XIX IMEKO World Congress
Fundamental and Applied Metrology
September 6−11, 2009, Lisbon, Portugal

TYPE A EVALUATION OF UNCERTAINTY DUE TO SYSTEMATIC EFFECTS IN DIGITAL OSCILLOSCOPES

Filippo Attivissimo[1], *Andrea Cataldo*[2], *Laura Fabbiano*[1], *Nicola Giaquinto*[1]

[1] Department of Electrics and Electronics, Polytechnic of Bari, 70125 Bari, Italy
e-mail: [attivissimo, fabbiano, giaquinto]@misure.poliba.it
[2] Department of Engineering for Innovation, University of Salento, 73100 Lecce, Italy
email: andrea.cataldo@unisalento.it

Abstract – The paper presents a type A (experimental) evaluation of the uncertainty due to systematic effects. After a brief discussion about the general problem of choosing a proper mathematical representation for systematic effects in the context of uncertainty evaluation, a pragmatic approach, based on the familiar random variable theory and the ISO 5725 norm is proposed. The approach is called "inter-instrument experiment", on the analogy of the "inter-laboratory experiment" of the ISO 5725. Preliminary experimental results, relevant to commercial digital oscilloscopes, are presented and discussed.

Keywords: uncertainty, systematic effects, digital instruments

1. INTRODUCTION

In recent years, much attention has been devoted by researchers to uncertainty evaluation in general, and to the contribution of systematic effects (SEs) to uncertainty in particular. The uncertainty due to SEs is indeed a source of many theoretical and practical problems, and also of simple misunderstandings, even among professionals in the measurement fields.

It is quite common, for example, to confuse between "component of uncertainty arising from a SE" and "type B component of uncertainty", i.e., uncertainty evaluated without actual statistical analysis of experimental data; so common, that the American National Institute of Standards and Technology warns explicitly against this mistake [1]. On the other hand, the systematic nature of an effect is generally an inherent obstacle to statistical evaluation, since, by definition, an effect is systematic if it does not change in a series of measurements.

Describing a SE, which "does not change", with a random variable (RV), a mathematical object that "changes", is conceptually a not so straightforward step (contrary to the case of random effects). The ISO/IEC Guide to the Expression of Uncertainty in Measurements (GUM) [2, 3, 4], however, yields clear directions on how to handle an unknown SE with RVs. The effect must be thought as an unknown additive correction to the measured value, with a proper distribution which can be inferred, for example, from the manufacturer's accuracy specifications (type B evaluation). As a consequence, a SE contributes to the combined uncertainty exactly as a random one, with a +1 sensitivity coefficient.

If the SE is *known* (and of meaningful importance), the GUM prescribes simply to subtract it from the measurement result. This sensible and quite obvious recommendation is another common source of misunderstandings, as many people seem to think that the GUM, recommending the compensation of known SEs, does not address the uncertainty contribution of *unknown* SEs. The Guide, on the contrary, provides explicitly a procedure, with examples, to deal with unknown SEs.

Apart from this confusion, some researchers found the GUM procedure to evaluate the uncertainty due to unknown SEs not totally convincing, and they have proposed some alternative methods that do not involve RVs. In [5], for example, it is stated that probability theory "handles, with its random variables, only that particular class of incomplete information due to random effects"; as a consequence, the use of the Dempster-Shafer theory of evidence is suggested, to take into account both systematic and random effects by means of random-fuzzy variables. The use of this theory in the measurement fields has been proposed also in relation to different problems, like sensor fusion [6].

The Dempster-Shafer theory is the subject of many lively discussions (see e.g. [7, 8]), and discussing it is beyond the scope of this work. It must be noted, however, that this theory is usually conceived as a mathematical tool to deal with decisions problems depending upon *subjective judgment*, such as, for example, a trial with contrasting testimonies. Shafer himself has expressed, very recently [9], the view that the theory is useful when "there is not a repetitive structure for the question and the data"; otherwise, "we can make good probability forecasts".

SEs in measurements certainly may have a repetitive structure, which is typically described by manufacturers through the accuracy specifications. It is true, on the other hand, that specifications (which usually only say that SEs won't exceed some specified limits) are far from giving a complete statistical information. Therefore, in the evaluation of SEs-related uncertainty there is, after all, some amount of subjective judgment, but, contrary to other situations (like that of contrasting testimonies), this judgment can be subject to extensive experimental verification.

ISBN 978-963-88410-0-1

Experimental evaluation of the uncertainty due to SEs, and comparison with manufacturers' specifications, is the subject and the goal of the present work. The work connects to others by some of the authors, like [10], in which an operational way of writing, interpreting and using the specifications of instruments is discussed, and to others of independent authors like [11], in which actual specifications about SEs are used to evaluate uncertainty of actual measurements. The key point of the present paper is that the uncertainty evaluation is experimental not only in the sense that it is made on actual measurements (like in [11]); also the specifications about SEs are derived experimentally. This involves the statistical analysis of many SE measurements on *many different instruments of the same kind*, situated in different laboratories (interlaboratory comparison). Therefore, the work makes use also of concepts illustrated in the ISO norms 5725 [12, 13, 14, 15].

The paper is organized as follows. Section 2 illustrates the theoretical fundamentals of the work. Section 3 presents some preliminary results, coming from experimental data obtained from two oscilloscopes, along with related considerations about future work. The last section contains final considerations and conclusions.

2. THEORETICAL FUNDAMENTALS

The theoretical background of the work consists, basically, in a re-formulation of well-established concepts of the ISO 5725 norms, in particular those of Part 2 [13], dealing with the determination of repeatability and reproducibility of a standard measurement method.

2.1. Basic concepts in ISO 5725

It is useful, without going deeply in the procedures and the mathematics of the ISO 5725 norm, to recall its basic conceptual scheme. In the norm, a number of laboratories co-operate in an "interlaboratory experiment" to determine the metrological performance of a measurement method, which must be "standard", i.e., must be properly described in an accepted written document. To this purpose, a set of samples are prepared, in order to test the measurement methods for different amounts of the quantity to be measured. For example, if the method is aimed at measuring the sulfur content in coal, many different coal samples, with different sulfur content, must be prepared.

Let p be the number of laboratories, and q the number of different samples, or *levels*. The i-th laboratory ($i = 1,..., p$) obtains, and publishes, n_{ij} repeated measurements on the j-th sample, ($j = 1,...q$). The overall result is a table with $p \times q$ cells, where each cell contains the measurement results y_{ijk} ($k = 1,..., n_{ij}$).

In the following, for the sake of clarity, the index j is omitted, since it identifies the particular sample considered and is inessential for other purposes. The obtained measurements can be written as:

$$y_{ik} = x + b + B_i + e_{ik} \tag{1}$$

where x is the "true value" (accepted reference value) of the quantity in the considered sample, b is the bias of the measurement method, B_i is the bias of the i-th laboratory, and e_{ik} is the random error component.

The essence of decomposition (1) lies in the basic assumption that the average of each indexed error term, both over i and over k, is zero. With this principle in mind, it is quite obvious to interpret the term e_{ik} as a "completely random" error, and the term B_i as an error that is systematic in the local context of a single laboratory, while it is random in the global context of the many different laboratories. The term b is irreducibly systematic and deterministic (it is determined simply by the nature of the measurement method, and not by the particular laboratory using the method).

In the context of the single i-th laboratory the variance of the measurements is of course the one of the sole term e_{ik}. This is called the *intralaboratory variance* (or intracell variance) σ_{Wi}^2 and is, ideally, independent on i (the same for all the laboratories). Since in actual measurements it is impossible to obtain exactly equal intralaboratory variances, their average is taken and given the name of *repeatability variance* σ_r^2. The error term B_i is, of course, systematic.

In the larger context of the set of participating laboratories, the term B_i is not systematic, but completely random, and its variance is the *interlaboratory variance* σ_L^2. The variance of the measurements in the larger context is the sum $\sigma_r^2 + \sigma_L^2$, which is given the name of *reproducibility variance*.

2.2. Transposition of ISO 5725 concepts to SE characterization in instruments: inter-instrument experiment

The error model represented by (1) can be readily applied to ordinary measurements performed by ordinary instruments, e.g. digital oscilloscopes. Instead of examining a standard measurement method described in a written document, the object of the analysis is a given instrument (identified by manufacturer and model number) in given operational conditions. The index i identifies the *instrument*, instead of the laboratory; likewise, *intra-instrument* and *inter-instrument* variances are evaluated; and so on. This way of operating may be called "inter-instrument experiment", on the analogy of the ISO 5725 locution.

Since a laboratory does not own, usually, a big number of identical instruments, interlaboratory collaboration is necessary. The difference, with respect to ISO 5725, is that a single laboratory contributes for many values of the index i, if many identical instruments are available in the laboratory.

A specific note about the error terms B_i (bias of the i-th laboratory) and b (bias of the measurement method) is appropriate. In an inter-instrument experiment, the term B_i is of course the SE of the i-th instrument, while b is an "essential" bias inherent in the instrument design. The GUM suggestion to model the unknown SE of an instrument as a zero-mean RV, typically with uniform distribution, is equivalent to supposing $b=0$, and B_i uniformly distributed. A properly conducted inter-instrument experiment, therefore, allows one to test experimentally these hypotheses, possibly determining a nonzero value for b, and a different distribution (e.g. Gaussian) for B_i. In general, an inter-instrument experiment is the tool to verify manufacturer's specifications on SEs of an instrument, or to write down new, more accurate specifications.

2.3. Basic scheme of an inter-instrument experiment

The adopted basic experimental setup to perform an inter-instrument experiment is depicted in Fig. 1.

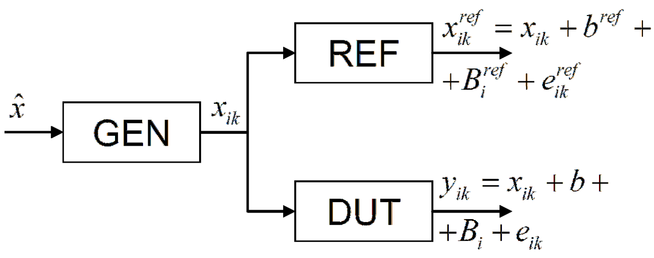

Fig. 1. Basic setup for an inter-instrument experiment. The tested instrument is the device-under-test DUT; GEN is a (non-ideal) generator, while REF is a reference instrument.

GEN is a non-ideal generator (for example, a voltage generator), set to produce a nominal voltage $\hat{x}$. Since it is non-ideal, in repeated measurements it produces the unknown voltages x_{ik}, being *i* the instrument index and *k* the measurement index. The instrument under test DUT yields the measurements $y_{ik} = x_{ik} + b + B_i + e_{ik}$, while a reference instrument REF yields the measurements $x_{ik}^{ref} = x_{ik} + b^{ref} + B_i^{ref} + e_{ik}^{ref}$. REF is a good reference if it is possible to approximate

$$b - b^{ref} \cong b,\ B_i - B_i^{ref} \cong B_i,\ e_{ik} - e_{ik}^{ref} \cong e_{ik}. \qquad (2)$$

The possibility to use these approximations can be inferred from the manufacturer's specifications of DUT and REF.

Instead of considering the measurements y_{ik}, like in the ISO 5725, it is convenient to consider the differences (errors)

$$E_{ik} = y_{ik} - x_{ik}^{ref} = b - b^{ref} + B_i - B_i^{ref} + e_{ik} - e_{ik}^{ref} \qquad (3)$$

that is

$$E_{ik} \cong b + B_i + e_{ik} \qquad (4)$$

The quantities E_{ik} can be treated exactly as the measurements y_{ik} in the ISO 5725 norm, being $\hat{x}$ the (approximate) level relevant to the measurements. This scheme of operations is justified by the fact that, in many kinds of common measurements, reference instruments (e.g. voltmeters) are more widely available than reference generators. If reference generators are available in the collaborating laboratories, it is possible to use the simpler scheme of Fig. 2.

It is important to note that (4) is applicable not merely to the absolute error affecting a single generated value, but also to errors with complex definition, more frequently employed in instrument characterization. In digital instruments (a typical case), gain, offset and nonlinearity errors are usually of primary interest [10]. If, for example, the gain error is of interest, each gain error measurement E_{ik} is obtained by using many measurements x_i^{ref}, y_i from DUT and REF, obtained from different values of $\hat{x}$. The obtained measurements are modeled by (4), where e_{ik} will represent a residual random error and $B_i + b$ is the actual gain error of the instrument, being B_i a zero-mean component and *b* a fixed error component, inherent to the instrument design.

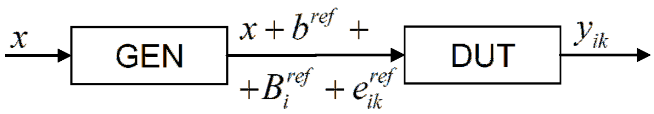

Fig. 2. Basic setup for an inter-instrument experiment, if a reference generator is available.

3. PRELIMINARY EXPERIMENTAL RESULTS

Actual experiments have been performed using the following instruments:

DUT: Agilent digital oscilloscope, model 54600B;
REF: Agilent digital multimeter, model 34401A;
GEN: Agilent generator, model 33220A;

From the specifications in the manuals, it is readily verified that the measurement uncertainty of REF (a 6 ½ digit multimeter), if correctly used, is negligible with respect to that of DUT (an 8-bit digital scope).

The experimental setup has been used to characterize the gain error of the instrument, defined on the basis of a straight line passing by two test points. Tab. 1 summarizes the instrument settings in the gain error measurements. The only relevant information available from manufacturer's specifications is that its value should be within ±1.9%. Therefore, without further investigation, one should suppose for the gain error a zero-mean uniform distribution with the same symmetric limits, or, in other terms, $b=0$, B_i with symmetric uniform distribution.

Table 1. Basic instrument settings in the gain error measurement.

Vertical range of the oscilloscope	[−4, 4] V
Test point 1	−3.6 V nominal
Test point 2	+3.6 V nominal
Vertical range of the multimeter	[−10, 10] V

Each gain error measurement has been obtained by averaging many measurements for each test point. This was done because in this experiment the quantities of interest (the systematic ones) are b and B_i, while the terms e_{ik}

(essentially due to noise) must be considered actually "random". For each instrument the gain error measurement must be repeated many times consecutively, and reproduced in many different instruments. In this preliminary phase of the work, measurements have been reproduced on two separate instruments only. Tab. 2 summarizes the data relevant to the actual number of measurements and of instruments involved.

Table 2. Parameters relevant to repeating/reproducing the measurements.

Number of measurements taken by each DUT per each gain error measurement and per each test point	10 000
Number of measurements taken by each REF per each gain error measurements and per each test point	5
Number n of repeated gain error measurements for each DUT	$k = 1, \ldots, 4000$
Number p of different DUTs considered	$i = 1, 2$
Time duration of the test	14 hours

Fig. 3 shows the result of this inter-instrument experiment. It must be highlighted that the measurements started immediately after powering on the oscilloscopes, since it was of interest to observe gain errors during the warm-up of the instruments. The figure shows that the drift of the gain error has the same pattern in the instruments, which is perfectly consistent with the fact that they worked in identical environmental conditions (particularly, at the same non-controlled temperature).

The plots in Fig. 3 can be clearly divided in at least three pieces with different characteristics, or three phases:

1) the warm-up phase, lasting about 50 minutes, in which the error has a fast negative slope;

2) a "quiet" phase, covering the period of approximately 9 hours (from 20.30 to 9.30), in which the error is substantially stable on a value (which is different, however, for the instruments);

3) a "variable" phase, covering a period of approximately 4 hours, in which the error is not stable on a single value, but varies, following similar patterns in separate instruments.

Of course the quiet phase seems to be associated to the constant environmental conditions occurring during the night, and the variable phase seems due to the daytime activities in the laboratory (in the same hours the laboratory was being intensively used). An average user should be more interested to the results relevant to the variable phase, which was obtained in conditions similar to those in which instruments normally operate.

Fig. 4 shows an histogram of the gain error measurements obtained during the variable phase. Of course the two Gaussian-like curves are each one relevant to one instrument. Fig. 5 shows a quantile-quantile plot of the same gain error measurements, relevant only to one of the instruments. From both the figures it is clear that errors relevant to a single instrument (although "systematic") are roughly Gaussian.

Fig. 4 is very helpful to understand a key question (and how to answer it). Assume that a user makes a measurement in normal working conditions, similar to those relevant to the variable phase (i.e. not in a perfecly controlled metrological environment), and that she/he does not know which particular instrument is being used. What is the probability that the measurement is affected by a gain error of a given value?

It is obvious that the answer should come exactly from the data in Fig. 4, which are relevant to the assumed conditions. This means to treat the systematic gain error as a random variable, with a Type A evaluation of the distribution. However, it is also clear that the bimodal nature of the actual distribution in Fig. 4 is not inherent to the instrument model, but is merely due to the limitation of having examined only two separate instruments.

Therefore, it is clear what the path to provide a correct answer to the question should be:

- test a larger number of instruments (the larger, the better), deriving curves similar to those in Fig. 3;
- repeat the measurements also for other values of the full-scale range;
- select the measurements relevant to normal working conditions;
- make an histogram and a statistical analysis of the kind illustrated by Fig. 4.

After having performed these steps, it is also possible to decide if the hypothesis $b = 0$ is acceptable, and if the hypothesis that B_i are normally distributed is acceptable. This, of course cannot be done on the basis of measurements on only two instruments. Therefore, the steps pointed out above may be regarded as the operative program leading to the final completion of the present work.

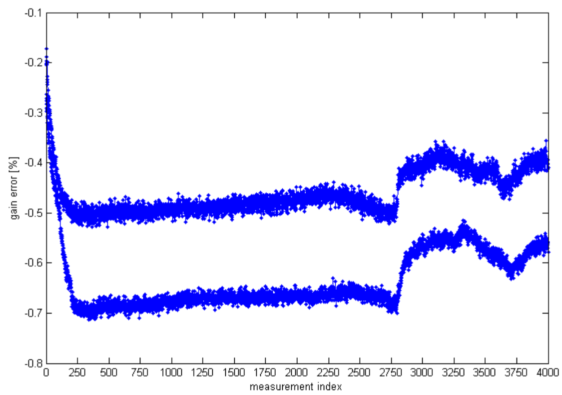

Fig. 3. Repeated gain error measurements for two identical oscilloscopes, operating simultaneously in the same laboratory for about 14 hours.

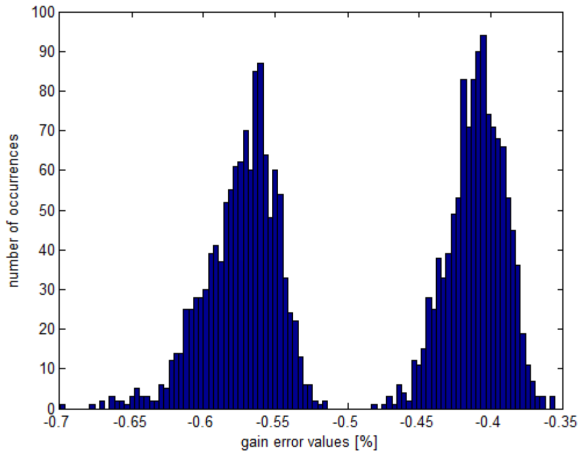

Fig. 4. Histogram of the gain error measurements in Fig. 3, taking into account only the data relevant to the “variable” phase.

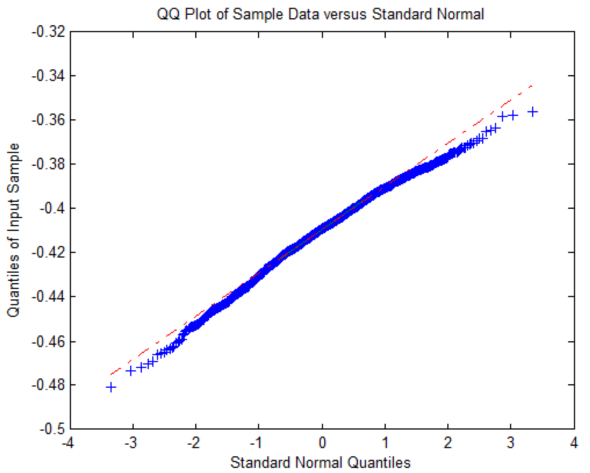

Fig. 5. Quantile-quantile plot of the gain error measurements in one of the tested instruments, taking into account only data relevant to the “variable” phase.

4. CONCLUSIONS

The paper depicts a theoretical and practical framework to obtain Type A evaluation of systematic errors affecting digital instruments. Preliminary results are presented, which must be completed with many more measurements, to be accomplished on many different instruments of the same model. This will possibly involve more laboratories in a networked team. Even if only partial experimental results are currently provided, the paper points out an approach to treat systematic errors and random errors in a perfectly homogenous way, both theoretically and experimentally.

REFERENCES

[1] B. N. Taylor and C. E. Kuyatt, *NIST Technical Note 1297 - Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. NIST, 1994. [Online]. Available: http://physics.nist.gov/ Pubs/guidelines/TN1297/tn1297s.pdf

[2] International Organization for Standardization, *Guide to the Expression of Uncertainty in Measurement (GUM)*, 1st ed. ISO, 1993, (withdrawn).

[3] ——, *ISO/IEC Guide 98:1995 - Guide to the Expression of Uncertainty in Measurement (GUM)*. ISO, 1995, (withdrawn).

[4] ——, *ISO/IEC Guide 98-3:2008 - Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)*. ISO, 2008.

[5] A. Ferrero and S. Salicone, “Modeling and processing measurement uncertainty within the theory of evidence: Mathematics of random-fuzzy variables,” vol. 56, no. 3, pp. 704–716, 2007.

[6] H. Wu, M. Siegel, R. Stiefelhagen, and J. Yang, “Sensor fusion using dempster-shafer theory [for context-aware HCI],” in *Instrumentation and Measurement Technology Conference, 2002. IMTC/2002. Proceedings of the 19th IEEE*, vol. 1, May 21–23, 2002, pp. 7–12.

[7] G. Shafer, “Perspectives on the theory and practice of belief functions,” *International Journal of Approximate Reasoning*, vol. 4, pp. 323–362, 1990.

[8] ——, “Rejoinder to comments on "Perspectives on the theory and practice of belief functions.",” *International Journal of Approximate Reasoning*, vol. 6, pp. 445–480, 1992.

[9] ——, “What is risk? What is probability? - Tutorial,” presented at the 10th International Conference on Information Fusion, Quebec City, Jul. 9, 2007.

[10] F. Attivissimo, N. Giaquinto, and M. Savino, “Worst-case uncertainty measurement in ADC-based instruments,” *Computer Standards & Interfaces*, vol. 29, pp. 5–10, 2007.

[11] E. Ghiani, N. Locci, and C. Muscas, “Auto-evaluation of the uncertainty in virtual instruments,” *IEEE Trans. Instrum. Meas.*, vol. 53, no. 3, pp. 672–677, Jun. 2004.

[12] International Organization for Standardization, *ISO 5725-1:1994 - Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions*. ISO, 1994.

[13] ——, *ISO 5725-2:1994 - Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*. ISO, 1994.

[14] ——, *ISO 5725-3:1994 - Accuracy (trueness and precision) of measurement methods and results – Part 3: Intermediate measures of the precision of a standard measurement method*. ISO, 1994.

[15] ——, *ISO 5725-4:1994 - Accuracy (trueness and precision) of measurement methods and results – Part 4: Basic methods for the determination of the trueness of a standard measurement method*. ISO, 1994.