v. 2, n. 20, 2022

Journal of
Engineering
Research

THE TRACEABILITY CHAIN IN METROLOGY: A STATISTICAL APPLICATION IN METROLOGY

Enrique Villa-Diharce
Researcher at the institution: Centro de Investigación en Matemáticas, A.C., Guanajuato, Guanajuato, Mexico

María Guadalupe Russell-Noriega
Professor-Researcher of the Faculty of Physical-Mathematical Sciences of the institution: Universidad Autónoma de Sinaloa, Culiacán, Sinaloa, México

Abstract: A fundamental element in the international metrological system, which makes it possible to relate any pair of measurements (measurement or metrological statements), is the traceability chain in metrology. A measurement result that cannot be traced to a national standard and therefore to the international metrological system cannot be trusted. Two commonly used procedures to ensure the traceability of measurements to the international metrological system are the calibration of measurement instruments and the inter-comparisons of laboratories, which allow reaching agreements between the measurement standards. In this work we illustrate the use of the statistical methodology in the traceability chain, in order to determine the accuracy and precision of the measurements.
Keywords: Calibration, Monte Carlo, Inter-laboratory studies, Traceability, Reference value.

INTRODUCTION

A basic principle in experimental studies is that measurements are not perfect, but have a certain uncertainty, even despite initially making an experimental design, to optimize the amount of information that the study can provide. Due to the magnitude of the studies and projects where measurements are made, it is very common that we must exchange measurement results, which leads us to consider the reliability of the measurement processes.

Inaccurate measurements, with a significant level of uncertainty, can generate economic losses when the measurements are related to a commercial exchange, or health risks, when they are measurements related to medical treatments. In order for measurement results to be comparable with each other, regardless of where and when the measurements are made, it is necessary that the measurement results be traceable to national or international standards, or to accepted international references.

Metrological traceability consists of an uninterrupted and documented chain of calibrations, up to a reference, which can be a measurement pattern, a measurement procedure, or the practical realization of the definition of a unit of measurement, through documented measurement procedures, that makes it possible to relate the measurement results, to the units of the International System, with a measurement uncertainty, known and documented.

Two procedures that allow traceability of measurements are the calibration of measurement instruments and the key comparison of metrology laboratories. Each stage in the chain of comparisons typically involves the calibration of a measurement device or equipment using a more accurate reference standard. The evaluation of the degrees of equivalence of the results of the participating laboratories, with respect to

Of the reference value that results, is important because it allows comparing the measurement capacity of the different laboratories. Some problems of inter-laboratory evaluations have been discussed in the literature, as can be seen in Kacker et al. (2002), and Kacker et al. (2003), Tomán and Possolo (2009), and Rukhin (2009). In the case of calibration processes, the analysis of measurement data has traditionally been developed using the delta method, which in the metrological community is identified as the procedure recommended by the Guide for the evaluation and expression of measurement uncertainty (2008), better known as "the GUM". A more general methodology to deal with these problems is distribution propagation, which can be implemented using Monte Carlo simulation techniques (Robert and Casella, 1999). The use of this methodology

has been growing, since the publication of the Supplement (2008) of the GUM. In Section 2, we describe a statistical technique for the analysis of data from inter-laboratory studies, based on weighted averages. A gauge block calibration problem is also presented, which is solved using a Monte Carlo procedure. In Section 3, some comments are made on the application of statistics in metrology.

COMPONENTS OF THE TRACEABILITY CHAIN

In the following subsections, we present a couple of examples that we use to illustrate the application of statistical procedures, in the analysis of data arising from inter-laboratory studies that are carried out to determine reference values, of three volume transfer standards. In the second example, we develop a longitudinal gauge block calibration process.

In the first example, we used a data set from an inter-laboratory comparison, known as a Key Comparison, or KC. Here the data are measurements (and their uncertainties) of transfer patterns, reported by the national laboratories participating in the study. In this example, three volume standards (20L) were used, as in the original data the observed scatter was very small, the values of the first artifact (TS 710-04) were replaced by a set of values with a larger scatter, with the objective of being able to show results, where there is a significant inter-laboratory variation.

REFERENCE VALUES OF A KEY COMPARISON

Key Comparisons (KCs) between laboratories are the technical basis for mutual recognition agreements between national metrology centres. The purpose of the inter-comparisons between the national metrology institutes is to test whether the measurements made by the participating countries are consistent, taking into account the uncertainties assigned to the measurements. If an inconsistency is detected, the participating countries must take the necessary corrective actions to have consistency. The purpose of a KC is to establish the key comparison reference value (KCRV), the degrees of equivalence and their associated uncertainties, based on the data provided by the participants.

The results of the laboratories are considered realizations of the random variables: x_1, x_2,..., x_n, where,

$$x_i = Y + \epsilon_i, i = 1,2, \dots, n, \quad (1)$$

and, Y is the measurand and ϵ_1, ϵ_2,..., ϵ_n, mutually independent random variables with zero mean and variances σ_1^2, σ_2^2,..., σ_n^2. We consider the measurand to be a physical quantity of stable value during the comparison. We further assume that the random variables have a normal distribution, that is, $x_i \sim N(Y, \sigma_i^2)$.

The least squares estimate of the parameter: Y is $x_R = \sum_{i=1}^{n} \frac{w_i x_i}{\sum w_j}$, where $w_i = 1/\sigma_i^2$. Besides, $E(x_R) = Y$ and $V(x_R) = \frac{1}{\sum w_i}$.

In practice the variances: σ_i^2 are unknown, so metrologists substitute these variances for their estimates: s_i^2. As a result of the statistical analysis of the data we have:

x_R = reference value,

$d_i = x_i - x_R$ degree of equivalence of the result: x_i,

$d_{ij} = d_i - d_j = x_i - x_j$ =degree of equivalence of the results x_i and x_j.

$u(x_R)$, $u(d_i)$=, $u(d_{ij})$ and are the standard uncertainties of x_R, d_i and d_{ij}, respectively.

It can be seen that considering the distribution of the x_i's, degrees of equivalence: d_i and d_{ij}, satisfy, $E(d_i)=E(d_{ij})=0$, $V(d_i)=u^2(x_i)-u^2(x_R)$ y $V(d_{ij})=u^2(x_i)-u^2(x_j)$.

In the considerations we have made about the relationship between the results: x_i of the laboratories and of the value: Y of the measurand, we have assumed that the variation of the results is a consequence only of the inter-

Journal of Engineering Research ISSN 2764-1317 DOI 10.22533/at.ed.3172202213092

laboratory variation, expressed in the value of $u(x_i)$. Sometimes this does not happen and we find that the variation between the results is greater than the dispersion explained by the inter-laboratory uncertainties. In this case we must assume that there is an external variation component, inter-laboratory, that can explain the excess in the variation of the results. To model this variation of the results and its relationship with the measurand, we consider a random effects model, where we include a variation component due to the laboratories,

$$x_i = Y + b_i + \epsilon_i, \qquad (2)$$

where $b_i = Y + X_i$ is the laboratory effect in: x_i y $\epsilon_1 = (x_i - X_i)$ is the intra-laboratory error, and X_i is the expected result of the i-th laboratory, that is, $E(x_i) = X_i$.

The biases: b_1, b_2,...,b_n due to laboratories, identically distributed random variables (vaiid), with normal distribution, with zero mean and variance are considered: σ_b^2, this is, $b_1 \sim N(0, \sigma_b^2)$. An assumption that metrologists usually make here is that the estimated variances: $s_1^2 + u^2(x_i), \ldots, s_n^2 + u^2(x_n)$, are taken as the true variances of the sampling distributions of the results: $x_1, x_2, \ldots, x_n$. Under this assumption, the best estimate of the value of the measurand Y is the weighted mean: $x_P = \sum_{i=1}^{n} \frac{w_i x_i}{\sum w_j}$, with a variance given by $V(x_P) = \frac{1}{\sum w_i}$, being in this case the weights given by: $w_i = 1/[s_b^2 + u^2(x_i)]$. The expressions for the estimates in the case of the simple model given by equation (1), and the random effects model (2), are similar, only that the variance of the results of the laboratories, in the second model, is greater, since an inflation term is added, given by the inter-laboratory variance: s_b^2.

To illustrate the procedure for estimating the reference value and the degrees of equivalence of the different participating laboratories in a key comparison that is carried out to establish the reference value of a measurement pattern, we use information from a report of a inter-comparison of national laboratories from 8 countries: CENAM (Mexico), NIST (United States), MC (Canada), SP (Sweden), PTB (Germany), IMGC (Italy), NMIA (Australia) and finally INMETRO (Brazil). Table 1 shows the values reported by the different laboratories. These values resulted from measurement processes carried out by each laboratory following their own measurement protocols. Participating laboratories determined the volume of water that each of three Transfer Standards (labeled TS 710-04, TS 710-05, and TS 710-06) of 20 l can deliver after a 60-second drip period, at a reference temperature of 20°C. The transfer patterns were three reservoirs with a nominal volume of 20 l.

There are different methods of estimating s_b^2, inter-laboratory variance, which have been proposed by Cochran (1954), Paule and Mandel(1982) and DerSimonian and Laird(1986). Here we use the iterative method of Paule and Mandel, which is included in the "metRology" computing package, developed within the R platform, Team, R. C. (2015). According to the data that we observe in Table 1, while the values reported for artifacts TS 710-05 and TS 710-06 show an extremely small variation, for artifact TS710-04, the values reported by the laboratories show a greater variation. This difference in dispersion is reflected in the estimation of the inter-laboratory variance, since the iterative method of Paule and Mandel(1982) gives a value $s_b^2 = 4.94$ for the first artifact (TS710-04), a value: $s_b^2 = 0.083$ for the second and a null value for the third.

In Table 1, at the bottom, the reference values (and their uncertainties) that result when taking the weighted average for the three transfer patterns are shown. In the last artifact, the uncertainty associated with the

		TS 710-04		TS 710-05		TS 710-06	
Metrological Center		$x_i[ml]$	$u(x_i)[ml]$	$x_i[ml]$	$u(x_i)[ml]$	$x_i[ml]$	$u(x_i)[ml]$
CENAM	(1)	20000.03	0.17	19997.31	0.17	20005.60	0.17
NIST	(2)	19999.86	0.38	19996.83	0.25	20005.04	0.37
MC	(3)	19998.64	0.31	19997.75	0.31	20005.98	0.31
SP	(4)	20001.93	0.36	19997.40	0.36	20005.63	0.36
PTB	(5)	19999.34	0.20	19997.44	0.20	20005.54	0.20
IMGC	(6)	20004.17	0.13	19998.00	0.15	20005.96	0.14
NMIA	(7)	19999.98	0.23	19997.16	0.22	20005.59	0.22
INMETRO	(8)	20001.42	0.15	19997.33	0.14	20005.54	0.15
KCRV	U(KCRV)	20000.68	0.62	19997.77	0.13	20005.67	0.07

Table 1. Results reported for 20 l. transfer standards. (artifacts 710-04, 710-05 and 710-06). In addition, the estimates of the reference values are presented at the end.

		TS 710-04		TS 710-05		TS 710-06	
Metrological Center		D_i	$U(D_i)$	D_i	$U(D_i)$	D_i	$U(D_i)$
CENAM	(1)	-0.65	2.09	-0.10	0.31	-0.03	0.15
NIST	(2)	-0.82	2.11	-0.58	0.36	-0.26	0.36
MC	(3)	-2.04	2.10	0.34	0.40	0.13	0.30
SP	(4)	1.25	2.11	-0.01	0.44	0.00	0.35
PTB	(5)	-1.34	2.09	0.03	0.33	-0.01	0.19
IMGC	(6)	3.49	2.08	0.59	0.30	0.24	0.12
NMIA	(7)	-0.70	2.09	-0.25	0.34	-0.05	0.21
INMETRO	(8)	0.74	2.08	-0.08	0.29	-0.03	0.13
KCRV	U(KCRV)	20000.68	0.62	19997.77	0.13	20005.67	0.07

Table 2. Degrees of equivalence for artifacts (710-04, 710-05 y 710-06).

Variable	Distribution	Parameters
l_s	$t(\mu,\sigma,\nu)$	$\mu = 50{,}000623nm, \sigma = 25nm, \nu = 18$
D	$t(\mu,\sigma,\nu)$	$\mu = 215nm, \sigma = 6nm, \nu = 24$
d_1	$t(\mu,\sigma,\nu)$	$\mu = 0nm, \sigma = 6nm, \nu = 4$
d_2	$t(\mu,\sigma,\nu)$	$\mu = 0nm, \sigma = 7nm, \nu = 8$
α_s	$Uniforme(a,b)$	$a = 9.5 \times 10^{-6\circ}C^{-1}, b = 13.5 \times 10^{-6\circ}C^{-1}$
θ_s	$N(\mu,\sigma)$	$\mu = -0.1°C, \sigma = 0.27°C$
Δ	$ArcoSeno(a,b)$	$a = -0.5°C, b = 0.5°C$
$\delta\alpha$	$TraCur(a,b,d)$	$a = -1.0 \times 10^{-6\circ}C^{-1}, b = 1.0 \times 10^{-6\circ}C^{-1}, d = 0.1 \times 10^{-6\circ}C^{-1}$
$\delta\theta$	$TraCur(a,b,d)$	$a = -0.05°C, b = 0.05°C, d = 0.025°C$

Table 3. Proposed distributions for the input variables of the proposed measurement model in the longitudinal pattern block calibration problem.

Method	δl	$u(\delta l)$	$IC(95\%)$	$IC(99\%)$
GUM	838	32	(775,901)	(756,920)
MMC-A	838	36	(767,908)	(745,932)

Table 4. Results of the estimation of the output variable, for the measurement model considered in the calibration example, taking into account the uncertainty propagation and distribution approaches.

measurand (0.07) is the result only of the intra-laboratory uncertainties, for the second artifact the standard uncertainty includes a small inter-laboratory variance (0.083) and instead, for the first artifact, it is has a greater uncertainty (0.63), thanks to the inter-laboratory variance, which has a considerable value, s_b^2 =4.94.

Table 2 presents the estimates of the degrees of equivalence (and their uncertainties), here we also find that the degrees of equivalence for the values of the first artifact have a greater dispersion (than the other two artifacts) with respect to the value of reference (the estimated measurand), as a product of having a higher inter-laboratory variance.

CALIBRATION PROCESSES

A calibration process establishes the relationship between the values of the quantities indicated by an instrument or measurement system and the values given by the measurement standards. Typically a calibration model establishes a relationship between an artifact or artifacts being calibrated, with measurement standards from a higher level in the traceability chain.

In the measurement model: $Y=f(X_1,X_2,...,X_n)$, we have in the vector of input quantities $(X_1,X_2,...,X_n)$, both reference standards used and environmental variables. To illustrate a measurement process, we present here the longitudinal gauge block calibration process. We consider the determination of the length of a calibrated block, of a nominal length of 50 mm, by comparison with a known standard of the same nominal length. This is an example from GUM (2002) Supplement 1 and is also included in GUM (2002).

The difference in their lengths is, $d=l(1+\alpha\theta)-l_s(1+\alpha_s\theta_s)$,

where,

l= length at 20°C, of the standard block to be calibrated (measurand),

l_s= gauge block length at 20°C, given in its calibration certificate,

α= coefficient of thermal expansion of the block to be calibrated,

α_s= coefficient of thermal expansion of the gauge block,

θ= temperature deviation of the block to be calibrated, with respect to 20°C,

θ_s = temperature deviation of the standard block, with respect to 20°C.

A more suitable expression of the model is,

$$l = l_s[1 + \alpha_s(\theta + \delta\theta)]/[1 + (\alpha_s + \delta\alpha)\theta].$$

A suitable expression for l is, $l=l_s+d-l_s(\alpha_s\theta_s-\alpha\theta)$.

Considering,

$\delta\theta=\theta-\theta_s$= temperature difference between blocks,

$\delta\alpha=\alpha-\alpha_s$= difference of the coefficients of thermal expansion.

Considering also,

$d=D=d_1-d_2$, where,

D= average of five measurements,

d_1= random effect associated with the comparator,

d_2= systematic effect associated with the comparator,

$\theta=\theta_0+\Delta$, where,

θ_0 = average deviation of 20°C, of the block to be calibrated,

Δ= Cyclic temperature variation.

Considering the previous expressions of d and θ, we have the following expression for l,

$$l = l_s + D + d_1 + d_2 - l_s[\delta\alpha(\theta_0 + \Delta) + \alpha_s\delta\theta]$$

Finally we take as measurand the deviation of l from its nominal length (l_{nom}=50mm),

$$\delta l = l - l_{nom} = l_s + D + d_1 + d_2 - l_s[\delta\alpha(\theta_0 + \Delta) + \alpha_s, \delta\theta] - l_{nom}$$

So, finally, we have the measurement model,

$$\delta l = f(l_s, D, d_1, d_2, \alpha_s, \theta_0, \Delta, \delta\alpha, \delta\theta).$$

Table 3 shows the distributions that

are assumed for the input variables of the measurement model presented here. Both for the lengths: l_s y D that result from a previous calibration, as from a measurement, respectively; As for the random and systematic effects of the comparator used for the measurements, Student's t-distributions are associated. The coefficient of thermal expansion of the gauge block (α_s), it is assumed that it follows a Uniform distribution, since it is only known that the value of the coefficient is in a given interval (a,b). In the case of temperature variation, around 20°C, (Δ) it is assumed to be cyclical (sinusoidal), thanks to the temperature conditioning system. For the differences between the temperature of the blocks ($\delta\theta$), and between the coefficients of thermal expansion ($\delta\alpha$), since there is inexact information about the limits of the intervals that contain them, then, according to the principle of maximum entropy, it is appropriate to assume a Curvilinear Trapezoidal distribution.

Traditionally, to obtain the distribution of the output variable (δl) of the measurement model, the GUM approach is used, based on the method of moments, and which also assumes a normal distribution for this variable. This approach works well when the measurement model is approximately linear, which is sometimes not the case. Thanks to the information that is available, or assumed, (the distributions), the Monte Carlo approach can be used to obtain the distribution of the output variable by simulation. Figure 1 shows the (normal) distribution that results according to the GUM, and a histogram that shows the distribution of the simulated values of the output variable, according to the Monte Carlo method.

The solution reported by the GUM, considering the propagation of uncertainty, and the solution based on the adaptive Monte Carlo method, have a very small difference, as can be seen in Table 4, and in Figure 1, which shows the density function reported by the GUM and the histogram of the values of the output variable (δl), corresponding to the final measurement model, taking the distributions of the input variables, from Table 3. The number of trials required to have a tolerance: $\delta=(1/2)10^{o}$ is M=130000.

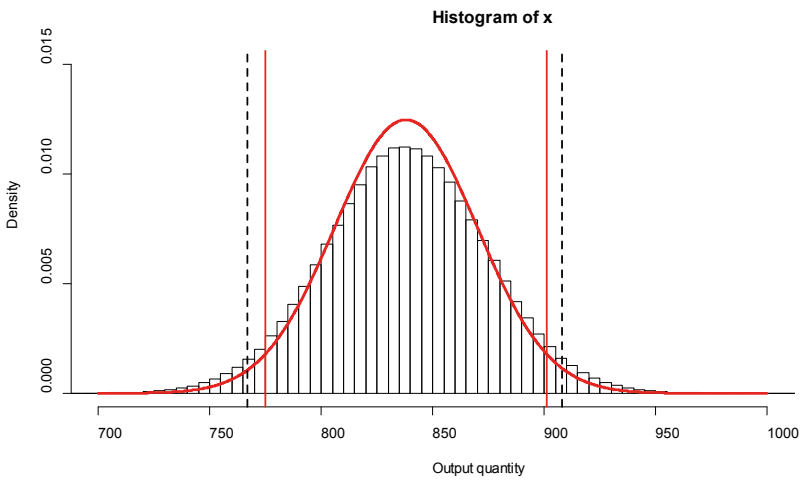

Figure 1. Distribution of the output variable, of the measurement model of the calibration example, according to the uncertainty propagation and distribution approaches.

CONCLUSIONS

In this work we have commented on two commonly used techniques in metrology, inter-laboratory studies, which are developed to establish reference values, which are basic in order to establish the equivalence relationship between different measurements and calibration processes, which are essential in any measurement system. measurement assurance. In apparently simple inter-laboratory comparison studies, there are a great variety of statistical problems because the characteristics of the different participating laboratories are very diverse. We have also presented a calibration data analysis process, following a Monte Carlo procedure, which has a greater range of application than the usual procedure, based on the method of moments. Metrology is an interesting area of opportunity, which requires the development of statistical models and the application of efficient solution techniques.

REFERENCES

Cochran, W.G. "The Combination of Estimates from different Experiments," *Biometrics*, 10: 101-129, 1954.

DerSimonian, R. y N. Laird. "Meta-analysis in clinical trails," *Controled Clinical Trials*, 7(3):177-188. 1986.

Joint Committee for Guides in Metrology. "*Evaluation of measurement data- Guide to the expression of uncertainty in measurement*," International Bureau of Weights and Measures (BIPM), Sevres, France 2008.

Joint Committee for Guides in Metrology. *Evaluation of measurement data-Supplement 1 to the "Guide to the expression of uncertainty in measurement"- Propagation of distributions using a Monte Carlo method*. International Bureau of Weights and Measures (BIPM), Sevres, France 2008.

Kacker, R.N., R.U. Datla. y A.C. Parr. "Combined result and associated uncertainty from inter-laboratory evaluation based on the ISO Guide," *Metrologia*, 39:279-293. 2003.

Kacker, R.N., R.U. Datla. y A.C. Parr. "Statistical interpretation of key comparison reference value and degrees of equivalence," *J. Res. Natl. Inst. Stand. Technol*., 39:279-293. 2003.

Paule, R.C. y J. Mandel. "Consensus values and weight factors," *J. Res. Natl. Bur. Stand*. 7: 177-87. 1982.

Robert, C.P. y G. Casella. "Monte Carlo Statistical Methods," Springer Texts in Statistics, ISBN 0-387-98707-X. 1999.

Rukhin,A.L. "Weighted means statistics in interlaboratory studies," *Metrología*, 46: 323-331. 2009.

Team, R.C. "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria:. 2015.