XVII IMEKO World Congress
Metrology in the 3rd Millennium
June 22–27, 2003, Dubrovnik, Croatia

UNCERTAINTY RELATED WITH THE USE OF LINEAR REGRESSION ANALYSIS FOR THE CORRECTION OF CALIBRATED INSTRUMENTS

A. S. Ribeiro[†], *J. O. Mimoso*[†], *J. A. Sousa*[‡] *and M. P. Castro*[†]

[†] *National Laboratory of Civil Engineering, Lisbon, Portugal*
[‡] *Madeira Regional Laboratory of Civil Engineering, Funchal, Portugal*

Abstract

Linear Regression Analysis (LRA) is a technique commonly applied in many different branches of science. The present study investigates the use of LRA in Metrology and aims to develop a mathematical approach to adequately take into account its contribution for the uncertainty budget in a measurement.

In a calibration involving many standards and measuring instruments, the LRA technique is an important tool for the estimation of conventional true values based on certificate results. This statistical treatment usually intends to reduce the errors measured in the calibration process in order to achieve lower residual errors. The operation, however, introduces statistical uncertainties, which can be of significance when compared with the uncertainty contributions from other input quantities.

This document also presents the results of a measurement uncertainty evaluation related to the calibration of a length measuring machine, including the LRA contribution based on the application of the mathematical expression proposed. The relative influence of this contribution is also investigated.

Keywords: Calibration, Linear Regression Analysis, Measurement Uncertainty

1. INTRODUCTION

The calibration of instruments used for metrological purposes results in measurement errors and related uncertainties as presented in calibration certificates. The subsequent use of such instruments usually suggests the correction of readings so that the systematic effects determined by the calibration may be reduced. To implement this procedure Linear Regression Analysis is usually applied.

As this technique affects the measurement results, it must be considered as an additional source of uncertainty to be taken into account in the measurement uncertainty budget and, therefore, a model for its evaluation is needed.

The present study investigates the use of LRA in Metrology and aims to develop a practical mathematical expression to adequately take into account its contribution for the uncertainty budget in a measurement. An application to the measurement uncertainty evaluation related to the calibration of a length measuring machine and the results obtained (namely, measurement uncertainties and the relative influence of this input quantity) are also discussed in the paper.

2. THE MATHEMATICAL MODEL

The further use of calibrated standards and measuring instruments for metrological purposes commonly involves the need of correction of the readings using LRA statistical treatment.

In this type of analysis it is common to use the reference values, x_i (conventional true values with standard uncertainty given by ± i_x), and the average values of readings, y_i (Figure 1), taken from an instrument under calibration, both expressed in a certificate, to estimate the correction of the instrument readings (within the interval of calibration) using the expression of the least squares method given by:

$$Y_i = \beta_0 + \beta_1 X_i \qquad (1)$$

Thus having an estimate of the coefficients (β_0 and β_1) enables an estimate of the corrected values to be evaluated, according to the next equation.

$$y_i = b_0 + b_1 x_i \qquad (2)$$

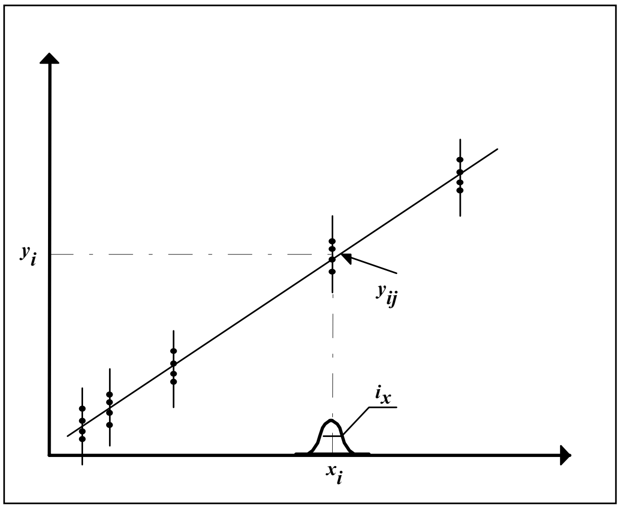

Fig. 1 - LRA Reference values, x_i, measurement averages, y_i, and reference values uncertainty, i_x.

Now substituting $b_0 = \bar{y} - b_1\bar{x}$ and imposing the translation of x=0 to x=$\bar{x}$ (in order to eliminate the covariance between the two coefficients b_0 and b_1), transforms equation (2) into expression (3).

$$y_i = \bar{y} + b_1(x_i - \bar{x}) \tag{3}$$

where the coefficient b_1 is given by the least squares method [1]:

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{4}$$

The correction of the readings of an instrument using a calibration curve based on LRA analysis (Figure 2) means that, from each reading y_i* (or from an average value based on *m* readings), the expected reference value, x_i*, can be found, i.e., the expectancy $\hat{E}[x_i * | y_i *]$.

The way to do this is to solve expression (3) for the quantity wanted (x_i*) and, therefore, obtain the functional relationship for the measurand:

$$x_i* = \bar{x} + \frac{(y_i * - \bar{y})}{b_1} \tag{5}$$

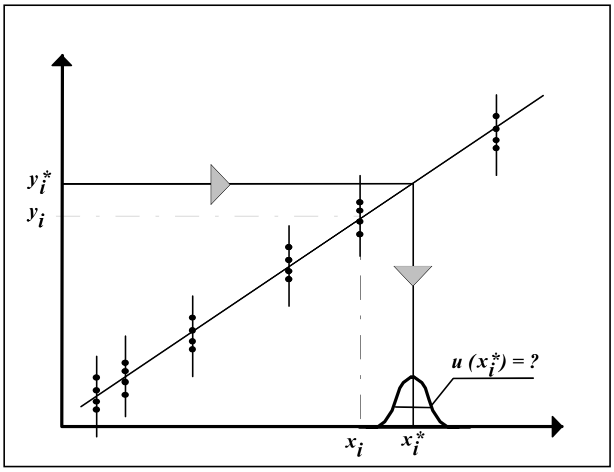

Fig. 2 - Expected reference value, x_i*, obtained from a measurand, y_i*, and LRA uncertainty to be found, $u(x_i*)$.

This functional relationship may be used as basis for the evaluation of the uncertainty component due to the mathematical process, using the Law of Propagation of Uncertainty [2,3] or by applying the variance operator. We have opted for the former.

3. UNCERTAINTY EVALUATION

The model presented in (5) contains four input quantities:

$$x_i* = f(q_i) = f(\bar{x}, y_i*, \bar{y}, b_1) \tag{6}$$

The Law of Propagation of Uncertainty applied to (6) yields the general formulation:

$$u^2(x_i*) = \sum_{i=1}^{4}\left(\frac{\partial f}{\partial q_i}\right)^2 u^2(q_i) + 2\sum_{i=1}^{4}\sum_{j=i+1}^{4}\left(\frac{\partial f}{\partial q_i}\right)\left(\frac{\partial f}{\partial q_j}\right)u(q_i)u(q_j)\ r(q_i, q_j) \tag{7}$$

where *r* is the correlation coefficient.

In order to compute this uncertainty it is necessary to evaluate the derivatives and to estimate the uncertainty of each input quantity and associated correlation coefficient. The derivatives related to the given model (5) are:

$$\left(\frac{\partial x_i*}{\partial \bar{x}}\right) = 1 \qquad \left(\frac{\partial x_i*}{\partial y_i*}\right) = \frac{1}{b_1}$$
$$\left(\frac{\partial x_i*}{\partial \bar{y}}\right) = -\frac{1}{b_1} \qquad \left(\frac{\partial x_i*}{\partial b_1}\right) = -\frac{(y_i * - \bar{y})}{b_1^2} \tag{8}$$

The uncertainty of the input quantities, $u(q_i)$, can be obtained based on the following arguments:

Since LRA is based on a set of *n* calibration pairs (x_i, y_i), each one based on *k* replicates, the uncertainty related to $\bar{y}$ is evaluated by,

$$u^2(\bar{y}) = \frac{s^2}{n} \quad (9)$$

whereas the evaluation of the uncertainty related to the coefficient b_1 [4] yields:

$$u^2(b_1) = \frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (10)$$

In the last two equations, s^2 is the variance of the residuals obtained by LRA, ε_{ij}, estimated using equation (11).

$$s^2 = \frac{1}{nk-2}\left(\sum_{i=1}^{n}\sum_{j=1}^{k}\varepsilon_{ij}\right)^2 \quad (11)$$

The uncertainty of each x_i can be obtained from the calibration certificate either directly or, if the expanded uncertainty is declared (with a coverage factor of k), using the expression: $u(x_i) = i_x / \mathrm{k}$. Applying the Law of Propagation of Uncertainty to the statistical expression of the average, $\bar{x} = \sum_{i=1}^{n} x_i / n$, the uncertainty of this quantity is given by: $u^2(\bar{x}) = \frac{i_x^2}{n}$.

Finally, the uncertainty related to y_i^*, which is taken as an average of *m* readings, can be obtained from:

$$u^2(y_i^*) = \frac{s^2}{m} \quad (12)$$

In the last case, notice the use of the same estimate *s*, resulting from the LRA curve, since one of the conditions required for its application is that the standard deviation of the population distribution must be constant throughout the domain. This assumption combined with the fact that y_i^* belongs to this domain, justifies the above result.

The last group of elements that need to be estimated are the correlation coefficients. The general expression used in statistics to evaluate these coefficients depends of the variances and covariance of the quantities under study, according with (13).

$$r(q_j, q_j) = \frac{\mathrm{cov}[q_i, q_j]}{\sigma(q_i)\,\sigma(q_j)} \quad (13)$$

The mathematical model (5) shows that there are four input parameters and, therefore, six pairs of combinations to be treated. However, two of these parameters can be immediately considered as independent from the others: y_i^*, as it represents the result of a set of independent measurements performed with the calibrated instruments; and $\bar{x}$, which results from the values of the reference standard, x_i. The remaining two input parameters give origin to a single coefficient, $r(\bar{y}, b_1) = \frac{\mathrm{cov}[\bar{y}, b_1]}{\sigma(\bar{y})\,\sigma(b_1)}$, which is null, since it can easily be proved that $\mathrm{cov}[\bar{y}, b_1] = 0$ [5].

Taking into account the results presented, the expression (6) applied to the functional relationship (5) finally gives the square of the LRA uncertainty:

$$u^2(x_i^*) = \frac{i_x^2}{n} + \frac{s^2}{b_1^2}\left(\frac{1}{m} + \frac{1}{n} + \frac{(y_i^* - \bar{y})^2}{b_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}\right) \quad (14)$$

which is in accordance with [6].

4. CALIBRATION EXAMPLE

The case under study is related with the calibration by direct comparison of the measurand, of a 1-dimensional co-ordinate measuring machine in the range of 0 mm to 100 mm, using as reference standards gauge blocks.

In addition, it can be referred that the instrument to be calibrated has a resolution of 0.1μm, that the block gauges were calibrated in an accredited laboratory and that the calibration certificate issued states an uncertainty of:

$$U_{95\%}(x_L) = \pm\left(4.8\ 10^{-2} + 1.9\ 10^{-4} L\right)\mu m \quad (15)$$

where *L* is the length expressed in mm and the uncertainty is given with a coverage factor of 2.00.

The study took as reference a set of eight block gauges (giving origin to *n*=8 calibration pairs), with nominal values between 1 mm to 100 mm, and each calibration pair was achieved from averages obtained from 10 replications

(k=10). The measurements were made in a laboratory under controlled temperature and humidity conditions (20ºC ± 0.3ºC and 50 %rh ± 5 %rh) in order to limit its influence in the length measurement. The results that were obtained are shown in table 1.

TABLE 1 - Calibration pairs and repeatability

Conventional true values x_i (mm)	Measurements average values y_i (mm)	Repeatability (mm)
0.99995	0.9999	$5.8 \cdot 10^{-6}$
2.00002	2.0000	$9.8 \cdot 10^{-6}$
5.00003	4.9999	$5.3 \cdot 10^{-6}$
10.00005	9.9998	$7.1 \cdot 10^{-6}$
25.00005	24.9997	$7.1 \cdot 10^{-6}$
49.99990	49.9991	$7.5 \cdot 10^{-6}$
74.99990	74.9987	$5.8 \cdot 10^{-6}$
99.99989	99.9984	$7.8 \cdot 10^{-6}$

The analysis of the residuals under Gauss-Markov conditions showed that LRA could be applied. With these pairs, the coefficient b_1 and the correction equation can be obtained using (4) and (5), respectively:

$$b_1 = 0.999985127 \qquad (16)$$

$$x_i{*} = 33.49997375 + \frac{(y_i{*} - 33.4994375)}{0.999985127} \qquad (17)$$

The expression (17) allows the correction of a single reading and also of the average of m readings performed with the calibrated instrument. However, this correction has an uncertainty contribution given by (14). To calculate it, two general steps must be followed.

The first step in evaluating the uncertainty is to find the values or the expressions that are taken into account. In this particular case, i_x is the gauge blocks contribution obtained from $\left[U_{95\%}(x_L)/2\right]\mu m$, n is the number of pairs used in the LRA (n = 8), the value of the coefficient b_1 is given by (16) and the s^2 estimate is based on (11), which gives:

$$s^2 = 5.4525 \cdot 10^{-9}\ mm^2 \qquad (18)$$

The second step is to use these results, and obtain the expression for the LRA uncertainty according with (14):

$$u^2(x_i{*}) = \left(7.2 \cdot 10^{-11} + 4{,}5 \cdot 10^{-15} L^2\right) mm^2 + \\ + 5.4527 \cdot 10^{-9}\left(\frac{1}{m} + \frac{1}{8} + \frac{(y_i{*} - 33.4994375)^2}{9.9016763 \cdot 10^3}\right) mm^2 \qquad (19)$$

where L is approx. equal to $x_i{*}$ and represents the length expressed in mm.

This uncertainty depends on two quantities, $y_i{*}$ and m. Therefore, further analysis should review their influence.

The quantity m represents the number of replicates, namely 10, therefore the use of (19) results in uncertainty values which, in this case, varies from 3.508x10^{-5} mm (minimum at 33 mm length) and 6.066x10^{-5} mm (maximum at 100 mm length). Considering 1 mm steps between consecutive computations, Figure 3 shows the results obtained in the range of 1 mm to 100 mm, with the expected minimum value at 33 mm ($\bar{x}$), the global maximum at 100 mm (and a local maximum at 1mm).

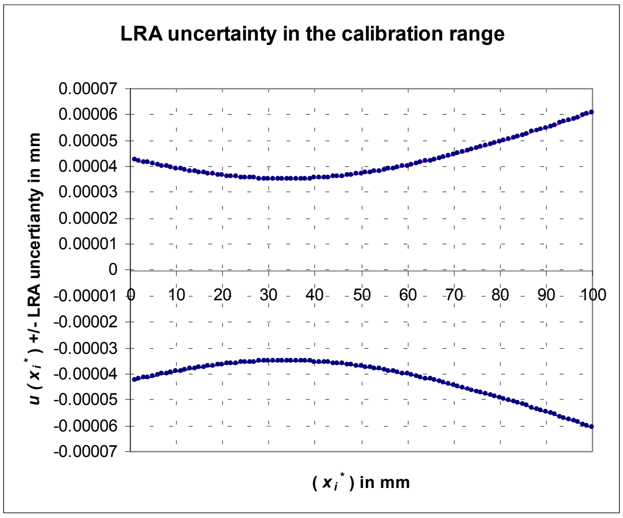

Figure 3 - LRA uncertainty in [1 mm,100 mm]

With the knowledge of these extreme values, a study can be carried out in order to evaluate the influence of the quantity m in the LRA uncertainty (19). Considering these 3 values (1 mm, 33 mm and 100 mm) and a parametric study of m between 1 and 80, Figure 4 shows that for readings between 1 and 10 an important gain can be achieved with a decrease in LRA uncertainty, whereas for m greater than 10 there is no significative gain in uncertainty decrease (for m higher than 10, the gain increase relative to a single measurement is lower than 10 %).

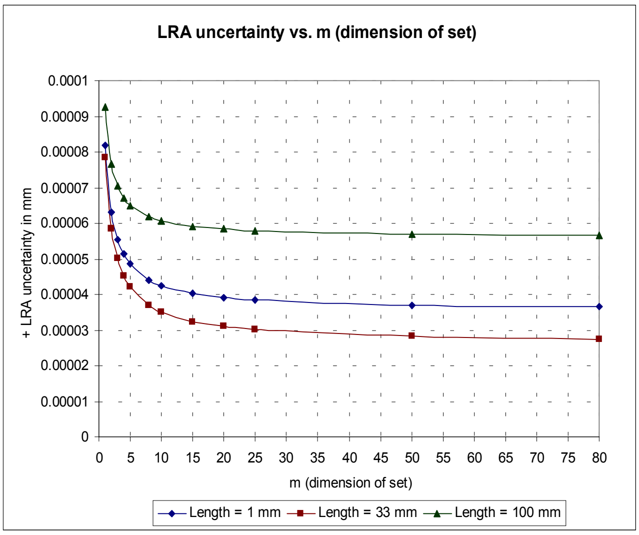

Figure 4 - LRA uncertainty vs. m

Another interesting analysis results from the comparison between the LRA contribution against other major contributions to the uncertainty evaluation, as is the case with the reference block gauge uncertainty, the instrument resolution and the instrument repeatability. Figure 5 shows the absolute values of these components for the case where m=10, expressed for the points taken in the regression analysis, showing that LRA contribution is the highest. The points related with the global uncertainty were estimated neglecting the temperature effect, since this contribution depends on factors that are external to the calibrated instrument (laboratorial room ambient temperature behavior).

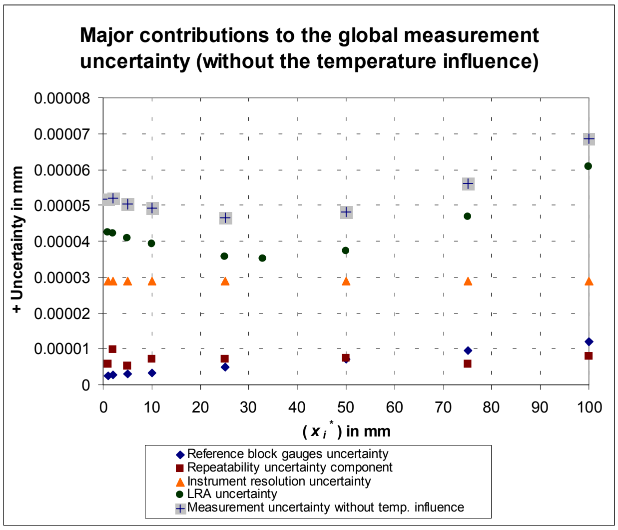

Figure 5 - Intrinsic contributions to measurement uncertainty

This simple example underlines the type of information that can be obtained by the evaluation of the LRA uncertainty contribution, namely, the range and selection of points and number of replicates that should be considered in the calibration process, the selection of the number of replicates to use in order to perform measurements with the instrument under specified accuracy and the ability to compute measurement uncertainties considering the LRA contribution.

5. CONCLUSION

Considering the wide use of LRA as a technique applied in the correction of values from measuring instruments, the implementation of the mathematical approach presented in this study is scientifically relevant and has a broad range of application.

The computation of the uncertainty related with the use of LRA can be nuclear, depending on the relative importance of its value in a measurement overall uncertainty budget, but its use has not been reported often.

The study carried out in this paper aims to discuss and find a practical expression for the LRA uncertainty contribution to the measurement standard uncertainty, and to study the influence of its parameters.

When it is intended to go further and obtain the measurement expanded uncertainty to a certain confidence level, this contribution can acquire a larger importance, namely, due to the role of the degrees of freedom in the determination of the coverage factor (Welch-Satterthwaite formula [2,3]).

Considering the example given in this work where the number of degrees of freedom is low: 6 ($= n - 2$), it was shown that the LRA uncertainty was the major contribution towards to the final uncertainty result. Thus, if this value is taken into account in the evaluation of the number of effective degrees of freedom (Welch-Satterthwaite formula), certainly a lower value would be obtained, leading to a coverage factor considerably higher than the usual k=2.00 (95% of confidence level). An important conclusion to be drawn is, therefore, that the number of pairs considered in the LRA analysis is an important parameter to be considered when specified levels of accuracy need to be achieved.

REFERENCES

[1] Sen, A. and Srivastava, M. "Regression analysis. Theory, methods and applications", *Springer-Verlag*, New York, USA, 1990.

[2] *"Guide for the expression of uncertainty in measurement"*, *ISO, BIPM, CEI, IFCC, IUPAC, OIML*, 1995.

[3] *"Expression of the uncertainty of measurement in calibration"*, EA- 4/02, EA - European Co-operation for Accreditation, 1999.

[4] Mood, A. *et al.*, "Introduction to the theory of statistics", 3rd Ed., *McGraw-Hill International Editions*, Singapore, 1974.

[5] Johnson, N. and Leone, F. "Statistics and experimental design in engineering and the physical sciences. Vol. I", 2nd Ed., *John Wiley & Sons*, USA, 1977.

[6] Mimoso, J. "Certificação de materiais de referência", (Research Study), *Laboratório Nacional de Engenharia Civil*, Lisboa, 1999.

Authors:
National Laboratory of Civil Engineering,
Av. do Brasil, 101, 1700-066 Lisboa, Portugal.
Tel. ++ 351.21 84430 00, Fax ++ 351 21 844 30 41,
E-mail: asribeiro@lnec.pt, jmimoso@lnec.pt, mpcastro@lnec.pt.

Madeira Regional Laboratory of Civil Engineering,
Rua Agostinho Pereira de Oliveira, S.Martinho, 9000-264 Funchal, Portugal. Tel. ++ 351.291 76 52 23 /4/5, Fax ++ 351.291 76 53 85,
E-mail: jasousa@lrec.pt