The evolution of calibration practices

How to optimize calibration intervals

Endress+Hauser
People for Process Automation

The evolution of calibration practices using data science

Why and when to calibrate?

Executive summary

Achieving the correct balance between too much calibration and too little has always posed a challenge for anyone reliant on critical measurements. This paper reviews the underlying need for calibrations and traditional methods applied to determine calibration intervals. It further evaluates the advantages and disadvantages of some better-known intervals optimization methods and argues that these methods still do not offer practical alternatives to existing practices.

It then proceeds to explain two complimentary methods developed by Endress+Hauser that are jointly able to cover a wide range of applications and are robust enough to produce reliable results. It argues that these methods have the potential to deliver the elusive cost / risk balance that other methods have failed to deliver. It further highlights the limitations and complexities inherent in applying a variable interval approach and provides recommendations for additional risk management mechanisms to be applied in high risk scenarios. It concludes with some final words about the future of interval optimization and "predictive reliability".

1. Introduction

Measuring devices play a vital role in industrial process engineering by delivering crucial process data from pipes and vessels that are otherwise inaccessible. Measured values provide insight to process conditions that directly impact on product quality and safety.

It should however be noted, that even the most accurate measuring devices have measurement errors, i.e. the difference between the measured value and a reference value. These errors differ depending on the type, function and condition of the measuring point and typically increase over time. Consequently, a general requirement exists to regularly quantify such errors to determine if measurements are reliable enough to fit the purpose they are used for.

Simplistically stated, a calibration is the process of comparing the output values of a measuring instrument with a known reference for a given measurand (see exact definition in the VIM[1]). The objective of a **calibration is to determine the measurement errors of a given device**, often called UUT (Unit Under Test). Once the measurement errors are known, a UUT could in turn be used as a reference to calibrate another instrument. Ultimately, this "unbroken chain" of calibrations leads back to national or international primary references thereby creating a guarantee of metrological traceability.

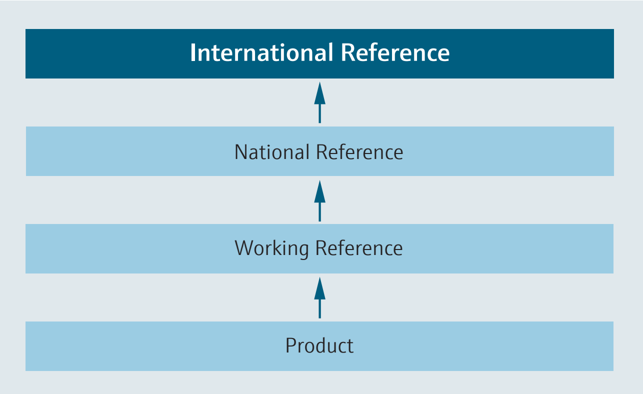

Figure 1: Traceable calibration chain between measurement devices

Calibration effectively creates a link between values acquired from measuring devices in the field with international measurement standards SI[2]. In practice this provides industrial companies a **common reference for**

measurements thereby allowing them to compare and transport manufacturing “recipes” to anywhere in the world. It also allows companies to **manage measuring devices and to detect measurement drifts** which may corrupt the “recipes” and therefore the related production.

Based on the above arguments it should be clear that calibrations are of vital importance for almost all industrial companies however this is not always the case. Many companies follow a reactive approach and **only calibrate when they are forced to** by local regulations or quality requirements. In even simple processes, a relatively **small measurement error can easily result in significant wasted resources** over time (energy, excessive input materials, wasted output), far in excess of calibration costs. When safety and environmental concerns are considered, the business case for regular calibration is overwhelming. Unfortunately, calibrations are often perceived as a costly administrative burden rather than a potential source of a value. It is no wonder that calibrations are often performed by “blindly following established rules” instead of adapting rules to reduce production risk and increase control. As stated in ISO 9001[3] calibration should “ensure valid and reliable results when monitoring or measuring is used to verify the conformity of products and services to requirements”.

This document is targeted at **companies and individuals that see the value in challenging the established rules for calibrations to improve product quality, reduce energy and raw material usage and to generally improve safety and control** in their plants.

2. Calibration intervals

Calibration interval selection is a typical illustration of established but outdated rules. An interval is defined as a pause or break between activities, in this case two subsequent calibrations. Calibration intervals should be chosen to reflect **an acceptable risk that the measurement error has not drifted outside of an acceptable range**. If we consider common practice, this is seldom the case.
Often **calibration intervals are set to one year** for the sake of convenience as it aligns well with annual planning cycles. This practice is so well entrenched that those responsible often claim this period is specified as a requirement in ISO9001. The standard however merely states that instruments should be: “...calibrated or verified, or both, at specified intervals, or prior to use, against measurement standards traceable to international or national measurement standards”. This means **intervals need to be determined and specified**. It does not state that intervals should all be the same across different instruments or that the period should be 1 year or multiples thereof. The only other relevant requirement from ISO 9001 for calibration intervals is to “ensure their continuing fitness for their purpose”.

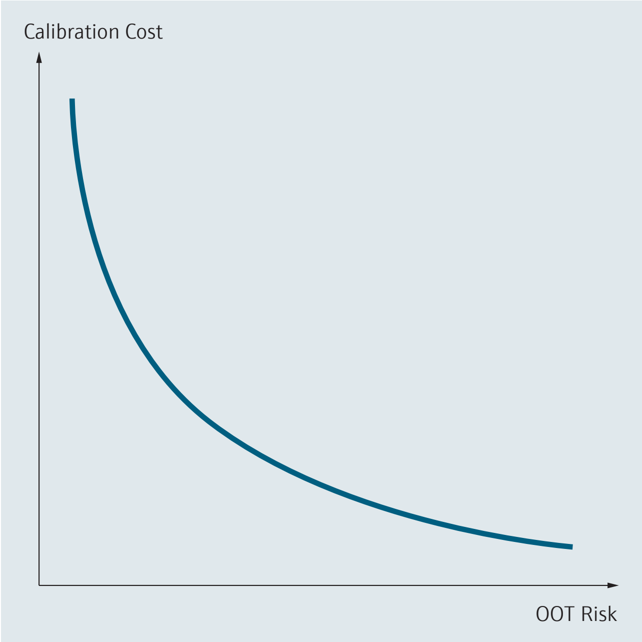

Figure 2: Inverse relationship between calibration costs and OOT risk

If we revert to the idea that an interval should reflect the risk of the measurement being “out of tolerance” (OOT) and that companies inherently wish to reduce operational costs, we can easily conclude that **the best interval is one which minimizes time OOT and cost at the same time**. Unfortunately, those two objectives are typically in opposition with each other. At one extreme we could eliminate calibrations to satisfy our cost objective with risk increasing over time. The opposite is true if we calibrated daily. This means **finding the appropriate interval is finding an optimal tradeoff between these two objectives** (Figure 2). The process of determining the optimal result is calibration interval optimization.

As stated, the optimal result is highly dependent on the definition of the acceptable risk of being OOT. This risk is typically reflected by the criticality defined for the measurement. As an example, for a non-critical device we would accept more OOT risk to help reduce calibration costs whereas for a highly critical device we would prefer a low OOT risk and would agree to invest more to manage this risk. **This leads to our first conclusion that interval optimization only makes sense when device criticality is properly defined**.

The definition of when a device is OOT is an equally important consideration. In this case the “Maximum Permissible Error” or MPE is an indicator when the measurement deviation exceeds operational requirements. In some companies/industries it is common practice to inflate MPE values to improve calibration conformity results. At the other extreme MPE´s are sometimes defined according to device’s theoretical capabilities under ideal condition due to this value being easily accessible. This results in MPE´s that are overly restrictive to be useful for interval optimization as they no longer reflect operational requirements. **This leads to the second conclusion that interval optimization only makes sense when a device´s MPE is properly defined**.

3. How to determine calibration intervals

Several methods exist to determine appropriate calibration intervals, most of them share a common underlying principle which is a “risk-based approach”. As previously highlighted, cost and risk need to be balanced however with a risk-based approach we typically start by defining the acceptable risk to derive the resultant cost and not vice-versa. A number of methods exist to assess the acceptable risk or device criticality, the most popular being Failure Mode, Effect and Criticality Analysis (FMECA)[4].

In relation to calibration intervals ILAC[5] and ISO 10012[6] provide no definitive guidance, however they do offer a better understanding of the relationship between risk and the chosen calibration interval.
A more detailed method is specified in the Good Automated Manufacturing Practice or GAMP5[7], however the proposed method suffers from a number of disadvantages:

- When calibrations fail, interval reduction is not considered
- Method is reactive and does not provide any predictions
- The maximum interval is arbitrarily fixed at two years
- Only conformity status (pass/fail) is considered and not measurement error size

The most holistic work done to date on interval determination is published by the NCSLI as their recommended practice RP1[8], here two groups of methods are described: **reactive and statistical methods**.

Reactive methods “are those in which calibration intervals are adjusted in response to data from previous calibrations without any attempt to model or “predict” measurement reliability behavior over time.”

The method presented in the GAMP5 is a good example of a reactive method. The drawbacks of these methods are well described by the NCLSI in the following statement: “Most reactive methods are, in general, less effective than statistical methods in terms of establishing intervals to meet reliability objectives. Additionally, reactive methods usually require long time (up to sixty years) to reach a steady state where the average in-tolerance rate attains a desired level.”

Statistical methods rely on complex mathematics to determine so called “maximum likelihood estimations”. While these methods offer some improvements over reactive methods the NCLSI acknowledge they are “expensive to design and implement.” They typically require a large installed base and a significant amount of data to be feasible and/or cost-effective.

The NCSLI-RP1 made significant advances towards a more scientific approach for interval selection. Unfortunately, due to the limitations mentioned above their methods have found limited practical applications since they released their first interval optimization paper in 1979! Their work has however created a solid basis for the development of more advanced methods that further maximize benefits and reduce drawbacks.

4. The Endress+Hauser approach

It has been established that **intervals should be optimized to find the best trade-off between cost and risk**. It should however be noted that the risk of an instrument operating OOT is not eliminated. Although the methods listed below have proven vastly superior to commonly used static interval approaches when balancing cost and risk they cannot reduce time OOT to zero in all cases.

Using established existing statistical methods as a base, **Endress+Hauser has developed several enhanced methods for calibration interval optimization**. Each method offers advantages and disadvantages and requires different inputs. The result is an optimization “toolbox” where the best “tool” can be applied depending on the situation, i.e. data availability, size of the installed base, device type, etc. Two of the most commonly applied methods are explained in more detail below:

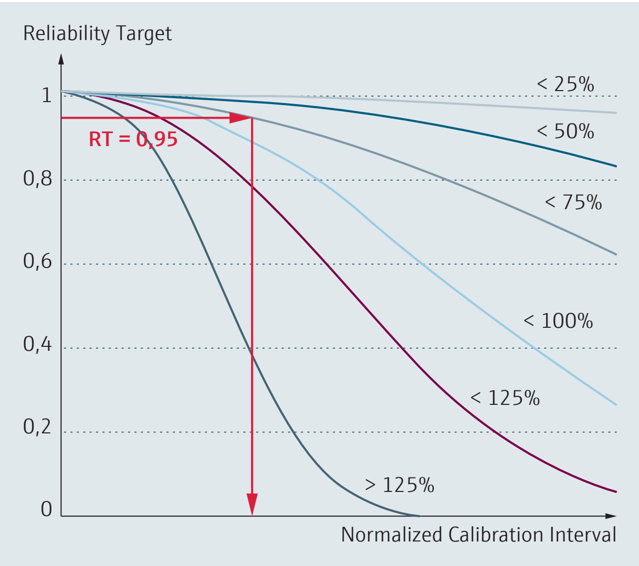

Figure 3: Deviation and reliability target mapping in a discretized space with the standardized reliability method

The first method is an enhancement of a statistical method defined by NCSLI-RP1. It has been specifically adapted to be dependent on less data so only the last calibration result is required to make a prediction (Endress+Hauser patent). The method relies various reliability models (e.g. Figure 3) which are selected according to the device type, measurement technology and extensive experience. This model selection defines how the normalized reliability space is segmented by the reliability curves (colored lines). The relevant reliability curve is then selected based on the measurement errors from the last calibration. Next, we apply a reliability target (determined by device criticality) to find the point where it intersects the reliability curve to determine the resulting normalized interval. Finally, the new calibration interval is calculated by transforming the normalized interval into days/months.

A significant advantage of this method is that it is relatively easy to apply. It requires little input data and is not dependent on a large quantity of devices. It utilizes readily available data such as criticality and calibration values. In contrast all methods previously mentioned only use pass/fail results and not the calibration values. It also offers the benefit that even with limited data, it's still able to offer a robust predictive approach to determine calibration intervals. This means calibrations do not have to fail before action is taken. Finally, it is applicable to any device type if the appropriate model (specific set of reliability curves) is applied.

The second method is based on Monté-Carlo simulations[9]. This method differs from the first as it factors in the two most recent calibration results along with measurement uncertainty. Since no measurement instrument is accurate enough to indicate the true measurement value it stands to reason that even reference meters have a degree of uncertainty which should be considered. This method determines the optimal calibration interval by simulating millions of calibrations and their potential drift based on actual results of the two last calibrations and their associated measurement uncertainty.

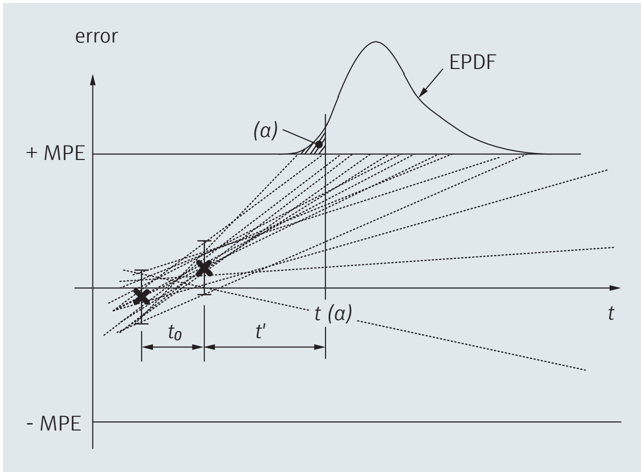

Figure 4: Simulation of drifts to determine the EDPF using the Monte Carlo method

The result of each simulation is projected forward in time (Figure 4) to determine when the measurement error would exceed the defined tolerance (MPE). When the results of all projections are considered they create an empirical probability density function (EPDF). Similar to the reliability target (first method) we use criticality to determine the required level of confidence which is in turn applied to the EPDF to determine the recommended time to perform the next calibration. The recommended interval reflects the projected time that the error will remain within tolerance with the selected level of confidence (Endress+Hauser patent).

This method offers the benefit that is **factors in additional information** (measurement uncertainty, past calibration values). It effectively **adapts to each instruments**

individual "metrological health" without relying on any large population statistics. Due to the inclusion of additional input data it **reacts faster to metrological device performance changes** than other methods. It does however suffer from a disadvantage that more information is required to apply it which may result in additional operational costs.

5. Validation of methods

Since calibrations intervals typically fall in the 6-36mo range rigorously testing a new method in practice would require both a significant amount of time (20+ years) and calibration results (+1Million). To overcome this challenge Endress+Hauser opted to fast track the validation process by employing a detailed theoretical model to simulate 22 Billion calibration results. These calibrations were simulated across 12 different drift scenarios both with and without interval optimization to determine the impact on risk, cost and process availability. After excluding results with extreme interval reductions (i.e. faulty devices) the overall results were extremely positive in comparison with the commonly applied fixed interval approach (Figure 5). While theoretical validation already provides compelling support for the methods employed, further validation in practice would eliminate any remaining doubts. Validating interval optimization methods under real conditions was initiated in 2018 and early indicators are starting to confirm theoretical results.

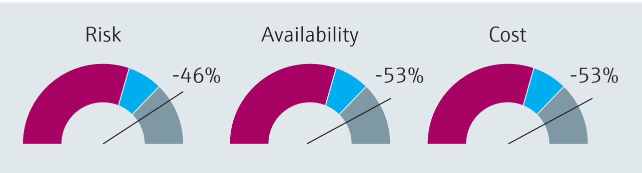

Figure 5: Calibration interval optimization benefits over a fixed interval approach for 22Billion simulated calibrations

6. Interval optimization considerations

While interval optimization inherently aims to **improve the balance between cost and risk**, it is important to note that variable intervals increase the complexity of calibration management and scheduling. **Optimization benefits could erode if processes are not adapted to effectively handle new intervals** generated after every calibration campaign.

This challenge is not method specific but rather inherent to the fact that fixed (e.g. yearly) calibrations are easier manage than ones that are constantly changing and distributed over time. For instance, distributed calibrations require additional effort to group them into practical campaigns that align with production shutdowns or maintenance plans. This implies that calculated intervals should not be applied "as is" and that intervals need to be mediated based on process availability, extreme conditions and practical considerations like geographic location. In practice **calculated intervals often differ from those applied**.

It should also be noted that interval optimization results are based on constraints defined by the input variables (MPE, Criticality, Deviations, etc.). This also means that if any of these variables are poorly defined (e.g. MPE purposefully increased to improve calibration conformity) this will negatively impact on the accuracy of calculated results and thus any calibration optimization plan.

7. Beyond interval optimization

Whether intervals are optimized or not, **observing measurement issues between two calibrations** remains a challenge unless the instrument fails.

As previously mentioned, interval optimization reduces the risk of running a process OOT between two calibrations however this risk is not reduced to zero. Even with the significant improvements in relation to a fixed interval strategy, risk management may still be **insufficient for the most critical devices**.

To address this need, Endress+Hauser has developed technologies such as TrustSens and Heartbeat that enable IIoT connected devices to self-diagnose their sensors metrological health in-between formal calibrations. In future many instruments will apply artificial intelligence to determine their own calibration needs. Endress+Hauser at the forefront of such developments having tested various "predictive reliability" prototypes already. To cover the full instrument spectrum companies will need to partner with vendors that have answers for existing devices and that a strong vision for the future. Significant benefits await those companies brave enough to embrace the change and leave outdated practices in the past where they belong.

By Dr. Dimitri Vaissière, Expert Data Scientist,
Endress+Hauser (Deutschland) AG+Co. KG

Gary Dreyer, Product Manager Optimization Services,
Endress+Hauser (Deutschland) AG+Co. KG

List of Abbreviations

OOT: Out of Tolerance
VIM: International Vocabulary of Metrology
UUT: Unit under Test
FMECA: Failure, Mode, Effect and Criticality Analysis
MPE: Maximum Permissible Error
IIoT: Industrial Internet of Things
EPDF: Empirical Probability Density Function

References

[1] ISO/IEC Guide 99, in International vocabulary of metrology – Basic and general concepts and associated terms (VIM) – 3rd edition. 2012: JCGM 200.
[2] BIPM, The International System of Units (SI), in Organization Intergouvernementale de la Convention du Mètre. 2006.
[3] ISO 9001, in Quality management systems – Requirements. 2015.
[4] Robert Borgovini, S.P., Michael Rossi, Failure Mode, Effects and Criticality Analysis (FMECA). 1993: Reliability Analysis Center. 134.
[5] ILAC-G24, in Guidelines for the determination of calibration intervals of measuring instruments. 2007.
[6] ISO 10012, in Measurement management systems – Requirements for measurement processes and measuring equipment. 2003.
[7] ISPE, GAMP Good Practice Guide. A risk-based approach to calibration management. 2nd ed. 2010.
[8] NCSLI. Establishment and adjustment of calibration intervals - Recommended practices RP-1. 2010.
[9] ISO/IEC Guide 98–3/S1, in Uncertainty of measurement – Part 3: guide to the expression of uncertainty in measurement (GUM:1995) – Propagation of distributions using a Monte Carlo method. 2008: JCGM 101.

www.addresses.endress.com

PU01334H/99/EN/01.19

Endress+Hauser

People for Process Automation