

Traceable measurements and calibration: a primer on uncertainty analysis

JANAE CSAVINA,^{1,†} JOSHUA A. ROBERTI,¹ JEFFREY R. TAYLOR,² AND HENRY W. LOESCHER^{1,3}

¹*Battelle–National Ecological Observatory Network (NEON), 1685 38th Street, Boulder, Colorado 80301 USA*

²*Institute of Technology Campus, Nova Scotia Community College, 5685 Leeds Street, Halifax, Nova Scotia B3K 2T3 Canada*

³*Institute of Alpine and Arctic Research (INSTAAR), University of Colorado, Boulder, Colorado 80301 USA*

Citation: Csavina, J., J. A. Roberti, J. R. Taylor, and H. W. Loescher. 2017. Traceable measurements and calibration: a primer on uncertainty analysis. *Ecosphere* 8(2):e01683. 10.1002/ecs2.1683

Abstract. Describing the quality of measurements is necessary to understand the level of confidence in any observation. Accuracy, precision, trueness, repeatability, reproducibility, and uncertainty are all used to describe quality of measurement, but the terms are inconsistently defined and measured and thus easily misunderstood. One purpose of quality parameters is for the comparison of observations, but when dissimilar methods for estimating quality terms are utilized, a comparison is misrepresented. A standardized approach to estimating uncertainty provides a basis for meeting measurement requirements and providing a level of confidence for observations. Here, we show the approach used by the National Ecological Observatory Network to estimate uncertainty of the calibration processes and measurements illustrated with an example of uncertainty assessment on a temperature sensor. Detailing the approach for uncertainty assessment provides the transparency necessary for network science and allows for the approach to be adopted in the scientific community. Reporting uncertainty with all measurements needs to become consistent and commonplace across disciplines.

Key words: accuracy; measurement quality; National Ecological Observatory Network (NEON); precision; Special Feature: Uncertainty Analysis; uncertainty.

Received 23 September 2016; revised 30 November 2016; accepted 1 December 2016. Corresponding Editor: Ruth D. Yanai.

Copyright: © 2017 Csavina et al. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† **E-mail:** jcsavina@battelleecology.org

INTRODUCTION

The concept of uncertainty is at the core of what certainty we have in scientific knowledge. However, the understanding and reporting of uncertainty is not consistent across disciplines (Aleksandrov 2001). Further, the lack of understanding about uncertainty can cause anxiety in those unfamiliar with this notion (Stemwedel 2014).

Accuracy and precision are commonly reported but the manner in which they are quantified differs among disciplines. For example, accuracy has been misreported as error, bias, or standard

deviation (SD) of one or more populations of data (Hickey et al. 2008, Prenesti and Gosmaro 2015). The concept of precision also has been ambiguously used, often describing repeatability and reproducibility or just repeatability alone (ISO 1994, Menditto et al. 2007, Prenesti and Gosmaro 2015). These terms are not standardized in metrology terminology as quantitative quality metrics, but are used qualitatively (ISO 1994).

The lack of consistency in reporting metrology terms in studies and literature does a disservice to the advancement of science. Take, for example, the simple circumstance of comparing sensor

performance from two different manufacturers who quote accuracy and precision but quantify the terms in different ways (Agilent 2006, Toro 2015). Having accuracy and precision estimated in different ways limits comparative study across observations from different sources and the evaluation of changes in measurement uncertainties over time.

Governing metrology bodies such as International Standards Organization (ISO) recognized the lack of consistency in reporting uncertainty and created a Working Group called the Joint Committee for Guides in Metrology (JCGM), made up of individuals nominated by their member organizations. An extensive metrological guide called the Evaluation of measurement data—guide to the expression of uncertainty in measurement (GUM) was the result of the working group (ISO 1995, JCGM 2008). The primary objective of the GUM was to standardize the evaluation methods for uncertainty. The GUM also chose to promote the use of “uncertainty,” rather than other terms (e.g., accuracy), for the quantified estimate of the quality of measurement. Standards for terminology used by the GUM are reported in the international vocabulary of basic and general terms in metrology (JCGM 2012). We recognize that many of these terms are foreign to the ecological community, so the standardized vocabulary used in metrology is defined in Table 1.

To assess uncertainty with a standardized approach requires traceability in measurement to international standards, national standards, or first principles. This traceability is sometimes lacking in autonomous, as opposed to network, science. Limited knowledge for estimates of uncertainty and high uncertainty in some findings were found to be knowledge gaps from the 2014 International Panel of Climate Change (IPCC) synthesis report (IPCC 2015). In part, these knowledge gaps are why network approaches are being integrated into science. One of the benefits of a network approach to data generation is the use of consistent and standardized methods applied to all measurements, which can reduce and provide better estimates of uncertainty.

A high-level requirement of many environmental networks or observatories (e.g., the National Ecological Observatory Network [NEON], the Integrated Carbon Observing System [ICOS], and the

Terrestrial Ecosystem Research Network [TERN]) is to provide consistent, long-term, multi-scaled ecological datasets used in the context of research and education (Peters et al. 2014). For scientific utility of network data, users need to have open access to all data with the associated documentation and metadata, for example, measurement methodology, sensor types, protocols, standard operating procedures, and documentation of the theoretical basis of the measurement. This transparency and utility is also achieved by providing uncertainty budgets with all measurements, which creates a benchmark to compare different measurements from other networks (network-level quality assurance).

National Ecological Observatory Network is an environmental observatory charged by the National Science Foundation with providing the data to enable long-term ecological forecasting. Thus, a core underlying mission of NEON is to assess, manage, and report the associated uncertainty and observations over a decadal-scale operational lifespan. Additionally, to provide ecological forecasting, quality control (QC), including automated checks of the data, is central to network function to assure consistency and assurance over its lifespan. Here, we will use NEON examples of the types of functions that are needed for long-term QC.

National Ecological Observatory Network has an in-house metrology laboratory (calibration, validation, and audit laboratory [CVAL]) to calibrate and validate all sensors used in the observatory. An alternative (and common) option is to outsource this activity to other laboratories, which creates additional challenges for managing a QC program. For instance, there would be an ongoing need to assess the data quality from different laboratories (at an additional cost). Further, the use of different standards introduces additional sources of uncertainty. Additionally, an outsourced operational model would have turnover of laboratories for the duration of NEON's 30-year lifespan whose QC would have to be managed individually. Hence, to meet network goals, in-house metrology by CVAL provides the most cost-effective means to provide these quality provisions and reduce uncertainty. These goals are achieved by consistently managing the quality of NEON's sensor network to meet quality requirements over a 30-year period; providing baseline

Table 1. Metrologic terminology and definitions used in the text.

Metrologic term	Definition
Accuracy	The closeness of the agreement between the result of a measurement and a true value of the measurand. "Accuracy" is a qualitative concept. The term precision should not be used for "accuracy" (JCGM 2008)
Calibration, validation, and audit laboratory	NEON's in-house metrology laboratory which calibrates and validates network sensors traceable to international and national standards, or first principles
Combined (standard) uncertainty	The standard uncertainty of the result of a measurement when that result is obtained from the values of a number of other quantities, equal to the positive square root of a sum of terms, the terms being the variances or covariances of these other quantities weighted according to how the measurement result varies with changes in these quantities (JCGM 2008)
Confidence level	The value of the probability associated with a confidence interval or a statistical coverage interval. Note: The value is often expressed as a percentage (JCGM 2008)
Coverage factor	The numerical factor used as a multiplier of the combined standard uncertainty in order to obtain an expanded uncertainty (JCGM 2008)
Degrees of freedom	A statistical term that refers to the number of terms in a sum minus the number of constraints on the terms of the sum (JCGM 2008)
Expanded uncertainty	The quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand (JCGM 2008)
Evaluation of measurement data—guide to the expression of uncertainty in measurement	The standardized approach to quantifying "uncertainty," that focused on the mathematical treatment of measurement uncertainty through an explicit measurement model under the assumption that the measurand can be characterized by an essentially unique value (JCGM 2012)
First principles	A fundamental law of physics that governs the measurement technology which a calculation is based
Fixture	The calibration system that includes both the hardware and software necessary to perform the calibration under stable, repeatable conditions
Intermediate precision condition of measurement	Condition of measurement, out of a set of conditions that includes the same measurement procedure, same location, and replicate measurements on the same or similar objects over an extended period of time, but may include other conditions involving changes (JCGM 2012)
International System of Units—the SI	The system of units, based on the International System of Quantities, their names and symbols, including a series of prefixes and their names and symbols, together with rules for their use, adopted by the General Conference on Weights and Measures (JCGM 2012)
Material measure	Measuring instrument reproducing or supplying, in a permanent manner during its use, quantities of one or more given kinds, each with an assigned quantity value (JCGM 2012)
Measurand	A particular quantity subject to measurement (JCGM 2008)
Measurement	The set of operations having the object of determining a value of a quantity (JCGM 2008)
Measuring system	Set of one or more measuring instruments and often other devices, including any reagent and supply, assembled and adapted to give information used to generate measured quantity values within specified intervals for quantities of specified kinds (JCGM 2012)
Metrological compatibility (of measurement results)	Property of a set of measurement results for a specified measurand, such that the absolute value of the difference of any pair of measured quantity values from two different measurement results is smaller than some chosen multiple of the standard measurement uncertainty of that difference (JCGM 2012)
Metrological traceability chain	The sequence of measurement standards and calibrations that is used to relate a measurement result to a reference (JCGM 2012)
Population	The totality of items under consideration (JCGM 2012)
Precision	The closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified (steady state) conditions. . . Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance, or coefficient of variation under the specified conditions of measurement. The "specified conditions" can be, for example, repeatability conditions of measurement, intermediate precision conditions of measurement, or reproducibility conditions of measurement. Measurement precision is often used to define measurement repeatability, intermediate measurement precision, and measurement reproducibility. Sometimes "measurement precision" is erroneously used to mean measurement accuracy (JCGM 2012)
Reference material	Material, sufficiently homogenous and stable with reference to specified properties, which has been established to be fit for its intended use in measurement or in examination of nominal properties (JCGM 2012)

Table 1. Continued.

Metrologic term	Definition
Relative (standard measurement) uncertainty	The standard measurement uncertainty divided by the absolute value of the measured quantity value (JCGM 2012)
Repeatability	The closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement (JCGM 2008)
Reproducibility	The closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement (JCGM 2008)
(Measurement) standard	The realization of the definition of a given quantity, with stated quantity value and associated measurement uncertainty, used as a reference. . . A "realization of the definition of a given quantity" can be provided by a measuring system, a material measure, or a reference material. A measurement standard is frequently used as a reference in establishing measured quantity values and associated measurement uncertainties for other quantities of the same kind, thereby establishing metrological traceability through calibration of other measurement standards, measuring instruments, or measuring systems (JCGM 2012)
Standard deviation	The positive square root of the variance (JCGM 2008)
Standard deviation of the mean	The positive square root of the variance of the mean (JCGM 2008)
State-of-health	A qualitative condition of a test object compared to its ideal conditions and generally compared to a threshold to determine the suitability for the measurement system
Trueness	The closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value. . . Measurement trueness is not a quantity and thus cannot be expressed numerically, but measures for closeness of agreement are given in ISO 5725 (JCGM 2012)
True value	The quantity value consistent with the definition of a quantity. . . In the Error Approach to describing measurement, a true quantity value is considered unique and, in practice, unknowable. The Uncertainty Approach is to recognize that, owing to the inherently incomplete amount of detail in the definition of a quantity, there is not a single true quantity value but rather a set of true quantity values consistent with the definition. However, this set of values is, in principle and in practice, unknowable. Other approaches dispense altogether with the concept of true quantity value and rely on the concept of metrological compatibility of measurement results for assessing their validity (JCGM 2012)
Tolerance	The upper and lower limits of the interval of variability (European Accreditation 2013)
True value (of a quantity)	The quantity value consistent with the definition of a quantity. . . In the Error Approach to describing measurement, a true quantity value is considered unique and, in practice, unknowable. The Uncertainty Approach is to recognize that, owing to the inherently incomplete amount of detail in the definition of a quantity, there is not a single true quantity value but rather a set of true quantity values consistent with the definition. However, this set of values is, in principle and in practice, unknowable. Other approaches dispense altogether with the concept of true quantity value and rely on the concept of metrological compatibility of measurement results for assessing their validity (JCGM 2012)
Type A (evaluation of uncertainty)	The method of evaluation of uncertainty by the statistical analysis of a series of observations (JCGM 2008)
Type B (evaluation of uncertainty)	The method of evaluation of uncertainty by means other than the statistical analysis of series of observations (JCGM 2008)
U (calibration trueness)	The uncertainty due to the difference between the standard and sensor
U (calibration reproducibility)	The uncertainty that describes how reproducible a calibration while experiencing differing conditions
U (DAS)	The uncertainty due to the data acquisition system used in the measurement
U (sensor repeatability)	The uncertainty due to the variance of the sensor in repeatable conditions
U (standard)	The uncertainty due to the incomplete knowledge of the true value of the measurement and is related to the standard used for the calibration
(Standard measurement) uncertainty	An estimate associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand (JCGM 2008)
Unit under test	The sensor being calibrated or validated
Variance	The measure of dispersion, which is the sum of the squared deviations of observations from their average divided by one less than the number of observations (JCGM 2008)

Notes: All definitions from the listed sources are provided verbatim with no modification by the authors. NEON, National Ecological Observatory Network.

calibrations and uncertainties that allow detection of sensor drift over time; calibrating to the same standard and protocols, which reduces uncertainty throughout the network; and being able to assess new technological advances against long-term and established sensors (Czaske 2008, Pendrill 2014).

Constructing uncertainty estimates to include all independently quantified component sources that contribute toward the overall uncertainty can be complex and costly. Doing this requires resources that many researchers or manufacturers may have not budgeted. NEON's CVAL has established procedures to estimate uncertainty across a wide range of sensors and observations. Here, we outline the process to develop an uncertainty budget for a simple sensor, and the approach can be applied to more complex measurements. We also outline the metrological method for calibration processes, the propagation of calibration-related sources of uncertainty, and the determination of the appropriate confidence intervals. Finally, the method by which these parameters are used to control the quality of the network of sensors and to report uncertainty is explained.

UNCERTAINTY OVERVIEW

Background

The GUM provides standardized rules to evaluate and express GUMs (JCGM 2008). Estimating the uncertainty of a measurement requires a traceable measurement system (ISO 2003, Bennet and Zion 2005, ISO/IEC 2005). Here, we outline our interpretation of the GUM approach to estimating uncertainty. Applications that involve correlated uncertainties and non-normal distributions are not covered here; readers are referred to the GUM for approaches appropriate to these cases (JCGM 2008).

The uncertainty of a measurement or observation characterizes the dispersion of the values that could reasonably be attributed to the measurand (JCGM 2008). Any measurement is an imperfect estimate of the value of the measurand because uncertainties arise from both random and systematic effects. Thus, a measurement is not a complete representation of the measurand without providing an uncertainty estimate.

Uncertainty component evaluation

There are two methods for quantifying uncertainty (JCGM 2008). Type A uncertainty evaluation uses an experimental approach by calculating the variance of independent observations made under identical measurement conditions (repeatability) or under varying measurement conditions (reproducibility). Type B uncertainty estimates are obtained from other sources, such as calibration certificates, manufacturer specifications, handbooks, or other previously reported estimates (JCGM 2008).

Type A evaluation is often based on repeated measurements in which case the average is taken by:

$$\bar{x}_i = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

where \bar{x}_i is the average value of the measured and calculated from n number of independent repeatable observations (x_i). For measurements reported as an average, the uncertainty, $u(x_i)$, in this estimator is best described by the unbiased SD of the mean, $s(\bar{x}_i)$, and representing a measure of the variability in the population from which the x_i are drawn, that is, $s(x_i)$, such that,

$$u(x_i) = s(\bar{x}_i) = \frac{s(x_i)}{\sqrt{n}}, \quad (2a)$$

and

$$u(x_i) = s(x_i) = \left[\frac{\sum_{i=1}^n (x_i - \bar{x}_i)^2}{v} \right]^{1/2} \quad (2b)$$

where v is the degrees of freedom equal to $n - 1$, in Eq. 2b assuming sample independence and that the sum of the deviations from the average is zero (Devore and Farnum 1999). The sources of $u(x_i)$ can be many; for example, reproducibility, repeatability, tolerance, and others are discussed in more detail below.

Combined uncertainty

The uncertainty of a measurement has more than one source. The individual components of uncertainty, each characterized by a Type A or Type B evaluation, can be combined to represent the combined uncertainty of a measurement, u_c

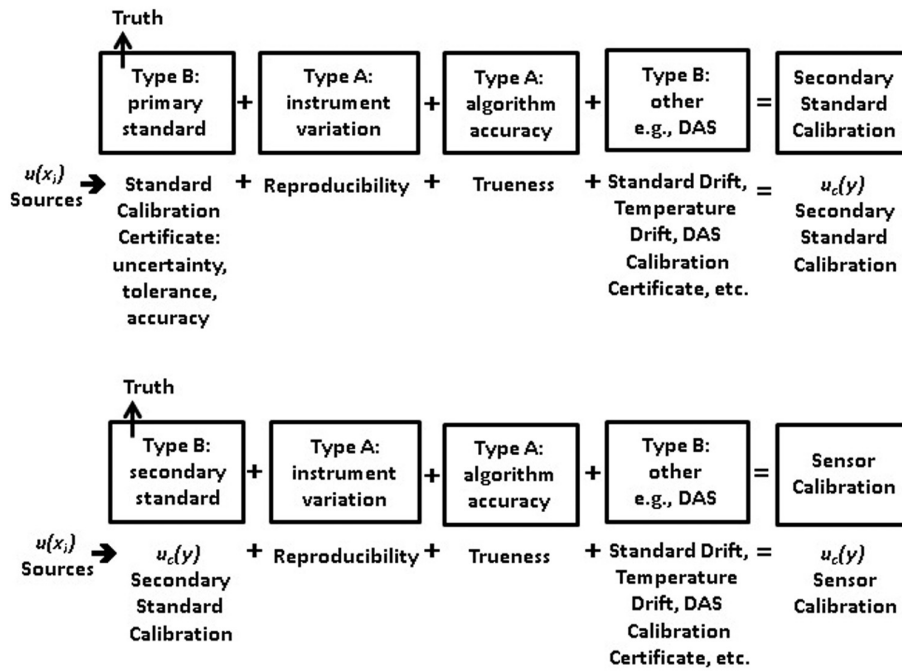


Fig. 1. Uncertainty assessment components for a calibration system that includes a transfer of standard from a primary to a secondary standard (top) and a sensor calibration from the secondary standard (bottom). The boxes provide basic concepts along with the typical type of assessment done for the component (Type A, statistical method and Type B, published materials). Below the boxes are the uncertainty component sources, $U(x_i)$.

(Fig. 1). The rule for propagation of uncertainty describes combined uncertainty from multiple sources,

$$u_c(y) = \sqrt{\underbrace{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i)}_I + 2 \sum_{i=1}^n \sum_{j=i+1}^n \left(\frac{\partial f}{\partial x_i}\right) \left(\frac{\partial f}{\partial x_j}\right) u(x_i, x_j)}_II} \quad (3a)$$

where the approximation for the measurement, y , is the resultant based on x_1, x_2, \dots, x_N input variables and f is an algorithmic function based on the input quantities (Taylor 1997). For example, for the calibration of a platinum resistance thermometer (PRT), f could be a second-order polynomial model, where temperature (y_i) is fit to a number of resistance measurements (x_i).

Term I in Eq. 3a applies to all uncertainty components (i), while Term II in Eq. 3a is important when input quantities x_i and x_j are correlated. It should also be noted that when the nonlinearity of f is significant, the higher-order terms of the Taylor series expansion of f should be included

in Eq. 3a. If we assume the input quantities are independent and only random effects contribute to the variance, Term II vanishes and Eq. 3a reduces to Term I, where uncertainty components are summed in quadrature:

$$u_c(y) = \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i)} \quad (3b)$$

Eqs. 1, 2a, and 2b provide the conceptual basis by which we view a mean quantity, analogous to accuracy, and a variance structure, analogous to precision and the collective propagation of all the sources of uncertainty (Eqs. 3a and 3b). For the remaining methods outlined in this paper, the different uncertainty components will be assumed independent and uncorrelated, such that the contributions in the u_c estimates will be summed in quadrature (Eq. 3b).

Often the uncertainty correlates with the magnitude of the measurement. For instance, a scale's uncertainty will increase as heavier weights are measured. In this case, the estimate of a

combined relative uncertainty ($u_{c,r}$) is used, where u_c can be expressed as a proportion of the estimate of the measurand, and defined as:

$$u_{c,r}(y) = \frac{u_c(y)}{|y|} \quad (4)$$

where $u_{c,r}(y)$ is the combined relative uncertainty of y , $|y|$ is the absolute value of y , and $y \neq 0$. For small values of y , it may be more appropriate to report uncertainty in the units of the measurement. For example, meaningful long-wave radiation signals can be near zero, and thus, measurement uncertainty is reported in W/m^2 . Short-wave measurements, on the other hand, are significant at higher signals and are expressed as a proportion or percentage (proportion multiplied by 100%) due to the uncertainty scaled with the amount of radiation measured.

Reporting uncertainty

For a normally distributed population of values Y , of which y is an estimate, $Y = y \pm u_c(y)$, or if using relative uncertainty, $Y = y \pm u_{c,r}(y)\%$, where $u_{c,r}(y)\%$ is $u_{c,r}(y)$ multiplied by 100%. Provided normal distribution, $u_c(y)$ or $u_{c,r}(y)$ is expected to encompass approximately 68% of the distribution of values in that population. However, having a measure of uncertainty that accounts for only 68% of the population of measurements may lack the stringency required to test an ecological hypothesis, where a confidence level of 90% or 95% is commonly required. The JCGM (2008) and the metrology community address this issue by expanding the uncertainty, that is, multiplying combined uncertainty with a coverage factor, k_p .

$$U_p = k_p u_c(y) \quad (5a)$$

$$U_{p,r} = k_p u_{c,r}(y) \quad (5b)$$

where the expanded uncertainty, U_p (Eq. 5a), or relative expanded uncertainty, $U_{p,r}$ (Eq. 5b), is used to represent a larger confidence interval at level p (metrological convention is to note p as subscript, where $p = 1 - \alpha$, where α is the statistical level of significance), and $k_p = t_p(v)$ where t_p is the t -statistic. The t -statistic can be calculated by defining the degrees of freedom and desired level of confidence automatically using software functions (e.g., TINV[probability,deg_freedom] in Excel) or from a simple lookup of t test statistic (e.g.,

Table G.2 in JCGM 2008). If using the table and the degrees of freedom is not represented (i.e., not a whole number), options are to round down or interpolate. As a rough approximation, $p = 0.95$ corresponds to $k_{95} = 2$, assuming a degrees of freedom around 60. However, a better approximation of k_p can be made by estimating from the effective degrees of freedom v_{eff} . The coverage factor then becomes $k_p = t_p(v_{\text{eff}})$. The effective degrees of freedom is calculated by using the degrees of freedom for each of the components of uncertainty for the computation of the Welch-Satterthwaite formula (Taylor and Kuyatt 1994):

$$v_{\text{eff}} = \frac{u_c^4(y)}{\sum_{i=1}^n (u_i^4(y)/v_i)} \quad (6)$$

Eq. 6 requires degrees of freedom to be defined for all components of uncertainty which can include a Type B evaluation, which may be difficult to assign degrees of freedom. If the source provides the degrees of freedom, this should obviously be used. If not provided, scientific judgment is necessary to approximate the degrees of freedom. For example, for sources that warrant a high level of confidence, such as National Institute of Standards and Technology (NIST), the degrees of freedom is conventionally set to 100, which may be a conservative estimate (Taylor and Kuyatt 1994). However, it may be reasonable to use a lower value in the cases where it is common to have a small sample size. An example of this is the level of confidence on DNA barcoding known for only a few taxonomically identified rare species.

SENSOR CALIBRATION UNCERTAINTY, MEASUREMENT UNCERTAINTY, AND QUALITY CONTROL

Sensor calibration

Periodic calibrations are necessary to maintain confidence in sensor measurements. Calibrations naturally change when sensors are subjected to environmental conditions and when materials degrade. For many environmental sensor networks, sensor calibrations are made on an annual basis (e.g., NEON; ICOS). Calibrations are made on a calibration fixture, each specific to the type of sensor, to assess the sensor functional performance (f in Eqs. 3a and 3b) against a traceable series of standards, transfer standards, etc., under

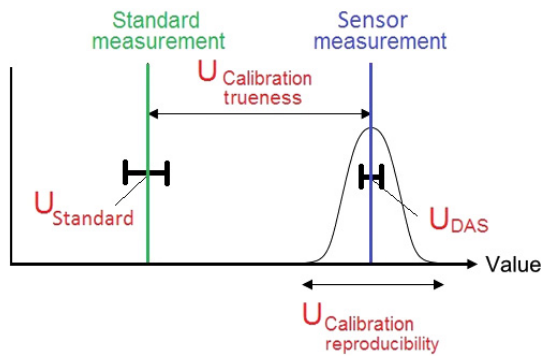


Fig. 2. Conceptual diagram of uncertainty components that have to be quantified for a given calibration where the sensor measurement is the measurement from the sensor being calibrated; standard measurement is the measurement from standard to which the sensor is being calibrated; $U(\text{calibration trueness})$ is the difference between the standard and sensor; $U(\text{calibration reproducibility})$ quantifies how reproducible a calibration is under differing conditions; $U(\text{DAS})$ is the uncertainty due to the data acquisition system used in the measurement; $U(\text{standard})$ is the uncertainty due to the incomplete knowledge of the true value of the measurement.

stable, repeatable conditions (Figs. 1, 2). Hence, each calibration will have its own set of sources of uncertainty (e.g., ambient conditions, transfer functions, multiple operators) which need to be included in the uncertainty assessment.

The objective of the calibration is to make the sensor's measurement as unified with the reference standard as possible by exposing the sensor to a standard under a controlled and stable, repeatable environment and calibrating its response with an algorithmic fit between the reference and sensor. However, even standards have an uncertainty that contributes to the overall combined uncertainty. Additional examples of associated calibration uncertainties are the algorithm fit, repeatability, data acquisition system (DAS), and reproducibility. These calibration uncertainty components are addressed below with a typical approach to quantifying them for a sensor. This approach can also be applied to method development of a non-sensor observation-based measurement.

Measurement standard.—Calibrating a sensor (sometimes referred to in metrology as a unit under test) in theory brings the measurement as

close to the true value as possible. However, the true value of a measurand can never be completely known. Therefore, a reference standard sensor is used to represent the true value, and the uncertainty associated with the reference standard measurand is a distribution (Eqs. 3a and 3b) within which the true value is estimated to occur. The reference standard can be a primary standard, such as those based on first principles. Many times, the standard is a secondary or higher standard in which case the calibration of this standard can be traced to the primary standard (Fig. 1). For example, a secondary standard of a PRT calibrated to the first principles (e.g., temperature defined by the melting point of gallium) is then used to calibrate masses of temperature sensors. Traceability should be based on nationally or internationally recognized standards from organizations such as World Meteorological Organization (WMO), NIST, International Atomic Energy Agency, and ISO. Metrology texts including the International Vocabulary on Metrology (JCGM 2012) and the GUM (JCGM 2008) utilize “measurement standard” to describe the true value and uncertainty associated with the reference measurement.

When assessing uncertainty for the standard, a Type B method is often used. The manufacturer or the calibration facility may provide uncertainty, which may need to be converted to a non-expanded uncertainty. However, sometimes they provide a calibration sheet reporting tolerance (t) or accuracy (a). These have to be transformed into uncertainty. Tolerance is modeled by a uniform (rectangular) distribution, in which case

$$u(x_i) = \frac{t}{\sqrt{3}} \quad (7a)$$

where t is the tolerance as quoted by the manufacturer or calibration facility.

The transformation of accuracy is less straightforward because it has multiple definitions; it may be necessary to contact the manufacturer to determine the meaning. One example of this transformation is:

$$u(x_i) = \frac{a}{3} \quad (7b)$$

where a represents accuracy as reported by the manufacturer or calibration facility and said to include nearly all possible values, that is, $p = 99.73\%$ and $k_p = 3$.

If little information is available,

$$u(x_i) = a \quad (7c)$$

where a represents accuracy reported as the SD, in which case Eq. 7c applies and $k_p = 1$ (European Accreditation 2013).

Eqs. 7b and 7c assume normal distribution of the measurand.

Calibration trueness.—Calibration trueness is determined under stable conditions by the sensor measurement (x_i) and the standard measurement (S_i) in a Type A evaluation, such that:

$$u(x_i) = \left[\frac{\sum_{i=1}^n (x_i - S_i)^2}{v} \right]^{1/2} \quad (8)$$

By normalizing the sensor data to the standard, that is, detrending, Eq. 8 inherently includes both the repeatability of measurements of the sensor and the goodness of fit of the modeled algorithm. Fig. 3 provides an example of a temperature sensor (PRT) and a standard held at stable, repeatable conditions. While stable, repeatable conditions here mean the temperature bath set point does not change from 50°C in this example, Fig. 3 does show the temperature variation (maximum of 50.195°C to minimum of 50.172°C) due to a limitation in the controls of the bath. The variation in the bath temperature does not impact the accuracy of the calibration, but rather, the difference between the standard and sensor is the impact on uncertainty. “Accuracy” is sometimes used to refer to this uncertainty term, and while historically this was acceptable, current standardized metrology vocabulary makes these terms improper to quantify, and thus, we are choosing “calibration trueness” to represent this term (ISO 1994, JCGM 2012).

It is important to check for normally distributed results for the difference between the sensor and standard. If non-normality exists, it could mean that the calibration is improperly modeled; that is, systematic effects exist in the calibration measurement system. If non-normality is observed, it must be evaluated and accounted for separately; this systematic error cannot be combined in quadrature (Eq. 3b). There are methods to incorporate systematic non-normality to uncertainty assessment, which are beyond the scope of this paper and the basic framework of the GUM (JCGM 2008, 2011).

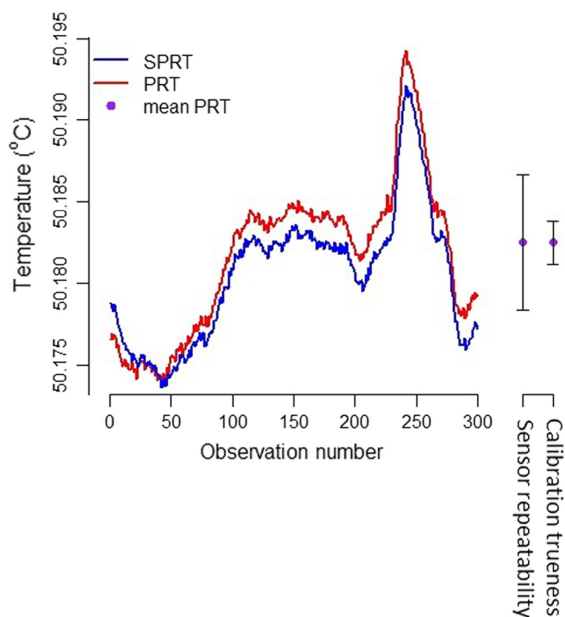


Fig. 3. Converted individual temperature readings of a single platinum resistance thermometer (PRT, the sensor being calibrated) and the standard platinum resistance thermometer (SPRT, the standard used to calibrate the PRT sensor) at a set point of 50°C. The average PRT result of the 300 readings with the designated uncertainty error bars is shown on the right side of the figure and is shown twice to illustrate the difference between the error bars representing sensor repeatability (Eq. 2b), that is, the standard deviation of the 300 PRT readings, and calibration trueness (Eq. 8), that is, the average difference in the SPRT and PRT observations among the 300 data points.

Calibration reproducibility.—Variations that occur during typical operating conditions of a calibration should be evaluated using the Type A approach and reported by SD (Eq. 2b). Here, the conditions to test depend on the sensor and the calibration protocol including the equipment hardware and software (i.e., calibration fixture). For example, the physical adjustment of a tipping bucket could vary with the operator, whereas for a PRT, the operator’s only influence is the placement of the sensor in a controlled environment. The analysis should include evaluating a sample of sensors that represents the population of sensors being evaluated. When multiple tests are involved, for example based on multiple operators, ambient conditions, and

seasonality, our policy is to report the largest SD as the reproducibility of the calibration.

Data acquisition system.—A DAS, such as a data logger, receives a signal from a sensor and passes it along or logs it. Because the data logger is typically calibrated by an external facility, uncertainty related to the DAS is typically estimated by a Type B evaluation. Radiofrequency interference, induced voltage spikes, and other electronic influences can affect a signal measured by a DAS and should be assessed for influence. For example, if analog signals from multiple sensors are multiplexed (sensed sequentially but through one input channel), a stable reference signal (i.e., precision resistor) can be switched (plexed) through the sensor input to assess the DAS including the effects of the multiplexor. Using this method to assess both calibration trueness and reproducibility, the variance terms for the signal influence should then be combined in quadrature (Type A) along with the DAS manufacturer's provided calibration (Type B). In some cases, the DAS uncertainty needs to be added twice in quadrature (two times the variance as a component in Eq. 3b) if both a standard and the sensor have an analog signal.

Combined calibration uncertainty.—Sources comprised in the combined uncertainty of a calibration include calibration trueness, calibration reproducibility, uncertainty of the reference standard, and the DAS. Similarly, evaluation of uncertainty for an observation-based method would need to include some standard or controlled representation of the measurement with estimates for trueness and reproducibility of the method. All of these terms are added in quadrature to represent the combined calibration or method uncertainty, cf. Eq. 3b. Expanded uncertainty is found by multiplying the calibration combined uncertainty by the coverage factor (Eqs. 5a or 5b) to provide an uncertainty estimate at the 95% confidence level. As an example, the combined and expanded uncertainty estimates of the standard reference temperature sensor (standard platinum resistance thermometer [SPRT]) and a temperature sensor (PRT) calibration are presented in Tables 2 and 3, respectively, and mirror the details found in Fig. 1 for a transfer of standard (SPRT) to a sensor calibration.

Calibration quality control.—The QC parameters and thresholds used for automated calibration

include sensor-specific state-of-health data, calibration trueness, sensor repeatability, drift, plausibility, and calibration reproducibility (Taylor and Loescher 2013). A gold standard or a sensor that remains in line with the calibration is used to monitor the calibration reproducibility through a redundancy test (Taylor and Loescher 2013). Calibration and sensor repeatability parameters can be determined and controlled for all calibration set points and for every sensor for complete QC of the population of sensors.

Measurement uncertainty

After a method has been developed or a sensor has been calibrated and used to provide in situ measurements, additional uncertainty components are necessary to estimate the measurement uncertainty; that is, measurement uncertainty differs from method or calibration uncertainty. If a DAS is used to log analog signals from the sensor, the uncertainty from the system is an additional component similar to that described in the section on DAS under *Sensor calibration*. Further, repeatability of the measurement is an additional uncertainty component. Drift associated with the sensor degradation may also need to be accounted for if no correction is in place. Methods for estimating measurement uncertainty are briefly discussed below, but a follow-up manuscript to this one will provide additional insight and the employed methods of Monte Carlo.

Measurement repeatability.—Measurement repeatability is assessed under stable, repeatable conditions and is represented by the SD (Eq. 2b). This measurement repeatability can be used as a proxy for automated quality thresholds because it represents the spread of possible measurements that could occur under identical conditions. If a measurement average is reported as the measurand, however, Eq. 2a should be used as the repeatability for the measurement because it provides a better estimate for the distribution of measurements.

Measurement drift.—Drift in the sensor's response or calibration can result from the deterioration of the sensor over time. Additionally, sensor readings that occur in the field can degrade due to issues such as sensor fouling. Here, we address drift in sensor response or calibration, which can be estimated for an individual sensor and used for QC based on the calibration change (Fig. 4). Based on environmental conditions and the relevant

Table 2. An example of calibration uncertainty components with combined and expanded uncertainties calculated based on effective degrees of freedom based on a temperature sensor (platinum resistance thermometer [PRT]).

X_i	Description	Model/methodology	df	Value (°C)	Type
Standard	First principles of the temperature of the physical point of known molecules	Standard holding body is International Temperature Scale of 1990 (ITS-90)	na	na	na
* U (primary standard)	TPHg, TPW, MPGa	Calibration ITS-90 certification and uncertainty certificates (Fluke 2011a, b, c), respectively	100	0.00010	B
* U (DAS)	DAS—digital multimeter	Calibration certificate traceable to the SI from National Instruments, Austin, Texas, USA	100	0.000010	B
* U (DAS)	DAS—multiplexer	Repeatability and reproducibility determined with 115-ohm precision resistor (model 115R4920-5102K, Vishay, Shelton, Connecticut, USA) combined in quadrature	32	0.0000000041	A
* U (SPRT reproducibility)	SPRT calibration reproducibility	Calibration reproduced eight times for a given SPRT and standard deviation of reproduced calibration determined over calibration range	8	0.00022	A
* U (SPRT trueness)	SPRT calibration trueness	Algorithm accuracy and repeatability of the SPRT estimated under steady-state conditions of 100 readings	100	0.00020	A
U (SPRT combined)	Combined uncertainty for the SPRT calibration	All of (*) terms are added in quadrature	29	0.00031	Eff
U (SPRT expanded)	Expanded uncertainty for SPRT calibration	An expansion factor of 2.05 was determined by a t -distribution, resulting in $Y = y \pm 0.00064^\circ\text{C}$ with 95% level of confidence	na	0.00064	k

Notes: Components mirror Fig. 1 for first principles to secondary standard calibration. The calibration method overview is the following: standard platinum resistance thermometer (SPRTs, 5626, Fluke Corp., Everett, Washington, USA) are calibrated against first principles: triple point of mercury, TPHg; triple point of water, TPW; and melting point of gallium, MPGa (Models 5900E, 5901D, and 5943E, respectively, Fluke Corp.). This follows the National Institute of Standards and Technology methodological procedures outlined in Strouse (2008).

science requirements, a QC plan generally calls for a re-calibration after 1-year field deployment. If there are at least two calibrations ($i \geq 2$), drift can be estimated by:

$$u(x_i) = \left[\frac{\sum_{i=1}^n (C_i - C_{i-1})^2}{v} \right]^{1/2} \quad (9a)$$

where C_i is the calibrated measurement at each instance i . This equation is modeled after SD (Eq. 2b). Since the reproducibility of the calibration contributes to the differences in the calibrated measurements, one way to quantify the uncertainty due to sensor drift is to subtract the uncertainty due to the reproducibility of the calibration ($u(\text{calibration_reproducibility})$) calculated from Eq. 3b:

$$u(x_i) = \sqrt{\left[\frac{\sum_{i=1}^n (C_i - C_{i-1})^2}{v} \right] - (u(\text{calibration_reproducibility}))^2} \quad (9b)$$

If the calibration interval changes (e.g., 1 yr is extended to 2 yr), it is important to adjust the drift estimate and state the assumptions for users of the uncertainty estimate. For example, the variance term for drift might be doubled in the second year (Eq. 3b) for the estimate of measurement uncertainty, assuming that field deterioration is the major cause of the observed drift, neglecting the effect of shipping the sensor. To the degree that shipping causes drift, uncertainty due to drift is overestimated in this example.

Other networks such as NASA's aerosol monitoring network utilize the approach of having a pre- and post-field calibration application (Giles and Holben 2014) to correct data products for drift by an applied linear degradation from pre- to post-field results. Of course, this approach has

Table 3. An example of calibration uncertainty components with combined and expanded uncertainties calculated based on effective degrees of freedom based on a temperature sensor (platinum resistance thermometer [PRT]).

X_i	Description	Model/methodology	df	Value (°C)	Type
Standard	SPRT	Calibrated to first principles (TPHg, TPW, MPGa)	na	na	na
* U (secondary standard)	Combined uncertainty for the SPRT calibration	Found by combined uncertainty components from the SPRT calibration uncertainty analysis	29	0.00031	B
* U (SPRT drift)	SPRT drift between calibration cycles	Change in calibration over four calibration cycles with reproducibility removed from difference	4	0.0010	A
* U (DAS)	DAS—digital multimeter	Calibration certificate traceable to the SI from National Instruments, Austin, Texas, USA (accounted for twice for SPRT and PRT)	100	0.000014	B
* U (DAS)	DAS—multiplexer	Repeatability and reproducibility determined with 115-ohm precision resistor (model 115R4920-5102K, Vishay, Shelton, Connecticut, USA) combined in quadrature (accounted for twice for SPRT and PRT)	32	0.000091	A
* U (PRT calibration reproducibility)	PRT calibration reproducibility	Calibration reproduced 11 times for a given PRT and standard deviation of reproduced calibration determined over calibration range	11	0.0034	A
* U (PRT calibration trueness)	PRT calibration trueness	Algorithm accuracy and repeatability of the PRT estimated under steady-state conditions from 60 readings	60	0.0011	A
U (PRT combined)	Combined uncertainty for the PRT calibration	All of (*) terms are added in quadrature	14	0.0037	Eff
U (PRT expanded)	Expanded uncertainty for PRT calibration	An expansion factor of 2.14 was determined by a t -distribution, resulting in $Y = y \pm 0.0080^\circ\text{C}$ with 95% level of confidence	na	0.0080	k

Notes: Components mirror Fig. 1 for secondary standard to sensor calibration. The calibration overview is the following: The standard platinum resistance thermometer (SPRT) is now the transfer of standard for the calibration of the PRTs (R032-00000038, Thermometrics Corp., Northridge, California, USA), which involves placing sensors, PRTs, and transfer SPRT, in three individual high-precision calibration baths (7341, Fluke Corp.) at temperatures that span the range of natural environmental temperatures. This follows the National Institute of Standards and Technology methodological procedures outlined in Strouse (2008).

shortcomings if the degradation is nonlinear. Accounting for the uncertainty in the correction is further necessary if this is the case.

Combined and expanded measurement uncertainty.— At a minimum, measurement uncertainty will include two components: repeatability and calibration or method uncertainty. A combined measurement uncertainty can be found by adding in quadrature (Eq. 3b) the measurement repeatability and the combined calibration uncertainty (not the expanded calibration uncertainty). If other components are contributing, such as drift and DAS uncertainty, these would also be included for the combined measurement uncertainty. Finally, the measurement can be reported properly when the results are accompanied by the

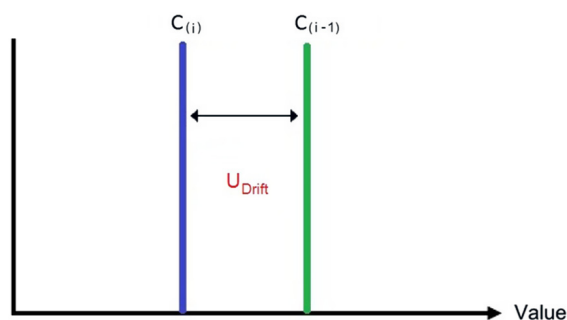


Fig. 4. Illustration of uncertainty due to drift in calibration where $C_{(i)}$ and $C_{(i-1)}$ denote the calibrated measurement with the current and previous calibrations, respectively, and the change in measurement is due to drift in the sensor calibration.

Table 4. An example of measurement uncertainty components with combined and expanded uncertainties calculated based on effective degrees of freedom based on a temperature sensor (platinum resistance thermometer [PRT]).

X_i	Description	Model/methodology	df	Value (°C)	Type
* U (PRT calibration)	Combined uncertainty for the PRT calibration	Found by combined uncertainty components from the PRT calibration uncertainty analysis	14	0.0037	B
* U (DAS)	DAS—digital multimeter	Calibration certificate traceable to the SI	100	0.00040	B
* U (PRT drift)	PRT drift between calibration cycles	Change in calibration estimated based on six calibration cycles with reproducibility removed from difference	6	0.00096	A
* U (PRT repeatability)	PRT repeatability	Standard deviation of the mean of the 60 readings used to provide average where standard deviation 0.0015°C	60	0.00020	A
U (combined measurement)	Combined measurement uncertainty	All of (*) terms are added in quadrature	17	0.0040	Eff
U (expanded measurement)	Expanded measurement uncertainty	An expansion factor of 2.11 was determined by a t -distribution, resulting in $Y = y \pm 0.0084^\circ\text{C}$ with 95% level of confidence	na	0.0084	k

Note: The measurement overview is the following: A PRT in the field takes readings every second and minute averages are reported for the measurand.

uncertainty with the associated level of confidence. If expanding the uncertainty to higher levels of confidence than one SD, then the combined measurement uncertainty is multiplied by the coverage factor (Eqs. 5a or 5b). Table 4 details the combined and expanded uncertainties for a PRT measurement.

Measurement quality control.—The QC approaches for networks of automated sensors have been well documented in the literature (e.g., Campbell et al. 2013, Taylor and Loescher 2013), but do not take into account the plethora of new sensors and emergent environmental research infrastructures that require more optimized approaches toward network-level QC. Automated quality flags and quality metrics have also been applied to sensor networks (Smith et al. 2014). Measurement repeatability determined under controlled stable, repeatable conditions can inform some automated controls, but preliminary field data are often required to estimate real-time, dynamic QC thresholds on field measurements.

CONCLUSION

In summary, we have introduced standardized methods to estimate uncertainty and illustrated its application to the calibration of a simple sensor. We also introduced the overall approach to calculate uncertainty, and noted that the same philosophy can be applied to many other ecological

quantities. The ISO provides additional metrology guides for standardized measurements (www.iso.org), for example, the collection of aquatic organisms; air, water, and soil quality; soil microorganisms; aquatic and terrestrial mammal trapping; CO₂; and environmental sensing managed by the WMO.

We recognize that metrological approaches are foreign to many ecologists. However, ecologists are being asked more and more to provide solutions to many of today's environmental problems, for example, ecosystem effects on sea level rise, controls of climate-induced movement of species, role of chronic disturbance of increasing air temperature, nitrogen deposition, and atmospheric CO₂ on ecosystem services. The use of Bayesian modeling approaches (data assimilation) is at the forefront of ecological forecasting, and it requires a priori estimates of uncertainty for all data-assimilated parameters. Robust cross-study comparisons, moving from correlative statistics to predicative and prognostic approaches, and defending science-based policy recommendations also require defensible uncertainty budgets. Determining small but important differences in ecological data (often stochastic) over large spatial areas and across decades require known signal-to-uncertainty ratio of the measurement systems. Ecological sciences have entered the world of "Big Science" (e.g., PCAST 2011, Holdren 2014), and thus, there is an increasing responsibility to use,

analyze, and report data with high fidelity in describing uncertainty.

There is still opportunity for progress in the field of uncertainty analysis, even though the GUM was produced in 1995 (Aleksandrov 2001). While organizations such as NIST, National Renewable Energy Laboratory, and others are using uncertainty as the JCGM intended (Taylor and Kuyatt 1994, Reda 2011), not all disciplines have followed suit, and broader adoption is ongoing (Aleksandrov and Belyakov 2002, Linko 2004). Uncertainty assessment is a healthy and necessary exercise in all sciences; without uncertainty, a measurement is meaningless. Fear of large uncertainties should not be a deterrent; understanding where knowledge is lacking aids in developing better measurements and observations.

ACKNOWLEDGMENTS

This paper is a product of QUEST (Quantifying Uncertainty in Ecosystem Studies, www.quantifyinuncertainty.org), a Research Coordination Network funded by the National Science Foundation (NSF). The National Ecological Observatory Network (neon-science.org) is a project sponsored by the NSF (EF-102980) and managed under cooperative agreement to Battelle. The authors would like to thank the NEON CVAL staff for the logistical support in uncertainty testing, Sarah Streett for methodological and statistical assistance, and Ruth Yanai for discussion and copy editing. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of our sponsors.

LITERATURE CITED

- Agilent. 2006. Understanding Reading Accuracy and Resolution in a 6 1/2-Digit DMM. 5989-4879 EN Technical Note. <http://cp.literature.agilent.com/litweb/pdf/5989-4879EN.pdf>
- Alekandrov, Y. I. 2001. Uncertainty of measurement: twenty years afterwards. *Fresenius' Journal of Analytical Chemistry* 370:690–693.
- Aleksandrov, Y. I., and V. I. Belyakov. 2002. Error and uncertainty in the results of chemical analysis. *Journal of Analytical Chemistry* 57:94–103.
- Bennet, K., and H. Zion. 2005. Metrology concepts: understanding test uncertainty ratio (TUR). Transcat White Paper. <http://www.transcat.com/media/pdf/TUR.pdf>
- Campbell, J. L., L. E. Rustad, J. H. Porter, J. R. Taylor, E. W. Dereszynski, J. B. Shanley, and E. R. Boose. 2013. Quantity is nothing without quality: automated QA/QC for streaming environmental sensor data. *BioScience* 63:574–585.
- Czaske, M. 2008. Usage of the uncertainty of measurement by accredited calibration laboratories when stating compliance. *Accreditation and Quality Assurance* 13:645–651.
- Devore, J. L., and N. R. Farnum. 1999. *Applied statistics for engineers and scientists*. Duxbury Press, Pacific Grove, California, USA.
- European Accreditation. 2013. Evaluation of the uncertainty of measurement in calibration. <http://www.european-accreditation.org/publication/ea-4-02-m-rev01-September-2013>
- Fluke, C. 2011a. Certification of the melting point of gallium cell. <http://us.flukecal.com/products/temperature-calibration/its-90-temperature-standards>
- Fluke, C. 2011b. Certification of the triple point of mercury cell. <http://us.flukecal.com/products/temperature-calibration/its-90-temperature-standards>
- Fluke, C. 2011c. Certification of the triple point of water cell. <http://us.flukecal.com/products/temperature-calibration/its-90-temperature-standards>
- Giles, D. M., and B. N. Holben. 2014. DATA—version 2 AOD descriptions DATA download information. http://aeronet.gsfc.nasa.gov/new_web/data_description_AOD_V2.html
- Hickey, J. M., M. G. Keane, D. A. Kenny, A. R. Cromie, H. A. Mulder, and R. F. Veerkamp. 2008. Estimation of accuracy and bias in genetic evaluations with genetic groups using sampling. *Journal of Animal Science* 86:1047–1056.
- Holdren, J. 2014. National plan for Earth observations. https://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/national_plan_for_civil_earth_observations_-_july_2014.pdf
- IPCC [International Panel on Climate Change]. 2015. IPCC AR5: lessons learnt for climate change research and WCRP. https://www.ipcc.ch/pdf/supporting-material/IPCC_WCRP_workshop_report_final.pdf
- ISO [International Organization for Standardization]. 1994. Accuracy (trueness and precision) of measurement methods and results. Part 1. International Organization for Standardization, Genève, Switzerland.
- ISO [International Organization for Standardization]. 1995. Guide to the expression of uncertainty in measurement. International Organization for Standardization, Genève, Switzerland.
- ISO [International Organization for Standardization]. 2003. Measurement management—requirements for measurement processes and measuring equipment.

- International Standard 10012. International Organization for Standardization, Genève, Switzerland.
- ISO/IEC [International Organization for Standardization/International Electrotechnical Commission]. 2005. General requirements for the competence of testing and calibration laboratories. International Standard 17025. International Organization for Standardization, Genève, Switzerland.
- JCGM [Joint Committee for Guides in Metrology]. 2008. Evaluation of measurement data—guide to the expression of uncertainty in measurement (GUM). International Organization for Standardization, Genève, Switzerland.
- JCGM [Joint Committee for Guides in Metrology]. 2011. Evaluation of measurement data—supplement 2 to the “Guide to the expression of uncertainty in measurement”. Extension to any number of output quantities. International Organization for Standardization, Genève, Switzerland.
- JCGM [Joint Committee for Guides in Metrology]. 2012. International vocabulary of metrology—basic and general concepts and associated terms. ISO Copyright Office, Geneva, Switzerland.
- Linko, S. 2004. From a single analyzer to multiple instruments—the GUM approach. *Accreditation and Quality Assurance* 9:137–140.
- Menditto, A., M. Patriarca, and B. Magnusson. 2007. Understanding the meaning of accuracy, trueness and precision. *Accreditation and Quality Assurance* 12:45–47.
- PCAST [President’s Council of Advisors on Science and Technology]. 2011. Sustaining environmental capital: protecting society and the economy. Report to the President. www.whitehouse.gov/ostp/pcast
- Pendrill, L. R. 2014. Using measurement uncertainty in decision-making and conformity assessment. *Metrologia* 51:S206.
- Peters, D. P. C., H. W. Loescher, M. SanClements, and K. M. Havstad. 2014. Taking the pulse of a continent: role of observatories and long-term research networks to fill critical knowledge gaps. *Ecosphere* 5:1–23.
- Prenesti, E., and F. Gosmaro. 2015. Trueness, precision and accuracy: a critical overview of the concepts as well as proposals for revision. *Accreditation and Quality Assurance* 20:33–40.
- Reda, I. 2011. Method to calculate uncertainties in measuring shortwave solar irradiance using thermopile and semiconductor solar radiometers (NREL/TP-3B10-52194). <http://www.nrel.gov/research/publications.html>
- Smith, D. E., S. Metzger, and J. R. Taylor. 2014. A transparent and transferable framework for tracking quality information in large datasets. *PLoS ONE* 9: e112249.
- Stemwedel, J. D. 2014. Brief thoughts on uncertainty. <http://blogs.scientificamerican.com/doing-good-science/brief-thoughts-on-uncertainty/>
- Strouse, G. F. 2008. Standard platinum resistance thermometer calibrations from the Ar TP to the Ag FP. <https://www.nist.gov/sites/default/files/documents/calibrations/sp250-81.pdf>
- Taylor, J. R. 1997. An introduction to error analysis. University Science Books, Sausalito, California, USA.
- Taylor, B. N., and C. E. Kuyatt. 1994. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. <https://www.nist.gov/pml/nist-technical-note-1297>
- Taylor, J. R., and H. L. Loescher. 2013. Automated quality control methods for sensor data: a novel observatory approach. *Biogeosciences* 10:4957–4971.
- Toro, A. D. 2015. Uniformity, repeatability, stability, and accuracy: knowledge base. <http://www.apogeeinstruments.com/uniformity-repeatability-stability-and-accuracy/>