

회귀식 사용에 따른 화학 분석 과정의 불확도 처리 연구

우진춘* · 서정기 · 임명철 · 박민수*

*한국표준과학연구원, 물질량표준부.

대전대학교, 환경공학과

(2002. 10. 11 접수, 2003. 4. 9 승인)

A Study on the Treatment of Uncertainty in Linear Regression Method for Chemical Analysis

Jin-Chun Woo*, JungKee Suh, MyungChul Lim and MinSu Park

Division of Chemical Metrology and Materials Evaluation, Korea

Research Institute of Standards and Science(KRISS), Daejeon, Korea

*Department of Environmental Engineering, Daejeon University, Daejeon, Korea

(Received Oct. 11, 2002, Accepted Apr. 9, 2003)

요 약 : 회귀식 사용에 따른 불확도 계산의 정확성을 조사하기 위하여, 수정된 방법의 최소제곱법 (Modified Least Square Method, MLS)과 회귀식에서 일반적으로 적용되는 불확도 처리 과정을 각각 1차 식에 적용하고 비교하였다.

회귀식에서 일반적으로 적용되는 불확도 처리 과정에서, 대부분의 경우 불확도 값이 크게 계산되고 있어 확률적으로 안전한 범위를 표기할 수 있는 것으로 평가되었다. 그러나, 표준시료 농도의 상대 표준불확도가 클 때 (교정점이 우발적으로 흩어지는 정도의 표준 편차가 5% 수준, 표준시료 농도의 상대 표준불확도가 7% 수준) 회귀식에서 일반적으로 적용되는 불확도 계산 방법으로 얻은 값이 고 농도 측정에서 매우 작게 평가되고 있어 확률적으로 매우 위험한 것으로 평가되었다. 이 경우, 통계학적으로 불확도를 정확히 계산하기 위하여, 수정된 방법의 최소제곱법이 유리하다고 판단하였다.

Abstract : We applied modified least square method (MLS) and ordinary least square method (OLS) to 1st order equation for the comparison of the uncertainties calculated by these methods. The uncertainty calculated by OLS covered statistically safe interval because it was over-estimated in many cases of measurement and concentration level. But, if the uncertainty of the concentration as a reference value was comparably large (about 5% of the relative standard deviation of random scattering from the regression line and about 7% of relative standard uncertainty of reference values), then uncertainty calculated by OLS was seriously under-estimated at high concentration level. It was revealed that the

★ Corresponding author

Phone : +82+(0)42-868-5364 Fax : +82+(0)42-868-5042

E-mail : jcw@krissol.kriss.re.kr

calculated uncertainty didn't cover statistically safe interval at the stated confidence level. It was found that the method, MLS, described in the previously article would be valid for this calculation of uncertainty.

Key words : uncertainty, linear regression, calibration curve

1. 서 론

측정 결과를 비교하기 위해서는 측정 결과의 생산자가 이에 대한 품질, 즉 정확성을 정량적인 숫자로 표시하는 것이 선결 조건이다. 이러한 필요성에 따라 1993년에 국제표준화기구(ISO)가 여러 국제기구와 합동으로 '측정 불확도 표현 지침서 (Guide to the Expression of Uncertainty in Measurement, GUM)'를 발행하였다. 이 후, 이 지침서는 측정 결과의 품질과 정확성을 나타내기 위한 절차서로서 측정 및 화학 분석의 전 분야에 적용되어 왔다.¹⁻⁵

GUM에 따라 측정 결과의 불확도를 계산하기 위해서는, 1 관계 모델식의 설정, 2 요인별 표준불확도의 산출, 3 합성표준불확도의 계산 4 확장불확도의 계산과 같은 4 단계를 거쳐야 한다. 이 단계에서, 측정학적인 관점에서 중요한 것은 1 단계와 2 단계이며, 3 및 4 단계는 순수하게 계산 과정이라고 볼 수 있다. 특히, 관계 모델식의 설정 과정에서, 각각의 불확도 요인들을 잘 검토한 후, 각 입력량과 해당 표준불확도가 잘 산출될 수 있도록 수학적 식을 설정하는 것은 측정학적인 관점에서 매우 중요한 일이다.

화학 분석 과정에서 많이 이용되는 교정식 또는 검정곡선식의 작성과 관련하여 회귀 분석을 이용하는 경우, 일반적인 최소제곱법 (Ordinary Least Square method, OLS)⁶의 불확도 계산처럼, 기준값 (화학 분석의 경우 표준 시료의 농도)에 불확실성이 없다고 가정할 수 없기 때문에 불확도 계산을 위한 수학적 식을 설정하는데 어려움이 있다. 본 논문은 측정 및 분석 (특히, 화학 분석) 과정의 불확도를 구하기 위하여 수정된 방법의 최소제곱법 (Modified Least Square method, MLS)⁷과 OLS의 불확도 처리 과정을 1차 회귀식에 대하여 적용하고 비교하였다. 이러한 비교 검토를 위하여 화학 분석 과정에서 얻을 수 있는 전형적인 몇 가지의 자료를 임의로 설정한 후, 각각의 불확도 처리 방법에 적용하고 결과들을 비교하였다.

2. 불확도 합성 방법의 검토

화학 분석에서 몇 개의 표준시료를 가지고 일반적인 최소제곱법 (OLS)을 이용하여 1 차의 회귀식을 작성하고 이 식을 이용하여 농도를 구하는 과정은 Fig. 1과 같으며, 이 과정에서 작성된 검정곡선의 식은 다음과 같이 설정할 수 있다.⁶

$$y = a + bx \quad (1)$$

여기서, a , b 는 회귀식의 계수

화학 분석에서 실질적으로 식 1을 이용하는 것은 역추정으로서, 측정 대상 시료의 반응량 (y_{obs})을 측정하여 식 2와 같이 농도 (x_{obs})를 구하는 것이다.

$$y_{obs} = a + bx_{obs} \quad (2)$$

여기서 y_{obs} 는 반응값

x_{obs} 는 농도

여기서 회귀식의 계수는 OLS에 따라 구할 수 있으며, 최종적인 식은 다음과 같다.⁶

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} \quad (3)$$

여기서 n 는 표준 시료의 개수

y_i 는 표준 시료 i 의 반응량

x_i 는 표준 시료 i 의 농도

이러한 과정에서 측정량은 측정 대상 시료의 농도와 반응량 (y_{obs})이고, 이것의 불확도를 구하는 것이 최종적인 목적이다. 먼저, 정량화하여 구하기 용이한 입력량들의 표준불확도를 정리하면 다음과 같다.

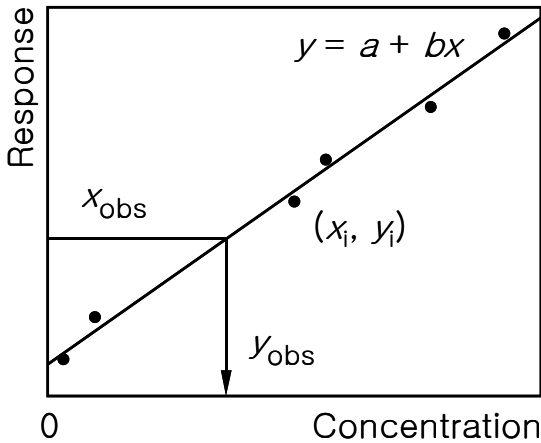


Fig. 1. Graphic illustration of a calibration curve.

- 1) 각 표준시료 농도 (x_i)의 표준불확도
- 2) 각 표준 시료 반응량 (y_i)의 표준불확도
- 3) 측정 대상 시료 반응량 (y_{obs})의 표준불확도이다.

그러나 OLS에서는 표준시료 농도에 불확실성이 없고, 표준 시료 반응량의 불확실성이 같은 분산을 보인다는 가정을 적용하게 된다. 이러한 가정에 의해 구해진 검정곡선식을 사용하여 분석하는 경우, 계산과정에서 고려하여야 하는 불확도 요인을 정리하면 다음과 같다.

- 1) 검정곡선 주위에서 각각의 교정점이 우발적으로 분산되는 불확도 요인
- 2) 모든 표준시료의 농도가 일률적으로 가지고 있는 계통적 불확도 요인

이 두 종류의 불확도 요인 중에서 각각의 교정점에서 우발적으로 분산되는 불확도 요인은 정량화 가능하고 이것을 반응량의 단위로 계산하면 다음과 같다.⁶

$$u(y_i) = s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} \quad (4)$$

여기서, \hat{y}_i 는 각 표준시료 농도값을 검정곡선식에 대입하여 구한 반응값이다. 그러나, 두 번째의 불확도 요인을 정량화하기 위해서는 표준 시료의 제작 과정을 검토하여야 하며, 실질적으로 이것을 정량화하는 것은 용이한 일이 아니다. 그러나, 표준 시료 농도의 불확실성이 반응값의 불확실성보다 상대적으로 매우 클 때는 이 요인이 첫 번째 불확도 요인에 포함되기 때문에 이 요인의 합성을 고려하지 않아도 무방할 것이다.

이러한 여러 종류의 불확도 요인을 이용하여 합성

표준불확도를 계산할 수 있는 가능한 3 가지 방법을 선택하여 불확도의 약식 계산방법 (SIM), OLS에 의한 불확도 계산방법 (OLS-1 및 2) 및 MLS에 의한 불확도 계산방법 (MLS-1 및 2)이라고 명명하였으며, 각각의 구체적인 적용 과정은 다음과 같다.

2.1. 불확도의 약식 계산방법 (SIM)

최종 측정량인 측정 대상 시료 농도의 합성표준불확도를 각각의 점들이 우발적으로 흩어지는 불확도와 같다고 가정하면, 합성표준불확도는 다음과 같을 수 있다.

$$u_c(x_{obs}) = \frac{1}{b} \cdot \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-2}} \quad (5)$$

모든 표준시료의 농도가 일률적으로 가지고 있는 계통적 불확도가 각각의 점들이 우발적으로 흩어지는 불확도에 비해 매우 작은 경우, 잘 적용할 수 있을 것이며, OLS 방법의 약식 계산으로 적용 가능하다.

2.2. OLS에 의한 불확도(OLS)⁶

일반적인 최소제곱법에서의 계수의 계산은 식 2 및 3이 이용되며, 불확도 계산을 위한 일반적인 적용 조건은 표준 시료의 농도가 기준량으로서 불확실성이 없고, 반응량들은 표준 시료별로도 같은 수준의 분산 상태를 보인다는 것이다. 그러므로 표준 시료 농도의 표준불확도 ($u(x_i)$)는 0이고, 표준시료들의 반응량에 대한 표준불확도 ($u(y_i)$) 및 미지 시료 반응량의 표준불확도 ($u(y_{obs})$)는 식 4와 같다. 따라서, 식 2와 3을 관계 모델식으로 설정하고, 입력량들의 표준불확도를 불확도의 전과 공식에 따라 합성하면, 출력량들(a , b 및 y_{obs})의 합성표준불확도를 구할 수 있다.⁴ 이들의 결과로부터 측정량인 대상 시료의 농도 (y_{obs})의 합성표준불확도는 다음과 정리할 수 있다.

$$u_c(y_{obs}) = \sqrt{\frac{s^2}{b^2} \left\{ \frac{1}{m} + \frac{1}{n} + \frac{(y_{obs} - \bar{y})^2}{b^2 \sum (x_i - \bar{x})^2} \right\}} \quad (6)$$

여기서 m 은 측정 대상 시료의 반응값을 결정하기 위한 측정 횟수이며 n 은 사용한 표준 시료의 수이다. 측정 대상 시료를 1회 측정한 경우와 무한히 측정할 경우에 대하여 검토하기 위하여 각각을 OLS-1 (식 7 참조) 및 OLS-2 (식 8 참조)라 명명하였다.

$$u_c(y_{obs}) = \sqrt{\frac{s^2}{b^2} \left\{ 1 + \frac{1}{n} + \frac{(y_{obs} - \bar{y})^2}{b^2 \sum (x_i - \bar{x})^2} \right\}} \quad (7)$$

$$u_c(y_{obs}) = \sqrt{\frac{s^2}{b^2} \left\{ \frac{1}{n} + \frac{(y_{obs} - \bar{y})^2}{b^2 \sum (x_i - \bar{x})^2} \right\}} \quad (8)$$

이러한 계산 방법도 SIM 방법과 같이, 모든 표준시료의 농도가 일률적으로 가지고 있는 계통적 불확도가 각각의 점들로부터 우발적으로 흩어지는 불확도에 비해 매우 작은 경우에 적용이 가능하다.

2.3. MLS에 의한 불확도(MLS)⁷

OLS에서 가정하는 분석 조건이 만족되는 과정의 불확도는 앞에서 설명한 바와 같이 계산될 수 있지만, 이러한 조건이 만족되지 않는 경우에는 불확도의 계산이 어렵게 된다. 이에 따라, OLS에 의해 교정식이나 검정곡선식을 구하여 이용하는 일반적인 화학분석 과정에서 표준시료 농도값의 불확실성이 큰 경우에 모든 불확실성이 합성될 수 있는 알고리즘을 발표한 바 있다.⁷

MLS에 의한 계산 방법에서는 식 2 및 3을 관계 모델식으로 설정하지만, 특별히 식 2에 오차량의 변수를 추가하여 식 9와 같이 관계식을 변경하였다.

$$y_{obs} = a + bx_{obs} + \tau \quad (9)$$

여기서, 검정곡선식의 변수는 입력량의 값으로부터 계산되며, 일반 최소제곱법의 변수들과 같은 값을 가진다. 측정량의 합성표준불확도의 계산을 위하여 모든 입력량, 즉, 각 표준시료의 농도와 표준시료 및 측정 대상 시료의 반응값의 표준불확도는 각각 정량화되어 있어야 한다. 특별히, 오차량, τ 는 검정곡선으로부터 위의 불확실성에 의한 양을 제외하고 구할 수 있기 때문에 그 기대값은 0 이고 표준불확도는 다음과 같이 구한다.

$$u^2(\tau) = \left(\frac{\sum (y_i - \hat{y}_i)^2}{n-2} \right) - \left(\frac{\sum \{u^2(y_i) \times v_{y_i}\}}{\sum v_{y_i}} \right) - b^2 \cdot \left(\frac{\sum \{u^2(x_i) \times v_{x_i}\}}{\sum v_{x_i}} \right) \quad (10)$$

여기서 $u^2(y_i)$ 는 표준시료 i의 농도의 표준불확도
 $u^2(x_i)$ 는 표준시료 i의 반응량의 표준불확도

v_{y_i} 는 표준시료 i의 농도의 자유도

v_{x_i} 는 표준시료 i의 반응량의 자유도

측정량 (yobs)인 시료의 농도를 구하고 불확도 전과 공식을 유도하기 위해서는 모든 입력량의 값과 표준편차를 가지고 관계 모델식 3 및 9를 이용하고 행렬식을 이용하여 계산한다.⁴

3. 합성 방법에 따른 합성표준불확도의 비교

앞에서 검토된 3가지 불확도 계산 방법에 대하여 숫자로 계산된 실질적인 결과를 비교하기 위하여 Table 1과 같이 임의로 자료를 작성하였다. 검정곡선 작성을 위한 교정점의 수는 5이고, 각각의 표준시료농도에 대응하는 반응값을 설정하였다. SIM 및 OLS-1 및 OLS-2의 경우, 표준시료 농도 및 반응량의 표준불확도 자료는 사용되지 않는다. 그러나, MLS 방법에 적용하기 위하여, 반응값의 표준불확도 수준은 대략 1%으로, 표준시료 농도의 표준불확도는 각각 3%와 7% 수준에서 2 별로 작성되었다. 표준시료 농도의 표준불확도는 각각 별개로 적용하였으며, 3%로 처리된 결과를 MLS-1로, 7%로 처리된 결과를 MLS-2로 각각 명명하였다.

Table 1. Artificial data for the comparison of uncertainty

| Standard sample, No. | Concentration | | | Response | |
|----------------------|-----------------------------|------------------------------|-------|---------------------|---------------------------|
| | Concentration value (mg/kg) | Standard uncertainty (mg/kg) | | Response value (AU) | Standard uncertainty (AU) |
| | | MLS-1 | MLS-2 | | |
| 1 | 0.52 | 0.02 | 0.036 | 334 | 3 |
| 2 | 5.10 | 0.25 | 0.34 | 822 | 8 |
| 3 | 9.95 | 0.32 | 0.7 | 1232 | 13 |
| 4 | 15.24 | 0.47 | 1.0 | 1911 | 20 |
| 5 | 20.31 | 0.69 | 1.4 | 2367 | 21 |

표의 자료를 가지고 최소제곱법을 적용하고 도시하면 관계식의 기울기는 103.7 kg/mg , 절편은 272.4 그

리고 상관계수는 0.983이다. 각각의 교정점이 식에서 분산되는 정도를 간단하게 식 4와 같은 방법으로 표시하면, 반응량의 단위로 55.4이고 농도 단위로 0.534 mg/kg 수준이다.

측정 대상 시료의 여러 농도준에 대하여도 각각의 불확도를 구하여 비교하기 위해서 여러 수준의 반응값을 임의로 선정하였다. 특히 MLS 방법에 의해서는 반응값과 함께 표준불확도의 값도 필요하므로 반응값의 1% 수준에서 표준불확도를 임의로 작성하였으며, 복잡한 불확도 계산을 위하여 보조 프로그램인 PUMA를 이용하였다.⁸ 여러 계산 방법, 즉, SIM, OLS-1, OLS-2, MLS-1 및 MLS-2에 의해 계산된 합성표준불확도는 Table 2과 같으며, 이것을 Fig. 2에 도시하였다.

Table 2. Combined standard uncertainties ($u_c(x_{\text{obs}})$) at various calculation procedures and response levels

| y_{obs} ,AU | $u(y_{\text{obs}})$,AU | Measurand (x_{obs}),mg/kg | Combined standard uncertainty, $u_c(x_{\text{obs}})$, mg/kg | | | | |
|-------------------------|----------------------------|---|--|-------|-------|-------|-------|
| | | | SIM | OLS-1 | OLS-2 | MLS-1 | MLS-2 |
| 0 | 0 | -2.63 | 0.534 | 0.730 | 0.498 | 0.412 | 0.514 |
| 100 | 1 | -1.66 | 0.534 | 0.711 | 0.469 | 0.394 | 0.460 |
| 300 | 3 | 0.27 | 0.534 | 0.676 | 0.414 | 0.365 | 0.359 |
| 500 | 5 | 2.19 | 0.534 | 0.646 | 0.363 | 0.345 | 0.274 |
| 800 | 8 | 5.09 | 0.534 | 0.611 | 0.296 | 0.336 | 0.221 |
| 1100 | 11 | 7.98 | 0.534 | 0.590 | 0.251 | 0.354 | 0.296 |
| 1400 | 14 | 10.87 | 0.534 | 0.586 | 0.240 | 0.396 | 0.438 |
| 1700 | 17 | 13.76 | 0.534 | 0.598 | 0.267 | 0.454 | 0.605 |
| 2000 | 20 | 16.65 | 0.534 | 0.625 | 0.324 | 0.523 | 0.773 |
| 2400 | 24 | 20.51 | 0.534 | 0.682 | 0.423 | 0.626 | 1.005 |

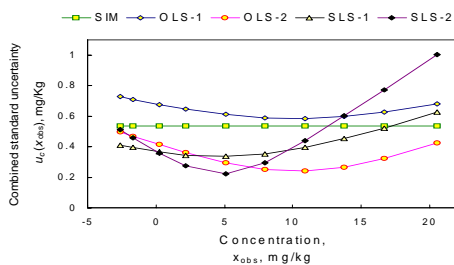


Fig. 2. Combined standard uncertainties ($u_c(x_{\text{obs}})$) of measurand (x_{obs}) at various calculation procedures and response levels.

표에서와 같이 측정 대상 농도가 -2.63 ~ 20.51 mg/kg에서 비교 검토하였다. 표준물첨가법에서와 같이 반응값이 음수인 경우에도 값과 합성표준불확도의 계산이 중요하므로 이 농도 수준의 결과를 비교 대상에 포함시켰다. Table 2 및 Fig. 2에 나타난 결과들을 다음과 같이 정리할 수 있다.

- 1) 교정점이 우발적으로 흩어지는 정도가 5% 수준이고 표준시료의 농도의 상대 표준불확도가 3% 수준, 반응값의 상대 표준불확도가 1% 수준일 때, OLS-1과 MLS-1을 비교하면, 모든 농도 범위에서 OLS-1의 결과가 크게 평가되는 것을 알 수 있다.
- 2) 교정점이 우발적으로 흩어지는 정도가 5% 수준이고 표준시료의 농도의 상대 표준불확도가 7% 수준, 반응값의 상대 표준불확도가 1% 수준일 때, OLS-1과 MLS-2를 비교하면, OLS-1이 낮은 농도 범위에서 과대하게 평가되며, 높은 농도에서 과소 평가되는 것을 알 수 있다.
- 3) 교정점이 우발적으로 흩어지는 정도가 5% 수준이고 표준시료의 농도의 상대 표준불확도가 3% 수준, 반응값의 상대 표준불확도가 1% 수준일 때, SIM과 MLS-1을 비교하면, 중간 농도 수준에서 SIM 방법이 OLS-2 보다 매우 과대하게 평가되는 것을 알 수 있다.
- 4) 결과 3과 불확도 표기의 안전성과 간편성을 고려해 볼 때 SIM 방법이 유효하나, MLS-2의 결과, 즉, 표준시료의 농도의 상대 표준불확도가 7% 수준으로 높을 때는 높은 농도에서 과소 평가되어 안전하지 못하다.
- 5) 표준물 첨가법은 반응값이 0이고 이 반응량의 표준 불확도가 0 인 경우에 해당되므로, 이것을 비교하기 위하여 SIM, OLS-2, MLS-1의 결과를 비교할 수 있으며, 모든 결과값들이 매우 유사한 값을 보여주고 있다.

4. 결 론

측정 및 화학분석 과정에서 회귀식을 이용하는 경우, 일반적인 최소제곱법의 불확도 계산 과정에서처럼, 표준 시료의 기준값에 불확실성이 없다고 가정할 수 없기 때문에 여러 불확도 산출 방법으로 계산된 결과값을 비교 연구하였다.

회귀식에서 일반적으로 적용되는 불확도 처리 과정에서, 대부분의 경우 불확도 값이 크게 계산되고 있어 확률적으로 안전한 범위를 표기할 수 있는 것으로 평가된다. 표준물 첨가법이 적용되는 화학분석에서는 여러 종류의 불확도 계산 과정에서 유사한 결과 값을 보여주고 있어 계산된 불확도가 비교적 정확하다고 생각된다. 그러나, 표준시료 농도의 상대 표준불확도가 클 때 (교정점이 우발적으로 흩어지는 정도의 표준 편차가 5% 수준이고, 표준시료 농도의 상대 표준불확도가 7% 수준), 회귀식에서 일반적으로 적용되는 불확도 계산 방법으로 계산된 불확도 값이 고 농도 측정에서 매우 작게 평가되고 있어 확률적으로 매우 위험한 것으로 평가된다.

최소제곱법에 의한 여러 종류의 불확도 계산 결과에서처럼, 표준 시료의 기준값에 불확실성이 없다고 가정할 수 없기 때문에 불확도를 보다 정확히 계산하기 위해서는 수정된 방법의 최소제곱법이 유리할 것으로 생각된다.

참고 문헌

1. "Guide to the Expression of Uncertainty in Measurements", ISO, (1993).
2. "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results," NIST, Technical Note 1297, NIST, (1993).
3. "Quantifying Uncertainty in Analytical Measurements," EURACHEM, (1995).
4. "Grundlagan der Messtechnik," DIN 1319-4
5. "측정불확도 표현 지침," KRIS-99-070-SP, 한국표준과학연구원, (1999).
6. N. L. Johnson and F. C. Leone, "Statistics and Experimental Design in Engineer and Physical Science," 2nd. Ed., Vol. I, p 128, Wiley, New York, U. S. A., (1977).
7. Jung-Ki Suh, Jin-Chun Woo, Dong-Min Moon, Minsoo Park, *Bull. Korean Chem. Soc.*, **22**, 247, (2001)
8. PUMA, 불확도 계산 컴퓨터 보조 프로그램, Version 2, KISCOM, (2002).