METHODS FOR THE DEFINITION OF CALIBRATION INTERVALS AND TO PERFORM COST-EFECTIVENESS ANALYSIS IN MANAGEMENT SYSTEMS

A. Silva Ribeiro[1], L. Lages Martins[1], J. Alves e Sousa[2] and R. Barreto Baptista[2]

[1] Laboratório Nacional de Engenharia Civil, Lisboa, Portugal. asribeiro@lnec.pt, lfmartins@lnec.pt
[2] Laboratório Regional de Engenharia Civil, Funchal, Portugal. jasousa@lrec.pt, rbaptista@lrec.pt

Résumé

La périodicité des opérations d'étalonnage est une décision technique qui apporte des implications considérables pour la performance du système de management. Bien que cette décision puisse être considérablement coûteuse, beaucoup d'entreprises et de laboratoires choisissent des périodicités d'étalonnage indépendant d'aucune analyse des données concernant l'utilisation et les différentes demandes d'exactitude associés à des applications spécifiques.

Une fois que qu'il y a une relation entre la périodicité d'étalonnage et l'attente de « défaillance » d'un instrument de mesure (ici, « défaillance » veut dire que l'instrument sera dehors tolérance quand on considère ses conditions d'utilisation), la combinaison de la nature stochastique de ce problème et des méthodes statistiques pour prévoir la meilleure périodicité d'étalonnage pose un défi pour beaucoup de systèmes de management.

Cet exposé présente des concepts et discute des modèles et leurs variables qui peuvent être utilisés pour définir des intervalles de temps entre deux étalonnages. Une approche de fiabilité pour une analyse coût-efficacité est aussi présentée. Un exemple illustratif de l'expérience des laboratoires métrologiques présentés au début est utilisé pour mettre en valeur quelques observations et les conclusions.

Abstract

The periodicity of the calibration operations is a technical decision with considerable implications in the system management performance. Although this decision can be significantly cost effective, many industries and laboratories apply recommended calibration intervals regardless of any type of data analysis concerning aspects such as the severity of use and the different accuracy requirements associated with specific applications.

Considering that there is a relation between the calibration intervals and the expectations of "failure" of a measurement instrument ("failure" meaning that the instrument will be out-of-tolerance considering its own usage requirements), to combine the stochastic nature of the problem with statistical methods to predict the optimized calibration interval is a challenge to many management systems.

This paper presents concepts, discusses methods, models and its variables that can be used to define calibration time intervals. Moreover, a reliability approach to a cost-effectiveness analysis is also presented. An illustrative example concerning the experience of the above mentioned metrological laboratories is used in order to enhance some of the remarks and conclusions.

1 – Introduction and motivations

Measurement systems are applied in a variety of contexts of management systems, and one of its major concerns is the quality of the measurement results. Quality requirements concerning measurement includes the measurement traceability obtained through instruments calibration and the assessment of the data obtained in this process.

The instruments performance is usually influenced by many different sources of uncertainty, according to stochastic behavior, not allowing a deterministic prediction of its metrological status in a specific moment to be made accurately. Therefore, in order to provide quality assurance of the measurements, it becomes necessary to perform, periodically, the calibration of instruments and to evaluate its conformity according to the intended use in a process designated by *metrological confirmation* [1]. This process leads to acceptance-rejection decisions with technical, commercial and human risks and also economical costs.

The definition of the calibration time interval has a major role in this process because the elapsed time between calibrations can affect the probability of accepting or rejecting the calibrated instrument due to its own reliability and, consequently, it has a direct influence in the decision risk and cost. In this context, the optimization of the calibration time interval should be considered as a major concern to the implementation of quality assurance in management systems[1], bearing in mind that under-specified quality can lead to the increase of technical, commercial or even human integrity risks; and over-specified quality leads to the increase of costs.

In a broad context including industry, testing and metrological laboratories, this process, however, encompass a framework of concepts and practices not always consensual, being a motivation for this paper the description of the approach that is applied in two secondary metrological laboratories and the analysis and discussion of some results.

Another motivation is due to the need to perform this type of analysis according to new metrology concepts oriented to a probabilistic approach of measurement [2], in contrast with the traditional approaches based on a deterministic point of view structured upon the error analysis and often using straightforward techniques.

---

[1] Legal metrology is considered outside the aim of this discussion as their calibration time intervals are defined according to regulations.

The economic and technical developments have introduced high expectancy regarding the high-performance of measurement systems obtained at lower risk and costs, being consistent with the main management systems ideal of continuous improvement of Quality. The aim of optimizing the relations established between these three management variables depends particularly upon the definition of the calibration time interval.

2 – Basic concepts and input information

The development of a procedure to optimize calibration time intervals specified in laboratory calibration plans is supported in a set of quality premises:

- traceability of calibration process to national or international standards;
- knowledge of measurement uncertainty regarding the calibration and the measurement process;
- knowledge of the measurement system intended use and the level of accuracy required;
- knowledge of the economical factors related to calibration and maintenance.

The parameter calibration time interval can be obtained using different approaches, as mentioned before. In the context of metrological laboratories (see Figure 1) it is considered that the relevant information required should be concerned with the measurement risk assessment, the metrological performance of the type of instrumentation in relation to the intended use and the cost analysis related to the calibration and maintenance processes.

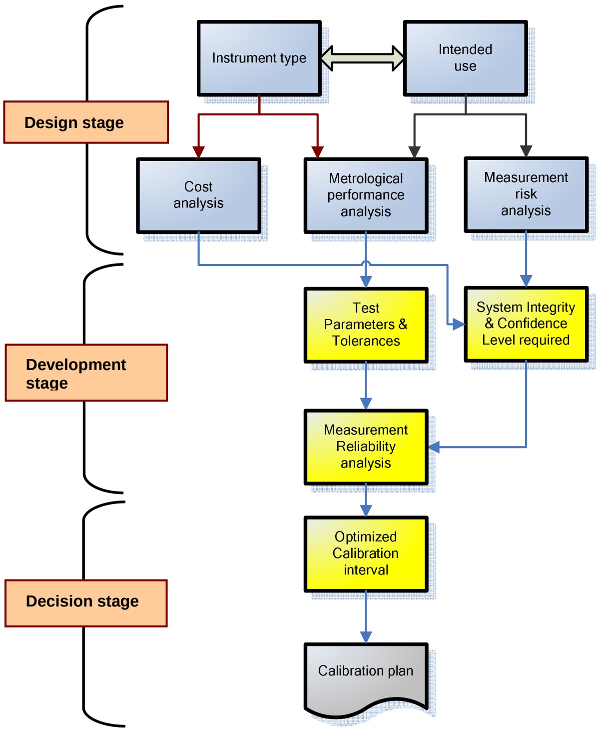

Figure 1 – Stages related to the definition of an optimized calibration time interval in a metrological laboratory.

Provided this information, metrological management should be able to establish the metrological parameters of interest, a *system integrity level* required and its confidence level, and to perform *reliability analysis*, emphasizing the need to determine the *reliability function* and *reliability target* depending on the methodology applied [3]. The optimization of the calibration time interval can be obtained combining the previous information, leading to the guidance established in the laboratory calibration plan.

In the following sections, some remarks are made concerning the processes used in order to obtain and apply such relevant set of information.

2.1 – Cost analysis

Cost analysis applied to the calibration interval of measurement instrumentation is a function of several management variables, needed to be identified and evaluated with respect to cost per unit of time.

There are two different types of costs to be considered, the direct costs (namely, those related to operational costs) and the indirect costs (namely, due to possible out-of-tolerance performance of instruments).

The first type of costs can usually be obtained from historical data concerning calibration and maintenance annual costs such as human resources occupation, materials and energy consumption, out-of-service daily costs and subcontract costs.

The second type of costs is obtained indirectly by estimating the costs related to non-conform work. This might have both objective and subjective contributions, being the first related to the time needed to identify non-conform work, to repeat calibration operations, or similar. The latter contributions arise from aspects such as the decrease of confidence of clients leading to possible loss of contracts.

Usually, the first type of costs decrease with longer calibration time intervals and the second type of costs has an opposite relation with the calibration time intervals (see Figure 2).

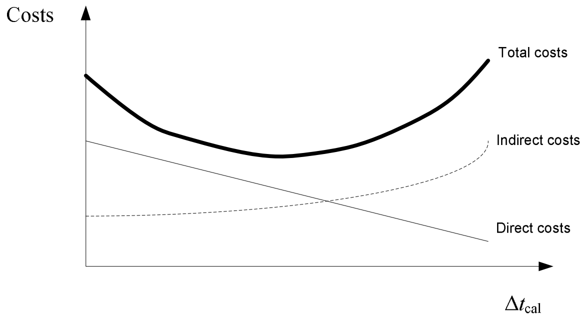

Figure 2 –Balance between direct and indirect costs.

From an economical point of view, the optimized calibration interval should be the longest possible. However, the decision should always be balanced considering that the indirect costs also increase with time and, therefore, need to be taken into account when establishing the criticality of the intended use.

2.2 – Measurement risk analysis

In a simplified approach adopted in the context of risk assessment, measurement risk analysis aims to identify the threats to measurement quality and their consequences within a probability frame. To achieve that purpose, a scale based on a parameter *SIL* (System Integrity Level) [4] is established so that a quantitative parameter can be used as guidance to the adoption of a confidence level and to select a method to obtain the optimized calibration time interval.

The major concern regarding measurement is related to the use of instrumentation in a condition of out-of-tolerance. In fact, this condition can be critical because it is often detected only during recalibration process after performing a large amount of non-conform work.

The approach proposed is applied to instrumentation independently of its function in a traceability chain (reference standard, transfer standard or working standard), however, this fact is quite relevant for the definition of the criticality related to the instrumentation intended use.

The *SIL* parameter has four categories (1 to 4) [4], being defined from the analysis of two input parameters[2]: the application **criticality,** p_{crit}; and the **complexity,** p_{comp}, of the measurement intended use. Each parameter has four levels as described in Table 1 and Table 2.

Criticality	Description
Low (1)	The economical and technical effects due to the measurement system performance are not significant (e.g., related with monitoring systems).
Moderate (2)	The economical and technical effects due to the measurement system performance can be significant but can be overcome by strategies based in data analysis (e.g., compensating influence quantities contributions).
High (3)	The economical and technical effects due to the measurement system performance can affect significantly the output data (e.g., commercial effect).
Very high (4)	The economical and technical effects due to the measurement system performance can be critical to the overall use (e.g. human integrity effect).

Table 1: Criticality parameter quantification.

Complexity	Description
Low (1)	Simple and direct functional relation of input data (e.g., linear relation)
Moderate (2)	Functional relation with explicit mathematical models (e.g., non-linear relations using known functions)
High (3)	Functional relations based in known algorithms requiring simple data analysis tools
Very high (4)	Complex, implicit functional relations requiring advanced data analysis tools

Table 2: Complexity parameter quantification.

The definition of the *SIL* parameter is based on the following expression

$$SIL = \text{INT}\left(w_{\text{crit}} \cdot p_{\text{crit}} + w_{\text{comp}} \cdot p_{\text{comp}}\right), \qquad (1)$$

where w_{crit} and w_{comp} represent, respectively, the weights of the criticality and complexity input parameters. Considering $w_{\text{crit}} = 0{,}7$ and $w_{\text{comp}} = 0{,}3$ (empirical values based on studied metrological applications), a combination matrix of the two input variables is presented in Table 3. The values of *SIL* are used in order to guide the user to the selection of the confidence level (as presented in Table 4) and to the method to apply (as presented in Chapter 3).

	Criticality			
Complexity	Low (1)	Moderate (2)	High (3)	Critical (4)
Low (1)	**1**	**1**	**2**	**3**
Moderate (2)	**1**	**2**	**2**	**3**
High (3)	**1**	**2**	**3**	**3**
Very high (4)	**1**	**2**	**3**	**4**

Table 3: SIL parameter definition.

The confidence level (1-α) is a parameter required to establish the *in-tolerance* and *out-of-tolerance* domains, expressing *the probability that the result obtained lies within the in-tolerance interval of the measurement*. Considering that gaussian conditions apply, a relation between the *SIL* parameter and the standard deviation parameter, σ, was defined in order to support the definition of the confidence levels required [3] (Table 4).

SIL	**Confidence level** (1-α) / %	**N. of standard deviations ($n \cdot \sigma$)**
1	86,64	1,5
2	95,45	2,0
3	99,73	3,0
4	99,994	4,0

Table 4: Relation between SIL parameter and confidence level under Gaussian assumptions.

2.3 – Metrological performance analysis

It is widely accepted that instrumentation changes its measurement uncertainty along the time, being permanently affected by sources of uncertainty with stochastic behavior that promote a progressive increase of the measurement uncertainty. This parameter is particularly suited to be considered the test parameter because it allows an analysis of the instrumental accuracy evolution. Figure 3 presents a typical behavior of the uncertainty interval changes obtained at different time between two consecutive calibration operations [5].

Moreover the intended use usually has its own requirements, namely, related to testing or calibration standards and experimental conditions. These requirements should serve to define the in-tolerance and out-of-tolerance limits in a consistent way with the test parameter target intervals (measurement uncertainty intervals).

---

[2] Adopting a process that is similar to the one used in software validation processes.

[3] Regarding Table 4, it should be mentioned that some authors (e.g., [6]) recommend higher discrimination of levels (namely, 5 to 7) and, consequently, of the confidence levels (from 1,0σ to 5,0σ, this last one corresponding to 99,99994 % of confidence level).

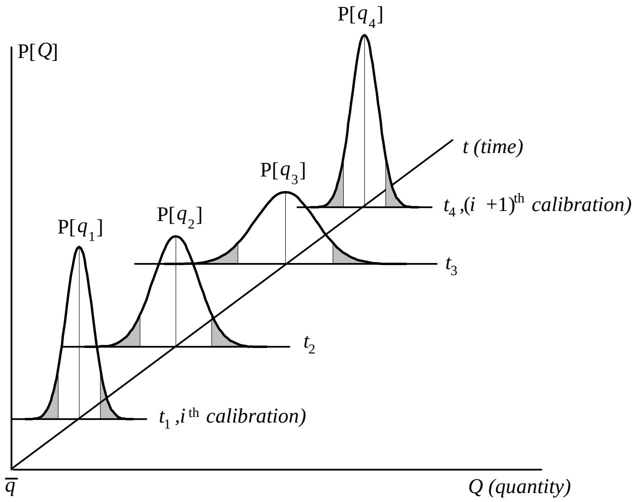

Figure 3 – Progressive increase of measurement uncertainty along time between two consecutive calibrations.

The measurement uncertainty can be expanded within a confidence interval established in conformity with measurement risk analysis (presented in previous section, see Table 4), establishing limits that can be used to define the in-tolerance and out-of-tolerance domains. In this way, it can be defined both the testing parameter (expanded measurement uncertainty) and the tolerance (as prescribed in Fig. 1).

Reliability analysis applied to a type of instruments used in certain type of application, uses the information provided by the comparison of test parameters with tolerance limits to define the reliability function and the reliability target, widely used in statistical methods [3] applied in the definition of the optimized calibration intervals (see Chapter 3).

3 – Methods and modelling

The determination of the optimized calibration interval is "*a complex mathematical statistical process requiring accurate and sufficient data taken during the calibration process. There appears to be no universally applicable single best practice for establishing and adjusting the calibration interval*" [7].

The same document, however, expresses that there is some consensus in a set of methods that are considered suitable for this purpose:

- automatic adjustment or "staircase";
- control-chart;
- "in-use" time;
- black-box testing;
- other statistical approaches.

Considering the nature of the measurements performed in the metrological laboratories that support this study, the first type of methods is considered particularly suitable to apply to most of working standards and the statistical approach, being able to provide a more rigorous analysis, is considered suitable to apply to reference and transfer standards, taking into account the criticality of its intended use.

Therefore, there are two main approaches that are considered in this context:

- reactive methods (adjustment of data based on calibration history, as a variation of automatic adjustment);
- statistical approach based on the reliability function and using the maximum likelihood method.

The first type is used for instrumentation with lower integrity requirements (typically of class 1 or 2) whereas the second type is usually applied to instrumentation having a higher level of integrity required (class 3 or 4).

Being these two the main approaches adopted, they will be described with some detail. Detailed reading of [3], [6] and [7] is recommended.

The *reactive methods* or *calibration interval adjustment methods* [3] are established upon the recent calibration information available, being the most direct approach the *Simple Response Method*, following the model (with the iteration counter n = 0, 1, …):

$$t_{n+1} = t_n\left(1+\Delta t_{\mathrm{a}}\right)y_n + t_n\left(1-\Delta t_{\mathrm{b}}\right)\bar{y}_n \qquad (2)$$

being t_n the duration of the n$^{\mathrm{th}}$ calibration interval, $\Delta t_{\mathrm{a}} > 0$ and $\Delta t_{\mathrm{b}} > 0$ correction time intervals, y_n a Boolean variable equal to 1 if the calibration result is "in-tolerance" and 0 if it is "out-of-tolerance" ($\bar{y}_n$ represents the complement of y_n). Considering the aim to achieve a reliability target, this method has a recognized very low convergence, being strongly recommended to use other methods such as *Incremental Response Method* or *Interval Test Method* in order to achieve better results. The major critic that is pointed out to this approach concerns the fact that these methods react to calibration data and do not attempt to model the uncertainty growth function. However, they are considered adequate for many low criticality applications.

The statistical approach is related to the idea of modeling the reliability function that best fit the calibration data and to estimate the optimized calibration time interval knowing the *reliability target*, R^* (Fig. 4).

The approach used to establish the value of the reliability target, R^*, follows the rough guidelines of [6], defining ranges of reliability target[4] according to specific conditions, namely, testing uncertainty impact in the output measurement uncertainty and its usefulness. Table 5 exhibits the relation proposed.

Condition	Reliability target %
Low impact on final measurement or large in-tolerance range	≤ 60
Redundant application or medium criticality (*SIL* 2-3)	60 to 90
Critical application or absence of backup procedures	≥ 90

Table 5: Conditions for the definition of reliability target.

---

[4] Reliability target is usually defined considering EOP (End–of–Period), AOP (Average–of–Period) related to the calibration interval period.

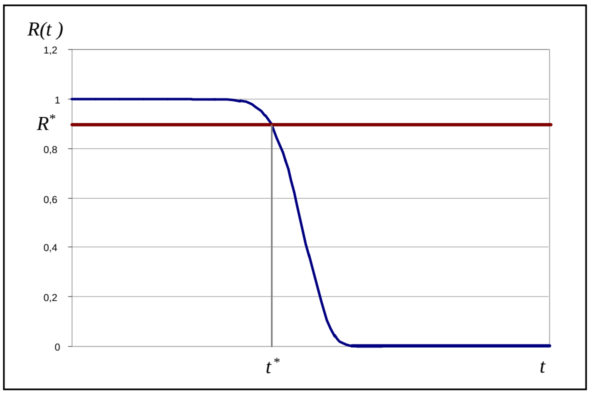

Figure 4 – Finding the optimized calibration interval from a Weibull reliability function and a defined reliability target.

The reliability function $R\left(t,\hat{\theta}\right)$ best fit to calibration experimental data is obtained from the Maximum Likelihood Estimation method (MLE). The optimized calibration time interval, t^* is obtained solving the equation,

$$R\left(t,\hat{\theta}\right)=R^* . \qquad (3)$$

Recommendations aiming this purpose can be found [3], namely, applying known mathematical functions that can fit the reliability function (Exponential; Weibull; Mixed Exponential; Random Walk; Restricted Random Walk; Modified Gamma; Mortality Drift; Warranty; Drift; and Lognormal), being selected accordingly to out-of-tolerance rates performance [6]. To solve equation (3) a two steps process is recommended [3]: firstly, an analytical approach or Newton-Raphson numerical method should be used; if convergence is not properly achieved, a trial-and-error approach should be applied.

4 – Metrological laboratory application

The study concerning an application refers to the determination of the heat capacity quantity related to a calorimeter (Fig. 5) used in the heat of combustion determination test, being carried out at a reaction to fire laboratory dedicated to the assessment of building products.

Figure 5 – Experimental apparatus related to the heat of combustion determination test at LNEC (from left to right: water supply equipment, refrigerator and calorimeter).

The heat capacity of the calorimeter is determined by performing, at least, five consecutive complete combustions of benzoic acid pellets at the calorimeter. The benzoic acid used is a NIST certified reference material in terms of its heat of combustion, being a relevant input quantity in the determination of the heat capacity of the tested calorimeter and, consequently, in the determination of the heat of combustion associated with a tested sample of a building product for which the calorimeter is used.

In this case, the costs associated with this metrological test are significantly increased by the use of certified benzoic acid. In addition, according to the test standard [8], the calorimeter shall be tested at regular intervals not greater than two months (considering that the main components of the calorimeter are not changed) regardless of the nature and frequency of the application. Considering this requirement, it becomes useful to estimate an optimized calibration interval for the laboratorial conditions and to discuss whether the above mentioned normative test interval is under or over specified.

The aforesaid study uses data (Table 6) based on the metrological tests performed with the LNEC's calorimeter.

Time (months)	Reliability (%)
1	0,99
2	0,98
3	0,94
4	0,82
5	0,68
6	0,62
7	0,39
8	0,31
9	0,18
10	0,16

Table 6: Experimental data.

This type of test is considered to be of very high criticality (level 4) since it is related with fire safety of building products, and moderate complexity (level 2), which gives a *SIL* of 3 according with Table 4 leading to the use of 3σ confidence interval to perform in-tolerance test. A EOP reliability target of 90 % was adopted due to the criticality of the intended use and according with Table 5.

Regarding the cost analysis, two main contributions are found: the (direct) cost of the reference material and the (indirect) cost related to a possible use of the calorimeter in out-of-tolerance conditions. The first one can be easily quantified and has a monotonic decrease along the period of observation (10 months). The second increases significantly (because of the critical impact of the testing) after reliability function finds reliability target.

In order to find best fit functions to experimental data, two types of functions were applied: Weibull (4) and exponential (5). In each case, the best fit function was found by the linearization of the functions and applying regression analysis. The reliability best fit functions parameters obtained are given in Table 7 and the approximations can be seen in Figure 6 within the reliability target limit,

$$R(t;\lambda,k)=e^{-(\lambda t)^k} \qquad (4)$$

$$R(t;\lambda)=e^{-(\lambda t)} \qquad (5)$$

being the best fit parameters to obtain: λ, in both cases; and *k*, in the Weibull function case (the exponential function can be considered as a particular case of the Weibull function with k = 1). In both cases, the reliability functions are 1 for $t = 0$, otherwise, it would be necessary to consider an R_0 value in expressions (4) and (5).

Function	Parameter	Value
Weibull	λ	2,5013
	k	0,1295
Exponential	λ	0,0581

Table 7: Best fits parameters.

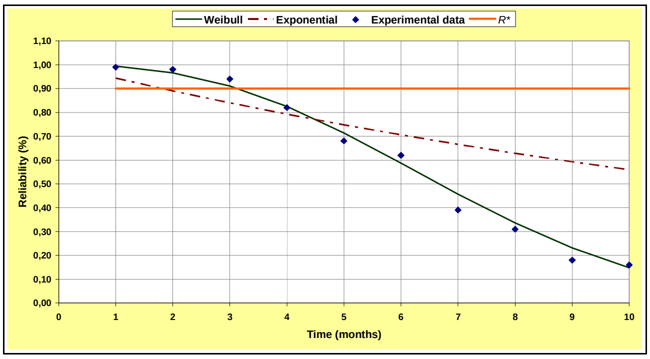

Figure 6 – Experimental data, Weibull and exponential reliability function best fits and reliability target.

The estimates of the optimized calibration time interval, for each best fit function, were obtained using the analytic approach. Table 8 presents those best estimates.

Function	Estimate of optimized calibration time interval (months)
Weibull	3,1
Exponential	1,8

Table 8: Optimized calibration time intervals estimates.

The analysis of the results achieved allows to conclude that both estimates are not far from the normative calibration time interval. Figure 6 shows that the Weibull model is a better fit to experimental data and, therefore, the best estimate obtained is 3,1 months.

According with these results and considering the specific testing application, the calibration time interval could be changed to three months instead of the conservative requirement of two months, meaning a time increase of 50 %. In fact, one of the aims of the total quality improvement is to apply scientific methods in order to obtain the best technical and economical decisions, allowing the management systems to develop appropriate strategies.

5 – Conclusions and future trends

The relation between technical and economical requirements is becoming a major concern in management systems, being the definition of the instrumentation calibration interval a type of decision that should account for, combining information from three main parameters: costs, risk and reliability.

Different experimental approaches can be used to achieve the aim of obtaining optimized calibration time interval, from a simple reactive method to a more complex statistical approach. In this context, this paper presents a strategy, adopted by two metrological laboratories, able to establish integrity categories based in criticality and complexity criteria, and use it to select specific methods considering the three main issues mentioned above, to find the optimized calibration time intervals, which is the main aim of this type of studies.

The illustrative study carried out enhances how these different steps can support the relevant decision of defining the calibration time interval, in a context of total quality improvement.

It should be emphasized that the results presented were obtained using the LNEC's calorimeter test data leading to optimized calibration interval estimates that can only be considered in this case. The main reason for this discussion is that this type of studies (based on experimental data) can be useful to establish different calibration intervals according with reasonable arguments as the nature of the use. In this case, the results agree with the normative recommendations.

6 – References

[1] *International vocabulary of metrology – basic and general concepts and associated terms (VIM),* Joint Committee Guides in Metrology (JCGM), 2008.

[2] *Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method,* Joint Committee Guides in Metrology (JCGM), 2008.

[3] *Establishing and adjustment of calibration intervals,* Recommended Practice RP-1, NCSL International, January 1996.

[4] Wichmann, B, Parkin, G. and Barker, R. – *Software Support for Metrology Best Practice Guide No. 1, Validation of Software in Measurement Systems*. National Physical Laboratory Report DEM-ES014, Jan. 2007.

[5] Castrup, H., Techniques for optimizing calibration intervals, ASNE Test & Calibration Symposium, Arlington, December 1994.

[6] *Metrology – Calibration and measurement processes guidelines*, NASA Reference Publication 1342, Jet Propulsion Laboratory (JPL), June 1994.

[7] ILAC-G24/OIML D10 – *Guidelines for the determination of calibration intervals of measuring instruments.* International Laboratory Accreditation Cooperation (ILAC) and Organisation Internationale de Métrologie Légale (OIML), 2007.

[8] EN ISO 1716 – *Reaction to fire tests for building products. Determination of the heat of combustion.* European Committee for Standardization (CEN), Bruxelles, February 2002.