Measurement uncertainty: historical perspective, present status and foreseeable future

Original
Measurement uncertainty: historical perspective, present status and foreseeable future / Bich, Walter. - (2021), pp. 305-324. [10.3254/ENFI210032]

Availability:
This version is available at: 11696/72802 since: 2022-09-06T09:56:24Z

Publisher:
Società italiana di fisica and IOS press

Published
DOI:10.3254/ENFI210032

Measurement uncertainty – Historical perspective, present status and foreseeable future

W. Bich

Istituto nazionale di ricerca metrologica, INRIM, Torino, I

Summary. —
In this paper I explore the very birth, due to Gauss, of today's views on measurement uncertainty and its quantitative expression and propagation. I also try to understand and explain the influence of Bridgman's operationalism on these views, and discuss the conflict between a certain misconception of his epistemology and the mathematical and probability tools necessary to treat measurement and measurement uncertainty quantitatively. I also give an overview of current trends in the field.

PACS 06.20 –
PACS . – 02.50....

Pauca sed matura

Karl Friedrich Gauss

1. – Laplace, Legendre and Gauss

Gauss' motto *pauca sed matura* (few but ripe) referred to publications is in strong contrast with the "publish or perish" imperative that rules nowadays research. Nevertheless, even if he had lived in modern times, he would not have had to fight hard to gain a recognition as a scientist. His prediction of the orbit of the asteroid Ceres [1] (now classified as a dwarf planet) gave him international reputation. This asteroid had already been observed by the Italian astronomer Giuseppe Piazzi in Palermo on New Year's Day of 1801, in the framework of a broad European effort aimed at discovering a planet supposed to orbit between the orbits of Mars and Jupiter. Piazzi took observations until 11 February, before the planetoid came too close to the Sun to be observed any further. Piazzi's data were circulated and some of the leading European astronomers made preliminary predictions of Ceres orbit, but observations in August and September failed to find the asteroid. Then came the 24-years-old Gauss with values of the six parameters necessary to determine an orbit and Ceres was found where predicted. The six parameters can be viewed, say, as the components of the position and velocity vectors $\vec{r}$ and $\vec{v}$ at a given epoch t_0. For further details about Gauss' method, see [2].

The reader might wonder what is the relevance of this story to the topic of measurement uncertainty. The relevance is that in his determination of the orbit of Ceres Gauss used the least squares method he had discovered (without publishing it) perhaps a decade earlier. It was typical of Gauss not to publish a scientific finding or result before he had completely and satisfactorily understood it, and in general not to care too much about making public his ideas. And this habit of his is responsible for the dispute about the priority in the discovery of least squares between himself and Legendre. He (Legendre) had published his discovery in 1805 in an appendix to a dissertation on the motion of comets [3], whereas Gauss expounded the theory in his subsequent (1809) *Theoria motus corporum cœlestium in sectionibus conicis solem ambientium* (Theory of the motion of the heavenly bodies moving about the sun in conic sections) [4]. A complete treatment is given only in his *Theory of the Combination of Observations Least Subject to Errors* [5] a work in three parts published in 1821, 1823 and 1826, respectively. It must be said that Gauss, apart from declaring that he was unaware of Legendre's work (which is plausible) did not insist on the issue of the priority.

> Our principle, which we have made use of since the year 1795, has lately been published by Legendre in the work *Nouvelles méthodes pour la détermination*

> *des orbites des comètes, Paris, 1806*, where several other properties of this principle have been explained, which, for the sake of brevity, we here omit. ([4], p. 270).

In addition, as it is often (if not invariably) the case with great scientific discoveries (or inventions), the scene is not limited to two actors. Also the great Laplace had entered the field, by deriving the least squares as the asymptotically optimal solution to the minimization of the mean absolute error, and using his central limit theorem to demonstrate his assertion ([6], ch. 4). See also [7].

2. – Theory of errors

Gauss' motivation in developing the tool of least squares was to find the most plausible value of one (or more) quantities of interest from a set of observations affected by errors. In this respect, his least squares are a (major) part of a wider framework generally known as the theory of errors. The same book in which Gauss gave an exhaustive description of least squares [5] starts with an admirable analysis of observation errors and the related classification into random and systematic.

> However carefully one takes observations of the magnitudes of objects in nature, the results are always subject to larger or smaller errors ... Two of these must be carefully distinguished. Certain causes of error are such that their effect on any one observation depends on varying circumstances that seem to have no essential connection with the observation itself. Errors arising in this way are called irregular or random, ... On the other hand, other sources of error by their nature have a constant effect on all observations of the same class. Or if the effect is not absolutely constant, its size varies regularly with circumstances that are essentially connected with the observations. These errors are called constant or regular. ([5], page 3).

We call today "systematic" Gauss' "constant or regular" errors. Apart from this minor terminological change, the concepts are unchanged.

Also, Gauss made it clear that the distinction is not rigorous.

> Now it is clear that this distinction is to some extent relative and depends on how broadly we take the notion of observations of the same class. For example, consider irregularities in the graduations of an instrument for measuring angles. If we need to measure a given angle again and again, then the irregularities produce constant errors, since the same defective graduation is used repeatedly. On the other hand, the same errors can be regarded as random when one is measuring unknown angles of arbitrary magnitude, since there is no table of errors in the individual graduations. (*ibid.*)

3. – From errors to uncertainty

> *PROBLEM. Given a function U of the unknown quantities V, V', V", etc., find the mean error M to be feared in estimating U when, instead of the true values of V, V', V", etc. one uses independently observed values having mean errors m, m', m", etc.* ([5], p. 33).

Gauss' problem above is exactly the same any experimenter has to face still today when performing a measurement, and is the same as described in the *Guide to the expression of uncertainty in measurement*, or GUM [8], the current authoritative document as concerns measurement uncertainty.

In modern terms, we write:

$$Y = f(X_1, X_2, \ldots, X_N), \tag{1}$$

to mean that the *measurand* Y is not observed directly, but as a function of other quantities that are estimated by observation or by other means.

The quantity estimates are denoted by x_i (often, especially in statistics, also $\widehat{X}_i$) and $\epsilon_i = x_i - X_i$ is the error in the i-th estimate.

Gauss' problem is to find "the mean error to be feared" for the estimate of the measurand – as we will see, the variance –, given the "the mean errors" of the estimates of the *input quantities.*

The solution given in the GUM is still, not surprisingly, the one found by Gauss, and given here

> Solution. Let e, e', e'', etc. denote the errors in the observed values of V, V', V'', etc., and let λ, λ', λ'', etc. be the differential quotients $\frac{\mathrm{d}U}{\mathrm{d}V}$, $\frac{\mathrm{d}U}{\mathrm{d}V'}$, $\frac{\mathrm{d}U}{\mathrm{d}V''}$, etc. at the true values of V, V', V'', etc. Then the resulting error in U can be represented by the linear function
>
> $$\lambda e + \lambda' e' + \lambda'' e'' + \text{etc.} = E,$$
>
> provided the observations are precise enough so that we can neglect squares and products of the errors. From this it follows first that the mean value of E is zero, since the observation errors are assumed to have no constant parts. Moreover, the mean error to be feared in this value of U is the square root of the mean value of EE; that is, MM is the mean value of the sum
>
> $$\lambda\lambda ee + \lambda'\lambda' e'e' + \lambda''\lambda'' e''e'' + \text{etc.} + 2\lambda\lambda' ee' + 2\lambda\lambda'' ee'' + 2\lambda'\lambda'' e'e'' + \text{etc.}$$
>
> Now the mean value of $\lambda\lambda ee$ is $\lambda\lambda mm$, the mean value of $\lambda'\lambda' e'e'$ is $\lambda'\lambda' m'm'$, etc. The mean values of the products $2\lambda\lambda' ee'$, etc. are all zero. Hence it follows that
>
> $$M = \sqrt{\lambda\lambda mm + \lambda'\lambda' m'm' + \lambda''\lambda'' m''m'' + \text{etc.}} \tag{2}$$

Equation (2) is, notation apart, the same as equation (10) in the GUM. What we today call pompously "Law of Propagation of Uncertainties", and for which we even invented an acronym, LPU, is little more than a minor corollary in Gauss' great *opus*. What is more notable is that this equation is still an excellent tool for a large number of experimental situation. Of course, as Gauss (and the GUM) point out, errors must be small, so that the linearization is meaningful and sufficient, and their expectation, *the mean value of* E, must be equal to zero (*the observation errors are assumed to have no constant parts*). Gauss is explicit about this:

> We explicitly exclude the consideration of regular errors from this investigation. Of course, it is up to the observer to ferret out all sources of constant error and remove them. Failing that, he should at least scrutinize their origins and magnitudes, so that their effects on any given observation can be determined and removed, after which it will be the same as if the errors had never occurred. Irregular errors are essentially different, since by their nature they are not subject to calculation. For this reason we have to put up with them in the observations themselves; however, we should reduce their effects on derived quantities as far as possible by using judicious combinations of the observations. ([5], p. 5).

and further on

> Thus the errors in the corrected observations have no constant part, a fact which is actually self-evident. ([5], p. 9).

And this is the main point. Gauss' equation holds for *corrected observation errors*. For this kind of errors, the mentioned assumptions are plausible. But there is a second, equally important concept in Gauss' reasoning: all sources of constant errors should be identified and removed. This statement is, again, echoed in the GUM:

> It is assumed that the result of a measurement has been corrected for all recognized significant systematic effects and that every effort has been made to identify such effects. ([8], 3.2.4).

To correct for all recognized systematic errors is precisely one of the roles of the measurement model (1). Any correction aimed to compensate an identified effect should appear in it as an input quantity.

There is a further condition for equation (2) to work, that is, "*The mean values of the products* $2\lambda\lambda' ee'$*, etc. are all zero.*", which means that there are no correlations between any of the input estimates. A generalization for correlated input estimates ([8], equation (13)) is straightforward [9]. A word of caution is needed here. Correlation is not between quantities (keeping apart quantum entanglement), nor their estimates, but rather between the random variables that describe one's state of knowledge about them. Therefore, when using the loose wording "correlated quantities" or "correlated estimates", it is understood that the meaning is as described above.

But the real milestone in Gauss' theory of errors is to have selected the mean squared error as the *loss function*, *i.e.*, the function to be minimized when searching for the value of a quantity from a set of data. The mean squared error of an estimator $\widehat{\theta}$ of an unknown quantity θ, $\mathrm{MSE}(\widehat{\theta})$, is

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{E}\left[(\widehat{\theta} - \theta)^2\right],$$

where E is the expectation $\mathrm{E}(X) = \int x f(x)\mathrm{d}x$, $f(x)$ being the probability density function, PDF, of the random variable (RV) X.

It is easily shown that $\mathrm{MSE}(\widehat{\theta}) = \mathrm{VAR}((\widehat{\theta}) + \left[\mathrm{E}(\widehat{\theta} - \theta)\right]^2$, that is, the variance of the estimator, usually denoted by $\sigma^2(\widehat{\theta})$, plus its squared *bias*, the constant part of errors in Gauss' language. But random errors as those considered in equation (2) have no constant part, so the mm, $m'm'$, ... are just the error variances.

Gauss' motivation is clear from his own words:

> ...the integral $\int xx\varphi x\mathrm{d}x$(1) taken from $x = -\infty$ to $x = +\infty$ (the mean square of x) *seems most appropriate to generally define and quantify the uncertainty of the observations.* Thus, given two systems of observations which differ in their likelihoods, we will say that the one for which the integral $\int xx\varphi x\mathrm{d}x$ is smaller is the more precise. ([5], p. 9).

This somewhat arbitrary decision about the loss function (arbitrariness of which Gauss was well aware) was dictated essentially by utilitarian considerations, as the simplest non-negative function of the errors. Laplace had chosen the absolute error as the loss function, a choice that "resists analytical treatment, while the results that my convention leads to are distinguished by their wonderful simplicity and generality." (*ibid.*, p.11).

4. – Modern times

The same choice as Gauss' was adopted in the Recommendation INC-1 (1980), *Expression of Experimental Uncertainties*, issued in 1980 by a Working Group on the Statement of Uncertainties convened by the BIPM (Bureau international des poids et mesures) upon request of the CIPM (Comité international des poids et mesures). This recommendation eventually led to the GUM. The authoritative French text is reproduced in [8], Annex A.1. For an excerpt in English of the most significant part, see *ibid.*, 0.7. In a nutshell,

> 1) The uncertainty in the result of a measurement generally consists of several components which may be grouped into two categories according to the way in which their numerical value is estimated:
>
> A. those which are evaluated by statistical methods,

(1) The x are here the errors and φx is $f(x)$, the PDF; the integral $\int_{-\infty}^{+\infty} xx\varphi x\mathrm{d}x = mm$. The italics are mine.

> B. those which are evaluated by other means.
>
> 2) The components in category A are characterized by the estimated variances s_i^2, ...
>
> 3) The components in category B should be characterized by quantities u_j^2, which may be considered as approximations to the corresponding variances, the existence of which is assumed. The quantities u_j^2 may be treated like variances ...

The recommendation marks a great generalization with respect to Gauss' original formulation. This generalisation consists in including uncertainty components belonging to category B in the same formal mathematical scheme conceived by Gauss for random errors, as I will show in the following part.

Type A evaluations concern cases in which there is a sample of data concerning the same quantity, so that statistical methods (such as the sample variance of the mean) can be applied. Errors are assumed to be random, with no constant part, *i.e.*, with expectation equal to zero. An instance of random error with non-zero expectation is the so-called cosine error. A method useful to shift its PDF so that the expectation is zero is exhaustively discussed in the GUM (see [8], F.2.4.4).

In conclusion, if there is a data sample, one can calculate a sample standard deviation (variance) of the mean, comfortably within Gauss' theory-of-errors framework.

The generalisation with respect to Gauss' scheme is in Type B evaluations. Type B evaluations enable the assignment of a variance even in those cases in which there is no sample. This possibility, counter-intuitive for many, is a feature of an interpretation of probability as *degree of belief.*

> Thus a Type A standard uncertainty is obtained from a probability density function derived from an observed frequency distribution, while a Type B standard uncertainty is obtained from an assumed probability density function based on the degree of belief that an event will occur [often called subjective probability]. Both approaches employ recognized interpretations of probability. ([8], 3.3.5).

So, in the GUM two views of probability coexist. The first sees the probability of an event as a limit of the relative frequency observed for that event, and is often labelled as *frequentist.* The second sees probability as a subjective opinion, or degree of belief, on the realization of the event; alternatively, and perhaps more generally, a probability distribution is viewed as as a representation of one's state of knowledge about the event. This view of probability is appropriately denoted as *subjective.*

Subjective probability has a long and rich history in the mainstream of mathematics and probability, dating back to Bayes [10] and Laplace [6], and developed in modern times by Keynes [11], Ramsey [12], de Finetti [13, 14, 15] and subsequently by Savage [16], Jeffreys [17, 18], Jaynes [19], Lindley [20] and many others.

In its most radical formulation, conceived by Bruno De Finetti and subsequently championed, among others, by Dennis Lindley, the subjective interpretation of probability has been connected with how much one is disposed to bet on an event. Interestingly, in De Finetti's view, the assignment of a subjective probability is related with acting optimally, by minimising a squared error loss function, thus connecting subjective probability with least squares [21]. Other formulations, mostly due to Jeffreys and followers, attempt to give an image of objectivity. In this view, the probability distribution describes a supposedly "objective", or shared, state of knowledge about the event.

It has become common practice to denote either approaches as *Bayesian*. The distinctive feature of the Bayesian view of probability with respect to the frequentist is that in the latter, probability is a sort of physical propensity inherent in the real world, whereas it is a state of mind in the former. For more on Bayesian epistemology, see [22].

A nice thing is that the mathematics of probability works the same, regardless of the interpretation of probability. This does not mean that the same problem tackled with frequentist and Bayesian methods leads necessarily to the same conclusion.

I see Bayesian probability more general than frequentist. First, it allows one to assign meaningful probabilities to unique events, such as "a nuclear war in the next ten years". Second, it nicely includes frequentism. For example, when dealing with a sample of observations, the observed dispersion of values, expressed quantitatively by the sample variance, is no doubt an expression of the propensity of the instrument to scatter its responses to a constant stimulus. At the same time, however, it represents an empirical piece of knowledge from which one can infer the (personal) uncertainty about the value of the stimulus. Thus, sample variance of the mean and (squared) uncertainty are completely different concepts. The sample variance pertains to a real-world population, the (squared) uncertainty refers to a state of mind. No wonder that their numerical values do not necessarily coincide. They might, for infinitely large sample, whereas it is well known that small samples give unreliable estimates of the population variance. Thus, identifying the sample variance of the mean with the (squared) uncertainty would yield an unreliable, or uncertain, uncertainty; honestly, a very strange beast. The Bayesian uses the sample variance of the mean and, taking into account its reliability (represented by the *degrees of freedom*), decides how much he would bet on the estimate. His belief is expressed quantitatively by a *standard uncertainty*, calculated from the sample variance and the degrees of freedom, according to a formally rigorous procedure.

In addition, Bayesian probability allows one to blend today's information (the sample variance) with prior knowledge (the usual scatter of the instrument, scatter well known from past usage). In conclusion, in the Bayesian scheme the sample variance is just one piece of information that contributes, alone or together with other pieces, to form one's opinion on how much to bet on the value of the quantity under measurement. For further discussion on Type A evaluations from a Bayesian viewpoint, see section **9**.

Whilst Bayesian probability has a long and respected tradition at large, its adoption in the comparatively narrow world of metrology is recent, and the path that eventually led to Recommendation INC-1 was by no means easy. One of the ideas was, and still is, that systematic errors should be identified and corrected, as already mentioned in section **3**. As we saw, the idea of "removing" systematic errors was already in Gauss' mind, but there is no mention in his work of the remaining doubt on how well the correction has been made. Only a subjective view of probability enables this doubt to be expressed quantitatively in a rigorous manner, the same that can (and should) be used to express contributions evaluated statistically. Unfortunately, consistency is realized in the GUM in the opposite direction, by artificially (and unconvincingly) converting state-of-knowledge variances into traditional ones. Technically, this task is accomplished by attaching (subjectively evaluated!) degrees of freedom to state-of-knowledge variances.

It should be evident that such a solution does not work. Yet, a 2014 attempt at revising the GUM in the "right" direction [23] failed due to rejection from the user's community [24].

Coming back to Resolution INC-1, it has great merits, as it prescribed a unified treatment of random and systematic contributions to uncertainty. To have an idea of the debate at those times, see [25]. However, a clear view of a unified treatment came only later, with a seminal paper on a Bayesian theory of uncertainty [26].

5. – True value

Gauss' concept of error given above implies that "objects in nature" possess "magnitudes". In the currently leading mathematical description of Nature this magnitude of an object, or *value of a quantity*, as we preferably write nowadays, is represented by a point on a one-dimensional, metric or coordinate space $\Re^1$, *i.e.*, by a real number multiplied by a unit. By writing m one means the generic quantity "mass", mathematically represented by a non-negative variable (presumably with an upper bound). By writing m_e one means the specific value "rest mass of the electron", mathematically represented by a specific value of the variable m, *i.e.*, by a point in $\Re^1$. This representation implies the the rest mass of the electron is believed to have a unique, albeit unknown, *true value*, or simply *value*. Incidentally, our present best estimate of it is $\widehat{m_\mathrm{e}} = 9.109\,383\,701\,5 \times 10^{-31}$ kg [27], where I have used the common "hatted" notation for an estimate, just to emphasize that this is **not** (the true value of) m_e. However, we are confident that m_e lies somewhere close to $\widehat{m_\mathrm{e}}$, and are able to quantify both how close it lies and how confident we are in this statement. Specifically, it is believed that $\widehat{m_\mathrm{e}} - 0.000\,000\,028 \times 10^{-31}\,\mathrm{kg} \leq m_\mathrm{e} \leq \widehat{m_\mathrm{e}} + 0.000\,000\,028 \times 10^{-31}\,\mathrm{kg}$ with a probability $\mathrm{P} \approx 0.68$. A good way to admit our incomplete knowledge, and to quantify it. This quantification is enabled by two key factors: first, data, many data, coming from several experiments around the world and suitably processed to obtain a consensus value (for an exhaustive discussion, see [28]). Second, probability theory, the master tool to quantify one's confidence in a statement.

6. – Digression on the International System of Units

The case of m_e is just one, among many others, illustrating the impact of the revision of the International System of Units, SI, on the fundamental constants. The revised SI came into force on 20 May 2019, following a landmark resolution of the XXVI General Conference on Weights and Measures [29]. In the revised system, all the seven base units are defined in terms of as many invariants of nature, most of them fundamental physical constants, such as the elementary charge e, the Planck constant h, the Boltzmann constant k, the Avogadro constant N_A [30]. These invariants were given exact values, with zero uncertainty. There was a precedent, when the speed of light in vacuum c_0 was given an exact value in 1983 and the unit of length, the metre, was defined in terms of it ([31], Appendix 1). The advantages introduced by the 2019 epochal change with respect to the past are immense and largely outperform a few minor disadvantages. Among the advantages, the uncertainties of many other fundamental constants, functionally connected to the fixed foundational constants, decreased considerably by the simple fact that other constants to which they are connected became exact. The uncertainty of the electron mass, prior to the SI revision, was $u(m_\mathrm{e}) = 0.000\,000\,11 \times 10^{-31}\,\mathrm{kg}$ [28]. Comparing the relative standard uncertainties, it is a shift from 1.2×10^{-8} to 3×10^{-10}, an improvement of almost two orders of magnitude. Of course, there is nothing magical in this, the uncertainties having been simply re-distributed in a way closer to our vision of nature. For a historical perspective, a review of the changes involved in the SI revision, and further relevant literature, see [32, 33].

7. – Bridgman and the operationalism

In 1927, P W Bridgman published his book *The Logic of Modern Physics* ([34], freely downloadable). Bridgman was an experimental physicist who was awarded the 1946 Nobel Prize "for the invention of an apparatus to produce extremely high pressures, and for the discoveries he made therewith in the field of high pressure physics" [35]. He was also a deep and analytical thinker and applied his qualities to the discussion of the foundations of physics, eventually becoming one of the most influential philosophers of science not only in the field of physics but also, perhaps especially and certainly unexpectedly to himself, in the fields of experimental psychology, social science and life sciences in general.

The essence of Bridgman's reasoning is that any concept, be it mental or physical, should be defined through the operations needed to identify (or *measure*) it. Therefore, a set of operations defines the concept. "The concept is synonymous with the corresponding set of operations." ([34], p. 5). Such a general attitude became known as *operationalism*, and the position of defining a concept (or a physical property or quantity) in terms of a set of operations as *operational definition*.

Bridgman explicitly acknowledges that his investigation was stimulated by Einstein's famous critique of the then current concept of simultaneity, analysis that led Einstein to the relativity theory. Einstein analyzed the concept of simultaneity in terms of the operations needed to assess it, coming to the conclusion that simultaneity is not an absolute, but a relative concept, depending on the reference frames of the observer(s) of the supposedly simultaneous events. It could be said that an operational definition of simultaneity triggered the development of the relativity theory.

Bridgman applied the same technique to the concept of length, highlighting the existence of many sets of operations depending on the size of the length to be measured, from the atomic level to stellar distances. One of the consequences of his analysis is that quantities (I would say *measurands*) so different in size should in principle be considered different in quality, as the set of operations necessary to measure them is different. The same consideration applies to the same measurand, measured according to different measurement principles and thus to different measurement setups and procedures.

Incidentally, this argument is relevant, again, to fundamental constants and, perhaps less interestingly, to inter-laboratory comparisons. In these cases, and especially in the former, various measurement principles happen to be used. Each of these corresponds to an individual realization of the measurand. It is not necessarily true that all these different realizations are equivalent, which can explain (a part of) the scattering of data that is usually observed in such exercises. For an exhaustive review of the techniques useful to form a consensus value and the associated uncertainty from a set of consistent or inconsistent data, see [36]. For a related software implementing some of these techniques, see [37].

The discussion of the concept of length (other quantities and concepts are discussed by Bridgman, but length became indisputably the most famous and frequently-cited) includes as well the case of an elementary measurement using a calibrated rod. Here, Bridgman shows what any practitioner knows (or should know), that is, that many effects influence the measurement and its result, so that even such a simple measurement, when scrutinised operationally, is not simple at all.

"From a very direct concept we have come to a very indirect concept with a most complicated set of operations." ([34], p. 16).

Eventually, following the operational definition through to its extreme consequences, the result is discouraging:

"If we stick to the concept of length by itself, we are landed in a vicious circle. As a matter of fact, the concept of length disappears as an independent thing, ..." ([34], p. 22).

The two quotations above embody the twofold relevance of Bridgman's views to our discussion on measurement uncertainty, and to measurement in general. First, the operational analysis *excludes the possibility of direct measurements.* Even when one compares a rod to the length on an item, a *measurement model* needs to be established containing the details of the measurement. Maybe the accuracy requested (sometimes called *target uncertainty*) is such that the effects of all these details are negligible compared to it, yet the measurement model conceptually is in place, and cannot be eliminated.

Second, the very concept of true value, so needed by the mathematics we use to obtain measurement results, is questioned. If the specification of the measurand cannot be carried to its ultimate step, it cannot be complete, and the concept of true value vanishes into "a twilight zone, a *penumbra of uncertainty*, into which we have not yet penetrated." ([34], p. 33). For sure, by *penumbra* Bridgman intended what we know as measurement uncertainty ("The penumbra is to be penetrated by improving the accuracy of measurement." ([34], p. 34). However, a misinterpretation of this passage might conclude that the penumbra derives from the impossibility of completely defining the measurand, thus constituting an impenetrable limit to any measurement.

Despite my search, I could not identify who came (and when) to the conclusion described above, crystallizing the concept in a *definitional uncertainty* defined as "component of measurement uncertainty resulting from the finite amount of detail in the definition of a measurand" ([38], definition 2.27). Note 1 to the definition reads "Definitional uncertainty is the practical minimum measurement uncertainty achievable in any measurement of a given measurand." The definition is taken from the authoritative *International Vocabulary of Metrology — Basic and General Concepts and Associated Terms*, known as the VIM.

Whilst the concept of definitional uncertainty may have a value in a philosophical context, it poses a problem in the evaluation of measurement uncertainty, of which it is claimed to be a component. Actually, it is the only component of measurement uncertainty for which there are no known tools able to quantify it, even subjectively. How can be evaluated the amount of *penumbra* implied by the incomplete definition of the measurand? How is it possible to quantify the incompleteness of the measurand definition? I tend to consider definitional uncertanty as an unknowable (idealised?) concept, that is much less useful than other idealised concepts. The GUM discusses to some extent the uncertainty arising from the incomplete definition of the measurand, ([8], 3.1.3 and Annex D), but there is not even a single mention of it in the mathematical treatment. The only appearance in an example ([8], H.6) looks weak and artificial.

Yet, a practical solution to the problem of an ill-defined measurand exists, is clearly indicated in the GUM (and regularly overlooked), and has been put in practice in at least one exemplary case. An ill-defined measurand, such as the length of a table, or the stature of blonde persons, or the distance from Varenna to Paris, or the atomic weight of some elements, implies a population of values that satisfy the definition. Then, in a correct setting of the problem two measurands would be defined, one being a location parameter for the frequency distribution of the population, the other being a scale parameter for that same distribution (see [8], 2.1). From the corresponding estimates, a state-of-knowledge probability distribution can be inferred. For an example of ill-defined measurand(s), see [39], concerning the atomic masses of some elements.

If the former conclusion of Bridgman's operational analysis – the need for measurement models – is a welcome one, the latter – that there is no true value of a quantity – poses problems. It would seem that trying to measure anything is nonsense, as the very quantity of interest does not exist.

"The 'absolute' therefore disappears in the original meaning of the word." ([34], p. 26).

Clearly this position, highly respectable in the context of epistemology, would be paralysing in practical measurements. Yet, it has been the rationale for criticisms against true value and, as a consequence, measurement error, labelled as unknowable and thus useless concepts. Echoes of this view can be found in the GUM (see [8], Annex E) and especially in the VIM, in which an Uncertainty Approach, based on operational concepts, is contrasted with a traditional Error Approach, based un unknowable concepts (see

[38], Introduction). For an analysis of the misconception implied by this non-existing contraposition, see [40] and especially [41], section 6.3.

A point that critics of true value and error do not seem to consider is that estimation theory [42, 43, 44, 45, 46, 47], on which in turn are based the evaluation of a measurand value and of the uncertainty about it, is entirely based on the idealised concepts of (true) value and error. In addition, to the best of my knowledge, no working counterpart based on operational concepts is known.

For more on Bridgman, see [48, 49].

8. – Definition of uncertainty

Measurement uncertainty as a superordinate concept is defined in the GUM essentially (although not explicitly) as a measure (any measure) of the scale, *i.e.*, the spread of a probability distribution. The measure formally adopted is specified in the definition of standard uncertainty, that reads

> standard measurement uncertainty: uncertainty of the result of a measurement expressed as a standard deviation ([8], 2.3.1)

This definition of standard uncertainty is clear in one respect and obscure in another. It is clear in that, following the Recommendation INC-1 mentioned in section **4**, it explicitly states that the measure chosen to express uncertainty in measurement, be it arising by the dispersion of experimental data or by incomplete knowledge of a datum, is the variance (the standard deviation is used only because it has the same units as the relevant quantity). It is obscure in that the specification of standard deviation (or variance) alone is meaningless. Actually, the variance is defined as the second moment (about the mean) of a probability distribution. It is evident that defining standard uncertainty as one standard deviation, without specifying the probability distribution to which it pertains, is meaningless. The superordinate definition of measurement uncertainty might suggest that the distribution is the-state-of-knowledge for the measurand, following measurement (a true Bayesian would express this as the *posterior* probability distribution for the measurand.)

Yet, some ambiguity remains. A frequentist would see the standard uncertainty as belonging to the frequency distribution *of the estimates* that would be obtained repeating the experiment a large number of times, an untenable position, as we saw, yet still widely followed.

Followers of operationalism, not accepting the concept of true value, interpret this distribution as a "set of values being attributed to the measurand" (see [38], definitions 2.9 and 2.26), which in turn implies the existence of a set of true values for the measurand. So, to remove the concept of true value of a measurand, an entire set of true values is introduced for it (see [38], definition 2.11 and especially NOTE 1.) Again, a position difficult to understand, yet, the one adopted in the VIM, and the main reason for the inconsistencies between the VIM and the GUM. Again, estimation theory gives the appropriate answer. The probability distribution for the measurand fully describes one's opinion on where the value of the measurand lies in $\Re^1$ (the extension to the case of more than one measurand being conceptually straightforward). Appropriate tools of probability theory make it possible to specify the (subjective) probability that the measurand lies in a given interval.

This interpretation is the same as that described in section **5**. The VIM (see NOTE 2 to its definition 2.11 of true value) considers fundamental constants as special cases, that somehow elude the otherwise comprehensive philosophical framework of operationalism, or are, in the least, limit cases in which the set of true values consists of just one value. This position is questionable, in that an operative definition of any measurable constant implies its realization by means of a real experiment. According to Bridgman, different measurement principles (and corresponding experimental systems) measure different measurands, and this is true also for fundamental constants. So, a true operationalist should consider fundamental constants as any other measurand, and attach to them as well a definitional uncertainty. He should accept that the concept of true value is an idealisation also for a fundamental constant. I prefer to extend and use the idealised concept for all measurands, and treat all measurands as if they were fundamental constants. If a measurand is ill-defined, there is often a way to find one or more well-defined, relevant measurands, as outlined in section **7**.

As a final comment on the definition of a measurand, it should be considered that an unambiguous, clear definition is represented by the measurement model given by expression (1). This is the honest, simple declaration of what the experimenter intends for the measurand. It might not be *the* measurand, it is simply *his* measurand. No more, no less.

9. – Measures of uncertainty

The appealing feature of mean and variance (the first moment about zero and the second moment about the mean of a probability distribution, respectively) is that they propagate in a comparatively simple way. This feature, as we saw in section **3**, was Gauss' original motivation and is still a major reason to adopt them as a summary of the measurement result (for some, as *the* measurement result.) In addition, the rule by which variances propagate [expression (2)] is distribution-free, that is, it works for any underlying distribution. This is of course a limitation as well. In principle, it is not possible to specify the probability that the measurand lies in a given interval, although often it is possible to specify an approximate probability, which in most cases is sufficient.

A negative feature of moments is that they may not exist, or be undefined. A classic example is the Cauchy distribution, which is encountered in many physical processes and is also the ratio of two Gaussian distributions, the latter occurrence being non-infrequent in modelling a measurement. Also the variance of the Student-t distribution, being equal to $\nu/(\nu-2)$, where ν is the (non necessarily integer) degrees of freedom, is finite only for $\nu > 2$, being otherwise infinite or undefined (see any textbook of mathematical statistics, for example, [43]).

In addition, the corresponding sample statistics, the sample mean and the sample variance of the mean, can be unreliable, especially when the sample is small. This is due, in general, to issues of robustness against outliers (see [50, 51, 52]) and for the sample variance, in particular, to its asymmetric distributional properties, by which, the smaller the sample, the more likely it is that the produced values are low with respect to the population variance.

There is a further, usually overlooked drawback, when the standard uncertainty is concerned rather than its square. Standard uncertainty is defined in terms of standard deviation, rather than variance, essentially to have uncertainty expressed in the same units as the estimate and the measurand. This choice responds to needs of practicality, such as enabling relative uncertainty, widely used, ease of visualisation and comparability. However, since the sample variance s^2 is an unbiased estimator of the population variance σ^2, *i.e.*, $\mathrm{E}(s^2) = \sigma^2$, and since the square root is a non-linear operator, $\mathrm{E}(s) \neq \sigma$, that is, the sample standard deviation is a biased estimator of the population standard deviation. In particular, since $\mathrm{Var}(s) = \mathrm{E}(s^2) - [\mathrm{E}(s)]^2$ and $\mathrm{Var}(s) \geq 0$, one has $\mathrm{E}(s) \leq \sigma$, *i.e.*, the sample standard deviation is on average biased downwards. Thus, identifying the standard uncertainty $u(x)$ with the sample standard deviation of the mean yields a too small standard uncertainty. In the frequentist framework, the bias can be eliminated with a distribution-dependent correction. In a Bayesian framework, the sample variance of the mean is used, *together with its degrees of freedom*, to calculate a standard deviation for the corresponding state-of-knowledge distribution, *i.e.*, a Student-t distribution (for a Gaussian population). In conclusion, $u(x) = \left(\frac{n-1}{n-3}\frac{s(x)^2}{n}\right)^{1/2}$, where n is the number of observations in the sample, is an appropriate Type A measure of uncertainty.

The choice of the state-of-knowledge standard deviation above as the standard uncertainty for Type A evaluations was negatively received by the metrological community for being too large with respect to the current one, the sample standard deviation of the mean, and for the constraint $n > 3$. As to the first criticism, it is not that expression that provides too large values, but rather the current one that gives too smal values, as we saw. As to the second criticism, it is true that most calibration laboratories, including National Metrology Institutes, do not take more than two or three observations during routine calibrations, unless automatic instruments are available. In such cases, however, the standard uncertainty associated with the mean of the observations is typically not based on the current sample, but rather on the known performance of the instrument that is used. The performance has been checked with extensive tests producing large samples, whose variance is reliable. The variance of the current, small sample is just used to check that the process is under control. The values reported in calibration certificates for a given measurand are usually fixed and coincide with the calibration and measurement capability of the laboratory for that measurand.

A formally rigorous way to blend the (well-known) historical performance of the instrument with its (poorly known) current one is through the Bayes-Laplace rule [45]. The implementation of Bayesian inference in routine laboratory activity is probably too difficult a task. Attempts to provide simple formulas have been made [53, 54, 55]. For a discussion of various methods of evaluating input uncertainties, especially concerning

Type A evaluations, see [56, 57].

Location and scale measures other than moments, such as quantiles, might also be considered. The appealing feature of quantiles as location and scale parameters of a distribution is that they always exist. In addition, the corresponding sample statistics are generally more robust against outliers than those for moments [50, 51, 52]. The median is widely used as location parameter in the scientific literature, especially in the life sciences. A negative feature is that no simple propagation law exists for quantile-based measure of uncertainty, similar to that found by Gauss for variances. This is a major obstacle to a wide adoption of such measures for estimate and uncertainty. However, work is in progress in this field [58].

10. – Broader view of measurement and measurement uncertainty

Quantities are considered a subset of the greater set of *properties*. Quantities are characterized by a magnitude, and are measurable in terms of a number and a reference, typically a measurement unit ([38], definition 1.1). *Nominal*, or, *qualitative* properties have no magnitude. The VIM specifies that measurement only applies to quantities ([38], NOTE 1 to definition 2.1). So, strictly speaking, the concept of measurement uncertainty as defined in the VIM and the GUM is not applicable to *examination*. However, examination, the process of assigning a (non-numerical) value to a property, *i.e.*, to classify it in a given category, is possibly as widespread as measurement.

The debate is alive on whether the definition of measurement should be extended to include examination of nominal properties – a prominent National Metrology Institute such as the National Institute of Standards and Technology, NIST, US, has no doubts on this, it should! – see [59, 60].

Independent of the debate on formal aspects, awareness is rapidly growing that examinations as well are not perfect, and errors are possible and should be quantified in some way. So the concept of examination uncertainty is gaining grounds and actions should be taken to develop or adapt suitable mathematical tools able to quantify examination uncertainty. These tools already exist: nominal properties can be described mathematically as *categorical variables*, whose discrete probability distributions are *categorical distributions*. There is a huge literature on categorical variables (an authoritative reference is [61]), so that an extension of measurement uncertainty to include examination would seem straightforward. This is not the case, essentially because of the current definition of measurement uncertainty as a number (*parameter* in the formal definition). The results of examinations are not numbers, as the sample space for an examination is not metric, although numbers can be assigned to its elements as convenient labels. Accordingly, the concepts of location and scale parameters do not apply to categorical probability distributions. Therefore, a different, broader definition of measurement uncertainty is needed. Definitions of this kind have been suggested in terms of the whole probability distribution for the measurand (or *examinand*) [62, 63]. Even more generally, uncertainty is defined as *doubt*, although not formally, in the GUM itself ([8], 2.2.1) and more explicitly in the *NIST Simple Guide for Evaluating and Expressing the Uncertainty* *of NIST Measurement Results* ([64], section 3). Under this general umbrella provided by a doubt-based definition of uncertainty, various options can be suggested to quantitatively express examination uncertainty. The full probability distribution is of course the most complete description. Entropy has been suggested as a possible summary ([64], 3c). Another possibility could be to specify the categories with the highest probabilities, thus including the mode as the estimate for the examinand, and the associated probabilities, so that the probabilities of the selected categories sum up to (no less than) a prescribed value, say, 0.68, or 0.95. This would probably be helpful in taking subsequent decisions, based on the examination. I suspect that subjectivity plays a great role in examinations. In this respect, *expert elicitation* is the ideal tool to give a rigorous shape to the resulting probability distribution [63].

11. – Conclusion

Sometimes I ask myself whether it is sensible to devote so much time and effort at the international level for something that usually, in its current form, is reported with just two significant digits. A similar question was posed to me years ago by a Nobel laureate also lecturing at the Varenna school. After so much time and effort, one would imagine that the topic is now mature and stable. Yet, even at the normative level inconsistencies and misconceptions remain unresolved. In addition, the fields in which the evaluation of measurement uncertainty is becoming a requirement are rapidly increasing and differentiating. There is therefore a continuous need for adapting and modifying the existing tools, and finding new ones in order to meet all these requirements.

At the end of the day, uncertainty remains an uncertain matter.

12. – Acknowledgements

The author is convener of the Joint Committee for Guides in Metrology (JCGM) Working Group 1 (GUM). He is grateful to his colleagues of Working Group 1 as well as to those of Working Group 2 (VIM), for the many fruitful discussions that contributed to form the ideas expressed in this paper. However, they remain personal views. Special thanks go to Tony O'Hagan, for sharing with me his acute insight and vast expertise on some specific topics treated in the paper.

13. – Afterword

When I started writing this paper, I had to face a difficulty which is, I imagine, typical of those who lecture repeatedly on the same topic at the Varenna summer school. My lecture was essentially the same as that of the previous edition of the school (things do not move that fast in the field of uncertainty), but the paper could not be the same. So, I decided to try something different. It is often heard that Gauss is the father of the theory of errors, but I had only confused ideas about his contribution beyond the probability distribution that carries his name. So, I went first to the original Gauss' writings. I was soon fascinated by the many, so often tightly intertwined contributions that led to our current views on uncertainty, and to the equally intricate misunderstandings about it. Therefore, I tried to understand the reasons for the demonisation of terms such as error and true value, and the resulting contraposition of the uncertainty approach against error approach. So, I went to Bridgman and his operationalism. I tried also to give an overview of more recent developments in the field. The reader is invited to take this as an intellectual journey of the author, in an attempt to clarify to himself the present situation. The reader curious about measurement uncertainty in general has at disposal a vast literature. The student who attended the 2019 Varenna summer school can refer to [65] for a good reproduction of my 2019 talk. Also see [50, 40, 66, 41].

My cats Gin&Fizz contributed in part to this paper, by walking at random on the keyboard. However, the responsibility for any blunder, typo, mistake or misconception is entirely mine.

References

[1] Donald Teets and Karen Whitehead. "The Discovery of Ceres: How Gauss Became Famous". In: *Mathematics Magazine* 72.2 (1999), pp. 83–93. DOI: 10.1080/0025570X.1999.11996710. eprint: https://doi.org/10.1080/0025570X.1999.11996710. URL: https://doi.org/10.1080/0025570X.1999.11996710.

[2] Giovanni F. Gronchi. "Classical and modern orbit determination for asteroids". In: *Proceedings of the International Astronomical Union* 2004.IAUC196 (2004), pp. 293–303. DOI: 10.1017/S174392130500147X.

[3] A.M. Legendre. *Nouvelles méthodes pour la détermination des orbites des comètes*. Nineteenth Century Collections Online (NCCO): Science, Technology, and Medicine: 1780-1925. F. Didot, 1805. URL: https://books.google.it/books?id=FRcOAAAAQAAJ.

[4] C. F. Gauss and C. H. Davis. *Theory of the motion of the heavenly bodies moving about the sun in conic sections: a translation of Gauss's "Theoria motus." With an appendix*. Little, Brown and company, 1857. URL: https://books.google.it/books?id=cspWAAAAMAAJ.

[5] C. Gauss and G. Stewart. *Theory of the Combination of Observations Least Subject to Errors, Part One, Part Two, Supplement*. Society for Industrial and Applied Mathematics, 1995. DOI: 10.1137/1.9781611971248. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611971248. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611971248.

[6] Pierre-Simon Laplace. *Œuvres complètes*. Vol. 7-2: *Théorie analytique des probabilités*. Gauthier-Villars (Paris), 1886. URL: https://iris.univ-lille.fr/handle/1908/4944 (visited on 07/03/2020).

[7] O. B. Sheynin. "- Origin of the Theory of Errors". In: *Nature* 211.5052 (1966), p. 1004.

[8] BIPM et al. *Evaluation of measurement data — Guide to the expression of uncertainty in measurement*. Joint Committee for Guides in Metrology, JCGM 100:2008. URL: `http://www.bipm.org/utils/common/documents/jcgm/JCGM\_100\_2008\_E.pdf`.

[9] W Bich. "Simple formula for the propagation of variances and covariances". In: *Metrologia* 33.2 (1996), p. 181. URL: `http://stacks.iop.org/0026-1394/33/i=2/a=9`.

[10] Thomas Bayes and Mr. Price. "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S". In: *Philosophical Transactions of the Royal Society of London* 53 (1763), pp. 370–418. DOI: `10.1098/rstl.1763.0053`. eprint: `https://royalsocietypublishing.org/doi/pdf/10.1098/rstl.1763.0053`. URL: `https://royalsocietypublishing.org/doi/abs/10.1098/rstl.1763.0053`.

[11] John Maynard Keynes. *A Treatise on Probability*. McMillan, 1921. URL: `https://archive.org/stream/treatiseonprobab007528mbp#page/n9/mode/2up` (visited on 07/03/2020).

[12] Frank Plumpton Ramsey. *The Foundations of Mathematics and other Logical Essays*. Vol. VII: *Truth and Probability*. Ed. by R. B. Braithwaite. electronic version, 1999. Kegan Paul, Trench Trubner & Co, 1931. URL: `https://core.ac.uk/download/pdf/7048428.pdf` (visited on 07/03/2020).

[13] D.V. Lindley. "Obituary: Bruno de Finetti, 1906-1985". In: *Journal of the Royal Statistical Society: Series A (General)* 149.3 (1986), pp. 252–252. DOI: `10.1111/j.2397-2327.1986.tb00410.x`. eprint: `https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2397-2327.1986.tb00410.x`. URL: `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2397-2327.1986.tb00410.x`.

[14] Donato Michele Cifarelli and Eugenio Regazzini. "De Finetti's contribution to probability and statistics". In: *Statist. Sci.* 11.4 (Nov. 1996), pp. 253–282. DOI: `10.1214/ss/1032280303`. URL: `https://doi.org/10.1214/ss/1032280303`.

[15] URL: `http://www.brunodefinetti.it/` (visited on 08/04/2020).

[16] Leonard Jimmie Savage. *The Foundations of Statistics, 2nd edition*. Dover Publications, 1972. ISBN: 0-486-62349-1. URL: `https://www.gwern.net/docs/statistics/decision/1972-savage-foundationsofstatistics.djvu` (visited on 07/03/2020).

[17] H. Jeffreys. *Theory of Probability.* International series of monographs on physics. Clarendon Press, 1983. ISBN: 9780198531937. URL: https://books.google.it/books?id=EbodAQAAMAAJ.

[18] Alan Hugh Cook. “Sir Harold Jeffreys, 2 April 1891 - 18 March 1989”. In: *Biographical Memoirs of Fellows of the Royal Society* 36 (1990), pp. 301–333. DOI: 10.1098/rsbm.1990.0034. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rsbm.1990.0034. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsbm.1990.0034.

[19] Edwin Thompson Jaynes. *Probability Theory – The Logic of Science.* Cambridge University Press, 2003. ISBN: 9780521592710.

[20] Dennis V. Lindley. *Understanding Uncertainty, Revised Edition.* Wiley, 2014. ISBN: 978-1-118-65012-7. (Visited on 07/03/2020).

[21] A. O'Hagan. *Private communication.* Aug. 20, 2020.

[22] William Talbott. “Bayesian Epistemology”. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Winter 2016. Metaphysics Research Lab, Stanford University, 2016.

[23] Walter Bich et al. “Revision of the “Guide to the Expression of Uncertainty in Measurement””. In: *Metrologia* 49.6 (2012), pp. 702–705. URL: http://stacks.iop.org/0026-1394/49/i=6/a=702.

[24] Walter Bich, Maurice Cox, and Carine Michotte. “Towards a new GUM - an update”. In: *Metrologia* 53.5 (2016), S149. URL: http://stacks.iop.org/0026-1394/53/i=5/a=S149.

[25] K Weise and W Wöger. “A Bayesian theory of measurement uncertainty”. In: *Measurement Science and Technology* 4.1 (Jan. 1993), pp. 1–11. DOI: 10.1088/0957-0233/4/1/001. URL: https://doi.org/10.1088%2F0957-0233%2F4%2F1%2F001.

[26] Jörg W. Müller. “Some second thoughts on error statements”. In: *Nuclear Instruments and Methods* 163.1 (1979), pp. 241–251. ISSN: 0029-554X. DOI: https://doi.org/10.1016/0029-554X(79)90055-7. URL: http://www.sciencedirect.com/science/article/pii/0029554X79900557.

[27] Eite Tiesinga et al. *CODATA Internationally recommended 2018 values of the Fundamental Physical Constants.* accessed 2020-07-05. 2018. URL: https://physics.nist.gov/cuu/Constants/index.html.

[28] Peter J. Mohr, David B. Newell, and Barry N. Taylor. “CODATA recommended values of the fundamental physical constants: 2014”. In: *Rev. Mod. Phys.* 88 (3 Sept. 2016), p. 035009. DOI: 10.1103/RevModPhys.88.035009. URL: https://link.aps.org/doi/10.1103/RevModPhys.88.035009.

[29] Conférence générale des poids et mesures, CGPM. *Resolution 1 of the 26th CGPM. On the revision of International System of Units (SI)*. 2018. URL: https://www.bipm.org/en/CGPM/db/26/1/.

[30] Bureau international des poids et mesures, BIPM. *The International System of Units (SI)* 9e *édition*. 2019. URL: https://www.bipm.org/utils/common/pdf/si-brochure/SI-Brochure-9.pdf.

[31] Bureau international des poids et mesures, BIPM. *Le Système international d'unités,* 8e *édition*. 2006. URL: http://www.bipm.org/utils/common/pdf/si_brochure_8.pdf.

[32] W. Bich. "The third-millennium International System of Units". In: *La rivista del Nuovo Cimento della Società italiana di fisica* 42 (2019), pp. 49–102. ISSN: 1826-9850. DOI: 10.1393/ncr/i2019-10156-2. URL: https://doi.org/10.1393/ncr/i2019-10156-2.

[33] Michael Stock et al. "The revision of the SI—the result of three decades of progress in metrology". In: *Metrologia* 56.2 (Feb. 2019), p. 022001. DOI: 10.1088/1681-7575/ab0013. URL: https://doi.org/10.1088%2F1681-7575%2Fab0013.

[34] P. W. Bridgman. *The Logic of Modern Physics*. McMillan, 1927. URL: https://archive.org/details/logicofmodernphy00brid (visited on 07/03/2020).

[35] *The Nobel Prize in Physics 1946*. URL: https://www.nobelprize.org/prizes/physics/1946/summary/ (visited on 07/03/2020).

[36] Amanda Koepke et al. "Consensus building for interlaboratory studies, key comparisons, and meta-analysis". In: *Metrologia* 54.3 (May 2017), S34–S62. DOI: 10.1088/1681-7575/aa6c0e. URL: https://doi.org/10.1088%2F1681-7575%2Faa6c0e.

[37] A. Koepke et al. *NIST Consensus Builder — User's Manual*. National Institute of Standards and Technology. Gaithersburg, MD, 2017. URL: https://consensus.nist.gov.

[38] BIPM et al. *International Vocabulary of Metrology — Basic and General Concepts and Associated Terms*. Joint Committee for Guides in Metrology, JCGM 200:2012. URL: http://www.bipm.org/en/publications/guides/vim.html.

[39] Antonio Possolo et al. "Interpreting and propagating the uncertainty of the standard atomic weights (IUPAC Technical Report)". In: *Pure and Applied Chemistry* 90.2 (2018), pp. 395 –424. DOI: https://doi.org/10.1515/pac-2016-0402. URL: https://www.degruyter.com/view/journals/pac/90/2/article-p395.xml.

[40] W. Bich. "From Errors to Probability Density Functions. Evolution of the Concept of Measurement Uncertainty". In: *IEEE Trans. Instrum. Meas.* 61 (2012), pp. 2153–2159. DOI: `10.1109/TIM.2012.2193696`.

[41] W. Bich. "Error, uncertainty and probability". In: *Proceedings of the International School of Physics "Enrico Fermi", Course CLXXXV "Metrology and Physical Constants"*. Ed. by E. Bava M. Kuehne and A. M. Rossi. Vol. 185. Amsterdam NLD: IOS Press and Società italiana di fisica, 2013, pp. 47–73. ISBN: 978-1-61499-325-4.

[42] Lewis F. L. *Optimal estimation*. John Wiley & Sons, 1986.

[43] Bickel P. J. and Docksum K. A. *Mathematical statistics: Basic Ideas and Selected Topics*. Prentice Hall, 1977.

[44] Rao C. R. *Linear Statistical Inference and its Applications, 2nd ed.* John Wiley & Sons, 1973.

[45] Lee P.M. *Bayesian Statistics: an introduction*. Edward Arnold, 1989.

[46] Luenberger D. G. *Optimization by Vector Space Methods*. John Wiley & Sons, 1969.

[47] Beck J. V. and Arnold K. J. *Parameter estimation in engineering and science*. John Wiley & Sons, 1977.

[48] Dudley Maurice Newitt. "Percy Williams Bridgman, 1882-1961". In: *Biogr. Mems Fell. R. Soc.* 8 (Nov. 1962), pp. 26–40. DOI: `10.1098/rsbm.1962.0003`. URL: `https://royalsocietypublishing.org/doi/10.1098/rsbm.1962.0003`.

[49] Edwin C. Kemble and Francis Birch. *Percy Williams Bridgman, 1882-1961 – A Biographical Memoir*. National Academy of Sciences, 1970. URL: `{http://www.nasonline.org/publications/biographical-memoirs/online-collection.html}`.

[50] W. Bich. "Estimation and uncertainty in metrology". In: *Proceedings of the International School of Physics "Enrico Fermi", Course CXLVI "Recent Advances in Metrology and Fundamental Constants"*. Ed. by S. Leschiutta T. J. Quinn and P. Tavella. Vol. 146. Amsterdam NLD: IOS Press and Società italiana di fisica, 2001, pp. 653–688. ISBN: 1-58603-167-8.

[51] P.J. Huber. *Robust statistics*. John Wiley & Sons, 1981.

[52] Wilcox Rand. *Introduction to robust estimation and hypothesis testing*. Academic Press, San Diego, 1997. DOI: `10.2307/2669876`.

[53] M Cox and K Shirono. "Informative Bayesian Type A uncertainty evaluation, especially applicable to a small number of observations". In: *Metrologia* 54.5 (2017), p. 642. URL: `http://stacks.iop.org/0026-1394/54/i=5/a=642`.

[54] Antonio Possolo and Olha Bodnar. "Approximate Bayesian evaluations of measurement uncertainty". In: *Metrologia* 55.2 (2018), p. 147. URL: `http://stacks.iop.org/0026-1394/55/i=2/a=147`.

[55] Adriaan M H van der Veen. “Bayesian methods for type A evaluation of standard uncertainty”. In: *Metrologia* 55.5 (2018), p. 670. URL: http://stacks.iop.org/0026-1394/55/i=5/a=670.

[56] Antonio Possolo and Adam L Pintar. “Plurality of Type A evaluations of uncertainty”. In: *Metrologia* 54.5 (2017), p. 617. URL: http://stacks.iop.org/0026-1394/54/i=5/a=617.

[57] Antonio Possolo and Clemens Elster. “Evaluating the uncertainty of input quantities in measurement models”. In: *Metrologia* 51.3 (2014), p. 339. URL: http://stacks.iop.org/0026-1394/51/i=3/a=339.

[58] A. O'Hagan and M.G. Cox. *Private communication.* Aug. 17, 2020.

[59] NIST. *NIST Quality Manual for Measurement Services – NIST QM-I.* National Institute of Standards and Technology, U.S. Department of Commerce. Gaithersburg, Maryland, Dec. 2016. URL: https://www.nist.gov/nist-quality-system/.Version10.

[60] Antonio Possolo. “Introducing a Simple Guide for the evaluation and expression of the uncertainty of NIST measurement results”. In: *Metrologia* 53.1 (2016), S17. URL: http://stacks.iop.org/0026-1394/53/i=1/a=S17.

[61] A. Agresti. *Categorical Data Analysis.* Wiley Series in Probability and Statistics. Wiley, 2003. ISBN: 9780471458760. URL: https://books.google.it/books?id=hpEzw4T0sPUC.

[62] Michael Thompson. “What exactly is uncertainty?” In: *Accreditation and Quality Assurance* 17.1 (Feb. 2012), pp. 93–94. ISSN: 1432-0517. DOI: 10.1007/s00769-011-0840-8. URL: https://doi.org/10.1007/s00769-011-0840-8.

[63] Anthony O'Hagan. “Eliciting and using expert knowledge in metrology”. In: *Metrologia* 51.4 (July 2014), S237–S244. DOI: 10.1088/0026-1394/51/4/s237. URL: https://doi.org/10.1088%2F0026-1394%2F51%2F4%2Fs237.

[64] A. Possolo. *Simple Guide for Evaluating and Expressing the Uncertainty of NIST Measurement Results.* NIST Technical Note 1900. Gaithersburg, MD: National Institute of Standards and Technology, 2015. URL: http://dx.doi.org/10.6028/NIST.TN.1900.

[65] W. Bich. “Uncertainty of measurement”. In: *Proceedings of the International School of Physics “Enrico Fermi”, Course CXCVI “Metrology: from Physics Fundamentals to Quality of Life”*. Ed. by P. Tavella M. J. T. Milton M. Inguscio and N. De Leo. Vol. 196. Amsterdam NLD: IOS Press and Società italiana di fisica, 2017, pp. 149–166. ISBN: 9788874381098.

[66] W. Bich. “Recent developments in uncertainty evaluation”. In: *Proceedings of the International School of Physics “Enrico Fermi”, Course CLXVI “Metrology and Fundamental Constants”*. Ed. by A. J. Wallard T. Hänsch S. Leschiutta and M. L. Rastello. Vol. 166. Amsterdam NLD: IOS Press and Società italiana di fisica, 2007, pp. 81–94. ISBN: 978-1-58603-784-0.