditions)을 의미한다.

### 3. 결과 및 토론

#### 환경자료의 이상자료 진단

표층과 저층, 그리고 계절로 구분된 부산 연안의 환경 자료를 Rosner 방법을 이용하여 이상자료를 진단하였다. 본 연구는 표층과 저층의 환경인자 차이 분석보다는 이상자료 진단에 중점을 두고 있기 때문에, 이상자료 진단 결과는 표층 결과만을 대상으로 이상자료의 개수 및 전체 자료에서 차지하는 비율을 제시하였다(Table 2 참조). 환경인자에서 이상자료로 진단된 자료가 차지하는 평균 비율은 COD 항목이 0.2%로 가장 낮았으며, NH<sub>4</sub>-N 항목이 12.3%로 가장 높았다. 질소 및 인 계열 항목에서 이상 자료가 차지하는 비율은 6-12% 정도로 다른 항목에 비하여 높은 비율을 보였으며, 염분, SS, SiO<sub>2</sub>, 엽록소 농도도 2-5% 정도로 다소 높은 비율을 보이고 있다. 반면 수온, pH, DO, COD 항목의 경우 이상자료로 진단되는 자료의 비율은 1.5% 또는 그 이하로 매우 낮은 것으로 파악되었다. 이러한 진단결과는 어떤 특정 계절의 낮은 빈도의 조사가

**Table 2. Numbers of total data (samples) and detected outliers and the ratio(%) of the number of outliers to the number of total samples for each water quality parameters of the coastal environmental data**

	Number of the samples														
	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl-a
Spring	116	116	116	116	116	116	116	116	116	100	116	100	88	116	104
Summer	116	116	116	116	116	116	116	116	116	100	116	100	88	116	104
Fall	116	116	116	116	116	116	116	116	116	96	116	96	88	116	104
Winter	116	116	116	116	116	116	116	116	116	100	116	100	88	116	104
Sum.	464	464	464	464	464	464	464	464	464	396	464	396	352	464	416
	Number of samples detected by rosner test														
	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl-a
Spring	2	13	5	0	0	12	6	10	14	12	13	6	1	2	0
Summer	0	6	1	0	1	21	8	9	9	8	22	11	2	1	10
Fall	0	2	0	3	0	11	11	8	8	17	10	2	1	2	2
Winter	0	1	1	1	0	13	5	9	11	8	10	11	8	3	6
Sum.	2	22	7	4	1	57	30	36	42	45	55	30	12	8	18
	Ratio (%) of no. of outliers to no. of samples														
	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl-a
Spring	1.7	11.2	4.3	0.0	0.0	10.3	5.2	8.6	12.1	12.0	11.2	6.0	1.1	1.7	0.0
Summer	0.0	5.2	0.9	0.0	0.9	18.1	6.9	7.8	7.8	8.0	19.0	11.0	2.3	0.9	9.6
Fall	0.0	1.7	0.0	2.6	0.0	9.5	9.5	6.9	6.9	17.7	8.6	2.1	1.1	1.7	1.9
Winter	0.0	0.9	0.9	0.9	0.0	11.2	4.3	7.8	9.5	8.0	8.6	11.0	9.1	2.6	5.8
Mean.	0.4	4.7	1.5	0.9	0.2	12.3	6.5	7.8	9.1	11.4	11.9	7.5	3.4	1.7	4.3

Ref. SD = Standard Deviation, MAD = Median Absolute Deviation, IQR = Inter-Quartile Range, R-Skewness and R-kurtosis = octile skewness and kurtosis, respectively. WT = Water Temperature (°C), S = Salinity (psu), DO = Dissolved Oxygen (mg/L), COD = Chemical Oxygen Demand (mg/L). NH<sub>4</sub>, NO<sub>2</sub>, NO<sub>3</sub>, DIN, TN, DIP, and TP = ammonium, nitrite, nitrate, dissolved inorganic, and total nitrogen and dissolved inorganic and total phosphorus (µg/L). SS = Suspended Solids (mg/L). chl-a = chlorophyll-a (mg/m<sup>3</sup>)

일시적인 환경변동의 영향을 적절하게 감지하지 못하기  
때문으로 판단된다.

한편 계절에 따른 이상자료 진단비율은 표층, 저층 각  
각 평균 5.5%, 3.3%로 표층이 보다 높게 나타났다. 이는  
관측빈도와 더불어 표층의 환경변화가 저층의 환경변화보

다 크기 때문에 판단된다.

**이상자료 제외 전·후의 통계모수 추정**

이상자료로 진단된 자료를 제외한 경우와 포함한 경우  
의 다양한 기본 통계정보를 계산하여 비교·분석하였다. 이

**Table 3. Statistical measures and ratios of their changes between the conditions before and after excluding outliers**

**(a) Statistical measures before excluding outliers**

Spring	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl-a
Mean	15.06	33.31	8.12	9.02	1.23	81.76	11.02	110.26	203.43	447.34	19.00	35.24	226.15	6.84	4.52
Median	14.97	33.78	8.15	9.11	1.10	30.87	7.58	63.50	110.50	367.30	11.59	26.19	211.60	5.41	3.98
SD	1.04	1.27	0.17	0.86	0.63	150.13	17.44	134.90	271.19	278.81	26.27	25.33	145.55	4.19	2.85
MAD	0.87	0.44	0.13	0.94	0.49	42.28	4.78	75.93	119.28	151.23	9.58	10.93	112.47	2.82	2.66
IQR	1.17	0.79	0.18	1.21	0.70	67.15	7.12	114.02	174.80	209.30	13.83	21.90	152.68	4.01	3.66
Skewness	0.67	-2.72	-1.75	0.20	1.05	3.37	6.45	2.15	2.57	2.01	3.30	2.81	1.87	1.84	1.03
R. Skewness	0.06	-0.43	-0.01	-0.25	0.15	0.26	0.25	0.28	0.25	0.20	0.11	0.43	-0.02	0.32	0.12
Kurtosis	1.62	7.98	4.13	-0.08	1.08	11.81	51.79	4.73	6.82	4.06	12.48	10.74	7.54	4.35	0.82
R. Kurtosis	0.20	0.29	-0.09	-0.14	0.37	0.62	0.38	0.23	0.33	0.39	0.36	0.00	0.29	0.21	0.12
Summer	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl-a
Mean	22.20	31.00	8.16	8.19	1.56	80.54	10.77	181.93	273.25	411.44	25.09	38.23	434.10	5.62	6.95
Median	21.98	31.72	8.17	8.06	1.50	21.80	4.84	85.20	114.45	308.64	9.32	27.17	234.86	4.91	5.28
SD	2.56	3.36	0.16	1.63	0.90	158.06	15.84	273.63	413.06	381.82	37.43	35.97	533.93	3.99	6.41
MAD	2.53	1.12	0.11	1.20	0.83	29.19	5.69	118.16	151.27	140.68	10.87	17.15	231.43	2.30	3.74
IQR	3.43	1.49	0.15	1.60	1.09	74.22	12.61	186.77	286.06	206.35	20.72	22.67	416.99	3.20	5.01
Skewness	0.15	-3.94	-0.90	-0.24	0.98	4.47	2.73	2.76	3.19	3.36	2.15	2.44	3.32	4.68	2.06
R. Skewness	0.19	-0.26	0.02	0.08	-0.10	0.63	0.45	0.34	0.52	0.15	0.39	0.09	0.44	0.12	0.00
Kurtosis	-0.47	17.88	1.74	1.37	1.46	25.73	8.03	8.11	12.60	14.05	3.92	6.34	16.09	34.88	4.71
R. Kurtosis	0.10	0.40	0.55	0.34	0.04	0.56	0.18	0.09	-0.10	0.82	1.16	0.62	0.64	0.22	0.43
Fall	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl-a
Mean	17.91	32.83	8.08	7.36	1.18	50.52	12.16	135.18	197.84	373.59	24.15	33.97	377.46	5.49	1.78
Median	18.05	32.94	8.10	7.14	0.95	25.23	9.85	100.28	145.07	283.87	18.76	28.41	365.51	4.60	1.53
SD	1.69	1.28	0.13	0.81	0.63	77.38	8.25	130.96	187.52	253.74	19.43	20.60	139.68	3.76	1.26
MAD	1.62	0.41	0.12	0.47	0.45	26.12	4.22	45.76	72.08	94.10	11.31	13.31	105.79	2.37	0.84
IQR	2.29	0.53	0.16	0.78	0.68	37.28	6.03	73.08	97.83	145.67	18.02	21.12	141.50	3.44	1.21
Skewness	0.08	-9.25	-0.24	1.65	1.18	3.75	2.28	3.16	3.25	2.30	2.45	2.47	1.25	2.82	2.77
R. Skewness	-0.14	0.12	-0.25	0.40	0.31	0.36	0.32	0.26	0.01	0.28	0.36	0.27	-0.01	0.14	0.14
Kurtosis	-0.34	94.29	-0.08	4.78	0.49	16.97	6.17	11.37	12.92	5.83	7.97	8.54	3.75	12.10	12.84
R. Kurtosis	0.03	0.11	0.10	0.24	0.53	0.25	0.45	0.39	0.69	1.54	0.20	-0.02	0.03	0.16	0.11
Winter	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl-a
Mean	11.70	33.90	8.08	8.80	1.14	105.22	13.38	130.38	248.95	485.13	30.32	35.06	271.77	7.56	1.14
Median	11.70	33.96	8.11	8.77	1.07	37.11	9.00	100.25	149.88	325.23	18.36	29.00	271.44	5.68	0.90
SD	0.79	0.48	0.13	1.09	0.62	211.97	15.54	109.81	299.18	601.76	42.64	26.55	111.67	5.24	0.93
MAD	0.79	0.36	0.13	0.72	0.69	39.22	4.45	37.19	80.57	158.93	7.86	15.93	66.60	3.02	0.38
IQR	1.05	0.46	0.17	1.08	0.93	65.67	9.41	59.53	151.53	251.62	14.50	22.59	84.35	5.17	0.49
Skewness	0.24	-2.45	-1.15	-0.04	0.38	3.77	4.81	3.54	4.08	4.86	4.37	3.15	0.53	1.83	2.32
R. Skewness	0.00	0.05	0.00	0.25	0.14	0.38	0.46	0.22	0.43	0.34	0.41	0.07	0.30	0.38	-0.06
Kurtosis	-0.43	11.23	2.61	1.95	-0.47	15.56	27.83	15.20	20.94	28.08	19.91	14.23	5.46	3.90	5.99
R. Kurtosis	-0.13	0.38	-0.30	0.30	0.03	0.52	-0.27	0.48	0.39	-0.15	0.14	0.17	0.41	0.31	1.54

**Table 3. Continued**  
**(b) Statistical measures after excluding outliers**

Spring	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl-a
Mean	15.00	33.68	8.15	9.02	1.23	39.32	8.03	75.74	116.29	359.16	11.25	30.27	217.42	6.54	4.52
Median	14.96	33.83	8.15	9.11	1.10	25.00	7.15	53.05	88.65	337.55	10.36	26.00	210.00	5.40	3.98
SD	0.93	0.48	0.12	0.86	0.63	42.01	5.52	70.12	92.19	124.85	7.91	14.33	121.01	3.52	2.85
MAD	0.87	0.34	0.13	0.94	0.49	33.58	4.23	60.44	95.91	130.17	8.36	10.38	112.68	2.82	2.66
IQR	1.15	0.56	0.16	1.21	0.70	56.39	5.49	105.73	141.71	179.09	10.82	18.70	152.06	3.89	3.66
Skewness	0.09	-1.05	-0.84	0.20	1.05	1.34	0.97	1.12	0.75	0.58	0.71	1.26	0.56	1.21	1.03
R. Skewness	0.05	-0.36	0.16	-0.25	0.15	0.25	0.07	0.39	0.32	0.23	0.04	0.34	-0.04	0.29	0.12
Kurtosis	0.05	0.45	1.20	-0.08	1.08	1.20	1.08	0.68	-0.15	-0.25	0.02	1.18	0.21	0.89	0.82
R. Kurtosis	0.17	0.11	0.08	-0.14	0.37	0.07	0.52	-0.19	-0.02	-0.01	0.11	0.00	0.14	0.18	0.12
Summer	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl-a
Mean	22.20	31.68	8.17	8.19	1.53	28.22	7.12	112.89	172.44	318.88	9.38	27.29	377.84	5.34	5.18
Median	21.98	31.81	8.17	8.06	1.50	18.11	4.21	72.32	105.00	283.00	6.54	24.00	231.97	4.90	5.03
SD	2.56	1.25	0.15	1.63	0.85	28.95	7.50	116.27	171.95	155.98	8.85	14.24	365.31	2.66	3.22
MAD	2.53	1.09	0.11	1.20	0.83	19.43	4.76	101.38	136.49	113.12	7.03	15.11	213.76	2.30	3.36
IQR	3.43	1.42	0.15	1.60	1.08	38.30	9.13	157.47	221.05	165.87	11.10	20.22	392.87	3.15	4.70
Skewness	0.15	-0.56	-0.63	-0.24	0.72	1.24	1.26	1.21	1.20	1.04	1.20	0.95	1.44	0.92	0.71
R. Skewness	0.19	-0.28	0.06	0.08	-0.10	0.44	0.35	0.33	0.46	0.18	0.34	0.20	0.41	0.11	-0.13
Kurtosis	-0.47	1.20	0.90	1.37	0.47	0.56	0.57	0.92	0.73	0.53	0.75	0.87	1.25	0.99	0.34
R. Kurtosis	0.10	0.18	0.43	0.34	0.02	0.08	0.22	-0.08	-0.03	0.34	0.07	-0.19	0.47	0.15	-0.07
Fall	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl-a
Mean	17.91	32.95	8.08	7.29	1.18	29.84	9.93	104.66	154.34	273.30	19.36	31.95	370.51	5.16	1.67
Median	18.05	32.95	8.10	7.12	0.95	22.12	9.21	97.69	136.11	266.73	17.93	28.22	363.50	4.55	1.52
SD	1.69	0.38	0.13	0.64	0.63	23.35	3.99	53.80	79.05	78.15	10.05	15.29	124.25	2.79	0.94
MAD	1.62	0.39	0.12	0.45	0.45	22.91	3.28	41.64	63.16	70.04	8.96	12.94	104.52	2.26	0.81
IQR	2.29	0.52	0.16	0.70	0.68	31.95	4.78	57.67	89.15	100.54	11.98	20.21	137.34	3.30	1.19
Skewness	0.08	-0.15	-0.24	0.71	1.18	0.94	0.48	0.86	1.24	0.51	0.63	1.20	0.60	1.27	1.07
R. Skewness	-0.14	0.10	-0.25	0.37	0.31	0.37	0.35	0.10	0.09	0.10	0.01	0.22	-0.01	0.12	0.12
Kurtosis	-0.34	-0.38	-0.08	1.22	0.49	0.27	0.21	0.98	1.50	0.39	-0.10	1.10	1.11	1.61	1.04
R. Kurtosis	0.03	0.04	0.10	0.18	0.53	-0.06	0.03	0.24	0.13	0.15	0.35	-0.03	-0.02	0.14	0.04
Winter	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl-a
Mean	11.70	33.92	8.09	8.77	1.14	44.59	10.57	103.81	170.21	344.34	19.86	27.76	281.30	7.05	0.95
Median	11.70	33.96	8.11	8.76	1.07	32.40	8.70	98.41	139.36	307.65	17.90	24.62	274.55	5.50	0.87
SD	0.79	0.39	0.12	1.02	0.62	39.52	5.50	41.50	89.66	138.93	8.42	11.68	61.84	4.22	0.53
MAD	0.79	0.36	0.13	0.71	0.69	31.28	4.00	32.93	68.30	131.80	6.41	11.31	57.96	2.67	0.34
IQR	1.05	0.45	0.17	1.08	0.93	52.85	7.25	42.24	110.31	198.90	9.86	18.00	79.78	4.63	0.44
Skewness	0.24	-1.05	-0.65	-0.50	0.38	1.22	1.20	1.02	1.22	0.83	1.03	0.70	0.19	1.22	1.05
R. Skewness	0.00	0.07	-0.06	0.25	0.14	0.31	0.37	-0.04	0.36	0.28	0.18	0.30	0.30	0.37	-0.04
Kurtosis	-0.43	1.27	0.28	1.03	-0.47	0.87	0.80	1.09	0.78	-0.07	0.36	-0.33	-0.02	0.82	1.43
R. Kurtosis	-0.13	0.36	-0.36	0.27	0.03	0.04	-0.10	0.45	0.03	-0.09	0.44	-0.24	0.26	0.35	0.92

상자료 제외 전·후의 통계량 변화는 평균 및 표준편차로 대표되는 위치(location, central tendency) 및 퍼짐(spreading) 척도는 환경자료와 단위가 같기 때문에 상대 변화비율로 무차원화하여 비교하였으며, 무차원 분포형태 척도에 해당하는 왜도 및 첨도는 이상자료 제외 전·후의 변화정도

를 절대 변화정도로 직접 비교하였다. 이상자료 제외 전·후의 모수변화는 표층의 연안환경자료만을 이용한 결과만을 제시하였다(Table 3 참조, Figs. 2-4 참조).

이상자료 제외 전·후의 통계정보 변화비율은 평균, 표준편차의 경우 각각 평균 12.6%, 32.9% 정도로 파악되었

Table 3. Continued

## (c) Ratio of changes between before and after excluding outliers' conditions

Spring	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl.a
Mean	0.4	-1.1	-0.3	0.0	0.0	51.9	27.1	31.3	42.8	19.7	40.8	14.1	3.9	4.4	0.0
Median	0.0	-0.2	0.0	0.0	0.0	19.0	5.7	16.5	19.8	8.1	10.6	0.7	0.8	0.1	0.0
SD	10.6	62.5	27.3	0.0	0.0	72.0	68.3	48.0	66.0	55.2	69.9	43.4	16.9	15.8	0.0
MAD	0.0	22.6	4.4	0.0	0.0	20.6	11.4	20.4	19.6	13.9	12.8	5.0	-0.2	0.0	0.0
IQR	1.3	28.4	12.8	0.0	0.0	16.0	22.9	7.3	18.9	14.4	21.8	14.6	0.4	3.1	0.0
Skewness	0.6	-1.7	-0.9	0.0	0.0	2.0	5.5	1.0	1.8	1.4	2.6	1.5	1.3	0.6	0.0
R. Skewness	0.0	-0.1	-0.2	0.0	0.0	0.0	0.2	-0.1	-0.1	0.0	0.1	0.1	0.0	0.0	0.0
Kurtosis	1.6	7.5	2.9	0.0	0.0	10.6	50.7	4.1	7.0	4.3	12.5	9.6	7.3	3.5	0.0
R. Kurtosis	0.0	0.2	-0.2	0.0	0.0	0.6	-0.1	0.4	0.3	0.4	0.3	0.0	0.1	0.0	0.0
Summer	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl.a
Mean	0.0	-2.2	-0.1	0.0	1.9	65.0	33.8	37.9	36.9	22.5	62.6	28.6	13.0	4.9	25.4
Median	0.0	-0.3	0.0	0.0	0.4	16.9	13.0	15.1	8.3	8.3	29.8	11.7	1.2	0.1	4.8
SD	0.0	62.9	5.9	0.0	6.1	81.7	52.6	57.5	58.4	59.1	76.4	60.4	31.6	33.3	49.8
MAD	0.0	3.1	2.7	0.0	0.3	33.4	16.4	14.2	9.8	19.6	35.3	11.9	7.6	0.0	10.1
IQR	0.0	5.0	-1.7	0.0	0.9	48.4	27.6	15.7	22.7	19.6	46.4	10.8	5.8	1.5	6.1
Skewness	0.0	-3.4	-0.3	0.0	0.3	3.2	1.5	1.5	2.0	2.3	1.0	1.5	1.9	3.8	1.4
R. Skewness	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.1	0.0	0.1	-0.1	0.0	0.0	0.1
Kurtosis	0.0	16.7	0.8	0.0	1.0	25.2	7.5	7.2	11.9	13.5	3.2	5.5	14.8	33.9	4.4
R. Kurtosis	0.0	0.2	0.1	0.0	0.0	0.5	0.0	0.2	-0.1	0.5	1.1	0.8	0.2	0.1	0.5
Fall	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl.a
Mean	0.0	-0.4	0.0	1.1	0.0	40.9	18.3	22.6	22.0	26.8	19.8	6.0	1.8	6.0	6.3
Median	0.0	0.0	0.0	0.2	0.0	12.3	6.5	2.6	6.2	6.0	4.5	0.7	0.6	1.1	0.5
SD	0.0	70.6	0.0	20.2	0.0	69.8	51.6	58.9	57.8	69.2	48.3	25.8	11.0	25.8	25.3
MAD	0.0	5.3	0.0	5.9	0.0	12.3	22.3	9.0	12.4	25.6	20.8	2.8	1.2	4.7	3.1
IQR	0.0	2.1	0.0	9.7	0.0	14.3	20.7	21.1	8.9	31.0	33.5	4.3	2.9	4.1	1.7
Skewness	0.0	-9.1	0.0	0.9	0.0	2.8	1.8	2.3	2.0	1.8	1.8	1.3	0.7	1.6	1.7
R. Skewness	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	-0.1	0.2	0.3	0.0	0.0	0.0	0.0
Kurtosis	0.0	94.7	0.0	3.6	0.0	16.7	6.0	10.4	11.4	5.4	8.1	7.4	2.6	10.5	11.8
R. Kurtosis	0.0	0.1	0.0	0.1	0.0	0.3	0.4	0.2	0.6	1.4	-0.2	0.0	0.1	0.0	0.1
Winter	WT	S	pH	DO	COD	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	DIN	TN	DIP	TP	SiO <sub>2</sub>	SS	chl.a
Mean	0.0	-0.1	-0.1	0.4	0.0	57.6	21.0	20.4	31.6	29.0	34.5	20.8	-3.5	6.8	16.4
Median	0.0	0.0	-0.1	0.1	0.0	12.7	3.3	1.8	7.0	5.4	2.5	15.1	-1.1	3.2	3.3
SD	0.0	17.4	8.5	5.9	0.0	81.4	64.6	62.2	70.0	76.9	80.3	56.0	44.6	19.6	42.8
MAD	0.0	0.0	-5.9	1.0	0.0	20.2	10.0	11.5	15.2	17.1	18.4	29.0	13.0	11.8	10.0
IQR	0.0	1.6	0.0	0.2	0.0	19.5	22.9	29.0	27.2	21.0	32.0	20.3	5.4	10.5	9.5
Skewness	0.0	-1.4	-0.5	0.5	0.0	2.6	3.6	2.5	2.9	4.0	3.3	2.4	0.3	0.6	1.3
R. Skewness	0.0	0.0	0.1	0.0	0.0	0.1	0.1	0.3	0.1	0.1	0.2	-0.2	0.0	0.0	0.0
Kurtosis	0.0	10.0	2.3	0.9	0.0	14.7	27.0	14.1	20.2	28.2	19.5	14.6	5.5	3.1	4.6
R. Kurtosis	0.0	0.0	0.1	0.0	0.0	0.5	-0.2	0.0	0.4	-0.1	-0.3	0.4	0.2	0.0	0.6

Ref. SD = Standard Deviation, MAD = Median Absolute Deviation, IQR = Inter-Quartile Range, R-Skewness and R-kurtosis = octile skewness and kurtosis, respectively. WT = Water Temperature (°C), S = Salinity (psu), DO = Dissolved Oxygen (mg/L), COD = Chemical Oxygen Demand (mg/L). NH<sub>4</sub>, NO<sub>2</sub>, NO<sub>3</sub>, DIN, TN, DIP, and TP = ammonium, nitrite, nitrate, dissolved inorganic, and total nitrogen and dissolved inorganic and total phosphorus (µg/L). SS = Suspended Solids (mg/L). chl.a = chlorophyll-a (mg/m<sup>3</sup>)

다. 반면 로버스트 추정 통계정보에 해당하는 Median, MAD, IQR 척도의 경우에는 각각 평균 3.7%, 8.5%,

10.5% 정도로 일반적으로 이용되는 통계정보에 비하여 그 변화비율이 매우 낮아 로버스트 통계정보를 적절하게 추



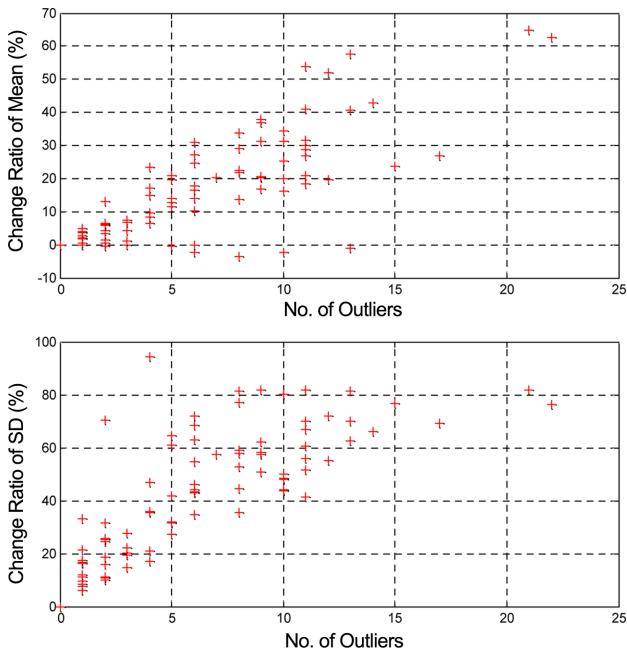


Fig. 2. Scatterplots between the change ratio of the mean and standard deviation (SD) and the number of outliers

정·제시하는 것으로 판단된다. 왜도와 첨도의 경우에도 각각 절대평균이 1.36, 8.11, 로버스트 왜도 및 첨도의 경우 각각 절대평균이 0.056, 0.17 정도로 그 변화정도가 일반적인 왜도 및 첨도 추정수치의 2~4% 정도로 매우 작게 나타나고 있는 것으로 파악되어 로버스트 통계정보를 적절하게 제시하는 것으로 판단된다. 또한 Fig. 2에서 볼 수 있는 바와 같이 이상자료의 개수가 증가할수록 일반적인 통계척도의 경우 그 변화정도가 증가하는 양상을 보이고 있음을 알 수 있다.

이상자료 제외 전·후의 평균과 이상자료에 둔감한 통계인자로 제시되는 Median 변화를 비교하면, Median 변화는 평균 변화의 약 30% 정도에 해당하고 있다. 더불어 표준편차와 이상자료에 둔감한 편차 통계인자에 해당하는 MAD, IQR 변화는 각각 37%, 33% 정도에 해당하고 있다. 변화정도는 회귀곡선의 경사를 변화에 대한 상대변화로 간주하여 추정하였다. 통계척도의 변화비율자료에 대한 변화정도 곡선추정과정에서도 이상자료의 영향을 저감하기 위하여 표층 환경자료에 대한 통계척도의 변화비율 추정함수에 대하여 로버스트 회귀분석도 더불어 수행하였다(Fig. 3 참조). 선형함수의 완만한 경사 및 로버스트

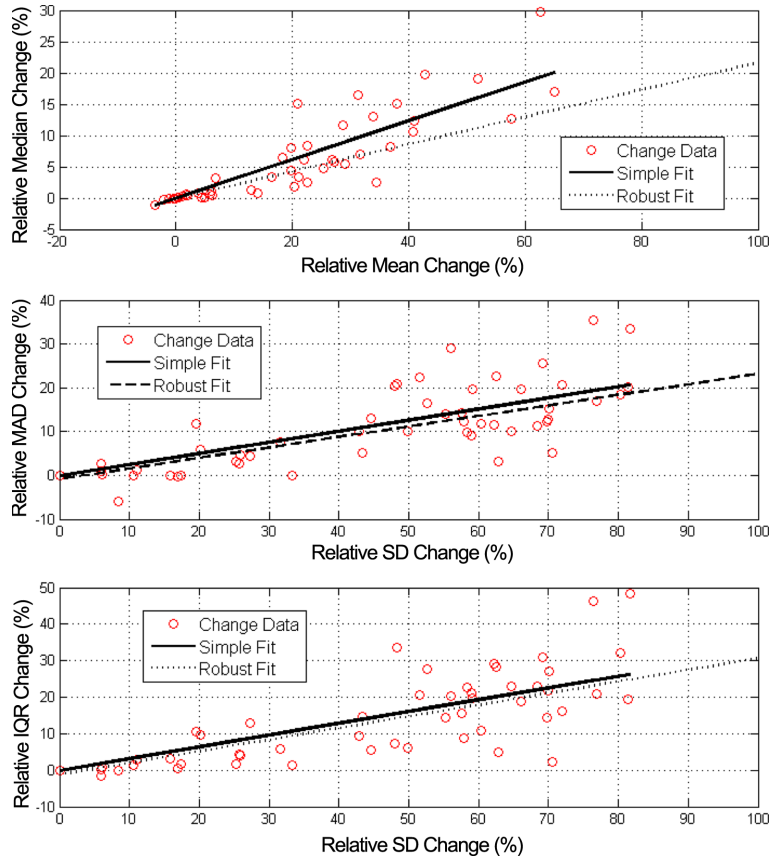


Fig. 3. Comparison of the change ratios between the robust and non-robust measures of location and spread before and after outlier removal conditions (○ = estimation data, Blue thick line = ordinary regression fitting line, Green thick line = robust regression fitting line)

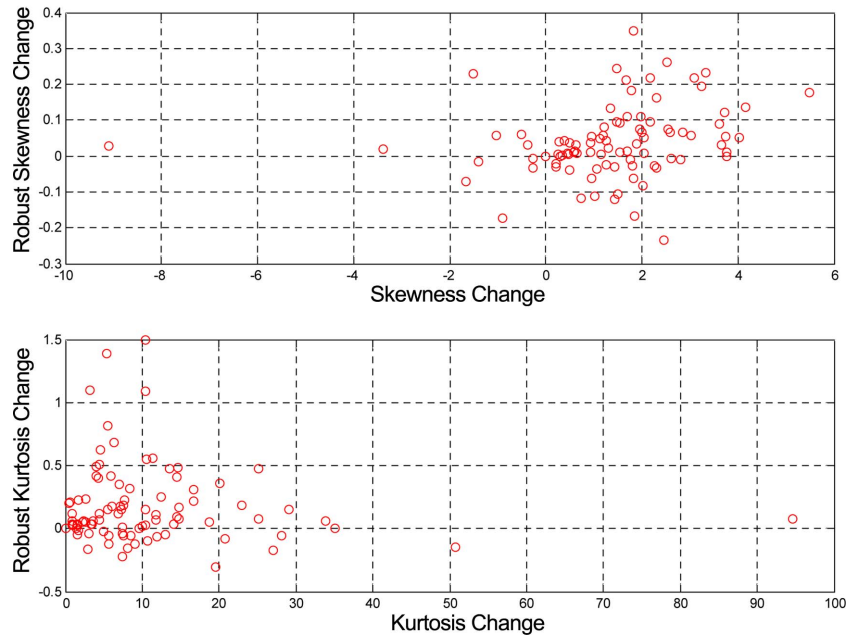


Fig. 4. Amount of change of the skewness and kurtosis before and after outlier removal conditions

기법을 적용한 자료의 작은 변동범위를 보면, 일반적인 통계척도의 변화에 비하여 로버스트 통계척도의 변화정도가 매우 작게 나타나고 있음을 알 수 있다(Figs. 3, 4 참고).

로버스트 회귀분석은 일반적인 선형회귀분석을 초기 조건으로 가정하고, 다음 단계에서는 최적 추정된 선형 회귀 곡선 추정 잔차(residuals)의 크기에 따라 가중계수를 달리 부여하는 방법이다. 잔차가 작으면 큰 가중계수를 주고, 잔차가 큰 경우에는 작은 가중계수를 주는, 일반적으로 잔차가 크게 나타나는 자료를 이상 자료로 간주하여 반복적으로 회귀분석을 수행하는 방법이다(Martinez and Martinez 2005; Rousseeuw and LeRoy 2003).

#### 4. 결론 및 제언

연안 환경자료의 이상자료를 진단하고, 이상 자료를 제외하기 전·후의 통계정보 변화를 비교 분석하였다. 또한 이상자료에 덜 민감한 로버스트 통계정보의 변화 정도도 비교분석하여 로버스트 통계정보로 제시되는 추정방법의 성능평가를 수행하였다. 본 연구를 통하여 얻어진 결과는 다음과 같다.

(1) 이상자료는 영양염류 자료에서 약 5~10% 정도 진단되는 것으로 파악되었으며, 수온, pH, DO 항목은 1% 정도 이하로 진단되었다. 이상 자료는 표층에서 저층보다 높은 비율로 진단되었다. 이러한 이상자료 진단비율은 자료의 계절변동 범위에 따른 영향으로 판단된다.

(2) 이상자료 제외로 인한 기본 통계정보의 변화는 평균

및 표준편차의 경우 13, 33% 정도, 왜도 및 첨도의 경우 1.36, 8.11 정도로 이상 자료로 진단된 개수가 증가할수록 그 변화정도가 크게 나타났으며, 로버스트 통계정보 추정인자로 제시되는 방법으로 통계정보를 추정하는 경우, 그 변화정도는 평균 및 분산의 경우 3.7~10.5% 정도로 매우 작았으며, 왜도 및 첨도의 경우에는 그 변화정도가 일반적인 왜도 및 첨도 변화정도의 2~4% 정도로 매우 미미한 수준으로 파악되어 로버스트 통계정보가 적절한 통계정보를 제시하고 있는 것으로 파악되었다.

(3) 또한 로버스트 통계량(robust statistics)을 이용하면 이상 자료에 대하여 정량적이고 객관적인 평가가 가능하며, 특수한 조건에 대한 정보 왜곡을 크게 저감할 수 있기 때문에 매우 유리한 방법으로 판단된다.

한편 이상 자료로 진단되는 자료는 자료의 분포 특성에 따라 크게 영향을 받기 때문에 보다 다양한 연안 환경자료를 이용하여 진단되는 이상 자료의 개수를 분석하고, 더불어 통계정보의 변화양상을 분석하여 자료를 대표할 수 있는 통계정보 추정에 대한 연구가 수행되어야 할 것으로 판단된다.

그리고 이상자료는 통상적인 관측범위를 벗어난 자료일 수 있기 때문에 이상자료로 진단되었다 할지라도 모두 제거하는 것은 부적절하다. 그러나 이상자료를 포함하여 정상자료의 통계정보를 추정하는 경우, 편향된(biased) 통계정보 추정이 되기 때문에 이상자료를 제외하고 분석을 수행하거나 로버스트 통계기법을 이용하여 분석을 수행하는

것이 실질적으로 바람직하다. 그리고 제외된 이상자료는 정상자료의 통계척도를 이용하여 정량적인 편향 정도를 추정할 수 있기 때문에 극한적인 자연현상 해석에 정량적인 의미를 부여하는 연구가 필요할 것으로 판단된다.

## 사 사

본 논문은 해양수산부의 재원으로 한국해양과학기술진흥원의 지원을 받아 수행된 KOOS 제2단계 연구사업 (PM-58951)의 연구 성과입니다. 연구비 지원에 감사드립니다. 또한 본 연구에서 사용한 환경자료는 국가해양환경측정망의 부산연안 환경자료로, 모니터링 자료를 제공해주신 해양환경관리공단에 감사드립니다. 그리고 관측정점 지도(Fig. 1)에 사용한 부산 연안 해안선 자료는 GIS for Ocean Research (<http://mgis.kiost.ac>)에서 제공받은 자료입니다. 자료제공에 감사드립니다.

## 참고문헌

- 국가해양환경정보통합시스템 (2016) 해양환경측정망 원본 자료. <http://www.meis.go.kr/> Accessed 18 Jan 2016
- 해양환경관리공단 (2015) 국가해양환경측정망 자료 - 부산연안. <http://www.koem.or.kr/> Accessed 18 Jan 2016
- Barnett V, Lewis T (1994) Outliers in statistical data. John Wiley Sons, 584 p
- Bonato M (2011) Robust estimation of skewness and kurtosis in distributions with infinite higher moments. *Financ Res Lett* **8**:77-87
- Brys G, Hubert M, Struyf A (2004) A robust measure of skewness. *J Comput Graph Stat* **13**(4):996-1017
- Dixon WJ (1951) Ratios involving extreme values. *Ann Math Stat* **22**(1):68-78
- Erceg-Hurn DM, Mirosevich VM (2008) Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *Am Psychol* **63**(7):591-601
- Grubbs FE (1950) Sample criteria for testing outlying observations. *Ann Math Stat* **21**(1):27-58

- Huber PJ, Ronchetti EM (2009) Robust statistics. John Wiley & Sons, New York, 380 p
- Kim T-H, White H (2004) On more robust estimation of skewness and kurtosis *Financ. Res Lett* **1**:56-73
- Martinez WL, Martinez AR (2005) Exploratory data analysis with MATLAB. Chapman & Hall/CRC, Boca Raton, 405 p
- Millard SP (2013) EnvStats: an R package for environmental statistics. Springer, New York, 291 p
- Moors JJA (1988) A quantile alternative for kurtosis. *J Roy Stat Soc D-Sta* **37**(1):25-32
- R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://cloud.r-project.org/index.html> Accessed 18 Jan 2016
- Rosner B (1983) Percentage points for a generalized ESD many-outlier procedure. *Technometrics* **25**(2):165-172
- Rousseeuw PJ, Croux C (1993) Alternatives to the median absolute deviation. *J Am Stat Assoc* **88**(424):1273-1283
- Rousseeuw PJ, LeRoy AM (2003) Robust regression and outlier detection. John Wiley & Sons, New Jersey, 329 p

## 국문참고자료의 영어표기

**English translation/Romanization of references originally written in Korean**

- Marine Environment Information System (2016) Raw data - marine environmental monitoring network. <http://www.meis.go.kr/> Accessed 18 Jan 2016
- Korea Marine Environment Management Corporation (2016) <http://www.koem.or.kr/> National marine environmental monitoring network data - Busan coastal sea Accessed 18 Jan 2016

*Received Jan. 18, 2016*

*Revised Apr. 6, 2016*

*Accepted Apr. 12, 2016*