Proceedings of the 13th International Symposium on Measurements for Research and Industry Applications - IMEKO TC-4. Athens: National Technical University of Athens, 2004. v. I. p. 386-391.

A SOFTWARE FOR THE EVALUATION OF THE STABILITY OF MEASURING STANDARDS USING BAYESIAN STATISTICS

G. M. Rocha[1], G. A. Kyriazis[2].

Instituto Nacional de Metrologia, Normalização e Qualidade Industrial – Inmetro.
Av. Nossa Senhora das Graças, 50 - Rio de Janeiro, Brazil – CEP: 25250-020
e-mail: [1]gmrocha@inmetro.gov.br, [2]gakyriazis@inmetro.gov.br

Abstract - The manufacturers specify the long term stability of their measuring standards. This parameter is sometimes the largest uncertainty contribution to the overall uncertainty of the standard. An automated system for the evaluation of the stability of measuring standards using Bayesian statistics was developed. For several standards it was verified that the stability estimate is much smaller than the value specified by the manufacturer. The software for data analysis was implemented in language C in the environment LabWindows/CVI.

I. Introduction

The classical approach considers the probability of an event as an objective property of that event, always subject, in principle, to empirical measurement by means of frequencies in a random experiment. On the other hand, the Bayesian approach considers probabilities as an expression of the human ignorance. The probability of an event is merely a formal representation of our belief that the event occurred or will occur, based on any available information. Bayesian statistics has been increasingly applied in measurement science and technology [1]-[4].

In this paper, the stability, drift and other parameters related to the measuring standards were estimated based on a database containing the historic of periodic calibrations. The main justification for using Bayesian techniques is to estimate both the stability parameters and their associated uncertainties based on small data samples. The specific statistical techniques used here allow one to know the actual behavior of the standard in the long term and even to make predictions of its future behavior. An automated system was developed that uses Bayesian statistics to estimate the standard drift (or long term stability), and to predict future measurement results, based on a data set obtained from previous periodic calibrations of the standard. This tool has been very useful in the control and maintenance of our measuring standards.

II. Bayesian Inference

In classical statistical theory, probabilities are assigned to observable events that occur at random as a result of some well-defined experiment. The probability of such an event is identified with the relative frequency of the event being observed after unlimited repetitions of the experiment performed under the same nominal conditions. However, there are events that cannot be observed, and yet a measure of degree of belief about their occurrence is needed. This measure is offered by Bayesian interpretation of probability as a number that satisfies certain axioms and that is used to describe the state of incomplete knowledge – derived from the available information – about the occurrence of an observable or no observable event.

Bayes' Theorem, expounded in many textbooks on probability and statistics [5][6], can be understood as a mathematical description of the learning process. Indeed, let θ represent the values of a random variable Θ. First, an initial assessment is made of the *probability density function* (pdf) that reflects the knowledge about Θ existing before measurement data are gathered. This is called the prior (to the data) pdf, denoted here as $p(\theta)$. Next, the prior pdf is modified according to the actual data x that are obtained during the course of the experiment. The result is the posterior pdf, $p(\theta|x)$, from which the best estimate and the standard or expanded uncertainty are calculated.

To construct $p(\theta)$, one should make use of all available information, such as calibration data, data from other similar experiments, personal experience, etc. However, since this information is usually personal and vague, the most conservative and general approach is to suppose that we start from "complete ignorance" about the quantity.

Proceedings of the 13th International Symposium on Measurements for Research and Industry Applications - IMEKO TC-4. Athens: National Technical University of Athens, 2004. v. I. p. 386-391.

To incorporate the measurement data, one needs to express the probability density of observing the data x given that the value of Θ is θ. Of course, once the data are at hand, they are fixed and can no longer be considered as variables. Therefore, it is more natural to think of $p(x|\theta)$ as a function of the unknown value of Θ given the data. Regarded in that sense, $p(x|\theta)$ is not necessarily a pdf. For this reason, this density (in x) is called the *likelihood* function (in θ), and is denoted $l(\theta|x)$ [5].

Assuming the vector of n available observations $\boldsymbol{x} = x_1, x_2, \ldots, x_n$, if the vector of m parameters $\theta = \theta_1, \theta_2, \ldots, \theta_m$, denotes the values that can be reasonably attributed to the vector of m measurands $\Theta = \Theta_1, \Theta_2, \ldots, \Theta_m$, then the Bayes' Theorem may be written in terms of probability densities as

$$p(\theta \mid \boldsymbol{x}) \propto p(\theta) \,.\, l(\theta \mid \boldsymbol{x}) \tag{1}$$

Here $p(\theta \mid \boldsymbol{x})$ is the posterior (to data $\boldsymbol{x}$) probability density of θ, $p(\theta)$ is the prior (to data $\boldsymbol{x}$) probability density of θ and $l(\theta \mid \boldsymbol{x})$ is the likelihood of θ (to given $\boldsymbol{x}$). The likelihood is numerically equal to the probability density of data $\boldsymbol{x}$ (given that θ is known). It represents the additional information provided by the data.

Bayes' theorem may be easily remembered as: *Posterior* $\propto$ *Prior x Likelihood.* This relation summarizes the way one should modify the degree of belief in order to consider the available data. In the Bayesian point of view, the prior and posterior densities are descriptions of the state of knowledge of Θ and, thus, there is no restriction as to Θ generate random data or not. The result of Bayesian inference is uniquely determined when the prior density is chosen.

III. Linear Regression Model

In the control and maintenance of measuring standards, each laboratory realizes periodic calibrations and from these measurements we may assume that we have a set of n ordered pairs of observations, the pair being independent of one another but members of the same pair being, in general, not independent. We denote these observations (x_i, y_i) and, as usual, we assume that these pairs have a bivariate normal distribution. The notation below is adopted.

$$\bar{x} = \Sigma x_i/n, \quad \bar{y} = \Sigma y_i/n, \quad S_{xx} = \Sigma\,(x_i - \bar{x})^2, \quad S_{yy} = \Sigma\,(y_i - \bar{y})^2, \quad S_{xy} = \Sigma\,(x_i - \bar{x})(y_i - \bar{y}) \tag{2}$$

The problem here is to use the values of one variable to explain or to predict the values of another. In Bayesian statistics this is commonly referred to as an explanatory and a dependent variable, although it is conventional to refer to an independent and a dependent variable.

A. Linear regression

The model can be written as $y_i \sim \mathrm{N}(\alpha + \beta\,(x_i - \bar{x}), \varphi)$. Because a key feature of the model is the regression line $y = \alpha + \beta(x - \bar{x})$ on which the expected values lie, the parameter β is usually referred to as the slope and α is sometimes called the intercept.

By using Bayes theorem and an uninformative prior, the posterior pdf is

$$p(\alpha, \beta, \varphi \mid \boldsymbol{x}, \boldsymbol{y}) \propto p(\alpha, \beta, \varphi)\,p(\boldsymbol{y} \mid \boldsymbol{x}, \alpha, \beta, \varphi) \propto \varphi^{-1}(2\pi\varphi)^{-n/2} \exp\left[-\sum\{y_i - \alpha - \beta(x_i - \bar{x})\}^2/2\varphi\right]$$

or

$$p(\alpha, \beta, \varphi \mid \boldsymbol{x}, \boldsymbol{y}) \propto \varphi^{-(n+2)/2} \exp\left[-\left\{S_{ee} + n(\alpha - a)^2 + S_{xx}(\beta - b)^2\right\}/2\varphi\right] \tag{3}$$

where S_{ee}, a and b are given by

$$S_{ee} = S_{yy} - S_{xy}^{\ 2}/S_{xx} = S_{yy}\,(1-r^2), \quad a = \bar{y} \text{ and } b = S_{xy}/S_{xx}, \tag{4}$$

Integrating out φ, we obtain

$$\frac{\alpha - a}{s/\sqrt{n}} \sim t_{n-2} \quad \text{and} \quad \frac{\beta - b}{s/\sqrt{S_{xx}}} \sim t_{n-2} \tag{5}$$

where t_{n-2} is a Student distribution with n-2 degrees of freedom, a and b are referred to as least squares estimates of α and β ,and $s^2 = S_{ee}/(n-2)$. The regression line

$$y = a + b(x - \bar{x}) \tag{6}$$

which can be plotted for all x as opposed to just those x_i observed, is called the line of best fit for y on x. The standard uncertainty associated with the best estimates of the intercept and slope are, respectively:

Proceedings of the 13th International Symposium on Measurements for Research and Industry Applications - IMEKO TC-4. Athens: National Technical University of Athens, 2004. v. I. p. 386-391.

$$u(a) = \sqrt{(n-2)/(n-4)}\,.\; s\big/\sqrt{n} \;>\; s\big/\sqrt{n}$$

and (7)

$$u(b) = \sqrt{(n-2)/(n-4)}.\; s\big/\sqrt{S_{xx}} > s\big/\sqrt{S_{xx}}$$

When n is large,

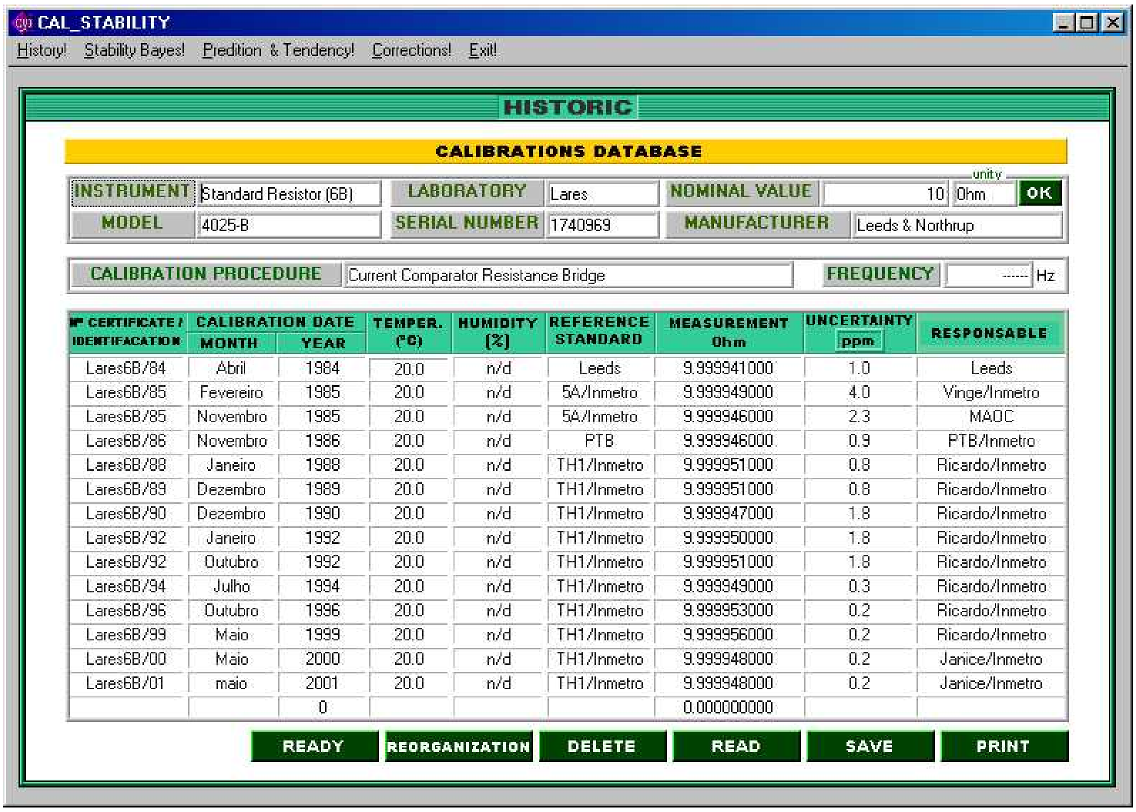
$$u(a) = s\big/\sqrt{n} \qquad \text{and} \qquad u(b) = s\big/\sqrt{S_{xx}} \tag{8}$$

The interested reader may find a detailed derivation of the above expressions in [5].

IV. The software for curve adjustment, calculation of stability and other parameters

A. The software

The software comprises four panels with several fields for identification, data storage and statistical calculations. The first panel is named "Historic" (Fig. 1) and contains the calibration database, where the calibration date, temperature, measurement results, associated uncertainties and serial numbers (or identifications) are stored among other data. The second panel is named "Bayesian Statistics Graph" (Fig. 2) and displays the stability and a graph containing the measurement results obtained from periodic calibrations as function of the calibration dates. The third panel is named "Prediction Graph" (Fig. 3) and displays the estimated values and the graph with the behavior of the standard as function of the time. The fourth panel (not shown) is named "Temperature Correction" and is used to correct the values obtained in different temperatures. The figures shown here were obtained from a 10-ohm resistor.



Nº CERTIFICATE / IDENTIFACATION	CALIBRATION DATE MONTH	YEAR	TEMPER. (°C)	HUMIDITY (%)	REFERENCE STANDARD	MEASUREMENT Ohm	UNCERTAINTY ppm	RESPONSABLE
Lares6B/84	Abril	1984	20.0	n/d	Leeds	9.999941000	1.0	Leeds
Lares6B/85	Fevereiro	1985	20.0	n/d	5A/Inmetro	9.999949000	4.0	Vinge/Inmetro
Lares6B/85	Novembro	1985	20.0	n/d	5A/Inmetro	9.999946000	2.3	MAOC
Lares6B/86	Novembro	1986	20.0	n/d	PTB	9.999946000	0.9	PTB/Inmetro
Lares6B/88	Janeiro	1988	20.0	n/d	TH1/Inmetro	9.999951000	0.8	Ricardo/Inmetro
Lares6B/89	Dezembro	1989	20.0	n/d	TH1/Inmetro	9.999951000	0.8	Ricardo/Inmetro
Lares6B/90	Dezembro	1990	20.0	n/d	TH1/Inmetro	9.999947000	1.8	Ricardo/Inmetro
Lares6B/92	Janeiro	1992	20.0	n/d	TH1/Inmetro	9.999950000	1.8	Ricardo/Inmetro
Lares6B/92	Outubro	1992	20.0	n/d	TH1/Inmetro	9.999951000	1.8	Ricardo/Inmetro
Lares6B/94	Julho	1994	20.0	n/d	TH1/Inmetro	9.999949000	0.3	Ricardo/Inmetro
Lares6B/96	Outubro	1996	20.0	n/d	TH1/Inmetro	9.999953000	0.2	Ricardo/Inmetro
Lares6B/99	Maio	1999	20.0	n/d	TH1/Inmetro	9.999956000	0.2	Ricardo/Inmetro
Lares6B/00	Maio	2000	20.0	n/d	TH1/Inmetro	9.999948000	0.2	Janice/Inmetro
Lares6B/01	maio	2001	20.0	n/d	TH1/Inmetro	9.999948000	0.2	Janice/Inmetro
		0				0.000000000		

Fig. 1 – Panel displaying the standard calibration historic.

B. Curve fitting

The problem considered in this section is a special case of least-squares fitting. It is concerned with the fitting of a curve to a two-dimensional set of points on a coordinate plane. The form of the curve is fixed and its parameters are to be determined. In our case, the points are distributed along a straight line and we use Bayesian statistics to estimate the slope, intercept and other parameters.

Proceedings of the 13th International Symposium on Measurements for Research and Industry Applications - IMEKO TC-4. Athens: National Technical University of Athens, 2004. v. I. p. 386-391.

In formal terms, we consider two quantities $X(\phi)$ and $Y(\phi)$ where ϕ represents the measurement conditions. Let these conditions be discretely variable and we define $x_i = X(\phi)$ and $y_i = Y(\phi)$, with $i = 1, 2, ..., n$. Then, considering the data shown in Fig. 1 our task is to estimate the slope, intercept and other parameters associated with the straight line that best accounts for the data. More specifically, given a set of measurements represented by x_i and y_i, our objective is to fit a straight line that pass near of the points (x_i, y_i) where x_i are the values that correspond to the month and year of the experiment and y_i are the calibration results along the time [8].

C. Stability

The stability of a measurement standard may be specified as short-term stability and long-term stability. The ***short-term stability*** is defined as **repeatability** (of results of measurements) - closeness of the agreement between the results of successive measurements of the same measurand carried out under the same measurement conditions . The ***long-term stability*** refers to the ability of a measuring instrument to maintain constant its metrological characteristics along the time. **Drift** is a slow change of a metrological characteristic of a measuring instrument or standard [7].

The stability and drift are estimated by applying Bayesian statistics to sample data obtained from a historic database obtained from successive calibrations. The data for the calculation of the stability are: the date and the measurement result in each calibration. As mentioned before, these data correspond in the graph (x_i, y_i) where x_i are the values that correspond to the month and year of the experiment and y_i are the calibration results along the time.

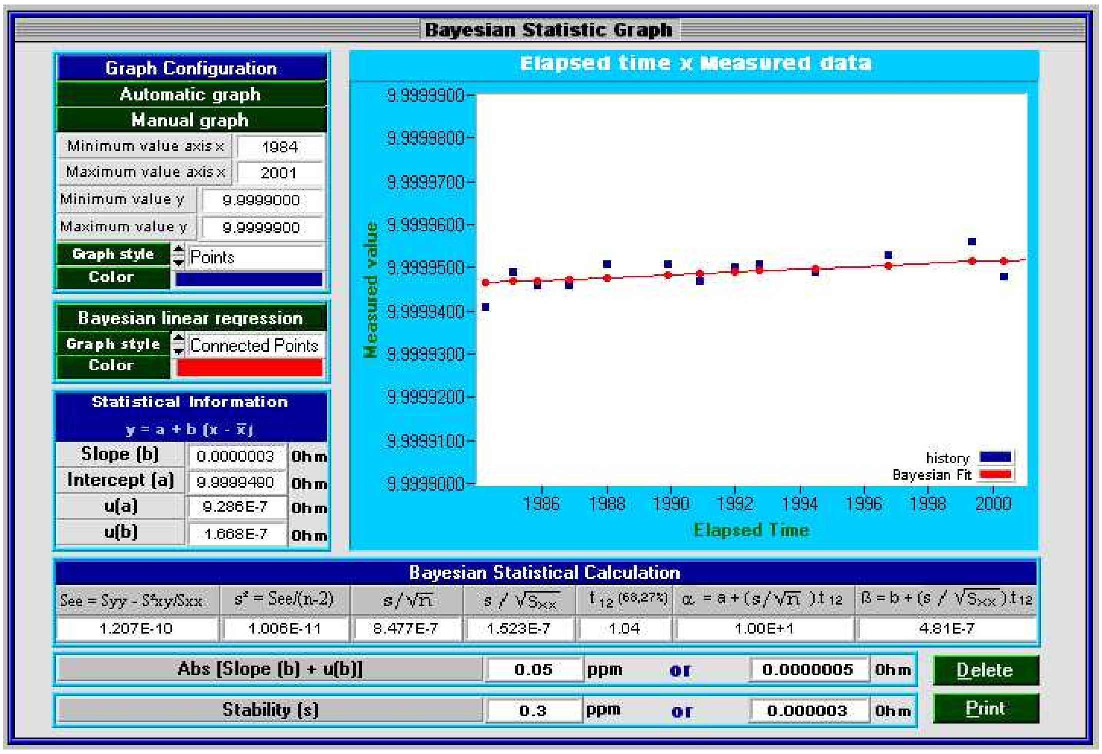


Fig. 2 – Display panel with the stability graph.

The drift was taken as the slope of the regression line. When the graph with the results of calibrations along the time presents a linear behavior the values obtained for the stability and drift are practically same.

The calculations of the stability, drift and the other parameters were totally automated. Then, using the rationale shown in section III, we obtained from (4) the stability (s), the slope (b) and the intercept (a). From (7) and (8) we obtained the uncertainties associated with the last two quantities. Equation (6) is used in the graph construction. These calculations are shown in Fig. 2, where in the upper left side there are some fields for graph plotting, while in the lower side some results obtained from the statistical calculation are shown. The graph also displays the regression line. These data are used to quantify the stability of the standards, as well as to calculate other important parameters.

Proceedings of the 13th International Symposium on Measurements for Research and Industry Applications - IMEKO TC-4. Athens: National Technical University of Athens, 2004. v. I. p. 386-391.

D. Prediction

Now we are interested in the distribution of a potential observation of a value $x = x_o$, that is, the predictive distribution, and then, the result is slightly different. The mean of such observations conditional on x, y and x_o is still $a + b(x_o - \bar{x})$, but since we have a new distribution, in addition to this new distribution for $\gamma = y_o - \alpha - \beta(x_o - \bar{x})$, it follows that [5]

$$\frac{y_0 - a - b(x_0 - \bar{x})}{s\sqrt{1 + n^{-1} + (x_0 - \bar{x})^2 / S_{xx}}} \sim t_{n-2} \tag{9}$$

The standard uncertainty associated with the predicted value is:

$$u(y_0) = \sqrt{(n-2)/(n-4)}\,.\,s\sqrt{1 + n^{-1} + (x_0 - \bar{x})^2 / S_{xx}}$$

When n is large, (10)

$$u(y_0) = s\sqrt{1 + n^{-1} + (x_0 - \bar{x})^2 / S_{xx}}$$

The results obtained are shown in the third panel below (Fig. 3), where the graph displays the regression line. The dashed lines represent the uncertainties associated with the predicted values. These data are used to quantify the standard drift as well as to predict future values. The prediction results are very useful for the maintenance of the standards.

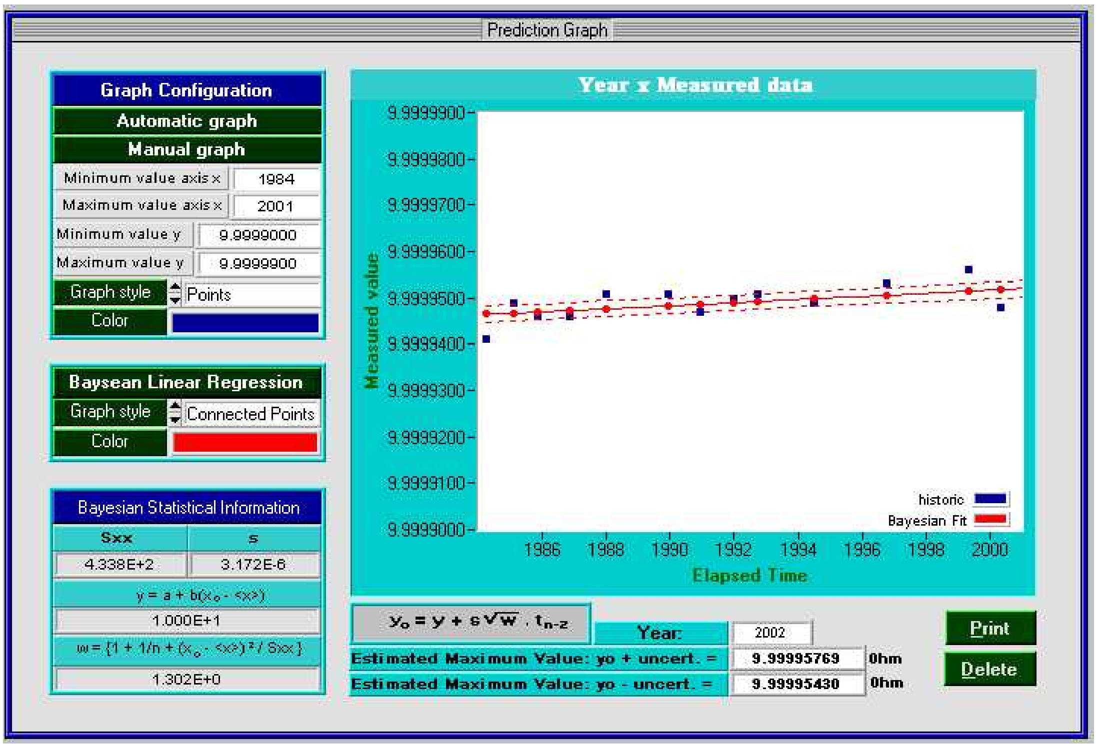


Fig. 3 – Display panel with the prediction graph

V. Results

As an example of the application of Bayesian statistics to the evaluation of stability of standards, tests were done with two different quantities. The following table shows the results obtained with standard inductors and standard resistors [8].

Proceedings of the 13th International Symposium on Measurements for Research and Industry Applications - IMEKO TC-4. Athens: National Technical University of Athens, 2004. v. I. p. 386-391.

Equipment	Model	Identification	Nominal value	Declared stability	Estimated stability
Standard Resistor	NBS/4020-B	R-5A	1 Ω	10 ppm	0.3 ppm
Standard Resistor	NBS/4020-B	R-5B	1 Ω	10 ppm	0.3 ppm
Standard Resistor	NBS/4020-B	R-6B	10 Ω	10 ppm	0.5 ppm
Standard Resistor	NBS/4020-B	R-7A	100 Ω	20 ppm	0.7 ppm
Standard Inductor	GenRad-1482L	I-B5	10 mH	0.01 %	0.01 %
Standard Inductor	GenRad-1482H	I-A6	100 mH	0.01 %	0.01 %
Standard Inductor	GenRad-1482T	I-A8	10 H	0.01 %	0.008 %

Table 1. Comparison between the stability declared by manufacturer and the stability estimated by Bayesian statistics.

The analysis of the results shows that:

1) The stability estimate of all standard resistors was smaller than that declared by the manufacturer. The results show that the manufacturer was very conservative;

2) The stability estimate of two standard inductors was equivalent to the manufacturer specification. The stability estimate of one of the standard inductors was smaller than that declared by the manufacturer. As the method works better for a large time interval, we believe that in the future we can have a reduction of the estimated value of the stability.

VI. Conclusion

The results obtained show that the automated system developed for the evaluation of the stability of measuring standards using Bayesian statistics is very useful. The calculations show that for several standards the stability estimate is much smaller than the value specified by the manufacturer. This reduced estimate leads to a smaller overall uncertainty of measurement and consequently to an improvement of the calibration process.

References

[1] Weise K and Wöger W, "A Baysian theory of measurement uncertainty", *Meas. Sci. Tech*, **4**, pp. 1-11, 1993.
[2] Lira I, Kyriazis G.A., "Bayesian inference from measurement information", *Metrologia*, **36**, pp. 163-169, 1999.
[3] Lira I and Wöger W, "Bayesian evaluation of the standard uncertainty and coverage probability in a simple measurement model", *Meas. Sci. Tech.* **12**, pp. 1172-9, 2001.
[4] Kyriazis G. A., Campos M.L.R., "Bayesian inference of linear sine-fitting parameters from integrating digital voltmeter data", *Meas. Sci. Tech*, **15**, pp. 337-346, 2004.
[5] Lee P. M., *Bayesian Statistics: An introduction*, Edward Arnold, London, 1989.
[6] Lindley D.V., *Bayesian Statistics: A Review. Philadelphia*, PA: S. I. A. M. Society for Industrial and Applied Mathematics, 1971.
[7] International Vocabulary of Basic and General Terms in Metrology, Geneva, International Organization for Standardization, 1993.
[8] Rocha G. M., Bulhões J. S., *Application of the Method of the Least-squares in the Study of the Stability of the Inductance Standards*, Monograph, Inmetro/CECO, Rio de Janeiro, Brazil, 2002. In Portuguese.