Scaling Factor Calculation with Robust Regression

J.W. Park [a], K.J. Lee [a], S.C. Lee [a], Y.H. Lee [a], J.K. Lee [a], S.Y. Kim [a], M.H. Ahn [a], T.W. Kim [b]
[a]*Nuclear Environmental Engineering Lab, KAIST*
[b]*Nuclear Engineering & Technology Institute, KHNP*

1. Introduction

The radioactivities of specific nuclides in waste packages have to be declared in accordance with limits and criteria derived from safety assessment of the disposal facility. Some of these nuclides are difficult to measure from the outside of the waste packages because they are alpha or beta emitting nuclides. The scaling-factor method is widely applied in order to evaluate these difficult-to-measure nuclides. The scaling-factor method is based on a correlation between easily measurable gamma emitting nuclides(key nuclides) and difficult-to-measure(DTM) nuclides. Scaling-factor has been calculated by the statistical processing of radiochemical analysis data. Current statistical methods to calculate scaling-factor are geometric mean and the linear regression based on least square. However both methods have a chance to distort the scaling-factor because they are vulnerable to outliers especially in small number of data points. In this study, robust regression (Least Median square, Least Trimmed Square and Reweighted Least Square) is suggested and applied to calculate scaling-factor. These results are compared with the scaling-factors calculated by current statistical methods.

2. Robust Regression

Least Median Square (LMS) replaces the sum of the Least Square by a median. It turns out that this estimator is very robust with respect to outliers in Y as well outliers X. Least Trimmed Square(LTS) formula is very similar to Least Square. The only difference being that the largest squared residuals are not used in the summation, thereby allowing the fit to stay away from the ouliers.

Statistical Method	Methodology
Geometric Mean (GM)	$f_{SF}=\sqrt[n]{(a_{d,1}/a_{k,1}\times\cdots\times a_{d,i}/a_{k,i}\times\cdots\times a_{d,n}/a_{k,n})}$
Least Square (LS)	$Minimize \sum_{i=1}^{n}(r_i)^2$
Least Median Sqaure (LMS)	$Minimize\ med\ (r_i)^2$
Least Trimmed Square (LTS)	$Minimize \sum_{i=1}^{h}(r^2)_{i:n}$

Table 1 Comparison of Statistical Methods

3. Application of Robust Regression

Before application of LMS and LTS, the radiochemical data set was roughly divided into 3 data groups (strong, weak and reasonable relationship).

3.1. LMS & LTS Methods

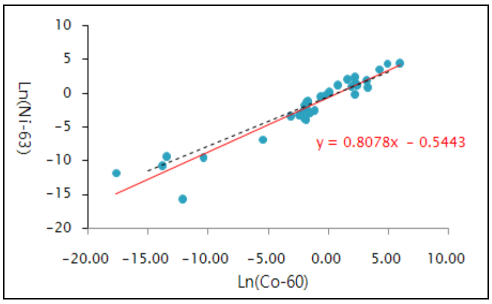

Fig. 1. SF Plot between ^{63}Ni and ^{60}Co
(33data points, SR, EPRI-5077, Correlation coefficient: 0.927)

Statistical Methods	Estimator
Geometric mean	$Y=X+\ln(0.812)$
Least square	$Y=0.808X+0.544$
Least median sqaure	$Y=0.737X+0.405$
Least trimmed square	$Y=0.737X+0.405$

Table 2. Estimators of statistical methods
(33data points, SR, EPRI-5077, Correlation coefficient: 0.927)

As could be expected in such data set, only marginal differences exist between the robust and those based on least squares. However if correlation coefficient between DTM and Key is unreasonably low value (under0.3), applicability of SF method should be reconsidered.

41 data points sampled from Uljin dry active waste have been used to determine the scaling-factor.

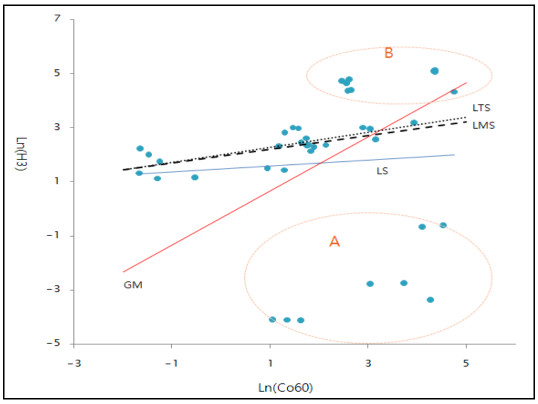

Fig. 2. SF Plot between ^{3}H and ^{60}Co
(41data points, DAW, Correlation coefficient : 0.35)

Statistical Methods	Estimator
Geometric mean	$Y = X + \ln(0.343)$
Least square	$Y = 0.108X + 1.474$
Least median sqaure	$Y = 0.250X + 1.995$
Least trimmed square	$Y = 0.276X + 1.994$

Table 3 Estimators of statistical methods
(41data points, DAW, Correlation coefficient: 0.350)

The basic principal of LMS and LTS is to fit the majority of the data, after which outliers may be identified as those points that lie far away from the robust fit, that is, the case with large positive or large negative residuals. Therefore both group A and B have been rejected during scaling factor calculation with LMS and LTS.

3.2. Reweighted Least Square(RLS) Method

The basic principal of LMS and LTS is to fit the majority of the data. However, in general the y_i may be in any unit of measurement, so in order to decide if a residual r_i is "large" we need to compare it to an estimate $\hat{\sigma}$ of the error scale. of course, this scale estimate $\hat{\sigma}$ has to be robust itself, so it depends only on the "good" data and does not get blown up by the outliers.

$$w_i = \begin{cases} 1 & \text{if } |r_i/\hat{\sigma}| \le 2.5 \\ 0 & \text{if } |r_i/\hat{\sigma}| > 2.5 \end{cases}$$

The bound 2.5 is, of course, arbitrary, but quite reasonable because in a Gaussian situation there will be very few residuals larger than $2.5\hat{\sigma}$. Reweighted least squares defined by

$$Minimize \sum_{i=1}^{n} w_i r_i^2$$

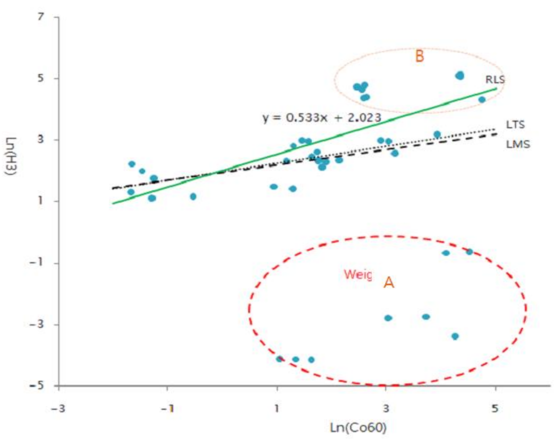

Fig. 3 SF Plot between ^{3}H and ^{60}Co
(LMS, LTS amd RLS)

The rejection of group A have an effect on correlation coefficient. Correlation coefficient increased from 0.350 to 0.766 and the efficiency (used data/total data) also increased from 50% to 78%. The relative error bound is also reduced to 23.4±12.4.

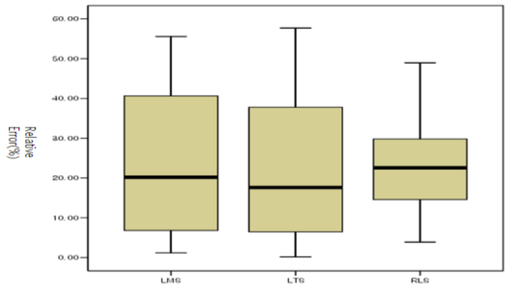

Fig. 4. Relative error of LMS, LTS and RLS

3. Conclusions

LMS and LTS have been suggested for the calculation of scaling factors, because these methods are less vulnerable to outliers than current methods. It can be used to determine more reliable scaling factors of the DTM nuclides that have relatively weak relationship with Key nuclide. However, LMS and LTS use only half data during scaling factor calculation. Therefore its maximum efficiency would be around 0.5. To compensate for the low efficiency of LMS and LTS methods, Reweighted Least Square method was also suggested. RLS method has the merits of both robust and current methods. So this method can be applicable to determine scaling factors of the important fission product nuclides(^{129}I, ^{99}Tc), the corrosion product nuclides(^{14}C, ^{3}H) and transuranic nuclides.

REFERENCES

[1] Peter j. Rousseeuw, Annick m. Leroy "Robust Regression and Outlier Detection" 1987

[2] W. T. Best and A. D. Miller, "Radionuclide correlations in low-level radwaste" Electric Power Research Institute Report, EPRI-4037, Impell Corporation, Califonia, 1985

[3] W. T. Best, "Update scaling factors in low level radwaste" Electric Power Research Institute Report, EPRI-5077, Impell Corporation, Califonia, 1987

[4] S. M. Ahn, "A Ststistical Study on Scaling Factors for Radionuclide Assay", M.S. Thesis, KAIST, 1993

[5] J. P. Park "Robust regression and multiple outliers detection " 정보통신연구소 제 2 집

[6] D. H. Lee, Y. S. Park and K. H. Kim, "로버스트 회귀추정에 의한 신뢰구간 구축" 응용통계연구 제 19 권 1 호, 2006