

부여 취수장의 NH₃-N 자료에 대한 평균 및 분산추정 Estimation of Mean and Variance for NH₃-N data of Puyeo Intake

김 형 수* / 정 건 희** / 김 응 석*** / 김 중 훈****

Kim, Hung Soo / Chung, GunHui / Kim, Eung Seok / Kim, Joong Hoon

Abstract

Sometimes the observed data is too small to discriminate it from noise of the instrument. Say, the data can be recorded as below DL(Detection Level) value. Even though the data below Detection Level(BDL) is small value, it can be resulted in wrong estimates for mean and variance. However, in practice, the BDL data is generally eliminated as N.D. (Not Detected) and do not record it in Korea. This study investigates the distributions according to the data values of ammonia concentration (NH₃-N) in Puyeo intake. Also we try to find out DL value and an appropriate method for the estimations of mean and variance of BDL values that can be discriminate the distributions. The DL is estimated by trial and error method. The appropriate method for the estimations of mean and variance of above the detection level(ADL)and BDL data sets is selected, and the mean and variance are estimated. As a result, it is found that the Bias Corrected Maximum Likelihood Estimator is the most accurate method for NH₃-N in Puyeo intake.

keywords : Detection Level, Below the Detection Level, Above the Detection Level, Maximum Likelihood Estimator

요 지

실험 또는 계측에 의해 관측된 관측치들은 종종 어떤 기준치 이하의 작은 값들이 기록되는데, 이들 기준치 이하의 값들이 크기는 미소할지라도 평균이나 분산 추정시 왜곡된 결과를 줄 수 있다. 그러나 우리 나라에서는 관측오차로 간주하여 N.D.(Not Detected)로 처리하는 것을 관례로 하고 있어 미소치들이 기록되지 않고 있다. 따라서 본 연구에서는 부여 취수장의 암모니아성 질소(NH₃-N)자료가 크기에 따라 분포형이 다름을 조사하고 그 분포를 구별할 수 있는 기준치와 기준치 이하의 자료들에 대한 평균과 분산 추정시 가장 적절한 기법을 찾고자 하였다. 즉, 기준치 이하의 값들과 이상의 값들을 구분하여 평균과 분산을 위한 적절한 기법을 선정하여 추정하였다. 분석 결과 부여취수장의 자료는 편기 수정된 최우도(Bias Corrected ML)법이 가장 적합한 것으로 결정되었으며, 시행착오법에 의하여 기준치를 설정하였다.

핵심용어 : 기준치, 기준치 이하의 자료, 기준치 이상의 자료, 최우도 추정자

* 선문대학교 토목공학과 조교수

Assistant Professor. Department of Civil Engineering. Sunmoon University(E-mail : sookim@mail.sunmoon.ac.kr)

** 고려대학교 토목환경공학과 석사과정

Graduate Student. Department of Civil and Environmental Engineering. Korea University

(E-mail : gunhui@korea.com)

*** 고려대학교 토목환경공학과 박사수료

Doctoral Candidate. Department of Civil and Environmental Engineering. Korea University

(E-mail : blue70@korea.ac.kr)

**** 고려대학교 토목환경공학과 부교수

Associate Professor. Department of Civil and Environmental Engineering. Korea University

(E-mail : jaykim@korea.ac.kr)

1. 서 론

실험 또는 계측에 의한 관측치는 가끔 기준치 (detection level, DL)보다 작은 값들이 존재한다. 이러한 값들은 수학적 관점에서 큰 영향을 주지 못하는 것처럼 생각되어 실무에서는 통상 N.D.(not detected)라고 처리하여 무시하는 경향이 있다. DL이라는 기준치를 사용하는 이유는 관측치가 아주 작은 값을 가질 경우에 이 값이 계측오차에 의한 값인지를 구별하기가 어려울 때가 있기 때문이다. 이와 같이 계측오차와 관측치를 구별할 수 없을 경우에 기준치보다 작은 값을 가지는 자료를 DL 이하의 자료(below the detection level, BDL)라고 하며, DL 이하의 관측치를 포함하고 있는 자료를 "Censored Data"라고 한다. 반면에 기준치보다 큰 값으로 구성되어 있는 자료를 DL 이상의 자료(above the detection level, ADL)라고 한다.

ADL 및 BDL에 대한 자료의 수가 주어져 있는 경우에 대하여 기준치 이하의 관측치를 포함하고 있는 자료들의 평균과 분산 추정은 다음과 같은 가정에 의하여 수행된다. 첫째는 ADL부분의 확률분포가 연속적이라는 것과, 둘째는 그 분포를 BDL부분까지 확장시킬 수 있다는 가정이며, 이들 자료를 분석하는 방법에는 크게 3가지가 있다. 즉, 첫째는 기계적인 오차나 실험오차로 간주하여 제거하거나, 0, DL 혹은 DL/2로 대체하여 사용하는 것이고, 둘째는 분포방법(distributional method), 셋째는 회귀방법(regression method)에 의하여 분석하는 것이다. 우리나라에서는 통상 첫번째의 방법을 이용하여 분석하는 것이 관계화되어 있다. 그러나, 이러한 방법은 모집단의 평균과 분산의 추정시 왜곡된 결과를 초래할 수 있으며, 특히, 미소한 값을 갖는 자료의 경우에는 그 영향이 더욱 크다고 할 수 있다. 따라서 본 연구에서는 기존의 삭제하거나 치환하는 방법 대신 최우도 추정(maximum likelihood estimation, MLE), one-step restricted MLE, Bias Corrected MLE 등과 같은 매개변수 추정방법을 적용하여 방법들간의 차이를 비교하고 자료계열에 가장 적합한 기법을 찾아 보고자 한다.

또한 현재 우리 나라에서는 이러한 자료를 분류하는 기준이 마련되어 있지 않은 점을 감안하여 서로 다른 형태의 분포형을 가지고 있는 자료들이 결합된 형태의 분포(censored distribution)를 2개의 분포형으로 구

분짓는 기준치 결정에 대한 방법론을 제시하고 이 방법론에 의해 결정되는 기준치를 제안하고자 한다.

기존의 연구를 살펴보면, censored data에 대한 초기의 최우도법의 적용은 Aitchison and Brown(1969), Cohen(1959), Newman et al.(1989), Schneider(1986)에 의하여 매개변수 추정이 이루어졌다. Cohen(1959)은 절단된 표본(truncated samples)과 censored samples의 분포형을 소개하고 최우도법에 의한 평균과 분산 추정을 제시하였다(1961). Newman 등(1989)은 censored distribution을 환경자료에 적용하였으며, Ahn(1998)은 미국 플로리다의 강우에 포함된 인자자료에 대하여 분석을 실시하였다.

2. BDL의 평균 및 분산 추정 기법

2.1. 분포방법

분포방법은 자료가 모두 주어진 분포형을 따르고 그 통계치는 자료의 확률분포함수를 구함으로써 산정할 수 있다는 가정하에 평균 및 분산과 같은 통계치들을 추정하는 것이다. 통상적으로 많이 이용되고 있는 최우도법(ML method)을 기술하면 다음과 같다.

관측치의 개수를 n 이라고 할 때, 시계열, $X = x_i, i = 1, \dots, n$ 을 고려하고 변량 X 는 대수정규분포(lognormal distribution)($LN(\mu, \sigma^2)$)를 따른다고 하자, 즉, $Y = y_i = \ln(x_i), i = 1, \dots, n$ 의 대수변환에 의해 $N(\mu_y, \sigma_y^2)$ 의 정규분포를 따르도록 한다. 계산상의 편의를 위해 내림차순으로 정리하고 계산을 시작한다. 그러나 우리는 DL 이하의 자료계열에 대한 분포형에 관심이 있으므로 전체 n 개의 자료들을 다음과 같이 나눌 수 있는데, 대수변환한 자료계열인 Y 에 대하여 $d = \ln(DL)$ 을 기준으로 다음과 같이 2개의 계열로 나눌 수 있다.

m 개의 $Y_B = \{y_i, i = 1, \dots, m, y_i < d\}$ 인 부분(censored observations)과 $(n-m)$ 개의 $Y_A = \{y_i, i = 1, \dots, m+1, \dots, n, y_i \geq d\}$ 인 부분(uncensored observations)으로 나누어진다. 여기서 d 보다 작은 값들의 계열은 BDL을 나타내고 d 보다 큰 자료들은 ADL을 의미한다. 모멘트법을 사용하여 구한 전체자료 Y 의 평균과 표준편차를 각각 \bar{y}, s 로 표시하고 역시 모멘트법에 의해 구한 ADL과 BDL부분의 평균과 표준편차는 각각 $\bar{y}_A, s_A, \bar{y}_B, s_B$ 라 정의하기로 한

다.

Y는 정규분포를 따르므로 확률밀도함수(probability density function; pdf)는 다음과 같은 형태를 가진다.

$$f(y) = \frac{1}{\sqrt{2\pi}s} \exp\left[-\frac{y_i - \bar{y}^2}{2s^2}\right] \quad (1)$$

누가분포함수(cumulative distribution function; cdf)는 $F(\xi) = \int f(y)dy$ 를 나타내며, 여기서

$\xi = d - \frac{\mu_y}{\sigma_y}$ 이다. 본 연구에서 BDL자료의 평균과 분산을 추정하기 위하여 이용한 최우도법들에 대한 식을 기술하면 다음과 같다.

2.1.1. 최우도 추정자

대수변환한 자료의 평균과 분산을 추정하기 위하여 Cohen(1959, 1961)은 최우도 추정자(ML estimator)인 식(2)을 이용하였다.

$$\hat{\mu}_y = \bar{y}_A - \lambda(\bar{y}_A - d), \quad \hat{\sigma}_A^2 + \lambda(\bar{y}_A - d)^2 \quad (2)$$

여기서, $\lambda = \lambda(h, \xi)$ 로 $h = \frac{m}{n}$ 의 함수이며 다음과 같은 함수들에 의하여 낼 수 있다.

$$\lambda(h, \xi) = \frac{Y(h, \xi)}{Y(h, \xi) - \xi},$$

$$Y(h, \xi) = \left[\frac{h}{1-h} \right] Z(-\xi),$$

$$Z(\xi) = \frac{f(\xi)}{1 - F(\xi)}$$

위의 식들은 Y에 대한 평균과 분산이므로 원자료 x의 평균과 분산은 식(3)에 의해 역변환을 함으로써 추정할 수 있다.

$$\hat{\mu} = \exp\left(\hat{\mu}_y + \frac{\hat{\sigma}_y^2}{2}\right), \quad \hat{\sigma}^2 = \hat{\mu}^2 \exp(\hat{\sigma}_y^2) - 1 \quad (3)$$

2.1.2. 편기 수정된 최우도 추정자

$n < 20$ 경우와 같이 자료의 수가 적은 경우에 시계열이 왜곡될 수 있으므로 그 영향을 줄이기 위하여 Shaw(1961)와 Schneider와 Weissfeld(1985)는 편기 수정된 최우도 추정자(bias corrected ML

estimator)인 식(4)과 (5)를 제안하였다.

$$\hat{\mu}_y = \bar{y} - \frac{s^2}{n+1} \left\{ 2,692 - 5.439 \left(\frac{n-m}{n-2m+1} \right) \right\} \quad (4)$$

$$\hat{\sigma}_y = s - \frac{s^2}{n+1} \left\{ 0.312 - 0.859 \left(\frac{n-m}{n+1} \right) \right\} \quad (5)$$

2.1.3. One - Step Restricted ML

BDL 표본들이 불연속적인 이항분포(binomial distribution)를 갖는다는 가정하에 Persson과 Rootzen(1977)은 다음과 같은 계열을 발생시켜 식(6)에 의하여 평균과 분산을 추정하였다.

$$C = c_i = y_i - d, \quad i = m+1, \dots, n$$

$$\hat{\mu}_y = \hat{\mu}_A - \alpha^* \sigma^* \sigma_y^2 = \frac{1}{k} \sum_{i=m+1}^N y_i^2 - \left(\frac{1}{k} \sum_{i=m+1}^N y_i \right)^2 - \alpha^* \varepsilon - (\alpha^*)^2 (\sigma^*)^2 \quad (6)$$

여기서,

$$\varepsilon = d - \frac{\bar{y}}{s}, \quad \alpha^* = \frac{nf(\varepsilon)}{k}, \quad k = n - m$$

$$\sigma^* = \frac{1}{2} \left[\varepsilon \frac{1}{k} \sum_{i=m+1}^N c_i + \left\{ \left(\varepsilon \frac{1}{k} \sum_{i=m+1}^N c_i \right)^2 + \frac{4}{k} \varepsilon \frac{1}{k} \sum_{i=m+1}^N c_i^2 \right\}^{\frac{1}{2}} \right]$$

2.2. ADL과 BDL의 조합 모집단에 대한 통계치

2.1절에서 구한 BDL자료들의 최우도법 결과치와 모멘트법을 사용하여 구한 ADL자료의 평균과 분산을 결합하여 각 시계열의 특성을 고려한 분포형의 평균과 분산을 아래의 식(7), (8)에 의하여 추정한다.

$$\hat{\mu} = h \hat{\mu}_B + (1-h) \hat{\mu}_A \quad (7)$$

$$\hat{\sigma}^2 = h \hat{\sigma}_B^2 + (1-h) \hat{\sigma}_A^2 + h(1-h)(\hat{\mu}_B - \hat{\mu}_A)^2 \quad (8)$$

여기서, $\hat{\mu}_A$ 와 $\hat{\sigma}_A$ 는 ADL자료의 평균과 표준편차로써 다음의 식(9)로 구할 수 있고 $\hat{\mu}_B$ 와 $\hat{\sigma}_B$ 는

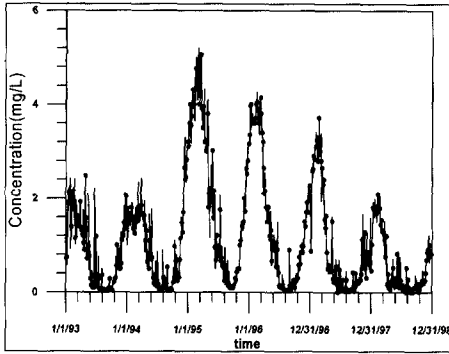


그림 1. 암모니아성 질소 자료의 시계열

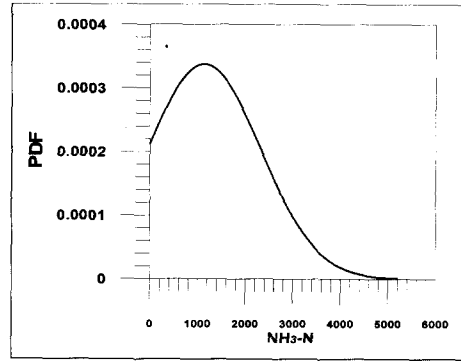


그림 2. 암모니아성 질소의 확률밀도함수

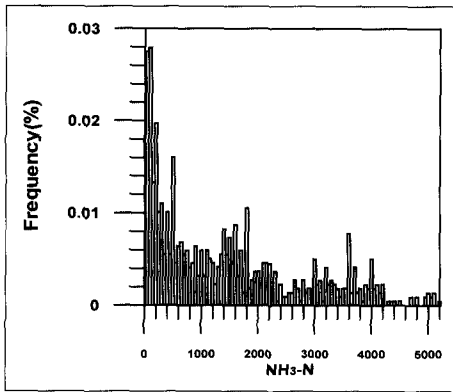


그림 3. 암모니아성 질소의 빈도분석

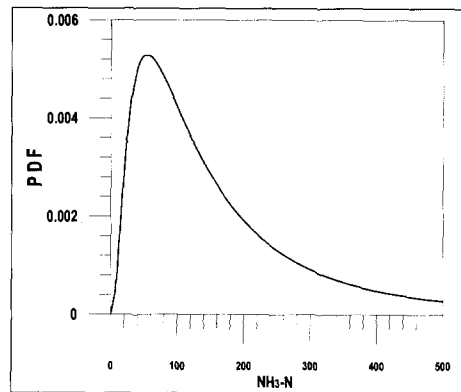


그림 4. 500 µg/L 이하의 값들에 대한 확률밀도함수

BDL자료의 평균과 표준편차를 나타낸다.

$$\hat{\mu}_A = \sum_{i=m+1}^N \frac{x_i}{n-m}, \quad \hat{\sigma}_A^2 = \sum_{i=m+1}^N \frac{[x_i - \hat{\mu}_A]^2}{n-m-1} \quad (9)$$

3. 부여 취수장의 NH₃-N자료에 대한 적용

3.1. 자료의 특성

우리 나라에서는 미소한 값들을 모두 N.D.로 처리하여 무시하기 때문에 기준치를 갖는 자료를 구하기가 어렵고, 앞에서 설명한 방법론들을 이용하여 분석한 예도 없는 것으로 사료된다. 따라서 본 연구에서는 부여 취수장에서 관측한 암모니아성 질소(NH₃-N)의 농도 자료에 대하여 그 값이 작은 자료들과 큰 자료들의 분포형태가 다르다는 점에 착안하여 작은 자료와 큰 자료의 분포형태를 구분할 수 있는 기준치를 결정하고자

한다.

본 연구에서는 자료의 수가 풍부한 부여 취수장의 암모니아성 질소(NH₃-N) 자료를 이용하기로 한다. 부여취수장의 수질자료 단위는 µg/L를 사용하였고 자료기간은 1993년 1월 1일부터 1998년 12월 31일까지 2,187개의 일 자료이며, 시계열은 그림 1에 나타내었다. 시계열을 그릴 때는 편의상 단위를 mg/L를 사용하여 도시하였다.

전체 자료의 분포를 살펴보면 최소값 0.007µg/L에서부터 최대값 5,200µg/L이고 평균과 표준편차가 각각 1,137.45µg/L과 1,182.33µg/L를 가지는 대수정규분포로 가정하여 확률밀도함수를 그려보면 그림 2와 같이 왼쪽 꼬리부분이 두껍게 절단된 형태의 분포를 보이고 있다. 실제 왼쪽 꼬리 부분이 길게 확장되어 있어야 하나 미소값들이 삭제되어 있는 상황을 감안할 때 충분히 가능한 분포형임을 알수 있다. 또한 그림 3의 빈도분석결과를 살펴보면 작은 값들의 빈도가 현저히 높

표 1. NH₃-N 자료의 기초 통계치

		전체자료($\mu\text{g/L}$)	BDL ($\leq 500\mu\text{g/L}$)	ADL ($\geq 500\mu\text{g/L}$)
원자료 X _i	평균	1137.531	177.956	1861.290
	표준편차	1182.137	136.241	1103.839
대수 변환한 자료 Y _i	평균	6.273	4.829	7.357
	표준편차	1.458	0.916	0.593

표 2. DL=500 $\mu\text{g/L}$ 를 가정한 경우 전체자료(BDL+ADL)의 평균과 표준편차에 대한 오차

방 법	결 과 ($\mu\text{g/L}$)		전체자료의 오차 (%) (=100Obs.-Est./Obs.)	
	평균	표준편차	평균	표준편차
관 측 치	1137.531	1182.137	-	-
최우도 추정자	1130.271	1189.251	0.813	0.602
편기 수정된 최우도 추정자	1143.109	1182.754	0.314	0.052
One-Step Restricted ML	1151.348	1182.892	1.037	0.064

음을 알 수 있다. 그러므로 질소자료에는 기준치가 존재하지 않지만 임의의 값인 500 $\mu\text{g/L}$ 을 이 자료의 기준으로 가정하여 분석을 시도하였다. 이런 가정의 타당성을 입증하기 위하여 500 $\mu\text{g/L}$ 보다 작은 값들만을 사용하여 확률밀도함수를 그려본 결과 전체 자료와는 매우 상이한 분포를 나타내고 있음을 알 수 있었다. 즉, 그림 4에서 볼 수 있는 것처럼 기준치 이하의 값들은 대수정규분포를 나타내고 있음을 알 수 있다. 그러므로 임의의 기준치인 500 $\mu\text{g/L}$ 보다 작은 고빈도의 값들의 분포형을 따로 분리하여 모의한 후에 결합하는 시도는 타당한 것으로 사료된다.

임의의 기준치인 500 $\mu\text{g/L}$ 를 기준으로 하여 분석한 기초통계량은 표 1과 같고 2장의 기법들을 적용하여 분석한 결과는 표 2와 같다. 표 2에서 ADL자료는 정규분포로, BDL자료는 대수정규분포로 가정하여 2장의 방법을 적용하여 결합한 모형의 결과를 보면 전체 자료에 대하여 편기 수정된 최우도(Bias Corrected ML)법이 가장 잘 맞는 것으로 나타나 있다. 그러나 그림 2와 3을 근거로 가정한 DL=500 $\mu\text{g/L}$ 는 인위적인 값이므로 정확한 DL을 산정하기 위하여 다음의 3.2절에서 시행착오법에 의하여 기준치 DL을 추정하였다.

3.2. 자료의 기준치 설정

부여 취수장의 자료를 사용하여 각각 기준치를 100 $\mu\text{g/L}$ 에서부터 750 $\mu\text{g/L}$ 까지 변화시켜가면서 2장에 언급된 세 가지 방법으로 평균과 표준편차를 구하여 실제 관측치와의 오차를 계산하였다. 이 결과 중에서 각 기준치와 적용방법별로 평균과 표준편차의 오차가 가장 작은 경우를 구하여 그 때의 기준치를 적절한 DL로 선정하도록 하였다. 각 방법별로 DL에 따라 평균 및 표준편차에 대한 오차를 산정한 결과 평균은 260 $\mu\text{g/L}$ 에서 가장 작은 오차를 나타내었고 표준편차는 490 $\mu\text{g/L}$ 에서 가장 작은 오차를 나타내었다 (그림 5(b) 참조). 두 경우 모두 앞에서 임의의 기준치에 적용하였을 경우와 같이 편기 수정된 최우도 추정방법에서 가장 작은 값을 나타내었다. 결과가 이렇게 산출되었으므로 이 두 가지 기준치에 대해서만 결과를 표로 작성하였다. 표 3은 각각의 경우에서의 평균과 표준편차 추정 값들이고 표 4는 각각의 경우에 계산결과인 BDL과 ADL자료를 결합한 전체 자료에 대한 오차의 값을 나타낸 것이다.

위의 표 4에서 알 수 있듯이 전체자료에 대하여 오차를 추정한 결과를 기준으로 생각할 때 260 $\mu\text{g/L}$ 와 490 $\mu\text{g/L}$ 의 두 기준치에서 모두 편기 수정된 최우도

표 3. 최소오차를 나타낼 때의 각 방법에 따른 결과비교

기준치가 260 $\mu\text{g/L}$ 일 경우의 방법별 평균과 표준편차의 결과				
방 법	BDL자료의 결과($\mu\text{g/L}$)		전체자료의 결과($\mu\text{g/L}$)	
	평 균	표준편차	평 균	표준편차
관측치	177.956	136.241	1137.531	1182.137
최우도 추정자	107.747	88.898	1135.850	1183.852
편기 수정된 최우도 추정자	118.971	106.542	1139.530	1181.143
One-Step Restricted ML	133.747	151.162	1144.374	1178.570
기준치가 490 $\mu\text{g/L}$ 일 경우의 방법별 평균과 표준편차의 결과				
방 법	BDL자료의 결과($\mu\text{g/L}$)		전체자료의 결과($\mu\text{g/L}$)	
	평 균	표준편차	평 균	표준편차
관측치	177.956	136.241	1137.531	1182.137
최우도 추정자	154.095	152.347	1133.047	1187.090
편기 수정된 최우도 추정자	176.405	194.289	1142.279	1182.052
One-Step Restricted ML	194.766	253.766	253.766	1180.542

표 4. 각 방법별 오차의 비교

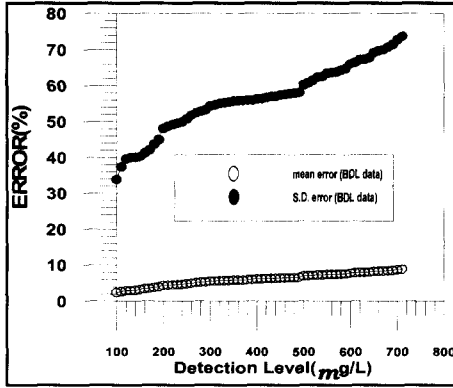
D.L.	자료형	방 법	평 균(%)	표준편차(%)
260 $\mu\text{g/L}$	BDL	최우도 추정자	4.85797	26.7065
		편기 수정된 최우도 추정자	5.05361	51.85498
		One - Step Restricted ML	18.10079	115.4509
	전체자료	최우도 추정자	0.32301	0.14509
		편기 수정된 최우도 추정자	0.00006	0.08416
		One - Step Restricted ML	0.42503	0.30171
490 $\mu\text{g/L}$	BDL	최우도 추정자	6.86761	23.95616
		편기 수정된 최우도 추정자	6.61566	58.08222
		One - Step Restricted ML	17.71318	106.48430
	전체자료	전체자료최우도 추정자	0.56897	0.41902
		편기 수정된 최우도 추정자	0.24116	0.00724
		One - Step Restricted ML	0.90794	0.13492

추정방법에 의한 오차가 가장 작은 것으로 나타났다. 이것은 앞의 임의의 기준치인 500 $\mu\text{g/L}$ 를 적용한 결과와 같은 결과이다.

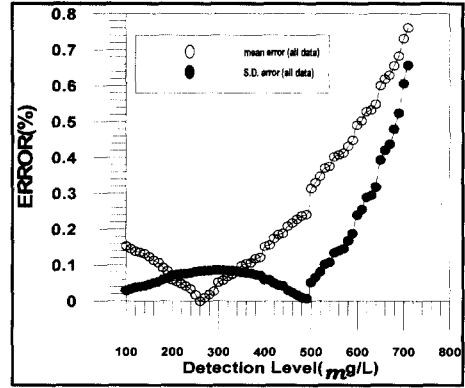
그러므로 그림 5에서 편기 수정된 최우도 추정방법

에 의한 오차를 BDL자료와 전체 자료에 대하여 각 기준치별로 그래프로 나타내어 보았다.

그림 5(a)를 보면, BDL자료는 DL이 작아져 BDL 자료의 수가 작아질수록 평균과 표준편차의 오차가 줄



(a) BDL자료



(b) 전체자료

그림 5. (a) BDL자료와 (b) 전체자료의 DL에 따른 오차비교

어둡고 있음을 알 수 있다. 이것은 내림차순으로 정리 되어 있는 상태에서 자료의 수가 작으면 작을수록 편차가 작아지므로 나타나는 당연한 결과라 할 수 있다. 그러나 그림 5(b)에서는 전체 자료의 평균과 표준편차의 추정치가 DL에 의해 현저히 차이가 남을 볼 수 있다. 즉, 평균의 경우에 $260\mu\text{g/L}$, 표준편차의 경우에는 $490\mu\text{g/L}$ 에서 가장 작은 결과를 보이고 있음을 알 수 있다. 이 때의 오차는 $260\mu\text{g/L}$ 에서 평균오차가 0.00006%, $490\mu\text{g/L}$ 에서 표준편차의 오차가 0.00724%로 매우 작은 값을 나타내므로, 전체 자료를 잘 모의하고 있음을 알 수 있다.

마지막으로, 여기서 제시된 2개의 기준치 중에서 하나를 선택하는 문제가 남아있다. 자료의 분포형을 2개로 나눈 것은 미소값을 가지는 BDL자료의 정확한 모의를 통한 전체자료의 분포형을 정확히 모의하기 위한 것이었으므로 2개의 기준치 중에서 전체자료의 평균오차가 작은 $260\mu\text{g/L}$ 가 더욱 적합한 것으로 사료된다. 또한 이것은 평균을 훨씬 잘 모의하고 있고, 표준편차의 오차 역시 상당히 작은 값을 나타내므로 타당할 것으로 판단된다. 그러므로 본 연구에서 이용한 $\text{NH}_3\text{-N}$ 자료에 대한 DL은 $260\mu\text{g/L}$ 임을 설정할 수 있을 것이다.

따라서 DL을 $260\mu\text{g/L}$ 로 결정하고 편기 수정된 최우도 추정자가 가장 적합한 방법으로 선정되었다. 이 때 표 3에서 전체자료에 대해 관측치와 편기 수정된 최우도 추정자에 의한 평균을 비교하여 보면 왜곡되어 있음을 알 수 있다. 즉, 관측치 자체에 의한 평균과 ADL과 BDL로 분리하여 분포형을 고려한 평균이 다

름을 알 수 있다. 미소 자료가 많은 자료에 대해서는 이러한 영향이 더욱 클 것으로 판단이 된다.

4. 결 론

본 연구에서 분석한 기법들은 대상이 되는 자료들이 DL을 갖는 자료들로서 기준이 되는 값을 경계로 하여 작은 크기의 자료(BDL)와 큰 값을 갖는 자료(ADL)가 각기 다른 통계적 특성을 가질 때 적합한 방법이라 하겠다. 그러나 본 연구에서 이용한 암모니아성 질소자료의 경우에 실제 정확한 DL을 갖는 자료는 아니나 자료의 특성을 살펴 본 결과, 자료 크기에 따라 통계적 특성을 달리하고 있어 연구의 가치가 있다고 판단하였다. 분석 결과, 질소자료의 분포를 가장 잘 표현해주는 기준치를 찾을 수 있었고 이 때에도 역시 임의의 기준치를 적용한 결과와 같이 편기 수정된 최우도법이 가장 적절한 방법으로 결정되었다.

이와 같은 분포형의 적용이 일반화되면 환경자료 같은 미소농도를 가지고 있는 자료에 대한 더욱 더 정밀한 적용이 이루어질 수 있고 미소농도의 기여도 역시 더욱 중요하게 다루어질 것으로 생각된다. 또한 우리나라에서도 미소한 값들을 N.D. 처리하지 않고 관측치를 기록한다면 기준치를 결정함에 있어서 본 연구의 내용들을 이용할 수 있을 것으로 사료된다.

본 연구 결과 하나의 분포형으로 잘 모의되지 않을 경우에는 어떤 기준치를 경계로 하여 2개의 분포형으로 분리한 후 모의를 시도하면 실제 자료의 분포를 거의 유사하게 모의할 수 있음을 알 수 있었다. 물론 아직은 실제로 직접 모의해보지 않고서는 그 기준치를

알 수가 없으나, 이런 매개변수를 추정된 후에 분포형을 모의한다면 더욱 정확한 모형을 구축할 수 있을 것으로 생각된다. 또한 평균과 표준편차가 본 연구의 분석에서 처럼 각기 다른 DL에서 최소의 오차를 가질 경우에 표준편차의 오차는 DL의 값이 작을수록 작아지기 때문에 평균 오차를 기준으로 하여 DL을 선정함이 옳을 것이며, 또한 DL을 고려하였을 때 자료의 평균에 영향이 있음을 알 수 있었다.

참 고 문 헌

- Aichison, J. and J. A. C. Brown, (1957). *The Lognormal Distribution*, Cambridge University Press, Cambridge, Massachusetts.
- Cohen, A. C., Jr.,(1959). Simplified Estimates for the Normal Distribution When Samples Are Singly Censored or *Truncated*. *Technometrics* 1(3), pp. 217~237.
- Cohen, A. C., Jr., (1961). Tables for Maximum Likelihood Estimates: Singly Truncated and Singly Censored Samples. *Technometrics* 3(4), pp. 535~541.
- Hosung Ahn, (1998), Estimating the Mean and Variance of Censored Phosphorus Concentrations in Florida Rainfall, *Journal of the American Water Resource Association*, Vol. 34 No. 3, pp. 583~593.
- Newman, M. C., P. M. Dixon, B. B. Looney, and J. E. Pinder, III, (1989). Estimating Mean and Variance for Environmental Samples With Below Detection Limit Observations. *Water Resources Bulletin* 25(4), pp. 905~916.
- Persson T. and H Rootzen, (1977). Simple and Highly Efficient Estimators for a Type I Censored Normal Sample. *Biometrika* 64(1), pp. 123~128.
- Shaw, J.G., (1961). The Bias of the Maximum Likelihood Estimates of the Location and Scale Parameters Given a Type II Censored Normal Sample. *Biometrika* 28, pp. 448~451.
- Schneider, H. and L. Weissfeld, (1986). Inference Based Type II Censored Samples. *Biometrics* 42, pp. 531~536.
- (논문번호:00-002/접수:2000.01.13/심사완료:2000.04.24)