The value of uncertainty: the evaluation of the precision of physical measurements and the limits of experimental knowledge

Fabien GRÉGIS
Univ Paris Diderot, Sorbonne Paris Cité, laboratoire Sphere UMR 7219,
F-75205 Paris, France
fabien.gregis@etu.univ-paris-diderot.fr

Short summary

My thesis is organized in three parts, preceded by an introductory chapter (chapter 1). Chapter 1: "Preamble: error and uncertainty" presents my subject matter by introducing the different concepts that will be used in the core of the thesis. These concepts are measurement error and measurement uncertainty, physical quantity and "measurand", measurement model and measurement function. Chapter 1 also explains the context of my study, drawing on the recent history of metrology, which has undergone major developments impulsed by national and international institutions.

The first part of my thesis is dedicated to the probabilistic treatment of measurement uncertainty. It takes a contemporary perspective and examines the technical foundations of uncertainty analysis. It is shown that the technical developments of the 20th century progressively bring out epistemological issues. The second part of my thesis capitalizes on the previous conclusions and proceeds to elucidate the philosophical viewpoints taken by metrologists themselves within their technical elaborations. This part is centred on a particularly controversial concept, the "true value" of a physical quantity. The third part of my thesis turns towards a specific context of use of measurement uncertainty: it extends the scope of parts 1 and 2 by tackling the issue of measurement uncertainty from the viewpoint of precision physics, more specifically through the activity of adjustment of the physical constants. This broader perspective enables me to offer new answers to the issues raised until then.

An epilogue (chapter 12) is dedicated to the recent history of metrology and briefly relates the genesis of the *Guide to the expression of uncertainty in measurement* (GUM). It ends up exploring the institutional and conceptual issues that were encountered during the 1980s and the 1990s in the course of the preparation of the GUM. It shows in particular that there are two levels of influence that guided the content of the document. The first one is institutional: international institutions like ISO, BIPM in particular drove the whole process. However, there is a second level of influence since the significant progresses were attained thanks to a series of individual initiatives.

Part 1. The probabilistic interpretation of measurement uncertainty

Part 1 tackles the concept of measurement uncertainty and its philosophical ramifications by studying how the GUM offers to conceptualize it. By investigating the technical structure elaborated in the GUM and in the metrological literature, I identify how different technical impediments bring out general issues that have a philosophical nature. My study concentrates on the nature and the function of two main statistical models of uncertainty analysis. I especially show that contemporary metrology is filled with open questions about the nature of probabilities that should be used in these models, and that the traditional attachment to a frequentist interpretation of probabilities has been seriously challenged since the beginning of the 1970s. Instead, many articles of the contemporary literature in metrology suggest to rely on a alternate interpretation of probability, called "epistemic". Even more so, an increasing number of metrologists, statisticians and practitioners defend the adoption of a fully Bayesian account of measurement. This move accompanies an outstanding evolution of the conception of measurement, that can be called an "epistemic turn" in metrology. In particular, some metrologists insist on the subjectivity of the measurement activity, standing against the classical perception of measurement as the ultimate objective means of investigation of the physical world. This generates nourished discussions in the field. I endeavour to elucidate the reasons why metrologists came to criticize the classical conception of measurement that was common by the middle of the 20th century, and to look for superior alternatives.

Chapter 2: "General foundations of uncertainty analysis" presents the common elements that serve as the basis to uncertainty analysis and to the introduction of probabilities in its calculus. In particular, the chapter discusses the dichotomy between so-called "random" and "systematic" errors.

This enables me to dive into the frequentist model of measurement uncertainty, which is the main matter of chapter 3: "The frequentist approach of measurement". One of the crucial points of my analysis consists in identifying which role probabilities play in this model, and in reconstructing the core set of hypotheses of the frequentist approach. I show that the frequentist calculus of measurement uncertainties cannot take into account all the components of measurement error, and that this has been regarded as a serious limit of the approach by metrologists since the beginning of the 1970s.

The use of epistemic probabilities in metrology is intended to overcome the limitations of the frequentist account, and it leads to the development of a Bayesian conceptualization of measurement. This is discussed in chapter 4: "From epistemic probabilities to a Bayesian model of measurement", which proceeds in two stages. First, I show how epistemic probabilities answer to a specific problem raised of the traditional approach, namely that it renders a probabilistic treatment of so-called "systematic errors" impossible. Secondly, I explain how metrologists and statisticians have proposed to extend this solution to so-called "random errors" so as to point to a fully Bayesian approach which displays for some parts an explicitly subjectivist epistemology.

In the end, two archetypal approaches of measurement are opposed: they interpret measurement uncertainty in different ways. One, frequentist, insists on the objectivity of measurement and revolves around an ideal of accuracy, this latter concept being defined in metrology as the closeness of agreement between the measurement result and the so-called "true value" of the quantity intended to be measured. The other one, Bayesian, underlines the subjectivity of the measurement activity and engages an ideal of rationality oriented towards the best expression of a given state of knowledge. This opposition generates a lot of debates in the specialized literature, that I review in chapter 5: "Discussion: the epistemological ramifications of the statistical debate". This leads me into raising two specific questions. The first one relates to the status of the "true value" of a physical quantity in the context of contemporary metrology. The second question concerns the dialogue between two essential concepts of measurement, "measurement uncertainty" for one part and "measurement accuracy" for the other part. These two questions set the stage for the two next parts of my thesis.

Part 2. The probabilistic interpretation of measurement uncertainty

Part 2 focuses on the notion of “true value” of a physical quantity, intuitively defined as the value that would be obtained out of a perfect measurement, *i.e.* of a measurement that is not subject to error. However, this definition is circular since, precisely, measurement error is quantitatively defined as the deviation to the true value. Indeed, the “true value” issue goes deeper into the foundations of measurement, and interrogates the possibility of a description of physical quantities by numerical values. Classical discussions in philosophy gather a whole span of viewpoints, from strongly realist positions where the physical world is quantitative by nature, to anti-realist or coherentist viewpoints. A classical account is the representational theory of measurement; while debated, this account enables to discuss the conditions under which the use of mathematics in the empirical sciences is meaningful, and also reminds how the value of a quantity is an element of representation. While not digging into deep philosophical developments, metrologists have themselves criticized the concept of true value on different occasions. This criticism has recently spread in the specialized literature where some propositions try to make this concept disappear from the vocabulary and from the formalism of uncertainty analysis and measurement in general. These attempts rejoin an underlying idea according to which avoiding “true value” enables to simplify the conceptualizations of metrologists and scientists in general by silencing metaphysical issues and other problems considered as purely philosophical. I distinguish two types of arguments developed by metrologists, that I study one after the other: an argument of “unknowability”, and an argument of “non-uniqueness”. I defend that the notion of “true value” can be accommodated to the valid points that these arguments make.

Chapter 6: “The “unknowability” argument” addresses the first attack against the concept of true value. The true value of a quantity is viewed as a an ideal, forever unknown and unknowable concept, since one can never guarantee that an experiment is free from error. Therefore, it could be useful to dispense with it. I offer two answers to this argument. Drawing on the developments of part 1, I show that the main statistical models of uncertainty analysis, including the Bayesian one, do not eliminate the true value from their technical formalism. As a consequence, the criticism of true value does not impact its use but rather its interpretation. Then, I explore the possible interpretations of the notion of “true value” by contrasting realist with empiricist conceptions of a physical quantity, and by exploring the different theories of truth in contemporary philosophy and the way they relate to the term “true” in “true value”. I conclude that the formulation of the unknowability argument by metrologists is rather weak, even if there is an actual philosophical issue behind it. I interpret the criticisms of metrologists as displaying essentially an anti-metaphysical stance that is not an anti-realist viewpoint. In the end, if the philosophical issue cannot itself be easily solved, as it relates to deep long-standing problems of general philosophy, it remains interesting to see how a philosophical positioning can impact the practices of scientists. To this regard, the attachment to the notion of “true value” has epistemic virtues, especially as it gives a framework for keeping open a never-ending process of correction that guides experimental progress.

Chapter 7: “True value: non-uniqueness and definitional uncertainty” scrutinizes the second argument that I identified against the notion of “true value”. This argument states that measurands are vague concepts, as they rely on a idealized representation that conceals lots of details in the behaviour of physical objects and physical phenomena. Indeed, it is rarely possible to conceive of a unique true value for a given measurand since its definition is often incomplete. But then, if there is no such thing as the “unique” “true value” of a quantity, how can this value even be said to be “true”? Here again the argument concerns the very possibility to use mathematics to represent physical phenomena, and in particular to use numbers to represent physical quantities. I suggest that the argument of non-uniqueness can be examined by mobilizing considerations about reductionism in science and about the epistemology of approximation. I claim that the notion of “true value” of a quantity continues to have a proper meaning provided that one acknowledges the approximate character of scientific knowledge and scientific theories (including the mathematical laws displayed in these theories). In that case, it should be preferable to talk about an “approximately true value”. According to this position, the notion of “definitional uncertainty” that metrologists have introduced to deal with the problem of non-uniqueness appears as an estimation of the limit of precision not of the experiments and the measurements performed, but of the theoretical models that are used at different levels of the description of the physical quantities in order to provide a framework in which these measurements can be done.

Part 3. The adjustment of the physical constants: a different perspective on the relationship between uncertainty and accuracy

In Part 3, I extend the scope of my research by looking for new answers to the previous issues. To this end, I analyze works of precision physics. I offer a study of a specific practice, called the “adjustment of the physical constants”, that was initiated in 1929 by Raymond T. Birge in a pioneer article, and which is now perpetuated under the auspices of an international institution, the CODATA (Committee on Data for Science and Technology). I raise two working questions: how can different measurement results, obtained in different conditions, be combined together? How can scientists agree on the value of a quantity? This leads me to question the use that scientists make of measurement uncertainty within this specific practice. I end up discussing the relationship between uncertainty and accuracy that was first raised in Part 1.

Chapter 8: “Combination of measurement results and least-squares adjustments” goes back to the 18th century and concentrates on two episodes of the history of astronomy that illustrates how scientists have progressively begun to accept the idea of a “combination of observations”. This expression, dubbed by Stephen Stigler, expresses the fact that measurement results are not anymore considered individually, independently of each other, but that scholars acknowledge that one can jointly use different measurement results obtained by different people in different conditions with different instruments, and even about different phenomena. In this episode, I identify what I think marks the beginnings of an idea of *reproducibility* of a measurement, attached to a careful survey of errors, which in turn suggests the foundations of an idea of measurement uncertainty. This episode also marks the beginnings of a statistical treatment of errors which was later on systematized by Legendre and Gauss with the “method of least-squares", essential to the practice of adjustments.

In chapter 9: "The adjustment of the physical constants: Birge's initiative", I proceed to discuss the adjustments of the physical constants themselves in the course of the 20th century. The adjustments of the physical constants are a practice which consists in collecting all the best-to-be-known measurement results obtained about a pool of important physical constants (often called "fundamental constants") which are linked together trough a web of overdetermined equations. The adjustment ends up in mixing all these empirical values in order to provide a set of "best values" or "recommended values" of these constants at a given time. I inspect Birge's initial proposal in his 1929 and 1932 articles. The combination of observations appears as constitutive of his initiative. I expose the role that is accorded to measurement uncertainty in the comparison and the combination of measurement results. It is shown that measurement uncertainty is essential to judge the discrepancy or the agreement between different results, but that it cannot be properly interpreted as estimating the amplitude of a measurement error or as expressing a degree of belief.

I pursue the study of the adjustments in Chapter 10: "The treatment of discrepant data in the adjustment of the physical constants" which is dedicated to a period of about thirty years following Birge's initiative. This periods ends around 1970, when physicists, metrologists and statisticians gather to discuss the practice of adjustments, its legitimacy, and the specific and problematic issue of the treatment of discrepant data. A conservative approach defends that the most important thing to ensure is that the "best values" obtained out of the adjustments are as accurate as possible, even if this means that the measurement uncertainties attached to these values have to be expanded. A second position, defended by influential actors of the field, claims on the contrary that results have to be *precise* even if this costs accuracy. This position rests on an argument according to which only precise statements can be properly tested and that only them offer a good insight on the coherence and the validity of scientific theories. Therefore, this position is oriented toward the *long-term* and focuses not on the immediate accuracy of results but on scientific *progress* by engaging a never-ending process of identification and correction of measurement errors. The CODATA still keeps important traces of this position, as shown by a contemporary example of a similar issue concerning the measurement of the proton radius.

Chapter 11: "Uncertainty and accuracy" then concludes the part 3 with a more systematic account of the relationship between uncertainty and accuracy. It is argued that the "long-term progress" position defended by metrologists enables to conciliate the metaphysical-frequentist and epistemic-Bayesian standpoints and to understand the interpretations given in both approaches not as exclusive but rather at complementary. It is also argued that this "long-term progress" position corresponds to a research perspective, which sees accuracy not as a static but as a dynamical concept, but which does not accommodate all uses of measurements. Indeed, one can think of an applied perspective (for engineers, industry, ...) where users focus on the immediate accuracy of measurement and have to think in terms of risks. In that case, users are not satisfied by the promise of a hypothetical progress in the future.