

Article

시화호 · 인천연안 환경자료의 오차범위 분석

조 흥 연*

한국해양연구원 해양환경 · 방제연구부
(425-600) 경기도 안산시 안산우체국 사서함 29

Error Bounds Analysis of the Environmental Data in Lake Shihwa and Incheon Coastal Zone

Hong Yeon Cho*

Marine Environment & Pollution Prevention Research Department, KORDI
Ansan P.O. Box 29, Seoul 425-600, Korea

Abstracts : The characteristic analysis of the estimated population parameters, i.e., standard deviation and error bound of coastal pollutant concentrations (hereafter PC, i.e., COD, TN, and TP concentrations), was carried out by using environmental data with different sampling frequency in Lake Shihwa and Incheon coastal zone. The results clearly show that standard deviation of the PC increases as its mean value increases. The error bounds of the annual mean values based on seasonally measured DO concentrations and PC data in Incheon coastal zone were estimated as ranges 2.26 mg/l, 0.68~0.86 mg/l, 0.62~0.80 mg/l, and 0.074~0.082 mg/l, respectively. In terms of annual mean of the DO concentration and PC in Lake Shihwa, the error bounds based on monthly measured data from 1997 to 2003 were also estimated as ranges 4.0 mg/l, 3.0 mg/l, 0.5~1.0 mg/l, and 0.05 mg/l, respectively. The error bound on the basis of real-time monitoring data is 7~13% only as compared to that of monthly measured data.

Key words : sampling frequency, error bounds, confidence level, environmental data, Incheon coastal zone

1. 서 론

현장 관측은 표본공간(sample space)에서 수행되기 때문에 현장 관측자료는 표본자료로 간주할 수 있다. 표본자료를 이용하여 통계적으로 추정하는 모수(population parameters; 표본자료의 평균, 표준편차(분산) 등)는 자료의 수(sample number, n)와 변동정도(표본분산, sample variance), 신뢰수준(confidence level, 보통 95%)에 따라 결정되는 신뢰구간(오차범위, error bound, margin of errors)을 가지게 된다(김 등 2007). 일반적으로 자료의 개수가 많고, 표준편차가 작고, 신뢰수준이 낮을수록 신뢰구간은 감소하고, 반대인 경우에는 신뢰구간이 증가하게 된

다. 기본적 추정 모수인 평균, 표준편차의 신뢰구간(confidence intervals)이 증가한다는 것은 표본자료를 이용하여 추정한 평균, 표준편차의 오차범위가 증가하는 것을 의미하며, 이 범위가 일정 범위를 넘어서는 경우에는 추정한 평균 및 표준편차가 실질적으로 무의미하거나 관측 목적에 부적합하게 되는 경우가 발생하게 된다.

연안 환경변화를 시간적 · 공간적으로 파악하기 위하여 수행되는 관측은 통계적인 측면에서는 연안 환경변화를 목표수준의 정확도로 파악(모수 추정)하기 위하여 표본을 추출(관측)하는 과정으로 간주할 수 있다. 따라서 정확한 환경변화 파악을 위해서는 관측 자료를 이용하여 분석되는 추정모수의 신뢰수준에 따른 오차범위가 일정 수준(목표 수준) 이내로 제한되어야 한다. 추정 모수의 오차범위는 단순한 가정을 통하여 확보되는 것이 아니라 표본자료

*Corresponding author. E-mail : hycho@kordi.re.kr

정보로부터 결정되는 것이기 때문에 추정 모수의 신뢰수준이 허용범위보다 낮은 경우에는 표본의 개수를 증가시켜 신뢰수준을 높이는 과정이 불가피하다. 그러나 이러한 통계적인 신뢰수준 제약이 있음에도 불구하고 제한된 표본을 이용하여 환경변화를 분석·추정·제시하는 연구, 즉 오차범위를 무시한 점 추정(point estimation) 자료를 이용한 분석이 빈번하게 수행되고 있다. 오차범위를 무시한 환경자료의 분석결과는 자료로부터 추정되는 오차범위 또는 연구목적에 따라 무의미하게 될 수도 있기 때문에 통계적인 신뢰수준 검토가 있어야 한다.

본 연구에서는 연안 환경변화를 분석하기 위하여 관측된 기존의 환경자료를 이용하여 환경변화 분석에 필요한 모수의 오차범위를 추정·분석하였다. 오차범위 분석은 다양한 관측빈도의 연안 환경자료가 가용한 시화호·인천연안의 자료를 이용하여 수행하였으며, 환경변화는 연안 환경인자의 평균 농도변화와 자료의 변동 정도를 반영하는 표준편차 변화를 중점적으로 분석하였다. 한편, 연안 환경인자 관측에서 필요로 하는 목표 오차범위 달성을 위한 최소한의 관측빈도도 본 연구에서 사용한 자료를 근거로 하여 추정·제시하였다.

2. 시계열 자료의 관측간격 및 대표기간

기본적으로 일정 시간 간격(Δt)으로 관측된 시계열자료(time-series data)에서 파악할 수 있는 한계빈도(최대빈도, 최소주기) 정보는 Nyquist 빈도(Nyquist frequency = $1/[2\Delta t]$)이며, 보다 작은 주기의 변화를 파악하는 것은 이론적으로 불가능하다(이 2001; Harvey 1993). 따라서 월별($\Delta t=1$ 달) 관측 자료에서는 $1/(2$ 달), 즉 2개월에 1회 발생하는 빈도(또는 2달 주기)보다 작은 규모의 변화(일일변화, 주별·순별 변화, 보름 동안의 변화 등) 등을 파악할 수 없다. 그리고 시계열 자료의 연속된 관측자료 사이 시간에서의 변화는 두 관측 자료의 변화 범위 내에서의(선형)변화를 함축하고 있으며, 각각의 관측 자료는 다음 관측시간, 이전 관측시간 동안의 자료를 대표하고 있다고 가정한다.

한편, 관측 자료가 관측시점에서의 정확한 자료라 할지라도, 일정 기간의 표본자료에 불과하기 때문에, 관측시점의 자료를 각각 연결하여 변화 양상을 분석할 수는 있으나, 관측자료 하나하나가 일정한 관측 기간의 관측 인자로서의 대표성이 있는가를 판단하기 위해서는 일정기간 동안 일정한(충분한) 개수를 가지는 표본으로 간주하여 분석할 필요가 있다. 표본은 모집단의 모수(parameters)인 평균, 분산 등을 통계적으로 추정하는 재료이므로, 모수 추정을 위한 구간에서의 표본자료의 개수가 충분하게 확보되어야 한다. 하나의 표본으로 한 시점을 대표할 수는 있으나, 표본의 개수($n=1$) 제한으로 말미암아 연속되는 시

간간격에서의 공백 기간까지도 대표할 수는 없다. 따라서 표본의 개수 증가에 따라 모수 추정의 정확도가 증가하는 이론에 근거하여 모수의 신뢰구간추정을 수행하는 경우, 대표본은 적어도 $n=30$ 개 이상의 표본 개수를 포함하는 기간이 필요하다. 소표본(small sample, $n=15\sim 25$)의 경우에도 모수 추정은 가능하지만 신뢰구간(오차범위)이 증가하게 된다. 통계적으로 원하는 정확도를 달성하기 위해서는 표본의 개수를 그 수준을 달성하는 한계까지 늘려야 한다(김 등 2007; Cryer 1986; Keith 1988; Spiegel 1997; Montgomery 2005; Agresti and Franklin 2007). 이미 관측된 시계열 자료에서는 원하는 신뢰구간을 달성하는 통계 모수 추정을 위해서는 신뢰구간 달성에 필요한 표본의 개수를 포함할 수 있는 기간으로 모수 추정(분석)이 제한된다. 새로운 관측을 수행하고, 분석을 수행하고자 하는 기간이 결정(제한)되어 있는 경우에는 표본의 개수를 증가시키는 것이 불가피하다. 비용, 인력, 시간 등의 실질적인 문제로 표본의 수를 증가하는 관측이 제한을 받는 경우에는 결정된 조건에서의 신뢰구간에 따른 오차발생 가능성을 배제할 수 없다. 따라서 모든 분석은 기본적으로 이 불확실성(uncertainty)을 포함한 상태에서의 분석으로 귀결된다.

시계열 자료 하나하나를 정확한 관측(관측 오차는 무시)이 이루어졌다고 가정하는 경우, 관측시점에서의 자료로서의 가치는 있으나 관측구간에서의 자료를 대표하기에는 관측항목의 변동성과 변동주기에 따라 다르지만 무리가 있다고 할 수 있다. 즉, 월별 간격으로 관측되는 자료는 관측시점에서 정확하게 관측되어도 관측시점에서의 자료로서의 가치는 있으나, 월별 이내에서도 변동성이 상당한 경우 월 자료를 대표하기에는 상식적으로 무리가 있으며, 통계적인 관점에서도 $n=1$ 조건에서의 표본에 해당하기 때문에 표본을 이용한 모수 추정도 곤란하다.

시간과 비용의 제약에 의하여 관측빈도를 결정하거나 통상적인 방법을 기반으로 관측을 설계·시행하는 경우가 빈번하나 통계적인 분석과정에 근거하지 않는다면 과학적인 근거는 없다고 할 수 있다. 이 경우에는 분석 자료의 오차한계에 대한 분석이 포함되어야 한다. 과학적으로 한 시점의 정확한 관측 자료는 분석 수준 검토에는 유용하지만 시간적인 변화분석에는 무의미하다. 시계열 관측 자료가 변화양상을 적절한 수준으로 재현하고, 어떤 관심기간의 보고자 하는 현상을 대표할 수 있는 자료를 산출할 수 있는가 여부가 가장 중요한 부분으로 판단된다. 실질적으로 연안 환경변화 분석은 관측 오차, 분석 오차, 표본 오차 등이 모두 반영되기 때문에 어느 한 부분의 개선만으로는 오차범위를 줄일 수 없다(Keith 1988; Ward *et al.* 1990; US EPA 1992; Berthouex and Brown 1994; Emery and Thompson 1997).

통계적인 측면에서 대표 자료를, 즉 모집단의 모수를 가능한 정확한 범위에서 추정하기 위해서는 표본의 수를 증가시키지 않고는 달성할 수 없다. 월별 대표자료, 즉 환경인자의 월평균 농도자료를 파악하고자 한다면 해당 월에 충분한 표본을 확보하여 평균의 구간추정을 수행하여야 한다. 해당 월 자료가 하나인 채로 수년 동안 지속하여 왔다면 변화추세를 제거한 후, 해당 월의 자료를 모아서 모수 추정을 한다면 가능할 수 있다. 표본의 개수가 작으면 보다 많은 시간 변화 양상을 파악할 수 있으나 파악하는 변화의 신뢰구간이 감소하는 상충하는 결과를 가져올 수 있다.

3. 인천연안의 수질변동 특성 및 오차구간

인천연안은 국립수산과학원에서 계절별(2월, 5월, 8월, 11월; 연 4회)로 해양환경을 관측하는 지역으로 장기간의 환경자료가 축적되어 있는 지역이다(국가해양환경통합정보시스템 2007; 물환경정보시스템 2007; 해양수산연구정보 2007; 관측지점은 Fig. 1 참조). 본 연구에서는 1995년~2006년 동안의 18개 지점 인천연안 환경자료를 이용하여 환경자료의 변동 정도, 즉 표준편차 변동양상 및 평균농도와의 상관관계를 분석하고, 분석된 결과를 이용하여 95% 신뢰수준에서 추정 모수, 즉 추정평균의 오차구간도 산정하였다. 오차구간 분석에서 중요한 인자는 모표준편차(population standard deviation)이나 모표준편차를 모를 경우에는 표본 표준편차를 사용하는 것을 권장하고 있으나, 실질적으로 표본의 개수가 작을수록 표준편차의 변동 정도가 커지기 때문에 오차구간은 크게 증가할 수 있다. 따라서 불가피한 경우 표본편차를 사용하는 것도 의미가 있으나 자료로부터 적절한 방법으로 신뢰할만한 표준편차를 추정

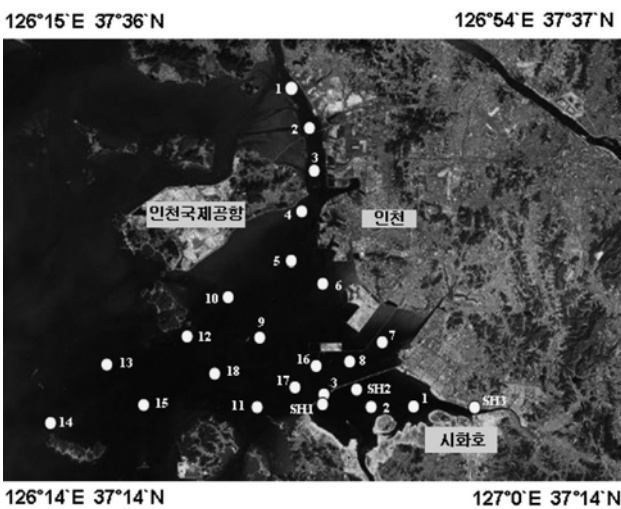
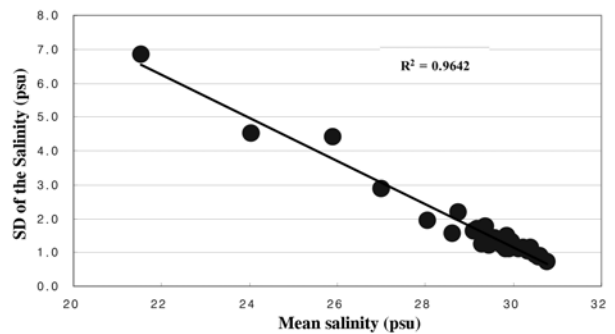


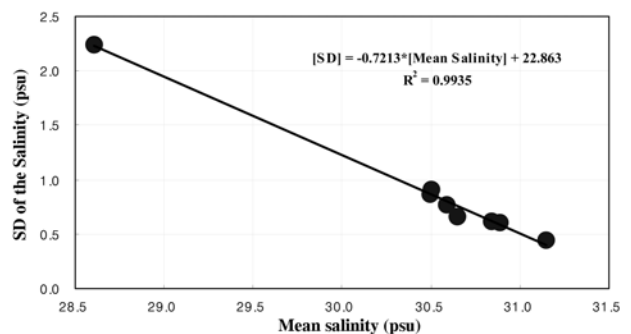
Fig. 1. Water quality measurement stations in Incheon coastal zone and Lake Shihwa.

하는 방법도 매우 유용하다(박 1985; Atkinson 1985). 본 연구에서는 회귀분석을 통하여 환경자료의 평균 농도와 표준편차의 상관관계도 분석하였다.

환경자료의 오차구간 결정에 영향을 미치는 인자는 표본의 개수, 표본의 표준편차, 신뢰수준이다. 일반적으로 사용되는 95% 정도로 신뢰수준을 결정하면 실질적으로 오차구간을 결정하는 인자는 표본의 개수와 표준편차가 된다. 인천 연안 환경자료는, 오염물질은 아니지만 추적자(tracers)로 판단되는 표층과 저층의 염분 농도변화를 보면 유역 담수 유출의 영향을 받는 것으로 판단되는 염분 농도가 낮은 영역에서 염분 자료의 표준편차가 크게 나타나는 경향도 볼 수 있다(Fig. 2). 즉 염분의 저하 및 오염물질 농도 증가는 하천 또는 배수갑문(방조제) 개방을 통하여 배출되는 수량의 영향으로 판단할 수 있으며, 오염물질 유출지점에서 가까운 지역일수록 수량의 큰 변동성이 오염부하에 반영되어 염분 및 오염물질의 표준편차를 크게 하는 것으로 추정할 수 있다. 이와 같은 경향은 아산만 염분변화 양상분석 결과(조 등 2008)와도 일치한다. DO 농도는 평균 농도와 관계없이 표준편차는 약 2.0~2.5 mg/l 정도를 유지하고 있는 것으로 파악되었다(Fig. 3). 반면 오염물질, 즉 COD, TN, TP 평균 농도와 표준편차는 강한 상관관계를 보이고 있는 것으로 파악되었다. 연안 환경은 육상 오염부하의 영향으로 오염물질이 유입되어 영향을 미치는 영역, 즉 오염물질의 농도가 높은 영역은 오염물질



(a) Incheon coastal zone



(b) Asan Bay (Cho et al. 2008)

Fig. 2. Regression curve between annual mean and standard deviation of the salinity.

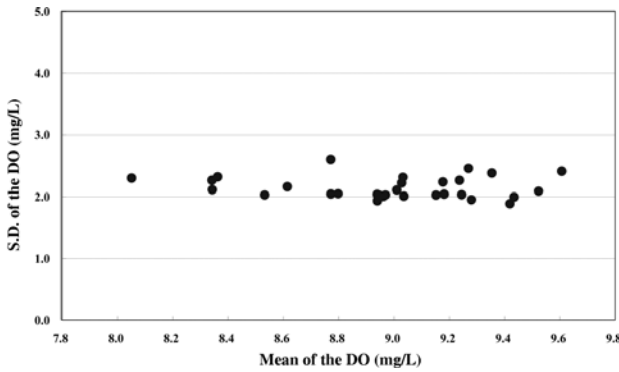


Fig. 3. Scatter plot between mean and S.D. of the DO concentration at each station in Incheon coastal zone.

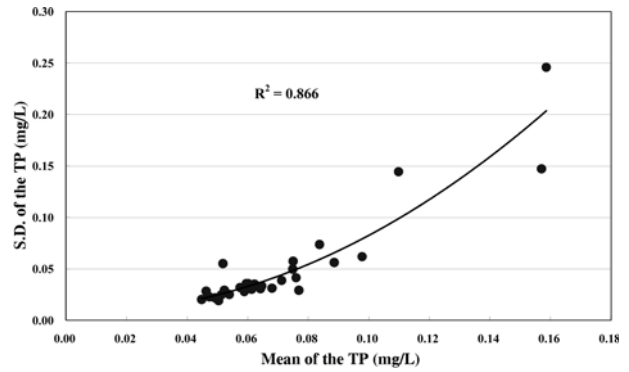


Fig. 6. Scatter plot between mean and S.D. of the TP concentration at each station in Incheon coastal zone (excluding the outlier, surface data of St. 12).

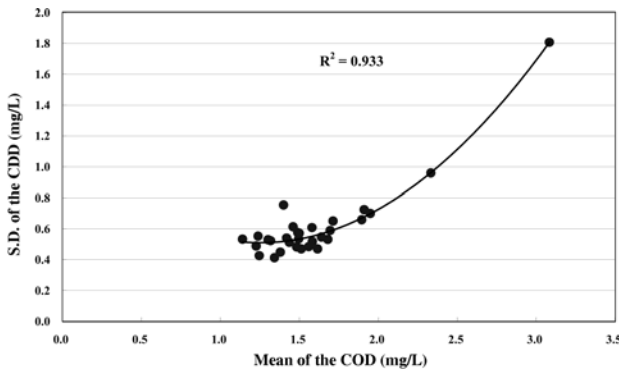


Fig. 4. Scatter plot between mean and S.D. of the COD concentration at each station in Incheon coastal zone.

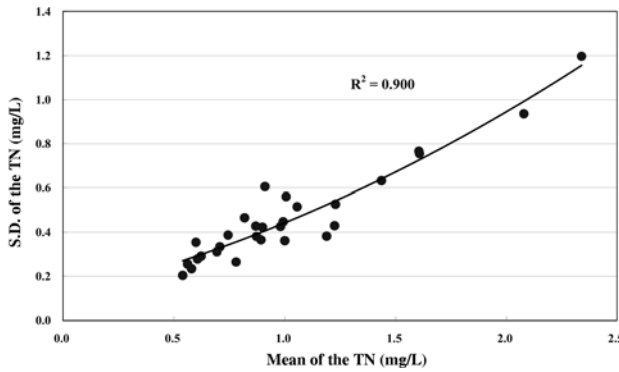


Fig. 5. Scatter plot between mean and S.D. of the TN concentration at each station in Incheon coastal zone (excluding the outlier, surface data of St. 12).

농도자료의 표준편차도 상대적으로 크게 나타나고 있다 (Figs. 4~6 참조). 인천 연안 12지점(무의도 남쪽) 등 이상 자료(outlier)로 판단되는 지점의 자료는 통계적인 기준에 따라 분석기준에서 제외하였다(Barnett and Lewis 1998; Goda 2000; www.graphpad.com 2008).

Table 1. Temporal standard deviation variations and error margins of the WQ Constituents in Incheon coastal zone

Month	S.D. of the DO		DO Margin of Error	
	Surface	Bottom	Surface	Bottom
2	1.46	1.35	[1.61] (4.04)	[1.49] (3.74)
5	0.93	0.85	[1.03] (2.59)	[0.94] (2.36)
8	1.49	1.18	[1.65] (4.14)	[1.30] (3.26)
11	0.60	0.65	[0.66] (1.66)	[0.72] (1.78)
Annual	2.13	2.13	[1.10] (2.26)	[1.10] (2.26)
Month	S.D. of the COD		COD Margin of Error	
	Surface	Bottom	Surface	Bottom
2	0.80	0.93	[0.88] (2.21)	[1.03] (2.59)
5	0.59	0.66	[0.65] (1.63)	[0.73] (1.83)
8	0.39	0.90	[0.43] (1.08)	[0.99] (2.48)
11	0.46	0.67	[0.51] (1.28)	[0.74] (1.86)
Annual	0.64	0.82	[0.33] (0.68)	[0.42] (0.86)
Month	S.D. of the TN		TN Margin of Error	
	Surface	Bottom	Surface	Bottom
2	0.54	0.52	[0.60] (1.51)	[0.57] (1.43)
5	0.81	0.63	[0.90] (2.26)	[0.70] (1.76)
8	0.97	0.71	[1.07] (2.69)	[0.78] (1.96)
11	0.52	0.43	[0.57] (1.43)	[0.48] (1.20)
Annual	0.75	0.58	[0.39] (0.80)	[0.30] (0.62)
Month	S.D. of the TP		TP Margin of Error	
	Surface	Bottom	Surface	Bottom
2	0.041	0.052	[0.045] (0.113)	[0.057] (0.143)
5	0.025	0.026	[0.028] (0.070)	[0.029] (0.073)
8	0.039	0.050	[0.043] (0.108)	[0.055] (0.138)
11	0.138	0.111	[0.153] (0.384)	[0.123] (0.309)
Annual	0.077	0.069	[0.040] (0.082)	[0.036] (0.074)

*Ref. S.D. = Standard deviation, [], () = Estimated error margins using the 15 and 4 sampling stations data, respectively.

오차범위 계산을 위한 자료의 수는 인천연안 수심 평균 계절별 환경자료의 오차범위로 15개 정점 자료(최근에 추가된 16, 17, 18의 3개 지점 자료는 자료의 수가 적어서 분석에서 제외)를 모두 이용하여, 즉 $n = 15$ 로 간주하여 오차범위를 계산하였다(Table 1 참조). 연평균 오차범위는 표본의 수를 $15(\text{지점}) \cdot 4(\text{개/년}) = 60$ 개로 간주하고, 인천 연안영역도 관측지점 4개 정도를 포함하는 4개의 영역(15개 관측지점을 4개의 영역으로 구분)으로 구분하여 환경변화를 분석하였다. 이 경우, 표본의 개수 감소로 계절별 오차한계는 2.51배 증가하고, 연평균 오차한계는 2.05배 증가한다(소표본의 경우, Student t-분포 이용, 95% 신뢰수준, 계절별 자료는 D.F.(Degree of freedom, 자유도) = $14 \rightarrow t = 2.14$ 적용; 연평균 자료는 D.F. = 59, $\rightarrow t = 2.00$ 이용). 따라서 한 지점의 농도변화 자료를 이용하여 환경변화 양상을 분석하는 경우는 변화범위가 오차범위에 포함될 가능성이 크며, 한 시점에서의 농도변화 양상 분석도 표본의 수 관점에서는 동일한 결과를 유발할 수 있다. 오차범위는 특정 시점에서의 공간적인 오차범위(spatial error bounds)와 특정 영역에서의 시간적인 오차범위(temporal error bounds)로 구분할 수 있다. 통계적인 관점 또는 실질적인 관점에서 특정 지점의 시간적인 오차범위 분석은 가능하나 공간적인 오차범위 분석은 불가하며, 특정 시점의 공간적인 오차범위 분석(공간적 자료가 2개 이상)은 가능하나 시간적인 오차범위 분석은 불가하다. 자료가 1개인 경우에는 분산이 무한대로 되어 오차범위가 무한대로 증가하기 때문이다.

연구목표 및 관심을 두는 환경 인자에 따라 오차한계(error margin) 범위가 결정될 수 있으나 실질적으로 12년 동안의 자료를 이용한 연평균농도의 관점에서 DO 농도 1.0(mg/l) 정도의 오차한계는 수용 가능하지만, COD, TN 농도는 0.5 mg/l 수준, TP 농도는 0.05 mg/l 수준 이상의 오차범위는 수용이 곤란할 것으로 판단된다. 모든 경우에서 오차 범위에 포함되는 정도의 변화양상을 분석하는 것은 통계적으로 불합리하며, 오차범위를 줄이기 위해서는 지점 또는 시점을 통합하여 표본의 수를 증가시켜 분석하는 것이 타당하다. 그러나 계절별 농도변화 분석을 위한 경우에는 자료의 수가 1/4 수준으로 감소하기 때문에 오차범위도 COD, TN, TP 항목은 연평균 농도의 오차범위는 각각 1.1~2.6 mg/l, 1.2~2.7 mg/l, 0.07~0.38 mg/l 정도로 크게 증가하게 되므로 오차범위보다 작은 변화 양상 분석은 실질적으로 무의미하며 잘못된 분석결과를 도출할 수도 있다.

4. 시화호의 수질변동 특성 및 오차구간

인천연안 환경자료에 비하여 시화호는 관측빈도가 다양

한 자료를 이용할 수 있다. 시화호 지역에서 가용한 환경 자료는 실질적으로 연속 관측(일 자료로 제공)되는 실시간 해양수질 자동측정망 자료(해양수산부)와 월별관측이 수행되는 환경부 자료, 계절별 관측이 수행되는 해양수산부(국립수산과학원) 자료이다. 환경부와 해양수산부 환경자료는 관측 업무 인수인계과정에서 관측빈도가 각각 연 12회(월별 관측), 연 4회(계절별 관측 - 표층, 저층 수행)로 차이가 있다. 시화호 자료 변동 특성은 국립수산과학원 자료(환경부 자료 포함)를 이용한 시화호 환경자료의 분산변화 분석, 표준편차와 평균 농도 관계분석 및 실시간 해양수질 자동측정망 자료를 이용한 SH1, SH2, SH3 지점의 분산변화 분석, 표준편차와 평균 농도 관계분석을 중심으로 수행하였다.

시화호 1, 2, 3지점의 공간적인 상관관계(결정계수 분석), 평균과 표준편차 정보의 상관관계 및 연평균 자료 신뢰구간은 1995년부터 2006년까지의 DO, COD, TN, TP 농도자료를 이용하여 분석하였다(Tables 2~4, Figs. 7~10). DO 농도자료를 제외하고는 COD, TN, TP 농도는 공간적으로 매우 큰 상관관계를 보이고 있는 것으로 파악되었다. 또한, 수질항목의 농도와 표준편차 변화도 DO 항목은 크지 않으나 TN, TP 항목은 매우 크게 나타났으며, COD 항목도 결정계수가 0.58 정도로 유의할 만한 수준으로 파악되었다. 한편, 표본자료를 이용한 추정 모수의 오차범위를 절반으로 줄이기 위해서는 표본의 개수를 4배로 증가시켜야 하는 기본적인 규칙에 의하면, 일단 2004년 이후 표본의 수가 12개에서 4개로 감소한 영향을 보완하기 위하여 기본적으로 월별 관측이 필요하며, 현재의 신뢰수준도 절반 이하로 줄이기 위해서는 중요 지점 또는 대표지점에 대해서는 주별 관측이 수행되어야 할 것으로 판단된다. 2004년 이후의 농도 오차한계 증가는 오염물질의

Table 2. Coefficient of determination between stations on the WQ concentrations in Lake Shihwa

Items	St. 1 vs. St. 2	St. 2 vs. St. 3
Annual-mean DO	0.6322	0.5712
Annual-mean COD	0.9752	0.9585
Annual-mean TN	0.9151	0.9365
Annual-mean TP	0.9655	0.9506

Table 3. Coefficient of determination between annual mean and standard deviations on the WQ concentrations at each station in Lake Shihwa

	DO	COD	TN	TP
Station 1	0.5670	0.6764	0.9847	0.8714
Station 2	0.0416	0.6230	0.9401	0.6967
Station 3	0.0058	0.5647	0.9673	0.9089
All Stations	0.2345	0.5803	0.9492	0.8101

Table 4. Temporal and spatial standard deviation variations of the monthly environmental data in Lake Shihwa

DO (mg/l)					COD (mg/l)				
Month	St.1	St.2	St.3	Mean	Month	St.1	St.2	St.3	Mean
1	3.32	2.97	2.87	3.05	1	4.15	5.76	4.26	4.72
2	1.65	1.80	2.39	1.95	2	6.47	4.69	5.26	5.47
3	1.96	1.64	1.64	1.75	3	8.50	6.99	7.33	7.61
4	2.07	2.19	2.64	2.30	4	6.31	6.47	6.26	6.35
5	3.09	3.38	2.37	2.95	5	6.12	7.21	6.76	6.70
6	2.75	1.98	1.81	2.18	6	7.43	7.23	7.16	7.27
7	2.60	1.53	2.30	2.14	7	4.75	3.53	4.74	4.34
8	2.99	2.58	2.42	2.66	8	3.95	3.99	4.05	4.00
9	4.43	3.65	1.45	3.18	9	4.45	4.58	1.91	3.64
10	1.92	1.94	1.76	1.87	10	3.47	5.53	5.23	4.74
11	2.82	3.12	2.81	2.92	11	2.39	5.79	5.45	4.54
12	1.51	1.38	1.43	1.44	12	4.80	4.18	3.96	4.31
Mean	2.59	2.35	2.16	2.37	Mean	5.23	5.50	5.20	5.31

TN (m/l)					TP (mg/l)				
Month	St.1	St.2	St.3	Mean	Month	St.1	St.2	St.3	Mean
1	3.74	3.65	3.73	3.71	1	0.060	0.281	0.199	0.180
2	4.48	3.37	2.92	3.59	2	0.273	0.101	0.179	0.184
3	2.93	2.33	2.13	2.46	3	0.218	0.133	0.274	0.208
4	2.29	2.11	2.60	2.34	4	0.119	0.124	0.092	0.111
5	3.87	2.56	2.93	3.12	5	0.124	0.086	0.095	0.102
6	2.25	3.75	3.38	3.13	6	0.130	0.121	0.109	0.120
7	1.55	3.02	1.94	2.17	7	0.106	0.126	0.110	0.114
8	2.16	2.50	2.05	2.24	8	0.066	0.063	0.083	0.071
9	1.73	2.12	1.58	1.81	9	0.065	0.064	0.064	0.065
10	1.91	2.68	2.28	2.29	10	0.052	0.069	0.061	0.061
11	4.03	5.07	1.51	3.54	11	0.080	0.097	0.093	0.090
12	2.89	3.07	3.53	3.16	12	0.063	0.059	0.071	0.064
Mean	2.82	3.02	2.55	2.80	Mean	0.110	0.110	0.120	0.114

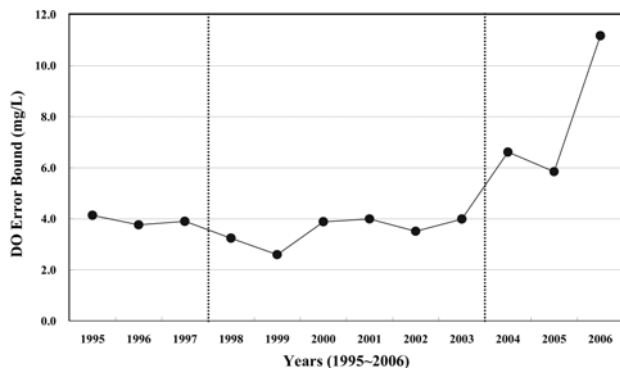


Fig. 7. Error bound of the estimated mean DO concentration in Lake Shihwa (95% confidence level).

농도증가에서 원인을 찾는 것보다는 표본의 수 감소에 따른 오차한계 증가 원인으로 판단하는 것이 더욱 타당하다. 한편, 시화호 연속(실시간, real-time) 관측자료 분석은

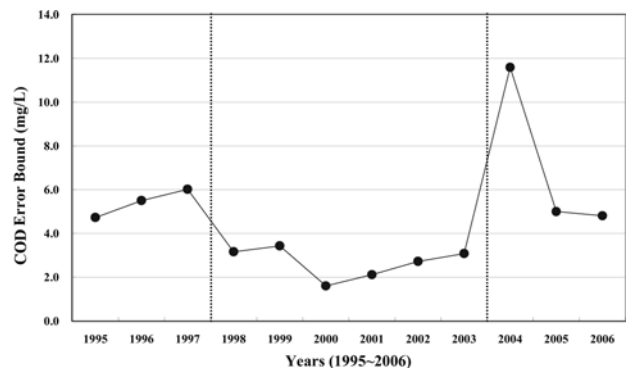


Fig. 8. Error bound of the estimated mean COD concentration in Lake Shihwa (95% confidence level).

관측빈도 약 10분~2시간 간격으로 측정된 자료를 평균한 일별 농도관측 자료를 이용하였으므로 연 관측빈도는 365회를 크게 넘고 있으나, 일 자료이므로 연 365회로 간주하

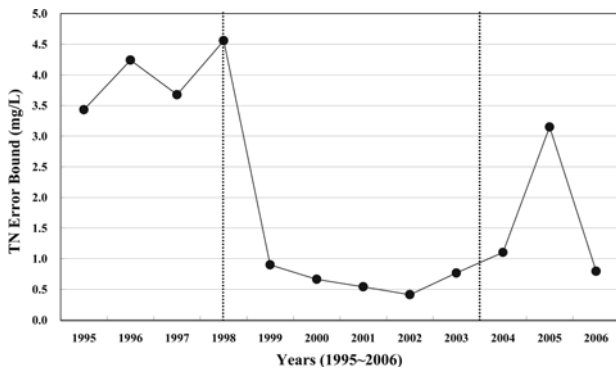


Fig. 9. Error bound of the estimated mean TN concentration in Lake Shihwa (95% confidence level).

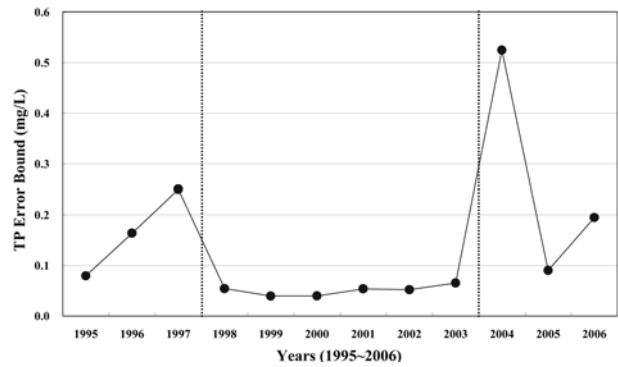


Fig. 10. Error bound of the estimated mean TP concentration in Lake Shihwa (95% confidence level).

Table 5. Mean, standard deviation, and error bounds of the annual mean values on the basis of the real-time COD, TN and TP concentration monitoring data in Lake Shihwa

COD (mg/l)	Mean			Data numbers			Standard deviation			Error bounds		
	SH1	SH2	SH3	SH1	SH2	SH3	SH1	SH2	SH3	SH1	SH2	SH3
2003	5.20	6.11	10.48	277	271	267	1.41	1.51	4.64	0.33	0.36	1.11
2004	4.29	3.74	10.03	277	277	251	2.69	1.68	5.95	0.63	0.40	1.47
2005	5.05	4.67	13.21	301	299	322	2.48	1.45	5.95	0.56	0.33	1.30
2006	3.45	3.71	12.06	350	331	352	2.11	1.58	6.25	0.44	0.34	1.31
2007	2.67	3.85	10.87	323	314	319	1.13	1.72	6.51	0.25	0.38	1.43
DO (mg/l)	SH1	SH2	SH3	SH1	SH2	SH3	SH1	SH2	SH3	SH1	SH2	SH3
2003	7.33	7.92	7.17	168	215	258	2.20	1.87	3.98	0.67	0.50	0.97
2004	7.33	8.51	7.55	297	313	284	3.00	3.68	4.33	0.68	0.82	1.01
2005	9.00	9.26	8.47	339	300	308	2.85	3.27	4.25	0.61	0.74	0.95
2006	8.42	8.83	8.28	352	347	360	2.40	3.27	4.73	0.50	0.69	0.98
2007	8.97	9.03	9.09	356	363	336	2.16	2.96	4.22	0.45	0.61	0.90
Salinity (psu)	SH1	SH2	SH3	SH1	SH2	SH3	SH1	SH2	SH3	SH1	SH2	SH3
2003	22.62	25.74	19.00	226	241	250	6.10	4.90	6.85	1.59	1.24	1.70
2004	24.65	24.83	22.44	289	314	311	4.52	5.44	7.58	1.04	1.20	1.68
2005	28.23	26.06	21.62	344	307	321	3.18	3.57	7.58	0.67	0.80	1.66
2006	29.30	26.66	22.32	352	352	362	4.39	6.29	8.16	0.92	1.31	1.68
2007	30.45	28.61	20.89	361	355	352	3.61	4.31	7.63	0.74	0.90	1.59
TN (mg/l)	SH1	SH2	SH3	SH1	SH2	SH3	SH1	SH2	SH3	SH1	SH2	SH3
2006	0.482	0.409	1.265	342	336	344	0.312	0.38	0.983	0.066	0.081	0.208
2007	0.484	0.414	2.092	324	291	350	0.138	0.236	0.899	0.030	0.054	0.188
TP (mg/l)	SH1	SH2	SH3	SH1	SH2	SH3	SH1	SH2	SH3	SH1	SH2	SH3
2006	0.041	0.043	0.140	355	335	350	0.029	0.034	0.104	0.006	0.007	0.022
2007	0.047	0.042	0.172	333	303	278	0.026	0.027	0.092	0.006	0.006	0.022

는 경우, DO, COD, TN, TP 항목의 오차범위는 SH1 지점과 SH2 지점은 0.45~0.82 mg/l, 0.25~0.63 mg/l, 0.03~0.08 mg/l, 0.006~0.007 mg/l 범위이며, 오염이 심한 SH3

지점은, 0.9~1.0 mg/l, 1.11~1.47 mg/l, 0.19~0.21 mg/l, 0.022 mg/l 범위로 SH1 지점과 SH2 지점에 비하여 2~5배 정도로 크게 나타났다. 즉 SH1 지점과 SH2 지점을 기준

Table 6. Error bound estimation methods based on the sample types

Parameters	Large sample (n ≥ 30)	Small sample (n < 30)
Mean*	<u>Normal Distribution</u> Student t-Distribution	Student's t-Distribution
Variance, Standard Deviation**	Chi-Squared (χ^2) Distribution	Chi-Squared (χ^2) Distribution

Ref. * = 김 등 (2000), ** = Kottogoda and Rosso (1997).

으로 하는 경우 시화호 실시간 수질자료의 오차범위는 연 12회 정도의 월별 관측 자료를 이용하여 추정된 오차범위에 비하여 DO, COD, TN, TP 항목은 각각 87~93% 정도 감소하는 것으로 파악되었다(Table 5~6). SH 3지점의 오차범위 증가는 오염물질 농도 증가에 따른 표준편차 증가에 의한 영향으로 파악되었다.

5. 목표 오차구간 확보에 필요한 표본 규모 추정

표본 하나에서 얻어내는 정보가 아무리 정확하다 할지라도, 표본을 이용하여 모집단의 특성(모수)을 정확하게 추정하기 위해서는 표본의 수가 증가함에 따라 모수 추정구간(오차범위)가 $1/\sqrt{n}$ 비율로 감소하기 때문에 표본의 개수를 증가시키는 것이 가장 우선적으로 수행되어야 한다. 즉, 연 4회 매우 정확한 측정(측정오차 ±0.1%)을 통한 연평균 농도(대표농도)의 신뢰구간 추정결과와 연 12회의 다소 부정확한 측정(측정오차 ±5%)을 통한 연평균 농도 신뢰구간 추정결과와는 정확도에 따라 차이가 있으나, 표본개수가 큰 경우가 다소 정확도가 떨어지더라도 모수 추정 오차한계는 작을 수도 있다.

기본적으로 큰 시간(장기간의) 규모의 대표 정보를 얻기 위해서는 작은 시간규모의 표본을 이용하여 추정하는 방법이 타당하지, 큰 시간규모의 관측으로는 곤란하다. 또한, 장기간의 추세분석 및 일관적인 자료의 품질관리 등도 매우 중요한 요소이다(Farnum and Stanton 1989; Haldar and Mahadevanl 2000; Ott and Longnecker 2001). 월 평균자료는 일별 또는 주별 관측자료(평균자료)에서 추정하는 것이 바람직하고, 연평균 자료는 적어도 월별 관측 자료를 이용하여 추정하는 것이 타당하다. 월 1회의 관측으로 한 달 동안의 자료를 대표할 수 없으며, 연 1회의 관측으로 1년 동안의 자료를 대표할 수 없다.

표본의 크기는 실질적으로 목표로 하는 오차한계, 자료의 표준편차 정보로부터 결정된다(김 등 2007; Agresti and Franklin 2007). 오차는 표본자료를 이용한 추정량과 모수의 차이를 의미하며, 오차의 한계는 $100 \cdot (1 - \alpha)\%$ 방식으로 표현한다. 과거의 자료로부터 표준편차(또는 모집단의 표준편차)를 알고 있는 경우에는, 오차한계(margin of error)를 d 수준 이내로 하기 위한, 즉 신뢰구간의 길이를 $2d$ 수준으로 하기 위한 표본의 크기 n (개수)은 다음 부

등식을 만족한다.

$$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{(n)}} \leq d \rightarrow n \geq (z_{\alpha/2} \cdot \sigma/d)^2$$

이 조건을 만족하는 최소의 정수(n)

이 공식은 모표준편차를 아는 경우에만 적용할 수 있는 결과이다. 그러나 실제로는 모표준편차를 모르는 경우가 대부분이며, 이때에는 작은 크기의 예비표본을 미리 추출하여 모표준편차를 추정하고, 이를 이용하여 n 값을 구할 수 있다. 90%, 95%, 99% 오차한계에 대한 표본크기는 각각 다음과 같다.

$$\begin{aligned} n &\geq (1.64 \sigma/d)^2 \\ n &\geq (1.96 \sigma/d)^2 \approx 4\sigma^2/d^2 \\ n &\geq (2.57 \sigma/d)^2 \end{aligned}$$

인천연안은, 지점에 따라 차이가 있으나 COD 표준편차 1.0(mg/l) 정도의 조건을 적용하는 경우, 95% 신뢰수준에서 오차한계 $d=1.0, 0.5, 0.1$ 조건을 적용하는 경우의 필요한 최소의 표본개수는 각각 4개, 16개, 385개이다.

즉 인천연안에서 0.1 mg/l 정도 이하의 연평균 오염물질 추정농도 오차한계를 확보하기 위해서는 적어도 1일 1회 이상의 관측 자료가 필요함을 의미하며, 실질적으로는 연속 관측을 의미한다. 또한, 월 1회 정도의 관측 자료를 이용하여 COD 평균농도가 전(前)년도에 비하여 0.5 mg/l 정도 증가하였다 또는 감소하였다 하는 정도의 분석은 통계적으로 무의미하며 기본적으로 오차범위에 포함되기 때문에 타당한 분석이 아니다.

본 연구 분석 결과에 의하면 월 12회 또는 연 4회 정도의 관측에 의한 오차한계는 각각 1.13 mg/l, 1.96 mg/l 정도가 된다. 즉 COD 추정오차가 1.0 mg/l 정도이므로 1.0 mg/l 이하의 COD 농도변화 양상 분석은 통계적으로는 오차범위에 포함되기 때문에 무의미한 분석 또는 분석오류를 의미한다고 할 수 있다. 또한, 오염물질의 평균 농도가 증가할수록 표준편차가 증가하는 경향이 매우 강하기 때문에 동일한 수준의 오차범위를 유지하기 위해서는 오염물질이 유입되는 지점 또는 오염부하의 직접적인 영향범위에서는 의해 영역보다 빈번하게 또는 공간적으로 더욱 조밀하게 관측을 수행하여야 한다. 즉 관측빈도(밀도)도 오염 정도에 따라 공간적으로 조정할 필요가 있음을 시사한다.

6. 결론 및 제언

본 연구에서는 다양한 빈도로 관측된 연안 환경자료의 평균 및 표준편차 변화양상을 검토하고, 자료로부터 추정되는 모수(평균 및 표준편차)의 구간추정을 수행하여 신뢰구간에 따른 오차범위를 분석하였다. 모수 추정은 관측빈도, 관측지점에 따라 수행할 수 있기 때문에 다양한 조건에서 분석을 수행하였으며, 분석을 통하여 도출된 주요 결론은 다음과 같다.

(1) 연안 오염물질 COD, TN, TP 항목의 평균농도가 증가할수록 표준편차가 증가하는 양상을 뚜렷하게 보이고 있는 것으로 파악되었다. 즉 동일한 관측빈도(동일한 표본의 수)를 유지하는 경우에는 오염물질 농도가 높은 오염이 심한 지역의 오차범위가 보다 크게 나타날 수 있다. 또한, 오염물질 농도가 높은 지역은 오염물질 농도의 시간적·공간적인 편차가 크게 나타나는 것으로 추정할 수 있는바, 오염물질 농도가 안정된 수준을 유지하는, 즉 농도의 표준편차가 일정 범위 이하를 유지하는 지역은 배경농도 지역 또는 오염물질 확산범위의 기준으로 간주할 수도 있다.

(2) 인천연안의 공간적인 오염물질 농도변화 양상을 분석하기 위하여 약 4개 지점의 공간자료를 평균한 연 4회 정도의 빈도(국립수산과학원에서 수행하는 연안 환경자료의 통상적인 관측빈도)로 관측한 연안 환경자료의 연평균 농도 오차범위는 95% 신뢰수준을 기준으로 DO, COD, TN, TP 농도 각각 2.26 mg/l, 0.68~0.86 mg/l, 0.62~0.80 mg/l, 0.074~0.082 mg/l 수준으로 파악되었다. 월 1회 정도로 관측빈도를 증가시키는 경우 오차범위는 57.7% 정도 감소하는 것으로 추정된다. 그러나 공간적인 평균을 수행하지 않고, 월 1회 정도의 관측빈도를 유지할 경우에는 동일한 수준의 오차범위를 유지하게 된다.

(3) 시화호의 공간적인 오염물질 농도의 상관관계는 0.86~0.96 정도로 매우 크게 나타났으며, 해수교환에 의한 환경개선 목적으로 배수갑문을 개방하기 전에는 오염물질 농도가 높고 표준편차도 증가하여 오차범위가 크게 증가하였으나, 1997년 7월 배수갑문 개방 이후에는 오염물질 농도가 낮아지고 표준편차도 감소하여 오차범위가 작게 나타났다. 그러나 2003년 이후에는 관측빈도가 월 1회(연 12회)에서 연 4회로 감소하여 오차범위가 다시 증가하는 양상을 보이는 것으로 파악되었다. 시기적으로 변동양상이 있으나 수문개방 이후를 기준으로 하면 DO, COD, TN, TP 항목의 연평균 농도 오차범위는 각각 4.0 mg/l, 3.0 mg/l, 0.5~1.0 mg/l, 0.05 mg/l 정도로 파악되었다.

(4) 한편 실시간 해양수질측정망 자료를 이용한 SH1~SH3 지점의 오차범위는 오염이 심한 SH3 지점이 가장 크게 나타났으며, SH1 지점과 SH2 지점은 상대적으로 작게

나타났다. 관측빈도는 약 10분~2시간 간격으로 측정한 자료를 평균한 일 농도자료를 이용하였으므로 연 관측빈도는 365회를 크게 넘고 있으나, 일 자료이므로 연 365회로 간주하는 경우, DO, COD, TN, TP 항목의 연평균 농도 오차범위는 SH1 지점과 SH2 지점은 0.45~0.82 mg/l, 0.25~0.63 mg/l, 0.03~0.08 mg/l, 0.006~0.007 mg/l 범위이며, 오염이 심한 SH3 지점은, 0.9~1.0 mg/l, 1.11~1.47 mg/l, 0.19~0.21 mg/l, 0.022 mg/l 범위로 SH1 지점과 SH2 지점에 비하여 2~5배 정도로 크게 나타났다. 즉 SH1 지점과 SH2 지점을 기준으로 하는 경우 시화호 실시간 수질 자료의 연평균농도 오차범위는 연 12회 정도의 월 관측자료를 이용하여 추정된 오차범위에 비하여 DO, COD, TN, TP 항목은 각각 87~93% 정도 감소하는 것으로 파악되었다.

따라서 통상적으로 개발 사업에 따른 환경영향평가에서 수행되는 연 1~2회 또는 계절 관측 자료를 이용하여 오염물질의 평균변화 양상을 분석하는 경우는 분석이 무의미할 정도의 분석 오차를 포함할 수 있으며, 현재 통상적으로 사용되는 연 4회 정도의 연안환경 관측 자료도 COD 항목은 연평균 농도 오차범위가 3.0 mg/l 수준이므로 정확한 변화양상분석을 위해서는 관측빈도를 늘리거나, 오차범위를 파악하는 공간적인 범위를 확장하여 더욱 많은 지점을 포함하거나 장기간의 관측 자료를 이용하여 변화경향을 분석하는 경우에만 목표수준의 오차범위 이내로 농도추정이 가능하다. 한편, 오염물질이 유입되는 또는 체류되어 오염이 심한 해역의 오염물질 농도는 변동성이 크기 때문에 오염부하의 영향이 적은 외해에 비하여 더욱 빈번한 관측을 수행하여야 동등한 수준의 오차범위 수준을 유지할 수 있다.

사 사

본 연구는 한국해양연구원 기본연구사업 “하구 관리 및 기능회복 기술개발 연구(PE-981-1A)” 사업의 지원으로 추진되었습니다. 연구비 지원에 감사드립니다.

참고문헌

- 김우철, 김재주, 박병욱, 박성현, 박태성, 송문섭, 이상열, 이영조, 전종우, 조신섭. 2007. 통계학 개론, 4판. 영지문화사. 535 p.
- 물환경정보시스템. 2007. <<http://www.nier.go.kr/weis/>> [2007-10-10]
- 박성현. 1985. 회귀분석, 개정판. 대영사. 654 p.
- 이상열. 2001. 시계열 분석의 원리. 자유아카데미. 329 p.
- 조흥연, 조범준, 김상준. 2008. 계절별 염분 자료를 이용한 아산만 연안의 담수 영향범위 추정. 한국해안·해양공학회

- 논문집. 20(2), 219-231.
- 해양수산부. 국가해양환경통합정보시스템. <<http://www.meps.go.kr/tmsinfo/default.asp>> [2007-10-10].
- 해양수산연구정보. 2007. 국가해양측정망-해양환경. <<http://portal.nfrdi.re.kr/envirodata>> [2007-10-10].
- Agresti, A. and C. Franklin. 2007. Statistics: The art and science of learning from data. Pearson Prentice Hall. 693 p.
- Atkinson, A.C. 1985. Plots, transformations, and regression, an introduction to graphical methods of diagnostic regression analysis. Clarendon Press, Oxford. 282 p.
- Barnett, V. and T. Lewis. 1998. Outliers in statistical data, 3rd ed. Wiley, Chichester. 584 p.
- Berthouex, P.M. and L.C. Brown. 1994. Statistics for environmental engineers. Lewis Publishers, Boca Raton. 335 p.
- Cryer, J.D. 1986. Time series analysis. Duxbury Press, Boston. 286 p.
- Emery, W.J. and R.E. Thomson. 1998. Data analysis methods in physical oceanography. p. 193-304. In: *Statistical methods and error handling*, ed. by W.J. Emery and R.E. Thomson. Pergamon, Kidlington.
- Farnum, N.R. and L.W. Stanton. 1989. Quantitative forecasting methods. PWS-Kent Publishing, Boston. 573 p.
- Goda, Y. 2000. Random seas and design of maritime structures. In: *Advanced series on ocean engineering, v. 15*. 443 p.
- GraphPad Software. 2008. Grubbs' test for detecting an outlier. Available from WWW:<<http://www.graphpad.com/quickcalcs/Grubbs1.cfm>> [cited 2008-5-1].
- Haldar, A. and S. Mahadevan. 2000. Probability, reliability, and statistical methods in engineering design. Wiley, New York. 304 p.
- Harvey, A.C. 1993. Time series models, 2nd ed. The MIT Press. 308 p.
- Keith, L.H. 1988. Principles of environmental sampling, ACS professional reference book, ed. by L.H. Keith. American Chemical Society, Washington D.C. 458 p.
- Kottegoda, N.T. and R. Rosso. 1997. Statistics, probability, and reliability for civil and environmental engineers. McGraw-Hill, New York. 735 p.
- Montgomery, D.C. 2005. Introduction to statistical quality control, 5th ed. John Wiley, Hoboken. 759 p.
- Ott, R.L. and M. Longnecker. 2001. An introduction to statistical methods and data analysis. 5th ed. Duxbury. 1152 p.
- Spiegel, M.R. 1988. Schaumis outline of theory and problems of statistics. McGraw-Hill, New York. 504 p.
- US EPA. 1992. Monitoring guidance for the national estuary program. UA Environmental Protection Agency, EPA 842-B-92-004. 65 p.
- Ward, R.C., J.C. Loftis, and G.B. McBride. 1990. Design of water quality monitoring systems. Van Nostrand Reinhold, New York. 231 p.

Received May 2, 2008

Revised May 19, 2008

Accepted May 21, 2008