MATHEMATICAL AND STATISTICAL TOOLS IN METROLOGICAL MEASUREMENTS

Franco Pavese
IMEKO Technical Committee 21 “Mathematical Tools for Measurements”
c/o Istituto Nazionale di Ricerca Metrologica (INRIM),10139 Torino, Italy

Keywords: Metrology, calibration, testing, statistics, data modeling, measurement, uncertainty, error.

Contents

1. Introduction
2. Metrology in Measurement Science
3. Modeling and Data Treatment: Mathematics, Statistics and Computational Tools
4. Specific Terms and Concepts in Metrology
5. Data Modeling in Metrological Measurements
6. Systematic Effects in Comparisons
7. Guidance to the Choice of a Data Model
8. Corrections as Treatment of Systematic Effects
9. Outliers
Glossary
Bibliography
Biographical Sketch

Summary

Measurement science is a very broad field, with already partial contributions to the EOLSS. In this Chapter the contents are restricted to the branch of measurement science called ‘metrology’, comprising both ‘calibration’ and ‘testing’, and concerning only quantitative measurements. However, even after these delimitations, the field remains extremely vast. In general, the aim of this Chapter is to supply the reader an introduction to the essential concepts and to direct them to a selected and ample bibliography. In addition, the reader is directed to the tables of contents of nine books, summing up to more than 360 papers and 3000 pages, dedicated over the last 20 years to the subject matter of advanced mathematical, statistical and computational tools in metrology can be found at http://joomla.imeko.org/index.php/tc21-homepage (AMCTM). In this chapter the reader is assumed to be acquainted with the basic terms used in measurement science.

1. Introduction

A measurement process in physical, chemical and biological science consists initially of experimental determinations of a value of the quantity intended to be measured, the measurand, generally requiring to employ some hardware (measuring instrumentation system).

In the preceding phase of planning the experiment, it is often required to build up a
model of the system subject to the measurement for several purposes: to ensure having a sufficiently complete picture of the measurand and understanding of its features; to determine the needs of the instrumentation; to possibly perform a simulation of the experiment, in order to detect possible problems and to optimize the instrumentation and its performance.

Subsequent to the phase of data acquisition from the instruments while performing the experiment, the acquired data will need analysis and the use of a data model in order to obtain the measurement results and an evaluation of its uncertainty, exploiting the full potential of the measurements.

The issues concerning instruments and hardware in general and the instrumental procedures are out of the scope of this Chapter. Also modeling of the systems, a very broad matter, will not be tackled in the following, except by providing some references. The contents will focus instead on data treatment, including data modeling, an issue requiring more and more sophisticated mathematical and statistical tools. Also sophisticated computer-aided tools are necessary today, which will also remain out of the scope of this chapter.

In this chapter double quotes are for literal citations, reported in italics, single quotes for highlighting a term used in this text. See the meaning of the terms and acronyms at the end of the Chapter.

2. Metrology in Measurement Science

Metrology is a branch of measurement science specifically devoted to achieve two main goals:

- At the top level of accuracy, a degree of ‘metrological traceability’ sufficient to obtain nationally and internationally ‘metrological compatibility’ at the needed level;

- At higher levels of uncertainty (lower levels in the metrological hierarchy), metrological traceability of the downward chain (the metrological chain, or pyramid) through an ‘unbroken chain of calibrations’.

In order to ensure the achievement and maintenance in time of these goals, the measurement standards are the competence of metrology, so are the measurements that disseminate the measurement units by means of operations called ‘calibration’ of standards of hierarchically low levels, including a statement about its uncertainty.

For the use of the International System of Units, SI, today adopted by the vast majority of countries all around the world, the BIPM constitutes the top of the metrological chain, followed by the National Metrology Institutes (NMI), and then the secondary laboratories accredited by the NMIs. The latter may be, in turn, organized in ranks at different hierarchical levels, down to the floor level of the workshops and individual users.
At the top level, metrology is a sophisticated and demanding science, as required by the most accurate determination of the values of the fundamental constants of the physical laws. Thus, the state-of-the-art knowledge is adopted or developed at the most sophisticated levels and beyond.

At lower levels, the metrological system relies on more and more simplified procedures and limited need for detailed knowledge, like it happens in most of the ‘testing’ field. In the latter context, the needs are usually covered by written standards, which are promulgated by ISO or similar international and national organizations; for the subject matter of this chapter, standards are mostly those provided by the ISO Technical Committee 69 (TC69).

The specificity of metrology with respect to the general field of measurement science can have a critical effect on the meaning of some basic concepts or on the methods preferred for data treatment.

In particular, some peculiarities of the measurement results more critically affect the decisions to be taken based on them. In this respect, after a general treatment of other aspects in metrology, requiring a careful selection of suitable mathematical and statistical tools, this Chapter will tackle in details the issue of the systematic effects that affect all experimental determinations, and are responsible for the possible disagreement of the results that, when occurring in different laboratories, affects the degree of compatibility and, consequently, the integrity of the metrological chain.

3. Modeling and Data Treatment

When top accuracy is aimed at, the experimenter needs to get acquainted with all the known aspects of the physical (chemical, biological) system that will be subjected to measurement.

Based on this knowledge, the experimenter lists the quantities that need to be measured, and those that must be quantitatively evaluated, in both cases because they have an influence on the (numerical) measurement results. They are called in fact ‘influence quantities’.

The required type of “measurement process” (clause 2.1 in (VIM3)) is different depending on the fact that one needs to measure an object (e.g. a mass standard), or an immaterial property (e.g. time), or a functional relationship (e.g. a physical ‘law’) between two or more quantities. Similarly is different when measuring an “extensive” or an “intensive” property.

In addition, measurements are usually grouped into two broad categories:

(a) “direct”, when the kind of quantity (VIM3) that is measured and that of the measurand intended to be measured coincide (e.g. mass for a mass standard):

(b) “indirect”, when the measured quantity is different from the aimed measurand (e.g. electrical resistance for a temperature standard). In the latter instance, the case can be multivariate, when the involved quantities are more than one (see also Section 6.3.1 and 6.3.2)—“corrections” (see Section 8) are excluded from
this distinction.

In both cases, one refers to one, or a few respectively, of the influence quantities. Invariably, the influence quantities are many more, the remaining being normally labeled ‘corrections’. This distinction can cause confusion in some circumstances—see Section 8.4.

In order to proceed, the experimenter needs to ‘model’ the different aspects of his knowledge.

A model can be implicit, as simple as a list of the input quantities:

\[ f(Y, X_1, \ldots, X_i, \ldots, X_N) = 0 \]  \hspace{1cm} (1)

where \( Y \) is the outcome of the measurement process, the output quantity, intended to realize the measurand, and the \( X_i \) are the input quantities (see Section 8 for details on this model in (GUM)). This is called the “measurement equation” (VIM3, GUM) and is a relationship between variables, which are random variables whose realizations are the experimental observations. They are written in capital letters, the usual notation convention; the observations are instead written using the same lower-case letter, and an additional subscript index is added in the usual case of replicated measurements: for example, a replicated realization of \( X_i \) is written \( x_{i,j} \)—the term “replicated” should not be confused with the term “repeated”.

Not necessarily all the \( X_i \) are subject to measurement in the experiment in question. For some—or even most—of them the values can be obtained from prior information, typically from the literature, from calibration certificates, or in other ways.

Not infrequently, in the case that an implicit model is used, the effect of an influence quantity included in the list of corrections is given as a “sensitivity coefficient” (GUM), i.e. as the value \( dY/dX_i \) in the specific range of interest. In addition, this value is often not obtained from a relationship expressed in analytical form, but from a sub-experiment consisting in varying \( X_i \) and observing the variation in \( Y \). However, from this subset of observations an empirical law can be subsequently drawn, e.g. a linear relationship or a more complicated one.

Obviously an implicit model cannot be used if the model is intended for studying the behavior of the measurand when the influence quantities change, as required in simulation studies, either performed analytically or using computed-aided methods like Monte Carlo simulation. In these cases the model must be written in an explicit analytical form, i.e. expressed by specifying the analytical relationships between the quantities. Often this results in complicated implicit expressions, e.g. differential or integral equations, not solvable in closed analytical form, but requiring numerical solutions, e.g., by the use of a finite-elements method. The illustration of this type of models, though not infrequent also in metrology, is out of the scope of this Chapter. The reader can consult the modeling literature, since the features for the models developed
for metrological applications are not specific, except for the concept of uncertainty—and the requirement for its evaluation. For modeling in metrology the reader can consult, e.g. (Sommer 2008), together with the involved computational tools (Steele and Douglas 2008a).

The model for data treatment differs from the one in (1) in some features. Instead of a model of the variables of the system, the data model is called ‘observation model’, and is written instead for the observations, or ‘input estimates’ \( x_{ij} \) of an influence quantity and ‘output estimate’ \( y \) of the output ‘realized quantity’, which are realizations of the respective random variables \( X_i \) and \( Y \), i.e. are members of the respective probability distributions.

Measurement and data models are based on statistical concepts, of critical importance in measurement science, and metrology in particular, which include modeling of uncertainty. They are treated in detail in Section 5. In modern treatments the related computational tools are very important also in data treatment and analysis.

4. Specific Terms and Concepts for Metrology

Before proceeding with the methods and associated tools to be used for exploiting the treatment of metrological data, let us clarify the meaning in metrology of some of the basic term: the Glossary will refer to Section 4 for these terms. In fact, terms are simply a synthetic way to express concepts. It is therefore essential to be acquainted with them and understand the underlying concepts, because the semantic of the terms can change in time and with the language, leading to possible ambiguities or misunderstandings. For this reason, the most recent Vocabulary of Metrology (VIM) includes also “Basic and General Concepts” (VIM3), and will be used as the basis throughout this Chapter.

4.1. Measurand

The measurand is defined as the “quantity intended to be measured”, because it may happen that a measurement fails the goal to provide a measure of the aimed quantity (VIM3).

The definition requires understanding the concept of “quantity”, indicated as “property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed as a number and a reference”. Therefore, this definition is valid only for quantitative measurements, and “a reference can be a measurement unit, a measurement procedure, a reference material, or a combination of such”.

What is not explicitly said above is that any measurand can only be defined by a finite list of details. Therefore, should the random component of a measurement even be reduced to almost nil, the uncertainty on any measurement result cannot decrease below what is called the “definitional uncertainty”. This residual uncertainty is not due to neither random nor systematic effects, the two traditional components of uncertainty (see later). The definitional uncertainty is not of statistical nature, but is due to the multiplicity of the possible definitions of the measurand, something that should better
be called the “non-uniqueness” of the definition. Though amply used in metrology (White et al. 2009 and 2010) this term is not (yet) included in VIM (VIM3) nor in GUM (GUM).

In top-level metrology, the non-uniqueness emerges very clearly in many fields as the limiting factor, and, in fact, only the addition of further specification factors to the properties of a measurand can allow further progress. A couple of examples are: the definition of the unit of time, the second, where it is now specified that “In this definition it is understood that the Cs atom at a temperature of $T = 0 \text{ K}$ is unperturbed by blackbody radiation”; the added specification of the isotopic composition of a substance in thermal metrology.

Measurement results aim at accurately representing the value of the measurand. However, although the measurement result is conceptually different from the value of the measurement, “for any scalar quantity, subtleties arise only when its uncertainty is not symmetric about that quantity’s reported value (Douglas et al. 2005), so usually no distinction needs to be made between the distribution of measurements (or ‘gedanken’ measurements) and the distribution of the measurand (to which formal uncertainty distributions refer). The formal justification for this is facilitated by the standard practice for metrologists of using the same value to be the best representation both of the (fully corrected) measurement, and of the measurand. This ‘fiducial value’ (Wang and Iyer 2006, Guthrie et al. 2008, ISO12) simplifies, and in our view strengthens, the fiducial argument” (Steele and Douglas 2008a).

4.2. Realized Quantity

The GUM [GUM] introduced the concept of “realized quantity”: “Ideally, the quantity realized for measurement would be fully consistent with the definition of the measurand. Often, however, such a quantity cannot be realized and the measurement is performed on a quantity that is an approximation of the measurand … Neither the value of the realized quantity nor the value of the measurand can ever be known exactly; all that can be known is their estimated values”.

4.3. Compatibility

Since the basic aim of metrology is to ensure as much as possible consistent results from independent measurements for the same measurand, compatibility is a pivotal requirement to be achieved. “Metrological compatibility” is defined as “property of a set of measurement results for a specified measurand, such that the absolute value of the difference of any pair of measured quantity values from two different measurement results is smaller than some chosen multiple of the standard measurement uncertainty of that difference” (VIM3).

Obviously, the definition bound depends on a “chosen multiple of the standard measurement uncertainty”, where the latter is the standard deviation, a basic statistical parameter related to the second moment of a probability distribution. The choice of the multiple is left to the contingent decision, since it can be different for different circumstances to fit the purpose. Clearly here the “differences” in question are not those
between measurement results caused by the variability due to random effects, but are those arising from systematic effects (see Section 6).

The importance of compatibility is so great that in the prescriptive document signed by the metrological Institutes to mutually recognize the validity of their calibrations (MRA) (MRA), it is specified: “If, as a result of a key comparison, a significant unresolved deviation from the key comparison reference value persists for the standard of a particular participating institute, the existence of this deviation is noted in Appendix C. ... In this case, the institute has the choice of either withdrawing from Appendix C one or more of the relevant calibration and measurement services or increasing the corresponding uncertainties given in Appendix C”, where Appendix C is the database of the calibration services of the NMI (the expression “significant unresolved deviation from the key comparison reference value” means ‘an inconsistency with respect to’).

4.4. Traceability

This term is defined as the “property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty” (VIM3). It “requires an established calibration hierarchy” and, since each calibration contributes to measurement uncertainty, the latter always increases while descending the ladder of metrological hierarchy. In metrology the term does not only mean ensuring to trace back to the origin by knowing the path, but has also a quantitative qualification.

In principle, traceability does not mean that a step of the metrological ladder should necessarily be compatible with the next. However this is desirable.

Two National Metrology Institutes cannot be said to be traceable with each other, since there is no hierarchy among them (for the standards not obtained from another NMI). They ensure, at the top of the “traceability chain” (VIM3), the relationships between them called “degree of equivalence” (MRA), by means of inter-comparisons, formally setup in the frame of the MRA, called “key comparisons”—see Section 4.7.

4.5. The Error and Uncertainty Approaches

An underpinning basic concept of science, and hence of measurement science, is that, due to imperfect knowledge of the observed phenomena, the numerical data that are the outcomes of measurement are affected by errors. Irrespective of the reasons that are the causes of these errors, the resulting dispersion of the measured numerical values that is generally observed is interpreted as evidence of the imperfect knowledge.

Thus, the dispersion of the measured values introduces an uncertainty in the measure of the observed phenomena. Uncertainty associated with data is specified according to models that are different according to the underpinning assumptions, which must adequately match the characteristics of the observed phenomena or measurement process.
Uncertainty cannot be determined, but only estimated with a certain ‘confidence’ or ‘degree of belief’. This estimate is, together with assigning values to measurement results, the most fundamental aim and task of metrology.

The definitions of uncertainty and their use are studied since the times of Gauß and are the subject matter of statistics. Information in this respect can be found also in several chapters of the EOLSS, e.g. (Viertl 2003a, 2003b, 2003c, Mari 2012). A restricted set of them is also used in metrology, e.g. (Rossi 2008). For example, probability theory is almost exclusively considered, while possibility theory is not (nor data “imprecision” or “measurement inexactness”); interval-related statistical techniques are not either (Kreinovich 2008). On the other hand, different specific approaches have been developed, namely in GUM (GUM).

Bibliography


Allevard T., Benoit E., and Foulloy L. (2004). Signal processing on fuzzy nominal scales, in Proceedings 10th IMEKO TC7 International Symposium, Saint-Petersburg, Russia, pp. 478–483. [This paper reports a good example of use of fuzzy statistics]

American Association for Laboratory Accreditation (A2LA) (2002) Guide for the estimation of measurement uncertainty in testing. [This is a reference Guide provided by one of the National Associations for Accreditation]

Ballico M. (2001). Calculation of key comparison reference values in the presence of non-zero mean uncertainty distributions, using the maximum likelihood technique, Metrologia 38, 155–159. [This paper is one supporting Approach A, Section 5.2.1]


Chunovkina A.G., Chursin A. (2001). GUM and MRA: some problems of data processing and measurement uncertainty evaluation, in Mathematical and Computations Tools in Metrology V (AMCTM V), P. Ciarlini, M.G. Cox, E. Filipe, F. Pavese and D. Richter, eds., (Series on Advances in mathematics for applied sciences) 57, Singapore: World Scientific Publ. Co., pp. 55–66. [This paper is about problems in using a ‘mixture’ distribution, Section 6.3.3]

Chunovkina A.G. (2003). Determining reference values and degrees of equivalence in key comparisons,
**Measurement Techniques** 46, 406–414. [This paper is about problems in using a ‘mixture’ distribution, Section 6.3.3]

Chunovkina A.G., Ciarlini P., Cox M. and Pavese F. (2004). Handling and treatment of measurement data from different groups or sources, in Proceedings 10th IMEKO TC7 International Symposium, pp. 176-181. [This paper is one supporting Approach A, Section 5.2.1]

Chunovkina A.G. and Cox M.G. (2003). A model–based approach to key comparison data evaluation, in Proceedings 17th IMEKO World Congress, Cavtat, Slovenia, pp. 1240–43. [This paper is one supporting Approach B, Section 5.2.2]


Ciarlini P., Cox M.G., Pavese F., Regoliosi G. (2004). The use of a mixture of probability distributions in temperature interlaboratory comparisons, *Metrologia* 41, 116–121. [This paper is an improved follow up of the previous]


Cox M.G. (2002). The evaluation of key comparison data. *Metrologia* 39, 589–595. [This paper contains a proposal for the evaluation of the key comparison data]


Cox M.G. (2007). The evaluation of key comparison data: determining the largest consistent subset, *Metrologia* 44, 187–200. [on the same subject, this paper proposes to screen out inconsistent data for the computation of the KCRV]


DIN: DIN 1319-1 1995 *Fundamentals of metrology Part I: basic terminology*, Berlin, Beuth, [A basic standard from the German Standardisation Organisation]


Duewer D.L. (2007). How to combine results having stated uncertainties, in *Combining and Reporting*


EA03: European Accreditation 2003 EA guidelines on the expression of uncertainty in quantitative testing, EA(4/16, rev00). [A guide on uncertainty in testing issued by the European Accreditation Organisation]


Elster C. and Toman B. (2010). Analysis of key comparisons: estimating laboratories’ biases by a fixed effects model using Bayesian model averaging, Metrologia 47, 113–119. [A Bayesian approach to the issue]


Giacomo P. (1982). News from the BIPM, Metrologia 18, 41–44. [Introductory document on INC-80, the precursor of GUM]


Chapter


Lira I. and Wöger W. (2006b). Comparison between the conventional and Bayesian approaches to evaluate measurement data, *Metrologia* 43, S249–S259. [Same argument as above]


MRA: CIPM 1999 Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, Bureau International des Poids et Mesures, Sèvres. http://www.bipm.org. [Formal document to be signed by all NMIs who agree to recognize reciprocally their calibration certificates]


Pavese F. (2007b). The definition of the measurand in key comparisons: lessons learnt with thermal measurements - Franco Pavese


Pavese F. (2011). Methods for the assessment of correction for chemical-impurity effects and related uncertainty in ITS-90 fixed points, namely of $e$-H$_2$, Ne, O$_2$ and Ar, *Metrologia* 48, 268–274. [A paper about the problems in correcting measured values, in a specific domain]


White D.R., Bonnier G., Diril A., Arai M., Ballico M., Chimenti V., Duris S., Filipe E., Ivanova A.,


Biographical Sketch

Franco Pavese, graduated in Engineering, Principal Scientist (Research Director) at the National Research Council of Italy, Institute of Metrology “G. Colomnetti” (IMGC-CNR) from 1967 to 2005, then at the National Institute of Research in Metrology (INRIM) until 2008.

- Chairman of IMEKO Technical Committee 21 (TC21) “Mathematical Tools for Measurements” (2004–)
- Coordinator of several research Contracts with CERN (Geneva) on statistical treatment of LHC calibration data (1994–2007)
- Coordinator of iMERA/EURAMET Project “Softools” (2005–2009)
- Italian Delegate to ISO Technical Committee 69 “Applications of statistical methods” (2005–)
- Affiliated Scientist of Institute for Low Temperature Research, Poland
- Member of the Comité Consultatif de Thermométrie (CCT) (1980–2010) and occasionally Italian Delegate
- Chairperson of the CCT Task Group “Strategy” (2008–2012; past Chairperson of Working Group 7 “Key Comparisons”; Member of Working Group 3 “Uncertainty”
- Member of Commission A2 of the International Association IIF/IIR
- Titular Member, then National Representative, of IUPAC Commission I-1 (1991–)
- President of the International non-profit Society “The evitherm Society”, originated from the European Project EVITHERM (2006–)
- Coordinator of 3 EU Research Projects (1999–2008) and of several Italian National Research Projects on thermal metrology and cryogenic engineering
- Past Italian Representative in VAMAS, Area 15 "Cryogenic Structural Materials" and Area 16
• "Cryogenic Superconducting Materials"
• Editor of 13 multi-author International Books and Journal Special Issues.
• Member of the Editorial Boards of the: Journal of Chemical Thermodynamics (past), Journal Measurement and International Journal of Metrology and Quality Engineering.

Author of more than 20 Books, Monographs and book Chapters, of more than 200 scientific papers, most international, and of 3 patents, in the field of thermal metrology, cryogenic engineering, HTC superconductors and mathematics & statistics in measurement science.