INTERNATIONAL STANDARD

ISO 18315

First edition
2018-11

Nuclear energy — Guidance to the evaluation of measurement uncertainties of impurity in uranium solution by linear regression analysis

Énergie nucléaire — Lignes directrices pour l'évaluation des incertitudes de mesure des impuretés en solution d'uranium par analyse de régression linéaire

Reference number
ISO 18315:2018(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 85, *Nuclear energy, nuclear technologies, and radiological protection*, Subcommittee SC 5, *Nuclear installations, processes and technologies*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Nuclear energy — Guidance to the evaluation of measurement uncertainties of impurity in uranium solution by linear regression analysis

1 Scope

This document provides a method for evaluation of the measurement uncertainty arising when an impurity content of uranium solution is determined by a regression line that has been fitted by the "method of least squares". It is intended to be used by chemical analyzers.

Simple linear regression, hereinafter called "basic regression", is defined as a model with a single independent variable that is applied to fit a regression line through n different data points (x_i, y_i) (i = 1,..., n) in such a way that makes the sum of squared errors, i.e. the squared vertical distances between the data points and the fitted line, as small as possible. For the linear calibration, "classical regression" or "inverse regression" is usually used; however, they are not convenient. Instead, "reversed inverse regression" is used in this document[2].

Reversed inverse regression treats y (the reference solutions) as the response and x (the observed measurements) as the inputs; these values are used to fit a regression line of y on x by the method of least squares. This regression is distinguished from basic regression in that the x_i's (i = 1,..., n) vary according to normal distributions but the y_i's (i = 1,..., n) are fixed; in basic regression, the y_i's vary but the x_i's are fixed.

The regression line fitting, calculation of combined uncertainty, calculation of effective degrees of freedom, calculation of expanded uncertainty, reflection of reference solutions' uncertainties in the evaluation result, and bias correction are explained in order of mention. Annex A presents a practical example of uncertainty evaluation. Annex B provides a flowchart showing the steps for uncertainty evaluation. In addition, Annex C explains the use of weighting factors for handling non-uniform variances in reversed inverse regression.

NOTE 1 In the case of classical regression, the fitted regression line is inverted prior to actual sample measurement[3]. In the case of inverse regression, the roles of x and y are not consistent with the convention that the variable x represents the inputs, whereas the variable y represents the response. For these reasons, the two regressions are excluded from this document.

NOTE 2 The term "reversed inverse regression" was suggested taking into account the history of regression analysis theory. Instead, it can be desirable to use some other term, e.g. "pseudo-basic regression".

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
calibration
fitting of a regression line of *y* on *x* through *n* data points using the method of least squares

Note 1 to entry: The *n* data points are typically obtained by measuring *n* different reference solutions. After the fitting, the fitted regression line is used as a measurement formula for determining the physical or chemical quantity of a sample.

3.2
calibration quality control factor
factor that is used to check the adequacy of the fitted regression line from the aspect of calibration quality

3.3
calibration uncertainty
uncertainty due to such possible variations of the slope and intercept supposing that the regression line fitting is repeated according to the same procedure using a "new set of *n* different reference solutions" each time

Note 1 to entry: In this case, the fitted regression line will be different each time, i.e. the slope and intercept of the fitted regression line will vary.

3.4
combined uncertainty
uncertainty obtained by combining the *calibration uncertainty* (3.3) and the *random uncertainty of sample measurement* (3.9) according to the error propagation rule

3.5
effective degrees of freedom
degrees of freedom calculated by Welch-Satterthwaite approximate formula

3.6
expanded uncertainty
multiplication of the combined uncertainty u_y by a coverage factor *k* given depending on the effective degrees of freedom of the *combined uncertainty* (3.4)

Note 1 to entry: The probability that the true value of the physical or chemical quantity will be within ± the "final expanded uncertainty" from the determined and bias-corrected value is "exactly" or "approximately" 95 %; in most cases, "approximately" is more suitable expression.

3.7
predicted *y* value
y value of a point on the fitted regression line

Note 1 to entry: The predicted *y* value given by the regression line formula $y = a + bx$ indicates the physical or chemical quantity that will be determined in response to the light or current signal intensity "*x*" measured by instrument. The square root of the estimate for the variance of the predicted *y* value is treated as the calibration uncertainty in this document.

3.8
prediction interval
possible vertical distance between the additional data point and the previously fitted regression line after a regression line has been fitted using a set of *n* data points

Note 1 to entry: Another data point is additionally produced by measuring "a new reference solution".

3.9
random uncertainty of sample measurement
<ICP-AES> possible uncertainty that may arise during such a sample measurement supposing that the intensity of the light with a specific wavelength emitted from a sample is measured to determine the sample's impurity content following the regression line fitting

Note 1 to entry: This uncertainty is typically estimated based on multiple measurements of the sample and should not be inferred from the mean squared error.

3.10
uncertainty
concept corresponding to the square root of "variance (or estimate for that variance)" that is handled mainly in statistics

3.11
weighting factors
numbers by which the variances are multiplied in weighted least squares (WLS) regression

Note 1 to entry: OLS (ordinary least squares) regression handles the uniform variances. However, in WLS regression, the variances are assumed to be non-uniform and the weighting factors are used to handle the non-uniform variances.

Note 2 to entry: Let $\sigma_{x1}^2,\ldots, \sigma_{xn}^2$ be the variances of the variables $x_1,\ldots, x_n$ respectively. Then, in OLS regression, $\sigma_{x1}^2 = \cdots = \sigma_{xn}^2$ $(= \sigma_x^2)$, whereas, in WLS regression, typically $\sigma_{x1}^2 \neq \cdots \neq \sigma_{xn}^2$. However, even in the case of WLS regression, the equality $w_1\sigma_{x1}^2 = \cdots = w_n\sigma_{xn}^2$ $(= \sigma_x^2)$ can be established by utilizing the weighting factors w_i's $(i = 1,\ldots, n)$.

4 Principle

In practice, a single regression line fitting will be run for calibration. Nevertheless, supposing that the regression line fitting is repeated using a "new set of n different reference solutions" each time, the slope and intercept of the fitted regression line will be different each time; as a result, it can be observed that the predicted y value varies according to an approximately normal distribution. Such a possible variation of the predicted y value is the cause of the calibration uncertainty that arises during the regression line fitting. In this regard, there is a helpful statistical theory. With the aid of that theory, although calibration experts do not repeat the regression line fitting, they can estimate the variance of the predicted y value using only one set of n data points that have been used for a regression line fitting. In practical calibrations, the square root of such an estimated variance of the predicted y value is treated as the calibration uncertainty, and its degrees of freedom is $n - 2$.

Two components contribute to the uncertainty of the physical or chemical quantity determined by the fitted regression line. One is the calibration uncertainty previously mentioned and the other is the random uncertainty of sample measurement. The two uncertainties are combined and then the effective degrees of freedom of the combined uncertainty is calculated. The effective degrees of freedom, V_{eff}, is obtained by substituting the degrees of freedom of the calibration uncertainty and the degrees of freedom of the random uncertainty together into Welch-Satterthwaite approximate formula.

As the next step, the expanded uncertainty is obtained by multiplying the combined uncertainty u_y by a coverage factor k that is given depending on the effective degrees of freedom of the combined uncertainty.

Finally, the uncertainties of the reference solutions that have been used for the fitting of the regression line are reflected in the uncertainty evaluation.

NOTE 1 The "new set of n different reference solutions" means newly manufactured and prepared n different reference solutions, however, their intended physical or chemical quantities (e.g. target chemical concentrations) and confidence levels are identical to those of the "previous set of n different reference solutions". Only in this case, the variances of the x_i's (i.e. $\sigma_{x1}^2,\ldots, \sigma_{xn}^2$) can be observed correctly. In many cases, it is not possible to use a new set of reference solutions (or standards) for calibration each time; that is, the same set of n different reference solutions (or standards) can be used each time. However, it does not cause a problem in the regression line fitting and calibration uncertainty evaluation. Here, the supposition of the repetition of regression line fitting is only to explain the concept of the calibration uncertainty.

NOTE 2 In basic regression, the slope, intercept and predicted y value follow exactly normal distributions; however, in reversed inverse regression, they follow approximately normal distributions[2].

NOTE 3 The square root of the estimated variance of the "predicted y value", not "prediction interval", is treated as the calibration uncertainty.

5 Uncertainty evaluation

5.1 Regression line fitting

As the first step, a regression line is fitted by OLS regression; basically, in this document, the variance of the observed x value is assumed to be uniform (or roughly uniform) over the calibration range of interest. For the regression line fitting, n data points, i.e. $(x_1, y_1),\ldots, (x_n, y_n)$, are used. In the fitted regression line formula $y = a + bx$, the independent variable x represents the light or current signal intensity to be measured by instrument and the dependent variable y represents the physical or chemical quantity, such as impurity content or chemical concentration, to be determined in response to the measured signal intensity x.

$$y = a + b\,x$$

$$a = \bar{y} - b\bar{x} = \sum y_i/n - b\sum x_i/n$$

$$b = S_{xy}/S_{xx} = \sum(x_i - \bar{x})(y_i - \bar{y})/\sum(x_i - \bar{x})^2$$

$$MSE = \sum(y_i - a - b\,x_i)^2/(n - 2)$$

Estimator for the variance of predicted y value = $\{1/n + (x - \bar{x})^2(S_{yy}/S_{xy}^2)\} \cdot MSE$[2]

Estimator for the variance of prediction interval = $\{1 + 1/n + (x_i - \bar{x})^2(S_{yy}/S_{xy}^2)\} \cdot MSE$[2]

where

a and b are the intercept and slope of the fitted regression line respectively;

$\sum$ denotes summation from $i = 1$ to n;

MSE denotes the mean squared error;

$\bar{x} = \sum x_i/n$;

$\bar{y} = \sum y_i/n$;

$S_{xx} = \sum(x_i - \bar{x})^2$;

$S_{yy} = \sum(y_i - \bar{y})^2$;

$S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y})$.

NOTE 1 At times, it is assumed that the variance of the observed x value is unduly non-uniform over the calibration range. A weighting method to handle the non-uniform variances in reversed inverse regression is explained in Annex C.

NOTE 2 In basic regression, the estimator for the variance of the predicted y value is $\{1/n + (x - \bar{x})^2 (1/S_{xx})\}\cdot MSE$, and the estimator for the variance of the prediction interval is $\{1 + 1/n + (x_i - \bar{x})^2(1/S_{xx})\}\cdot MSE$[4]. Thus, the two estimators for basic regression are different from those for reversed inverse regression, and the magnitudes of the differences depend on the correlation coefficient between x and y, i.e. $r(x, y) = S_{xy}/(S_{xx}S_{yy})^{1/2}$.

5.2 Adequacy check of fitted regression line

The fitted regression line is checked to identify its adequacy. For this purpose, the predicted y values corresponding to the x_i values (i = 1,..., n) are calculated and then the differences between the "y_i values" and the "predicted y values" are checked.

$$|\Delta y_i| = |y_i - (a + bx_i)|, (i = 1,..., n)$$

All of these differences shall be smaller than the square root of the mean squared error (MSE) multiplied by 1,2, i.e. $(1,2)(MSE)^{1/2}$. If any $|\Delta y_i|$ is greater than this limit, the fitted regression line is considered to be inadequate for further use as a measurement formula. In this case, the reference solutions and instrument need to be investigated and a new regression line fitting is performed. If necessary, the calibration range is readjusted and/or new reference solutions are prepared.

NOTE 1 This check is to identify whether there have been any serious errors during reference solution measurements and whether the measured values imply any possibility of unduly non-uniform variance in the calibration range.

NOTE 2 In $(1,2)(MSE)^{1/2}$, the value "1,2" can be regarded as a calibration quality control factor. This factor is determined in line with the required calibration quality level and can be smaller than 1,2 or greater. For example, 1,05, 1,1 or 1,25 can be selected as the factor.

NOTE 3 The correlation coefficient between x and y, i.e. $r(x, y)$ $\{ = S_{xy}/(S_{xx}S_{yy})^{1/2}\}$, can also be estimated. If this estimated coefficient meets a limit that was established in advance, it can be said that the linear relationship between the physical or chemical quantity and the measured signal intensity is being maintained in the instrument to the extent required. In general, the more precise the instrument is, the closer to 1 the coefficient will be.

5.3 Combined uncertainty

After the regression line fitting, the physical or chemical quantity of a sample is determined by substituting the measured light or current intensity x into the regression line formula $y = a + bx$. To evaluate the uncertainty of the physical or chemical quantity thus determined, the calibration uncertainty occurred during regression line fitting and the random uncertainty of sample measurement are first combined as follows.

$$u_y^2 = u_{cal}^2 + u_{ran}^2$$

$$u_{cal}^2 = \{1/n + (x - \bar{x})^2(S_{yy}/S_{xy}^2)\}\cdot MSE$$

$$u_{ran}^2 = b^2 u_x^2$$

$$\therefore u_y^2 = \{1/n + (x - \bar{x})^2(S_{yy}/S_{xy}^2)\}\cdot MSE + b^2 u_x^2 \quad (1)$$

where

u_y is the combined uncertainty;

u_{cal} is the calibration uncertainty occurred during the regression line fitting;

u_x is the uncertainty of the measured light or current signal intensity;

u_{ran} is the random uncertainty of sample measurement ($u_{ran} = bu_x$);

MSE is the mean squared error;

$\bar{x}$ is the mean of x_i values ($\bar{x} = \sum x_i/n$, $i = 1,\ldots, n$);

x is the light or current signal intensity.

NOTE The random uncertainty of the sample measurement, i.e. u_{ran} (= bu_x), is typically obtained by repeating the sample measurement m times according to the same procedure. If the standard deviation of the m measured signal intensities is S_x, u_x is $S_x/(m)^{1/2}$ and the degrees of freedom of u_x is $m - 1$. If only one sample is taken and the sample is measured only once, Type B evaluation of uncertainty can be applied[1]. This random uncertainty of sample measurement should not be inferred from the mean squared error, because the mean squared error involves all the statistical fluctuations arising when manufacturing and preparing the reference solutions.

5.4 Effective degrees of freedom

The effective degrees of freedom of the combined uncertainty, V_{eff}, is calculated using Welch-Satterthwaite approximate formula[1].

$$(1/V_{eff})u_y^4 = (1/V_{cal})u_{cal}^4 + (1/V_x)(bu_x)^4$$

$$\therefore V_{eff} = u_y^4/\{(1/V_{cal})\, u_{cal}^4 + (1/V_x)(bu_x)^4\} \quad (2)$$

where

V_{eff} is the effective degrees of freedom of the combined uncertainty u_y;

V_{cal} is the degrees of freedom of the calibration uncertainty u_{cal} ($V_{cal} = n - 2$);

V_x is the degrees of freedom of the measured signal intensity's uncertainty u_x ($V_x = m - 1$) (V_x is equal to the degrees of freedom of bu_x.).

For reference, the maximum of the possible effective degrees of freedom is $(n - 2) + (m - 1)$, and the minimum is whichever is the smaller of $n - 2$ and $m - 1$.

