Supplementary Material

Quality Assurance of Oceanographic Observations: Standards and Guidance Adopted by an International Partnership

Mark Bushnell, Christoph Waldmann, Steffen Seitz, Earle Buckley, Mario Tamburri, Juliet Hermes, Emma Henslop, Ana Lara-Lopez

Corresponding Author: mark.bushnell@noaa.gov

1 Quality Assurance Discussions and Examples

As previously noted, quality assurance (QA) practices must cover a broad span of activities. These may include, but are not limited to, a) manufacturer guidance and user manual, b) calibrations, c) acceptance, use, and maintenance of instruments and equipment, d) biofouling and corrosion strategies, e) pre- and post-deployment procedures, and record-keeping strategies. Checklists must be employed to ensure adherence to standardized techniques. These issues are considered, and examples provided, in the following discussion.

The following lists suggest ways to ensure QA by using specific procedures and techniques:

- Perform pre-deployment calibrations on every sensor.
- Perform post-deployment calibrations on every sensor, plus in-situ comparison before recovery.
- Calibrate ready-to-use spares periodically.
- Monitor with redundant sensors whenever possible.
- Collect *in-situ* water samples to compare with the sensor.
- Take photos of sensor fouling for records.
- Record all actions related to sensors—calibration, cleaning, deployment, etc.
- Compare the first day or less of readings from newly deployed sensor to last sensor deployed. Large shifts in median values can indicate a problem with one of the sensors. A post calibration of a previously deployed sensor may help to determine if it is the source of the discontinuity in readings.
- Monitor battery voltage and watch for unexpected fluctuations.

When selecting sensors to ensure they are fit for purpose, consider these factors:

- Selection of a reliable and supportive manufacturer and appropriate model
- Measurable data concentration range (including detection limit)
- Lowest and highest possible readings
- Operating range (i.e., some instruments won't operate at certain temperatures or depth / pressure ranges
- Salinity correction
- Resolution/precision required

- Sampling frequency – how fast the sensor can take measurements
- Reporting frequency – how often the sensor reports the data
- Response time of the sensor – sensor lag – time response
- Power source limitations
- Clock stability and timing issues
- Internal fault detection and error reporting capabilities

When evaluating which specifications must be met:

- State the expected accuracy.
- Determine how the sensor compared to the design specifications.
- Determine if sensor met those specifications.
- Determine whether the result is good enough (fit for purpose: data are adequate for nominal use as preliminary data).

General comments regarding QA procedures:

A diagram (http://www.ldeo.columbia.edu/~dale/dataflow/, Dale Chayes, LDEO) provides a visual representation of proper QA procedures. These include:

- Require serial numbers and model ID from the supplier.
- Develop useful checklists and update them as needed.
- Do not assume the calibration is perfect (could be a calibration problem rather than a sensor problem).
- Keep good records of all related sensor calibrations and checks (e.g., conductivity and temperature).
- Use NIST-traceable standards or equivalent when conducting calibrations or calibration checks.
- Keep good maintenance records. Favor sensors that maintain an internal file of past calibration constants, which is very useful since it can be downloaded instead of transcribed manually, thus introducing human error.
- Plot calibration constants or deviations from a standard over time to determine if the sensor has a drift in one direction or another. A sudden change can indicate a problem with the sensor or the last calibration.
- Do not presume that anomalous values are always problems with a sensor. Compare measurements with other sensors to help determine if the reading is real; then examine the possibility of problems with a sensor.

Follow the manufacturer's recommendations and best practices established by knowledgeable users to ensure proper sampling techniques. For example, in a non-pumped sensor in a turbulent environment, bubbles can adhere to the surface of a sensor resulting in anomalous readings. Cycle the wipers or shutter before the reading to brush off the bubbles from the face of the instrument. For a pumped system in a turbulent environment, a degassing "Y" may limit bubbles adhering to the face of the sensor.

1.1 Manufacturer Guidance and User Manual

Each manufacturer provides an instrument-specific user manual, and many also offer online support. Accessing online support is especially helpful for maintaining the latest updates to software, firmware, and hardware. Manufacturer guidance is essential but not sufficient in all cases. It is also important to communicate with others who have experience using a specific instrument—to share knowledge and learn from others.

1.2 Sensor Calibration

Observations must be traceable to one or more accepted standards (such as the U.S. National Institute of Standards and Technology or NIST) through a calibration performed by the manufacturer and/or the operator. If the calibration is conducted by the manufacturer, the operator must also conduct some form of an acceptable calibration check.

An often-overlooked calibration or calibration check can be performed by choosing a consensus standard. For example, deriving the same answer (within acceptable levels of data precision or data uncertainty) from four different sensors of four different manufacturers, preferably utilizing several different technologies, constitutes an acceptable check. Because of the trend toward corporate acquisitions and mergers, those wishing to employ a consensus standard should ensure that the different manufacturers are truly independent.

The following principles apply to calibrations efforts:

- All equipment that affects the quality of test results is calibrated according to the minimum frequency suggested by the manufacturer, by regulation, by method, or as needed.
- These calibrations are traceable to international standards of measurement where available over the entire range of use.
- All equipment is calibrated or checked before being placed into use.
- Intermediate calibration confirmation checks are performed to maintain confidence in the calibration status.
- Where calibrations give rise to a set of correction factors, these factors are communicated to operators and data users.
- If equipment is sent outside the operator’s direct control for either maintenance or calibration, the function and calibration status of the equipment is checked upon return and shown to be satisfactory before the equipment is returned to service.
- A full calibration history is retained by the operator and made available to data users.

For equipment returned after calibration, the calibration certificates should be reviewed to ensure the calibration status (Table 1).

Table 1. Criteria for ensuring calibration status.

❒	a statement of conformity to relevant specification after calibration/verification (such as the manufactures specifications)
❒	item name, type, or description
❒	identification number
❒	location
❒	calibration interval

❒	calibration procedure used
❒	calibration source (both the standard used and the laboratory providing the service)
❒	date of calibration
❒	corrections, conditions of use (including environmental conditions) necessary to achieve the required performance
❒	specific results of each calibration, prior to adjustment or repair, if the item was found out of tolerance
❒	details of any maintenance such as servicing, adjustments, repairs or modifications carried out
❒	any limitations on use
❒	identification of person(s) performing the calibration
❒	if accuracy of an item is described in a calibration report or certificate, a means of traceability between that item and the report or certificate
❒	accreditation logo

1.3 Acceptance, Use, and Maintenance of Instruments and Equipment

The requirements for the acceptance, use, and maintenance of all field and laboratory instruments and equipment are found in the manufacturers' recommendations, instruction manuals, and field or laboratory Standard Operating Procedures. A list of recommended procedures for proper acceptance, maintenance, and use include:

- Manufacturers' manuals for each instrument and piece of equipment are located within proximity of the instrument.
- When new instruments are installed, the identification and the subsequent documentation for each installed item details normal operating instructions, routine user maintenance, preventive maintenance, and cleaning procedures.
- All equipment should be visually inspected daily for damage or dirt and repaired or cleaned if needed before use. If meters are stored for long periods (more than 1 week) without being used, they are calibrated and inspected at least weekly to keep them in good working order.
- No instrumentation or equipment is used until it is in a safe, reliable operational state.
- Equipment is operated only by authorized and trained personnel.
- Equipment and its associated reference materials are identified and labeled, marked, or qualified by calibration data to indicate their calibration status.
- Each item of equipment is uniquely identified, and records of performance checks, maintenance, and calibration results are maintained.
- All routine maintenance and non-routine repairs are documented and kept with the instrument/equipment. The information recorded includes the analyst's initials, date maintenance was performed, a description of the maintenance activity, and (if the maintenance was performed in response to a specific instrument performance problem) the result of retesting to demonstrate that the instrument was returned to acceptable standards prior to reuse.

- Equipment that has been subject to overloading or mishandling, has given suspect results, or has been shown to be defective or outside specifications needs to be taken out of service, isolated to prevent its use, or clearly labeled as being out of service. Out-of-service equipment must be calibrated or verified prior to use. Equipment is not returned to service until performance checks and verification have been performed and documented.

1.3.1 Biofouling/Corrosion Strategies

In 1969 B.F Brown wrote "…in both wet corrosion and dry oxidation of alloys the processes are so complex that theory does not now have predictive capability for corrosion technology, and the technology must depend heavily – almost exclusively – upon enlightened empiricism." (Myers et al., 1969).

The electro-chemical reaction causing metallic corrosion can result in the detrimental wasting of critical components, or it can be employed to protect them using sacrificial anodes. While the science of corrosion is well understood, in practice it can be challenging. A change in the ambient conditions can result in surprising changes in corrosion. For example, if the surrounding seawater suddenly becomes anoxic, stainless steel elements will quickly corrode.

Operators are strongly encouraged to follow the manufacturers recommendations regarding sacrificial anodes and to document their use so that when unexpected corrosion does occur, they have a defensible argument.

Whenever dissimilar metals are electrically connected and immersed in water, the less noble metal will corrode. Consequently, the differing metals must be electrically isolated. Whenever possible, use of non-metallic elements is desired.

When coated metallic components are employed, any penetration of the coating may result in focused localized corrosion. For example, a fish bite penetrating the nylon jacket of a wire rope can cause a focal point for both corrosion and fatigue.

Biofouling is a frequent cause of sensor failure, so the following strategies may be useful for ameliorating the problem:

- Use anti-fouling paint with the highest copper content available (up to 75%) where permissible (but not on aluminum).
- Tributyltin oxide anti-foulant systems, often used in conjunction with a pumped system, are highly effective (e.g., Sea-Bird SBE 43)
- To help with post-deployment clean-up (but not as an anti-foulant), wrap the body of the sensor with clear packing tape for a small probe or plastic wrap for a large instrument, followed by polyvinyl chloride (PVC) pipe wrap tape. (This keeps the PVC tape from leaving a residue on the sensor.) Wrap the sensor body with copper tape (again, beware of aluminum).
- Coat with zinc oxide (Desitin ointment).
- Use brass door/window screen around opening to sensor. The combination of copper and zinc is a great anti-foulant and is significantly cheaper than copper screen.
- Remember that growth is sensor, depth, location, and season dependent.
- Maintain wipers on sensors per manufacturers' recommendation.

- Flush out with chlorine gas pumped through the system. This technique requires a lot of battery power.
- Plan for routine changing or cleaning of sensor as necessary.
- Check with calibration facility regarding which anti-foulants will be handled (allowed) by the calibrators.
- Use copper plates as shutters, which keep the sensor open for limited time. This is ideal over wipers in oceanic environments with encrusting organisms like barnacles. Wipers may not work well for some applications, where sediment and particles that become embedded in the wipers can scratch the lens on optical sensors.
- Store the sensor in the dark when not in use.
- Avoid or isolate dissimilar metals.
- Maintain sacrificial anodes and ensure they are properly installed (good electrical contact).
- Maximize the use of non-metallic components.
- Use ultraviolet or UV-stabilized components that are not subject to sunlight degradation.
- Mount sensors vertically to minimize sediment buildup; employ filters for sensors with flow-through tubes.
- Where applicable, maintain sensor surfaces by gentle cleaning (e.g., using a baby toothbrush).
- Store the device above the surface between measurements.
- Use a pumped system where the sensor is kept above water and the sample is pumped through a flow chamber just before a reading is required.
- Use petroleum-based lubricants as biocides (using care in the vicinity of optics and other sensitive components).
- Carefully maintain and clean filters.
- Obtain mid-deployment validation field samples.

1.4 Pre- and Post-Deployment Procedures

Good QA needs to follow important procedures and checks that need to be performed before, during, and after sensor deployment (Figure 2). Below we present an example of those procedures and checks extracted from QARTOD manuals.

1.4.1 General pre-deployment QA checklist

- ❒ Read the manual.
- ❒ Establish, use, and submit (with a reference and version #) a documented sensor preparation procedure (protocol). Clean sensor according to the manufacturer’s procedures.
- ❒ Calibrate sensor against an accepted standard and document (with a reference and version #).
- ❒ Compare the sensor with an identical, calibrated sensor measuring the same variable in the same area (in a calibration lab).
- ❒ View calibration specifications with a critical eye (don’t presume the calibration is infallible). Execute detailed review of calibrated data.
- ❒ Check the sensor history for past calibrations, including a plot over time of deviations from the standard for each (this will help identify trends such a progressively poorer performance). Check the sensor history for past repairs, maintenance, and calibration.

- ❒ Consider storing and shipping information before deploying. Sensors may suffer degradation if improperly stored or shipped (e.g.m heat, cold, vibration, etc.)
- ❒ Record operator/user experiences with this sensor.
- ❒ Search the literature for information on your particular sensor(s) to see what experiences other researchers may have had with the sensor(s).
- ❒ Establish and use a formal pre-deployment checklist.
- ❒ Ensure that technicians are well-trained. Use a tracking system to identify those technicians who are highly trained and then pair them with inexperienced technicians for training purposes.

Figure 1. Doug Wilson prepares a NOAA Pacific Marine Environmental Laboratory ocean acidification buoy for deployment in the Chesapeake Bay. Checklists are an important component of the QA process and help to ensure necessary steps are conducted in the proper sequence.

1.4.2 General deployment checklist

- ❒ Scrape biofouling off platform.
- ❒ Verify sensor serial numbers.
- ❒ Perform visual inspection; take photos if possible (verify position of sensors, connectors, fouling, and cable problems).
- ❒ Verify instrument function at deployment site just prior to site departure. Monitor sensors for issues (e.g., freezing, fouling).
- ❒ Use established processes to confirm that the sensor is properly functioning before departing the deployment site.
- ❒ Specify date/time for all recorded events. Use Coordinated Universal Time, also known as Greenwich Mean Time or GMT.
- ❒ Check software to ensure that the sensor configuration and calibration coefficients are correct. Also check sampling rates and other timed events like wiping and time averaging.
- ❒ Visually inspect data stream to ensure reasonable values.
- ❒ Compare up and down casts and/or dual sensors (if available).
- ❒ Note weather conditions and members of field crew.

1.4.3 General post-deployment checklist

- ❒ Take pictures of recovered sensor prior to cleaning.
- ❒ Check to make sure all clocks agree or, if they do not agree, record all times and compare with an accepted time standard.
- ❒ Post-calibrate sensor before and after cleaning, if possible. Perform *in-situ* side-by-side check using another sensor, if possible.
- ❒ Use standard procedures to provide feedback about possible data problems and/or sensor diagnostics.
- ❒ Clean and store the sensor properly or redeploy.
- ❒ Visually inspect physical state of instrument.
- ❒ Verify sensor performance.
    - o Check nearby stations.
    - o Make historical data comparisons (e.g., long-term time-series plots, which are particularly useful for identifying long-term bio-fouling or calibration drift).

1.5 Record-keeping Strategies

Technical records are original observations, calculations, derived data, ancillary information, calibration records, and copies of test reports. Technical records provide sufficient information for traceability for all actions from sample collection to reporting the data. In addition, the records should contain sufficient information to facilitate identification of factors affecting the uncertainty and allow historical reconstruction of all activities that produced the data. ISO (2017) provides standards for technical record keeping.

Personnel must maintain complete, accurate, and legible records as they perform field operations, sampling, and analysis. Observations, data and calculations are recorded at the time they are made and are identifiable to the specific task. Some of the record-keeping strategies identified below may not apply in all instances. Some are dated, new techniques are rapidly being developed, and new standards are emerging. The concepts identified below should serve as guidance. At a minimum:

- Raw (original) data collected in the field or laboratory should be recorded such that samples collected and data generated are complete and traceable throughout their history. (Note: Raw data are defined as any original factual information from a measurement activity or study recorded in a field/laboratory notebook, worksheets, records, memoranda, notes, or exact copies thereof that are necessary for the reconstruction and evaluation of the report of the activity or study. Raw data may include photography, computer printouts, magnetic media [including dictated observations], and recorded data from automated instruments).
- Data should be recorded in standardized formats, e.g., data collection forms, bound and paginated laboratory and field logbooks, laboratory record books, spreadsheets, computer records, and output from instruments (both electronic and hardcopy). All test records shall carry minimum identification pertaining to title, responsible person or author, and date.
- All manual entries shall be entered using ink and initialed and dated by the individual recording the entry.
- Changes to original (raw) data should not obliterate the original entry and should be corrected using a single line and annotated with the new data, and the date/initials of the person who modified the record. A short explanation will be added to non-obvious corrections. For electronic data, all changes are tracked by the audit trail requiring date/time stamping, identification of the person making the change, and providing a reason for the change.
- Electronic data collected by field or laboratory instruments should be backed up daily or transcribed daily onto a hard copy data form and verified 100% by another person.
- Instrument logs should be used to document use and maintenance. Calibration records should be maintained as part of the test project file.
- Laboratory and field records must be completed and reviewed in real time.

Best practices for records management address eight specific requirements: record generation, identification, collection, indexing, filing, access and retrieval, storage, retention, and disposition of quality and technical records, including electronic records.

1.5.1 Records generation

- All work performed is recorded legibly and in such a manner that another individual, competent in the same field, may repeat the work described solely from the description written without additional explanation.
- Entries must permanently identify the person performing the work, and all records are signed and dated.
- Entries are made in ink. No erasures are made. Corrections will be made by drawing a single line through the incorrect entry, enter the correct information, initial and date the change. Under no circumstances should the mistake be erased or made illegible.
- Data or information is not discarded without explanation. To discard, the data or information is crossed out, initialed, dated and the reason for discarding indicated.

1.5.2 Identification of records

- Each record is uniquely identified through a number, code, title, and date and traceable to the procedure, product, or person to which they pertain.

1.5.3 Collection of records

- The individuals who are responsible for collecting records are identified in each specific standard operating procedure or work instruction.

1.5.4 Indexing records

- Records are indexed and organized in series based on activity/function to facilitate their retrieval. Binders, drawers, cabinets, computer disks and tapes, etc., containing records are clearly labeled with identification of their content.

1.5.5 Filing records

- Technical records are filed and maintained at (name location) by (name person or title). Electronic records are stored on a network (e.g., local area network). Electronic records and data files are backed up on a regular basis to safeguard against the loss of information.

1.5.6 Access to records

- There is restricted access to all records to prevent unauthorized use and amending of information except for routine use.
- Routine Use: The use of records in ways that are compatible with the purpose for which the records were collected is considered to be routine.
- Electronic records have password or field protection, or read-only capabilities Passwords should be known only to authorized individuals and changed at regular intervals.

1.5.7 Storage of records

- Paper records will be stored in a secure, dry, and clean location with limited access.
- Storage areas and cabinets are labeled.
- The storage of electronic data is on a server at known locations with the location name identified. The server used for the storage should be in a facility to which there is limited physical and electronic access and in which the climate is controlled. The requirements concerning prevention of damage or deterioration and to prevent loss need to be fulfilled.

1.5.8 Retention

- Record retention times may vary depending on program. They may range from 3–10 years, and some records may be permanent archives.

1.5.9 Disposition (disposal) of records

Records shall be retired and/or purged from active filing systems when no longer current, inactive, when no longer needed, and when projects are completed.

2 Measurement Uncertainty

2.1 Definitions

Coefficient of variation (CV)
Also termed relative standard measurement uncertainty-Standard measurement uncertainty (u(x)) divided by the absolute value x of the measured quantity. CV = u(x)/x.

Combined standard measurement uncertainty (uc)
Standard measurement uncertainty that is obtained using the individual standard measurement uncertainties associated with the input quantities in a measurement model.

Coverage factor (*k*)
Number larger than one by which a combined standard measurement uncertainty is multiplied to obtain an expanded measurement uncertainty.

Coverage interval
Interval containing the set of true values of a measurand with a stated probability, based on the information available.
https://www.iso.org/obp/ui/#iso:std:iso-iec:guide:98:-4:ed-1:v1:en, section 3.2.7

Expanded measurement uncertainty (U)
Product of a combined standard measurement uncertainty and a coverage factor larger than the number one.

Measurand
Quantity intended to be measured.

Measurement uncertainty
Non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used.

Quality
The totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs.

Standard measurement uncertainty (u)
Measurement uncertainty expressed as a standard deviation.

Trueness
Closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value.
https://metrology.wordpress.com/measurement-process-index/3-specification-of-demands-on-a-measurement-method/accuracy/

True value
Quantity value consistent with the definition of a quantity.
https://www.roma1.infn.it/~dagos/cern/node7.html

2.2 Estimation of measurement uncertainty

Even though GUM intends to provide practical guidance for daily use, the calculation of a comprehensive and consistent uncertainty budget can still be a very time-consuming task. Moreover, it often requires information and technical expertise that are not readily available. As a consequence, estimation of uncertainties is either often reduced to the calculation of standard deviations of several measurements or to the statement of device specifications. Unfortunately, this simple approach is inadequate to quantify the reliability of a measured quantity value and may lead to erroneous assessment of measurement data. At least some fundamental principles must be applied to estimate useful uncertainty values, despite the additional effort it may take.

The GUM discusses the evaluation of uncertainty according to a "Type A" or "Type B" method of evaluation. The Type A method evaluates standard uncertainty by the statistical analysis of a series of repeated observations. The Type B method evaluates uncertainty based on scientific judgment using all of the relevant information *available*, which may include:

- Previous measurement data
- Experience with or general knowledge of the behavior and properties of relevant materials and instruments
- Manufacturer's specifications
- Data provided in calibration and other certificates
- Uncertainties assigned to reference data taken from handbooks

Both types of evaluation are based on probability distributions, and the uncertainty components resulting from either type are quantified in terms of a "standard uncertainty." That is a known or reasonably assumed probability distributions assigned to an estimated uncertainty range of each component. The uncertainty range is multiplied with a defined factor so that the resulting uncertainty value corresponds approximately to the standard deviation of a normal distribution, yielding the standard uncertainty. In this way the standard uncertainty expresses a range in which the true value can be found with level of confidence of 68 percent. All identified standard uncertainty components, whether evaluated by Type A and Type B methods, are combined to produce an overall value of uncertainty to be associated with the result of the measurement, known as the combined standard uncertainty. Uncertainty contributions must be expressed in similar terms before they are combined, i.e., all the uncertainties must be given in the same units, and at the same level of confidence.

A generally applicable succession of steps is provided to estimating the overall standard uncertainty $u(x)$ of a measured quantity value x:

1) Specify the quantity to be measured and the measurement procedure for the determination of its value x. In addition to the actual measurement, the procedure comprises all preparative steps, e.g., sampling and sample preparation, calibration, potential adjustment or compensation for measurement error, the conditions that have to be maintained during preparation and measurement, as well as data processing. Identify the uncertainty sources.
2) List possible sources of uncertainty for each process or each parameter of the method and assess whether their contribution to the uncertainty of the result is significant. Do not double count uncertainty components, e.g., repeatability specification of a manufacturer and environmental fluctuations during the measurement.
3) Assign to each source of uncertainty an uncertainty range and the corresponding probability distribution. See supplemental material for the most common probability distributions.
4) Correct the quantity value for systematic errors if they can be quantified. Most often this is only the measurement error measured in a calibration.

5) Assign standard uncertainties to each identified component by multiplying the uncertainty range with the factor corresponding to the identified probability distribution. For details refer to section 4 of the GUM. However, factors of the most common distributions are summarized in the supplemental material.
6) Calculate the combined standard uncertainty $u_c(x)$ of x by building the square root sum of all standard uncertainties.
7) Very often, a quantity value y is calculated from different measured quantities x_j: $y = f(x_1, x_2, \ldots x_N)$, with f being the function used to calculate y from the input quantities x_j. Then, each of the input quantities x_j should be treated according to steps 1 to 6 and the combined standard uncertainty $u_c(y)$ of y is calculated from the individual uncertainties $u(x_j)$ according to

$$u_c^2(y) = \sum_{j=1}^{N} \left(\frac{\partial f}{\partial x_j} u(x_j)\right)^2 \qquad (1)$$

8) Note that f could imply constants or coefficients that might be the result of measurements, even if they had been measured before by someone else. Nevertheless, uncertainties have to be assigned to those values as well and they have to be considered as input variables. The derivative $\partial f/\partial x_j$ is called the sensitivity coefficient of x_j.
9) Express the combined uncertainty of the quantity value y (or x, respectively) in terms of a coverage factor k: $U_c(y) = k\, u_c(y)$. A coverage factor of k=2 is the usual default value, which corresponds to a 95 % confidence interval[1]. The measurement result is given by

$$\boldsymbol{y \pm U_c(y)},$$

10) whereas $U_c(y)$ should be noted with only two relevant digits at maximum and y should be rounded to the same number of digits.

Correlations between input quantities x_j may increase or decrease measurement uncertainty significantly, and therefore they must be considered. A general step-by-step procedure about how to include correlations in uncertainty calculation is given in the supplementary material.

It must be emphasized that the uncertainty calculation presented here and in the supplementary material are the essence of the GUM. Adequate uncertainty calculation should actually be based on the comprehensive guidance given there. However, the recommendations given here can be considered as the most essential requirements to provide a minimum of appropriate measurement uncertainties. To exert less effort is contrary to any good measurement practice. Measurement results with uncertain measurement uncertainties might be acceptable to some extent; however, measurement results without uncertainties are meaningless.

2.3 How to reduce uncertainty in measurement

There are some good practices that generally can help to reduce uncertainties in making measurements.

- Calibrate measuring instruments (or have them calibrated for you) and use the calibration to compensate for measurement error.

[1] Strictly speaking, k=2 corresponds to 95.45 % confidence level; however, it is common to round to zero digits.

- Apply corrections, which are given on calibration the certificate.
- Make corrections to compensate for any (other) errors you know about.
- Make your measurements traceable to national standards by using calibrations that can be traced to national standards via an unbroken chain of measurements.
- Choose the best measuring instruments and use calibration facilities with the smallest uncertainties.
- Check measurements by repeating them, or by getting someone else to repeat them periodically.
- Use an uncertainty budget to identify the worst uncertainties and address these.
- Be aware that in a successive chain of calibrations, the uncertainty increases at every step of the chain, since each calibration must include the uncertainty of the preceding calibration.

The most effective way to explain how measurement uncertainty can be calculated is with the following example for a temperature measurement.

2.4 Example of Uncertainty Calculation: Temperature Measurement

An SBE 39 temperature recorder (Table 6) on a mooring is presented as a simple example to demonstrate the most common aspects of uncertainty calculation. Pressure dependence greatly complicates the calculation and is not considered, yet such observations are commonly conducted (Figure 4). The best estimate is calculated from the mean of ten subsequent temperature measurements. More details on the various uncertainty contributions are given in the supplementary material.

The following represent uncertainty contributions for this example:

Indicated value

Sensor specification states 0.0001 °C resolution
Uncertainty range: 0.0001 °C
Assigned probability distributions: rectangular, corresponding factor $\frac{1}{2\sqrt{3}}$
Standard uncertainty: $\frac{0.0001}{2\sqrt{3}}\,\mathrm{K} = 0.029\ \mathrm{mK}$
Note that this contribution is not considered in the combined uncertainty, since the contribution due to unstable indications is significantly larger (see ‘stability of the measurement’ below)

Measurement error estimated by calibration

Calibration certificate states a measurement error of Δt_{me}=+5.4 mK (difference between indicated value and a measurement standard). The expanded (k=2) uncertainty is stated as 6.3 mK
Assumed probability distribution: normal.
Standard uncertainty: $\frac{6.3\ \mathrm{mK}}{2} = 3.15\ \mathrm{mK}$
For simplicity it is assumed that Δt_{me} applies to the complete temperature range of interest.

Stability of the measurement

Standard deviation of 10 measurements: 0.3 mK

Assumed probability distribution: normal, corresponding factor (is $\frac{1.06}{\sqrt{10}}$)
Standard uncertainty: 0.1 mK

Long-term sensor stability

Sensor specification states 0.2 mK/month (with unknown sign).
Uncertainty range: 0.4 mK (corresponding to ±0.2 mK).
Assumed probability distribution: triangular, corresponding factor $\frac{1}{2\sqrt{6}}$
Standard uncertainty $\frac{0.4\text{ mK}}{2\sqrt{6}} m = 0.49$ mK after six months
m: the number of months elapsed since calibration.

Self-heating

Sensor specification: 0.2 mK,
Uncertainty range: 0 to 0.2 mK, (asymmetric, *t* only increases)
Probability distribution: rectangular, corresponding factor $\frac{1}{2\sqrt{3}}$
Standard uncertainty: $\frac{0.0002}{2\sqrt{3}}\text{K} = 0.058\text{ mK}$

Sensor installation

Estimated uncertainty: +/-1 mK, corresponding to an uncertainty range of 2 mK.
Assumed probability distribution: triangular, corresponding factor $\frac{1}{2\sqrt{6}}$
Standard uncertainty: $\frac{2\cdot\text{mK}}{2\sqrt{6}} = 0.41\text{ mK}$

Repeatability

Repeatability 0.11 mK (measured with an independent experiment)
Assumed probability distribution: normal, corresponding factor: 1
Standard uncertainty: 0.11 mK
Since this value is similar to the stability contribution, it is assumed that repeatability is mostly determined by stability. Therefore, only the stability contribution is considered in the combined uncertainty.

Reproducibility

Not considered, since it is assumed to be covered by the combined uncertainty of all other contributions.

Assuming the temperature is measured six months after calibration of the SB39 and that the measured mean value is 15.1427 °C, the temperature value *t* corrected for measurement error is

$$t = 15.1427\text{ °C} - 0.0054\text{ °C} = 15.1373\text{ °C}$$

having a combined standard uncertainty of

$$u_c(t) = \sqrt{0 + 3.15^2 + 0.1^2 + 0.49^2 + 0.058^2 + 0.41^2 + 0}\text{ mK} = 3.22\text{ mK}.$$

Hence, the measurement result is (15.1373 ± 0.0064) °C.

Here, the expanded uncertainty is noted with a coverage factor of 2, indicating the true value is in the stated uncertainty interval with 95% probability.

Uncertainty Propagation

Very often several measurement results x_i are used to calculate a quantity *y*. Therefore, uncertainty propagations of all input quantities x_i have to be considered to calculate the uncertainty of the output quantity *y*. To demonstrate uncertainty propagation speed of sound v_{SoS} is calculated from a temperature (*t*) measurement (13.601 ± 0.0064) °C, from a corresponding salinity (*S*) measurement (38.7 ± 0.002) and a depth (*Z*) measurement (500 ± 2) m. To this end a simplified equation of state is assumed[2] that implies the latitude Φ of the measurement location.

$$\boldsymbol{v_{SoS}(t,S,Z,\Phi) = 1402.5 + 5t - 5.44 \cdot 10^{-2}t^2 + 2.1 \cdot 10^{-4}t^3 + 1.33S - 1.23 \cdot 10^{-2}St + 8.7 \cdot 10^{-5}St^2 + 1.56 \cdot 10^{-2}Z + 2.55 \cdot 10^{-7}Z^2 - 7.3 \cdot 10^{-12}Z^3 + 1.2 \cdot 10^{-6}Z(\phi - 45) - 9.5 \cdot 10^{-13}tZ^3 + 3 \cdot 10^{-7}t^2Z + 1.43 \cdot 10^{-5}SZ}$$

The corresponding uncertainty contributions to the uncertainty of v_{SoS} at *t*=13.601°C, S=38.7, Z=500m and Φ=35°) are

$$u_t = c_t \cdot u_c(t) = \left.\frac{\partial v_{SoS}}{\partial t}\right|_{(13.601°C,38.7,500m,35°)} \cdot u_c(t)$$
$$= [5 - 0.0123S - 0.1088t + 0.000174St + 0.00063t^2 + 6 \cdot 10^7 tZ - 9.5 \times 10^{-13}Z^3]_{(13.6,38.7,500,35)} \cdot 0.0064°C = 3.2564\,\frac{\mathrm{m\,s^{-1}}}{°C} \cdot 0.0064°C$$

$$u_S = c_S \cdot u_c(S) = \left.\frac{\partial v_{SoS}}{\partial S}\right|_{(13.601°C,38.7,500m,35°)} \cdot u_c(S)$$
$$= [1.33 - 0.0123t + 0.000087t^2 + 0.0000143Z]_{(13.6,38.7,500,35)} \cdot \ 0.002$$
$$= 1.186\ \mathrm{ms^{-1}} \cdot \ 0.002$$

$$u_Z = c_Z \cdot u_c(Z) = \left.\frac{\partial v_{SoS}}{\partial Z}\right|_{(13.601°C,38.7,500m,35°)} \cdot u_c(Z)$$
$$= [0.0156 + 0.0000143S + 10^7 3t^2 + 5.1 \times 10^{-7}Z - 2.19 \times 10^{-11}Z^2 - 2.85 \times 10^{-12}tZ^2 + 0.0000012(-45 + \phi)]_{(13.6,38.7,500,35)} \cdot 2\mathrm{m} = 0.01644\,\frac{\mathrm{m\,s^{-1}}}{\mathrm{m}} \cdot 2\mathrm{m}$$

The uncertainty of the latitude is neglected here. Moreover, the authors of (2) state an upper uncertainty limit of ±0.2 m s^{-1} of the simplified equation which has to be considered in terms of a standard uncertainty of the model equation

[2] Leroy et al, New ocean sound speed equation, J. Acoust. Soc. Am., Vol 124, No 5, 2008

$u_{model} = \frac{0.4\mathrm{m}s^{-1}}{2\sqrt{6}}$,

assuming a triangular probability distribution.

Consequently, the combined standard uncertainty of the speed of sound result is given according to equ. (1)

$$u_c(v_{SoS}) = \sqrt{(c_t u_c(t))^2 + (c_S u_c(S))^2 + (c_Z u_c(Z))^2 + u_{model}^2} = 0.090$$

Table 6. Specification List of the SBE 39 Thermometer (https://www.seabird.com).

	Temperature (°C)	Strain-Gauge Pressure (optional)
Measurement Range	-5 to +45	0 to full scale range: 20 / 100 / 350 / 600 / 1000 / 2000 / 3500 / 7000 m *Pressure expressed in meters of deployment depth capability.*
Initial Accuracy	± 0.002 (-5 to 35 °C) ± 0.01 (35 to 45 °C)	± 0.1% of full scale range
Typical Stability	0.0002/month	0.05% of full scale range/year
Resolution	0.0001	0.002% of full scale range
Sensor Calibration	-1 to +32	Ambient pressure to full scale range in 5 steps
Memory	64 Mbyte non-volatile FLASH memory	
Data Storage	Per sample: temperature 3 bytes, time 4 bytes, pressure (optional) 5 bytes. **Recorded Parameters** — **Memory Space (total samples)** T and time — 9,500,000 T, P, and time — 5,500,000	
Real-Time Clock	32,768 Hz TCXO accurate to ±1 minute/year	
Internal Battery Pack	7.2 V, 5.2 Amp-hour pack consisting of 4 AA Saft LS 14500 AA lithium cells (3.6 V and 2.6 Amp-hours each) (see *Shipping Precautions* in *Section 1: Introduction*)	
Sampling Speed	User-programmable 0.5-sec to 6-hour intervals	
Power Consumption	*Sample acquisition (per sample)*: T only (2.5 mA, 280 msec): 0.00070 A-sec T and P (3 mA, 280 msec): 0.00084 A-sec *Real-time data transmission:* 6 mA. Transmission time varies, depending on output format and baud (**shown at 9600 baud;** for other bauds, multiply by 9600/baud): T and Time: Raw – 62 msec, Converted decimal – 40 msec, XML – 250 msec T, P, and Time: Raw – 96 msec, Converted decimal – 50 msec, XML – 280 msec *Quiescent*: 25 µA	
Battery Pack Endurance (for 39*plus* with pressure sensor)	>12 million samples at 0.5-second sampling rate; > 11 million samples at 5-second sampling rate Notes: 1. This endurance is achieved if 39*plus* is deployed in **recommended orientation: thermistor end down or horizontal** (see *Deployment Orientation* in *Section 2: Description of SBE 39plus*). 2. **Deployment length may be limited by memory** (see Data Storage specification). 39*plus* continues sampling and transmitting real-time data after memory is full and does not overwrite data in memory.	
Optional External Power (with external connector)	9-30 VDC	
Housing, Depth Rating	PET plastic, 600 meters (1960 feet) Titanium, 10,500 meters (34,400 feet)	
Weight (without external connector, clamp, or fairing/net fender)	*Plastic housing with embedded thermistor*: In water: 0.25 kg (0.6 lb) In air: 0.6 kg (1.2 lb) *Titanium housing with thermistor in sheath*: In water: 0.7 kg (1.6 lb) In air: 1.2 kg (2.6 lb)	

Figure 2. A CTD, typically used for temperature measurements as described here, is deployed following the Deepwater Horizon incident. Temperature profiles from such instruments are among the most common oceanographic observations conducted. Photo credit: Mark Bushnell

2.5 Standard Uncertainties of Common Probability Distributions

In many cases little information is available on the true probability distribution of a measurement quantity, and it is impractical, if not impossible, to perform a sufficiently large number of measurements in order to determine the probability distribution and to estimate the standard uncertainty on solid statistical grounds. Nevertheless, reasonable assumptions most often can be made on the probability distribution to assign approximate standard uncertainties to measurement results.

a) Assuming normal distribution for Type A uncertainties

The standard uncertainty of the best estimate *x* of a measurement quantity is given by the standard deviation of the mean of several quantity values x_i, measured under the same conditions (repeatability conditions)

$$u(x) = \frac{a}{\sqrt{N}} \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - x)^2} \qquad \text{(A1)}$$

x is calculated as the arithmetic mean of the results x_i.

a is a factor that depends on the number *N* of measurements and is given in the table below. Basically, a small number of measurements provides a worse estimate of the true spread of the results than a large number which is considered by *a*.

Table 1. Examples of the relationship between the number of measurements N and the factor *a*.

N	2	3	4	5	6	8	10	20	30	∞
a	1.84	1.32	1.20	1.15	1.11	1.08	1.06	1.03	1.02	1

b) Assuming normal distribution for Type B uncertainties

If the uncertainty value $u(x)$ originates from an external source, e.g. instrument documentation, calibration certificate or literature, and a level of confidence is stated, the stated uncertainty value has to be divided by the corresponding coverage factor as given in the following table to get the standard uncertainty, even if it has not been explicitly mentioned that the given uncertainty is based on a normal distribution.

Table 2. Commonly used values of the level of confidence and the corresponding coverage factor.

Level of confidence	68 %	95,45 %	99 %
Coverage factor k	1	2	2.576

c) Rectangular probability distribution

If no information of the probability distribution is available at all, but upper and lower limits a_+ and a_- can be assumed for the uncertainty range, it is reasonable to assume a constant probability distribution within these limits. The standard uncertainty is then given by

$$u(x) = \frac{a_+ - a_-}{2\sqrt{3}} \quad \text{(A2)}$$

Note that the equation should also be applied, if a_+ and a_- are not symmetrically arranged around *x*. The limited resolution of digital displays is a common example of a rectangular distribution.

d) Trapezoidal probability distribution

Step function discontinuities in a probability distribution like in c) are often unphysical. In many cases, it is more realistic to expect that values near the bounds are less likely than those near the midpoint. It is then reasonable to replace the symmetric rectangular distribution with a symmetric trapezoidal distribution having equal sloping sides, with a base of width $a_+ - a_-$, and a top of width $\beta(a_+ - a_-)$, where $0 < \beta < 1$. The standard uncertainty is then given by

$$u(x) = \frac{\sqrt{1+\beta^2}(a_+ - a_-)}{2\sqrt{6}} \qquad \text{(A3)}$$

If no further information is given it is reasonable to assume a triangular probability distribution (β=0) with

$$u(x) = \frac{a_+ - a_-}{2\sqrt{6}} \qquad \text{(A4)}$$

It is not always easy to identify the correct probability function based on the available information and sometimes the choice must be arbitrary to some extent. However, even if the eventually chosen distribution does not reflect the true probability distribution the resulting uncertainty of the uncertainty is usually acceptable, and it is in any case preferable to omitting relevant contributions at all.

2.6 How to Consider Correlations

Uncertainties might be significantly under- or overestimated if correlations are not considered. Not all uncertainty components of two correlated quantities x_1 and x_2 are affected by correlations. If, for instance, two measurement instruments are calibrated with the same standard, the uncertainty of the standard used for calibration will affect the measurement error of both instruments likewise. Consequently, the uncertainty components due to calibration are correlated, while repeatability components might not be correlated if the devices are eventually used in different environments. The following rather formal steps can be applied to account for such correlations:

1. Identify significant uncertainty contributions as described above. Assign formal input variable names x_{j_vvv} for each uncertainty contribution assumed to be significant for an input quantity x_j, e.g. x_{j_repeat} for repeatability of the measurement of x_j, x_{j_ls} for long term stability of the sensor used to measure x_j, etc. This must be conducted for each input quantity x_j.
2. Assign quantity values to each formal input variable. Note that usually there is a formal input variable assigned to the value indicated by the instrument, named x_{j_ind} for instance. Another input variable x_{j_me} is assigned to the measurement error, which is usually indicated in the calibration certificate. The quantity values of all other formal input variables are usually set to zero, unless they correspond to quantifiable deviations from the indicated value that could be compensated.
3. Assign the corresponding standard uncertainty $u(x_{j\_vvv})$ to each formal input quantity value x_{j_vvv}. Note that in this way each formal input quantity value has only a single uncertainty contribution assigned which is the reason for splitting each input quantity x_j into several formal input quantities.
4. Identify all pairs (x_{m_vvv1};x_{n_vvv2}) of formal input quantities that are correlated and assign a correlation coefficient $r(x_{m\_vvv1};x_{n\_vvv2})$, with $-1 \leq r \leq 1$. The correlation coefficients of all other pairs are set to zero. The most common rules to assign correlation coefficients are summarized below.
5. Express each x_j formally as sum of the corresponding formal input variables x_{j_vvv} and insert the expressions in f. Hence, f is now a function of all formal input variables x_{j_vvv}.
6. Calculate the combined standard uncertainty of y according to:

$$u_c^2(y) = \sum_{j\_vvv=1}^{M} \left(\frac{\partial f}{\partial x_{j\_vvv}} u(x_{j\_vvv}) \right)^2 +$$
$$2 \sum_{m\_vvv1=1}^{M-1} \sum_{n_{vvv2}=m\_vvv1+1}^{M} r(x_{m\_vvv1}; x_{n\_vvv2}) \frac{\partial f}{\partial x_{m\_vvv1}} u(x_{m\_vvv1}) \frac{\partial f}{\partial x_{n\_vvv2}} u(x_{n\_vvv2}) \quad \text{(A5)}$$

with M expressing the total number of formal variables x_{j_vvv}. Note that usually just a small subset of all potential formal input quantities is really correlated, so that the expression is less extensive as it seems.

7. Express the combined uncertainty of the quantity value y in terms of a coverage factor k: $U_c(y) = k\, u_c(y)$.

From a practical point of view, it is often very difficult to estimate correlation coefficients of any two input quantity values v_1 and v_2 in order to account for correlations. However, some basic rules can be applied. Note that v_1 and v_2 are representative variable names for x_{m_vvv1} and x_{n_vvv2} chosen here for formal simplicity

a) v_1 and v_2 can be considered uncorrelated, i.e. $r(v_1;v_2) = 0$, if
   - measured in independent experiments, i.e. at different times, with different instruments, by different operators, etc.
   - one of them is a constant
   - information on potential correlation is insufficient

b) Statistical evaluation

   If K values of two different input quantities are measured at quasi the same time, e.g. temperature and conductivity, the resulting pairs $(v_{1i};v_{2i})$ can be used to calculate the correlation coefficient $r(v_1;v_2)$

$$r(v_1; v_2) = \frac{\sum_i^K (v_{1i} - v_1)(v_{2i} - v_2)}{\sqrt{\sum_i^K (v_{1i} - v_1)^2 \sum_{i=1} (v_{2i} - v_2)^2}} \quad \text{(A6)}$$

   with v_1 and v_2 being the arithmetic means of the individually measured values v_{1i} and v_{2i}, respectively.

c) Analytical correlation through a joint third quantity
   In practice, input quantities are often correlated because the same physical measurement standard, measuring instrument, reference date, or the same measurement method is used in the estimation of their values. If v_1 and v_2 can be expressed in terms of a joint quantity q such that $v_1=g(q)$ and $v_2=h(q)$ the correlation coefficient is given by

$$r(v_1; v_2) = \frac{\partial g}{\partial q} \frac{\partial h}{\partial q} \frac{u(q)^2}{u(v_1)u(v_2)} (\quad \text{A7)}$$

d) Same measurement conditions

If v_1 and v_2 are of the same kind of quantity, of compatible magnitude, measured using the same instruments and within a time period in which the equipment is reasonably stable it can be assumed that $r(v_1;v_2)=1$. This is especially applicable in differential measurements for instance.

2.7 Example of Temperature Measurement

The uncertainty contributions mentioned in the example of a temperature measurement in the *Measurement Uncertainty* section are discussed here in more detail. Furthermore, the formalism of variable splitting to consider correlations adequately is implemented in the example. It must be emphasized that, apart from sensor specification data, the numbers given here are not the result of an actual measurement but have been arbitrarily chosen just for the purpose of demonstrating uncertainty calculation.

2.8 Indicated value

Sensor specification states 0.0001 °C resolution. To account for the limited resolution a standard uncertainty of

$$u(t_{ind}) = \frac{0.0001}{2\sqrt{3}}\,\mathrm{K} = 0.029\ \mathrm{mK}$$

is assigned to the temperature value t_{ind} indicated by the instrument. The term “indicated” refers to the temperature value provided by the instrument, no matter how it is actually provided (display, printed, automatically saved in an internal memory or in a data base, etc.).

Assumed probability distribution: rectangular, since the true value must be somewhere within the resolution range.

Note: If fluctuation of the reading is significantly larger than the uncertainty due to limited resolution, t_{ind} should be calculated from the mean of several readings, however, $u(t_{ind})$ can be set to zero. The fluctuation will be considered in the stability or repeatability contribution (see below). In turn, repeatability and instability need not to be considered if they are smaller than the resolution.

2.9 Measurement error estimated by calibration

It is assumed that a calibration has been performed at atmospheric pressure prior to its usage. The calibration certificate states a value of t_{actual}=(15.0024 +/- 0.00056) °C, measured with the SBE 39, and a reference value of t_{ref}=(14.9970 +/- 0.0062) °C, measured with a temperature measurement standard. The uncertainties are stated as expanded (k=2) uncertainties.

Assumed probability distribution: normal, since a k factor that corresponds to a confidence level of about 95% is stated.

The measurement error $\Delta t_{me} = t_{actual}- - t_{ref}$ the device is +5.4 mK and its (combined) standard uncertainty is

$$u(\Delta t_{me}) = \sqrt{\left(\frac{0.00056\ °\mathrm{C}}{2}\right)^2 + \left(\frac{0.0062\ °\mathrm{C}}{2}\right)^2} = 3.15\ \mathrm{mK}$$

For simplicity it is assumed that this error applies to the complete temperature range of interest.

Note that the "Initial Accuracy" statement of the manufacturer is irrelevant for uncertainty calculation, since it gives no information about the measurement error of the sensor by the time of its usage. It just gives information about the general quality of the (new) sensor (compared to a less accurate sensor technique for instance).

2.10 Stability

Stability has to be calculated from the standard deviation of several readings without changing the measurement conditions. It is an estimate for the stability of the instrument and the environmental conditions during the measurement. Here, it assumed that the standard deviation of 10 measurements is 0.3 mK. The measurement error □tstab due to stability must be (formally) set to zero, since the actual measurement error introduced by instability at the time of measurement cannot be quantified. Instead, a standard uncertainty of

$$u(\Delta t_{stab}) = \frac{0.3 \cdot 1.06 \text{ mK}}{\sqrt{10}} m = 0.1 \text{ mK}$$

is assigned to Δt_{stab}.

Assumed probability distribution: normal, corresponding factor is $\frac{1.06}{\sqrt{10}}$.

Note that $u(\Delta t_{stab})$ can be set to zero, if it is either smaller than the resolution of the device or significantly smaller than measurement repeatability. If repeatability is worse than the spread of the indicated results there are other effects that determine the repeatability of the measured quantity than just instability of the measurement device and the environment during the measurement, e.g. uncertainties due to preparation steps before a measurement. If a drift is observed during repeated readings, it should be considered like the long-term stability contribution described in section 4.4.

2.11 Long-term sensor stability

Sensor specification states 0.2 mK/month. No information is given if the sensor signal increases or decreases. Consequently, the actual uncertainty range must be assumed ±0.2 mK/month, resulting in an uncertainty range of 0.4 mK. The measurement error $\Delta t_{\mathrm{long\_stab}}$ due to long-term stability must be (formally)[3] set to zero, since the actual measurement error by the time of measurement cannot be quantified. Instead, a standard uncertainty of

$$u\left(\Delta t_{long\_stab}\right) = \frac{2 \cdot 0.2 \text{ mK}}{2\sqrt{6}} m = 0.49 \text{ mK}$$

is assigned to Δt_{long_stab} to account for instability, with *m* indicating the number of months passed since the calibration. Here *m* is assumed to be 6 months.

[3] The procedure to assign formal input variables to the various uncertainty contributions is described in more detail with respect to correlations in section 3 of this supplementary material.

Assumed probability distribution: triangular. Here it supposed that the actual drift is more likely to be less than the value stated by the manufacturer. Therefore, a rectangular distribution is unlikely.

2.12 Self-heating

Sensor specification: 0.2 mK. The deviation Δt_{sh} from the true value is (formally) set to zero, because it is not known. To account for self-heating a standard uncertainty of

$$u(\Delta t_{sh}) = \frac{0.0002}{2\sqrt{3}}\,\mathrm{K} = 0.058\ \mathrm{mK}$$

is assigned to Δt_{sh}.

Assumed probability distribution: rectangular. Actually, self-heating is an asymmetric uncertainty contribution, since self-heating will always result in a positive offset from the true value. Nevertheless, symmetric distribution can be applied to the uncertainty.

Note that a triangular uncertainty would be as reasonable. However, pondering on the most adequate contributions is often not worth the effort, if the uncertainty contribution is rather small. Basically, it is more important to come up with a more or less reasonable estimate at all, in order to assess its effect compared to the contributions.

2.13 Sensor installation

Experience has shown that the installation of the sensor into the platform affects the heat exchange between the temperature sensor and the surrounding medium, e.g. flushing of the sensor. However, the sign of the deviation cannot be predicted. This results in an estimated uncertainty of +/-1 mK. The deviation Δt_{inst} from the true value is (formally) set to zero, because it is not known. Instead, a standard uncertainty of

$$u(\Delta t_{inst}) = \frac{2 \cdot 1\ \mathrm{mK}}{2\sqrt{6}} = 0.41\ \mathrm{mK}$$

is assigned to Δt_{inst}.

Assumed probability distribution: triangular. Here it supposed that the actual deviation is more likely to be centered within the uncertainty range.

2.14 Repeatability of the actual measurement

Frequent repeatability measurements, i.e., spread of temperature values of several subsequent measurements using the same instrument and applying the same measurement procedure under nominal the same conditions, can hardly be realised in long term field measurements. Instead, an independent, representative measurement series can be performed once, to estimate measurement repeatability. Here, preliminary measurements are assumed that have shown a typical standard deviation of 0.1 mK. The deviation δt_{repeat} of a measured, single value from the true value, i.e. the mean of a hypothetic infinite number of measurements, is (formally) set to zero, because it is not known. Instead, a standard uncertainty of

$$u(\delta t_{repeat}) = 0.1\ \mathrm{mK}$$

is assigned to δt_{repeat}.

Assumed probability distribution: normal.

Note that repeatability must not be confused with stability. The later estimates the just spread of several subsequent readings while nothing is changed. Repeatability estimates the spread when the whole measurement procedure is repeated, including preparation of the measurement setup, water bath preparation, etc., but using the same instruments, samples, etc.

In this example, repeatability is assumed to be significantly larger than the contributions due to instrument resolution, but similar to the value estimated for stability. Hence, the stability is the only contribution considered in the below calculation of the combined standard uncertainty.

2.15 Reproducibility

Reproducibility, i.e., the spread of temperature values measured by different people, with different (compatible) instruments under nominally the same conditions, is rather difficult to quantify, since it can only be estimated by an interlaboratory comparison that is adequately designed to reflect the actual measurement conditions. However, if the calibration laboratory that performs the calibration is part of the international accreditation system it has to participate regularly in comparison measurements. The minimal uncertainty it may assign to its calibration results is determined by its performance in such comparisons. The same holds for the reference values of certified calibration standards. Hence, the uncertainty of the measurement error estimated by such a calibration already includes, at least to some extent, the reproducibility contribution. Since certified calibration is assumed here and all other potential uncertainty sources are considered, no further reproducibility contribution is included in this calculation. However, in practice, this must be decided on a case by case basis. In fact, occasional comparison measurements between oceanographic laboratories and measurement facilities could validate if assigned uncertainties are appropriate.

The temperature value t calculates from

$$t = t_{ind} - \Delta t_{me} - \Delta t_{stab} - \Delta t_{long\_stab} - \Delta t_{sh} - \Delta t_{inst} - \Delta t_{repeat} \qquad \text{(A8)}$$

using the numbers given in the example

$$t = 15.1427\ °\mathrm{C} - 0.0054\ °\mathrm{C} = 15.1373\ °\mathrm{C}$$

Eq. A8 constitutes the measurement function f for the output quantity t. Since most potential deviations Δt_{vvv} are set to zero the final temperature value is only determined by the indicated value t_{ind} and the compensation for bias Δt_{me} determined in a calibration. f allows to formally introduce all uncertainty contributions according to eq. (A1). Consequently, applying eq. (A1) gives the uncertainty of t.

$$u(t) = \sqrt{u(t_{ind})^2 + u(\Delta t_{me})^2 + u(\Delta t_{stab})^2 + u(\Delta t_{long\_stab})^2 + u(\Delta t_{sh})^2 + u(\Delta t_{inst})^2 + u(\Delta t_{repeat})^2}$$

In this example the uncertainty of t, six months after calibration, is calculated as follows:

$$u_c(t) = \sqrt{0 + 3.15^2 + 0.1^2 + 0.49^2 + 0.058^2 + 0.41^2 + 0}\ \text{mK} = 3.22\ \text{mK}\ .$$

and $U_c(t) = 6.4$ mK.

2.16 Example on How to Consider Correlations

The example presented here describes a simple correlation and its effect on uncertainty. The temperature difference $\Delta t = t_2 - t_1$ between two measurements at two different sites using the same instrument is defined as the measurand. Both temperature values t_1 and t_2 are the means of several measurements. They are both split into various formal variables as described above, having the same uncertainties as mentioned in the previous section. An analysis of the uncertainty contributions identifies the following correlations:

- (Δt_{1_me};Δt_{2_me}), since the same instrument is used. Therefore, $r(\Delta t_{1\_me};\Delta t_{2\_me})=1$.
- ($\Delta t_{1\_long\_stab}$;$\Delta t_{2\_long\_stab}$), since the same instruments is used and t_1 and t_2 are measured within a reasonably small period. Therefore, $r(\Delta t_{1\_long\_stab};\Delta t_{2\_long\_stab})=1$.
- It is supposed for this example that a small drift within the instrument occurs after starting a measurement. Therefore, the measurement values forming the means t_1 and t_2 are suspected to be correlated. Application of eq. A6 (not shown) is assumed to reveal a correlation coefficient of $r(t_{1\_stab};t_{2\_stab})=0.8$.

All other formal input quantities are assumed independent. This is not necessarily true. Unfortunately, the estimation of adequate correlation coefficients is often too difficult and too elaborate for practical use, if not impossible. In this case a more conservative uncertainty estimation could assume a correlation coefficient of 1 if correlation increases the combined uncertainty of Δt and, as it is the case here, it could assume a correlation coefficient of 0, if correlation decreases the combined uncertainty of Δt. However, given uncertainty contributions are small it is advisable to assess first if further elaborate correlation calculation is worth the effort.

The quantity value of the measurand Δt calculates from

$$\Delta t = t_{2\_ind} - \Delta t_{2\_me} - \Delta t_{2\_stab} - \Delta t_{2\_long\_stab} - \Delta t_{2\_sh} - \Delta t_{2\_inst} - \Delta t_{2\_repeat}$$

$$-(t_{1\_ind} - \Delta t_{1\_me} - \Delta t_{1\_stab} - \Delta t_{1\_long\_stab} - \Delta t_{1\_sh} - \Delta t_{1\_inst} - \Delta t_{1\_repeat})$$

Based on the given numbers this expression reduces to

$$\Delta t = t_{2\_ind} - t_{1\_ind} \ ,$$

since the values of Δt_{1_me} and Δt_{2_me} cancel and all other contributions are zero.

Applying equation A5 the uncertainty of Δt gives

$$u(\Delta t)^2 = 2\left[u(t_{ind})^2 + u(\Delta t_{me})^2 + u(\Delta t_{stab})^2 + u(\Delta t_{long\_stab})^2 + u(\Delta t_{sh})^2 + u(\Delta t_{inst})^2 + u(\Delta t_{repeat})^2\right] - 2\left[r(\Delta t_{1\_me}; \Delta t_{2\_me}) u(\Delta t_{me})^2 + r(\Delta t_{1\_stab}; \Delta t_{2\_stab}) u(\Delta t_{stab})^2 + r(\Delta t_{1\_long\_stab}; \Delta t_{2\_long\_stab}) u(\Delta t_{long\_stab})^2\right]$$

$$u(\Delta t) = \sqrt{2 \cdot 3.22^2 \ - 2[1 \cdot 3.15^2 + 0.8 \cdot 0.1^2 + 1 \cdot 0.49^2]} = 0.60 \text{ mK}$$

Obviously, the uncertainty of the difference is much smaller compared to an individual *t* value, since the main uncertainty contribution, i.e. the uncertainty of the measurement error, cancels. A similar effect can be shown for ratios of two quantity values of similar size, which is the main reason for the definition of Practical Salinity in terms of a conductivity ratio.

3 References and relevant documents

American Association for Laboratory Accreditation. 2014. G104 - Guide for Estimation of Measurement Uncertainty in Testing.

Bell, S. 1999. A Beginner's Guide to Uncertainty of Measurement. National Physical Laboratory, Teddington, Middlesex, United Kingdom.

BIPM, JCGM 100:2008. 2008. Evaluation of measurement data - Guide to the Expression of Uncertainty in Measurement (GUM). International Organization for Standardization. Geneva, Switzerland.

BIPM JCGM 200:2012. 2012. International vocabulary of metrology - Basic and general concepts and associated (VIM) 3rd edition (2008 version with minor corrections). Geneva, Switzerland.

EUROLAB Technical Report 1/2006. 2006. Guide to the Evaluation of Measurement Uncertainty for Quantitative Test Results. Paris.

International Organization for Standardization (2017). General requirements for the competence of testing and calibration laboratories. ISO/IEC 17025:2017, Section 7.5.

ISO/TC 8/SC 13 Marine technology, https://www.iso.org/committee/5317919/x/catalogue, in development. Accessed November 27, 2018.

Magnusson, B., T. Näykki, H. Hovind, M. Krysell, and E. Sahlin, 2017. Handbook for calculation of measurement uncertainty in environmental laboratories, Nordtest Report. TR 537 (ed. 4) 2017.

Magnusson, B., T. Näykki, H. Hovind, M. Krysell, and E. Sahlin, 2017. Handbook for calculation of measurement uncertainty in environmental laboratories, Nordtest Report. TR 537 (ed. 4) 2017.

Myers, J.J., Holm, C.H., and McAllister, R.F. (1969). Handbook of ocean and underwater engineering. United States: N. p., 1969. Web. p. 7-2.

Ramsey, M.H. and S L R Ellison (eds.) 2007. Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide: Measurement uncertainty arising from sampling: a guide to methods and approaches.

USEPA. 2016. Field Measurement Uncertainty Standard Operating Procedure, https://www.epa.gov/sites/production/files/2016-04/documents/field_measurement_uncertainty014_af.r2.pdf, accessed Nov 27, 2018.