METRICA
METRology Institutional CAtalog

ISTITUTO NAZIONALE DI RICERCA METROLOGICA
Repository Istituzionale

A Bayesian statistical method for large-scale MEMS-based sensors calibration: a case study on 100 digital accelerometers

Publisher copyright

(Article begins on next page)

17 May 2024

A Bayesian statistical method for large-scale MEMS-based sensors calibration: a case study on 100 digital accelerometers

To cite this article: Andrea Prato *et al* 2024 *Metrologia* **61** 015005

View the article online for updates and enhancements.

A Bayesian statistical method for large-scale MEMS-based sensors calibration: a case study on 100 digital accelerometers

Andrea Prato[1],[*] ⓘ, Francesca Pennecchi[1] ⓘ, Gianfranco Genta[2] ⓘ and Alessandro Schiavi[1] ⓘ

[1] INRIM—Istituto Nazionale di Ricerca Metrologica, Applied Metrology and Engineering Division, Str. delle Cacce 91, 10135 Torino, Italy
[2] Department of Management and Production Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

E-mail: a.prato@inrim.it

CrossMark

Abstract

Low-cost sensors and in particular micro-electro-mechanical systems (MEMS) devices are widely used in many applications, including consumer electronics, healthcare, automotive, and industrial automation. Their large-scale production (typically in the order of millions per week in a single factory) would require the calibration of a huge number of devices that would be costly and time-consuming. A solution can be found in the use of statistical methods in order to (at least partially) substitute for the typical calibration procedures. In this work, we propose a Bayesian method to statistically calibrate large batches of sensors using probabilistic models and prior knowledge. The method involves experimentally calibrating only a small sample of sensors, then infer the number of reliable sensors in the entire batch and assign an appropriate uncertainty to all the sensors. Therefore, it can be considered as a statistical calibration of the batch. The Bayesian nature of this approach allows reducing the number of experimental calibrations by incorporating the prior knowledge coming from the previous calibration of a 'benchmark' batch, which is performed 'once and for all' and is representative of the whole production process. The application and validation of the method are performed through the calibration of 100 digital MEMS accelerometers. Validation results showed an acceptable agreement between experimental-based bootstrap and theoretical values, with relative differences within $\pm 7\%$.

Keywords: MEMS calibration, large-scale calibration, statistical calibration, Bayesian inference, low-cost sensors

---

[*] Author to whom any correspondence should be addressed.

1. Introduction

According to the International Vocabulary of Metrology (VIM) [1], the sensitivity of a measuring instrument/transducer is defined as the quotient of the change in an indication of the measuring system/transducer and the corresponding change in a value of a quantity being measured and is univocally attributed, based on specific standard calibration procedures. Any transducer is individually calibrated, against primary or secondary standards, and a single calibration certificate, guaranteeing its metrological traceability, is issued one-to-one. This process allows each specific transducer to be linked to the relevant SI units through a proper metrological traceability chain. Nevertheless, while this process is (and must remain) unavoidable to uniquely provide traceability to measuring instruments, it cannot be applied in the case of new generation sensors (e.g. based on micro-electromechanical systems (MEMS) technology), due to the large-scale production rate (a single manufacturer can produce millions of sensors per week [2]). This poses new challenges for metrology research. Nowadays, low-cost and low-power new-generation sensors are largely employed in many current daylife technological facilities, such as in smart systems and sensor networks for health, safety, automotive, buildings, energy, industry, environmental control, structural monitoring, as well as many more, in the framework of the evolving digitalization [3, 4].

To ensure the safety and functionality of operations managed by these systems, it is crucial that the data provided by the supporting sensors is both reliable and sufficiently accurate. Consequently, there is a need to establish comprehensive calibration methods on a large scale.

As suggested in the literature, this could be achieved by either in-line calibration systems or statistical methods, but none has been fully implemented in practice [5–7].

Ideally, in-line calibration systems should offer the advantage of calibrating all MEMS sensors during the production process, using traceable methods to be defined by the manufacturer. However, they come with certain drawbacks, including difficulty in ensuring standard calibration procedures and impartiality.

Conversely, statistical methods provide the advantage of drastically reducing the number of experimental calibrations, hence allowing that those are carried out by an accredited calibration laboratory, ensuring adherence to standard calibration procedures. Nevertheless, they also have drawbacks, such as not experimentally calibrating all MEMS sensors, which may lead to higher uncertainties and lower reliability. The possibility to exploit these methods is also indicated in the strategy document 2021–2031 of the Consultative Committee for Acoustics, Ultrasound and Vibration of the Bureau International des Poids et Mesures [8]. It states that «to maintain an acceptable reliability factor» of these sensors, «industry has moved from testing and calibrating every device towards statistical sampling to reduce manufacturing costs». Such approaches (aimed to provide «statistically acceptable levels of performance and reliability» of sensors from manufacturers [8]) involve a probabilistic determination of key metrological attributes, traditionally defined by experimental quantitative values as a result of a standard calibration. Therefore, these new approaches need a careful investigation of their feasibility and suitability, in the metrological perspective. In particular, this paper elaborates one of such approaches, discussing its potentialities but also its limitations, from both computational and practical point of view.

A Bayesian approach [9] is here explored. The method involves first the experimental calibration of a small subset of sensors from a larger batch and then, by means of a statistical calibration process, the estimation of the total number of reliable sensors in the entire batch and the assignment of an appropriate uncertainty to all batch sensors. The significant benefit of this statistical method is the ability to incorporate prior knowledge gained from the previous calibration of a 'benchmark' batch, which is experimentally performed 'once and for all'. The benchmark batch is chosen as representative of the entire production process, whose main characteristics are assumed to remain stable for all the time the statistical calibration of the produced batches is in place. In case of evidence or suspicion of significant changes in the production process, the calibration of a new benchmark batch is required.

After performing the experimental calibration of the benchmark batch, for all subsequent batches of that production process, the statistical method requires the experimental calibration of only a few sensors within each batch. In this way, the number of in-the-lab calibrations is reduced, enabling the statistical calibration of large batches of sensors at affordable expenses in terms of time and costs.

This process implies a merge of quantitative and probabilistic information, such as the calibration uncertainties, the precision of the provided indications, the repeatability and the reproducibility of results, but also predefined admissible tolerance limits, based on suitable hypotheses related to prior information and likelihood functions, to provide the metrological attributes on the basis of probability distributions.

As a case study, the proposed statistical procedure is applied to 100 nominally identical digital MEMS accelerometers and it is validated through a bootstrap technique.

2. The proposed statistical method

2.1. Ontology and taxonomy

Before delving into the methodology it is important to clarify some terminologies used, their relationships with the fundamental and general metrological concepts, and some necessary lexical expansions. Although, in general, terms and definitions used in this work are based on the VIM [1] (indicated in the following within single quotation marks), some exceptions must be pointed out in order to avoid misunderstandings. Indeed, often, terms and definitions currently used in engineering and electronic applications (indicated within double quotation marks) do not follow the VIM, and it is therefore necessary to address those definitions correctly.

Commonly speaking, the term 'sensor' is comprehensively attributed to a device producing output signals by sensing a

physical phenomenon, as a reaction to changes in that physical quantity. However, an acceleration sensor, for example, is neither strictly an accelerometer (i.e. the 'measuring instrument') nor the element directly affected by the phenomenon carrying the quantity to be measured (i.e. the 'sensor'). It is properly a 'measuring chain', i.e. the set of one or more measuring instruments and other devices constituting a single path of the signal from the sensor to the output element [1]. However, along this definition, it is preferable to consider the common term 'sensor' more properly as 'sensing system', with a weaker semantic degree than 'measuring chain'. Indeed, an acceleration sensor is a component of an electronic circuit, namely a chip embedding a vibration sensor with a transducer (generally based on MEMS technology), an amplifier, and an analogue-to-digital converter. Only for the sake of simplicity, hereinafter we use the term 'sensor', albeit intended as 'sensing system'.

In applied metrology, the 'sensitivity' can be attributed to a measuring chain, as a calibration result, in terms of the quotient of the change in an indication of the measuring system, and the corresponding change in the value of the reference standard quantity being measured, which is the stimulus/input to the measuring system. According to the IEEE 2700-2017 standard [10], the 'sensitivity' of a 'sensor' is opportunely indicated in terms of 'scale factor', as reported in the application datasheets provided by the manufacturers. However, the 'scale factor' is determined on the basis of 'adjustment' procedures, rather than 'calibration', and the 'adjustment' procedures, applied at the manufacturer level, are not traceable with respect to a reference standard, and the whole uncertainty budget is generally unknown or disregarded. Hence, in this work, when proper calibration methods against a reference measurement standard are implemented, we refer to the 'sensitivity' of a sensor, as properly defined.

Given the impossibility, in the current state of available technologies, of providing a sensitivity value for each individual sensor of very large batches, an approach based on Bayesian statistics to be applied for large-scale calibration and able to provide a sufficiently representative sensitivity value to a large set of (supposed) identical sensors is here proposed. In this context, it is therefore necessary to extend the concept of sensitivity, traditionally associated with each individual sensor on the basis of its experimental calibration. Through the proposed statistical approach, sensitivity is assessed by means of a statistical calibration performed by combining the sensitivities of a benchmark batch, representative of the entire production, and the experimental calibration results of a small portion of sensors of an unknown batch. Finally, an expression for the uncertainty to be associated with the sensors of the unknown batch, deduced from the same probabilistic assumptions on which the statistical model is based, is proposed.

2.2. Use of a benchmark batch

The first step of the proposed method consists in the usual experimental calibration of a benchmark batch of N sensors: each sensor is provided with a measurement of its sensitivity S_i

($i = 1, \ldots, N$) and the associated expanded uncertainty $U(S_i)$ at a confidence level of 95 %. We call it 'benchmark' not in the sense of the best-products batch, but in the sense of a batch faithfully representing the typical results of a production process, including possible defects.

To model the distribution of the benchmark batch sensitivities including their calibration uncertainties, a mixture distribution [11] can be applied. This distribution is derived from a set of N variables, each following a normal distribution that models the knowledge about the sensitivity of each sensor (i.e. having the mean and the standard deviation equal to the measured sensitivity value and the associated uncertainty, respectively). The mixture distribution is numerically obtained by generating data from the multivariate normal distribution of those N variables (allowing possible correlation among the several normal distributions). The parameters of the multivariate normal distribution are a vector of mean values equal to the N sensitivities of the benchmark batch and an $N \times N$ covariance matrix with squared standard uncertainties on its diagonal. A Monte Carlo simulation is employed [12, 13] that randomly generates 10^5 vectors made of N values each, from the multivariate normal distribution. The simulated values are combined into a $10^5 \times N$ matrix, whose columns represent the (marginal) probability density functions of individual sensors. By combining these density functions by equal weights, i.e. merging the simulated $10^5 \times N$ sensitivity values, the mixture distribution representing the potential sensitivity values of the entire benchmark batch is obtained. It is intended as a representation of the distribution of the sensitivity values typical of the production process from which the benchmark batch is taken.

The characterization of the benchmark batch in terms of a distribution allows defining an interval of 'good' sensitivity values according to some desired level of coverage probability. With such intent, by setting a bearable probability p of finding out-of-tolerance sensor sensitivities in the benchmark batch, the mixture distribution is used to find the lower and upper sensitivity limits, $S_{\text{low}}(p)$ and $S_{\text{up}}(p)$, as the limits of an interval encompassing the $(1 - p)$ fraction of acceptable sensors. The sensors whose sensitivity falls outside the $[S_{\text{low}}(p), S_{\text{up}}(p)]$ interval are considered as out-of-tolerance. In this work, the coverage interval is determined as a probabilistic symmetric interval, i.e. leaving out, both to its left and right side, a $p/2$ fraction of the values. The number of out-of-tolerance sensors in the benchmark batch is then equal to $C_{\text{bench}} = pN$.

For other future unknown batches of the same kind of sensors, only a subset of $n < N$ devices is required to be experimentally calibrated, hence reducing calibration time and cost efforts. Also for such batches, the requirement for acceptable sensors is related to the $[S_{\text{low}}(p), S_{\text{up}}(p)]$ interval. The sensitivities of the n calibrated devices S_j ($j = 1, \ldots, n$) are checked whether, together with their associated expanded uncertainty $U(S_j)$, are within or outside the tolerance limits, i.e. if the condition $S_j - U(S_j) > S_{\text{low}}(p) \ \Lambda \ S_j + U(S_j) < S_{\text{up}}(p)$ is satisfied or not. The number of the out-of-tolerance sensors among the calibrated n sensors is indicated by k.

2.3. Bayesian statistical calibration

The statistical calibration of the unknown batch of N sensors (actually, of the $N - n$ yet unknown ones in the batch) is based on the following Bayesian model [14]. Assuming that the expected number of out-of-tolerance sensors in the unknown batch is equal to C_{bench}, a binomial prior distribution $f_{\text{prior}}(C; N, p)$, with mean value equal to $C_{\text{bench}} = pN$, is used to model the actual number C of out-of-tolerance sensors in the unknown batch, as detailed in equation (1). This prior probability mass function (assuming that the experimental calibration and the manufacturing process do not change from batch to batch) models the state of knowledge on the number of out-of-tolerance items in a typical batch on N items from that production. The likelihood function $f_{\text{like}}(k; n, C, N)$, reported in equation (2), is defined as a hypergeometric distribution of the k out-of-tolerance sensors in the sample of size n, where N is the size of the unknown batch and C is the number of defective sensors in the batch. The hypergeometric distribution is a discrete probability distribution that describes the probability of k successes (in this case the number of defective sensors) in n draws without replacement. Multiplying the prior and the likelihood yields an un-normalized posterior $f_{\text{post,un}}(C; k, n, N, p)$, function of C out-of-tolerance sensors in the unknown batch. Normalizing this distribution, i.e. dividing it by the summation of $f_{\text{post,un}}(C; k, n, N, p)$ for C ranging from 0 to N, leads to the probability mass function $f_{\text{post,norm}}(C; k, n, N, p)$ reported in equation (3), which is the posterior distribution of the number C of out-of-tolerance sensors in the unknown batch, given that k out-of-tolerance devices are found in the small calibrated sub-batch. Detailed analytical calculations are reported in appendix,

$$f_{\text{prior}}(C; N, p) = \binom{N}{C} p^C (1-p)^{N-C} \qquad (1)$$

$$f_{\text{like}}(k; n, C, N) = \frac{\binom{C}{k}\binom{N-C}{n-k}}{\binom{N}{n}} \qquad (2)$$

$$f_{\text{post,norm}}(C; k, n, N, p)$$
$$= p^{C-k}(1-p)^{N-n-C+k}\binom{N-n}{C-k}. \qquad (3)$$

As an example, the probability mass function $f_{\text{post,norm}}(C; 0, 3, 100, 0.05)$ (i.e. at $k = 0$, $n = 3$, $N = 100$ and $p = 0.05$) as a function of C, is depicted in figure 1. It represents the probability that in the unknown batch of $N = 100$, C sensors are out-of-tolerance, knowing that $k = 0$ calibrated sensors from the subsample $n = 3$ are defective, when the out-of-tolerance sensors probability in the benchmark batch is $p = 0.05$.

2.4. Proposed criterion for the unknown batch performance assessment and uncertainty evaluation

As required by [8], in order to provide the batches with «statistically acceptable levels of performance and reliability»,
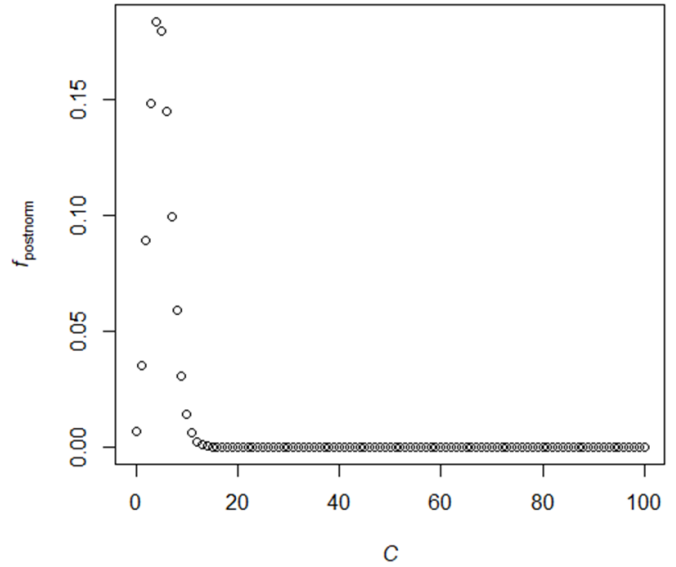
Figure 1. Probability mass function $f_{\text{post,norm}}(C; 0, 3, 100, 0.05)$ on the range of C values in [0, 100].

the posterior cumulative probability function can be used to define appropriate metrics for batch reliability. In this work, the focus is only on cases when $k = 0$ defective sensors are found among the n calibrated in the sub-batch. In the proposed approach, when $k > 0$ the unknown batch is deemed to be discarded and no reliability assessment is needed. Corresponding metrics could in principle be defined also for $k > 0$, but the attention is here on the more stringent requirement of no out-of-tolerance calibrated sensors.

For this purpose, it is here proposed to define the reliability of the unknown batch as the probability $P_{\text{reliab,unk}}$ to get a number of out-of-tolerance sensors in the whole batch not larger than the number of out-of-tolerance sensors C_{bench} in the benchmark batch. This probability is given by the posterior cumulative function:

$$P_{\text{reliab,unk}} = P_{\text{post,norm}}(C \leqslant C_{\text{bench}}; 0, n, N, p)$$

$$= \sum_{C=0}^{C_{\text{bench}}} f_{\text{post,norm}}(C; 0, n, N, p)$$

$$= \sum_{C=0}^{C_{\text{bench}}} p^C (1-p)^{N-n-C}\binom{N-n}{C}. \qquad (4)$$

From equation (4), in the limit case of $n = N$, i.e. when all sensors in the unknown batch are calibrated and none is out-of-tolerance, one has that $P_{\text{reliab,unk}} = P_{\text{post,norm}}(C \leqslant C_{\text{bench}}; 0, N, N, p) = 1$, which is the current requirement in the typical traceability chain when each and every device is experimentally calibrated.

This metric, defined for each unknown batch which is accepted, i.e. when $k = 0$, provides the fraction $P_{\text{reliab,unk}}$ of the future unknown batches that will have no more than C_{bench} out-of-tolerance sensors. In other terms, it is the probability that the sensitivity of at least $(N - n - C_{\text{bench}})$ uncalibrated

sensors of the unknown batch lies between $S_{low}(p)$ and $S_{up}(p)$. This is an additional information that characterizes the reliability of the batches that will be released. Based on this metric, a sensors producer can state that $P_{reliab,unk}$ % of the released batches will have a sufficiently large number of acceptable items, that is at least $N(1-p)$.

The n calibrated sensors get their sensitivity value and the associated uncertainty from the usual calibration process in-the-lab. Although the remaining $(N-n)$ sensors are not experimentally calibrated, it is nonetheless necessary to provide them with an estimate of their sensitivity, $S_{unk}(p)$, and an associated uncertainty, $u(S_{unk}(p))$. This can be done based on the information derived from the mixture distribution of the benchmark batch (always working under the hypothesis that the unknown batches are of the same kind as the benchmark one). In this sense, it is proposed to assign to the $(N-n)$ uncalibrated sensors of the unknown batch a statistically-averaged

sensitivity, equal to the mean value of the benchmark batch, e.g. when the mixture distribution is nearly symmetric, equal to

$$S_{unk}(p) = \frac{S_{up}(p) + S_{low}(p)}{2}. \tag{5}$$

The associated squared standard uncertainty $u^2(S_{unk})$ is evaluated based on the weighted mean of the variance associated with a sensitivity lying within $[S_{low}(p), S_{up}(p)]$ and that associated with the sensitivity outside that interval, the weights being the fractions of acceptable and out-of-tolerance sensors in the batch, respectively. Then, this weighted mean is evaluated for each possible value of C and multiplied by the posterior's probability to find exactly C out-of-tolerance sensors in the unknown batch (for C varying across all its possible values). The final expression for the uncertainty is given by

$$u(S_{unk}) = \sqrt{\sum_{C=0}^{N-n} f_{post,norm}(C;0,n,N,p) \left(\left(1 - \frac{C}{N-n}\right) \frac{(S_{up}(p) - S_{low}(p))^2}{12} + \frac{C}{N-n} \frac{(S_{max} - S_{min})^2}{8} \right)}. \tag{6}$$

When the unknown sensor sensitivity lies between the limits, the variance of a rectangular distribution on that interval is used. Conversely, when the sensitivity is out of the interval, the variance of a U-shaped distribution defined on a larger

interval $[S_{min}, S_{max}]$ is taken, where S_{min} and S_{max} are the minimum and the maximum value of the mixture distribution of the benchmark batch.

Substituting (3) in (6), with $k=0$, and after some algebra, standard uncertainty $u(S_{unk})$ becomes

$$u(S_{unk}) = \sqrt{(1-p) \frac{(S_{up}(p) - S_{low}(p))^2}{12} + p \frac{(S_{max} - S_{min})^2}{8}} = \sqrt{\frac{(S_{up}(p) - S_{low}(p))^2}{12} + p \left(\frac{(S_{max} - S_{min})^2}{8} - \frac{(S_{up}(p) - S_{low}(p))^2}{12} \right)}. \tag{7}$$

Equation (7) is independent of n and N, depending only on p and on the characteristics of the benchmark batch mixture distribution, which influences the value of lower and upper bounds and the minimum and maximum values. This uncertainty evaluation might result in a precautionary and conservative practice that, however, addresses the need to provide an uncertainty to the sensitivity values of the statistically calibrated sensors. Examples of uncertainty calculations are reported in section 3 with actual experimental data.

2.5. Discussion

To study the behaviour of the proposed metric, $P_{reliab,unk}$ values are depicted in figures 2–4 as function of the ratio n/N, for different N, and different probabilities p to find out-of-tolerance sensors in the benchmark batch. As expected, at increasing n/N, also $P_{reliab,unk}$ increases. When the ratio n/N tends to zero, $P_{reliab,unk}$ values tend to 50 %, i.e. pure randomness, since there is no evidence that the batch quality satisfies the requirement of a limited number of out-of-tolerance sensitivities, i.e.
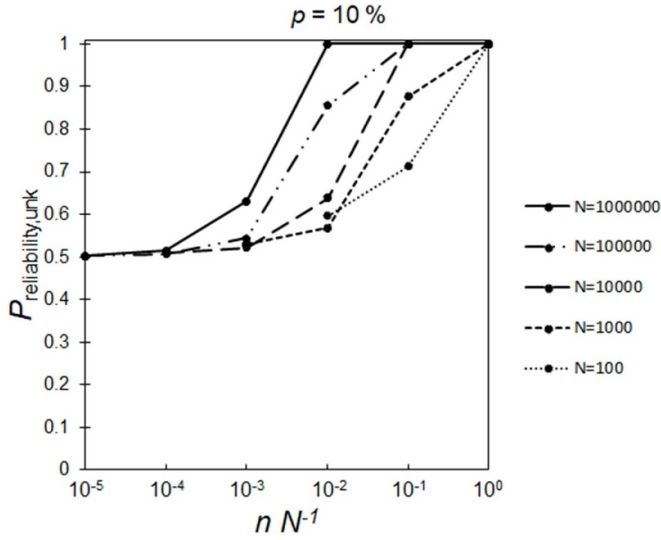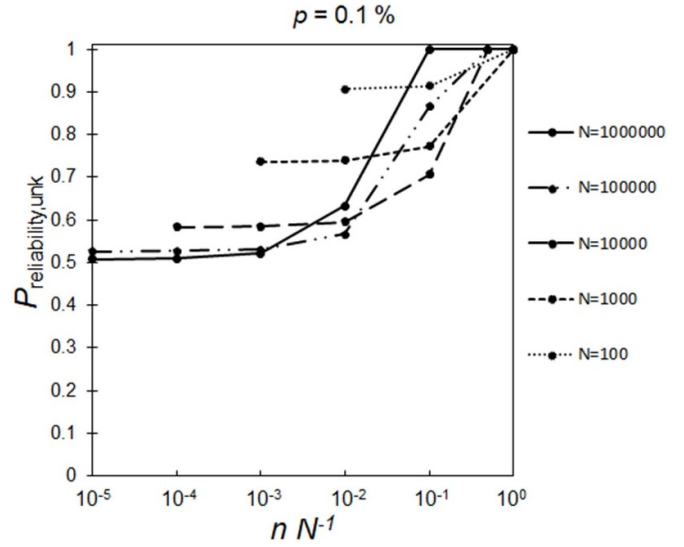
Figure 2. $P_{\text{reliab,unk}}$ as function of n/N, for different N values, with fixed probability $p = 0.1$ of out-of-tolerance sensors in the benchmark batch. Connecting lines are for representation purposes only.
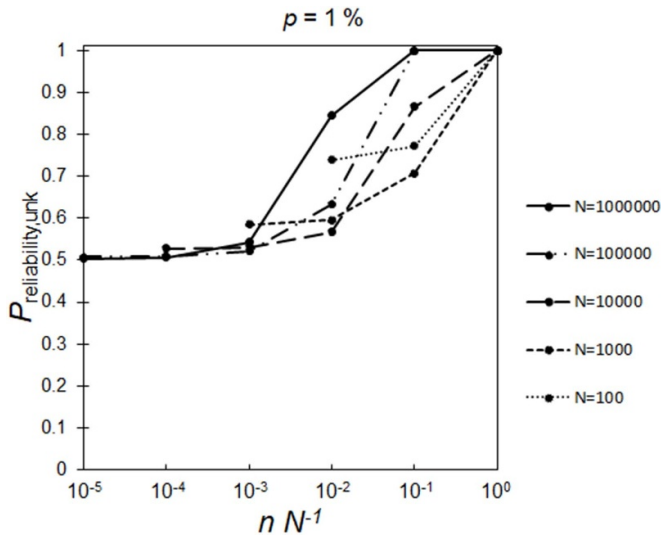
Figure 3. $P_{\text{reliab,unk}}$ as function of n/N, for different N values, with fixed probability $p = 0.01$ of out-of-tolerance sensors in the benchmark batch. Connecting lines are for representation purposes only.


Figure 4. $P_{\text{reliab,unk}}$ as function of n/N, for different N values, with fixed probability $p = 0.001$ of out-of-tolerance sensors in the benchmark batch. Connecting lines are for representation purposes only.

$C \leqslant C_{\text{bench}}$. Indeed, even if none out-of-tolerance sensor is found in the sample ($k = 0$), this too small sample becomes unrepresentative with respect to the whole batch and cannot convey further information with respect to the prior probability (which is a binomial distribution with median equal to C_{bench}, so that $P_{\text{prior}} (C \leqslant C_{\text{bench}}) = 0.5$). Note that when $k = n = 0$, equations (1) and (3) coincide, i.e. the posterior mass function degenerates to its prior.

Moreover, it is important to note that, for constant n/N ratios, $P_{\text{reliability,unk}}$ tends to rise as N increases, particularly with higher values of p. However, there are instances where it increases as N decreases, particularly with lower values of p. The general increase of $P_{\text{reliability,unk}}$ with larger N, more often when p is high, suggests that, in order to statistically calibrate

a whole population of items with higher reliability, it would be preferable to split the population into a few large batches rather than into many small batches, as the former ones result in a greater reliability for the same amount of effort (i.e. the same overall number of experimental calibrations). This, however, would imply the preliminary experimental calibration of a benchmark batch of a larger sample size N which can still demand a substantial experimental effort. By the way, the feasibility of the calibration of a large benchmark batch is highly dependent on several factors such as the sensor type, measurement technique, calibration conditions, and the calibration system. In order to further alleviate the workload, one might also consider the possibility of skipping the measurement of the benchmark batch and simply resorting to the producer's information. However, this would require complete trust in the manufacturer's ability to furnish the distribution of calibration values for the benchmark batch sensors. Although this could be a feasible option, it raises concerns regarding the required impartiality within the metrology chain, which may not be guaranteed.

In order to reduce the number of required calibrations, especially for the benchmark batch, it becomes important to find a balance by selecting batch sizes that are not excessively large. A good trade-off between high reliability values and experimental workload is to choose low values of p, so that batch size N can be decreased, while still choosing a very small size for the unknown batch. In fact, at decreasing p, it can be shown that lower N values entail higher reliability at decreasing n/N values (figure 4). This is an optimal solution since it guarantees, at the same time, a lower workload for calibration and acceptable levels of reliability. The downside of a small p, however, is that a wide acceptance interval $[S_{\text{low}}(p), S_{\text{up}}(p)]$ is allowed with the consequence that the end-user of those sensors should be aware of (and happy with) a large variability in the accepted sensors.
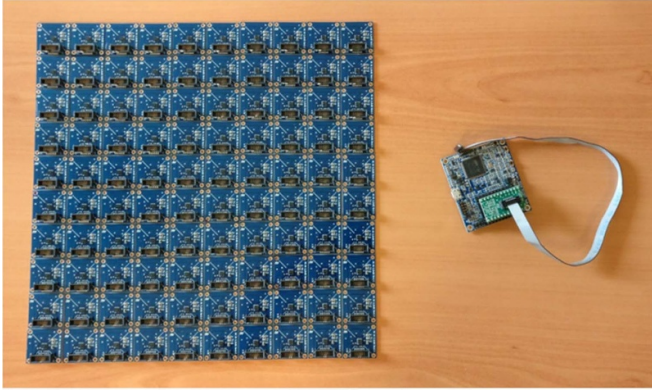
Figure 5. The 100 digital MEMS accelerometers (left) calibrated at INRiM and the external microcontroller (right).
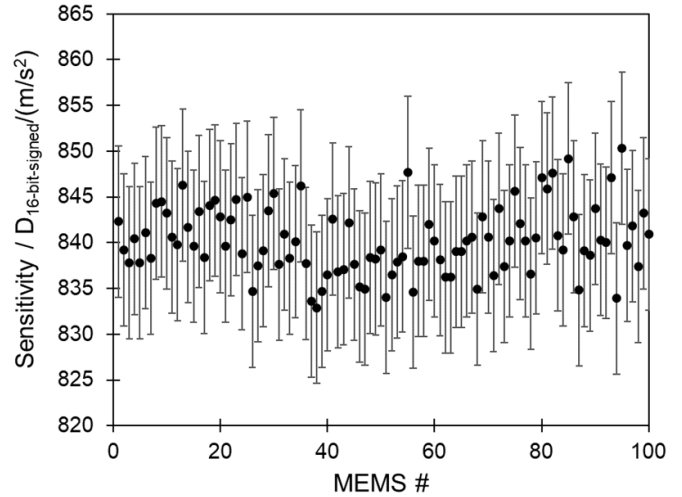

Figure 6. Sensitivities along *z*-axis of the 100 MEMS at 10 Hz with relevant calibration uncertainties (at 95 % confidence level).

3. Application and validation of the method

In order to assess the effectiveness of this Bayesian statistical approach and to validate it, 100 nominally identical digital MEMS accelerometers (STM, model LSM6DSR) connected to an external microcontroller (STM, model STEVAL MKIGIBV2) (figure 5) are calibrated at INRiM, one at a time according to prescribed procedures [15–19].

Calibration is performed using a single frequency of 10 Hz. A vibrating table (specifically, the PCB Precision Air Bearing Calibration Shaker) generates a reference acceleration with nearly constant amplitude (10 m s^{-2}) along the vertical axis for a duration of 10 s. This reference acceleration is detected by a single-axis reference transducer (PCB model 080A199/482A23), integrated within the shaker's stroke. Its output is collected using an acquisition board (NI 4431) in the PC. LabVIEW® software processes this data to provide the reference value in m s^{-2}, at a sampling rate of 50 kHz.

The digital MEMS accelerometer is fixed to the shaker along the vertical axis using an ultra-thin double-sided adhesive tape, typically employed in practical applications. The external microcontroller records the digital output of the MEMS sensor at a maximum sampling rate of 6.66 kHz and saves the data as binary files. The outputs of the MEMS are given in Decimal$_{16\text{-bit-signed}}$ (hereinafter abbreviated as D$_{16\text{-bit-signed}}$) where the digit unit is a signed 16-bit sequence converted into a decimal number. These files are subsequently processed with MATLAB® software. Firstly, a first-order Butterworth band-pass filter with a centre frequency matching the frequency of interest and a fractional bandwidth of 10 % is applied to the temporal digital signals. Secondly, the root mean square is computed to eliminate gravitational offsets and the influence of background vibrations. Results are depicted in figures 6 and 7. Values range between 832.9 D$_{16\text{-bit-signed}}$/(m s^{-2}) and 850.3 D$_{16\text{-bit-signed}}$/(m s^{-2}). Calibration standard uncertainties are in the order of 4.2 D$_{16\text{-bit-signed}}$/(m s^{-2}), which in relative terms, corresponds to about 0.50 %.

Once all MEMS have been calibrated in order to have a ground truth, the Bayesian method is validated using a bootstrap technique [20, 21]. All analyses are performed using *R*
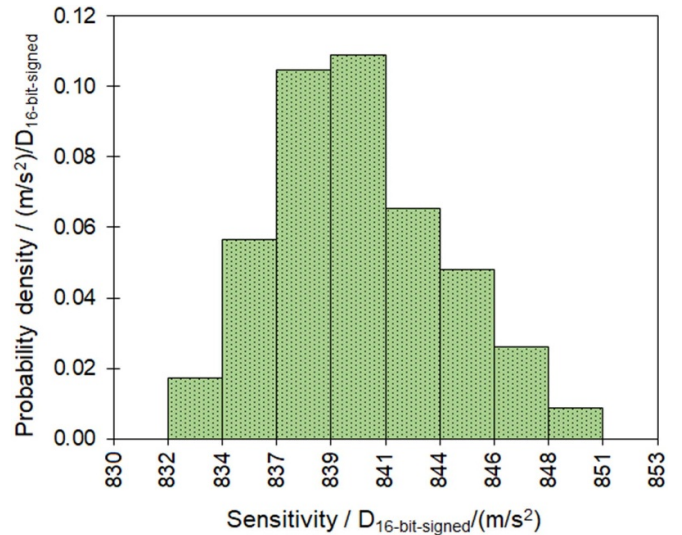

Figure 7. Distribution of the amplitude sensitivities along *z*-axis of the 100 MEMS at 10 Hz.

statistical software. The procedure involves generating multiple samples from the original dataset by randomly sampling observations with replacement. After generating the bootstrap samples, the statistical metric of interest, i.e. the reliability of the unknown batch $P_{\text{reliab,unk}}$, is experimentally evaluated. By replicating the bootstrap process, an accurate estimate of the experimental variability of the proposed metric is provided and compared with the expected theoretical result according to equation (4).

For each bootstrap cycle, the proposed Bayesian statistical method is implemented in the following way: from the 100 experimentally calibrated MEMS sensors, two groups of 50 MEMS each are randomly drawn, the first simulating the benchmark batch and the second simulating the unknown batch. A numerosity of 50 MEMS is chosen so as to have a sufficiently large number of values on which to perform the statistical control. From the benchmark batch, the distribution of
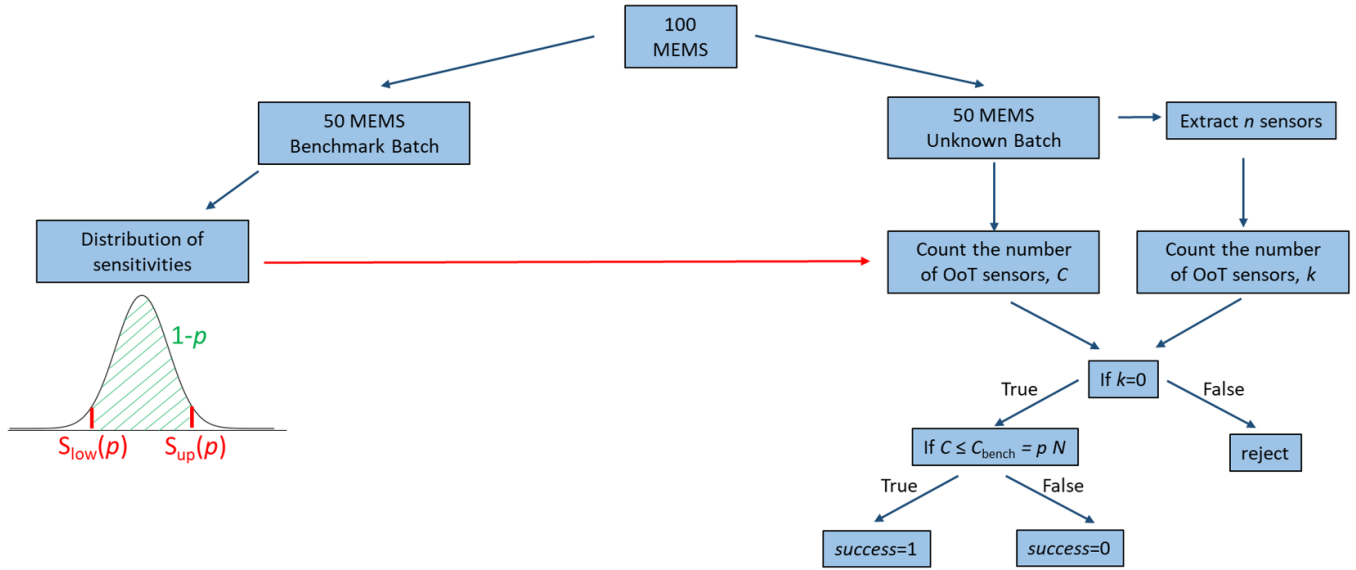
Figure 8. A single cycle of the bootstrap technique iterated r times (where 'success' is a counter variable). 'OoT' stands for 'out-of-tolerance'.

sensitivities is modelled through a mixture distribution which takes into account also the calibration uncertainties. The mixture distribution is obtained using a Monte Carlo simulation, as described in section 2 and, in more detail, in [12]. Once the benchmark batch is characterized and the desired probability p of out-of-tolerance sensors is set, the lower and upper tolerance limits of sensitivity, $S_{low}(p)$ and $S_{up}(p)$, are determined. At this point, in the unknown batch, the overall number C of sensors having out-of-tolerance sensitivities and the k ones in a subsample of n randomly drawn sensors, are counted. If in the sub-sample there are $k = 0$ out-of-tolerance sensors and, in the whole unknown batch, there are $C \leqslant C_{bench} = pN$ out-of-tolerance sensors, a success is counted. Iterating this process $r = 100$ times, the number of successes in r attempts is counted.

The ratio between the number of successes and the number of cases when $k = 0$ represents an estimate of the reliability metric defined in equation (4) based on experimental results, i.e. $P_{reliab,unk,exp}$. A brief scheme of a single cycle, to be iterated r times, is shown in figure 8.

Replicating this entire process $t = 1000$ times, the experimental histogram of the t reliability values $P_{reliab,unk,exp}$ is obtained and compared with the theoretical probability $P_{reliab,unk,theor}$ provided by equation (4).

The validation is performed by applying different boundary conditions, i.e. with n ranging from 5 to 20 (i.e. $n = 5, 10, 20$) and p ranging from 0.02 to 0.16 (i.e. $p = 0.02, 0.08, 0.16$), in order to simulate different scenarios by a factorial design. Boundary conditions and obtained results are summarized in table 1. $\overline{P_{reliab,unk,exp}}$ values represent the mean values from the $t = 1000$ repetitions. In the last column, relative differences between experimental and theoretical probabilities are computed. As an example, the experimental reliability probability histograms of $P_{reliab,unk,exp}$ values obtained for three boundary conditions are depicted in figures 9–11. From these, it is clear that the expected theoretical values fall within the bin of the
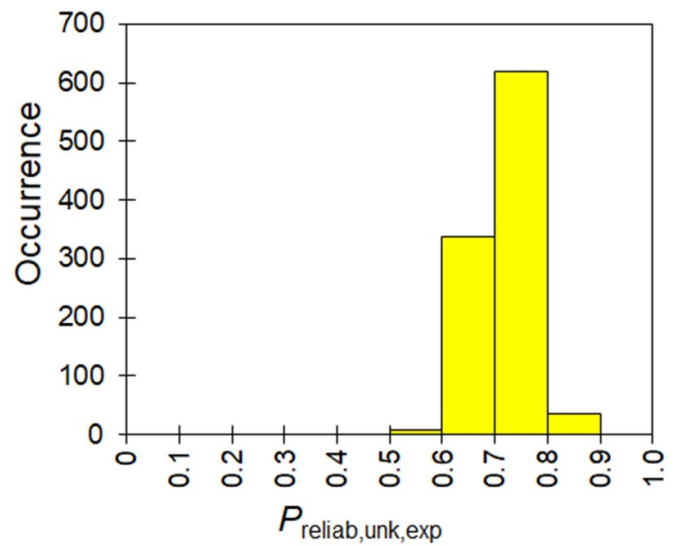
Figure 9. Histogram of $P_{reliability,unk,exp}$ from $t = 1000$ repetitions of the bootstrap process with $n = 5$ and $p = 0.02$.

histogram corresponding to the values with the highest occurrence. In addition, relative differences reported in table 1 lie within $\pm 7\%$. This value can be used to compare the suitability of the method with others that may be developed in the future. Anyway, it seems there is an acceptable agreement with the expected values. These evidences provide a first validation of the proposed method.

Standard uncertainty $u(S_{unk}(p))$ associated with the unknown MEMS according to equation (7) is also calculated for the previous boundary conditions with the addition of two more conditions with $p = 0.32$ and $p = 0.64$ in order to increase the number of cases. As benchmark batch mixture distribution, the one coming from the first bootstrap cycle is used (figure 12). Minimum and maximum values are $S_{min} = 819.6$
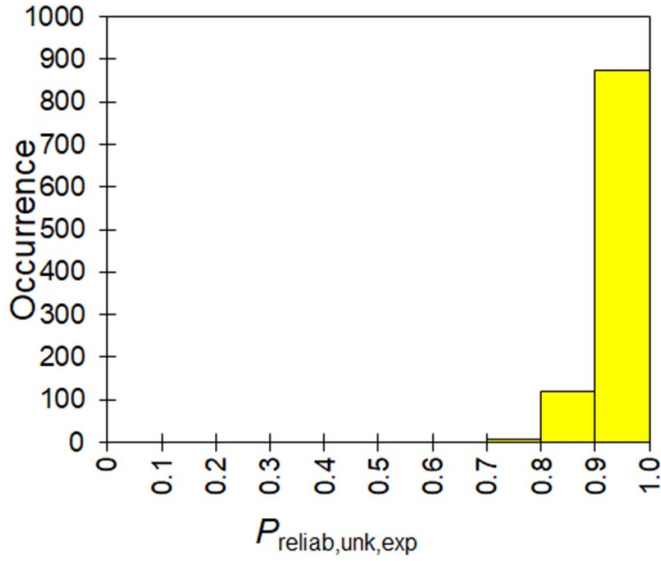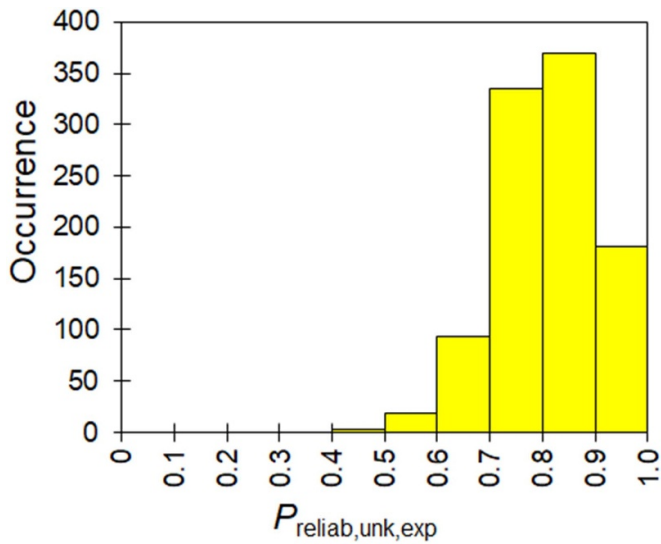
Figure 10. Histogram of $P_{\text{reliability,unk,exp}}$ from $t = 1000$ repetitions of the bootstrap process with $n = 20$ and $p = 0.08$.


Figure 11. Histogram of $P_{\text{reliability,unk,exp}}$ from $t = 1000$ repetitions of the bootstrap process with $n = 10$ and $p = 0.16$.

$D_{\text{16-bit-signed}}/(\text{m s}^{-2})$ and $S_{\text{max}} = 861.3\ D_{\text{16-bit-signed}}/(\text{m s}^{-2})$, respectively. Lower and upper bounds, $S_{\text{low}}(p)$ and $S_{\text{up}}(p)$, are found based on p. Results are reported in table 2.

Sensitivities of the unknown MEMS sensors, calculated according to equation (5), lie between 840.5 $D_{\text{16-bit-signed}}/(\text{m s}^{-2})$ and 841.0 $D_{\text{16-bit-signed}}/(\text{m s}^{-2})$, indicating a rather symmetrical distribution of the mixture distribution benchmark batch despite different p. Standard uncertainties lie between 6.9 $D_{\text{16-bit-signed}}/(\text{m s}^{-2})$ and 11.8 $D_{\text{16-bit-signed}}/(\text{m s}^{-2})$ which in relative terms correspond to about 1.1 %. Comparing this value with the calibration standard uncertainty (0.5 %, see section 2), it is two times larger. This is the price to pay for avoiding a huge number of experimental calibrations. However, for many practical applications where very low uncertainties are not required, this represents a good

Table 1. Results of the repeated bootstrap process with different boundary conditions.

Exp. #	n	p	$\overline{P_{\text{reliab,unk,exp}}}$	$P_{\text{reliab,unk,theor}}$	Rel. diff./%
1	5	0.02	0.7190	0.7729	−7.0
2	10	0.02	0.7665	0.8095	−5.3
3	20	0.02	0.8490	0.8795	−3.5
4	5	0.08	0.7524	0.7098	6.0
5	10	0.08	0.8426	0.7868	7.1
6	20	0.08	0.9521	0.9126	4.3
7	5	0.16	0.6695	0.7128	−6.1
8	10	0.16	0.8144	0.8203	−0.7
9	20	0.16	0.9549	0.9597	−0.5
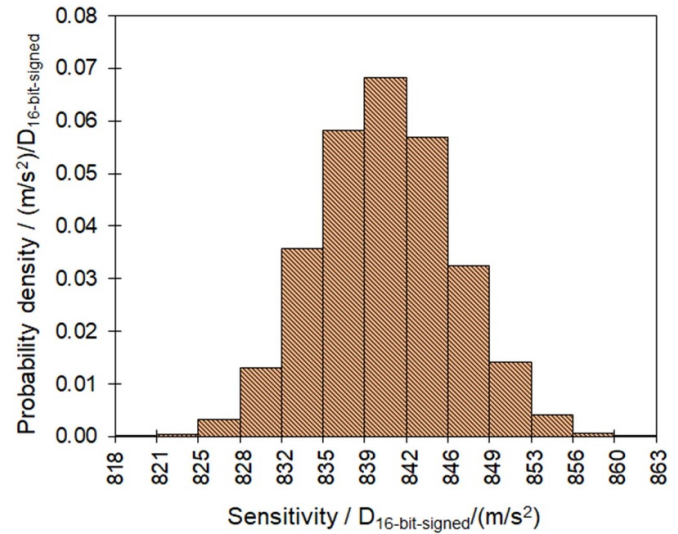
Figure 12. Benchmark batch mixture distribution of the $10^5 \times N$ simulated sensitivities coming from the first bootstrap cycle ($N = 50$).

Table 2. Evaluation of the uncertainty associated to the statistically calibrated MEMS sensors using equation (7). Sensitivities are reported in $D_{\text{16-bit-signed}}/(\text{m s}^{-2})$.

p	$S_{\text{low}}(p)$	$S_{\text{up}}(p)$	$S_{\text{unk}}(p)$	$u(S_{\text{unk}}(p))$
0.02	828.0	854.0	841.0	7.7
0.08	830.9	850.6	840.8	6.9
0.16	832.7	848.6	840.7	7.2
0.32	834.9	846.3	840.6	8.8
0.64	837.8	843.2	840.5	11.8

compromise that addresses the need for traceability with lower efforts compared to the traditional metrological chain. It is also worth noting that the minimum uncertainty, with the considered boundary conditions, is found for $p = 0.08$. However, this result does not hold universally but depends on the characteristics of the mixture distribution of the benchmark batch, which influences the value of lower and upper bounds. Depending on the needs, therefore, an appropriate trade-off between batch numerosity, N, number of sensors to be calibrated, n, reliability $P_{\text{reliab,unk}}$ and uncertainty of the

statistically calibrated MEMS, $u(S_{unk}(p))$, should be decided each time.

4. Conclusions

In this paper, a Bayesian approach to provide a statistical calibration of large batches of MEMS sensors is described and proposed. The need for this work arises from the impossibility of calibrating each MEMS one by one following the traceability standards of traditional measurement instruments.

It is important to underline that the proposed method is to address the need to extend traceability to these types of sensors, without affecting the current metrology chain for traditional measurement transducers currently used (one-to-one calibration). This aspect is crucial since this statistical methodology is in no way intended to replace the current rigorous metrological chain, nor to be used in applications that require measurements with high accuracy and low uncertainty. In fact, MEMS-based sensors are not intended as replacements for current traditional instrumentation but as complementary to them for the detection of physical phenomena, typically on a large scale through the use of extended sensor networks [22]. In practical applications, an extended sensor network, containing both statistically calibrated MEMS technology-based sensors and traditional higher-quality measurement instruments in a few nodes, is certainly much safer and more trustworthy (and traceable) than any other network containing sensors of unknown or undefined sensitivity.

Given this premise, the proposed idea is to experimentally calibrate only a small subset of sensors from an unknown batch and verify that their sensitivities lie within previously defined limits, which depend on the calibration of a benchmark batch representative of the whole production process, and on a desired probability $(1 - p)$ of acceptability. This entails a dramatic decrease in calibration efforts. Depending on the number of sensors in the unknown batch N, the number n of sampled sensors, and the number k of out-of-tolerance sensors found in the subsample, it is possible to evaluate the probability of having a number C of out-of-tolerance sensors in the unknown batch. Assuming as a desirable feature that of having in the unknown batch a number of out-of-tolerance sensors C not larger than the number of defective sensors in the benchmark batch C_{bench}, a batch reliability $P_{reliab,unk}$ can be defined as a suitable metric. This is an additional information that characterizes the reliability of the batches that will be released. Of course, calibrating a lower than N number of sensors implies a reliability below 100 %, which is instead the current metrological chain requirement. This signifies the potential for an impact on subsequent measurements conducted with these sensors. In practical terms, there exists a non-zero probability that the sensitivity of certain statistically calibrated sensors may fall outside the specified limits. However, the emergence of cutting-edge machine learning and deep learning methods [23] might offer a possible means to mitigate this issue by assessing the congruence of the physical signals detected by sensors within the same network.

Another challenge lies in the calibration of the benchmark batch, which can still demand a substantial experimental effort. The extent of this effort depends on the desired reliability. In fact, to achieve the highest attainable reliability, in general it is advisable to work with larger batch sizes, even though this approach needs a heightened calibration effort for the benchmark batch. Consequently, this might result in a considerable volume of measurements. Nevertheless, this outcome is highly contingent on factors such as the sensor type, measurement technique, calibration conditions, and the calibration system. To minimize the number of necessary calibrations, a favourable compromise between high reliability and reduced experimental workload involves selecting lower values of p, allowing a decrease in batch size N while still choosing very small samples from the unknown batches. This strategy might offer an optimal solution as it ensures a reduced calibration workload while maintaining acceptable levels of reliability simultaneously.

In addition to the batch reliability metric, this work gives also indiction on how to provide the $(N - n)$ statistically calibrated sensors in the unknown batch with a (statistical) sensitivity value and a projection of the associated uncertainty, both depending on the desired value of p and the characteristics of the benchmark batch mixture distribution. The proposed method is applied and validated with a repeated bootstrap technique through the calibration of 100 nominally identical digital MEMS accelerometers. The results showed an acceptable agreement between experimental-based bootstrap and theoretical values within ± 7 % in terms of relative difference, confirming the validity of the method.

In the future, it will be necessary to extend this methodology to encompass larger batches of sensors to determine if the level of agreement can be enhanced during the validation process.

Appendix. Mathematical formulation

Assuming that experimental calibration and manufacturing process do not change from batch to batch so that the expected number of out-of-tolerance sensors in the unknown batch is equal to that in the benchmark batch, C_{bench}, a binomial prior distribution representing the state of knowledge on the number of out-of-tolerance sensors C in a typical batch from that production can be given by

$$f_{prior}(C; N, p) = \begin{pmatrix} N \\ C \end{pmatrix} p^{C} (1 - p)^{N - C} \qquad \text{(A.1)}$$

where N is the number of sensors in a batch and p is the probability of out-of-tolerance sensors in the benchmark batch.

According to Bayes' theorem, the prior distribution is multiplied by a likelihood function $f_{like}(k; n, C, N)$, which is defined as a hypergeometric distribution with n calibrated sensors from the unknown batch, k of which are defective, drawn from the unknown batch of N sensors, C of which are defective, according to:

$$f_{\text{like}}(k;n,C,N) = \frac{\begin{pmatrix} C \\ k \end{pmatrix} \begin{pmatrix} N-C \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}. \tag{A.2}$$

This product yields an un-normalized posterior $f_{\text{post,un}}(C; k, n, N, p)$ for C out-of-tolerance sensors in the unknown batch according to:

$$
\begin{aligned}
&f_{\text{post,un}}(C;k,n,N,p) \\
&= f_{\text{prior}}(C;N,p)\,f_{\text{like}}(k;n,C,N) \\
&= \begin{pmatrix} N \\ C \end{pmatrix} p^C (1-p)^{N-C} \frac{\begin{pmatrix} C \\ k \end{pmatrix}\begin{pmatrix} N-C \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} \\
&= p^C (1-p)^{N-C} \frac{N!}{C!\,(N-C)!}\frac{C!}{k!\,(C-k)!} \\
&\quad \frac{(N-C)!}{(n-k)!\,(N-C-n+k)!}\frac{n!\,(N-n)!}{N!} \\
&= p^C (1-p)^{N-C} \begin{pmatrix} n \\ k \end{pmatrix} \begin{pmatrix} N-n \\ C-k \end{pmatrix}. \tag{A.3}
\end{aligned}
$$

Normalizing this distribution by dividing it by the summation of $f_{\text{post,un}}(C; k, n, N, p)$ for C ranging from 0 to N, i.e. $f_{\text{norm}}(k, n, N, p)$, defined as

$$
\begin{aligned}
&f_{\text{norm}}(k,n,N,p) \\
&= \sum_{C=0}^{N} f_{\text{post,un}}(C;k,n,N,p) \\
&= \sum_{C=0}^{N} p^C (1-p)^{N-C} \begin{pmatrix} n \\ k \end{pmatrix}\begin{pmatrix} N-n \\ C-k \end{pmatrix} \\
&= \sum_{C=0}^{N} p^C (1-p)^{N-C} \begin{pmatrix} n \\ k \end{pmatrix}\begin{pmatrix} N-n \\ C-k \end{pmatrix} \frac{p^{-k}(1-p)^{k-n}}{p^{-k}(1-p)^{k-n}} \\
&= p^k (1-p)^{n-k} \begin{pmatrix} n \\ k \end{pmatrix} \underbrace{\sum_{C=0}^{N} p^{C-k}(1-p)^{N-n-C+k}\begin{pmatrix} N-n \\ C-k \end{pmatrix}}_{1} \\
&= p^k (1-p)^{n-k} \begin{pmatrix} n \\ k \end{pmatrix}. \tag{A.4}
\end{aligned}
$$

The probability mass function $f_{\text{post,norm}}$ of the number C of out-of-tolerance sensors in the unknown batch, given that there are k out-of-tolerance sensors in the small calibrated sub-batch, can found be according to:

$$
\begin{aligned}
&f_{\text{post,norm}}(C;k,n,N,p) \\
&= \frac{f_{\text{post,un}}(C;k,n,N,p)}{f_{\text{norm}}(k,n,N,p)} \\
&= \frac{p^C (1-p)^{N-C} \begin{pmatrix} n \\ k \end{pmatrix}\begin{pmatrix} N-n \\ C-k \end{pmatrix}}{p^k (1-p)^{n-k} \begin{pmatrix} n \\ k \end{pmatrix}} \\
&= p^{C-k}(1-p)^{N-n-C+k}\begin{pmatrix} N-n \\ C-k \end{pmatrix}. \tag{A.5}
\end{aligned}
$$

ORCID iDs

Andrea Prato ⓘ https://orcid.org/0000-0003-3733-4189
Francesca Pennecchi ⓘ https://orcid.org/0000-0003-1328-3858
Gianfranco Genta ⓘ https://orcid.org/0000-0002-2983-1196
Alessandro Schiavi ⓘ https://orcid.org/0000-0003-4168-3605

References

[1] JCGM 200 2012 *International Vocabulary of Metrology Basic and General Concepts and Associated Terms (VIM 3rd edition)* (Joint Committee for Guides in Metrology)

[2] STMicroelectronics Q4 & FY 2022 Financial Results (available at: https://investors.st.com/static-files/4566e70d-3e92-4bda-80c0-944cfc24e2ee) (Accessed 24 April 2023)

[3] Ciuti G, Ricotti L, Menciassi A and Dario P 2015 MEMS sensor technologies for human centred applications in healthcare, physical activities, safety and environmental sensing: a review *Sensors* **15** 6441–68

[4] Eichstädt S, Gruber M, Vedurmudi A P, Seeger B, Bruns T and Kok G 2019 Toward smart traceability for digital sensors and the industrial internet of things *Sensors* **21** 2019

[5] Prato A, Mazzoleni F and Schiavi A 2020 Traceability of digital 3-axis MEMS accelerometer: simultaneous determination of main and transverse sensitivities in the frequency domain *Metrologia* **57** 035013

[6] Smorgon D, Fernicola V, Sousa J A, Ribeiro L, Tamburini E and Catto M 2020 Assuring measurement traceability to ATE systems for MEMS temperature sensors testing and calibration *Proc. IEEE Metrology for Industry 4.0 & IoT* (https://doi.org/10.1016/j.foodchem.2020.128174)

[7] Dorst T, Ludwig B, Eichstädt S, Schneider T and Schütze A 2019 Metrology for the factory of the future: towards a case study in condition monitoring *2019 IEEE Int. Instrumentation and Measurement Technology Conf. (I2MTC)* (*Auckland, New Zealand*) pp 1–5

[8] BIPM 2021 *Consultative Committee for Acoustics, Ultrasound, and Vibration (CCAUV), Strategy Plan 2021–2031* (available at: www.bipm.org/documents/20126/2070930/CCAUV+Strategy.pdf/30b85b6d-6d30-6a0f-b716-a95c7171484) (Accessed 24 July 2023)

[9] Meija J, Bodnar O and Possolo A 2023 Ode to Bayesian methods in metrology *Metrologia* **60** 052001

[10] IEEE 2700-2017 *IEEE standard for sensor performance parameter definitions*

[11] Seidel W 2011 Mixture models *International Encyclopedia of Statistical Science* ed M Lovric (Springer) pp 827–9

[12] Prato A, Pennecchi F R, Genta G and Schiavi A 2022 Mixture distribution modelling of the sensitivities of a digital 3-axis MEMS accelerometers large batch *IEEE Int. Workshop on Metrology for Industry 4.0 & IoT* pp 223–8

[13] van der Veen A and Cox M G 2021 Getting started with uncertainty evaluation using the Monte Carlo method in R *Accreditation Qual. Assur.* **26** 129–41

[14] van de Schoot R *et al* 2021 Bayesian statistics and modelling *Nat. Rev. Meth. Primers* **1** 1

[15] ISO 16063-21 2003 Methods for the Calibration of Vibration and Shock Transducers—Part 21: Vibration Calibration by Comparison to a Reference Transducer (International Organization for Standardization)

[16] Prato A, Mazzoleni F and Schiavi A 2019 Metrological traceability for digital sensors in smart manufacturing: calibration of MEMS accelerometers and microphones at

INRiM *II Workshop on Metrology for Industry 4.0 and IoT* pp 371–5

[17] Prato A, Schiavi A, Mazzoleni F, Touré A, Genta G and Galetto M 2020 A reliable sampling method to reduce large set of measurements: a case study on calibration of digital 3-axis MEMS accelerometers *IEEE Int. Workshop on Metrology for Industry 4.0 and IoT* pp 537–41

[18] Prato A, Mazzoleni F, Pennecchi F R, Genta G, Galetto M and Schiavi A 2021 Towards large-scale calibrations: a statistical analysis on 100 digital 3-axis MEMS accelerometers *IEEE Int. Workshop on Metrology for Industry 4.0 & IoT* pp 578–82

[19] Galetto M, Schiavi S, Genta G, Prato A and Mazzoleni F 2019 Uncertainty evaluation in calibration of low-cost digital MEMS accelerometers for advanced manufacturing applications *CIRP Ann.* **68** 535–8

[20] Bradley E 1979 Bootstrap methods: another look at the jackknife *Ann. Stat.* **7** 1–26

[21] Possolo A 2023 Measurement science meets the reproducibility challenge *Metrologia* **60** 044002

[22] Prato A, Mazzoleni F, D'Emilia G, Gaspari A, Natale E and Schiavi A 2021 Metrological traceability of a digital 3-axis MEMS accelerometers sensor network *Measurement* **184** 109925

[23] Chen A, Jia W, Hwang F, Liu G, Song F and Pu L 2022 Machine learning and deep learning methods for wireless network applications *J. Wirel. Commun. Netw.* **2022** 115