ISO Linear Calibration and Measurement Uncertainty of the Result Obtained With the Calibrated Instrument

Jakub Palenčár[1], Rudolf Palenčár[1], Miroslav Chytil[2], Gejza Wimmer Jr.[3], Gejza Wimmer[4], Viktor Witkovský[3★]

[1]*Faculty of Mechanical Engineering of the Slovak University of Technology in Bratislava, Námestie slobody 17, 812 31 Bratislava, Slovak Republic, jakub.palencar@stuba.sk, rudolf.palencar@stuba.sk,*

[2]*Slovak Institute of Metrology, Karloveská 63, 842 55 Bratislava, Slovak Republic, chytil@smu.gov.sk,*

[3]*Institute of Measurement Science of the Slovak Academy of Sciences, Dúbravská cesta 9, 841 07 Bratislava, Slovak Republic, umerwimm@savba.sk, witkovsky@savba.sk (★corresponding author),*

[4]*Mathematical Institute of the Science of the Slovak Academy of Sciences, Štefánikova 49, 814 73 Bratislava, Slovak Republic, wimmer@mat.savba.sk.*

Abstract: We address the problem of linear comparative calibration, a special case of linear calibration where both variables are measured with errors, and the analysis of the uncertainty of the measurement results obtained with the calibrated instrument. The concept is explained in detail using the calibration experiment of the pressure transducer and the subsequent analysis of the measurement uncertainties. In this context, the calibration and the measurements with the calibrated instrument are performed according to ISO Technical Specification 28037:2010 (here referred to as ISO linear calibration), based on the approximate linear calibration model and the application of the law of propagation of uncertainty (LPU) in this approximate model. Alternatively, estimates of the calibration line parameters, their standard uncertainties, the coverage intervals and the associated probability distributions are obtained using the Monte Carlo method (MCM) based on the law of propagation of distributions (LPD). Here we also obtain the probability distributions and the coverage interval for the quantities measured with the calibrated instrument. Furthermore, motivated by the model structure of this particular example, we conducted a simulation study that presents the empirical coverage probabilities of the ISO and MCM coverage intervals and investigates the influence of the sample size, i.e. the number of calibration points in the measurement range, and the different combinations of measurement uncertainties. The study generally confirms the good properties and validity of the ISO technical specification within the considered (limited) framework of experimental designs motivated by real-world application, with small uncertainties in relation to the measurement range. We also point out the potential weaknesses of this method that require increased user attention and emphasise the need for further research in this area.

Keywords: Linear comparative calibration, ISO Technical Specification 28037:2010, Monte Carlo method, measurement uncertainty, calibrated instrument, empirical coverage probability.

1. INTRODUCTION

Calibration is at the heart of measurement science and covers a wide range of industries and services. Calibration is an essential component of many measurement procedures, the first step of which is to fit a calibration function that best describes the relationship between the variables of interest. In particular, the *International vocabulary of metrology* (VIM) [4] specifies that „*calibration is defined as an operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication*".

The process of measuring, evaluating and expressing uncertainties in measurement results is described in detail in basic metrological documents, see the *Guide to the expression of uncertainty in measurement* (GUM) [1] and its supplements [2] and [3]. For further details see also the *International vocabulary of metrology* [4] and the other relevant documents, e.g. [6] and [7].

The main objective of calibration is to describe the theoretical relationship between the response variable (dependent variable) and the stimulus (independent variable), by a calibration function that belongs to a certain class of functions that can be assumed to best capture this relationship (as the reviewer notes, this does not exclude situations with multiple stimuli and responses). The simplest, but from a practical point of view most commonly used type of calibration function is the linear calibration function.

Comparative calibration, see e.g. [28], is a special type of calibration in which both variables included in the calibration experiment are subject to error (including the used measure-

ment standard). The appropriate modelling approach is formally based on the well-established errors-in-variables (EIV) modelling and total least squares estimation methods, see e.g. [10], [13], [15], [21], [22], and [25]. As the reviewer noted, error-in-variables models are typically used to compare methods with approximately equal variance (method comparison). However, this type of calibration is also used to compare two measuring devices or measurement methods, one of which is considered more accurate (the gold standard) and the other less accurate (a cheaper alternative recommended for typical applications with frequent measurements).

The calibration function expresses the relationship between the true values of the measured quantities expressed in the units of the measuring instruments used, e.g. $\mathscr{X}$ and $\mathscr{Y}$. An important goal of the calibration experiment is to express the true quantity values measured by the measuring instrument $\mathscr{Y}$ (response) as a function of the true quantity values measured by the measuring instrument $\mathscr{X}$ (stimulus). In the first phase of the calibration experiment, we are primarily interested in finding the legitimate and correct (in the specific sense, the best) estimators for the parameters of the calibration function. Since the measured data are subject to uncertainties, there are of course also uncertainties in the estimated parameters of the calibration function. The estimated parameters of the calibration function are further used to predict the values of the unobservable stimuli. These predictions are based on new readings received from the calibrated measurement instrument.

In linear comparative calibration based on the EIV modelling approach, it is possible to reverse the traditional roles of the stimulus variable (which is measured with the more accurate measurement device and usually represented by values on the x axis) and the response variable (which is measured with the less accurate measurement device and usually represented by values on the y axis). In the first stage of the calibration experiment, i.e. when fitting the respective calibration function by estimating its coefficients and assessing the associated uncertainties, this change has no negative numerical impact on the estimated parameters and their uncertainties if they are based on the first-order linearisation of the original nonlinear models (as with a linear model, there is a one-to-one correspondence between the coefficients of the direct and inverse linear calibration functions). In the second stage of calibration, this estimation approach allows direct prediction of the unknown stimulus from the observed response (rather than indirect prediction as in the traditional method). This brings some numerical advantages in calculating the associated uncertainty and the probability distribution of the values that can reasonably be attributed to the stimulus.

As the reviewer noted, it is quite natural to generalise the EIV modelling approach to polynomial calibration as in [16]. However, in this paper we focus only on the linear (straight-line) calibration problem and the comparison of the standard approach based on ISO Technical Specification 28037:2010 [5] (here referred to as ISO) with the Monte Carlo method (here referred to as MCM). By using MCM to propagate the distributions, see [2] and [3], we can derive a more general and accurate probability distribution of the values of the cal-

ibration function parameters and the stimulus value based on the measurement results with a calibrated instrument. Both approaches are illustrated and compared using an example with real data, the calibration of a pressure transducer performed at the Slovak Institute of Metrology. In the subsequent simulations we also investigate the empirical coverage probabilities of the ISO and MCM coverage intervals and the influence of the sample size, i.e. the number of calibration points in the measurement range, as well as the behaviour of the results with respect to different combinations of measurement uncertainties.

Another approach to estimating the calibration parameters and associated uncertainties could be based on the use of the local best linear unbiased estimator (BLUE) in a properly linearised model (an approximate first-order regression model) with linear constraints on its parameters, as proposed in [19]. This approach allows the development of criteria for testing the adequacy of the first-order approximation of the model near a particular point in parameter space, as proposed in [20], as well as the development of a generalised estimation approach based on the locally best linear unbiased estimator in the second-order approximate regression model. The advantage of these approaches is that the parameter estimator can be expressed as a locally linear function of the observed data, which allows the calculation of the associated (but still approximate) probability distributions based on the characteristic function approach (CFA) as described and implemented in [32], [34], and [36]. A detailed description of the proposed approach for polynomial comparative calibration can be found in [33] and [35].

The paper is structured as follows. In Section 2 we specify the linear comparative calibration model under consideration as a nonlinear regression model (in fact, the model could be represented as a linear regression model with nonlinear constraints on its parameters) and the approach to estimating the parameters and associated uncertainties as proposed in [5]. Formally, however, the measurement model required by GUM can be defined implicitly as a solution to the generalised Gauss-Markov regression problem. In Section 3 we present an example of linear calibration of a pressure transducer using a pressure standard. Here we specify the models of direct measurements with both instruments (statistical models of quantities representing the response of the instruments). Further, we derive estimates of the parameters of the calibration function and their uncertainties and coverage intervals according to [5]. We also give the estimate of the stimulus corresponding to the new indication observed by the calibrated instrument, together with the associated standard uncertainty and the derived coverage interval. Section 4 presents an alternative approach to determine the distribution of values that could be attributed to the parameters of the calibration function obtained by the Monte Carlo method and the distribution of values that could be attributed to the quantity Y, where $Y = a + bX$ given the distribution of X, and compares the results of the two approaches. Section 5 compares the empirical coverage probabilities of the ISO and MCM methods by means of a small simulation study motivated by the model structure of this particular example. Here we examine

the influence of the sample size, i.e. the number of calibration points in the measurement range, and the various combinations of measurement uncertainties of the measured quantities. Section 6 provides conclusions and recommendations for practitioners in measurement and metrology.

2. LINEAR COMPARATIVE CALIBRATION MODEL

Here we consider a linear calibration function that expresses the quantity values in units of the more accurate measurement instrument $\mathscr{Y}$ (now called stimulus) as a linear function of the quantity values in units of the less accurate measurement instrument $\mathscr{X}$ (now called indication or response). We consider the following setup of the calibration experiment, where the measured quantities are represented by a set of m artefacts with given properties (suitably chosen objects representing the measured quantities), say $V_1, \ldots, V_m$, such that their true values μ_i, $i = 1, \ldots, m$, expressed in units of the calibrated instrument $\mathscr{X}$, span the required calibration range. The true values of the measured quantities $V_1, \ldots, V_m$, expressed in units of the standard instrument $\mathscr{Y}$, are denoted by v_i, $i = 1, \ldots, m$. Each object could be measured repeatedly and independently, say $n \geq 1$ times, by both measuring instruments. The j-th measurement of the i-th object is thus represented by the random variable $X_{i,j}$, where $i = 1, \ldots, m$ and $j = 1, \ldots, n$. At the same time, each object is also measured with $\mathscr{Y}$. The j-th measurement of the i-th object is represented by the random variable $Y_{i,j}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$. The calibration experiment is thus based on measurements provided by the measuring devices under consideration and represented by pairs of random variables $(X_{i,j}, Y_{i,j})$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

Often the calibration experiment is presented in its aggregate form (by appropriately averaging the repeated measurements and combining them with other expert knowledge) by the pairs of random variables (X_i, Y_i) representing measurements of the quantity values μ_i and v_i for $i = 1, \ldots, m$. Their joint probability distribution is specified by the assumed statistical model which reflects knowledge about the quantity values μ_i and v_i, their functional relation, and the applied measurement process. The information received from the calibration experiment is given by the observed values from the direct measurements, denoted by (x_i, y_i), their associated uncertainties $u(x_i)$ and $u(y_i)$, and the covariances $u(x_i, x_j)$, $u(y_i, y_j)$. Technically, it is possible to consider also non-zero covariances $u(x_i, y_j)$, but as the reviewer noted, in a well-designed experiment it is very unlikely that the associated measurement errors are correlated, so here we assume that $u(x_i, y_j) = 0$ for $i, j = 1, \ldots, m$. According to [1], note 3 in subclause 2.2.3, it is understood that the result of the measurement is the best estimate of the value of the measurand, and that all components of uncertainty, including those arising from systematic effects, such as components associated with corrections and reference standards, contribute to the dispersion. In general, this modelling approach provides a sufficiently flexible tool to incorporate expert knowledge about the distribution of measurement error and systematic effects through a set of input variables.

The possible values of the input quantities are fully specified by the probability distributions based on a detailed uncertainty analysis of the results of the direct measurements. This allows to combine type A and type B methods of evaluation of the measurement uncertainties and to include dependency structures between input variables, e.g. correlated measurements, see [1], subclauses 3.3.4–3.3.6.

Given observed (x_i, y_i), the preliminary estimates of (μ_i, v_i) based on the direct measurements (together with their associated uncertainties), we are interested in finding the linear calibration function of the following form,

$$v = a + b\mu, \tag{1}$$

where v is the true (error-free) quantity value expressed in units of the measuring device $\mathscr{Y}$ (the more accurate standard instrument), μ is the true (error-free) quantity value expressed in units of the measuring device $\mathscr{X}$ (the less accurate calibrated instrument), and a and b represent the unknown coefficients (intercept and slope) of the linear calibration function. The purpose of a calibration procedure is to determine estimates for the coefficients a and b. Since the measured data are subject to uncertainties, the estimated coefficient values of a and b are also subject to uncertainties.

Although the calibration function (1) is linear in parameters a and b, the corresponding calibration model is expressed as a nonlinear regression model in the set of all unknown model parameters a, b and μ_i, $i = 1, \ldots, m$, due to errors in both measured quantities. In particular, the following statistical model is specified by the expectation and the covariance matrix of the measurements,

$$E\begin{pmatrix}\mathbf{X}\\\mathbf{Y}\end{pmatrix} = \begin{pmatrix}\mu\\a\mathbf{1} + b\mu\end{pmatrix}, \text{ and } Cov\begin{pmatrix}\mathbf{X}\\\mathbf{Y}\end{pmatrix} = \mathbf{U}, \tag{2}$$

where $\mathbf{X} = (X_1, \ldots, X_m)'$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ represent the direct measurements, $\mathbf{1}$ denotes the $(m \times 1)$-dimensional column vector of ones, a, b and $\mu = (\mu_1, \ldots, \mu_m)'$ are the model parameters and $\mathbf{U}$ is the known variance-covariance matrix specified by its blocks $\mathbf{U}_x$, $\mathbf{U}_y$, and $\mathbf{U}_{x,y}$ (here we assume $\mathbf{U}_{x,y} = \mathbf{U}'_{y,x} = \mathbf{0}$). Note that $E(\mathbf{Y}) = v$ with $v = a\mathbf{1} + b\mu$. Thus, we can represent (2) as a linear regression model with nonlinear constraints on its parameters.

One possible definition of the measurement model (which is considered here) can be based on the solution of the generalized Gauss-Markov regression problem,

$$\begin{pmatrix}\hat{a}\\\hat{b}\\\hat{\mu}\end{pmatrix} = \min_{a,b,\mu} Q(\mathbf{X}, \mathbf{Y}, a, b, \mu), \tag{3}$$

where

$$Q(\mathbf{X}, \mathbf{Y}, a, b, \mu) = \begin{pmatrix}\mathbf{X} - \mu\\\mathbf{Y} - a\mathbf{1} - b\mu\end{pmatrix}' \mathbf{U}^{-1} \begin{pmatrix}\mathbf{X} - \mu\\\mathbf{Y} - a\mathbf{1} - b\mu\end{pmatrix}. \tag{4}$$

Hence, the measurement model can be expressed as

$$\frac{\partial Q(\mathbf{X}, \mathbf{Y}, a, b, \mu)}{\partial (a, b, \mu')'} = \mathbf{0}, \tag{5}$$

where **X**, and **Y** are the input quantities and a, b, and μ are the output quantities. The measurement model (5) is nonlinear and implicit, see [17].

As one reviewer suggested, an alternative choice of measurement model is based on maximising a correctly specified likelihood function, which in turn allows statistical inference about the parameters of the calibration line, which is asymptotically optimal and correct for large m. From this point of view, model (3) is inherently an optimal choice under the assumption of normality of the inputs **X** and **Y**, with the resulting maximum likelihood estimates having known asymptotic optimality properties, see [16] and [11]. For example, if the inputs **X** and **Y** are sampled from independent joint m-dimensional multivariate Student t-distributions with k and l degrees of freedom, say $t_k(\mu, \mathbf{U}_x)$ and $t_l(\nu, \mathbf{U}_y)$ where $\nu = a\mathbf{1} + b\mu$, see [18], then the suggested optimum measurement model based on maximising properly specified likelihood function would be

$$
\begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{\mu} \end{pmatrix} = \min_{a,b,\mu} \left\{ (k+m)\log\left(1 + \frac{1}{k}(\mathbf{X}-\mu)'\mathbf{U}_x^{-1}(\mathbf{X}-\mu)\right) \right.
$$
$$
\left. + (l+m)\log\left(1 + \frac{1}{l}(\mathbf{Y}-a\mathbf{1}-b\mu)'\mathbf{U}_y^{-1}(\mathbf{Y}-a\mathbf{1}-b\mu)\right) \right\}.
$$
(6)

Furthermore, if X_i and Y_j can be modelled as independent random variables with the shifted and scaled Student t-distributions with k_i and l_j degrees of freedom, respectively, here denoted as $t_{k_i}\left(\mu_i, u(x_i)^2\right)$ and $t_{l_j}\left(\nu_j, u(y_j)^2\right)$, for $i, j = 1, \ldots, m$, then according to [16] the proposed optimal measurement model based on maximising the correctly specified likelihood function would be

$$
\begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{\mu} \end{pmatrix} = \min_{a,b,\mu} \left\{ \sum_{i=1}^{m}(k_i+1)\log\left(1 + \frac{(x_i-\mu_i)^2}{k_i u(x_i)^2}\right) \right.
$$
$$
\left. + (l_i+1)\log\left(1 + \frac{(y_i-a-b\mu_i)^2}{l_i u(y_i)^2}\right) \right\}. \quad (7)
$$

As the reviewer noted, the maximum likelihood approach immediately provides a large sample approximation to the standard uncertainties and correlation of the estimated parameters via the Hessian of the log-likelihood function, evaluated at the maximum likelihood estimates.

In more complex situations, and this also applies to the calibration example discussed in this article, deriving a properly specified likelihood function can be complicated (as it requires convolution of multivariate distributions), which we did not address in this article.

Therefore, in this paper we are concerned with the measurement model (3), despite its potential suboptimality if the normality assumptions are violated, since such parameter estimation is implemented in ISO/TS and we want to investigate its properties in situations that do not fully correspond to the optimal assumptions. Proper likelihood inference in the errors-in-variables model is beyond the scope of this paper. For more details see [16] and [27].

Thus, given the calibration data $\mathbf{x} = (x_1, \ldots, x_m)'$ and $\mathbf{y} = (y_1, \ldots, y_m)'$ together with the associated covariance matrix **U**, the WTLS estimates of the coefficients of the linear calibration function, say $\hat{a}$ and $\hat{b}$, are given as a solution to the generalised Gauss-Markov regression problem by minimising the weighted total least squares

$$
\min_{a,b,\mu} \left\{ \begin{pmatrix} \mathbf{x}-\mu \\ \mathbf{y}-\nu \end{pmatrix}' \mathbf{U}^{-1} \begin{pmatrix} \mathbf{x}-\mu \\ \mathbf{y}-\nu \end{pmatrix} \right\}, \quad (8)
$$

where $\mu = (\mu_1, \ldots, \mu_m)'$ represents the vector of true values of the measured quantities in units of the less accurate instrument and $\nu = (\nu_1, \ldots, \nu_m)'$ represents the vector of true values of the measured quantities in units of the more accurate instrument, such that

$$
\nu = a\mathbf{1} + b\mu. \quad (9)
$$

We would like to emphasise that the model has $2m$ measurements for m calibration points, $\mathbf{x} = (x_1, \ldots, x_m)'$ and $\mathbf{y} = (y_1, \ldots, y_m)'$ with $2 + m$ parameters, namely a, b and $\mu = (\mu_1, \ldots, \mu_m)'$. If the uncertainties cannot be assumed to be known, there are of course more parameters that need to be estimated.

Since the explicit solution is only available for very special cases, the parameters of the calibration function are usually determined by iterative numerical procedures that minimise the criterion function (8) based on the total sum of the weighted squares of the residues.

In recent years, several estimation strategies and efficient algorithms have been proposed that are suitable for nonlinear regression and calibration models in measurement and metrology, see e.g. [13], [21], [26], [23], [22], [24], and [9]. The estimated covariance matrix is usually based on the information matrices or their observed version based on the calculated Hessian matrix, although simplifications are possible, see for example [8] and [17] for a detailed discussion of the available approaches and comparisons in nonlinear measurement models.

A. Calibration according to the ISO Technical Specification 28037:2010

In metrology, the estimation of parameters a, b and the evaluation of the associated standard uncertainties and covariances is described in ISO Technical Specification 28037:2010 [5]. Recently, the ISO working group *Statistical Methods to Support the Evaluation of Measurement Uncertainty* (ISO /TC 69/ SC 6/ WG 7) has officially started the development of the new standard ISO 28037 *Determination and Use of Linear Calibration Functions*, which will potentially address known shortcomings of ISO/TS. In particular, the problem is that the ISO method does not correctly apply the LPU for the implicit measurement model according to GUM and its Supplements, which was recently pointed out in [17]. The ISO approach is based on a generalised (weighted) least squares estimation as described in (8)-(9) and on the application of the law of uncertainty propagation (LPU), albeit approximate, using a first order linearisation of the associated statistical model. Technical details on this approach can be found in section 10 of [5].

Since the ISO method and GUM use different modelling approaches to the calibration problem (implicit nonlinear measurement model in GUM versus statistical regression model with nonlinear constraints on the parameters in ISO/TS) and their subsequent linearisation, the associated uncertainty matrices of the parameter estimates are potentially different.

The numerical calculation of the uncertainty matrices could be a challenging task, especially the LPU approach, however, as the reviewer has noted, there are algorithms that accurately implement the calculation of uncertainties based on the ISO approach as well as the GUM LPU approach. The MATLAB version of the ISO algorithm was developed by the National Physical Laboratory (NPL), the UK's National Metrology Institute, as NPL's SOFT-WARE TO SUPPORT ISO/TS 28037:2010(E), see [30], and it is freely available at `https://www.npl.co.uk/resources/software/iso-ts-28037-2010e`.

Despite the above-mentioned shortcoming of the ISO method, it can be considered as a reasonable approximation method, as it is based on a specific linearised calibration model, resulting in a simpler calculation of the covariance matrix. Therefore, the main objective of this paper is to investigate and compare the statistical properties of the ISO estimators (including the coverage probabilities of the true parameters and the true values of the stimuli) in the case of typical calibration experiments.

As specified in [5], subclause 10.2.3 paragraph v), if the input quantities $\mathbf{X}$ and $\mathbf{Y}$ are random variables characterised by a multivariate normal distribution, then the probability distributions of the values that can be attributed to the parameters of the calibration function can be approximated by a (multivariate) normal distribution specified by the estimates and the associated uncertainties (uncertainty matrix), however, we remind that even in this case it is necessary to verify the adequacy of this approximation. In such cases, we can use approximate ISO 95% coverage intervals based on the use of the normal distribution approximation. However, in this paper we want to investigate the statistical properties of such intervals even in the presence of weak violations of the normality assumptions, as is the case in our calibration experiment.

It is recommended that assumptions about the normality of the distribution of input values and/or estimated parameters be checked (through formal statistical residual tests). This is in line with the reviewer's comment that the examination of residuals is an important means of validating a model and determining whether it is appropriate for the observed data. There are two types of residuals in the regression of errors in variables: horizontal residuals and vertical residuals. For a discussion of these residuals, see [12]. In case of a serious violation of the assumption about the normality of the distribution of the estimated parameters, it is recommended to use the alternative methods described in [1] or [2] and [3] to determine appropriate coverage intervals. By using a normal distribution approximation, the ISO approximate 95% coverage interval of values that could reasonably be attributed to the parameter a is given by

$$\mathrm{CI}_{a,0.95}^{(\mathrm{ISO})} = \left\langle \hat{a} - 1.96 \times u(\hat{a}); \hat{a} + 1.96 \times u(\hat{a}) \right\rangle, \quad (10)$$

the approximate 95% coverage interval of values that could

reasonably be attributed to the parameter b is given by

$$\mathrm{CI}_{b,0.95}^{(\mathrm{ISO})} = \left\langle \hat{b} - 1.96 \times u(\hat{b}); \hat{b} + 1.96 \times u(\hat{b}) \right\rangle. \quad (11)$$

The joint 95% coverage region of values that could reasonably be attributed to the vector of parameter (a,b) is given as an elliptical region such that the area under the joint PDF is equal to $p = 0.95$, see [3], clause 6.5.2.3.,

$$\mathrm{CR}_{\substack{(a,b) \\ 0.95}}^{(\mathrm{ISO})} = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} : \begin{pmatrix} \hat{a} - a \\ \hat{b} - b \end{pmatrix}' \mathbf{U}_{\hat{a},\hat{b}}^{-1} \begin{pmatrix} \hat{a} - a \\ \hat{b} - b \end{pmatrix} \leq \chi_{2,0.95}^2 \right\}. \quad (12)$$

Here, $\hat{a}$ and $\hat{b}$ are the estimates of the coefficients of the linear calibration function with the uncertainty matrix $\mathbf{U}_{\hat{a},\hat{b}}$ with its elements $u^2(\hat{a})$, $u^2(\hat{b})$, and $u(\hat{a},\hat{b})$ estimated by the ISO method [5], and $\chi_{2,0.95}^2$ is the 0.95-quantile of the chi-squared distribution with 2 degrees of freedom. In particular, we get $\chi_{2,0.95}^2 = 2.45$.

The information about the parameters of the calibration function is further used in the second calibration step to establish a relationship that allows the measurement result to be obtained from a new indication (by using the calibrated instrument). In doing so, we assume that x_0 is the estimate of the true (unknown) quantity value μ_0 measured and expressed in units of the calibrated instrument (less accurate instrument $\mathscr{X}$), with the associated uncertainty $u(x_0)$. In general, we assume that the uncertainty budget for this measurement result provides complete information in the form of a probability distribution.

According to [5], using direct prediction, we determine the estimate $\hat{v}_0$ of the true (unknown) stimulus v_0 to be

$$\hat{v}_0 = \hat{a} + \hat{b}\hat{\mu}_0, \quad (13)$$

where we use $\hat{\mu}_0 = x_0$ with $u(\hat{\mu}_0) = u(x_0)$. By applying LPU we set the associated (squared) uncertainty as

$$u^2(\hat{v}_0) = u^2(\hat{a}) + \hat{\mu}_0^2 u^2(\hat{b}) + 2\hat{\mu}_0 u(\hat{a},\hat{b}) + \hat{b}^2 u^2(\hat{\mu}_0). \quad (14)$$

Using the normal distribution approximation we specify the 95% coverage interval for the values that can be attributed to the (unknown) stimulus v_0 based on the new measurement obtained by using the calibrated instrument as follows,

$$\mathrm{CI}_{v_0,0.95}^{(\mathrm{ISO})} = \left\langle \hat{v}_0 - 1.96 \times u(\hat{v}_0); \hat{v}_0 + 1.96 \times u(\hat{v}_0) \right\rangle. \quad (15)$$

3. CALIBRATION OF THE PRESSURE TRANSDUCER USING A PRESSURE STANDARD ACCORDING TO THE ISO TECHNICAL SPECIFICATION 28037:2010

The present example focuses on the linear calibration of a pressure transducer (MERET TSZ) using a reference device as a pressure standard (YOKOGAWA 2655 electronic pressure gauge) with measurements at different pressure points. For each pressure point a measurement result of the standard is available together with a measurement result of the calibrated pressure transducer. The uncertainties in the calibration due to variations in the measurement data were determined by four repeated measurements with the standard and

Table 1. Measuring conditions, measuring devices and measured values in the calibration experiment of the pressure transducer carried out at the Slovak Institute of Metrology.

Measurement conditions:	Ambient temperature $(20 \pm 2)\,°C$;
	Atmospheric pressure (100 ± 5) kPa;
Measurement devices:	Calibrated instrument $\mathscr{X}$: Pressure transducer MERET TSZ (production number
	6526F06) with unified output $(4-20)$ mA, measuring range $(0-60)$ kPa;
	Reference instrument $\mathscr{Y}$: Electronic pressure gauge YOKOGAWA 2655 (production
	number 4DJ1019) with standard uncertainty of 0.006 kPa (in the range $(0-60)$ kPa)
	which is linked to the Slovak national pressure standard (standard pressure gauge SMU
	PTV-06, production number 045, certificate number 115/220/17/04);
	Digital multimeter connected to the calibrated device: standard uncertainty 0.0014 mA,
	(measuring range 20 mA);
Measured values:	Table 2. shows the measured values of the properties μ_i of the artefacts V_i obtained by
	direct measurements with the calibrated instrument $\mathscr{X}$ and the measured values of the
	properties ν_i of the artefacts V_i obtained by direct measurements with the reference
	instrument $\mathscr{Y}$.

the calibrated pressure transducer at a single pressure point (30 kPa) and are considered as known values for simplicity. As the reviewers noted, this is an oversimplified approach that can lead to underestimating the uncertainties of the output quantities. This simplification may be approximately true if there is indeed long experience with this particular transducer, and this experience supports the assumption that such uncertainty is constant across the whole range of relevant pressures. A correct alternative approach (not considered here) would be to consider the squares of the estimated measurement uncertainties (at each of the calibration points considered) as realisations of random variables whose distribution is proportional to the chi-squared distribution with three degrees of freedom.

The calibration of the pressure transducer was done by directly comparing the output data of the calibrated pressure transducer and the output data of the reference pressure transducer at the time when both instruments to be compared are subjected to a sufficiently constant equilibrium pressure.

Calibration of the pressure transducer was performed at seven pressure points by gradually increasing the pressure and then at six points by gradually decreasing the pressure. The pressure points were evenly distributed over the entire measurement range, including the lower and upper limits of the measurement range. At each pressure point, the measured pressure was kept at a constant value.

A. Measurement data and specifications

The details of the measurement conditions, the measuring instruments and the measured values can be found in Table 1. and Table 2. Based on an expert assessment of the influencing factors, we consider the following statistical model for the responses of the measuring instrument $\mathscr{X}$ (the calibrated device):

$$X_i = \mu_i + \varepsilon_{X,i} + \Delta_X, \quad i = 1, 2, \dots, 13, \quad (16)$$

where

- $\varepsilon_{X,i} \sim N(0, u_x^2)$, for $i = 1, 2, \dots, 13$, are the random variables representing our knowledge about the measurement errors (data fluctuation) in measurements X_i, having independent normal distributions with zero-mean and known dispersion $u_x^2 = 0.00001444\,(\text{mA})^2$ (i.e. $u_x = 0.0038$ mA).

- Δ_X is the random variable representing our knowledge about the systematic error of the multimeter (when measuring with the calibrated device), with zero-mean distribution; $\Delta_X \sim R(0, u_{\Delta_X})$, where $R(0, u_{\Delta_X})$ means rectangular distribution with mean 0 and standard deviation $u_{\Delta_X} = 0.0014$ mA (i.e. rectangular distribution over the interval $\langle -\sqrt{3}u_{\Delta_X}; \sqrt{3}u_{\Delta_X} \rangle = \langle -0.002424; 0.002424 \rangle$),

- ε_X and Δ_X are mutually independent random variables.

Similarly, the considered model of the responses of the measuring instrument $\mathscr{Y}$ (the standard) is given by

$$Y_i = \nu_i + \varepsilon_{Y,i} + \Delta_Y, \quad i = 1, 2, \dots, 13, \quad (17)$$

where

- $\varepsilon_{Y,i} \sim N(0, u_y^2)$, for $i = 1, 2, \dots, 13$, are the random variables representing our knowledge about the measurement errors (data fluctuation) in measurements Y_i, having independent normal distributions with zero-mean and known dispersion $u_y^2 = 0.000036\,(\text{kPa})^2$, (i.e. $u_y = 0.006$ kPa).

- Δ_Y is the random variable representing our knowledge about the systematic error of the reference device, with zero-mean distribution; $\Delta_Y \sim N(0, u_{\delta_Y}^2)$, where $u_{\delta_Y}^2 = 0.000036\,(\text{kPa})^2$, (i.e. $u_{\delta_Y} = 0.006$ kPa),

- ε_Y and Δ_Y are mutually independent random variables. Moreover, we shall also assume that ε_X, Δ_X, ε_Y, and Δ_Y are mutually independent.

Table 2. Measured values observed during the calibration experiment of the pressure transducer.

Calibration point i	x_i (mA) (calibrated device)	y_i (kPa) (reference device)
1	4.0030	0.0000
2	6.7160	10.191
3	9.3710	20.102
4	12.053	30.170
5	15.266	42.230
6	17.351	50.050
7	20.036	60.070
8	17.369	50.080
9	14.718	40.115
10	12.039	30.089
11	9.3760	20.095
12	6.6970	10.070
13	4.0080	0.0000

Based on the reviewers' comments, we also present the following possible alternative specifications for measurement errors, which we did not consider further for our analysis:

If u_x^2 is the same for all x_i and is considered to be the realisation of a random variable with a probability distribution proportional to a chi-squared random variable with $k = 3$ degrees of freedom, then we should adjust our knowledge of measurement error by assuming that approximately $\varepsilon_X = (\varepsilon_{X,1}, \ldots, \varepsilon_{X,13})' \sim t_k(\mathbf{0}, u_x^2\mathbf{I})$ (has a multivariate Student t-distribution with $k = 3$ degrees of freedom).

If $u(x_i)^2$ are possibly different estimates, considered as realisations of independent random variables proportional to chi-squared distributions with $k_i = 3$ degrees of freedom, then we should adjust our knowledge of measurement error by assuming that approximately $\varepsilon_{X,i} \sim t_{k_i}(0, u(x_i)^2)$ (Student t-distributions with $k_i = 3$ degrees of freedom),

Similarly, if u_y^2 is the same for all y_i and is considered to be the realisation of a random variable with a probability distribution proportional to a chi-squared random variable with $l = 3$ degrees of freedom, then we should adjust our knowledge of measurement error by assuming that approximately $\varepsilon_Y = (\varepsilon_{Y,1}, \ldots, \varepsilon_{Y,13})' \sim t_l(\mathbf{0}, u_y^2\mathbf{I})$.

If $u(y_i)^2$ are possibly different estimates, considered as realisations of independent random variables proportional to chi-squared distributions with $l_i = 3$ degrees of freedom, then we should adjust our knowledge of measurement error by assuming that approximately $\varepsilon_{Y,i} \sim t_{l_i}(0, u(x_i)^2)$,

B. The calibration model

The calibration model in a matrix form is specified as

$$\mathbf{X} = \mu + \varepsilon_X + \Delta_X\mathbf{1}, \tag{18}$$

$$\mathbf{Y} = \nu + \varepsilon_Y + \Delta_Y\mathbf{1}, \tag{19}$$

$$\nu = a\mathbf{1} + b\mu, \tag{20}$$

where $\varepsilon_X = (\varepsilon_{X,1}, \varepsilon_{X,2}, \ldots, \varepsilon_{X,13})'$, $\varepsilon_Y = (\varepsilon_{Y,1}, \varepsilon_{Y,2}, \ldots, \varepsilon_{Y,13})'$, and $\mathbf{1}$ denotes the (13×1)-dimensional column vector of ones.

The expectation (mean vector) is specified by

$$E\begin{pmatrix}\mathbf{X} \\ \mathbf{Y}\end{pmatrix} = \begin{pmatrix}\mu \\ \nu\end{pmatrix}, \tag{21}$$

and the blocks of the joint covariance matrix $\mathbf{U}$ are specified by

$$
\begin{aligned}
Cov(\mathbf{X}) = \mathbf{U}_x &= u_x^2\mathbf{I} + u_{\Delta_X}^2\mathbf{J} \\
&= 0.00001444 \times \mathbf{I} + 0.00000196 \times \mathbf{J}, \\
Cov(\mathbf{Y}) = \mathbf{U}_y &= u_y^2\mathbf{I} + u_{\Delta_Y}^2\mathbf{J} \\
&= 0.000036 \times \mathbf{I} + 0.000036 \times \mathbf{J}, \\
Cov(\mathbf{X}, \mathbf{Y}) = \mathbf{U}_{x,y} &= \mathbf{0},
\end{aligned}
$$

where $\mathbf{I}$ denotes the (13×13)-dimensional identity matrix and $\mathbf{J} = \mathbf{1}\mathbf{1}'$ denotes the (13×13)-dimensional matrix of ones. Thus,

$$Cov\begin{pmatrix}\mathbf{X} \\ \mathbf{Y}\end{pmatrix} = \begin{pmatrix}\mathbf{U}_x & \mathbf{0} \\ \mathbf{0}' & \mathbf{U}_y\end{pmatrix} = \mathbf{U}. \tag{22}$$

The realizations of $\mathbf{X}$ and $\mathbf{Y}$, say $\mathbf{x} = (x_1, \ldots, x_m)'$ and $\mathbf{y} = (y_1, \ldots, y_m)'$ specified in Table 2., represent the estimates of the properties $\mu = (\mu_1, \ldots, \mu_m)'$ (based on direct measurements with the calibrated device $\mathscr{X}$) and the estimates of the properties $\nu = (\nu_1, \ldots, \nu_m)'$ (based on direct measurements with the reference device $\mathscr{Y}$).

C. Estimation of the linear calibration function parameters and evaluation of the measurements with the calibrated device

Given the calibration model (18)-(20), the measurement data represented by the estimates $\mathbf{x}$ and $\mathbf{y}$, specified in Table 2., and the uncertainty matrix $\mathbf{U}$ associated with the measurement data, specified in (22), then the estimates of the coefficients of the linear calibration function, say $\hat{a}$ and $\hat{b}$, and the associated approximate uncertainty matrix $U_{\hat{a},\hat{b}}$, with its
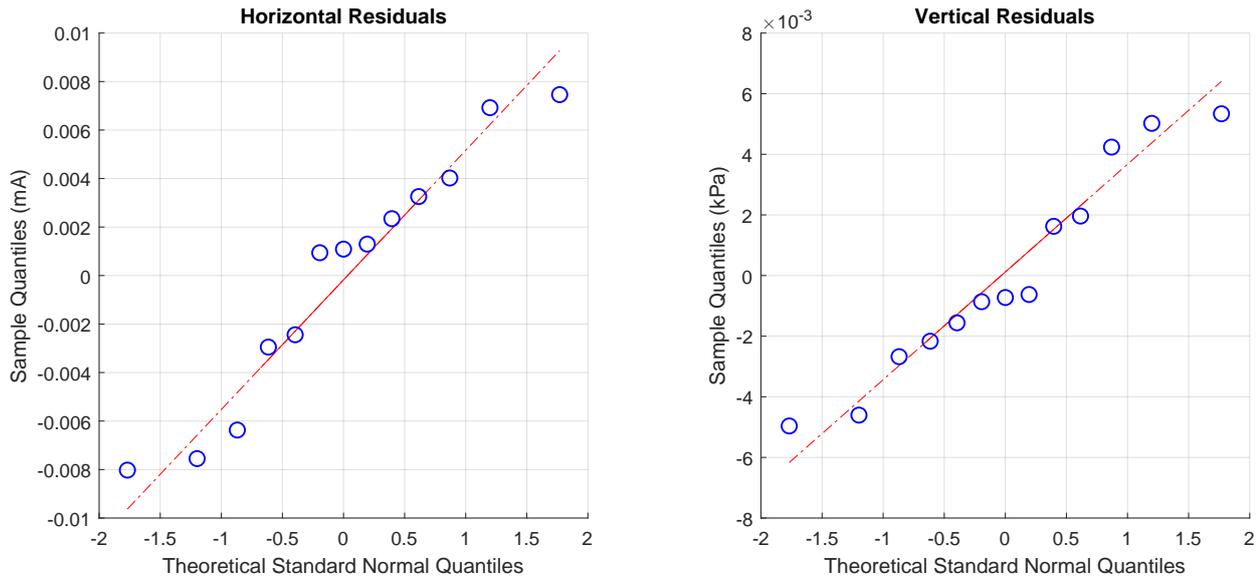
Fig. 1. QQ-plots of horizontal residuals, $\mathbf{x} - \hat{\mu}$, and of vertical residuals, $\mathbf{y} - \hat{v}$. In accordance with [12], both presented QQ-plots do not contradict the hypothesis of the normality of the distribution of the input variables and lend support to the adequacy of the model to the data.

Table 3. Uncertainty budget related to the direct measurement of the new indication μ_0 by using the calibrated device. Here, we consider the measurement equation $\tilde{\mu}_0 = x_0 + \varepsilon_{X_0} + \Delta_{X_0}$, where x_0 is the observed value from the calibrated device and ε_{X_0} and Δ_{X_0} are the considered corrections.

Quantity	Estimate (mA)	Standard uncertainty (mA)	Probability distribution	Sensitivity coefficient	Coefficient of contribution (%)
x_0	7.4970	0	Dirac	1	0%
ε_{X_0}	0	0.0038	Normal	1	6%
Δ_{X_0}	0	0.0150	Rectangular	1	94%
New indication	7.4970	0.0155	Convolution		100%

elements given by $u^2(\hat{a})$, $u^2(\hat{b})$, and $u(\hat{a}, \hat{b})$, are given by solving the generalised Gauss-Markov regression problem (8) as specified in [5]. The adequacy of the model to the data was checked using the QQ-plot, a graphical technique for determining if two data sets come from populations with a common distribution, as suggested in [12], see Fig.1. In particular, the estimates $\hat{a}$ and $\hat{b}$ of the straight-line calibration function parameters a and b and their approximate uncertainty matrix $U_{\hat{a},\hat{b}}$ are calculated by using the MATLAB algorithms from [30].

The estimates $\hat{a}$ and $\hat{b}$ of the straight-line calibration function parameters a and b and their uncertainty matrix $U_{\hat{a},\hat{b}}$ are given by

$$\hat{a} = -15.0167 \text{ kPa}, \quad \hat{b} = 3.7481 \text{ kPa/mA}, \quad (23)$$

$$\mathbf{U}_{\hat{a},\hat{b}} = \begin{pmatrix} 1.7586 \times 10^{-4} & -8.1970 \times 10^{-6} \\ -8.1970 \times 10^{-6} & 7.1516 \times 10^{-7} \end{pmatrix}. \quad (24)$$

Hence, the measurement result for the intercept of the calibration line is specified by the estimate $\hat{a} = -15.0167$ kPa

with the associated uncertainty $u(\hat{a}) = 0.0133$ kPa (the expanded uncertainty with the coverage factor $k = 2$ is $U(\hat{a}) = 0.0265$ kPa) and the slope of the calibration line is specified by the estimate $\hat{b} = 3.7481$ kPa/mA with the associated uncertainty $u(\hat{b}) = 10^{-4} \times 8.4567$ kPa/mA (the expanded uncertainty with the coverage factor $k = 2$ is $U(\hat{b}) = 0.0017$ kPa/mA). By using a normal distribution approximation, the approximate 95% coverage interval of values that could reasonably be attributed to the parameter a is given by (10),

$$\text{CI}_{a,0.95}^{(\text{ISO})} = \langle -15.0427; -14.9907 \rangle, \quad (25)$$

and the approximate 95% coverage interval of values that could reasonably be attributed to the parameter b is given by (11),

$$\text{CI}_{b,0.95}^{(\text{ISO})} = \langle 3.7464; 3.7498 \rangle. \quad (26)$$

Similarly, using (23), (24) and (12), we obtain the joint (approximate) 95% coverage region of the values that could reasonably be attributed to the vector of parameter (a, b) as
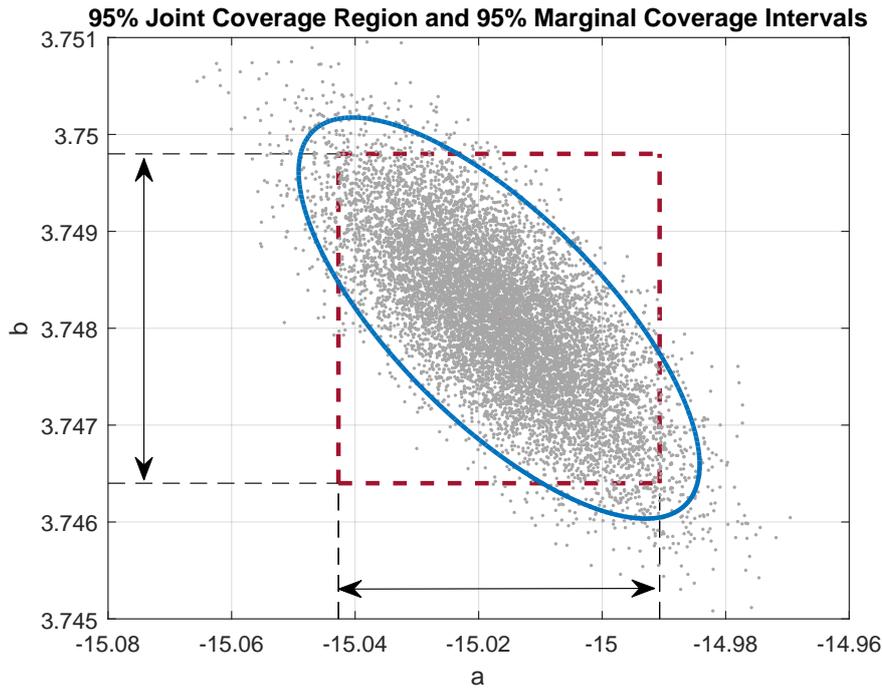
Fig.2. Plot of the joint 95% coverage region (12) (blue solid line) of the values that could reasonably be attributed to the vector of calibration line parameters (a,b) calculated from the estimates (23) and (24) based on ISO/TS 28037:2010, plotted together with the 95% coverage interval of the values that could reasonably be attributed to parameter a, given in (25) and with the 95% coverage interval of values that could reasonably be attributed to parameter b, given in (26) (the coverage intervals are specified by the sides of the rectangle with red dashed lines). The grey dots represent $N = 10000$ estimates $(\hat{a}, \hat{b})$ of the calibration line parameters (a,b) from the Monte Carlo simulations explained in section 4.

an elliptical region, so that the area under the joint PDF (approximated by the fitted bivariate normal PDF) is equal to $p = 0.95$, see Fig.2.

This information is further used to obtain the measurement result from a new indication estimated by using the calibrated pressure transducer (connected to a different multimeter). For illustration, let us assume that the estimate of a new indication obtained by using the calibration device is $x_0 = 7.4970$ mA with its combined uncertainty (derived based on an expert assessment of all influencing factors) given by $u(x_0) = 0.0155$ mA. In addition, for this measurement result, we have full information in the form of a detailed uncertainty budget, which also specifies the type of the state-of-knowledge distributions of the influencing quantities, see Table 3. Then, using direct prediction (13) and application of LPU (14) we stipulate the estimate $\hat{v}_0$ of the true (unknown) stimulus v_0 as

$$\hat{v}_0 = 13.0829\,\text{kPa}, \quad u(\hat{v}_0) = 0.0588. \tag{27}$$

Hence, the expanded uncertainty is given as $U(\hat{v}_0) = 2u(\hat{v}_0) = 0.1176$. Further, using (15), the 95% coverage interval of values that can be attributed to the unknown stimulus v_0 is given by

$$\text{CI}_{v_0,0.95}^{(\text{ISO})} = \langle 12.9676; 13.1981 \rangle. \tag{28}$$

4. MONTE CARLO METHOD FOR DETERMINING THE PROBABILITY DISTRIBUTION OF THE PARAMETERS OF THE CALIBRATION FUNCTION AND THE NEW STIMULUS

Using the calibration example above, we determine here the distribution (probability density function) of the values that could be attributed to the calibration function parameters a and b and the (unknown) stimulus v_0 (based on the measurement result obtained with the calibrated instrument), by using the MCM proposed in [2] and [3] and comparing it with the measurement results (23-28) based on the ISO approach proposed in [5].

For each $i = 1, 2, \ldots, N$ (with $N = 10000$) we have generated new data set, the estimates $\mathbf{x}^{(i)} = (x_1^{(i)}, \ldots, x_m^{(i)})'$ and $\mathbf{y}^{(i)} = (y_1^{(i)}, \ldots, y_m^{(i)})'$, with the common uncertainty matrix $\mathbf{U}$, specified in (22), where $\mathbf{x}^{(i)}$ and $\mathbf{y}^{(i)}$ are realizations of the random vectors $\mathbf{X}^{(i)}$ and $\mathbf{Y}^{(i)}$, generated by

$$\mathbf{X}^{(i)} = \mathbf{x} + \varepsilon_X^{(i)} + \Delta_X^{(i)}\mathbf{1}, \tag{29}$$

$$\mathbf{Y}^{(i)} = \mathbf{y} + \varepsilon_Y^{(i)} + \Delta_Y^{(i)}\mathbf{1}, \tag{30}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the measured values observed during the calibration experiment of the pressure transducer.
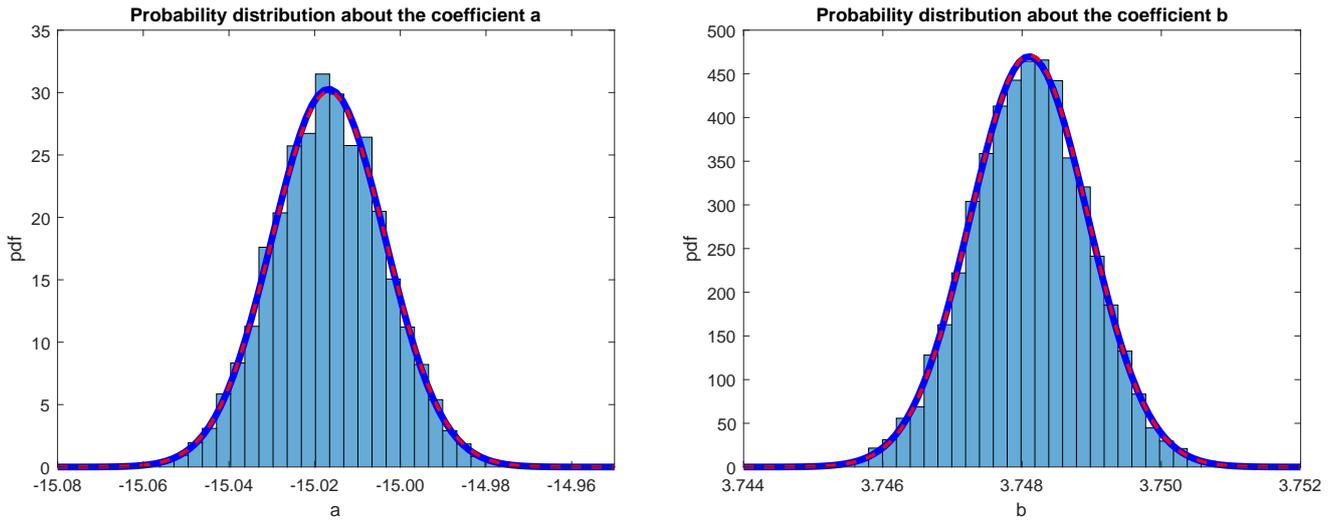
Fig. 3. Probability distribution over the parameters of the calibration function, a and b based on the Monte Carlo method and on ISO/TS 28037:2010. The plots represent histograms from Monte Carlo simulations together with the approximate normal fit (blue solid line) and the approximate normal fit based on ISO/TS 28037:2010 (dashed red line).

- $\varepsilon_X{}^{(i)}$ are mutually independent random vectors representing our knowledge about the measurement errors (data fluctuation) in measurements with the calibrated device, $\varepsilon_X{}^{(i)} \sim N(\mathbf{0}, u_x^2 \mathbf{I})$, where $u_x^2 = 0.00001444\,(\text{mA})^2$, (i.e. $u_x = 0.0038\,\text{mA}$),

- $\Delta_X^{(i)}$ are mutually independent random variables representing our knowledge about the systematic error of the multimeter connected to the calibrated device, $\Delta_X \sim R(0, u_{\Delta_X})$, where $u_{\Delta_X} = 0.0014\,\text{mA}$,

- $\varepsilon_Y{}^{(i)}$ are mutually independent random vectors representing our knowledge about the measurement errors (data fluctuation) in measurements with the reference device, $\varepsilon_Y{}^{(i)} \sim N(\mathbf{0}, u_y^2 \mathbf{I})$, where $u_y^2 = 0.000036\,(\text{kPa})^2$, (i.e. $u_y = 0.006\,\text{kPa}$),

- $\Delta_Y^{(i)}$ are mutually independent random variables representing our knowledge about the systematic error of the multimeter connected to the reference device, $\Delta_Y^{(i)} \sim N(0, u_{\delta_Y}^2)$, where $u_{\delta_Y}^2 = 0.000036\,(\text{kPa})^2$, ($u_{\delta_Y} = 0.006\,\text{kPa}$).

The empirical distributions about the calibration function parameters a and b, their estimates, the associated uncertainties and the related 95% coverage intervals, were derived from the estimates $(\hat{a}^{(i)}, \hat{b}^{(i)})$, $i = 1, \ldots, N$, computed from the data generated by the Monte Carlo method.

In an earlier version of this paper, we generated new samples based on WTLS estimates fitted from the original data, $\hat{\mu}$ and $\hat{\nu}$. However, as one reviewer pointed out, this method does not comply with JCGM 101:2008 [2] and JCGM 102:2011 [3]. We fully respect this expert opinion and have therefore recalculated all simulations around the observed values and found that the differences in the presented results are minimal.

A. Probability distributions over the parameters of the calibration function a and b based on the Monte Carlo method

Here we present the derived probability distributions of values that could be reasonably attributed to the calibration function parameters a and b, based on MCM combined with the ISO estimation approach.

The left panel in Fig.3. plots the histogram and the normal fit of the state-of-knowledge distribution about the parameter value a based on the Monte Carlo method (solid blue line), with the parameters $\hat{a}^{(\text{MCM})} = -15.0167$ ($\hat{a}^{(\text{MCM})} = \frac{1}{N}\sum_{i=1}^{N} \hat{a}^{(i)}$) and $u(\hat{a}^{(\text{MCM})}) = 0.0132$ ($u(\hat{a}^{(\text{MCM})}) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\hat{a}^{(i)} - \hat{a}^{(\text{MCM})})^2}$), plotted together with the normal fit based on ISO/TS 28037:2010 (dashed red line), with the parameters $\hat{a}^{(\text{ISO})} = -15.0167$ and $u(\hat{a}^{(\text{ISO})}) = 0.0133$.

The approximate 95% coverage interval of values that could reasonably be attributed to the parameter a (based on using the Monte Carlo empirical quantiles[1]), is given by

$$\text{CI}_{a,0.95}^{(\text{MCM})} = \left\langle \hat{a}_{0.025}^{(\text{MCM})}; \hat{a}_{0.975}^{(\text{MCM})} \right\rangle$$
$$= \langle -15.0424; -14.9909 \rangle. \qquad (31)$$

The interval (31) is close to the approximate 95% coverage interval (25) based on the ISO/TS 28037:2010.

The right panel in Fig.3. plots the histogram and the normal fit of the state-of-knowledge distribution about the parameter value b based on the Monte Carlo method (solid blue line), with the estimated parameters $\hat{b}^{(\text{MCM})} = 3.7481$ and

---

[1] If $\hat{a}_{(1)}, \ldots, \hat{a}_{(N)}$ are the ordered Monte Carlo estimates of the parameter a, here we define the required empirical quantiles by $\hat{a}_{0.025}^{(\text{MCM})} = \hat{a}_{(\lfloor N \times 0.025 \rfloor)}$ and $\hat{a}_{0.975}^{(\text{MCM})} = \hat{a}_{(\lceil N \times 0.025 \rceil)}$, where $\lfloor \cdot \rfloor$ rounds the element to the nearest integer less than or equal to that element and $\lceil \cdot \rceil$ rounds the element to the nearest integer greater than or equal to that element.
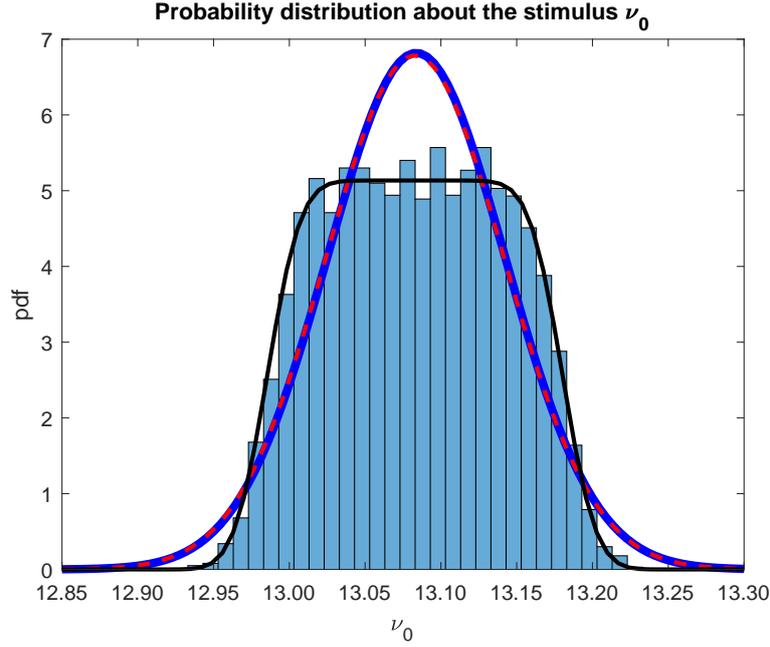
Fig. 4. The probability distribution of the values that could be attributed to the stimulus value v_0 based on the Monte Carlo method and on ISO/TS 28037:2010. The graph shows the histogram from the Monte Carlo simulations together with the approximate normal fit (blue solid line) and the approximate normal fit based on ISO/TS 28037:2010 (dashed red line). The solid black line shows the PDF derived by the characteristic function approach.

$u(\hat{b}^{(\text{MCM})}) = 10^{-4} \times 8.4896$, plotted together with the normal fit based on ISO/TS 28037:2010 (dashed red line), with the estimated parameters $\hat{b}^{(\text{ISO})} = 3.7481$ and $u(\hat{b}^{(\text{ISO})}) = 10^{-4} \times 8.4567$.

The approximate 95% coverage interval of values that could reasonably be attributed to the parameter b (based on using the Monte Carlo empirical quantiles) is given by

$$\text{CI}_{b,0.95}^{(\text{MCM})} = \left\langle \hat{b}_{0.025}^{(\text{MCM})}; \hat{b}_{0.975}^{(\text{MCM})} \right\rangle = \langle 3.7464; 3.7497 \rangle. \quad (32)$$

The interval (32) is perfectly close to the approximate 95% coverage interval (26) based on the ISO/TS 28037:2010.

B. Monte Carlo method for measurement uncertainty of the result obtained with a calibrated instrument

For each simulation step $i = 1, 2, \ldots, N$ we also generated a realisation $x_0^{(i)}$ of the indication μ_0 measured with the calibrated instrument under the following assumptions,

- $x_0^{(i)}$ are realisations of mutually independent random variables $X_0^{(i)}$ representing our knowledge about the new indication μ_0 measured with the calibrated instrument, $X_0^{(i)} \sim \hat{\mu}_0 + \varepsilon_{X_0}^{(i)} + \Delta_{X_0}^{(i)}$, where $\hat{\mu}_0 = x_0$ is the observed (measured) value from the calibrated instrument as specified in Table 3. and the random variables ε_{X_0} and Δ_{X_0} represent the imposed corrections (due to the considered measurement errors and the systematic error of the currently used multimeter). They are mutually independent random variables whose distributions are given in Table 3. In particular, $\varepsilon_{X_0}^{(i)} \sim N(\mathbf{0}, u_x^2)$,

where $u_x^2 = 0.00001444 \ (\text{mA})^2 \ (u_x = 0.0038 \ \text{mA})$, and $\Delta_{X_0}^{(i)} \sim R(0, u_{\Delta_{X_0}})$, where $u_{\Delta_{X_0}} = 0.015 \ \text{mA}$.

Then, for each $i = 1, 2, \ldots, N$ we calculated also the estimated value of the new stimulus $\hat{v}_0^{(i)}$,

$$\hat{v}_0^{(i)} = \hat{a}^{(i)} + \hat{b}^{(i)} \hat{\mu}_0^{(i)}, \quad (33)$$

where we use $\hat{\mu}_0^{(i)} = x_0^{(i)}$ to emphasise that the observed value $x_0^{(i)}$ is our estimate of the new indication μ_0 measured by the calibrated instrument.

Fig. 4. plots the histogram and the normal fit of the probability distribution about the stimulus value v_0 based on the Monte Carlo method (solid blue line), with the estimated parameters $\hat{v}_0^{(\text{MCM})} = 13.0837$ and $u(\hat{v}_0^{(\text{MCM})}) = 0.0585$, plotted together with the normal fit based on ISO/TS 28037:2010 (dashed red line), with the estimated parameters $\hat{v}_0^{(\text{ISO})} = 13.0829$ and $u(\hat{v}_0^{(\text{ISO})}) = 0.0588$.

Finally, the approximate 95% coverage interval of values that could reasonably be attributed to the new stimulus v_0, based on using the Monte Carlo empirical quantiles, is given by

$$\text{CI}_{v_0,0.95}^{(\text{MCM})} = \left\langle \hat{v}_{0,0.025}^{(\text{MCM})}; \hat{v}_{0,0.975}^{(\text{MCM})} \right\rangle$$
$$= \langle 12.9812; 13.1855 \rangle. \quad (34)$$

The interval (34) is narrower than (28), with visible differences between the histogram and the approximate (normal fit) distribution.

Table 4. Empirical coverage probabilities (the observed relative frequencies of inclusion of the true parameter value) of the considered approximate 95% coverage intervals of the values that could reasonably be attributed to parameters a, b and the new stimulus v_0, calculated for $m = 5$ calibration points and different combinations of the input parameters, namely the intercept $a = 0$ the slope $b = 1$, the combined uncertainty of the x measurements $u_c(x) \in \{0.002, 0.022, 0.224\}$ and the combined uncertainty of the y measurements $u_c(y) \in \{0.002, 0.022, 0.224\}$, using ISO/TS 28037:2010 (ISO) and the Monte Carlo method (MCM). The probabilities shown in bold deviate significantly from the stated nominal significance level of 0.95.

m	$u_c(x)$	$u_c(y)$	$\text{CI}_{a,0.95}^{(\text{ISO})}$	$\text{CI}_{a,0.95}^{(\text{MCM})}$	$\text{CI}_{b,0.95}^{(\text{ISO})}$	$\text{CI}_{b,0.95}^{(\text{MCM})}$	$\text{CI}_{v_0,0.95}^{(\text{ISO})}$	$\text{CI}_{v_0,0.95}^{(\text{MCM})}$
		0.002	**0.961**	0.948	0.945	**0.943**	0.953	0.946
	0.002	0.022	**0.961**	0.953	0.952	0.952	**0.970**	0.950
		0.224	**0.957**	0.948	**0.958**	**0.957**	**0.974**	0.952
		0.002	**0.981**	0.950	0.951	0.951	**0.961**	0.954
5	0.022	0.022	**0.962**	0.954	0.948	0.947	0.953	0.949
		0.224	**0.960**	0.950	0.949	0.950	**0.973**	**0.957**
		0.002	**0.984**	0.951	0.949	0.952	**0.964**	0.951
	0.224	0.022	**0.982**	0.948	0.948	0.948	**0.968**	0.953
		0.224	**0.964**	0.948	0.950	0.946	**0.967**	0.950
		0.002	**0.968**	0.955	0.944	0.944	**0.959**	0.953
	0.002	0.022	**0.967**	0.953	0.950	0.950	**0.973**	0.950
		0.224	**0.965**	0.950	0.949	0.948	**0.983**	0.951
		0.002	**0.984**	0.950	0.950	0.952	**0.956**	0.950
9	0.022	0.022	**0.960**	0.947	0.948	0.948	0.955	0.951
		0.224	**0.968**	0.956	0.953	0.954	**0.976**	0.953
		0.002	**0.989**	0.947	**0.943**	0.945	**0.957**	0.945
	0.224	0.022	**0.989**	0.953	0.955	0.952	**0.962**	0.951
		0.224	**0.962**	0.948	**0.943**	0.956	**0.961**	0.945
		0.002	**0.967**	0.950	0.949	0.949	**0.959**	0.953
	0.002	0.022	**0.972**	0.952	0.953	0.955	**0.975**	0.948
		0.224	**0.971**	0.951	0.945	0.945	**0.990**	0.952
		0.002	**0.988**	0.952	0.949	0.947	**0.960**	0.953
13	0.022	0.022	**0.964**	0.947	0.951	0.952	**0.961**	0.956
		0.224	**0.970**	0.951	0.950	0.951	**0.979**	0.950
		0.002	**0.993**	0.951	0.954	0.950	**0.960**	0.949
	0.224	0.022	**0.992**	0.954	0.954	**0.957**	**0.959**	0.948
		0.224	**0.965**	0.948	**0.944**	**0.962**	0.956	0.944

5. SIMULATION STUDY

Inspired by the linear comparative calibration of a pressure transducer with a pressure standard presented above, we conducted a simulation study in which we investigated the empirical coverage probabilities (i.e. the frequencies of coverage of the true parameters) of ISO and MCM coverage intervals and also examined the influence of the sample size and the effect of different measurement uncertainties of the input variables.

For each $i = 1, 2, \ldots, N$ (with $N = 5000$) and each $m \in \{5, 9, 13\}$ we have set the true values of the calibration line parameters as $a = 0$, and $b = 1$ and generated new measurements $\mathbf{x}^{(i)} = (x_1^{(i)}, \ldots, x_m^{(i)})'$ and $\mathbf{y}^{(i)} = (y_1^{(i)}, \ldots, y_m^{(i)})'$ as realisations of the random variables $\mathbf{X}^{(i)}$ and $\mathbf{Y}^{(i)}$,

$$\mathbf{X}^{(i)} = \mu + \varepsilon_X^{(i)} + \Delta_X^{(i)} \mathbf{1}, \qquad (35)$$

$$\mathbf{Y}^{(i)} = v + \varepsilon_Y^{(i)} + \Delta_Y^{(i)} \mathbf{1}, \qquad (36)$$

where $v = a\mathbf{1} + b\mu$, with the common uncertainty matrix $\mathbf{U}$ given in (22), together with $x_0^{(i)}$, a new indication measured by the calibrated instrument and assumed to be a realisation of the random variable $X_0^{(i)}$,

$$X_0^{(i)} = \mu_0 + \varepsilon_{X_0}^{(i)} + \Delta_{X_0}^{(i)}, \qquad (37)$$

were we set

- $\mu = (\mu_1, \ldots, \mu_m)'$ as an m-dimensional vector of equidistant values in the interval $(0, 1)$ and $\mu_0 = 0.5$.

- $\varepsilon_X^{(i)}$ and $\varepsilon_{X_0}^{(i)}$ as mutually independent random variables representing the measurement errors of the calibrated instrument, $\varepsilon_X^{(i)} \sim N(\mathbf{0}, u_{A,x}^2 \mathbf{I})$ and $\varepsilon_{X_0}^{(i)} \sim N(0, u_{A,x}^2)$, where $u_{A,x} \in \{0.001, 0.01, 0.1\}$,

- $\Delta_X^{(i)}$ and $\Delta_{X_0}^{(i)}$ as a mutually independent random variables representing the systematic errors of the calibrated instrument during its calibration and during measurement of the new indication, $\Delta_X \sim R(0, u_{B,x})$ and $\Delta_{X_0} \sim R(0, u_{B,x})$, where $u_{B,x} \in \{0.002, 0.02, 0.2\}$,

- $\varepsilon_Y^{(i)}$ as mutually independent random vectors representing the measurement errors of the reference instrument, $\varepsilon_Y^{(i)} \sim N(\mathbf{0}, u_{A,y}^2 \mathbf{I})$, where $u_{A,y} \in \{0.001, 0.01, 0.1\}$,

- $\Delta_Y^{(i)}$ as a mutually independent random variables representing the systematic error of the reference instrument, $\Delta_Y \sim R(0, u_{B,y})$, where $u_{B,y} \in \{0.002, 0.02, 0.2\}$.

On this basis, we further calculated the combined uncertainties $u_c(x) = \sqrt{u_{A,x}^2 + u_{B,x}^2} \in \{0.002, 0.022, 0.224\}$ and $u_c(y) = \sqrt{u_{A,y}^2 + u_{B,y}^2} \in \{0.002, 0.022, 0.224\}$.

For each $\mathbf{x}^{(i)}$ and $\mathbf{y}^{(i)}$ we fitted the calibration function $v = a + b\mu$ by estimating the parameters $\hat{a}$, $\hat{b}$ and $\hat{\mu}$ (together with $\hat{v} = \hat{a}\mathbf{1} + \hat{b}\hat{\mu}$), and evaluated the corresponding 95% coverage intervals: $\mathrm{CI}_{a,0.95}^{(\mathrm{ISO})}$ as given in (10), $\mathrm{CI}_{b,0.95}^{(\mathrm{ISO})}$ as given in (11), and $\mathrm{CI}_{v_0,0.95}^{(\mathrm{ISO})}$ as given in (15), using the approach and algorithms as given in [5].

Moreover, for each $i = 1, 2, \ldots, N$ (with $N = 5000$), we applied the Monte Carlo method, and for each $j = 1, 2, \ldots, M$ (with $M = 2500$) generated another new set of measurements, $\mathbf{x}^{(j)} = (x_1^{(j)}, \ldots, x_m^{(j)})'$ and $\mathbf{y}^{(j)} = (y_1^{(j)}, \ldots, y_m^{(j)})'$, as realisations of the random variables $\mathbf{X}^{(j)}$ and $\mathbf{Y}^{(j)}$,

$$\mathbf{X}^{(j)} = \mathbf{x}^{(i)} + \varepsilon_X^{(j)} + \Delta_X^{(j)}\mathbf{1}, \qquad (38)$$

$$\mathbf{Y}^{(j)} = \mathbf{y}^{(i)} + \varepsilon_Y^{(j)} + \Delta_Y^{(j)}\mathbf{1}, \qquad (39)$$

together with $x_0^{(j)}$, a new indication by the calibrated instrument generated as a realisation of the random variable $X_0^{(j)}$,

$$X_0^{(j)} = x_0^{(i)} + \varepsilon_{X_0}^{(j)} + \Delta_{X_0}^{(j)}, \qquad (40)$$

where $\varepsilon_X^{(j)}$, $\Delta_X^{(j)}$, $\varepsilon_Y^{(j)}$, $\Delta_Y^{(j)}$, $\varepsilon_{X_0}^{(j)}$, and $\Delta_{X_0}^{(j)}$ are distributed as above.

Then, for each $\mathbf{x}^{(j)}$ and $\mathbf{y}^{(j)}$ we fitted the calibration function and based on that evaluated the corresponding 95% MCM coverage intervals: $\mathrm{CI}_{a,0.95}^{(\mathrm{MCM})}$ as given in (31), $\mathrm{CI}_{b,0.95}^{(\mathrm{MCM})}$ as given in (32), and $\mathrm{CI}_{v_0,0.95}^{(\mathrm{MCM})}$ as given in (34).

Table 4. show the empirical coverage probabilities (the observed relative frequencies of inclusion of the true parameter values) of the considered approximate 95% coverage intervals that could reasonably be attributed to the parameters a, b and the new stimulus v_0, calculated for $m = 5$, $m = 9$ and $m = 13$ calibration points and different combinations of the input parameters based on ISO/TS 28037:2010 and the Monte Carlo method. The probabilities in bold deviate significantly (in a statistical sense, based on the use of a standard normal approximation) from the nominal significance level of 0.95. Here, we interpret the stated empirical coverage probabilities as estimates of the parameter p of the binomial distribution $Bino(N, p)$, where N represents the size of the simulation.

Under the null hypothesis that the considered coverage interval covers the true parameter with the prescribed probability p, and using the normal approximation, we determined a 95%-confidence interval for the true value of the probability p. In our situations, we had $N = 5000$ and $p = 0.95$, resulting in a 95% confidence interval $\langle 0.944, 0.956 \rangle$. The fact that the estimated probability is outside this confidence interval indicates that with a high probability the actual probability of coverage of the parameter differs from the set nominal value at the 95% level.

The ISO coverage intervals for the intercept a and also for the new stimulus v_0 are somewhat conservative in our experimental setup, i.e. they include the true parameter value more often than expected. On the other hand, the observed empirical coverage probabilities of the ISO coverage intervals for slope b suggest that these interval estimators are exact in the sense that the difference of the observed coverage probabilities from the nominal level 0.95 is not statistically significant. Similarly, it can be observed that all MCM coverage intervals for a, b as well as for v_0 are exact for all considered combinations of the input parameters.

The results are consistent for all designs considered. The reported empirical coverage probabilities are neither influenced by the number of calibration points nor by the considered measurement uncertainty levels (small, medium, large).

As suggested by one reviewer, investigation of the empirical coverage probability of the joint coverage regions (for slope and intercept) and the closeness of the estimate to the simulated value could provide further valuable insights. We plan to conduct such and other detailed investigations in future work.

6. Conclusions

The main objective of this work was to study the problem of linear comparative calibration and to analyse the uncertainty of the measurement results obtained with the calibrated instrument. We have described and illustrated this concept in detail by analysing the actual calibration experiment of the pressure transducer and the subsequent analysis of the uncertainty of the measurement results at the Slovak Institute of Metrology in Bratislava. All required procedures, calibration, parameter estimation and subsequent uncertainty analysis of the measurement result with the calibrated instrument were performed according to ISO Technical Specification 28037:2010 and furthermore compared with the Monte Carlo method.

Our analysis shows that under experimental conditions, as described in this paper, the reported probability distributions of the values that can be reasonably attributed to the parameters of the calibration function derived using both approaches are very close.

In this particular example, some visible differences between the Monte Carlo distribution and the approximate distribution based on ISO were found in the measurement results obtained with the calibrated instrument. However, these differences are still acceptable, although the ISO method in combination with the normal distribution approximation cannot be recommended here and MCM is preferable. On the other

hand, it should be emphasised that these differences were intentionally caused by the use of a lower accuracy multimeter in combination with a calibrated pressure transmitter (resulting in higher uncertainty and a non-normality effect due to the correction of its systematic error).

Our simulation study (using a similar calibration model as in the example considered) focused on the empirical coverage probabilities of ISO and MCM coverage intervals, as well as on exploring the effects of sample size and different combinations of measurement uncertainties. As shown, the ISO coverage intervals for the intercept a and also for the new stimulus v_0 in our experimental setup are somewhat conservative, i.e. they include the true parameter value more often than expected. On the other hand, the observed empirical coverage probabilities of the ISO coverage intervals for slope b suggest that these interval estimators are exact in the sense that the difference of the observed coverage probabilities from the nominal level 0.95 is not statistically significant. From this perspective, the MCM coverage intervals for a, b as well as for v_0 are exact for all considered combinations of the input parameters. The reported empirical coverage probabilities are neither influenced by the number of calibration points nor by the considered measurement uncertainties.

However, it should be clear that these findings have limitations and that there are still open questions that require further research in this area. It should be emphasised that in the above example of comparative linear calibration, a simple calibration model with only two sources of uncertainty was considered, while the combined measurement uncertainties were relatively small compared to the measurement range.

As discussed in [17], the uncertainty matrix of parameter estimates derived from ISO does not strictly follow the law of propagation of uncertainty as defined in GUM and its supplements, which could lead to obvious differences especially in situations with an implicit measurement model and large combined measurement uncertainties of the variables involved. The authors argue that the uncertainties correctly derived based on the LPU are always larger than the uncertainties based on the ISO approach proposed in [5]. This requires further investigation and comparison of the empirical coverage probabilities and other statistical properties of the proposed approaches for a much wider range of designs.

As our study shows, the MCM coverage intervals had good statistical properties for all models and combinations of input parameters considered. However, the question arises of how to properly generate independent samples for MCM calculations.

In an earlier version of this paper, we generated new samples based on WTLS estimates fitted from the original data, $\hat{\mu}$ and $\hat{v}$, which, in the reviewer's opinion does not comply with JCGM 101:2008 [2] and JCGM 102:2011 [3]. However, we still believe that in situations where the expected values of the input variables are functionally constrained, the correct procedure for generating new Monte Carlo samples should be based on generating data around the estimates which satisfy the constraints required by the model, here $\hat{v} = a\mathbf{1} + b\hat{\mu}$ for some intercept a and some slope b, which of course does not generally hold for the observations $\mathbf{x}$ and $\mathbf{y}$ obtained by direct measurement in the calibration experiment.

As the second reviewer noted, the uncertainty evaluations, including the coverage intervals obtained by application of the Monte Carlo method described in JCGM 101:2008 [2] and JCGM 102:2011 [3], are not universally valid, as illustrated in [29] and [31], and the sufficient conditions for them to be valid in the context of errors-in-variables regression should be specified. At this moment, we cannot specify such conditions. We believe that this will help to open a new discussion on how to correctly generate calibration data in order to use the law of propagation of distributions through the Monte Carlo method.

REFERENCES

[1] JCGM. (2008). *Evaluation of measurement data — Guide to the expression of uncertainty in measurement.* JCGM 100:2008 (GUM 1995 with minor corrections).

[2] JCGM. (2008). *Evaluation of measurement data — Supplement 1 to the Guide to the expression of uncertainty in measurement — Propagation of distributions using a Monte Carlo method.* JCGM 101:2008.

[3] JCGM. (2011). *Evaluation of measurement data — Supplement 2 to the Guide to the expression of uncertainty in measurement — Extension to any number of output quantities.* JCGM 102:2011.

[4] JCGM. (2012). *International vocabulary of metrology – Basic and general concepts and associated terms (VIM), 3rd Edition.* JCGM 200:2012.

[5] ISO. (2010). *Determination and use of straight-line calibration functions.* ISO/TS 28037:2010.

[6] ISO. (2006). *Statistics – Vocabulary and symbols – Part 1: General statistical terms and terms used in probability.* ISO 3534-1:2006.

[7] European co-operation for Accreditation. (2013). *Evaluation of the uncertainty of measurement in calibration.* EA-4/02 M:2013.

[8] Balsamo, A., Mana, G., Pennecchi, F. (2006). The expression of uncertainty in non-linear parameter estimation. *Metrologia*, 43 (5), 396. https://doi.org/10.1088/0026-1394/43/5/009

[9] Bartel, T., Stoudt, S., Possolo, A. (2016). Force calibration using errors-in-variables regression and Monte Carlo uncertainty evaluation. *Metrologia*, 53 (3), 965. https://doi.org/10.1088/0026-1394/53/3/965

[10] Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. (2006). *Measurement Error in Non-*

linear Models: A Modern Perspective. Chapman and Hall/CRC, ISBN 9781584886334.

[11] Casella, G., Berger, R. L. (2002). *Statistical Inference, Second Edition*. Duxbury, ISBN 0-534-24312-6.

[12] Cecelski, C. E., Toman, B., Liu, F. H., Meija, J., Possolo, A. (2022). Errors-in-variables calibration with dark uncertainty. *Metrologia*, 59 (4), 045002. https://doi.org/10.1088/1681-7575/ac711c

[13] Cox, M., Forbes, A., Harris, P., Smith, I. (2004). *The classification and solution of regression problems for calibration*. NPL Report CMSC 24/03, National Physical Laboratory, Teddington, UK.

[14] Cox, D. R., Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30 (2), 248–265.

[15] Fuller, W. A. (2009). *Measurement Error Models*. John Wiley & Sons, ISBN 9780470317334.

[16] Guenther, F. R., Possolo, A. (2011). Calibration and uncertainty assessment for certified reference gas mixtures. *Analytical and Bioanalytical Chemistry*, 399 (1), 489–500. https://doi.org/10.1007/s00216-010-4379-z

[17] Klauenberg, K., Martens, S., Bošnjaković, A., Cox, M. G., van der Veen, A. M., Elster, C. (2022). The GUM perspective on straight-line errors-in-variables regression. *Measurement*, 187, 110340. https://doi.org/10.1016/j.measurement.2021.110340

[18] Kotz, S., Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge University Press, ISBN 9780511550683. https://doi.org/10.1017/CBO9780511550683

[19] Kubáček, L. (1988). *Foundations of Estimation Theory*. Elsevier, ISBN 978-0444989413.

[20] Kubáček, L. (1995). On a linearization of regression models. *Applications of Mathematics*, 40 (1), 61–78.

[21] Kukush, A., Van Huffel, S. (2004). Consistency of elementwise-weighted total least squares estimator in a multivariate errors-in-variables model $AX = B$. *Metrika*, 59 (1), 75–97. https://doi.org/10.1007/s001840300272

[22] Krystek, M., Anton, M. (2007). A weighted total least-squares algorithm for fitting a straight line. *Measurement Science and Technology*, 18 (11), 3438. https://doi.org/10.1088/0957-0233/18/11/025

[23] Lira, I., Elster, C., Wöger, W. (2007). Probabilistic and least-squares inference of the parameters of a straight-line model. *Metrologia*, 44 (5), 379. https://doi.org/10.1088/0026-1394/44/5/014

[24] Malengo, A., Pennecchi, F. (2013). A weighted total least-squares algorithm for any fitting model with correlated variables. *Metrologia*, 50 (6), 654. https://doi.org/10.1088/0026-1394/50/6/654

[25] Markovsky, I., Van Huffel, S. (2007). Overview of total least-squares methods. *Signal Processing*, 87 (10), 2283–2302. https://doi.org/10.1016/j.sigpro.2007.04.004

[26] Milton, M., Harris, P., Smith, I., Brown, A., Goody, B. (2006). Implementation of a generalized least-squares method for determining calibration curves from data with general uncertainty structures. *Metrologia*, 43 (4), S291. https://doi.org/10.1088/0026-1394/43/4/S17

[27] Murphy, S. A., Van Der Vaart, A. W. (1996). Likelihood inference in the errors-in-variables model. *Journal of Multivariate Analysis*, 59 (1), 81–108. https://doi.org/10.1006/jmva.1996.0055

[28] Osborne, C. (1991). Statistical calibration: A review. *International Statistical Review/Revue Internationale de Statistique*, 59 (3), 309–336. https://doi.org/10.2307/1403690

[29] Possolo, A., Iyer, H. K. (2017). Invited article: Concepts and tools for the evaluation of measurement uncertainty. *Review of Scientific Instruments*, 88 (1), 011301. https://doi.org/10.1063/1.4974274

[30] National Physical Laboratory. (2010). *Software to support ISO/TS 28037:2010E*. https://www.npl.co.uk/resources/software/iso-ts-28037-2010e

[31] Stoudt, S., Pintar, A., Possolo, A. (2021). Uncertainty evaluations from small datasets. *Metrologia*, 58 (1), 015014. https://doi.org/10.1088/1681-7575/abd372

[32] Witkovský, V. (2016). Numerical inversion of a characteristic function: An alternative tool to form the probability distribution of output quantity in linear measurement models. *Acta IMEKO*, 5 (3), 32–44. http://dx.doi.org/10.21014/acta_imeko.v5i3.382

[33] Witkovský, V., Wimmer, G. (2018). Generalized polynomial comparative calibration: Parameter estimation and applications. In *Advances in Measurements and Instrumentation: Reviews, Vol. 1*. IFSA Publishing, 15–52. ISBN 978-84-09-07321-4.

[34] Witkovský, V., Wimmer, G. (2021). Exact confidence intervals for parameters in linear models with parameter constraints. In *Measurement 2021: 13th International Conference on Measurement*. Bratislava, Slovakia: Institute of Measurement Science, SAS, 22–25. https://doi.org/10.23919/Measurement52780.2021.9446783

[35] Witkovský, V., Wimmer, G. (2022). PolyCal – MATLAB algorithm for comparative polynomial calibration and its applications. In *Advanced Mathematical and Computational Tools in Metrology and Testing XII: Series on Advances in Mathematics for Applied Sciences – Vol. 90*. World Scientific, 501–512. https://doi.org/http://dx.doi.org/10.1142/9789811242380_0033

[36] Witkovský, V. (2022). *CharFunTool: The characteristic functions toolbox (MATLAB)*. https://github.com/witkovsky/CharFunTool