XXI IMEKO World Congress "Measurement in Research and Industry"
August 30 - September 4, 2015, Prague, Czech Republic

THE COSINE ERROR: AN EXAMPLE OF A MODEL WITH A NON-REPETITIVE SYSTEMATIC EFFECT

Ignacio Lira [1] *and Dieter Grientschnig* [2]

[1] Department of Mechanical and Metallurgical Engineering, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile. Email: ilira@ing.puc.cl
[2] Metrolytik, Kernstockstr. 4, 8600 Bruck an der Mur, Austria. Email: d.grientschnig@metrolytik.at

Abstract **-** An inconsistency with respect to variable transformations in our previous treatment of the cosine error example with repositioning (*Metrologia*, vol. 47, pp. R1–R14) is pointed out and three remedial procedures are proposed. The problem refers to the measurement of the vertical height of a column of liquid in a manometer. A systematic effect arises because of the possible deviation of the measurement axis from the vertical, which may be different each time the measurement is performed. The proposed procedures are: a) merging of independent probability density functions for the measurand and the non-repetitive systematic effect transformed to a common parameterization; b) straightforward application of Bayesian statistics using a conditional reference prior with partial information; and c) an ad hoc procedure based on interpreting the measured quantities and the different influences of the systematic effect as points in Cartesian coordinates. Of these, procedure b) is the most sound theoretically.

Keywords: Cosine error, probability density function, Bayesian statistics, reference prior, Borel's paradox

1. INTRODUCTION

One of the examples appearing in a previous paper of ours [1] was the measurement of the vertical height of a column of liquid in a manometer. Our model was

$$Z = X\,Y, \tag{1}$$

where X is the height indicated by the instrument and Y is a correction factor for a systematic error caused by a possible deviation of the measurement axis from the vertical. This correction factor was in turn modeled as

$$Y = \cos W, \tag{2}$$

where W – the deviation angle – was supposed not to have been measured.

We assumed that indications $\boldsymbol{x} = (x_1, \ldots, x_n)$ for X were obtained while keeping the instrument in a fixed position, that the maximum possible absolute value of W was a given angle w, and that all angles between $-w$ and w were equally likely. Formulae for obtaining the best estimate of the measurand Z and its associated standard uncertainty under these conditions are given in the GUM [2, F.2.4.4]. In [1], these formulae were complemented by a derivation of the probability density function (PDF) for Z using the Bayesian methodology.

However, it is more realistic to assume that the data $\boldsymbol{x}$ were obtained by repositioning the device after every observation while keeping it in a fixed vertical plane. In this case, the measurement model changes to

$$Z = X_i\,Y_i, \qquad i = 1, \ldots, n, \tag{3}$$

where $Y_i = \cos W_i$ and the X_i's are the measured heights linked with angles W_i, which can again be supposed to be equally likely to assume any value within the interval $[-w, w]$. The data are still the set $\boldsymbol{x}$, but they are now interpreted as a single datum for each of the quantities $(X_1, \ldots, X_n)$, which we designate by $\boldsymbol{X}$.

In [1] we also analyzed this alternative formulation of the problem. However, at that time an inconsistency in our solution went unnoticed. In section 2 of the present paper, a summary of that solution is given and the origin of the inconsistency is pointed out. In section 3, three approaches that are free from this shortcoming are suggested. Results for a numerical example are presented in section 4 and conclusions are drawn in section 5.

2. THE ORIGINAL ANALYSIS

In our previous paper we denoted by $f_Q(\phi \,|\, \mathcal{K})$ the PDF for a generic quantity Q with possible values ϕ, where $\mathcal{K}$ represents the given model, data, assumptions and other information. Here we shall keep the same notation, using ω_i, ξ_i, η_i and ζ for the possible values of W_i, X_i, Y_i and Z, respectively. However, instead of $\mathcal{K}$ we shall indicate only the data $\boldsymbol{x}$ in the 'given' part of the argument of the PDFs. Thus, our goal will be to derive an expression for $f_Z(\zeta \,|\, \boldsymbol{x})$. Also, in [1] we used the delta function approach for effecting variable transformations; in the present paper we shall use instead the change-of-variables theorem (see e.g. [3]).

The original procedure started by considering each of the models (3) at a time. The joint PDFs for the input quantities of these models are of the form

$$f_{X_i,Y_i}(\xi_i, \eta_i \,|\, x_i) = f_{X_i}(\xi_i \,|\, x_i) f_{Y_i}(\eta_i), \tag{4}$$

where we used the fact that X_i and Y_i are independent *a priori*. (This is because, before knowing Z, having only

information about X_i does not yield any information about Y_i, and vice versa.) Change of variables produces

$$f_{X_i,Z}(\xi_i,\zeta\,|\,x_i) = \xi_i^{-1} f_{X_i}(\xi_i\,|\,x_i)\, f_{Y_i}\left(\zeta\,\xi_i^{-1}\right) \qquad (5)$$

and multiplication of all these PDFs gives

$$f_{\boldsymbol{X},Z}(\boldsymbol{\xi},\zeta\,|\,\boldsymbol{x}) \propto f_{\boldsymbol{X}}(\boldsymbol{\xi}\,|\,\boldsymbol{x}) \prod \xi_i^{-1} f_{Y_i}\left(\zeta\,\xi_i^{-1}\right), \qquad (6)$$

where

$$f_{\boldsymbol{X}}(\boldsymbol{\xi}\,|\,\boldsymbol{x}) = \prod f_{X_i}(\xi_i\,|\,x_i). \qquad (7)$$

Here and in the rest, the product is from $i = 1$ to n.

We then proceeded to derive expressions for PDFs $f_{\boldsymbol{X}}(\boldsymbol{\xi}\,|\,\boldsymbol{x})$ and $f_{Y_i}(\eta_i)$. The former were easily obtained by assuming that the indications x_i are drawn from Gaussian distributions with a common unknown standard deviation S whose possible values σ range from zero to infinity. The individual likelihoods must then include this nuisance parameter, so they become

$$\ell(\xi_i,\sigma\,;\,x_i) \propto \sigma^{-1} \exp\left(-0.5\,\sigma^{-2}(\xi_i - x_i)^2\right) \qquad (8)$$

and therefore the joint likelihood is

$$\ell(\boldsymbol{\xi},\sigma;\boldsymbol{x}) \propto \sigma^{-n} \exp\left(-0.5\,\sigma^{-2}\Gamma\right), \qquad (9)$$

where

$$\Gamma = \sum (\xi_i - x_i)^2 \qquad (10)$$

and the summation is from $i = 1$ to n.

In this likelihood, the quantities $\boldsymbol{X}$ are of equal inferential interest, which is higher than that of S. A reasonable non-informative prior in these circumstances is the two-group reference prior [4], where one group is made up of the location parameters $\boldsymbol{\xi}$ whereas the other contains only the scale parameter σ. This reference prior for the multidimensional location-scale statistical model (9) reads $f^o_{\boldsymbol{X},S}(\boldsymbol{\xi},\sigma) \propto \sigma^{-1}$. According to Bayes' theorem, the product of this prior with the likelihood (9) yields the joint posterior

$$f_{\boldsymbol{X},S}(\boldsymbol{\xi},\sigma\,|\,\boldsymbol{x}) \propto \sigma^{-(n+1)} \exp\left(-0.5\,\sigma^{-2}\Gamma\right), \qquad (11)$$

from which, if the number of measurements is greater that two, the parameter σ can be integrated out. The result is

$$f_{\boldsymbol{X}}(\boldsymbol{\xi}\,|\,\boldsymbol{x}) \propto \Gamma^{-n/2}. \qquad (12)$$

The PDFs $f_{Y_i}(\eta_i)$ were immediately obtained by a change of variable as

$$f_{Y_i}(\eta_i) = \left(1 - \eta_i^2\right)^{-1/2} f_{W_i}\left(\pm \arccos \eta_i\right), \qquad (13)$$

where for W_i a uniform distribution in the interval $[-w, w]$ was used. Therefore

$$f_{Y_i}(\eta_i) = w^{-1}\left(1 - \eta_i^2\right)^{-1/2} \qquad \text{for } \cos w \le \eta_i \le 1. \qquad (14)$$

Substitution of (12) and (14) in (6) produces

$$f_{\boldsymbol{X},Z}(\boldsymbol{\xi},\zeta\,|\,\boldsymbol{x}) \propto \Gamma^{-n/2} \prod \left(\xi_i^2 - \zeta^2\right)^{-1/2}, \qquad (15)$$

from which the desired PDF $f_Z(\zeta\,|\,\boldsymbol{x})$ can be obtained by marginalization and normalization.

Equivalently, transformation of (15) into the $(\boldsymbol{Y}, Z)$ and $(\boldsymbol{W}, Z)$ parameterizations gives, respectively,

$$f_{\boldsymbol{Y},Z}(\boldsymbol{\eta},\zeta\,|\,\boldsymbol{x}) \propto \Gamma^{-n/2} \prod \eta_i^{-1}\left(1 - \eta_i^2\right)^{-1/2} \qquad (16)$$

and

$$f_{\boldsymbol{W},Z}(\boldsymbol{\omega},\zeta\,|\,\boldsymbol{x}) \propto \Gamma^{-n/2} \prod (\cos\omega_i)^{-1}, \qquad (17)$$

where Γ must be written as

$$\Gamma = \sum \left(\zeta\,\eta_i^{-1} - x_i\right)^2 \qquad (18)$$

in (16) and as

$$\Gamma = \sum \left(\zeta\,(\cos\omega_i)^{-1} - x_i\right)^2 \qquad (19)$$

in (17). Note that, in the notation of the present paper, (17) is exactly equal to equation (45) in [1].

Is there anything wrong with this analysis? Basically, it is perfectly legitimate to multiply the PDFs $f_{X_i}(\xi_i\,|\,x_i)$ to give $f_{\boldsymbol{X}}(\boldsymbol{\xi}\,|\,\boldsymbol{x})$ and to multiply the PDFs $f_{Y_i}(\eta_i)$ to give $f_{\boldsymbol{Y}}(\boldsymbol{\eta})$. This is because, in the case considered, the quantities $\boldsymbol{X}$ and $\boldsymbol{Y}$ are all pairwise independent. Multiplying $f_{\boldsymbol{X}}(\boldsymbol{\xi}\,|\,\boldsymbol{x})$ and $f_{\boldsymbol{Y}}(\boldsymbol{\eta})$ to give $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{\xi},\boldsymbol{\eta}\,|\,\boldsymbol{x})$ would also be legitimate. But the latter PDF *would not* represent the state of knowledge about all input quantities *after* considering the measurement models. If used in that sense, an inconsistency known as "Borel's paradox" would arise. In our case, avoiding this paradox precludes a distribution of dimension lower than $2n$ to be derived uniquely from $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{\xi},\boldsymbol{\eta}\,|\,\boldsymbol{x})$, as e.g. PDFs (15), (16) and (17), whose dimension is $n + 1$. However, the effects of this paradox are only revealed if further transformations designed as consistency checks are carried out [5]. For this reason, it was not noticed in our original analysis.

The operation above that causes the trouble is the step from (5) to (6). This represents the multiplication of PDFs that share a common variable. Such an operation is not justified by any rule, so it is not statistically sound.

3. ALTERNATIVE PROCEDURES

In this section, we propose three alternative procedures to obtain the PDF $f_Z(\zeta\,|\,\boldsymbol{x})$ that are free from the unwarranted step mentioned above. The first procedure is to use the logarithmic pooling technique for merging PDFs. The second is to use orthodox Bayesian statistics with a reference prior. And the third is to produce a single model for Z out of those in (3) by least-squares adjustment.

3.1. The merging procedure

The merging procedure is somewhat similar to our original treatment of the problem. However, to remedy the inconsistency mentioned above, all PDFs have to be transformed beforehand to a common parameterization by means of the models (3) and only subsequently merged

rather than vice versa, as was pointed out in [5, 6]. The parameterization we choose for this purpose is $(\boldsymbol{X}, Z)$. Thus, we start from the multiplication of (12) for $\boldsymbol{X}$ – which is based on the entire information imparted by the measurement data – with the PDF for a single Y_i. Such an operation is allowable, because these PDFs are independent. Hence, the transformation of

$$f^{(i)}_{\boldsymbol{X},Y_i}(\boldsymbol{\xi}, \eta_i \,|\, \boldsymbol{x}) \propto f_{\boldsymbol{X}}(\boldsymbol{\xi} \,|\, \boldsymbol{x})\, f_{Y_i}(\eta_i) \tag{20}$$

into the selected parameterization gives

$$f^{(i)}_{\boldsymbol{X},Z}(\boldsymbol{\xi}, \zeta \,|\, \boldsymbol{x}) \propto \xi_i^{-1} f_{\boldsymbol{X}}(\boldsymbol{\xi} \,|\, \boldsymbol{x})\, f_{Y_i}\left(\zeta\, \xi_i^{-1}\right), \tag{21}$$

where superscript (i) indicates which information underlies the distribution.

By substitution of (12) and (14) into (21), follows

$$f^{(i)}_{\boldsymbol{X},Z}(\boldsymbol{\xi}, \zeta \,|\, \boldsymbol{x}) \propto \Gamma^{-n/2} \left(\xi_i^2 - \zeta^2\right)^{-1/2}, \tag{22}$$

which is valid for $\xi_i \cos w \le \zeta \le \xi_i$.

All these PDFs for $i = 1, 2, \ldots, n$ need to be merged, but not by simple multiplication because, as (21) reveals, they share common information encoded by $f_{\boldsymbol{X}}(\boldsymbol{\xi} \,|\, \boldsymbol{x})$. The merging method to be employed must satisfy two requisites: first, it must have the distributive property [5], i.e. it must not repeatedly consider this common information. Second, the ensuing PDF needs to be consistent with the PDFs obtained by choosing any other parameterization for the same merging method. The logarithmic pooling technique satisfies both requisites. It consists in first raising each PDF to some exponent and then taking their product [7]. The exponents – known as pooling weights – are in principle arbitrary. However, as shown in [5, 6], to forestall Borel's paradox and to provide this technique with the distributive property, all weights must add up to one. They are usually selected such that greater weights are given to those PDFs that are considered to be more reliable. In our case, since there is no reason to prefer some of the PDFs in (22) over the others, we take all weights as equal to $1/n$. Thus, we obtain

$$f_{\boldsymbol{X},Z}(\boldsymbol{\xi}, \zeta \,|\, \boldsymbol{x}) \propto \Gamma^{-n/2} \prod \left(\xi_i^2 - \zeta^2\right)^{-1/(2n)}. \tag{23}$$

The equivalent expressions in the $(\boldsymbol{Y}, Z)$ and $(\boldsymbol{W}, Z)$ parameterizations are

$$f_{\boldsymbol{Y},Z}(\boldsymbol{\eta}, \zeta \,|\, \boldsymbol{x}) \propto \zeta^{n-1} \Gamma^{-n/2} \prod \left(1 - \eta_i^2\right)^{-1/(2n)} \eta_i^{-2+1/n} \tag{24}$$

and

$$f_{\boldsymbol{W},Z}(\boldsymbol{\omega}, \zeta \,|\, \boldsymbol{x}) \propto \zeta^{n-1} \Gamma^{-n/2} \prod |\sin \omega_i|^{1-1/n} (\cos \omega_i)^{-2+1/n}. \tag{25}$$

Evidently, the factor Γ must be written as (10) in (23), as (18) in (24) and as (19) in (25).

3.2. The Bayesian procedure

In the Bayesian procedure, we use the likelihood (9) to update the prior

$$f^{o}_{\boldsymbol{Y},Z,S}(\boldsymbol{\eta}, \zeta, \sigma) = f^{o}_{Z,S|\boldsymbol{Y}}(\zeta, \sigma \,|\, \boldsymbol{\eta}) f_{\boldsymbol{Y}}(\boldsymbol{\eta}), \tag{26}$$

where $f^{o}_{Z,S|\boldsymbol{Y}}(\zeta, \sigma \,|\, \boldsymbol{\eta})$ is a conditional reference prior with partial information [8]. In appendix 1, this prior is derived along the lines of [9] and [10]. The result is

$$f^{o}_{Z,S|\boldsymbol{Y}}(\zeta, \sigma \,|\, \boldsymbol{\eta}) \propto (\sigma\, \mu)^{-1}, \tag{27}$$

where

$$\mu = \min(\eta_1, \ldots, \eta_n). \tag{28}$$

Thus, using Bayes' theorem and integrating out σ, we obtain

$$f_{\boldsymbol{Y},Z}(\boldsymbol{\eta}, \zeta \,|\, \boldsymbol{x}) \propto \mu^{-1} \Gamma^{-n/2} \prod \left(1 - \eta_i^2\right)^{-1/2} \tag{29}$$

with Γ as in (18) or, in the $(\boldsymbol{W}, Z)$ parameterization,

$$f_{\boldsymbol{W},Z}(\boldsymbol{\omega}, \zeta \,|\, \boldsymbol{x}) \propto \nu^{-1} \Gamma^{-n/2} \tag{30}$$

with Γ as in (19) and

$$\nu = \cos(\max(\omega_1, \ldots, \omega_n)). \tag{31}$$

3.3. The least-squares procedure

Still another procedure follows by interpreting the quantities X_i and Y_i as a set of points in the Cartesian plane, to which the rectangular hyperbola $Z = X\,Y$ is fitted. This is done by minimizing the sum

$$\sum (Z - X_i\, Y_i)^2, \tag{32}$$

which leads to the single model

$$Z = \frac{1}{n} \sum X_i\, Y_i. \tag{33}$$

This model can be analyzed with the standard procedure in GUM Supplement 1 [11]. It consists in drawing a set of $\boldsymbol{\xi}$ values from the normalized joint PDF (12) in conjunction with (10), together with a set of $\boldsymbol{\omega}$ values from uniform PDFs with support on the interval $[-w, w]$. We then compute

$$\zeta = \frac{1}{n} \sum \xi_i \, \cos \omega_i \tag{34}$$

and repeat a large number of times. The histogram of the ζ values is an approximation to the PDF for the measurand [12].

4. EXAMPLE

Let the measurement information consist of the five values $\boldsymbol{x} = \{39.88, 39.93, 40.00, 40.09, 40.12\}$ (in centimeters) for the height of the liquid column in the manometer, and assume that the maximum deviation angle is $w = 5^\circ$ to both sides of the vertical. These are the values used in the original example in [1].

Two approaches were used for the evaluation of $f_Z(\zeta \,|\, \boldsymbol{x})$ using the merging and Bayesian procedures. One was direct numerical integration and marginalization of PDFs (25) and (30). (The $(\boldsymbol{W}, Z)$ parameterization was chosen in order to avoid the singularities at $\eta_i = 1$ exhibited by PDFs (24) and (29).) The other approach was the application of Markov chain Monte Carlo (MCMC) by means of the Metropolis-Hastings algorithm [13], see appendix 2.

Results for the first two procedures are depicted in Figs. 1 and 2. They show excellent agreement between these two numerical approaches, which lends credibility to both. The standard deviations of the two PDFs are quite similar (about 0.06 cm), but the mean for the merging procedure (39.932 cm) is a little smaller than that for the Bayesian procedure (39.951 cm).

The similarity between the PDFs in Figs. 1 and 2 is an unexpected finding, because usually logarithmic pooling yields a 'compromise distribution' of all PDFs pooled, whereas the ordinary Bayesian procedure generally produces a PDF that is narrower than the individual ones [5]. As (21) suggests, the latter are shifted with respect to one another along the ζ-axis, but apart from that they are similar. Hence, logarithmic pooling leads to a PDF of approximately the same standard deviation as the PDFs merged, albeit at an intermediate ζ-position. But how can we explain the fact that in the circumstances at issue the PDF resulting from the orthodox Bayesian procedure is not narrower than that produced by the logarithmic pooling technique? We speculate that this behavior might be a consequence of the considerable shifts between the individual PDFs.

The PDF obtained with the least-squares procedure is quite different from the former two, see Fig. 3. It was computed by first using MCMC to draw samples from (12) and then using these samples in conjunction with samples from the uniform distributions assigned to the quantities W_i to compute samples ζ in accordance with (34). Whereas the mean of the PDF in Fig. 3 is similar to those obtained with the other two procedures, the seemingly very long tails of this distribution cause an extremely large standard deviation, whose value we were unable to estimate reliably. The explanation is that, by chance, the products $X_i\, Y_i$ in (33) may all assume very large or very small values in the same random draw. We found that the same phenomenon arises when the manometer is assumed to be always positioned perfectly upright. Applying direct numerical integration to the latter scenario corroborated the respective MCMC calculations.

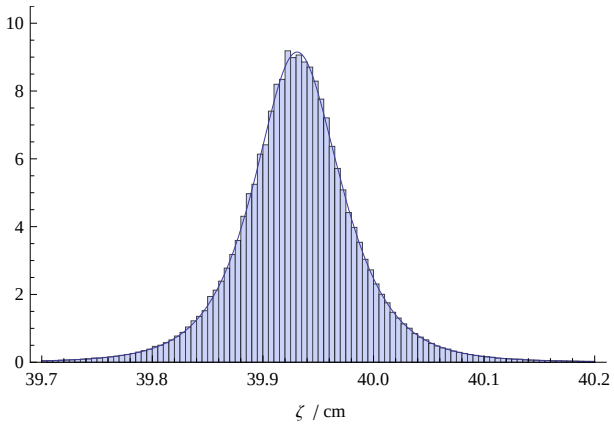

Fig. 1: PDF for the measurand Z using the merging procedure. It was obtained by direct numerical integration and marginalization of (25) (solid line) and by application of MCMC (histogram).

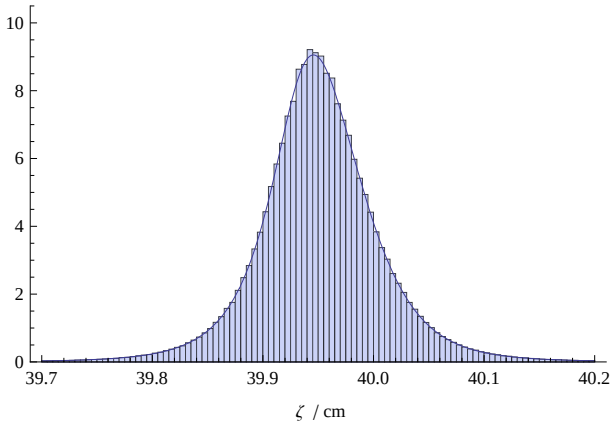

Fig. 2: Same as Fig. 1, but based on the Bayesian procedure (equation (30)).

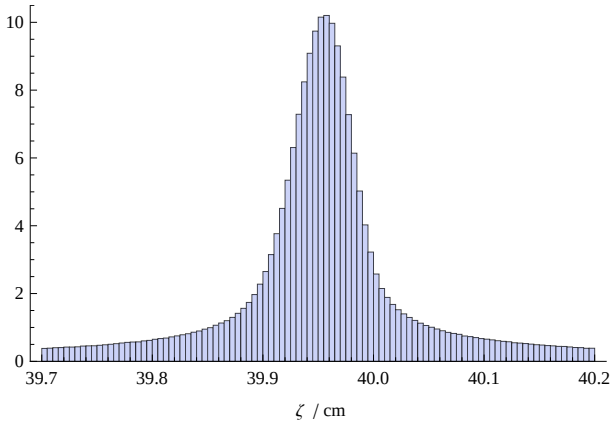

Fig. 3: PDF for the measurand Z using the least-squares procedure. It was obtained as described in subsection 3.3 using MCMC to draw samples from (12).

5. CONCLUSIONS

A deficiency in our previous treatment of the cosine error example with repositioning, given in [1], has been pointed out and three remedial procedures have been offered. These procedures are: a) logarithmic pooling of the joint PDFs (22) for the measurand Z and the measured quantities $\boldsymbol{X}$, where each of the merged PDFs is based on the consideration of just one embodiment Y_i of the systematic effect; b) orthodox Bayesian statistics using a

conditional reference prior with partial information derived in accordance with the principles in [8]; and c) an ad hoc procedure based on interpreting the measured quantities X_i and the influences Y_i of the systematic effect as Cartesian coordinates to which a rectangular hyperbola is fitted.

The three procedures were tested with the same data we used in [1]. The results of procedures a) and b) are acceptable and resemble each other. They were implemented using both, direct marginalization and normalization of the joint PDFs for the measurand Z and the unmeasured deviation angles W_i, and an MCMC algorithm to obtain a large number of samples from the PDF for Z. These two computation methods were found to be in perfect agreement with each other. However, MCMC can be used with any number of measurements of the height X, whereas, at present, integration is limited by the capacity of the numerical integration routine to handle a limited number of measurements.

Procedure c) is interesting in that its implementation consists of a combination of MCMC and the MC method in GUM Supplement 1 [11]. However, with this procedure we obtained a PDF exhibiting an unacceptably large standard deviation. We think that this is not a numerical artifact but is rather due to the characteristics of model (33) itself.

For a procedure to be advisable, the results it yields for the scenario where the manometer is known not to deviate from the vertical shall concur with the well-known solution that applies in this instance, viz. a scaled and shifted t-distribution of $n-1$ degrees of freedom [1]. The latter has a mean of 40.004 and a standard deviation of 0.065. Eq. (30) shows that in the circumstances described the Bayesian procedure reduces exactly to the aforementioned t-distribution, which is not the case for the other two procedures. In this respect, the Bayesian one is preferable.

The general lesson to be learned is that grafting the least-squares principle on Bayesian statistics, however reasonable and promising it may seem at first sight, is not recommended to be used in practice. The merging procedure is also ad hoc. While in the present case it produces acceptable results, it should also be avoided in favor of the sound, theoretically proven and widely accepted technique of Bayesian statistics with a reference prior.

APPENDIX 1: Derivation of reference prior (27)

With $\theta_1 = \zeta$ and $\theta_2 = \sigma$, the Fisher information matrix for the statistical model (9) (with Γ as in (18)) becomes

$$I_{\theta\theta}(\boldsymbol{\eta}, \boldsymbol{\theta}) = \mathrm{diag}\left(\theta_2^{-2} \sum \eta_i^{-2}, 2\,n\,\theta_2^{-2}\right), \tag{35}$$

(cf. (4) in [9]). Using the notation of [14] this implies

$$\left|I_{\theta\theta[\sim 00]}(\boldsymbol{\eta}, \boldsymbol{\theta})\right| = 2\,n\,\theta_2^{-4} \sum \eta_i^{-2} \tag{36}$$

$$\left|I_{\theta\theta[\sim 11]}(\boldsymbol{\eta}, \boldsymbol{\theta})\right| = 2\,n\,\theta_2^{-2} \tag{37}$$

$$\left|I_{\theta\theta[\sim 22]}(\boldsymbol{\eta}, \boldsymbol{\theta})\right| = 1, \tag{38}$$

so Lemma 2.1 of [14] directly yields

$$|h_1(\boldsymbol{\theta})| = \theta_2^{-2} \sum \eta_i^{-2} \tag{39}$$

$$|h_2(\boldsymbol{\theta})| = 2\,n\,\theta_2^{-2}. \tag{40}$$

According to the notation of [15] these expressions factorize as

$$|h_j(\boldsymbol{\theta})| = h_{j1}\left(\boldsymbol{\theta}_{(j)}\right) h_{j2}\left(\boldsymbol{\theta}_{(j)}^C\right). \tag{41}$$

Only $h_{11}\left(\boldsymbol{\theta}_{(1)}\right)$ and $h_{21}\left(\boldsymbol{\theta}_{(2)}\right)$ are required to calculate with the help of Theorem 1 of [15] or Proposition 1 of [16] the conditional reference prior that results from compact rectangular subsets (i.e. subsets that are products of compact sets in the two subspaces whose bounds are independent). This prior is

$$p^*(\boldsymbol{\theta}|\boldsymbol{\eta}) = h_{11}^{1/2}\left(\boldsymbol{\theta}_{(1)}\right)\, h_{21}^{1/2}\left(\boldsymbol{\theta}_{(2)}\right). \tag{42}$$

It is plausible that the η_i's have to be assigned to the factors that matter, because otherwise they could never take effect. However, this is just a supposition, not a reliable remedy for the ambiguity of the assignment. Since this calls the ensuing simplified calculation into question, it may be checked by a more cumbersome computation method referenced below. Nonetheless, in order to test the presumption made, let us proceed with the resulting factorization

$$h_{11}\left(\boldsymbol{\theta}_{(1)}\right) = \sum \eta_i^{-2}, \tag{43}$$

$$h_{12}\left(\boldsymbol{\theta}_{(1)}^C\right) = \theta_2^{-2}, \tag{44}$$

$$h_{21}\left(\boldsymbol{\theta}_{(2)}\right) = 2\,n\,\theta_2^{-2}, \tag{45}$$

$$h_{22}\left(\boldsymbol{\theta}_{(2)}^C\right) = 1. \tag{46}$$

This yields the conditional reference prior

$$p^*(\boldsymbol{\theta}|\boldsymbol{\eta}) = \theta_2^{-1}\left(\sum \eta_i^{-2}\right)^{1/2}. \tag{47}$$

By following the more fundamental procedure in [17] one obtains the same result, which retrospectively proves the above simplified derivation.

We continue in accordance with section 3 of [10]. Since the quantities X, Y and Z are all non-negative, the subsets of the original parameter space can be defined as e.g.

$$\Xi_{1,m} \times \ldots \times \Xi_{n,m} \times \Sigma_m = \{(\boldsymbol{\xi}, \sigma) : \\ \mathrm{e}^{-m} \le \xi_1 \le \mathrm{e}^m, \ldots, \mathrm{e}^{-m} \le \xi_n \le \mathrm{e}^m, \mathrm{e}^{-m} \le \sigma \le \mathrm{e}^m\}, \tag{48}$$

where $m = 1, 2, \ldots$. (Note that in this expression Σ does not stand for a summation sign.) In order that none of these limits is exceeded, the subsets restricted to $\boldsymbol{\theta}$ take the form $\boldsymbol{\theta}_m = \{\boldsymbol{\theta} : L_{\inf} \le \theta_1 \le L_{\sup}, \mathrm{e}^{-m} \le \theta_2 \le \mathrm{e}^m\}$, where $L_{\inf} = \mathrm{e}^{-m} \max_{1\le l\le n}(\eta_l)$ and $L_{\sup} = \mathrm{e}^m \min_{1\le l\le n}(\eta_l)$. In analogy to (6) of [10] follows

$$K_m(\boldsymbol{\eta}) = \left[2m\left(\sum \eta_i^{-2}\right)^{1/2} \int_{L_{\inf}}^{L_{\sup}} \mathrm{d}\theta_1\right]^{-1}, \tag{49}$$

hence

$$\lim_{m\to\infty}\frac{K_m(\boldsymbol{\eta})}{K_m(\boldsymbol{\eta}^{\times})} \propto \left[\min_{1\le l\le n}(\eta_l)\left(\sum \eta_i^{-2}\right)^{1/2}\right]^{-1}. \tag{50}$$

Thereby we get (27):

$$p(\boldsymbol{\theta}|\boldsymbol{\eta}) \propto \left[\theta_2 \times \min_{1\le l\le n}(\eta_l)\right]^{-1}. \tag{51}$$

APPENDIX 2: The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm [13] is an especially useful MCMC method for drawing random samples from a 'target' univariate or multivariate PDF $f(\boldsymbol{\nu})$. The algorithm is quite simple. It starts by choosing an arbitrary 'proposal' distribution $g(\boldsymbol{\mu}|\boldsymbol{\nu})$ together with an arbitrary starting point $\boldsymbol{\nu}_1$. Then, for $i = 1, 2, \ldots, N$:

- Draw a random sample $\boldsymbol{\mu}$ from $g(\boldsymbol{\mu}|\boldsymbol{\nu}_i)$ and a probability u from the uniform distribution supported on $(0, 1)$.
- Compute the probability of accepting $\boldsymbol{\mu}$:

$$p = \min\left\{\frac{f(\boldsymbol{\mu})}{f(\boldsymbol{\nu}_i)} \times \frac{g(\boldsymbol{\nu}_i|\boldsymbol{\mu})}{g(\boldsymbol{\mu}|\boldsymbol{\nu}_i)}, 1\right\}.$$

- If $u < p$ set $\boldsymbol{\nu}_{i+1} = \boldsymbol{\mu}$, else set $\boldsymbol{\nu}_{i+1} = \boldsymbol{\nu}_i$.

The values $\{\boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_N\}$ constitute a Markov chain. For sufficiently large N the chain will eventually converge to the target distribution. However, a 'burn-in' period is necessary, after which the initial state is considered to have been 'forgotten'. This means that an initial set $\{\boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_{M<N}\}$ is thrown away and the remaining set is assumed to represent a sample from $f(\boldsymbol{\nu})$.

Note that the acceptance probability involves only the ratios of the target and proposal distributions, so neither of them need to be normalized. Also note that since we used a symmetric proposal distribution, $g(\boldsymbol{\nu}_i|\boldsymbol{\mu}) = g(\boldsymbol{\mu}|\boldsymbol{\nu}_i)$, the acceptance probability simplifies to $p = \min\{f(\boldsymbol{\mu})/f(\boldsymbol{\nu}_i), 1\}$.

For implementing the merging and Bayesian procedures we chose a t-distribution with location parameter ζ_i, scale parameter 0.13 and five degrees of freedom, along with uniform distributions on the interval on $[-w, w]$, as the proposal distributions to generate candidate values $\boldsymbol{\mu} = (\zeta, \omega_1, \ldots, \omega_5)$. The starting point was $\zeta_1 = \overline{x}$ (the mean of the measurements $\boldsymbol{x}$), the chain was of length $N = 10^7$ and the burn-in period was $M = N/2$. Such a large N is probably an exaggeration, but as in other walks of life, there is no harm in excess.

Because of the singularities of PDF (12), located along a line of finite length, obtaining samples from this PDF – as necessary for the implementation of the least-squares procedure – was more difficult. A 5-variate Gaussian distribution with mean vector $\boldsymbol{x}_i$ and diagonal covariance matrix $\sigma\,\boldsymbol{I}$ was chosen as the proposal distribution, where $\boldsymbol{I}$ is the identity matrix. After some fine-tuning, acceptable results were obtained with $\sigma = 0.0075$ cm. Again, we let the chain run for $N = 10^7$ iterations (which in this instance was not exaggerated at all), throwing away the first $M = 2 \times 10^6$ vectors. The starting vector was $(\overline{x}, \ldots, \overline{x})$.

ACKNOWLEDGMENTS

The first author acknowledges the financial support of Fondecyt, research grant 1141165.

REFERENCES

[1] I. Lira and D. Grientschnig. Bayesian assessment of uncertainty in metrology: a tutorial. *Metrologia*, 47:R1–R14, 2010.

[2] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Guide to the Expression of Uncertainty in Measurement. JCGM 100:2008, Joint Committee for Guides in Metrology.

[3] P. D. Lax. Change of variables in multiple integrals. *Am. Math. Mon.*, 106:497–501, 1999.

[4] C. Fernández and M. F. J. Steel. Reference priors for the general location-scale model. *Stat. Prob. Lett.*, 43:377–384, 1999.

[5] D. Grientschnig and I. Lira. Combining probability distributions by multiplication in metrology: A viable method? *Int. Stat. Rev.*, 82:392–410, 2014.

[6] I. Lira and D. Grientschnig. Deriving PDFs for interrelated quantities: what to do if there is 'more than enough' information? *IEEE T. Instrum. Meas.*, 63:1937–1946, 2014.

[7] Clemen R. T. and Winkler R. L. Combining probability distributions from experts in risk analysis. *Risk Anal.*, 19:187–203, 1999.

[8] D. Sun and J. O. Berger. Reference priors with partial information. *Biometrika*, 85:55–71, 1998.

[9] D. Grientschnig and I. Lira. Reassessment of a calibration model by Bayesian reference analysis. *Metrologia*, 48:L7–L11, 2011.

[10] D. Grientschnig and I. Lira. Revision of 'Reassessment of a calibration model by Bayesian reference analysis'. *Metrologia*, 49:L1–L3, 2012.

[11] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Supplement 1 to the 'Guide to the Expression of Uncertainty in Measurement'. JCGM 101:2008, Joint Committee for Guides in Metrology.

[12] I. Lira. Monte Carlo evaluation of the uncertainty associated with the construction and use of a fitted curve. *Measurement*, 44:2156–2164, 2011.

[13] S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *Am. Stat.*, 49:327–335, 1995.

[14] G. S. Datta and M. Ghosh. On the invariance of noninformative priors. *Ann. Stat.*, 24:141–159, 1996.

[15] G. S. Datta and M. Ghosh. Some remarks on noninformative priors. *J. Am. Stat. Assoc.*, 90:1357–1363, 1995.

[16] E. Gutiérrez-Peña and R. Rueda. Reference priors for exponential families. *J. Stat. Plann. Inference*, 110:35–54, 2003.

[17] J. O. Berger and J. M. Bernardo. Reference priors in a variance components problem. In P. K. Goel and N. S. Iyengar, editors, *Bayesian Analysis in Statistics and Econometrics (Lecture Notes in Statistics*, volume 75, pages 177–194). Springer, New York, 1992.