

# Uncertainty-based determination of recalibration dates

Janik Schaudé\*  and Tino Hausotte 

Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Chair of Manufacturing Metrology, Nögelsbachstr. 25, 91052 Erlangen, Germany

Received: 4 November 2022 / Accepted: 15 November 2023

**Abstract.** Traceability is of vital importance in metrology and is achieved by an unbroken chain of calibrations that relate the measurement result to a reference. Less clear is the temporal aspect of traceability, namely the determination of recalibration dates. Relevant standards require the conduction of recalibrations in a planned manner, and thus in general there is a calibration history consisting of a number of past calibrations for measurands that are part of the traceability chain. Nevertheless, commonly only the results of the last calibration of the standards of the traceability chain are considered when determining a measurement result. Furthermore, recalibration dates are often determined in a rather unscientific manner. Within this paper, a method is proposed to predict the current value of a measurand along with its uncertainty, taking into account all past calibrations. Based on the predetermined target measurement uncertainty of the measurand, it is possible to recognize the need for and thus to date recalibrations. The applicability of the method is investigated by examining the metrological compatibility of the predicted results with the results of subsequent calibrations of the calibration histories of a number of different standards. There is an explainable limitation of the applicability of the method to calibration histories that exhibit a correlation of time and the chosen calibration laboratory. However, the paper leads the way of future research to remedy the currently unsatisfying state regarding the handling of calibration histories and the determination of recalibration dates.

**Keywords:** calibration / calibration interval / measurement uncertainty / traceability

## 1 Introduction

In the era of globalization and interchangeability, accurate and internationally comparable measurement results are indispensable in metrology [1]. On the downside, a measurement might always be subject to wholly or in part unknown impacts, and therefore a measurement result is only an estimate of the measurand's value [2]. To indicate the reliability or the quality of a measurement, the measurement result is accompanied by the measurement uncertainty evaluated according to the *Guide to the expression of uncertainty in measurement* (GUM) [2]. Metrological comparability of measurement results is ensured if the measurement results are traceable to the same measurement unit, where metrological traceability is defined by the *International vocabulary of metrology* (VIM) as “property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty” [3, 2.41]. Therefore, measurement uncertainty and metrological traceability are linked inextricably [4]. It is common to depict the traceability

chain as pyramid, where the broadening downwards indicates the necessarily increasing measurement uncertainty [5–9].

The importance of metrological traceability is also reflected in internationally recognized standards like the ISO 9001, which states: “When measurement traceability is a requirement, or is considered by the organization to be an essential part of providing confidence in the validity of measurement results, measuring equipment shall be: a) calibrated or verified, or both, at specified intervals, or prior to use, against measurement standards traceable to international or national measurement standards; (...)” ([10], 7.1.5.2). ISO 17025, which states the general requirements for the competence of testing and calibration laboratories, demands the metrological traceability of the measurement results reported by the laboratory as well ([11], 6.5). Therefore, calibration of the measuring equipment that is required to establish the metrological traceability, as well as the establishment of “a calibration program, which shall be reviewed and adjusted as necessary in order to maintain confidence in the status of calibration” ([11], 6.4.7) is mandatory. In clause 6.4.13 e) it is stated to retain records for the equipment on “calibration dates, results of calibrations, adjustments, acceptance criteria, and the due date of the next calibration

\* Corresponding author: [janik.schaude@fmt.fau.de](mailto:janik.schaude@fmt.fau.de)

or the calibration interval” ([11], 6.4.13 e). For dimensional measurements such equipment might be dimensional artefacts like gauge blocks, spheres, hole plates or surface texture artefacts [9,12]. While both standards, ISO 9001 and ISO 17025, call for the specification of calibration intervals, i.e., the planned periodic recalibration of measuring equipment that is required to establish metrological traceability, nothing is said within the standards about the way to determine the corresponding intervals or dates.

In Section 2, literature on the determination of recalibration dates or calibration intervals is reviewed. Some of the literature does not originate from the field of metrology and therefore does not stick to basic metrological terminology as defined within the VIM. Unfortunately, often it is not possible to “translate” this literature into metrological terminology, as the underlying concepts are completely different. Consider the term “reliability”, which has a well defined meaning and is well established within some scientific fields like integrated logistics support ([13], 4.1), but it is just not comparable to the concept of measurement uncertainty. Both terms might be seen as incommensurable, to use this term brought into scientific theory by Thomas S. Kuhn [14]. Similar is true for the maximum permissible error (MPE), although it is defined in the VIM as “extreme value of measurement error, with respect to a known reference quantity value, permitted by specifications or regulations for a given measurement, measuring instrument, or measuring system” ([3], 4.26). While there is some guidance on how to transform a MPE into a standard uncertainty ([2], F.2.3.3), no such guidance exists for the opposite direction in the majority of cases. On the contrary, in Annex E of the GUM it is stated explicitly that the method of the GUM stands in contrast to the “idea (...) that the uncertainty reported should be ‘safe’ or ‘conservative’, meaning that it must never err on the side of being too small” ([2], E.1.2). In particular, the normal distribution, which might be seen as the most relevant probability distribution due to the central limit theorem ([2], G.2) neither has an upper nor a lower boundary and therefore there is no maximum or minimum possible measurement error (although physics itself might set boundaries to reasonable measuring results).

In Section 3, a method to determine recalibration dates based on measurement uncertainty is proposed. The method is especially suited for calibration laboratories, as there is no need to cross the boundaries of the field of metrology. But it is just as suitable for the measuring equipment management in industrial facilities. Doubt about the current value of a measurand increases with time passed since the last calibration. The presented method deals with this doubt, as usual in metrology, as a source of uncertainty. The method can therefore also be used in the sense of risk management required by ISO 9001. Although examples are taken from the field of dimensional metrology, the method is not limited to this particular field of metrology.

## 2 Literature review

The most simple calibration program is to determine a calibration interval for each measuring equipment initially and to keep this interval fixed [15]. The German

Accreditation Body (Deutsche Akkreditierungsstelle GmbH, DAkkS) recommends calibration intervals for measuring equipment for laboratories of some specific areas [16,17], but only few equipment falls within the range of dimensional metrology. While the DAkkS recommends the calibration intervals based solely on the type of equipment, in a document published jointly by the International Laboratory Accreditation Cooperation (ILAC) and the International Organization of Legal Metrology (OIML), several factors which should be considered when determining the initial calibration interval are listed [18]:

- The instrument manufacturer’s recommendation
- Expected extent and severity of use
- The required uncertainty in measurement
- Maximum permissible errors (e.g., by legal metrology authorities)
- Adjustment of (or change in) the individual instrument
- Influence of the measured quantity (e.g., high temperature effect on thermocouples)
- Pooled or published data about the same or similar devices

The fifth bullet point relates to the stability of the measuring instrument and the instrumental drift as defined within the VIM ([3], 4.19 and 4.21). Nevertheless, no guidance is given on how to incorporate these factors when determining the calibration interval. The decision is left to a “person with general experience of measurements, or of the particular instruments to be calibrated, and preferably also with knowledge of the intervals used by other laboratories” [18].

In a document applicable for coordinate metrology issued by the DAkkS it is stated to set the initial calibration interval in the absence of a calibration history to no longer than one year. It might be adjusted when analysing the drift behaviour and considering the measurement uncertainty. Nevertheless, no guidance is given on how to conduct such an analysis and how to incorporate the uncertainty [19].

A standard published jointly by the Association of German Engineers (Verein Deutscher Ingenieure, VDI), the Association for Electrical, Electronic & Information Technologies (Verband der Elektrotechnik, Elektronik & Informationstechnik, VDE), the German Quality Society (Deutsche Gesellschaft für Qualität, DGQ) and the German Calibration Service (Deutscher Kalibrierdienst, DKD) states that there “are no generally binding specifications for the timing of a recalibration” ([20], 5.5). For the determination of the first recalibration date, an empirical value (based on experience with similar equipment) might be necessary. Furthermore, the presence of a calibration history enables the adjustment of the calibration interval, where “the most important criterion for the adjustment is the change in the characteristic value of the measuring and test equipment” ([20], 5.5). ISO 10012 and DIN 32937 are referenced for procedures for the determination of calibration intervals [20].

ISO 10012, which states the requirements for measurement processes and measuring equipment, also calls for the analysis of the calibration history, as “data obtained from calibration (...) histories (...) may be used for determining

intervals (...)” ([21], 7.1.2). However, also within this document no further guidance is given on how to incorporate the historic data into the decision about the calibration interval. DIN 32937 states several factors to be considered when determining the calibration interval of a test equipment [22]:

- Experiences from past tests or calibrations
- Stability of the test equipment
- Test process control
- Technical specifications (e.g., standards, customer requirements or other technical regulations)
- Wear (e.g., caused by different materials of the test equipment and the tested object)
- Experiences with similar test equipment
- Manufacturer’s recommendation
- Safety regulations
- Economical considerations and risk assessments
- Type of stress, static or dynamic measurements
- Experiences from the engagement of the test equipment (need for repairs, malfunctions, environmental influences, frequency of use, functional reliability)

But also within this document, no further guidance is given on how to incorporate these many factors when determining the calibration interval.

The already mentioned guideline jointly issued by the ILAC and the OIML contains five possible methods to review calibration intervals [18]:

Method 1: Automatic adjustment, where calibrations are conducted on a routine basis and the subsequent interval is extended or reduced based on the outcomes of the calibration. Nevertheless, no clear guidance is given on how to assess if the interval should be extended or be reduced.

Method 2: Control chart (calendar-time), where the results of calibrations are plotted against time to derive the dispersion of the results and drifts. It is recommended to calculate the optimum calibration interval from these plots, but no further guidance is given on how to do this.

Method 3: In-use time, which is similar to the previous method but instead of plotting the calibration results against time they are plotted against hours of use. This approach to consider the time of use rather than the passed time since the last calibration is common in preventive maintenance in general ([23], p.219) and also for the maintenance of measuring machines [24].

Method 4: In service checking, where critical parameters of an instrument are frequently checked against a portable calibration gear and a full calibration of the instrument is conducted only if errors are observed. Nevertheless, this in somehow outsources rather than solves the problem of determining a recalibration interval, as now the recalibration of the portable calibration gear is of vital importance.

Method 5: Other statistical approaches, that are based on statistical analysis. [25] is mentioned as an example.

The basic statement of [25] is that the uncertainty about the value of a measurement standard’s measurand grows with time and it is necessary to predict this value along with its uncertainty at the time of use. The value and its uncertainty are calculated based on the calibration history of the particular standard. Given an upper limit for a standard’s uncertainty in-use, it is possible to derive the

date of recalibration. Although the method to predict a standard’s value and its uncertainty is rather complicated, it does not take the uncertainties of the past calibration values into account. Nevertheless, one might reasonably argue that these uncertainties do have an influence on the uncertainty of the prediction, especially if the number of past calibrations is rather small, which is usually the case.

The approaches described in [26–28] are quite similar and rely on the determination of measurement reliability. For the measuring equipment a target reliability and tolerance limits are defined. Grouped measurement devices are tested after some time since the last calibration and the percentage of out-of-tolerance devices is determined. Based on this percentage, the target reliability and the time passed since the last calibration, the next calibration date is calculated. Nevertheless, it remains unclear how to incorporate the measurement uncertainty into the decision about the out-of-tolerance state of a device. Also the remarks about the determination of the reliability target and the tolerance limits remain vague. Furthermore, the approach described in [26] relies on homogeneous past calibration data, e.g., regarding the calibration procedure and the reported uncertainty. But this requirement might not be met in practice as calibrations are often conducted externally and the procedure as well as the uncertainty might change over the years. In some ways the requirement for an unvarying calibration procedure contradicts the concept of metrological comparability, which states that measurement results are comparable if they are traceable to the same unit regardless of the measurement procedure or the measurement uncertainty ([3], 2.46). But the most serious shortcoming of the approaches is the waste of information: The calibrated values are discretised into the two groups within-tolerance and out-of-tolerance and further calculations are based on the number of items per group. Nevertheless, much information is lost by this discretization, as for example a drift might be recognizable with the (more or less value-continuous) calibrated values before the discretization but not after discretization.

In [29] an uncertainty-based method for the determination of recalibration dates is presented. Based on the last two calibration results the systematic drift of the measurement equipment is evaluated as linear polynomial. Furthermore, the uncertainty of this prediction is evaluated by a Monte Carlo simulation considering the calibration uncertainties. The date of the next calibration is determined considering the pre-determined MPE of the device. After the next calibration, the procedure is either repeated with the results of the last two calibrations and a linear polynomial, or a non-linear polynomial is fitted to the results of more than the last two calibrations. While the linear model is recommended for robust measurement devices, non-linear models are recommended for devices which exhibit more rapidly changing characteristics. Nevertheless, especially in the case of robust measurement devices or standards like end-gauges, short-term deviations might rather be associated to changes of the calibration instrument than to changes of the standard itself [30]. It would therefore be beneficial to use as many data points as possible to predict the behaviour of the standard and the linear polynomial should be fitted to the results of all past calibrations.

### 3 Uncertainty-based determination of recalibration dates

Dimensional artefacts are common to achieve metrological traceability in dimensional metrology [9]. A calibration of such an artefact is hardly more than the (traceable) measurement of a particular measurand of the artefact. Thus, the result of the calibration is the best estimate of the value of this measurand along with the associated uncertainty about this value. In the presence of a calibration history, i.e., an amount of measurements of the same measurand, it might be possible to state the best estimate of the value of this measurand with more confidence and therefore with a decreased uncertainty compared to the uncertainty stated in the last calibration. On the other hand, this confidence diminishes with the time since the last calibration. The basis of an uncertainty-based determination of recalibration dates is therefore a method to obtain a best estimate about the current value of a measurand based on all available information (and therefore not merely on the last calibration) along with the associated uncertainty. Given a target measurement uncertainty  $u_T$  as defined by the VIM as “measurement uncertainty specified as an upper limit and decided on the basis of the intended use of measurement results” ([3], 2.34) for the value of the measurand, it is possible to recognize the need for recalibration. In general,  $u_T$  is one of the key characteristics for the management of measuring and test equipment [22].

#### 3.1 Allocation of the target measurement uncertainty

The starting point is the definition of  $u_T$  as standard uncertainty for the measurements to be conducted. The determination of  $u_T$  is a strategic decision and for calibration laboratories the choice might be based on factors like customer needs or available resources. For testing laboratories the tolerances to be tested might be considered, since a general rule states that the uncertainty should not exceed one tenth to one fifth of the tolerance [31]. Based on the measurement uncertainty budget,  $u_T$  of the complete measurement is allocated to the uncertainties of the  $n$  input quantities  $x_i$  ( $i = 1, \dots, n$ ) of the measurement. In practice, the standard uncertainty of most  $x_i$ ,  $u(x_i)$ , is fixed due to the availability of the measuring equipment, environmental conditions, etc. and it is primarily the uncertainties of the dimensional artefacts that might be altered if needed for example by choosing a different calibration laboratory able to provide calibrations with a smaller measurement uncertainty.

To show the determination of  $u_T$  for the value of a measurand of a standard, the example of the GUM for the calibration of an end-gauge is taken [2, H.1]. A sufficiently exact measurement function  $f$  for the determination of the length of the end-gauge at a temperature of 20 °C, denoted as  $l$ , by a comparison measurement with a standard with the known length  $l_s$  and the standard uncertainty  $u(l_s)$  is given by

$$l = f(l_s, h, \alpha_s, \theta, \Delta\alpha, \Delta\theta) = l_s + h - l_s(\Delta\alpha \cdot \theta + \alpha_s \cdot \Delta\theta), \quad (1)$$

where  $h$  denotes the difference in length between the measurements of the end-gauge and the standard,  $\alpha_s$  denotes the coefficient of thermal expansion of the standard,  $\theta$  denotes the deviation in temperature from 20 °C of the end-gauge to be calibrated,  $\Delta\alpha$  denotes the difference between the coefficient of thermal expansion of the end-gauge and of the standard and  $\Delta\theta$  denotes the difference in temperature between the end-gauge and the standard. The combined standard uncertainty  $u_c$  of  $l$  under the assumption of uncorrelated input quantities  $x_i$  is given by ([2], 5.1.2)

$$u_c(l) = \sqrt{\sum_{i=1}^6 \left( \frac{\partial f}{\partial x_i} \right)^2 \cdot u^2(x_i)}, \quad (2)$$

where  $u(x_i)$  is the standard uncertainty of  $x_i$  and  $x \in \{l_s, h, \alpha_s, \theta, \Delta\alpha, \Delta\theta\}$ .

Let's assume that due to strategic decisions,  $u_T$  for the calibration of an end-gauge with a similar coefficient of thermal expansion like the standard and with a nominal length of 10 mm has been determined to be no more than 20 nm, and therefore  $u_T^2(l) < 400 \text{ nm}^2$ . Furthermore,  $\Delta\theta$  is estimated to be zero, measurements are taken within an environment that might have a deviation from 20 °C of up to 0.5 K and  $\alpha_s$  is given as  $11.5 \times 10^{-6} \text{ K}^{-1}$ . The available measuring equipment allows the determination of  $h$  with an uncertainty of 11 nm.

In Table 1, the corresponding measurement uncertainty budget is shown. Note that  $-l_s\theta$  equals  $-5 \text{ mmK}$  and that  $-l_s\alpha_s$  equals  $-0.115 \times 10^{-6} \text{ mK}^{-1}$ . To achieve  $u_T(l) < 20 \text{ nm}$ ,  $u(l_s)$  must be less than 15 nm. If the standard is calibrated for the first time, then obviously the standard uncertainty  $u(l_s)$  of this calibration needs to be lower than  $u_T(l_s)$ . The first recalibration date might be determined considering factors already discussed in Section 2. In the following, a method to obtain a best estimate of the current value of a measurand along with the associated  $u$  in the presence of a calibration history, i.e., more than one past calibration of the particular measurand, is presented.

#### 3.2 Obtaining the best estimate of the current value of a measurand along with the uncertainty

For the sake of vividness, we take a real example to present the issue and the method to obtain a best estimate of the current value of a measurand along with the associated  $u$  in the presence of a calibration history. In Table 2, the calibration history of the diameter  $d$  of a reference sphere with a nominal  $d$  of 15 mm is listed and it is depicted in Figure 1. The calibrations have been conducted by a calibration laboratory accredited by the DAkkS or formerly by the DKD.

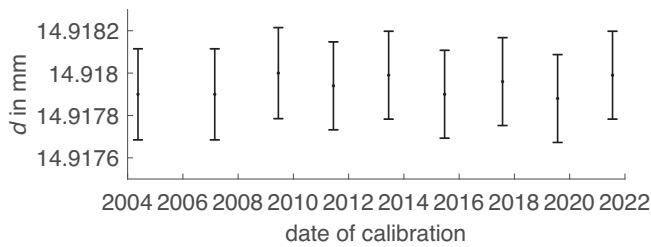
The diameter of the sphere has been calibrated about every two years since 2004, so in sum  $d$  has been calibrated nine times ( $N=9$ ). If there is no reason to assume that there is a fundamental change about the value of the measurand (e.g., caused by a crash), then to obtain the best estimate about the current value of the measurand all available data should be used and therefore all past

**Table 1.** Measurement uncertainty budget with  $u_T(l_s)=15$  nm to achieve  $u_T^2(l) < 400$  nm<sup>2</sup>.

$i$	$x_i$	$c_i \equiv \frac{\partial f}{\partial x_i}$	$u(x_i)$	$ c_i  \cdot u(x_i)$	$\sum_{j=1}^i [ c_j  \cdot u(x_j)]^2$
1	$h$	1	11 nm	11 nm	121 nm <sup>2</sup>
2	$a_s$	0	–	0	121 nm <sup>2</sup>
3	$\theta$	0	–	0	121 nm <sup>2</sup>
4	$\Delta a$	$-l_s \theta$	$0.8 \times 10^{-6}$ K <sup>-1</sup>	4 nm	137 nm <sup>2</sup>
5	$\Delta \theta$	$-l_s a_a$	0.05 K	6 nm	173 nm <sup>2</sup>
6	$i_s$	1	15 nm	15 nm	398 nm <sup>2</sup>

**Table 2.** Calibration history of a reference sphere’s diameter.

Date	17.05.2004	26.02.2007	16.06.2009	14.06.2011	13.06.2013
$d$ in mm	14.917 90	14.917 90	14.918 00	14.917 94	14.917 99
$u$ in $\mu\text{m}$	0.11	0.11	0.11	0.10	0.10
Date	23.06.2015	28.07.2017	26.07.2019	21.07.2021	
$d$ in mm	14.917 90	14.917 96	14.917 88	14.917 99	
$u$ in $\mu\text{m}$	0.10	0.10	0.10	0.10	



**Fig. 1.** Calibration history of a reference sphere’s diameter. The length of the error bars equals twice the expanded uncertainty  $U$  with the coverage factor  $k = 2$ .

calibrations should be taken into account. Nevertheless, there might be some systematic time-dependent drift, so it is reasonable to fit a straight line, i.e., a linear polynomial dependent upon the time  $t$ , given by

$$y(t) = a_1 t + a_0, \tag{3}$$

with  $a_1$  and  $a_0$  denoting the polynomial coefficients, to the calibrations instead of just taking the (weighted) mean value. Since the uncertainty of the past calibrations is variable (cf. Tab. 2), it must be taken into account when calculating the polynomial coefficients. Thus, to a calibration with a lower uncertainty a greater weight should be attached than to a calibration with a higher uncertainty when calculating the polynomial coefficients. Furthermore, to obtain the uncertainty associated with the best estimate of the current value of a measurand  $y$  calculated by equation (3), it is mandatory to have the uncertainties of the polynomial coefficients and their covariance as well.

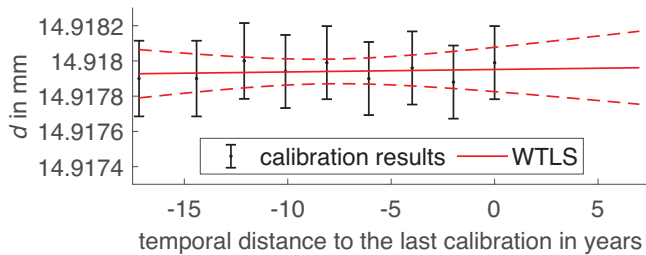
Fitting a linear polynomial to a number of data points following the least squares method minimizing the deviations in the horizontal, the vertical or both directions is well known [32,33]. In [34] also the uncertainties of the polynomial coefficients are calculated. Nevertheless, the individual uncertainties of the input data points are not taken into account for the calculation of these coefficients. In [35] a so-called weighted total least squares algorithm (WTLS), which calculates the polynomial coefficients taking into account the uncertainties of the input data, has been presented. Apart from the polynomial coefficients, it provides their uncertainties and covariance  $\text{cov}(a_1, a_2)$  as well. The implementation of the algorithm in MATLAB has been published by the authors at the MATLAB central file exchange<sup>1</sup>.

$y$  is calculated by equation (3). According to [2], the standard uncertainty of  $y$  is obtained by

$$u(y) = \sqrt{t^2 [u(a_1)]^2 + [u(a_0)]^2 + 2t \cdot \text{cov}(a_1, a_0)}, \tag{4}$$

where  $u(a_1)$  and  $u(a_0)$  denote the standard uncertainties of the polynomial coefficients and  $\text{cov}(a_1, a_0)$  denotes the covariance of  $a_1$  and  $a_0$ . In Figure 2, the calibration history of the reference sphere’s diameter along with the results of the WTLS are shown. In this figure, the abscissa is the temporal distance to the date of the last calibration. The uncertainty of these temporal values is negligible and thus was set to zero. Remarkably, the uncertainty of  $d$  calculated by the WTLS is lower than the uncertainty of a single calibration, at least for the time range within the

<sup>1</sup> <https://de.mathworks.com/matlabcentral/fileexchange/17466-weighted-total-least-squares-straight-line-fit>.



**Fig. 2.** Calibration history of the reference sphere's diameter (cf. Fig. 2) and the value of  $d$  calculated by the WTLS (solid line). The dashed red lines indicate the expanded uncertainty  $U$  of this value ( $k = 2$ ).

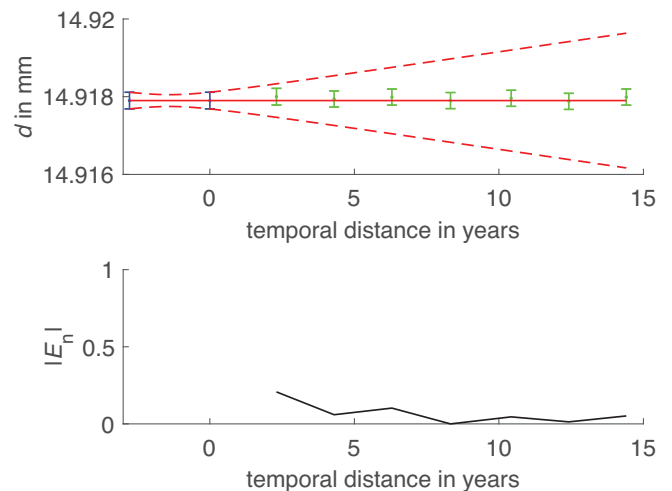
calibrations. Nevertheless, we may indeed state an estimation about the value of a measurand with more confidence and thus with lower uncertainty if several measurements and not just one single measurement are taken into account. The uncertainty of the value calculated by the WTLS is minimal for the mean of the calibration dates. However, the main intention is to apply the WTLS to predict a current or future value of a measurand along with its uncertainty, considering the calibration history of the particular measurand. Thus, the relevant values of WTLS are the ones for a positive temporal distance.

### 3.3 Validation

There might be some objections about the validness of these predicted values and their uncertainties. One might argue that it is not appropriate to treat a calibration from years ago the same way as a recent calibration. Instead, such old calibrations should be weighted less or disregarded completely. However, on the other hand the great time range of the calibrations yields a greater likelihood that also systematic influences are randomized and thus accessible to statistical analyses like the WTLS. Furthermore, if there is no reason to assume a fundamental change of the measurand, also past calibrations should keep their validness, because a simple drift of the value of a measurand is reflected by the WTLS as  $a_1 \neq 0$ . However, probably the most serious issue about the application of the WTLS to a calibration history is the fact that there is no knowledge about a causal relation between time and the measurand's value. Nevertheless, we will postpone the criticism from a theoretical standpoint to Section 4 and instead will evaluate the appropriateness of the application of the WTLS to a calibration history to predict a measurand's value and its uncertainty from an empirical standpoint by the assessment of the metrological compatibility of the results in this subsection. The method of validation will be described in the subsequent first part of this subsection and will be applied to a number of different standards in the second part of this subsection.

#### 3.3.1 Method

Metrological compatibility of measurement results is defined within the VIM ([3], 2.47). A frequently applied metric to assess the metrological compatibility of



**Fig. 3.** Upper panel: WTLS (shown in red) applied to the earliest two calibrations. The subsequent calibrations (shown in green) are used as reference to check the metrological compatibility of  $d$  and its uncertainty predicted by the WTLS. Lower panel:  $E_n$ -values ( $k = 2$ ) obtained by comparing the reference calibrations with the predicted results.

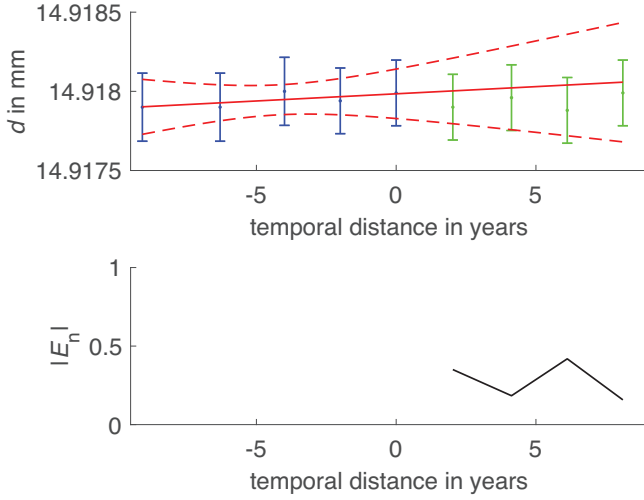
two measurement results is the normalized error [36]

$$E_n = \frac{1}{k} \frac{y_{test} - y_{ref}}{\sqrt{[u(y_{test})]^2 + [u(y_{ref})]^2}}, \quad (5)$$

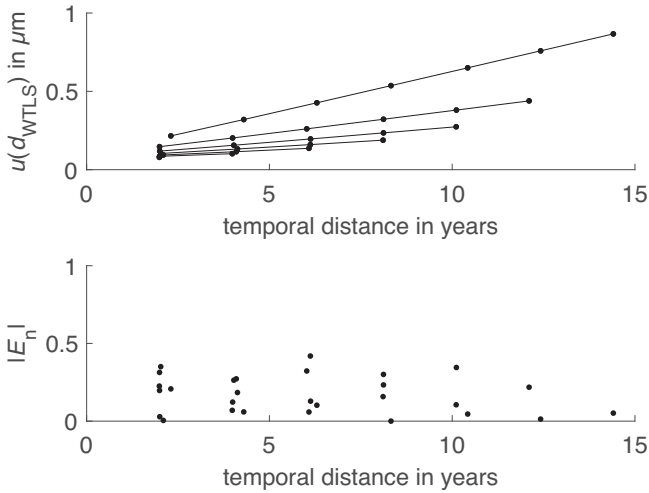
with  $k$  denoting the coverage factor, which is 2 within this paper,  $y_{ref}$  denoting the reference value and  $y_{test}$  the further value to be compared with the reference value. Metrological compatibility of the two measurement results is claimed for  $|E_n| \leq 1$ .

In the upper panel of Figure 3, the WTLS was applied to the earliest two calibrations, only. The seven subsequent calibrations are used as reference to check the metrological compatibility and thus the validity of the predicted results. All reference calibrations are compatible to the predicted results, because  $|E_n| \leq 1$  (see lower panel of Fig. 3). This is not surprising, since due to the limited number of calibrations considered by the WTLS the prediction's uncertainty is quickly increasing. However, also in Figure 4, where the WTLS was applied to the earliest five calibrations and thus the prediction's uncertainty is lower than the uncertainty of the first subsequent calibration, the prediction results and the reference calibrations exhibit metrological compatibility (upper panel of Fig. 4).

For a calibration history consisting of  $N$  calibrations, the WTLS might be applied to the first 2, 3, ...,  $N-1$  calibrations and the predicted results afterwards can be compared to  $N-2$ , ..., 1 reference calibrations. Thus, for the  $N = 9$  past calibrations of the sphere's diameter, in sum 28  $E_n$ -values can be obtained. In Figure 5, the standard uncertainty of  $d$  calculated by the WTLS and the corresponding  $E_n$ -values are shown over the temporal distances to the last calibration considered by the WTLS. As expected,  $u$  increases with time due to the uncertainty of



**Fig. 4.** Upper panel: WTLS (shown in red) applied to the earliest five calibrations. The subsequent calibrations (shown in green) are used as reference to check the metrological compatibility of  $d$  and its uncertainty predicted by the WTLS. Lower panel:  $E_n$ -values ( $k = 2$ ) obtained by comparing the reference calibrations with the predicted results.



**Fig. 5.** Standard uncertainty of  $d$  calculated by the WTLS (upper panel) and the corresponding  $E_n$ -values (lower panel) over the temporal distance to the last calibration considered by the WTLS.

$a_1$ , cf. equation (4). Since all absolute  $E_n$ -values are below 1 (the maximum value is 0.43), 100 % of the 28 results of the WTLS are compatible to the subsequent calibrations.

### 3.3.2 Examination of several standards

In the following, this method is applied to several standards: Two spheres with a nominal diameter of 15 mm (including the one of the previous subsection), two spheres with a nominal diameter of 30 mm, a step gauge with 52 measuring surfaces, a hole plate with 60 holes and a ball plate with 25 balls. All calibrations have been

**Table 3.** Overview of the standards.

Standard	Number of measurands	Number of calibrations
$d = 15$ mm-sphere 1	4	9 within 17 yr
$d = 15$ mm-sphere 2	4	7 within 12 yr
$d = 30$ mm-sphere 1	4	8 within 19 yr
$d = 30$ mm-sphere 2	4	7 within 18 yr
Step gauge	51	6 within 16 yr
Hole plate	1770	9 within 19 yr
Ball plate	300	4 within 6 yr

conducted either by the Physikalisch-Technische Bundesanstalt (PTB), which is the German national metrology institute, or calibration laboratories accredited by the DAkkS or formerly by the DKD. Each standard comprises more than one measurand. In the case of the spheres, apart from the diameter also the form deviation has been calibrated on three different orthogonal cross sections. For the step gauge, the distances of each measuring surface to the measuring surface number 0 (so in sum 51 distances) have been calibrated. For the hole plate and the ball plate, the distances of the centres of two arbitrarily chosen holes or balls have been calibrated, leading to 1770 (hole plate) and 300 (ball plate) measurands. An overview of the standards is given in Table 3.

In Table 4, the results of the validation of the WTLS by the calibration history of these standards are listed. The number of the results tested,  $N_{res}$ , is given by

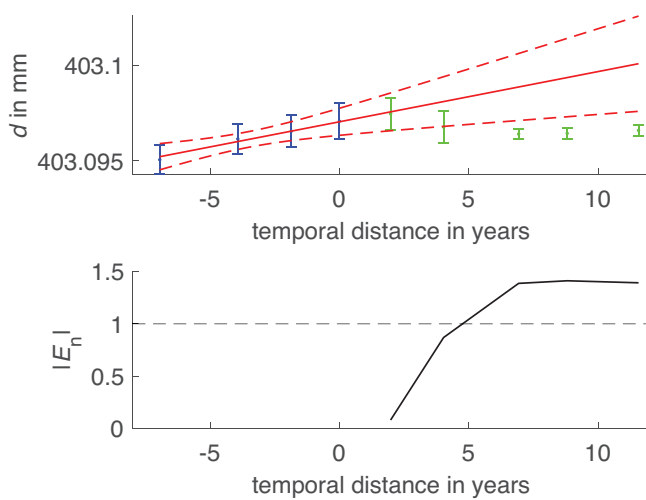
$$N_{res.} = N_{meas.} \cdot \sum_{i=1}^{N_{cali.}-2} i, \quad (6)$$

where  $N_{meas}$  denotes the number of measurands and  $N_{cali}$  denotes the number of calibrations of the particular standard. As can be seen, the validation is successful, meaning that the share of results with  $|E_n| > 1$  is below 5%, for six of the seven standards. However, in case of the hole plate, over 10% of the absolute values of  $E_n$  are above 1. This necessitates a more detailed investigation.

We take the calibration history of the measurand of the hole plate that exhibits the largest share of results with  $|E_n| > 1$ , which is about 46%, as an example. The calibration history of this measurand is shown in Figure 6. As an example, WTLS applied to the first four calibrations and the resulting  $E_n$ -values are shown as well. As can be seen, there is a linear drift of the measurand's value from the first until the fifth calibration. This drift starts to break with the sixth calibration and is abandoned completely for calibrations 7 to 9. Noteworthy, the calibrations 7 to 9 have been conducted by a different laboratory with a much smaller uncertainty. Unfortunately, we are not able to determine the reason for the break of the linear drift of the measurand's value. It might be a fundamental change of the standard which would cause the older calibrations to become invalid. On the other hand, it might also be related to the change of the calibration laboratory. However,

**Table 4.** Results of the validation of the WTLS by the calibration history of the standards.

Standard	Number of results tested	Number of results with $ E_n  > 1$	Share
$d = 15$ mm-sphere 1	112	0	00.0%
$d = 15$ mm-sphere 2	60	0	00.0%
$d = 30$ mm-sphere 1	84	0	00.0%
$d = 30$ mm-sphere 2	60	0	00.0%
Step gauge	510	1	00.2%
Hole plate	49 560	5 077	10.2%
Ball plate	900	16	01.8%

**Fig. 6.** Upper panel: Calibration history and WTLS applied to the first four calibrations of the worst measurand of the hole plate. Lower panel:  $E_n$ -values ( $k = 2$ ) obtained by comparing the reference calibrations with the predicted results.

it does lead to results predicted by the WTLS that do not exhibit metrological compatibility to the subsequent calibrations. This issue will be discussed in the following section.

## 4 Discussion

As has been stated in Section 3.3, there might be objections to treat calibrations from years ago the same way as a recent calibration. On the other hand, the great time range of the calibrations yields a greater likelihood that also systematic influences of the calibrations are randomized and thus accessible to statistical analyses like the WTLS and thus might improve the validness of its results. However, this is only true if there is no correlation between time and possible systematic influences. We might reasonably argue that it is likely that at least some systematic influences are related to the calibration laboratory. Thus, if one wants to apply the WTLS, the calibration laboratory should be chosen randomly for each calibration to randomize the systematic influences.

Care should be taken when applying the WTLS to a calibration history where the calibration laboratory has not been changed at all or where there is a correlation between time and the chosen calibration laboratory. Unfortunately, the calibration histories of all the standards analysed within this paper exhibit such a strong correlation between time and calibration laboratory. We might thus indeed question the appropriateness of the application of the WTLS from a theoretical point of view. And also the empirical findings point to a certain inappropriateness of the application of the WTLS to these particular calibration histories. To validate the appropriateness of the application of the WTLS to calibration histories that do not exhibit a correlation between time and calibration laboratory, it is probably necessary to create such calibration histories by purpose by deliberately choosing calibration laboratories randomly. Because it is common practice to rarely change the calibration laboratory. Noteworthy, the demand to choose the calibration laboratory randomly is also quite the contrary to the approach described in [26], which calls for homogeneous past calibration data. Nevertheless, take ISO 15530-3 [37] as an example, it is common knowledge that systematic influences should be randomized to yield their accessibility to subsequent statistical analyses.

## 5 Summary and outlook

This article started with a thorough review about the determination of recalibration dates or calibration intervals. Although there are some guidelines of internationally recognized organizations like the ILAC, the OIML or the DAkkS on this issue, the recommendations remain vague. As stated by a guideline published jointly by the VDI, the VDE, the DGQ and the DKD, there “are no generally binding specifications for the timing of a recalibration” ([20], 5.5). However, most of the literature recommends to take the calibration history into account when determining a calibration interval or a recalibration date. But only a few references give a clear guidance on how to process a calibration history to yield the next calibration date. The proposed procedures might be criticised from a theoretical standpoint, because either the uncertainty of the past calibrations is neglected or the number of considered past calibrations is limited arbitrarily. Nevertheless, the fundamental principle of those procedures, namely that the



uncertainty about the value of a measurand grows with time since the last calibration and it is necessary to predict this value along with its uncertainty at the time of use taking the calibration history into account, is adopted by the approach of an uncertainty-based determination of recalibration dates shown within this paper.

This approach is described in the third part of the paper. By the application of a so-called weighted total least squares algorithm (WTLS) to the calibration history it enables the prediction of the current value of a measurand along with its uncertainty, taking into account the past calibration values and their uncertainties. Based on the predetermined target measurement uncertainty of the measurand, it is possible to recognize the need for and thus to date recalibrations. To validate the applicability of the WTLS to a calibration history in order to predict a measurand's future value along with its uncertainty, the calibration history of the measurands of several standards has been investigated. There are cases where the results predicted by the WTLS did not exhibit metrological compatibility to subsequent calibrations. This issue has been discussed in Section 4. It might be related to the correlation between time and the chosen calibration laboratory. Since at least some systematic influences are highly likely related to the calibration laboratory, a correlation between time and these systematic influences might lead to biased results predicted by the WTLS. Unfortunately, such a correlation existed for all calibration histories examined within this paper, which might explain the lack of metrological compatibility.

In future works, the applicability of the WTLS should be investigated on calibration histories with randomly chosen calibration laboratories. As it is common practice to rarely change a calibration laboratory, it is probably necessary to create such calibration histories by purpose by deliberately choosing calibration laboratories randomly. It might also be investigated to what extent other approaches, like calculating a weighted mean value of all past calibration results, maybe including a coefficient to weight older calibrations less, give more accurate results. Because the currently most often applied procedure, which only considers the result of the last calibration and its uncertainty, in the presence of a calibration history seems to be a huge waste of information.

## Conflicts of interest

The authors declare no conflict of interest.

## Author contributions

Conceptualization, JS; methodology, JS; software, JS; validation, JS; formal analysis, JS; investigation, JS; resources, TH; data curation, JS; writing-original draft preparation, JS; writing-review and editing, TH; visualization, JS; supervision, TH; project administration, TH; funding acquisition, TH. All authors have read and agreed to the published version of the manuscript.

The authors would like to thank the students Jingjie Yang and Avian Kain for their support.

## References

1. R. Leach, M. Ferrucci, H. Haitjema, Dimensional metrology, in: *CIRP Encyclopedia of Production Engineering*, Springer, 2020, pp. S.1–S.11
2. Joint Committee for Guides in Metrology, JCGM 100: 2008. *Evaluation of measurement data – Guide to the expression of uncertainty in measurement*, 2008
3. DIN Deutsches Institut für Normung e. V. (Hrsg.), Internationales Wörterbuch der Metrologie: *Grundlegende und allgemeine Begriffe und zugeordnete Benennungen (VIM) – Deutsch-englische Fassung ISO/IEC-Leitfaden 99:2007*. 3. Beuth Verlag, 2010. - ISBN 978-3-410- 20070-3
4. H. Haitjema, Measurement uncertainty, in: L. Laperrière (Hrsg.), G. Reinhart (Hrsg.), *CIRP Encyclopedia of Production Engineering*, Springer, 2014, pp. S.852–S.857
5. *DAkkS 71 SD 0 006, Rückführung von Mess- und Prüfmitteln auf nationale Normale*, 2010. – 1. Neuauflage
6. F. Härtig, K. Wendt, Messunsicherheit und Rückverfolgbarkeit von Messwerten, in: A. Weckenmann (Hrsg.), *Koordinatenmesstechnik – Flexible Strategien für funktions- und fertigungsgerechtes Prüfen*. 2. Carl Hanser Verlag, 2012. - ISBN 978-3-446-40739-8, Kapitel 9, S.359–385
7. D. Imkamp, J. Berthold, M. Heizmann, K. Kniel, E. Manske, M. Peterrek, R. Schmitt, J. Seidler, K.-D. Sommer, Challenges and trends in manufacturing measurement technology – the “Industrie 4.0” concept, *J. Sens. Sens. Syst.* **5**, S.325–S.335 (2016)
8. C.P. Keferstein, M. Marxer, C. Bach, *Fertigungsmesstechnik – Alles zu Messunsicherheit, konventioneller Messtechnik und Multisensorik*. 9. Springer Vieweg, 2018. - ISBN 978-3-658- 17755-3
9. S. Carmignato, L. De Chiffre, H. Bosse, R.K. Leach, A. Balsamo, W.T. Estler, Dimensional artefacts to achieve metrological traceability in advanced manufacturing, *CIRP Ann.* **69**, S.693–S.716 (2020)
10. DIN EN ISO 9001:2015 *Quality management systems – Requirements (ISO 9001:2015)*
11. DIN EN ISO/IEC 17025:2017. *General requirements for the competence of testing and calibration laboratories (ISO 17025:2017)*
12. C.P. Keferstein, M. Marxer, R. Götti, R. Thalmann, T. Jordi, M. Andräs, J. Becker, Universal high precision reference sphere for multisensor coordinate measuring machines, *CIRP Ann.* **61**, S.487–S.490 (2012)
13. J.V. Jones, *Integrated logistics support handbook* (McGraw-Hill Companies, 2006), 3, ISBN 0071471685
14. T.S. Kuhn, *The structure of scientific revolutions* (The University of Chicago Press, 2012), 4
15. M. Salemink, F. Mager, Das richtige Kalibrierintervall – Ermittlung und Festlegung, *Pharm. Ind.* **76**, S.1931–S.1935 (2014)
16. *DAkkS 71 SD 4 027, Leitlinien und Beispiele für Überwachungsfristen von Prüfeinrichtungen für Laboratorien in den Bereichen Gesundheitlicher Verbraucherschutz, Agrarsektor, Chemie und Umwelt sowie Veterinärmedizin und Arzneimittel*. Oktober 2017, – Revision 1.2
17. *DAkkS 71 SD 3 026, Leitlinien und Beispiele für Kalibrier- und Überwachungsfristen von Einrichtungen für Laboratorien in der Kriminaltechnik und Forensik*. Juni 2018. – Revision 1.2

18. ILAC- G24: 2007/ OIML D 10:2007 – *Guidelines for the determination of calibration intervals of measuring instruments*. 2007
19. DAkkS 71 SD 5 004, *Regel zur Akkreditierung von Prüflaboratorien nach DIN EN ISO/IEC 17025:2005 für den Bereich "Koordinatenmesstechnik"*. Juli 2017. – Revision 1.3
20. *VDI/VDE/DGQ/DKD 2618 -1*. 1, Inspection of measuring and test equipment; Instructions to inspect measuring and test equipment for geometrical quantities; *Basic principles*. Berlin: Beuth Verlag, 2021
21. DIN EN ISO 1001 2: 2003, Measurement management systems – Requirements for measurement processes and measuring equipment (ISO 10012:2003)
22. DIN 3293 7: 2018, Mess- und Prüfmittelüberwachung – Planen, Verwalten und Einsetzen von Mess- und Prüfmitteln
23. J. Levitt, *The handbook of maintenance management* (Industrial Press, 2009), 2, ISBN 9780831133894
24. C. Grieser, D. Imkamp, "Onboard Diagnostics" – Ein innovatives System zur Messmaschinenüberwachung, in: VDI-Fachtagung "Sensoren und Messsysteme", VDI-Berichte 1829, 2004, pp. S.785–S.790
25. A. Lepek, Software for the prediction of measurement standards, in: NCSL Int. Workshop & Symposium, 2001
26. D.W. Wyatt, H.T. Castrup, Managing calibration intervals, in: NCSL Annual Workshop & Symposium, 1991
27. C. De Capua, S. De Falco, A. Liccardo, R. Morello, A virtual instrument for estimation of optimal calibration intervals by a decision reliability approach, in: IEEE International Conference on Virtual Environments, Human-Computer Interfaces, and Measurement Systems, 2005, pp. S.16–S.20
28. B. Pesch, Festlegung der Kalibrier- und Nutzungsintervalle von Messmitteln, in: VDI-Fachtagung "Prüfprozesse in der industriellen Praxis", VDI-Berichte 2319, 2017, pp. S.85–S.97
29. D. Vaissiere, Method of determining a calibration time interval for a calibration of a measurement device (European Patent Specification 2 602 680 B1, 2014)
30. C. Croarkin, An extended error model for comparison calibration, *Metrologia* **26**, S.107–S.113 (1989)
31. G. Berndt, E. Hultsch, H. Weinhold, Funktionstoleranz und Messunsicherheit, in: Wissenschaftliche Zeitschrift der Technischen Universität Dresden, 1968, 17, Nr. 2, pp. S.465–S.471
32. P. Glaister, Least squares revisited, *Math. Gaz.* **85**, S.104–S.107 (2001)
33. W. Kolaczia, Das Problem der linearen Ausgleichung im R<sup>2</sup>, *tm – Tech. Mess.* **73**, S.629–S.633 (2006)
34. M. Matus, Koeffizienten und Ausgleichsrechnung: Die Messunsicherheit nach GUM – Teil 1: Ausgleichsgeraden, *tm – Tech. Mess.* **72**, S.584–S.591 (2005)
35. M. Krystek, M. Anton, A weighted total least-squares algorithm for fitting a straight line, *Meas. Sci. Technol.* **18**, S.3438–S.3442 (2007)
36. W. Wöger, Remarks on the  $E_n$ -criterion used in measurement comparisons, *PTB-Mitteilungen* **109**, S.24–S.27 (1999)
37. DIN EN ISO 1553 0-3: 2011, Geometrical product specification (GPS) – Coordinate measuring machines (CMM): Technique for determining the uncertainty of measurement – Part 3: Use of calibrated workpieces or measurement standards (ISO 15530-3:2011)

**Cite this article as:** Janik Schaude, Tino Hausotte, Uncertainty-based determination of recalibration dates, *Int. J. Metrol. Qual. Eng.* **15**, 2 (2024)