Atmospheric
Measurement
Techniques

Truth and uncertainty. A critical discussion of the error concept versus the uncertainty concept

Thomas von Clarmann[1], **Steven Compernolle**[2], **and Frank Hase**[1]

[1]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany
[2]Department of Atmospheric Composition, Royal Belgian Institute for Space Aeronomy (BIRA-IASB),
1180 Brussels, Belgium

Correspondence: Thomas von Clarmann (thomas.clarmann@kit.edu)

Abstract. Contrary to the statements put forward in "Evaluation of measurement data – Guide to the expression of uncertainty in measurement", edition 2008 (GUM-2008), issued by the Joint Committee for Guides in Metrology, the error concept and the uncertainty concept are the same. Arguments in favor of the contrary have been analyzed and found not to be compelling. Neither was any evidence presented in GUM-2008 that "errors" and "uncertainties" define a different relation between the measured and true values of the variable of interest, nor does this document refer to a Bayesian account of uncertainty beyond the mere endorsement of a degree-of-belief-type conception of probability.

1 Introduction

It has long been recognized that quantitative characterization of the reliability of a measurement is essential for drawing quantitative conclusions from the measured data. Various and often contradictory methods and terminologies emerged over the years. The Towards Unified Error Reporting (TUNER) activity aims at a unification of the reporting of errors in estimates of atmospheric state variables retrieved from satellite measurements (von Clarmann et al., 2020). At the request of the Bureau International de Poids et Mesures (BIPM), the Joint Committee for Guides in Metrology (JCGM) issued a guideline for how measurement uncertainty should be dealt with (JCGM, 2008, this source is henceforth referenced as GUM-2008). A Supplement in the context of GUM is found in JCGM (2012), and several supplements to GUM-2008 are found on the BIPM website (https://www.bipm.org/

en/publications/guides, last access: 28 February 2022). The concept of uncertainty was developed long before the GUM-2008 was issued and has seen several refinements since then (e.g., Eisenhart and Collé, 1980; Collé, 1987; Colclough, 1987; Schumacher, 1987). A key claim of GUM-2008 is that the terms "error" and "uncertainty" connote different things and that the underlying concepts are different. GUM-2008 has been critically discussed by, e.g., Bich (2012), Grégis (2015), Elster et al. (2013), and The European Centre for Mathematics and Statistics in Metrology (2019) and more favorably by, e.g., Kacker et al. (2007).

In this paper we critically assess some of the claims made in GUM-2008 and, as part of the TUNER activity, discuss its applicability to remote sensing of the atmosphere. We start by analyzing GUM's claim about the differences between error and uncertainty (Sect. 2), whereby it is important to distinguish between terminological (Sect. 2.1) and conceptual (Sect. 2.2) issues. We find that the concept of the "true value of the measurand" makes up the alleged key difference. That is to say, the uncertainty concept endorsed by GUM is claimed, contrary to traditional error analysis, to be able to dispense with the concept of the true value that is neither known nor knowable (GUM-2008, pp. 3 and 5). This leads to the question of whether, and, if so, how the measured or estimated value along with the estimated error (or uncertainty) are related to the true value the measurand has in reality and what the problems related to the ignorance of the true value are in the context of error estimation (Sect. 3). In this context, we first address the question of whether it is adequate to use the true value, which is typically unknown and unknowable, in the definition of the term "error" and

to base error analysis on such a definition (Sect. 3.1). Second, we investigate the implications that the inverse nature of a measurement process has on the probabilistic relationship between the measured value, the true value, and the measurement (Sect. 3.2). In this context we discuss the problem of the base-rate fallacy. Further, we investigate whether the alleged difference between the error concept and the uncertainty concept can be explained by a Bayesian turn in metrology. Third, we assess the degree to which the nonlinearity of the relationship between the measured signal and the target quantity viz. the radiative transfer equation poses additional problems (Sect. 3.3), and fourth, we scrutinize the claim that there will always be unknown sources of uncertainty and that it is thus impossible to relate the measured value along with its uncertainty estimate to the true value (Sect. 3.4). After these more general considerations we critically discuss the applicability of the GUM-2008 concept to indirect measurements of atmospheric state variables (Sect. 4). There we discuss the problems of measurands that are not well defined in the sense of GUM-2008 (Sect. 4.1) and whether it is really adequate to report the combined error only (Sect. 4.2). Finally (Sect. 5), we conclude the degree to which the arguments put forward by the JCGM are conclusive and what the differences between the error concept and the uncertainty concept actually are.

2 Error and uncertainty

GUM-2008 endorses a new terminology compared to that of traditional error analysis. In the context of the work undertaken by the TUNER activity, a project aiming at unification of error reporting of satellite data (von Clarmann et al., 2020), terminological and conceptual divergence is particularly problematic. Without agreement on the concepts and the terminology of error versus uncertainty assessment, any unification is out of reach.

According to GUM-2008, the concept of uncertainty analysis should replace the concept of error analysis. The International Vocabulary of Metrology document (JCGM, 2012) points in the same direction. Thus, some conceptual and terminological remarks seem appropriate. While, on the face of it, this is quibbling about words, actually, conceptual differences between the errors and uncertainties are claimed to exist. This issue is discussed in the following.

A key claim of GUM-2008 is that the "concept of *uncertainty* as a quantifiable attribute is relatively new in the history of measurements, although *error* and *error analysis* have long been part of the practice of measurement science or metrology" (JCGM, 2008, Sect. 0.2, l. 1–2; emphases in the original.). In a note to their Sect. 3.2.3, GUM-2008 states that "The terms 'error' and 'uncertainty' should be used properly and care taken to distinguish between them." The discussion of these issues is occasionally led astray, because it often does not distinguish between two different questions: first, whether the terms "error" and "uncertainty" have the

same connotation, and second, whether the underlying concepts are indeed different. In the following, we try to shed some light on these issues.

2.1 Terminological issues

Already in the pre-GUM language there have been subtle linguistic differences between the terms "error" and "uncertainty". The error has been conceived as an attribute of a measurement or an estimate, while the term "uncertainty" has been used as an attribute of the true state, or, more precisely, an attribute of an agent's knowledge about the true state. We perfectly know the result of our measurement – even if it is erroneous – but we are uncertain about the true value. Because of the measurement error there is an uncertainty as to what the true value is. The uncertainty thus describes the degree of ignorance about the true value, while the estimated error describes the degree to which the measurement is thought to deviate from the true value. In this use of language, both terms still relate to the same concept. This notion seems, as far as we can judge, to be consistent with the language widely used in the pre-GUM literature since Gauss (1809), who used the Latin terms *error* and *incertitudo* in this way. Thus, both terms referred to the same thing but from a different perspective[1]. The estimate of the total error includes both measurement noise and all known components of further errors, random or systematic, caused by imperfections in the measurement and data analysis system.

It must, however, be noted that the term "error" is an equivocation. It has been used for both the unknown and unknowable signed actual difference between the measured value and the true value of the measurand and for a statistical estimate of it. The statistical estimate is mostly understood to be the square root of the variance of the probability density function of the error[2][3] and thus does not carry any information about the sign of the error. Nonlinear error propagation may in some cases make asymmetric error estimates necessary, but typically these do not carry any information about the actual sign of the error either. The ignorance of the sign of the error entails that the true or most probable value can-

---

[1] Possolo (2021) expresses this construal in more colorful words: "[... M]easurement uncertainty surrounds the true value of the measurand like a fog that obfuscates it, while measurement error is both the source of that fog and part and parcel of the measured value. Measurement uncertainty thus describes the doubt about the true value of the measurand, while measurement error quantifies the extent to which the measured value deviates from the true value."

[2] When we use variances and standard deviations, we do not mean sample variances and sample standard variations but simply the second central moment of a distribution or its square root. In accordance with GUM-2008, this distribution can represent a probability in the sense of personal belief and thus can also include systematic effects. See also Sect. 2.2.

[3] Other estimates are also used, e.g., robust ones like the interquartile range.

not simply be determined by subtracting the estimated error from the measured value.

One of the first major documents where the term "*error*" was used with this statistical connotation is, to the best of our knowledge, "Theoria Motus Corporum Celestium" by Gauss (1809). Since then, the term "error" has commonly been used to signify a statistical estimate of the size of the difference between the measured and true values of the measurand. Seminal publications by Gauss (1816), Pearson (1920), Fisher (1925), Rodgers (1990), and Mayo (1996) furnish evidence of this use of the term "error". The "estimated error" (as a composite term) is often understood to be a measure of the width of a distribution around the measured value which tells the data user the probability density of a certain value to be measured if the value actually measured was the true value. One might criticize equivocation of the traditional language, but one can equally well consider this to be a non-issue and trust that the context will make it clear what is meant. Often, some attributes are used for clarification and specification, e.g., "probable error" (Gauss, 1816; Bich, 2012), "statistical error" (Nuzzo, 2014), "error estimation" (Zhang et al., 2010), or "error analysis" (Rodgers, 1990, 2000; Hughes and Hase, 2010).

More recently, GUM-2008 presented a narrower definition for how we should conceive the term "error" and stipulated a new terminology, where the term "measurement uncertainty" is used in situations where one would have said "measurement error" before. According to GUM-2008 (p. 2), the uncertainty of a measurement is defined as "a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand". Conversely, GUM-2008 (Annex B.2.19) allows for the term "error" only the connotation "signed difference", but their use of the terms "error and error analysis" in the first sentence of their Sect. 0.2 or "possible error" in their Sect. 2.2.4 only makes sense if the statistical meaning of the term "error" is conceded.

In spite of the explicit definition in GUM-2008, there seems to be no unified stance among GUM-2008 endorsers as to what "error" is. For example, Merchant et al. (2017) maintain that "error" connotes only the signed difference, while Kacker et al. (2007) or White (2016) refer to "error" as a statistical estimate. Kacker et al. (2007) complain that GUM-2008 is often misunderstood, and we suspect that the cause of this might be that GUM-2008 is indeed not sufficiently clear with respect to the differences between the underlying error and uncertainty concepts.

The use of the term "uncertainty" in GUM-2008 seems inconsistent: the general GUM-2008 concept seems to be that the "error" has to include all error sources and thus cannot be known; "uncertainty" is weaker: it is only an estimate of quantifiable errors, excluding the unknown components. This view is supported by the following quotation (GUM-2008, p. viii): "It is now widely recognized that, when all of the known or suspected components of error have been evaluated and the appropriate corrections have been applied, there still remains an uncertainty about the correctness of the stated result, that is, a doubt about how well the result of the measurement represents the value of the quantity being measured." It is not fully clear what this means. One possible reading is that they use the term "error" in the redefined sense, i.e., as a quantity which measures the actual deviation from the true value. Then this statement would be a mere truism, just saying that after all correction and calibration activities there is still a need for error (in the error concept terminology) estimation. The only other possible reading is that they want to say that, due to unknown (unrecognized and/or recognized but not quantified[4]) error sources, error estimation will always be incomplete and there remains an additional uncertainty not covered by the error estimation. This often is very true, but the use of the term "uncertainty" would then be inconsistent in their document, because here the connotation of "uncertainty" is the unknown (unquantified or even unrecognized) part of the error, which by definition cannot be assessed, while in the main part of their document, the connotation of "uncertainty" seems to be a quantified statistical estimate. In summary, it is not clear whether the "uncertainty" includes the unknown error terms or not.

The introduction of the term "uncertainty of measurement" seems to us a mere linguistic revision of an established terminology which does not connect to any further insights. The issue of whether the term "error" should also be used for a statistical estimate cannot be judged on scientific grounds. It is a matter of stipulation, although in the main body of GUM-2008 this stipulation is presented as if it were a factual statement ("In this Guide, great care is taken to distinguish between the terms 'error' and 'uncertainty'. They are not synonyms, but represent completely different concepts; they should not be confused with one another or misused.", Sect. 3.2, Note 2). The synonymity of "error" and "uncertainty" is thus neither true nor false but adequate or inadequate. Instead of quibbling about words, we will, in the next section, concentrate on the concepts behind these terms.

2.2 Conceptual issues

Although GUM-2008 (Sect. 0.2) claims the "concept of uncertainty as a quantifiable attribute to be a relatively new concept in the history of measurement", we uphold the view that it has long been recognized that the result of a measurement remains to some degree uncertain even when a thorough measurement procedure and error evaluation are performed. Investigators realized already in the 19th century that measurement results always have errors. Gauss (1809) and Legendre (1805) formalized the required procedure of balancing imperfect astrometric measurements by least squares fitting, in support of orbital calculations from overdetermined

---

[4]Rigorously speaking, within the concept of subjective probability, recognized but unquantified uncertainties should not exist.

data sets, and there is no reason to believe that earlier investigators were unaware of the fact that they were not working on perfect observational data. The conclusion of Kepler (1609) concerning the elliptical shape of the orbit of Mars based on the rich observational data set collected by Brahe would have been impossible without proper implicit assumptions concerning the limited validity of the reported values. A rich methodological toolbox for error estimation and uncertainty assessment has been developed since then, including systematic errors and error correlations.

GUM-2008 not only presents traditional error analysis in a revised language, but also suggests that there is more to it. That is to say, the entire concept is claimed to be replaced (see, e.g., GUM-2008, Sect. 3.2.2, Note 2). We understand that GUM-2008 grants that the classical concept of error analysis deals with statistical quantities, but these are statistical estimates of the difference between the measured or estimated value and the true value. We take GUM-2008 to be saying that the reference of even this statistical quantity to the true value poses certain problems, because the true value is unknown and unknowable. As a solution of this problem, the uncertainty concept is introduced, which allegedly makes no reference to the true value of the measurand and is thus hoped to avoid related problems. GUM-2008 (particularly Sect. 2.2.4) unfortunately leaves room for multiple interpretations, but our reading is that an error distribution is understood by GUM-2008 to be a distribution whose dispersion is the estimated statistical error and whose expectation value is the true value, while an uncertainty distribution is understood to be a distribution whose dispersion is the estimated uncertainty and whose expectation value is the measured or estimated value.

GUM-2008 (p. 5) characterizes error as "an idealized concept" and states that "errors cannot be known exactly". This is certainly true, but it has never been claimed that errors can be known exactly. Since not all relevant error sources are necessarily known, any error estimate remains fallible, but still it is and has always been the goal of error analysis to provide error estimates that are as realistic as possible. To use the statistical conception of "error" and, conceding the fallibility of its estimated value, it is not necessary to know the true value. It is only necessary to know the chief mechanisms which can make the measured value deviate from the true value and to have estimates available of the uncertainties of the input values to these mechanisms.

Some GUM-2008 endorsers (e.g., Kacker et al., 2007) try to draw a borderline between error analysis and uncertainty assessment in a way that they associate error analysis with frequentist statistics while uncertainty is placed in the context of Bayesian statistics. Frequentist statistics, we understand, is a concept where the term "probability" is defined via the limit of frequencies for a sample size approaching infinity. This definition is challenged because it involves a circularity: it is based on the large number theorem, according to which (strong version) a frequency distribution will almost certainly converge towards its limit. This limit is then associated with the probability. "Almost certainly" means "with probability 1". The circularity is given by the fact that the definiendum appears in the definiens (see, e.g., Stegmüller, 1973, pp. 27). Also, the weak version of the large number theorem involves the concept of probability and thus poses a similar problem to the definition of the term "probability". We concede that many estimators in error estimation rely on frequency distributions. It is, however, a serious misconception to conclude from this that error analysis is based on a frequentist definition of "probability". This is simply a non sequitur. Frequency-based estimators are consistent with any of the established definitions of probability, and their use does not allow any conclusion about the definition of "probability" in use.

The conceptual differences between error analysis and uncertainty analysis seem to come down to the different relations between the measured and true values of the measurand. In GUM-2008 (pp. 3 and 5), the claim is made that the uncertainty concept can be construed without reference to the unknown and unknowable true value while the error concept cannot (GUM-2008, p. 3) and that the uncertainty concept is more adequate because there can always exist unknown error sources which entail that an error budget can never be guaranteed to be complete (GUM-2008, p. viii). It is stated that the uncertainty concept is not inconsistent with the error concept (GUM-2008, p. 2/3). There are, however, certain inconsistencies and shortcomings, which are discussed in the following.

One of the major purposes of making scientific observations, besides triggering ideas on possible relations between quantities, is to test predictions based on theories on the real world (Popper, 1935). To decide whether an observation corroborates or refutes a hypothesis, it is necessary to have an estimate of how well the observation represents the true state, because it must be decided how well any discrepancy between the prediction and the observation can be explained by the observational error (e.g., Mayo, 1996). Any concept of uncertainty that is not related to the true state cannot serve this purpose.

On page 3, GUM-2008 says that the attribute "true" is intentionally not used within the uncertainty concept because truth is not knowable. In GUM-2008 (p. 59) it is claimed that the uncertainty concept "uncouple[s] the often confusing connection between uncertainty and the unknowable quantities 'true' value and 'error' ". The term "measurand" in their definition, however, is defined as the quantity intended to be measured (JCGM, 2009); GUM-2008 (p. 32) says basically the same; GUM-2009 (p. 20) says that the "quantity" is the same as the "true quantity value". Inserting this definition into the GUM-2008 definition of uncertainty shows that, through the back door, uncertainty still refers to the true value. Thus it is not clear what the difference between the traditional concept of error analysis and the uncertainty concept is. Further, it is stated that systematic effects can contribute

to the uncertainty. GUM-2008 falls short of clarifying how a systematic effect can be understood other than as a systematic deviation between the measurement and the true value that the concept GUM-2008 apparently tries to avoid. In order to justify the attribution of an uncertainty distribution to the systematic effects without relying on frequentist statistics, they invoke the concept of subjective probability. With this it becomes possible to assign an uncertainty distribution to the combined random and systematic uncertainty, but still it is not clear how the systematic effect is defined without reference to the unknown truth.

Subjective probability reflects the personal degree of belief (GUM-2008, p. 39). Thus, a knowledge-dependent concept of probability is used in GUM-2008. This approach has been chosen to allow the treatment of systematic errors as dispersions, although the systematic error does not vary and cannot thus be characterized by a distribution in a frequentist sense (GUM-2008, p. 60). If we construe "estimated error" and "estimated value" as parameters of a distribution assigning to each possible value the probability (in a Bayesian context) or the likelihood (in a maximum likelihood context[5]) that it is the true value, no knowledge of the true value is required. This is because, by definition, the subjective probability distribution merely represents the knowledge of the person generating it. In GUM the error concept is discarded because the capability of conducting an error estimate allegedly depends on the knowledge of the true value. However, once the concept of subjective probability has been invoked, no objective knowledge of the unknowable true value is needed any longer. The subjectivist can work with the value they believe to be true. This solves the alleged problem of the error concept, namely, that the true value is unknown.

There is nothing wrong with the subjectivist concept of probability, nor are we attacking the possibility of combining random and systematic errors in a single distribution. This concept, however, makes the knowledge of the true value and the true error unnecessary, and still the estimated error can be conceived as a statistical estimate of the absolute difference between the measured value and the true value. We consider it untenable and inconsistent to refer to the concept of subjective probability when it comes in handy and to deny it when it would solve the conflict between the error and uncertainty concepts.

Our skepticism about the possibility of dispensing with the concept of the true value is shared by, e.g., Ehrlich (2014), Grégis (2015), and Mari and Giordani (2014). Note that in the International Vocabulary of Metrology (known as VIM) (JCGM, 2012), although also issued by the JCGM, the concept and definition of the true value are explicitly retained.

In GUM-2008 (p. 2/3), it is claimed that the concept of uncertainty "is not inconsistent with other concepts of uncertainty of measurement, such as a measure of the possible error in the estimated value of the measurand as provided

by the result of a measurement [or] an estimate characterizing the range of values within which the true value of the measurand lies[6] (VIM:1984 definition 3.09). Although these two traditional concepts are valid as ideals, they focus on *unknowable* quantities: the 'error' of the result of a measurement and the 'true value' of the measurand (in contrast to the estimated value), respectively. Nevertheless, whichever *concept* of uncertainty is adopted, an uncertainty component is always *evaluated* using the same data and related information ..." (emphases in the original). It remains unclear how the concepts can, on the one hand, be consistent, while, on the other hand, it is claimed that the error approach and the uncertainty approach are actually conceptually different and not only with respect to terminology. Since both concepts, however, are consistent, it is not clear what the difference in the concepts consists of.

Interestingly enough, early documents of the history of GUM (Kaarls, 1980; Bureau International des Poids et Mésures, 1980) provide evidence that the terminological turn from "error" to "uncertainty" was triggered only by linguistic arguments, based upon the fact that in common language the term "uncertainty" is often associated with "doubt", "vagueness", "indeterminacy", "ignorance", or "imperfect knowledge". These early documents provide no evidence that "error" and "uncertainty" were conceived as two different technical terms connoting different concepts. Any reinterpretation of the terms "error" and "uncertainty" as frequentist versus Bayesian terms or as operational versus idealistic concepts came later.

The answer to the terminological differences was found to be contingent upon the underlying stipulation and that any statement about their equivalence or difference without reference to a definition is a futile pseudo-statement. The answer to the question of conceptual differences is less trivial and deserves some deeper scientific discussion. The main question still seems to be how the true value, the error or uncertainty, and the measured value are related to each other. This question will be addressed in the following section.

3   The unknown true value of the measurand

The alleged key problem of the error concept is, in our reading of GUM-2008, that the value of the true value of the measurand is not known and that this true value must appear neither in the definition of any term nor in the recipes to estimate it. To better understand this key problem, we decompose it into four sub-problems.

1. Quantities whose value cannot be determined must not appear in definitions.

2. The error distribution must not be conceived as a probability density distribution of a value to be the true value.

---

[5]See Sect. 3.2 for a deeper discussion of this issue.

[6]It is not clear how this can be achieved without explicit consideration of the Bayes theorem.

3. Nonlinearity issues pose problems in error estimation if the true value is not known, at least in approximation.

4. One can never know that the uncertainty budget is complete because it can always happen that a certain source of uncertainty has been overlooked; thus, the full error estimate is an unachievable ideal, and thus the estimated error does not provide a link between the measured value and the true value.

Some of these sub-problems are in some way formulated in GUM-2008, but it is not exactly specified there why the fact that the true value of the measurand is unknowable poses a problem to the scientist applying traditional error estimation. We have formulated others as devil's advocates, which are intended to serve as working hypotheses to critically discuss the error and uncertainty concepts in the context of indirect measurements. In the following we will scrutinize these theses one after the other.

3.1 The operational definition

GUM-2008 tries to avoid using the true value of the measurand in the definition of the term "uncertainty". This strategy is employed because the true value of the measurand is "not knowable" (GUM-2008, p. 3). It may be puzzling why it should be necessary to know the value of a quantity to use it in the definition of a term. The heights of the Colossus of Rhodes or the Lighthouse of Alexandria are well-defined quantities, although we have no chance to measure them today[7]. Also, we might have a clear physical conception of what the temperature in the center of the Sun might be, although it may not be practicable to put a thermometer there, and we might even be unable to figure out any other, more sophisticated, method to assign an accurate observation-based value to this quantity. Intuitively, we conceive the definition of a quantity and the assignment of the value to a quantity as quite different things.

In GUM-2008 it is claimed that the definition of "uncertainty" is an operational one (p. 2). An operational definition defines a quantity by stipulating a procedure by which a value is assigned to this quantity. The concept of operational definitions was suggested by Bridgman (1927) in order to give terms in science a clear-cut meaning. This operationalism, or at least a narrow conception of it, has its own problems, has received considerable criticism, and has led to deep philosophical discussions (see, e.g., Chang, 2019). To summarize these is beyond the scope of this paper, and for here it must suffice to mention that there are alternatives, such as theoretical definitions or the reduction of the definiendum to previously defined terms.

GUM-2008's claim that the uncertainty concept is based on an operational definition leads to two further inconsistencies. First, no unambiguous operation is stipulated on which the definition can be based, but multiple operations are proposed, which might give different uncertainty estimates. Thus, the definition is void. Our critical attitude with respect to operationalism in the context of GUM-2008 is shared, e.g., by Mari and Giordani (2014).

The other problem with the operational definition is the following: in GUM-2008 (pp. 2–3), it is claimed that the uncertainty concept is not inconsistent with the error concept, and a few lines later it says that "an uncertainty component is always *evaluated* using the same data and related information" (emphasis in the original). The latter suggests that within the error concept the same operations are used as within the uncertainty concept. Since the operations define the term and the related concept, the uncertainty concept and the error concept must be the same.

In summary, the fact that the true value of the measurand is unknowable is a problem for the definition of the term "error" and its statistical estimates only if we commit ourselves to the doctrine that only operational definitions must be used. If we abandon this dogma, there is nothing wrong with conceiving of the estimated error as a statistical estimate between the measured or estimated value and the true value, and the problem is restricted to the assignment of a value to this quantity. Related issues are investigated in the following.

3.2 Measurements as inverse processes

Many conceptions of measurement models exist which relate the measured value to the true value and, depending on the context, one can be more adequate than another (Possolo, 2015). GUM-2008 recommends a model that conceives the estimate of the true value of the measurand as a function of the measured value. Since in remote sensing of the atmosphere multiple atmospheric states can cause the same set of measurements and the measurement function would thus be ambiguous, we prefer a different concept, as outlined in the following.

The causal error points from the true value to the measured signal. Thus, the estimation of the true value from a measured value can be conceived as an inverse process. An argument along this line of thought, but in a context wider than that of remote sensing of the atmosphere, has been put forward by Possolo and Toman (2007). The inverse characteristic of the estimation problem is particularly true for indirect measurements, e.g., remote sensing, but direct measurements can easily be conceived as indirect measurements. When reading the thermometer, we actually read the length of the mercury column (the measured value), apply inversely the law of thermal expansion, and get an estimate of the temperature. In trivial cases, when a measurement device has a calibrated scale from which the target quantity can be directly read, the inverse process is effectively pre-tabulated in the scale. Only

---

[7]We owe this illustrative example to Possolo (2021).

in these cases are the measured value and the estimate of the measurand the same.

With a transfer function F available that approximately describes the process that links the true value x of interest to the measured value, the expected measured signal $y_{\text{expected}} = F(x)$ can be estimated. The distribution of the measurement error around y_{expected} describes the probability of any value y to be measured.

Conversely, for a given measurement y_{measured}, the inversion of the transfer function allows us to estimate the true value x. If a genuine inversion of the transfer function is not possible due to ill-posedness of the inverse problem in the sense of Hadamard (1902), workarounds like least squares methods or regularized inversion schemes are available (see, e.g., von Clarmann et al., 2020, for a summary of some methods of particular relevance for remote sensing). Counterintuitively, however, in general, the estimate will not be the most probable value of x, nor will the mapping of the measurement error distribution into the x space yield the probability distribution of any value as the true value. This holds even if the error distribution is extended to also include systematic effects and if all error correlations are adequately taken into account in the case of multi-dimensional measurements. It is the theorem of Bayes (1763) which makes the difference. The only inverse scheme where such a probabilistic interpretation is valid in the x space is the maximum a posteriori method (Rodgers, 2000), which employs a Bayesian estimator.

The non-consideration of the Bayes theorem goes under the name of "base-rate fallacy"; 50 % of people suffering from Covid-19 have fever (Robert Koch Institut, 2020), but this does not imply that the probability is 50 % that a person with fever has Covid-19. To estimate the latter probability requires knowledge of the percentage of people being infected with the Corona virus and the probability that a person will suffer from fever for any reason. In metrology the situation is quite analogous. There are three possible solutions to cope with this problem.

The first solution is to apply a retrieval scheme that is based on a Bayesian estimator. Examples are found, e.g., in Rodgers (2000) or von Clarmann et al. (2020). On the supposition that the error budget is complete, the interpretation of the error bar as the dispersion of a distribution representing the probability density that a certain value is the true value is correct. The problem with this approach is that often there is no firm a priori knowledge about the value of the measurand available.

The second solution is the application of the principle of indifference, as applied, e.g., by Gauss (1809). That is, the same a priori probability is assigned to all possible values of the measurand. With this, e.g., in the application to a linear inverse problem and normal distributions of uncertainties, the Bayesian solution collapses back to a simple unconstrained least squares solution. Due to the assumption of the equidistribution of the a priori probabilities, the estimated uncer-

tainty of the estimate can still be interpreted as the width of the probability density function of the true value of the measurand. This concept of "non-informative a priori", however, has its own problems. Even if we ignore some more trivial problems for the moment, e.g., that some quantities cannot, by definition, take negative values, this concept can lead to absurdities: if we assume that we have no knowledge about, say, the volume density of small-particle aerosols in the atmosphere and describe this missing knowledge by an equidistribution of probabilities, this would correspond to a non-equidistribution of the surface densities, due to the nonlinear relationship between surface and volume. It strikes us as absurd that information can be generated just by such a simple transformation from one domain into another. The principle of indifference, upon which the concept of noninformative priors is built, is critically but still favorably discussed, e.g., by Keynes (1921, chap. IV). The concept of non-informative priors is still criticized even in the Bayesian community (e.g., D'Agostini, 2003).

The third solution is the likelihood interpretation, which has been introduced by Fisher (1922). The likelihood that the true value is x if the measured signal is y equals the probability density that y is measured if the true value is x. No prior information is considered. Solution of the inverse problem by maximizing the likelihood of x does not provide the most probable estimate of x, and accordingly the error bar of the solution must not be interpreted as the width of a probability distribution of the true value. Application to a linear inverse problem and normal distributions of uncertainties renders formally the same estimator as the Gaussian least squares solution, but its interpretation has changed. It can no longer be interpreted as the maximum of a probability density function of the true value. If need be, in some cases, i.e., if the inverse problem is well posed enough to allow an unconstrained solution, the maximum likelihood estimate can be post factum transformed into a Bayesian estimate by application of the Bayes theorem.

We concede that the interpretation of a measured value as the most probable true value is problematic. This implies that the interpretation of the error estimate as the width of a distribution around the true value is also not generally valid. These problems could justify some reluctance with regard to the concept of the true value. This argument, involving the base-rate fallacy, however, is not invoked in GUM-2008.

Some interpretations of GUM-2008 (e.g., White, 2016; Kacker et al., 2007) associate it with a Bayesian conception of probability and seem to suggest that error estimation and uncertainty analysis are best distinguished in the sense that the former relies on frequentist statistics while the latter is founded on Bayesian statistics. Thus one might suspect that "uncertainty" is simply the Bayesian replacement of error. Here the following remarks are in order.

1. Many of the methods presented in GUM-2008, including their "Type-A evaluation (of uncertainty)", which is

the "method of evaluation of uncertainty by the statistical analysis of series of observations" (p. 3), are from the frequentist toolbox. Gleser (1998) finds that the methods suggested in GUM are neither fully frequentist nor fully Bayesian. Furthermore, it is not quite clear which of Bayes' methods and principles a scientist has to use to be a Bayesian (cf. Fienberg, 2006), since the Bayes theorem is also accepted by non-Bayesians, and the use of maximum likelihood methods, introduced by the almost "militant" frequentist (Fisher, 1922), does, as far as we can judge, not commit one to using a frequentist definition of the term "probability". The GUM-2008 does not provide a clear reference to a specifically Bayesian uncertainty analysis method. GUM-2008 makes reference to Jeffreys (1983) as an authority of the degree-of-belief concept of probability. Jeffreys, however, offers no clue as to what the difference between "error" and "uncertainty" might be. In the context of measurements or observations, Jeffreys always uses the term "error" (e.g., *op. cit.*, p. 72), and often we find statements like "the probable error [...] is the uncertainty usually quoted" (op. cit., p. 72), "no uncertainty beyond the sampling errors" (op. cit., p. 389), or "treat the errors as independent" (op. cit., p. 443). With the statement that errors are not mistakes (op. cit., p. 13), Jeffreys explicitly contradicts the GUM pioneers (Kaarls, 1980) and GUM-2008 endorsers (Merchant et al., 2017). Also, Press (1989) is referenced by GUM-2008 only to defend the use of a subjective concept of probability but not in a context aiming at the clarification of the alleged difference between "error" and "uncertainty".

2. If the uncertainty concept was indeed based on a Bayesian framework, it would be astonishing why it does not in the first place require one to apply the Bayes theorem to convert the likelihood distributions to a posteriori probability distributions. The methodology proposed in GUM-2008 is uncertainty propagation. This is a mere forward (or direct) problem: given that x_{true} is the true value and a measurement procedure with some error distribution, it returns a probability distribution for values x_{measured} that might be measured. However, GUM-2008's definition of uncertainty, "parameter, associated with *the result of a measurement*, that characterizes the dispersion of *the values that could reasonably be attributed to the measurand*" (emphasis added by us), seems associated with another meaning: given a measured value ("result of a measurement") and a measurement procedure with some error distribution, what is the probability density distribution of the "values that could reasonably be attributed to the measurand" being the true one? This is an inverse problem for which the Bayes theorem is applicable rather than uncertainty propagation.

3. Interestingly enough, Willink and White (2012), who also use the term "uncertainty" in a frequentist framework, report that the turn to the new terminology happened already in 1980/81 and make a strong case that various allegedly purely Bayesian concepts of GUM-2008 can be given a valid frequentist interpretation.

Thus, we reject the hypothesis that the uncertainty concept as presented in GUM-2008 is a Bayesian concept. Bayesianism does not help to understand the claimed differences between the error concept and the uncertainty concept.

3.3　Nonlinearity issues

The uncertainty concept relies on the possibility of evaluating uncertainties caused by measurement errors and "systematic effects" without knowledge of the true value. This is certainly granted for linear problems. Here the uncertainty estimates do not depend on the value of the measurand. This is because in the linear case Gaussian error propagation of the type

$$\sigma_b^2 = \left(\frac{\partial b}{\partial a}\right)^2 \sigma_a^2 \tag{1}$$

or its multi-dimensional variant

$$\mathbf{S}_b = \mathbf{J}\mathbf{S}_a\mathbf{J}^T \tag{2}$$

holds. Here a and b are scalar variables, σ_a^2 and σ_b^2 are the variances characterizing their errors, $\mathbf{S}_a$ and $\mathbf{S}_b$ are the error covariance matrices of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, and $\mathbf{J}$ is the Jacobian with elements $\frac{\partial b_i}{\partial a_j}$.

For nonlinear problems the situation is more complicated because $\frac{\partial b}{\partial a}$ or $\mathbf{J}$ depends on a or $\boldsymbol{a}$, respectively, and Gaussian error propagation is valid only in approximation. Within the concept of error propagation, the concept of moderate nonlinearity (Rodgers, 2000) can be invoked. That is to say, the estimated value of the measurand is assumed to be a reasonably good approximation of the measurand, and the partial derivatives needed for Gaussian error estimation are evaluated at this estimate. If the resulting error bars are small enough to ensure that the range covered by the interval defined by the estimated value plus/minus the error bar is confined to the range where linear approximation is justifiable, then the error estimates are, while less than perfect, still far better than useless.

The endorser of the uncertainty concept has a problem if they want to stay consistent with their doctrine. Since knowledge of the true value is denied, it is not clear how Gaussian error estimation can be applied to the propagation of uncertainties, because it is not clear for which value of the measurand the required partial derivatives shall be evaluated.

On the face of it, Monte Carlo error estimation or other variants of ensemble-based sensitivity studies can serve as an alternative. These, however, also invoke the nonlinear model

that links the measured signal with the measurand, and uncertainty estimates thus still depend on the choice of the estimate that represents the true value; any choice of this value which is not closely related to the true value of the measurand will produce uncertainty estimates which are recalcitrant against any interpretation. Monte Carlo and related methods, however, are apt for the estimation of the error budget, including the systematic effects if f is too nonlinear to justify Gaussian error estimation if approximate knowledge of the measurand is conceded.

In summary, the evaluation of uncertainties in the case of nonlinearity poses a problem to the scientist who denies approximate knowledge of the true value of the measurand, because the uncertainty estimate depends on the assumed value of the measurand, and it must be assumed that it represents the true value reasonably well. Within the framework of error analysis this assumption is allowed, and measurement errors as well as systematic effects can thus also be evaluated for nonlinear inverse problems.

3.4 Incompleteness of the error budget

The arguments put forward above are based on the supposition that the error budget is complete. Beyond measurement noise, the total error budget includes systematic effects in the measured signal, uncertainties in parameters other than the measurand that affect the measured signal, and effects due to the chosen inverse scheme. If our reading of GUM-2008 is correct, then the most severe criticism by GUM-2008 of the "error concept" is that one can never be sure that the error budget is indeed complete and that the error estimate thus does not characterize the difference between the value estimated from the measurement value and the true value.

The precision of a measurement is a well-behaved quantity in the sense that it is testable in a straightforward way: from at least three sets of collocated measurements of the same quantity, where each set is homogeneous with respect to the expected precision of its measurements, the variances of the differences provide unambiguous precision estimates (see, e.g., McColl et al., 2014, or Stoffelen, 1998). The situation is more difficult for biases. Persistent differences between different measurement systems do not tell us what the bias of one measurement system with respect to the – unfortunately unknowable – truth is, because the comparison measurement system may be biased as well. Even if the number of measurement systems is quite large, it is not guaranteed that the mean bias of all of them is zero, and an infinite number of measurement systems is out of reach in the real world. Up to that point we concede that a positive proof of the completeness of the error budget is impossible, but this is not the end of the story.

A falsificationist (Popper, 1935) approach is more promising. It follows the rationale that it will never be possible to prove that our assumptions about the bias of a measurement system are correct. Instead, we estimate the bias as well as

we can and use it as a best estimate of the bias until some test provides evidence that the estimate is incorrect. Such a test typically consists of the intercomparison of data sets from different measurement systems. If the bias between these data sets is larger than the combined systematic error estimates, at least one of the systematic error estimates is too low and has to be refuted. Further work is then needed to find out which of the measurement systems is most likely to underestimate its systematic error. Conversely, as long as the mean difference of the measurements of the same measurand can be explained by the combined estimate of the systematic errors of both measurement systems, the systematic error estimates can be maintained, although this is, admittedly, no proof of the correctness of the error estimates. However, as long as severe tests as described above are executed and the error estimates cannot be refuted, it is rational to believe that they are sufficiently complete.

4 The applicability of GUM-2008 to remote sensing of the atmosphere

In this section we identify issues where GUM-2008 clashes with the needs of error or uncertainty estimation in the field of remote sensing of atmospheric constituents and temperature. These issues are (1) that since the atmospheric state varies quasi-continuously in space and time, the measurand is not well defined, and (2) there are applications of atmospheric data where the total uncertainty estimate alone does not help.

4.1 What if the measurand is not well defined?

On macroscopic scales, atmospheric state variables vary continuously in space and time. On microscopic scales, the typical target quantities, concentrations, or temperature are not even defined. A typical example of this problem is the volume mixing ratio (VMR) of a certain species at a point in the atmosphere (see also von Clarmann, 2014). The determination of a quantity like this requires a canonical ensemble of air, but in the real, inhomogeneous, atmosphere, this quantity does not exist. It is an uninstantiated ideal. Due to these inhomogeneities, the air volume sounded must be infinitesimally small; i.e., it must approach a point. In the real atmosphere there is either a target molecule at this point (VMR = 1) or another molecule (VMR = 0) or no molecule at all (undefined VMR due to division by zero). Thus, one measures only averages over finite inhomogeneous air volumes. This approach, supposedly the only possible approach, clashes with the premise of GUM-2008[8] that the measurand needs to be well defined. Measuring atmospheric state variables requires the specification of the region the average is

---

[8]In GUM-2008 this problem is recognized, but no solution is offered; the term "definitional uncertainty" is introduced in this context but is not applied in practice.

made over. The relevant toolbox of atmospheric data characterization includes concepts like resolution and averaging kernels (see Rodgers, 2000, for details). Since this type of measurement is apparently out of the scope of GUM-2008, the latter is quite silent with respect to solutions to the problem of the characterization of measurements of quantities that are not well defined. Broadening the scope and applicability of the GUM-2008 framework to include less than ideally defined measurands and measurements that demand inverse methods would significantly increase the value and utility of the GUM-2008 approach. Relevant recommendations on data characterization developed within the TUNER activity (von Clarmann et al., 2020) aim at helping to reach this goal.

4.2 The combined error

One of the positive aspects of GUM-2008 is that it breaks with the misleading concept of characterizing systematic errors with "safe bounds" (Kaarls, 1980; Kacker et al., 2007; Bich, 2012). This concept was sometimes endorsed by error statisticians subscribing to frequentism. Within a frequentist concept of probability, a probabilistic treatment of systematic errors was not easily possible because, due to its systematic nature, a systematic error cannot easily[9] be characterized by a frequency or probability distribution. The concept of subjective probability solves this problem. With the subjectivist's toolbox, it is no longer a problem to assign probability density functions, standard deviations, and so forth when characterizing systematic errors. This possibility is a precondition for aggregating systematic and random errors to give the total error. GUM-2008, however, goes a step further and even denies the necessity of reporting random and systematic errors independently. Here we have to raise severe objections.

von Clarmann et al. (2020) explicitly recommend that error estimates be classified as random or systematic[10]. In contrast, GUM-2008 (E.3.3/E3.7) states that "In fact, as far as the calculation of the combined standard uncertainties [...] is concerned, there is no need to classify uncertainty components and thus no real need for any classificational scheme." If indeed meant as written, we challenge the claim that a total combined error budget is sufficient and that therefore no classification scheme is needed at all. Characterizing the measurement of a unique quantity, e.g., the value of a natural

constant agreed upon by the calibration authorities, by a single error margin might be sufficient. However, most measurements, and particularly those of atmospheric state variables such as temperature and concentrations of trace species, deal with quantities varying with time and space. Any sensible use of the resulting data sets requires a clear distinction between statistical and systematic error budgets. For example, for time series analysis targeted at the determination of trends, the total error budget is of no use, but the random error budget is needed instead. This is because any purely additive systematic error component cancels out in this application, and its consideration in the error budget would unduly distort the weights of the data points available. In summary, the denial of the importance of distinguishing between random errors and systematic errors does not provide proper guidance and altogether is a strong misjudgment. The data users must be provided with all information required to tailor the relevant error budget to the given application of the data.

Benevolent readers of GUM-2008 take the GUM authors to be saying only that the aggregation of estimated errors to give the total error budget follows the same rules for systematic and random errors and that the criticized statement is not meant to deny the importance of distinguishing between random and systematic errors beyond the mere aggregation process. If this reading is correct, we agree, but here GUM-2008 leaves room for interpretation.

5 Conclusions

We have mentioned above that the uncertainty concept depends on the acceptance of the subjective probability in the sense of degree of rational belief. Without that, an error budget including systematic effects would make no sense because systematic effects cannot easily be conceived as probabilistic in a frequentist sense; that is to say, the resulting error cannot be conceived as a random variable in a frequentist sense. Being forced to adopt the concept of probability as a degree of rational belief, it makes perfect sense to conceive, after consideration of the Bayes theorem (see Sect. 3.2), the distribution with expectation $\hat{x}$ and covariance based on $\sigma_{x,\text{total}}$ as the probability distribution which tells the rational agent the probability of any value being the true value.

The denial that a valid connotation of the term "error" is a statistical characterization between a measured or estimated value and the true value of the measurand would be an attempt to brush away centuries of scientific literature. This is, however, a matter of stipulation or convention and thus beyond the reach of a scientific argument. We thus take GUM-2008 to be conceding that both the concepts, error analysis and uncertainty assessment, aim at providing a statistical characteristic of the imperfectness of a measurement or an estimate. We understand GUM-2008 in the sense that the problem of the error concept is that it conceives the estimated

---

[9]The qualification not *easily* was chosen because frequentists still might sample over multiple universes or apply other measures to squeeze systematic errors in a frequentist concept.

[10]In this context it is important to note that, in contrast to some older conceptions, von Clarmann et al. (2020) define "systematic errors" as bias-generating errors and "random errors" as variance-generating errors. To avoid confusion with the older conceptions, one can instead use the descriptive terms "persistent" and "volatile" errors as suggested by Possolo (2021). This is not done here to maintain consistency with von Clarmann et al. (2020).

error as a statistical measure of the difference between the measured or estimated values and the true value. Since the true value is unknowable, according to GUM-2008 the term "error" can neither be defined nor its value known.

It has been shown that the problem of the unknown true value of the measurand is a problem for the definition of terms like "error" or "uncertainty" only if the concept of an operational definition is pursued. This concept, however, has its own problems and is by no means without alternatives. As soon as the concept of an operational definition is given up, problems associated with defining the estimated error as a statistical estimate of the difference between the measurement or estimate and the true value of the measurand disappear, and the problem remaining is only one of assigning a reasonable value to this now well-defined quantity.

Since GUM-2008 did not provide many reasons why, in the context of indirect measurements, the error allegedly cannot be estimated without knowledge of the true value or why an uncertainty distribution does not tell us anything about the true value, we list the most obvious ones one could put forward to bolster this claim. These are the problem of the base-rate fallacy, the problem of nonlinearity, and the problem that one can never know that the error budget is complete. The problem of the base-rate fallacy can be solved by either performing a Bayesian inversion or by conceiving the resulting distribution as a likelihood distribution. Astonishingly enough, the GUM-2008's "dispersion or range of values that could be reasonably attributed to the measurand" is determined without explicit consideration of prior probabilities and thus cannot be interpreted in terms of posterior probability. The problem of nonlinearity can be solved either by assuming that the estimate is close enough to the true value and linearizing around this point or by Monte Carlo studies. A GUM-oriented scientist, who has to avoid referring to the true value, is at a loss in the case of nonlinearity because any estimate of the uncertainty of the estimate will be correct only when evaluated at the true value or an approximation of it. The problem of the unknown completeness of the error budget can be tackled by performing comparisons between measurement systems. While this will never provide positive proof of the completeness of the error budget, it still justifies rational belief in its completeness, and if error or uncertainty distributions are conceived as subjective probabilities in the sense of degrees of rational belief, this is good enough. In summary, if (a) our reading of GUM-2008 is correct in the sense that the traditional error analysis can deal with a statistical quantity and that the key difference between the "error" and "uncertainty" concepts is their relation to the true value of the target quantity and (b) that our list of arguments against the error concept is complete, and finally if (c) our refutation of these arguments is conclusive, then the claim that the "error" concept and the "uncertainty" concepts are fundamentally different is untenable.

Beyond this, reasons have been identified that bring the applicability of the GUM-2008 concept to atmospheric mea-

surements into question. At least we can state that GUM-2008, by presenting its terminological stipulation about the terms "error" and "uncertainty" in the guise of a factual statement, has triggered a linguistic discussion that distracted attention away from the more important issues of how the principles of error or uncertainty estimation, whatever one prefers to call them, could be made better applicable to measurements beyond the idealized cases covered by their document.

Review statement. This paper was edited by Nathaniel Livesey and reviewed by Antonio Possolo and two anonymous referees.

References

Bayes, T.: An essay towards solving a problem in the doctrine of chances, Phil. Trans., 53, 370–418, https://doi.org/10.1098/rstl.1763.0053, 1763.

Bich, W.: From Errors to Probability Density Functions. Evolution of the Concept of Measurement Uncertainty, IEEE T. Instrum. Meas., 61, 2153–2159, https://doi.org/10.1109/TIM.2012.2193696, 2012.

Bridgman, P. W.: The Logic of Modern Physics, Macmillan, New York, OCoLC 609483443, 1927.

Bureau International des Poids et Mésures: Report on the BIPM enquiry on error statements, Tech. Rep. BIPM-80/3, Bureau International des Poids et Mésures, F-92310, https://www.bipm.org/en/publications/rapports-bipm-pre-1990 (last access: 1 March 2022), 1980.

Chang, H.: Operationalism, in: The Stanford Encyclopedia of Philosophy, edited by: Zalta, E. N., Metaphysics Research Lab, Stanford University, Winter 2019 edn., https://plato.stanford.edu/archives/win2019/entries/operationalism/ (last access: 28 February 2022), 2019.

Colclough, A. R.: Two theories of experimental error, J. Res. Nat. Bur. Stand., 92, 16–185, https://doi.org/10.6028/jres.092.016, 1987.

Collé, R.: Minutes of the meeting on measurement uncertainties, NCSL Newsletter, 27, 52–55, 1987.

D'Agostini, G.: Bayesian Reasoning in Data Analysis: A critical Introduction, World Scientific, Singapore, ISBN 9789812383563, 2003.

Ehrlich, C.: Terminological aspects of the Guide to the Expression of Uncertainty in Measurements, Metrologia, 4, S145, https://doi.org/10.1088/0026-1394/51/4/S145, 2014.

Eisenhart, C. and Collé, R.: Postscript to expression of the uncertainties of final results, in: NBS Communications Manual for Scientific, Technical, and Public Information, edited by: Solomon, C. W., Bograd, R. D., and Tilley, W. R., Exhibit 2-E, 2-30–2-32, U.S. Dept. of Commerce, National Bureau of Standards, Gaithersburg, MD, https://catalog.hathitrust.org/Record/011389799 (last access: 9 September 2021), 1980.

Elster, C., Klauenberg, K., Bär, M., Allard, A., Fischer, N., Kok, G., van der Veen, A., Harris, P., Cox, M., Smith, I., Wright, L., Cowen, S., Wilson, P., and Ellison, S.: Novel mathematical and statistical approaches to uncertainty evaluation in the context of regression and inverse problems, in: 16th International Congress of Metrology, Paris, France, edited by: Filtz, J.-R., Larquier, B., Claudel, P., and Favreau, J.-O., 7–10 October 2013, 04003, https://doi.org/10.1051/metrology/201304003, 2013.

Fienberg, S. E.: When did Bayesian Inference become "Bayesian"?, Bayesian Anal., 1, 1–40, https://doi.org/10.1214/06-BA101, 2006.

Fisher, R. A.: On the Mathematical Foundations of Theoretical Statistics, Phil. Trans. R. Soc. Lond. A, 222, 309–368, 1922.

Fisher, R. A.: Statistical Methods For Research Workers, Oliver and Boyd, London, http://psychclassics.yorku.ca/Fisher/Methods/ (last access: 28 February 2022), 1925.

Gauss, C. F.: Theoria Motus Corporum Coelestium, F. Perthes and I. M. Besser, Hamburg, https://www.worldcat.org/title/theoria-motus-corporum-coelestium, last access: 28 Feb 2022, 1809.

Gauss, C. F.: Bestimmung der Genauigkeit von Beobachtungen, Z. Astronom. Verwandte Wis., 1–2, 185–196, 1816.

Gleser, L. J.: Assessing uncertainty in measurement, Statistical Science, 13, 277–290, 1998.

Grégis, F.: Can we dispense with the notion of "true value" in metrology?, in: Standardization in Measurement: Philosophical, Historical and Sociological Issues, edited by: Schlaudt, O. and Huber, L., Pickering & Chatto, London, 81–93, ISBN 9780367598761, 2015.

Hadamard, J.: Sur les Problèmes aux Dérivées Partielles et Leur Signification Physique, Princeton University Bulletin, 13, 49–52, 1902.

Hughes, I. G. and Hase, T. P. A.: Measurements and their Uncertainties: A practical guide to modern error analysis, Oxford University Press, USA, Oxford, UK, ISBN 9780199566334, 2010.

Jeffreys, H.: Theory of Probability, Oxford University Press, Oxford, 3rd edn., original edition: 1939, ISBN 9780198531937, 1983.

Joint Committee for Guides in Metrology (JCGM): Evaluation of measurement data – Guide to the expression of uncertainty in measurement, JCGM 100:2008(E) – in English Evaluation of measurement data, Pavillon de Breteuil, F-92312 Sèvres CEDEX, 1st edn., https://www.bipm.org/en/publications/guides (last access: 28 February 2022), 2008.

Joint Committee for Guides in Metrology (JCGM): Evaluation of measurement data – An Introduction to the "Guide to the expression of uncertainty in measurement" and related documents, JCGM 104:2009 An introduction to the "GUM" and related documents, Pavillon de Breteuil, F-92312 Sèvres CEDEX, 1st edn., https://www.bipm.org/en/publications/guides (last access: 28 February 2022), 2009.

Joint Committee for Guides in Metrology (JCGM): International vocabulary of metrology – Basic and general concepts and associated terms (VIM), 3rd edn., Pavillon de Breteuil, F-92312 Sèvres CEDEX, https://jcgm.bipm.org/vim/en/ (last access: 27 July 2021), 2012.

Kaarls, R.: Report of the BIPM working group on the statement of uncertainties (1st meeting – 21 to 23 October 1980) to the Comité International des Poids et Mésures, Tech. rep., BIPM, Paris, http://www.bipm.org/utils/common/pdf/WGUncertainties1980.pdf (last access: 12 November 2020), 1980.

Kacker, R., Sommer, K.-D., and Kessel, R.: Evolution of modern approaches to express uncertainty in measurement, Metrologia, 44, 513–529, https://doi.org/10.1088/0026-1394/44/6/011, 2007.

Kepler, J.: Astronomia Nova aitiologetos seu physica coelestis, tradita commentariis de motibus stellae Martis ex observationibus G.V. Tychonis Brahe, Pragae, https://archive.org/details/ioanniskepplerih00kepl (last access: 28 February 2022), 1609.

Keynes, J. M.: A Treatise on Probability, MacMillan and Co., Limited, London, https://www.gutenberg.org/files/32625/32625-pdf.pdf (last access: 28 February 2022), 1921.

Legendre, A.-M.: Nouvelles méthodes pour la détermination des orbites des comètes, F. Didot, Paris, https://openlibrary.org/works/OL13109117W/Nouvelles_m%C3%A9thodes_pour_la_

d%C3%A9termination_des_orbites_des_com%C3%A8tes (last access: 28 February 2022), 1805.

Mari, L. and Giordani, A.: Measurement error and uncertainty, in: Error and uncertainty in scientific practice, edited by: Boumans, A., Horn, G., and Petersen, A., Pickering and Chattoo, London, 79–96, ISBN 9781138662278, 2014.

Mayo, D. G.: Error and the growth of experimental knowledge, The University of Chicago Press, Chicago, ISBN 0226511987, 1996.

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., and Stoffelen, A.: Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target, Geophys. Res. Lett., 17, 6229–6236, 2014.

Merchant, C. J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A. C., Reuter, M., Sathyendranath, S., Sandven, S., Sofieva, V. F., and Wagner, W.: Uncertainty information in climate data records from Earth observation, Earth Syst. Sci. Data, 9, 511–527, https://doi.org/10.5194/essd-9-511-2017, 2017.

Nuzzo, R.: Scientific method: Statistical errors, Nature, 506, 150–152, https://doi.org/10.1038/506150a, 2014.

Pearson, K.: The fundamental problem of practical statistics, Biometrika, 13, 1–16, 1920.

Popper, K.: Logik der Forschung, Julius Springer Verlag, Wien, mohr Siebeck, Tübingen, 2002; English Edition, "The Logic of Scientific Discovery", Routledge, London, 2002, ISBN 9780415278447, 1935.

Possolo, A.: Simple Guide for Evaluating and Expressing the Uncertainty of NIST Measurement Results, Tech. Rep. NIST Technical Note 1900, U.S. Department of Commerce, National Institute of Standards and Technology, Gaithersburg, MD, https://doi.org/10.6028/NIST.TN.1900, 2015.

Possolo, A.: Truth and Uncertainty at the Crossroads, https://doi.org/10.5194/amt-2021-157-RC1, 2021.

Possolo, A. and Toman, B.: Assessment of measurement uncertainty via observation equations, Metrologia, 44, 464–475, https://doi.org/10.1088/0026-1394/44/6/005, 2007.

Press, S. J.: Bayesian statistics: principles, models, and applications, Wiley, New York, NY, ISBN 9780471637295, 1989.

Robert Koch Institut: SARS-CoV-2 Steckbrief zur Coronavirus-Krankheit-2019 (COVID-19), https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html#doc13776792bodyText2, last access: 18 May 2020.

Rodgers, C. D.: Characterization and error analysis of profiles retrieved from remote sounding measurements, J. Geophys. Res., 95, 5587–5595, 1990.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, vol. 2 of Series on Atmospheric, Oceanic and Planetary Physics, F. W. Taylor, World Scientific, Singapore, New Jersey, London, Hong Kong, ISBN 981-02-2740-X, 2000.

Schumacher, R. B. F.: A dissenting position on uncertainties, NCSL Newsletter, 27, 55–59, 1987.

Stegmüller, W.: Probleme und Resultate der Wissenschaftstheorie und Analytische Philosophie, Band IV, Personelle und Statistische Wahrscheinlichkeit, Studienausgabe Teil D 'Jenseits von Popper und Carnap:' Die logischen Grundlagen des statistischen Schließens, Springer-Verlag, Berlin Heidelberg New York, ISBN 0387060413, 1973.

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, J. Geophys. Res., 103, 7755–7766, 1998.

The European Centre for Mathematics and Statistics in Metrology: Regression and Inverse Problems, http://mathmet.org/research/regression/, last access: 18 February 2019.

von Clarmann, T.: Smoothing error pitfalls, Atmos. Meas. Tech., 7, 3023–3034, https://doi.org/10.5194/amt-7-3023-2014, 2014.

von Clarmann, T., Degenstein, D. A., Livesey, N. J., Bender, S., Braverman, A., Butz, A., Compernolle, S., Damadeo, R., Dueck, S., Eriksson, P., Funke, B., Johnson, M. C., Kasai, Y., Keppens, A., Kleinert, A., Kramarova, N. A., Laeng, A., Langerock, B., Payne, V. H., Rozanov, A., Sato, T. O., Schneider, M., Sheese, P., Sofieva, V., Stiller, G. P., von Savigny, C., and Zawada, D.: Overview: Estimating and reporting uncertainties in remotely sensed atmospheric composition and temperature, Atmos. Meas. Tech., 13, 4393–4436, https://doi.org/10.5194/amt-13-4393-2020, 2020.

White, D. R.: In pursuit of a fit-for-purpose uncertainty guide, Metrologia, 53, 107–124, https://doi.org/10.1088/0026-1394/53/4/s107, 2016.

Willink, R. and White, R.: Disentangling Classical and Bayesian approaches to uncertainty analysis, Tech. Rep. CCT/12-07, Comité Consultatif de Thermométrie, BIPM, Sèvres, https://www.bipm.org/documents/20126/28435677/working-document-ID-5058/8fd49ede-0e53-66b1-050a-edeae2c2f62c (last access: 28 February 2022), 2012.

Zhang, Z., Lin, T., and Chu, J.: Remote sounding of atmospheric pressure profile from space, part 3: error estimation, J. Appl. Rem. Sens., 4, 043536, https://doi.org/10.1117/1.3458870, 2010.