

Role of measurement uncertainty in conformity assessment

Stéphane Puydarrieux^{1*}, *Jean Michel Pou*², *Laurent Leblond*³, *Nicolas Fischer*⁴, *Alexandre Allard*⁴, *Max Feinberg*⁵, and *Driss El Guennouni*⁶.

¹ Orano cycle La Hague, 50444 Beaumont Hague, France

² DELTAMU 63800 Cournon d’Auvergne, France

³ PSA Groupe, France

⁴ LNE, 78197 Trappes, France

⁵ retired from INRA, France

⁶ DGA essais propulseurs, France

Abstract. Conformity assessment, focused on risks quantification, evolved from being binary to being probabilistic. In fact, new methods described by the standard ISO CEI guide 98-4 [1] and FD X 07-039 [2] show how measurement uncertainty is integrated in the conformity assessment (FD X 07-039 : “Role of measurement uncertainty in conformity assessment – implementation of NF ISO/IEC Guide 98-4 – illustration through industrial case studies”). In such a probabilistic framework, the conformity assessment relies on two quantities: the measurement with its related uncertainty and the measurand prior knowledge. Both quantities are combined statistically to determine measurand posterior knowledge, and to quantify a global or specific risks either for the producer or the consumer. The standard also defines how integrating those two quantities improves the reliability of decisions based on measurements. This work will first present a clarification of the probabilistic approach described by the FD X07-039. Then, the measurand prior knowledge use will be explained. Finally, future metrology applications caused by such an approach will be listed.

* Stéphane PUYDARRIEUX: stephane.puydarrieux@orano.group

Preamble:

The J.C.G.M (Joint Committee for Guides in Metrology), is an organization of the BPIM (Bureau International des Poids et Mesures) which aims at maintaining and promoting general documents dealing with metrology. It is composed of two Working Groups. The first one (JCGM-WG1) is in charge of the GUM and its supplements, while the second one (JCGM-WG2) is in charge of the VIM (International Vocabulary of basic and general terms in metrology).

In 2012, Working Group 1 published the document JCGM 106:2012 dealing with the role of measurement uncertainty in conformity assessment. This document was considered by the ISO as the international standard ISO/IEC GUIDE 98-4 [1]. Then, in 2013, France has translated it and published the new French version NF ISO/IEC GUIDE 98-4.

Considering the new concepts of metrologist practices given by this document and the resulting mathematical difficulties, the AFNOR commissions Metrology X07b and Statistic X06E have built a working group in order to write a document explaining and clarifying the international standard using illustrations and examples of some applications.

In 2018, AFNOR published the results of its reflexions in the form of the document FD X 07-039 [2].

The working group wished to take advantage of the International Congress of Metrology 2019 and to present these documents as they should acutely modify the decision making process in the measurement field in the coming years....

1. Nothing new at the beginning...

The GUM has been established in the industrial landscape since its publication in 1995. Awareness of the need to evaluate the measurement uncertainties is increasing, although this is a slow process. This acceptance of measurement uncertainties is difficult, especially because they highlight the existence of risks in decision making that the measurements induce.

“We don’t measure for the pleasure of measuring, but to know!” In the context of conformity, we measure to know how a given property (the measurand) is positioned in comparison with its tolerance limits.

It is not the measurement of diameter of the cup that will make it ensuring its function, but the true value of the diameter. In the same way, it’s not the measured value of the concentration of a marker in a patient’s blood that makes him sick or not, but the real value of this concentration.

Obviously, all the metrologists know that there is a difference between reality (the true value) and the measured value. In a way, this difference is translated by the fundamental relationship of metrology:

$$\text{Value}_{\text{Measured}} = \text{Value}_{\text{True}} + \text{Error}_{\text{Measurement}} \quad (1)$$

The main difficulty is that neither the measurement error nor the true value can be perfectly known. Since the GUM, metrologists have focused on evaluating the term “Error Measurement” which must be considered, for a given measurement, in the statistical sense, as a realization of a random variable. The dispersion of this random variable is then called “measurement uncertainty”.

For at least thirty years, metrologists have been trying to evaluate this probability distribution by different methods (GUM, GUM S1, inter-laboratory tests, etc.). Finally, they express the measurement result by combining two pieces of information: the measured value and the measurement uncertainty.

The measurement result is often represented as follows:

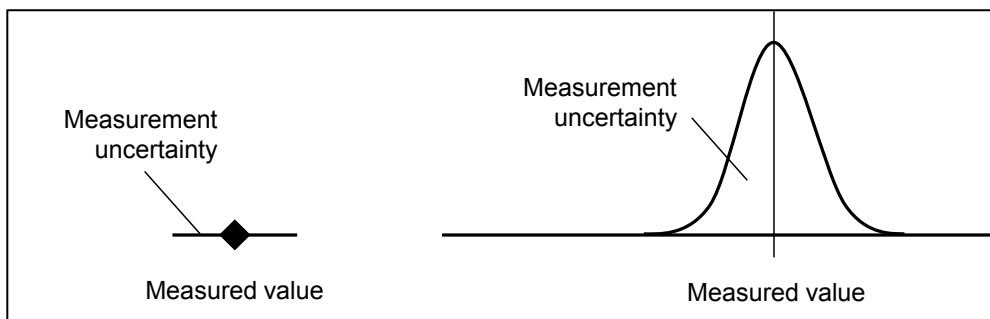


Fig. 1 : Traditional vision of measurement result

It should be noted in the Figure 1 that the measured value is positioned in the center of an uncertainty interval, as if the measured value (often a single measurement) was systematically the most likely value of the measurand. It only takes a few repetitions to understand this is rarely the case. Therefore, this representation does not comply with equation (1) which involves that, in the center of the interval, we find the true value. Concerning the measured value, it only belongs to the interval (if there is no measurement bias or after the correction of the bias).

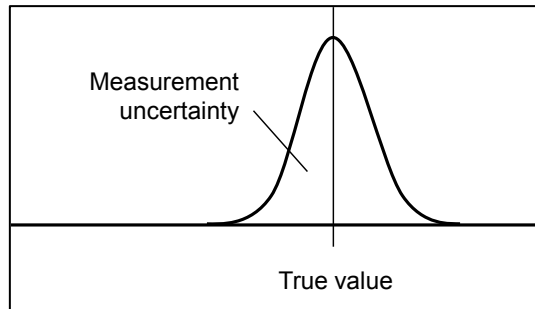


Fig. 2: Representation of the uncertainty

These figures introduce an important notion: the space of the measured values is not the space of the true values. These spaces are neither stackable nor interchangeable. The measured values allow acquiring an uncertain knowledge of the true values, and this relationship can be represented by a two-dimensional graph (Figure 3), the X-axis representing the space of the true values, the Y-axis representing the space of the measured values:

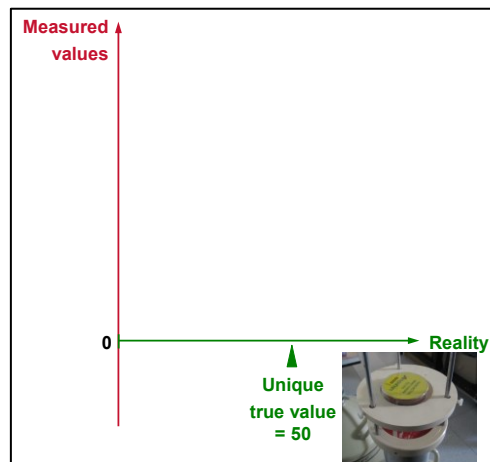


Fig. 3: Space of the true values and space of the measured values

In a perfect world, without bias, without uncertainty, the measurement process can be represented by the bisector of the orthonormal reference mark $\overline{x, 0, y}$. In this context, the measured value is exactly the true value by projection through the measurement process. The knowledge of reality is then perfect: $Y=X$, the true value is equal to the measured value.

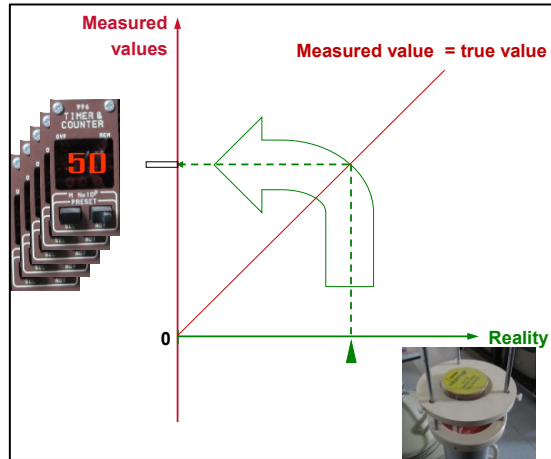


Fig. 4: Measured values with a perfect measurement system

Unfortunately, the measurement process is not perfect, so this bisector introduces a blur. Therefore, the measured value only provides an estimate of the true value and this estimate has an area of uncertainty...

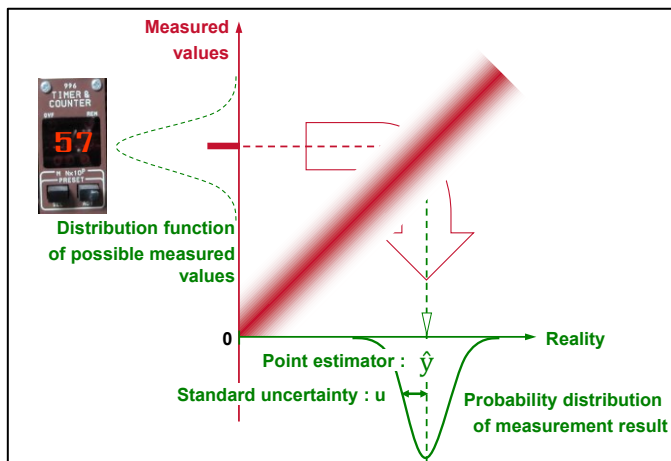


Fig. 5: From measured values dispersion to uncertainty

Therefore, the standard invites us to consider not only the measured values and associated measurement errors, but also the knowledge of reality, i.e. the measurand itself. There is often knowledge available of the measurand prior to the measurement. Which metrologist, in fact, has never said to himself “well, this value is strange”? Even the fact of being questioned by the supposed oddity of the measure indicates that he expected something else, that he already had an idea of the possible measured values for this measurand. This knowledge of the reality of the measurand prior to the measurement is called, in Bayesian terminology, prior knowledge.

It is this prior knowledge that the standard proposes to formalize and use to make more rational decisions than those based only on measured values, without taking into consideration what can be known elsewhere... The quantification of the prior knowledge,

or even of prior distributions is not straightforward and will become one of the new challenges of the mission of the metrologist... this will be his main added value.

In the simplest case of an industrial company that produces entities (vaccines, bolts, etc.) in series, the prior information can be found, for example, by analyzing the control charts of the processes under surveillance. In the case of some complex processes, the expected value is determined by computation codes, and confirmed by analyses or measurements. In the case of inspection laboratories upon receipt of certain products, the expected value is given by the supplier. In the case of calibration laboratories, prior information on a particular instrument can be obtained by analysing the historical calibration data. For a medical biology laboratory, the prior information on red blood cell or platelet counts in the blood may be the matter of the genetic heritage of the “client” populations, healthy subjects or not ...

The determination of this prior distribution may seem, and sometimes it is, complex. It is nevertheless crucial: on the basis of this prior knowledge, and the measured values, the Bayesian approach proposes to determine the posterior distribution, which is a statistical “combination” of the information contained in the prior knowledge AND in the measurements. This calculation is the basis of the standard.

It is possible to summarize the construction of the posterior distribution as follows:

- The true value, unique and unknown, is characterized by the prior distribution, in the space of true values.
- The measured value for this true value is characterized by the distribution of possible measured values in the measured value space.
- The true value is estimated using a posterior distribution, a Bayesian combination of the prior distribution and the distribution of possible measured values.

The interest of this posterior distribution is clearly visible in Figure 6: this posterior distribution is “better” than the “best” of both distributions used to build it: the uncertainty is reduced.

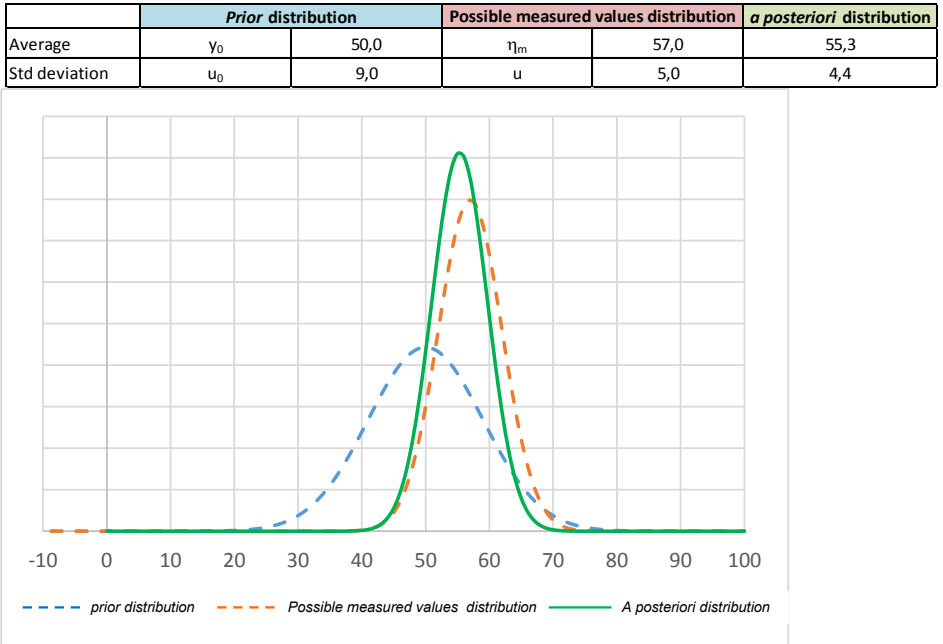


Fig. 6: Prior distribution, probability distribution of possible measured values and a posteriori distribution

2. From uncertainty to risks, in the plural!

When the decision is made, this doubt generates a risk of an incorrect decision:

- The **consumer’s risk**: the cap is declared conforming when in fact it is not.
- The **producer’s risk**: the cap is declared non-conforming when in fact it is.

In current representation these risks are schematized wrongly as follows:

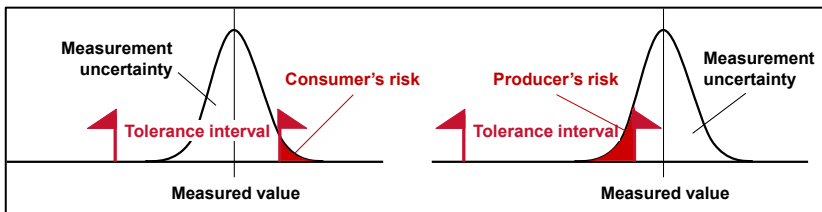


Fig. 7: Consumer’s and Producer’s risks according to the traditional vision.

The closest representation to these risk situations proposed by NF ISO/IEC GUIDE 98-4 standard and FD X 07-039 is given in two dimensions by:

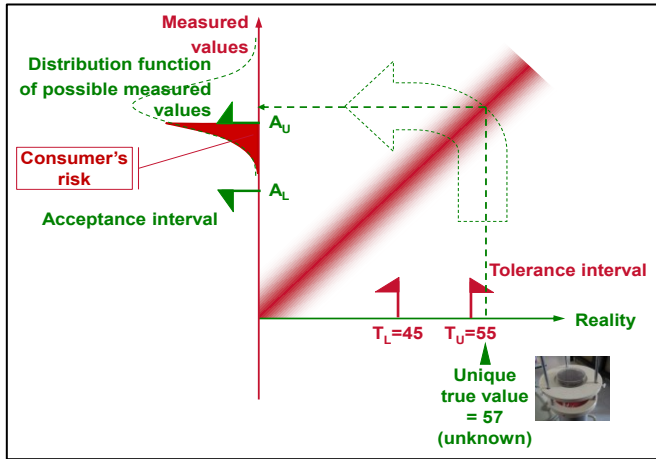


Fig. 8: Consumer's risk according to the FD X 07-039 vision.

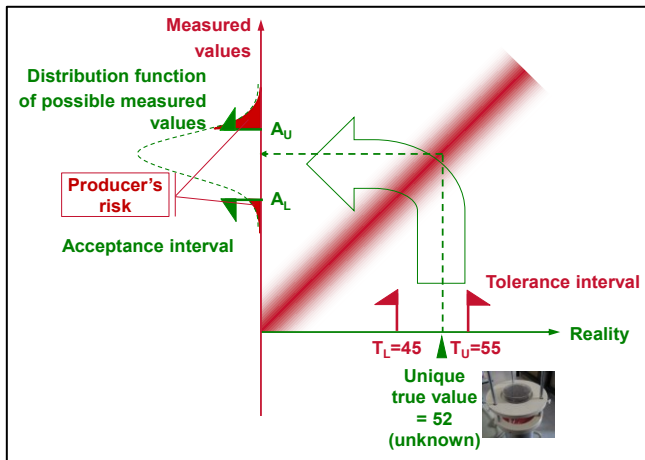


Fig. 9: Producer's risk according to the FD X 07-039 vision.

For the existence of consumer's risk, the real value must be non-conforming and the measured value must be “seen” in conformity. However, and even if the measurement uncertainties are significant, the consumer's risk is always equal to 0 if there is no true non-conforming value!

Then, we understand the risk calculation should involve two elements:

- The probability that the true value is conforming or not.
- The probability of seeing a given true value in conformity or non-conformity.

Before discovering the other declinations proposed by the standard, we can wonder about these meanings. It is easy to see that these two risks are, in many cases, assumed by the end consumer. The producer's risk, which leads to retouching and repelling the conforming entities, necessarily increases the global cost of the entities. It is therefore the end consumer who pays...

The standard introduces two new concepts: the global risk and the specific risk. For reasons of clarification, FD X 07-039 has allowed itself some adjustments and proposes the following definitions:

- Global risk: the global risk represents the proportion of entities that would be subject to a decision error, i.e. non-conforming entities measured conforming (mistakenly accepted) or conforming entities measured non-conforming (mistakenly refused).
- Specific risk: the specific risk is defined as the risk that a decision error will be made for a particular measured entity.

3. From definitions to calculations

Since the global and specific risks are of very different natures, the calculation approaches are different too.

3.1. Global Risk

It makes it possible to estimate the rate of decision errors when all entities have been measured. Such a concept is particularly interesting for a medical biology laboratory or a calibration laboratory.

This would allow these laboratories to estimate, for a given period of time, the average number of wrong decisions. This is an interesting indicator to develop as part of contract reviews.

FD X 07-039 proposes to calculate the global risk using an original approach. The numerical simulation allows this calculation to be carried out from the following information:

- Probability distributions of the prior process
- Probability distributions of the measurement result
- Limits of the tolerance interval T_L and T_U
- Limits of the acceptance interval A_L and A_U

Numerical simulation makes it possible to fully understand the global risk issues. First, possible true values are drawn from the prior distribution, making it possible to know whether the entity is conforming to the requirements or not. Second, possible measured values are drawn from the measurement distribution (which depends on the simulated true value). Finally, it is possible to see whether VALUE true and VALUE true + ERROR Measurement (see (1)) lead to the same decision, or if a wrong decision has been taken, in one direction (Consumer's risk) or in another (Producer's risk).

The ability to distinguish the tolerance limits (T_L and T_U in the space of true values) from the acceptance limits (A_L and A_U in the space of measurements) makes it possible to introduce the concept of "guard bands". These guard bands enable to manage risks.

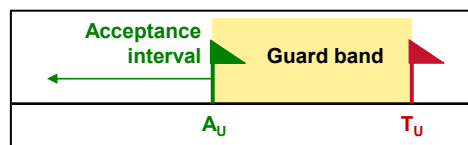


Fig. 10: Principle of guard band.

However, it should be noted that risks do not change symmetrically and that improving on one can result in significant losses on the other. It is not uncommon to choose counterproductive strategies by focusing on one without paying attention to the other. The Figure 11 presents a particular situation but illustrates the general case well and shows the comparative evolutions of the two risks according to the guard band.

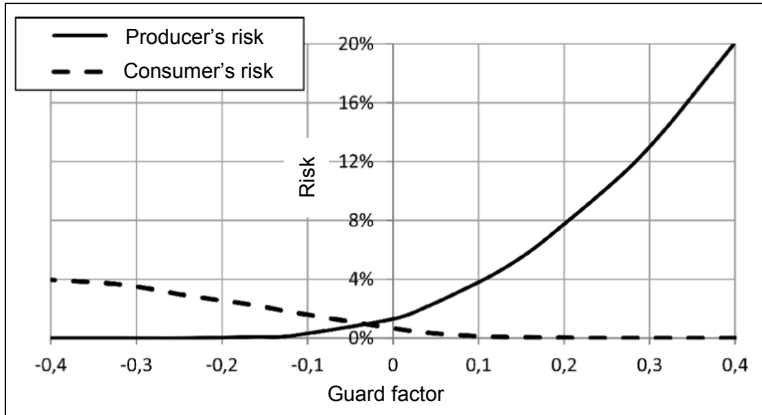


Fig. 11: Evolution of consumer's and producer's risks.¹

To take into account this reality, which means that consumer's and producer's risks are always assumed by the end consumer (or even the company as a whole), it is possible to determine guard bands based on an optimization of the weighted sum of customer and producer's risks. Information is provided in Appendix B of the standard.

3.2. Specific risk

The specific risk enables to estimate the probability of a decision error for a given entity. To estimate this risk, one use Bayes' theorem to estimate a posterior distribution, i.e. what one can know about the true value of a particular entity after making a measurement. The use of a posterior distribution can be summarized as follows: "Given what I know i.e. prior distribution and the measurement uncertainty u , if I find a measured value equal to X , then, the most probable true value underlying is equal to $Y \pm u$ ', not $X \pm u$ ".

¹ This graph represents the following particular situation:

- The probability law of the process is normal, mean 0 and standard deviation 1;
- The probability law of the measurement error is normal, mean 0 and standard deviation (standard uncertainty) 0.25;
- The tolerance is such that $T_U - T_L = 4$

On the other hand, the guard factor is defined as being the ratio: $1 - \frac{Au - Al}{Tu - Tl}$. So, the interval $[T_L, T_U]$ contains $[A_L, A_U]$ when the guard factor is positive and is included in $[A_L, A_U]$ when the guard factor is negative.

The decision is then made from the value of Y (average of the posterior distribution) and no longer from X (measured value). The risk of error, according to the decision taken from Y , is calculated by taking into account u' and not u . Finally, the posterior distribution which can be easily calculated when the involved phenomena are Gaussian, allows the traditional vision (Figure1) to be transformed into a more realistic vision.

4. Conclusion:

The standard NF ISO/IEC Guide 98-4 and the documentation booklet FD X 07-039 have paved the way for a change in decision-making process “within the metrological community.

- The measured value is not at the center of the measurement result,
- The measured value and the measurement result do not belong to the same space,
- The uncertainty is not “around” the measured value,
- The measurement result and the associated uncertainty are defined by the *a posteriori* distribution,
- This *a posteriori* distribution is defined by taking into account measured values, measurement uncertainty due to measurement system, and *prior* distribution, *prior* knowledge, and all pre-measurement knowledge of the measurand.

The metrologist in his company has now an opportunity to consider the prior distributions in order to estimate the measurement uncertainties and risks associated with decisions in a relevant way.

One can wonder why, till now, all these calculations have not been necessary. The answer is simple: in case of doubt, it is likely that we have had a natural tendency to tighten tolerances in order to limit consumer risk. However, as already written, both risks are assumed by the end consumer and their respective concomitant evolutions are not symmetrical; one lose a lot by privileging one over the other. This standard can therefore encourage manufacturers to ask themselves the right questions (what risks am I willing to accept?), to estimate the risks associated with their decisions (dimensioning, management and conforming) and, probably, to find more efficient balances than those based on empiricisms that are sometimes open to criticism

References:

1. Joint Committee for Guides Metrology, ISO/IEC GUIDE 98-4, Evaluation of measurement data – The role of measurement uncertainty in conformity assessment, 2012.
2. AFNOR, FD X 07-039, Role of measurement uncertainty in conformity assessment - Implementation of NF ISO/IEC Guide 98-4 – Illustration through industrial case studies, 2018.