

Interactive comment on “A guide for upper-air reference measurements” by F. Immler et al.

F. Immler et al.

franz.immler@dwd.de

Received and published: 20 August 2010

The review from Thierry Leblanc provides a number of very helpful comments that we like to thank him for. The issues he raised (quoted here in italic) are addressed in detail in the following:

1. Title: The title is too general to my opinion. Not being a native English speaker, I unfortunately cannot provide a much better solution, but in any case, it should reflect the fact that the paper treats more specifically the proper handling of measurement and data processing errors rather than being a comprehensive review of all aspects of the data processing chain (from measurement to end user). My best two suggestions for a revised title are: “Towards robust upper air reference measurements” or “Optimizing the accuracy and long-term stability of upper air measurements”. Any well-designed

C1213

combination of these two should work. There are enough native English-speakers among the co-authors to address this efficiently. Also, if acronyms are allowed in the title, it would be appropriate to include “GRUAN”.

There is certainly a point about changing the title to something more specific. With the approval of the editor we would like to change the title to "Reference Quality Upper-Air Measurements: Guidance for developing GRUAN data products."

2. Paragraph 2.1: “expected value of zero”. Should this instead be “expected averaged value is zero”?

"expected value" or "expectation value" is a term used in probability theory (http://en.wikipedia.org/wiki/Expected_value) which describes the first moment of a random variable. We used this expression instead of "mean" because "mean" is easily confused with "sample mean" which provides an estimator for the expected value but is not the same (except for when N , the size of the sample, goes to infinity).

The paragraph has been rephrased to make its relation to probability theory more obvious and also to define more clearly what the term "uncertainty" refers to:

The lack of exact knowledge of the value of the measurand is characterized by a random variable, the uncertainty U , which is evaluated from the uncertainties of all input quantities, including the uncertainties of all corrections that were applied for systematic effects. Provided that proper corrections have been made for all systematic effects, the expectation value (or "expected value") of uncertainty U is zero. In this case, the uncertainty of the measurement result can therefore be expressed by one single value, the standard uncertainty u which is the estimated standard deviation of the random variable U .

3. Paragraph 2.1: Please define “accuracy” in the glossary

The following paragraph is added to the glossary

Accuracy: closeness of agreement between the result of a measurement and a true value of the measurand

C1214

4. Fig.1: To my opinion, this figure does not convey its message efficiently, i.e., it does not illustrate clearly the data processing chain. Some improvement would be needed to make it clearer.

Fig.1 is not meant to illustrate a data processing chain. To express this more clearly the caption is changed to:

Conceptual traceability chain illustrating how the calibration of a sensor for deployment is tied to the realization of a SI unit. Each calibration step is defined by a comparison between two measurements with a stated, realistic uncertainty. All relevant details of the measurement comparison that can influence the measurement result must be recorded.

Also the following line was added to the text on page 1815, line 21:

Figure 1 shows the conceptual traceability chain for an upper air measurement, indicating the steps required to link the measurement to the fundamental SI units.'

5. Paragraph 2.4 (top of page 1817): "reproducible by the end-user" I do not agree with this statement. The end-user does not necessarily have to reprocess from scratch. Instead, it is the role of the data provider and his/her successors to be able to reprocess the data from scratch. Unlike the data provider, the end-user does not have a full knowledge of the instruments and techniques; this is the reason why a full error assessment and a traceable metrology are required, and why they should be provided to the end-user by the data provider.

The data users in the future do not have to reprocess everything from scratch, but they might want to. One reason could be that it is suspected that something was processed incorrectly. In this case they indeed should be provided with all the information necessary to do such a reprocessing.

6. End of paragraph 2.5: "if one of the two measurements does not provide uncertainties" This paragraph is unclear. Please expand to justify how a meaningful consistency analysis is still possible in this case.

C1215

We do think that statistical testing is still meaningful in the case where one measurement does not provide uncertainty estimates. Setting the uncertainty to zero for the sake of the test does not mean it is assumed to be zero in reality. In fact the test is more powerful this way rather than if some finite values for the uncertainty were assumed. The paragraph is rephrased to make the procedure clearer:

If one of the two measurements does not provide uncertainties, the same methodology can still be used. The consistency analysis is made in the same way by setting u_2 in equation (6) to zero (making the test statistically more powerful, i.e. the risk of yielding false confidence lower, than by arbitrarily assigning some finite value to u_2). This is equivalent to the notion that this value "does or does not lie within the error bars with a specified coverage factor of the reference measurement".

7. Paragraphs 3.3 and 3.4: There is no details given on the way validation should be handled. Do the authors have any specific ideas already, and can these be included in paragraphs 3.3 and 3.4?

Considering the handling of validation a new subsection was formed from a part from former section 3.3. with the title "Defining and validating a GRUAN data product" and the following paragraph is inserted to the text :

The operational concept that describes measurement method, calibration, procedures, and algorithms, including those used for corrections and estimation of uncertainties, establishes a data product for GRUAN. Such a data product needs to be validated before implementation as a product of the GRUAN network.

The validation will be made using redundant measurements and testing for agreement as described in section 2.5. Validation is first and foremost a validation of the uncertainty estimates. Agreement of two independent measurements, preferably based on different measurement principles, provides a high degree of confidence that no significant systematic effect was disregarded and uncertainties were not underestimated. As a larger number of comparisons become available, statistical analysis permits the uncertainty estimates to be evaluated further.

C1216

Referring to the significance levels indicated in table 1, one can deduce that if the measurements agree ($|m_1 - m_2| < 2 \cdot \sqrt{u_1^2 + u_2^2}$) in more than 95% of all cases, the uncertainties are likely to be smaller than estimated.

One question of particular interest is: Is there any efficient way of combining results from lab tests and in-field campaigns?

We think that the best way of combining results from lab tests and in-field campaigns is by using them for establishing and validating uncertainty estimates. To work this out more clearly the last paragraph of the section 3.4 was altered as follows:

Laboratory tests and intercomparisons are fundamental methods for establishing and confirming uncertainty estimates of data products. Laboratory tests provide an opportunity to investigate in detail the performance of sensors under controlled conditions and to measure differences against certified references or other standards. Data from these experiments can be used to detect biases that may be corrected for and to determine calibration uncertainties.

Field intercomparisons allow multiple in situ sensors and remote sensing data to be directly compared under the actual atmospheric conditions of the required measurement, and include all of the complex environmental conditions (temperature, humidity, pressure, wind/flow rate, radiation, and chemical composition) that cannot be fully reproduced in the laboratory. These complementary activities increase confidence that measurements are subject to neither unanticipated effects nor undiscovered systematic uncertainties. Therefore field experiments are particularly useful for validating GRUAN data products.

Also, should there be a "GRUAN-Certified" seal on measurements that have passed standardized tests (that is to say, within the framework of GRUAN). Should there be specific "GRUAN-trained" operators?

The way measurements are made, ground tests, correction schemes and uncertainty estimates, including the results of lab and field experiments that were made for establishing and validating them, must be published and reviewed in accessible literature. This will be the major step for a data-product in order to get the "GRUAN-certified"

C1217

seal. Some revisions were made to sections 3.3 and 3.4 to work out this issues more clearly. In section 3.5 the following paragraph was introduced:

From the required steps to develop a GRUAN data product described in the previous section, detailed procedures result. For a data product to be considered established and successfully validated, it needs to be documented in detail including a description of method, algorithms, and the in-field procedures for ensuring and controlling data quality. Results from validation experiments should be published in the peer-reviewed literature or technical notes with a strong preference toward the former. Once this is accomplished, and the data have been shown to meet the requirements of GRUAN regarding accuracy, operability, and stability the product can be considered for operational use at GRUAN sites and suitable for scientific applications. The description of the method and the measurement procedures will constitute an essential part of the GRUAN regulatory materials and procedures.

Implementation of a GRUAN data product at a site involves the installation of the required equipment and training of operators. Essential requirements for GRUAN operations are:

- pre-deployment recalibration to the GRUAN site working standard,
- the routine collection of all relevant meta-data for measurements (e.g reference values for recalibration, environmental conditions, etc.) and
- on-site quality assurance in general by consistency analysis of redundant measurements.

The latter may be provided by the data processing facility or the Lead Centre.

8. Radiosonde example First I would like to echo referee 1's voice in his comment posted on 4/16 on time-lag correction. However, I understand that the radiosonde example has a mainly illustrative purpose, and a quick mention of the time-lag correction could be simply added to the present text without altering the rest of the manuscript. This radiosonde example illustrates well, however, how far the authors are from their final objectives. For this reason I would suggest to add a last paragraph (before the conclusion) specifically dedicated to current outstanding/unresolved issues, and how these are going to be addressed in the future months/years.

C1218

Please, see the reply to referee #1 addressing this issue.

Interactive comment on Atmos. Meas. Tech. Discuss., 3, 1807, 2010.