Update on Task Team 3 : Measurement Scheduling and Related Activities

Tom Gardiner, Dave Whiteman

Reinout Boers, Tony Reale , Carl Mears, Howard Diamond

GRUAN ICM4, 6th March 2012

Contents

- Water vapour trends (DW)
- Uncertainty definitions
- Measurement scheduling
- Questions and issues

TT3 Scheduling - Objectives

- to develop defensible, quantifiable, scientifically-sound guidance for GRUAN sites on measurement schedules and associated site requirements, in order to meet all GRUAN objectives including :
  - climate trend detection
  - satellite calibration/validation
  - studies of local meso-scale processes and events
- main information sources are from peer-reviewed literature, GRUAN documentation, and currently unpublished studies of which the group is aware. Some limited new analyses where critical gaps exist, using existing data sets.

Recent work by Dave Whiteman on water vapour trends

- How long does it take to reveal a trend ?
- Using the statistical approach summarized in Weatherhead et. al., 1998

$$Y = \mu + \omega T + N$$

μ constant term

ω trend

T time (months)

N noise

$$n^* \approx \left[\frac{3.3\,\sigma_N}{|\omega_0|} \sqrt{\frac{1+\phi}{1-\phi}} \right]^{2/3}$$

n^* the number of years

ω_0 trend magnitude

σ_N standard deviation

ϕ autocorrelation

- Given the difficulty of measuring water vapor accurately, the challenge in evaluating these formulas for water vapor is determining the autocorrelation and standard deviation of the noise
- Autocorrelation and standard deviation of noise must be separately determined for UT and LS to make statements about these two vary differing regimes
- Anticipated trends obtained from GEOS5 model predictions

What we have done to determine these values...

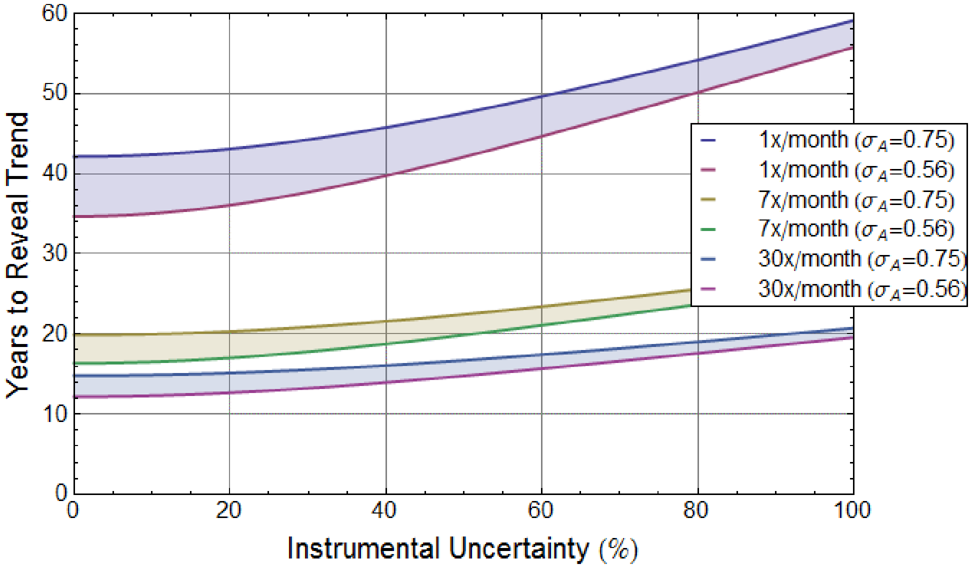
- **Upper troposphere** (published in JGR)
- Use radiosonde data from the ARM SGP site (Vaisala RS9X sensor) to calculate autocorrelation and standard deviation of noise
- Separate instrumental and atmospheric contributions to the noise in time series by making very conservative assumptions about the error characteristics of the Vaisala sensor
- **Lower Stratosphere** (Work in progress at preliminary stage)
- Consider time series of MLS data and compare autocorrelation and standard deviation of noise on a monthly basis to those obtained from Boulder FPH data record
- Develop confidence in MLS time series to provide reliable values for autocorrelation and noise and generate high temporal resolution time series of LS water vapor data from MLS.

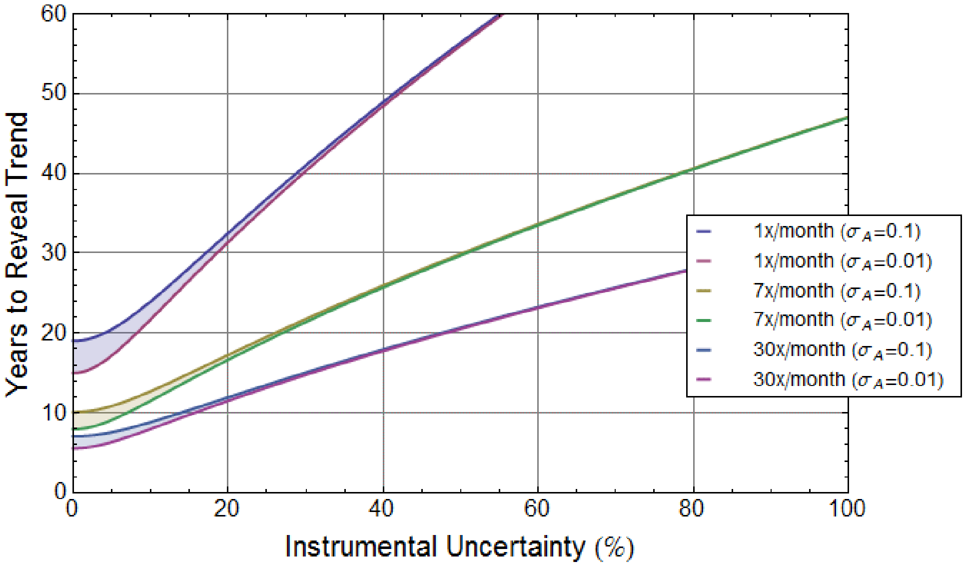
But…

- None of the work we are doing addresses gaps, drifts, jumps in time series.
- Needs to be done but ... (How many ways can you spell $$$??)
- Because of sensitivity of these results to assumed trend, we need to test robustness of results to assumed trend magnitude also (add another $)

Results - Trend Detection in UT vs LS (remember, LS results preliminary!)

- Upper Troposphere
  - Trend used was 1%/yr
  - High atmospheric variability
  - Low sensitivity to random error in measurement
- Lower Stratosphere
  - Trend used was 0.4%/yr
  - Low atmospheric variability
  - High sensitivity to random error in measurement





Results – UT Water Vapor Trend Detection

- High natural variability in UT water vapor implies trend detection in UT relatively insensitive to random errors in measurements
- High random error can hide small systematic errors so procedures used should attempt to randomized known sources of systematic errors
- e.g. frequent instrument recalibrations may be a good idea, not a bad idea
- To decrease time to detect trend in the UT, it is much more efficient to increase the frequency of measurement than reduce the noise of measurement.
- Quality profile extending into UT every 3-4 days a good compromise between efficiency of detection and level of effort.

Whiteman, D. N., K. C. Vermeesch, L. D. Oman, and E. C. Weatherhead (2011), The relative importance of random error and observation frequency in detecting trends in upper tropospheric water vapor, J. Geophys. Res., 116, D21118, doi:10.1029/2011JD016610

Results – LS Water Vapor Trend Detection (Preliminary!)

- Much lower natural variability implies need for much lower noise measurements to reveal trends accurately
- Increase in measurement frequency still pays large benefit to decreasing time to detect trend (but $$$ ...)
- A related note concerning trend detection in the LS
- Recent work (Hurst et al., 2011) shows that differing trends can exist in the LS in 2 km thick layers and that such studies are of value to understanding LS processes.
- This supports requirement that vertical resolution of <~ 1km optimum for LS water vapor time series
- Calculations of trends in 2km layers in LS from MLS data would be complicated by the instrument resolution.

Conclusions (preliminary)

- Given highly quality measurements in the LS, trends can be revealed there about as efficiently with 1 measurement per month as with 7 quality measurements per month in the UT.
- Total error of less than 10% desired in the LS.

Updates on other GRUAN related activities

- MOHAVE 2009 Analysis
  - Wet bias developed in Raman lidar due to deposition of biological material on receiver during campaign
  - Correction technique developed and applied consistent with the GUM
  - "It is assumed that the result of a measurement has been corrected for all recognized significant systematic effects and that every effort has been made to identify such effects."
- Recommendations for Raman Lidar
  - Window covering system can prevent accumulation of biological material on lidar optics. Window should be washed regularly
  - Data analysis should include checks for the existence of biases. Corrections should be applied when the biases are significant, with uncertainty of the correction accounted for, consistent with recommendations of the GUM
- *Whiteman et al., "Correction technique for raman water vapor lidar signal dependent bias and suitability for water vapor trend monitoring in the upper troposphere." Under discussion at AMTD.*

What is uncertainty?

Metrology definition

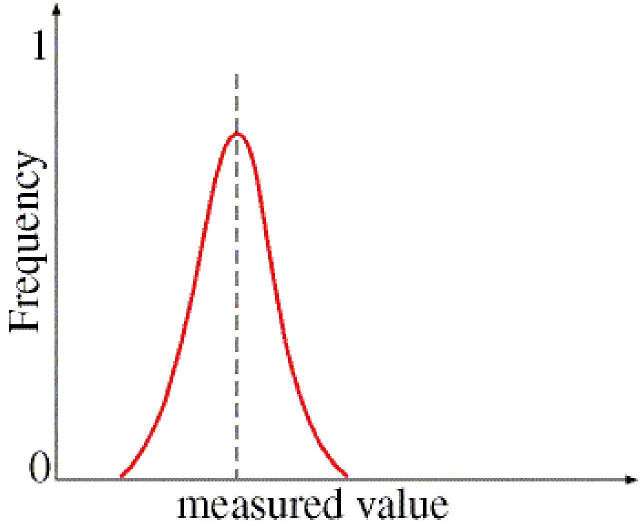
‘Parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand’

From which we can conclude:

- Uncertainty is a topic which seems to attract the most obscure and convoluted definitions
- Uncertainty is a **property of a result**
- Indicates the likely range within which we think the ‘true’ value of a measured quantity lies, **given all the information we have**
- Measurement uncertainty is a single value, expressed in terms of the measurand, either as a percentage or in units or the measurement

$x \pm U$

(with a given confidence interval defined by a coverage factor, k)



What isn’t it

- Mistakes

  Uncertainty doesn’t (can’t) cover mistakes and missing information

- The error in the result

  An error is the difference between a result and the true answer – we don’t (can’t) know what the ‘true’ answer is

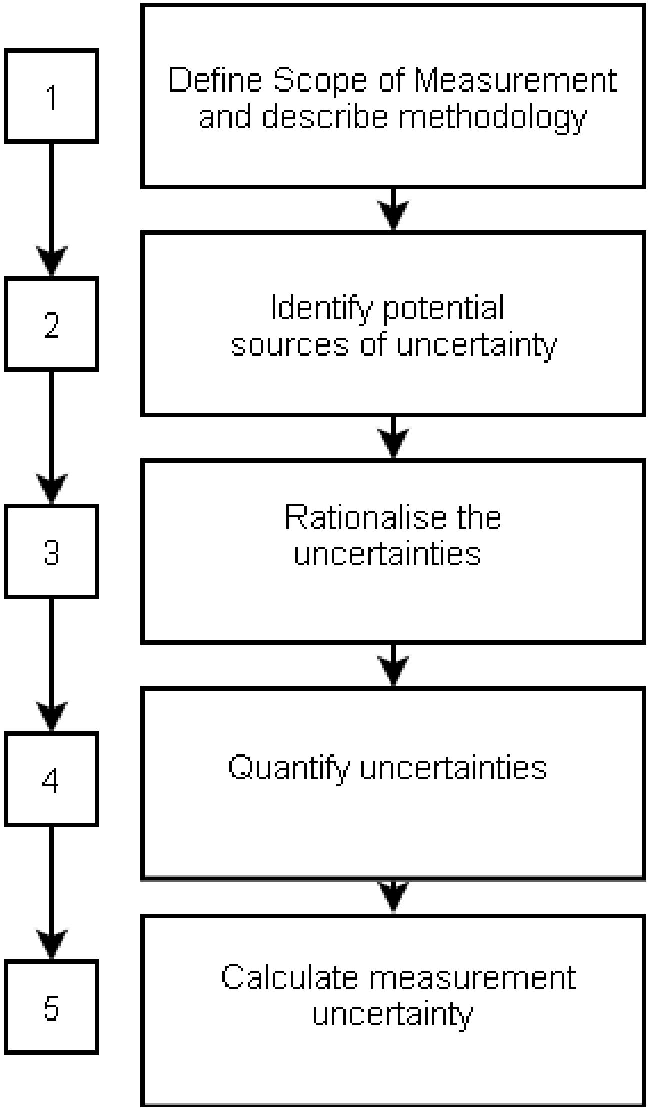
  Better to think of measurement uncertainty as a figure of merit, an indication of what values the true answer might have

- An absolute fact

  It is an estimate, at best we are saying that 95 times out of a 100 the result is (probably) within our uncertainty bounds.

Guide to Uncertainty in Measurement (GUM)

- GUM has been adopted in Metrology as an overarching methodology
- Approach can be summarised as
  - Describe measurement steps
  - Identify uncertainties associated with these and all inputs
  - Combine them
  - Assign known level of confidence to this uncertainty



1 Define Scope of Measurement and describe methodology

2 Identify potential sources of uncertainty

3 Rationalise the uncertainties

4 Quantify uncertainties

5 Calculate measurement uncertainty

GUM approach to determining uncertainty

- Define the measurement process

  In principle we should know the 'measurement equation'

  $$Y = f(X_1, X_2..X_N)$$

- Quantify uncertainties of each X_i these as standard uncertainties (in units of measurand)

  By statistical assessment (repeated measurement) - Type A

  By other methods (e.g. estimation) - Type B

  Insignificant contributions may be ignored

- Combine these

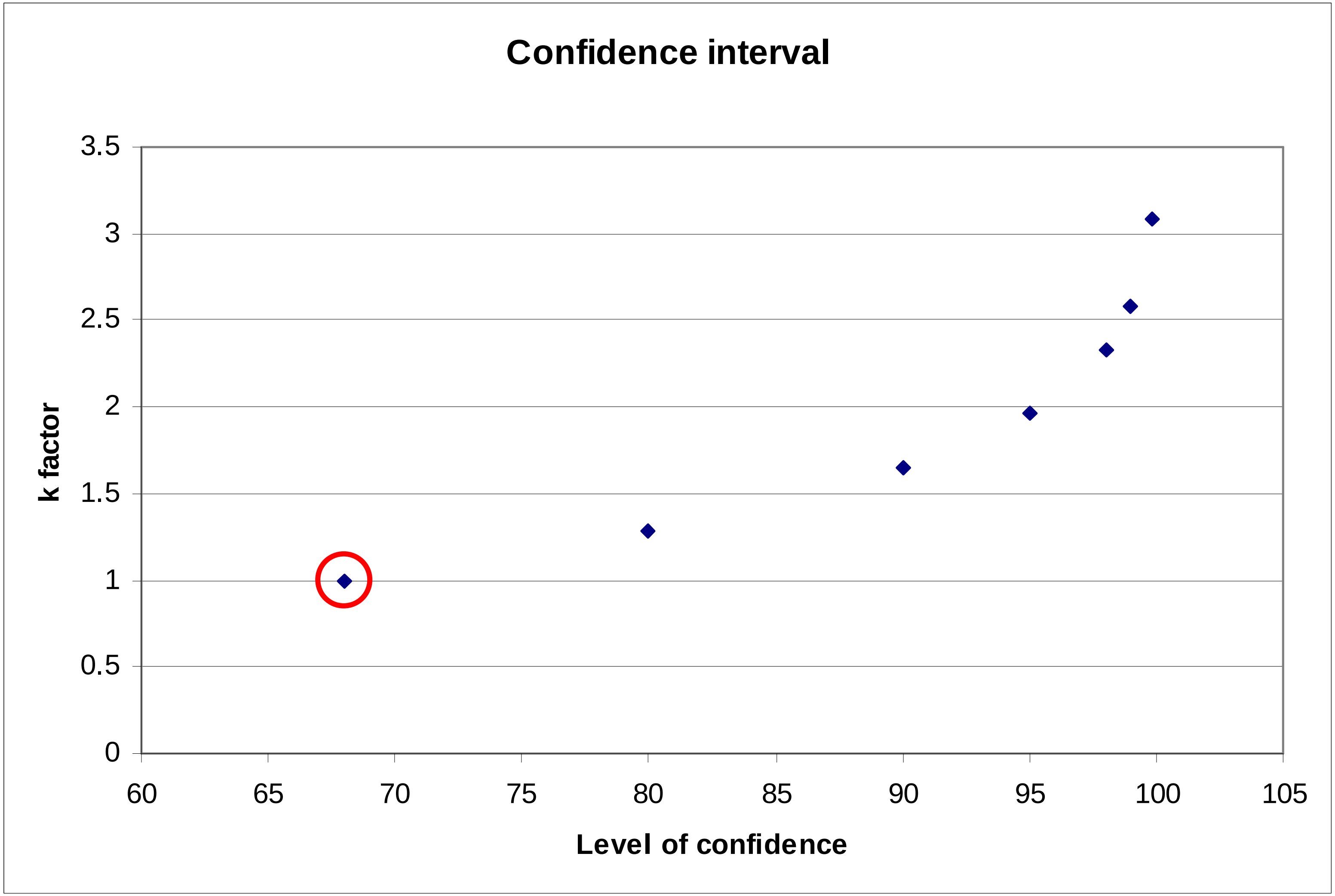
  For random (normally distributed) terms this is the standard deviation of the set of repeated results.

  For quantities which we believe lie within a range, but with equal probability of being anywhere in that range (often things like drift or certain bias corrections). This is a rectangular distribution, width R (the equivalent standard deviation is R/1.732).

- Expand the combined uncertainty to give an estimate of the uncertainty with a required level of confidence by multiplying by a coverage factor (k)

Relationship between coverage factor and level of confidence

National Physical Laboratory



Repeatability in atmospheric measurements

- One key issue in atmospheric measurements is that in general we can’t make repeated measurements of the measurand.
- Need to consider what it is we are actually measuring and then consider the contributions to the uncertainty in this parameter.
- The output parameter which we want the uncertainty of will vary with the application for the data.
- For example, if we want monthly means for trend assessment, then we can’t just look at the scatter of individual results as a measure of the uncertainty in the monthly measurement, or simply treat as a average of independent measurements.
- Don’t confuse variability of the measurand (the atmosphere) with uncertainty of the measurement.
- Need to determine (estimate) the natural variability to separate it from the experiment uncertainty.
- Also need to determine which factors in the overall uncertainty would reduce through averaging (generally Type A) and which would not (generally Type B).

Randomising uncertainties

- Always define the scope of the measurement that you are determining the uncertainty of
- What may appear as a systematic term in one context may be a random term in another (and vice versa)
- For example over a year the use of different calibrations will randomise some uncertainties.
- If you can randomise a systematic term then it can be reduced through averaging (e.g. use multiple independent calibration artefacts)

Terminology, e.g. Reproducibility rather than Stability

- **Reproducibility (of results of measurements)**
- closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement
- NOTE 1 A valid statement of reproducibility requires specification of the conditions changed.
- NOTE 2 The changed conditions may include:
  — principle of measurement
  — method of measurement
  — observer
  — measuring instrument
  — reference standard
  — location
  — conditions of use
  — time.
- NOTE 3 Reproducibility may be expressed quantitatively in terms of the dispersion characteristics of the results.
- NOTE 4 Results are here usually understood to be corrected results.

Measurement scheduling and GRUAN objectives

- **Trend detection**
- **Satellite validation**
- **Process studies**

- Need to consider at sampling requirements for these very different scientific objectives.
- Have started to do this (focus on water vapour), but more work needed.
- Will depend on individual site locations and site team objectives, so have to focus on generic issues and guidance.
- More details, including examples of actual measurement requirements, are given in the Guide but a few key issues discussed on the following slides.

Scheduling for Trend Detection

- The usual measurement unit is monthly mean value.
- Key uncertainty parameter is long term reproducibility.
- Statistical (Type A) uncertainties and location/timing uncertainties less important.
- There are a number of factors that should guide sampling requirements for trend detection
  - The variability of the measurements, both natural and instrumental.
  - The autocorrelation in the measurements, both day-to-day and month to month.
  - The measurement uncertainty, although this is often not included explicitly in the trend analysis.
  - The magnitude of the likely trend to be measured.
  - The dependence of all of the above on species, altitude (UT vs LS), location and season.

Study of previous datasets, model outputs, etc. will help in determining (estimating) these values for different sites.

Satellite validation

- The measurement unit may be the profile of the ECV or be in radiance space.
- Both datasets mapped onto the same vertical distribution (vertical resolution / averaging kernels).
- Should consider how the profile measurement uncertainty is affected by this conversion and/or mapping.
- The single profile uncertainty used for individual comparisons, but again may need to separate influence of Type A (statistical) and Type B uncertainties in combined datasets.
- Co-location uncertainty, driving by local spatial and temporal variability (and balloon drift), is key in determining the co-incidence requirements and needs to be included in overall uncertainties. Measurement footprint may also be an issue.
- As GRUAN aims to provide validation continuity for the long-term satellite record, long-term reproducibility is as important here as for trend detection.

Combining Measurements for Satellite Validation

- There will be "sweet times" when combining measurements into a single data product is optimal, for example, during sonde ascent co-incident with overpass.
- At such times a ‘site atmospheric state best estimate (SASBE)' can be routinely processed into a single data product through a combination of ancillary and sonde data. This can be a routine, quasi real-time output from each site that can be directly utilized in satellite product validation.
- SASBE would require that a minimal combination of measurement products be available.
- This could even be useful to monitor site "performance".

Process studies

- Requirements heavily dependant on process being studied, so difficult to come up with generic guidance.
- Typically looking at short term processes, so measurement unit will usually be individual profiles.
- High frequency sampling often required, so rapid ancillary data may be more important that sonde data.
- Long-term reproducibility less important than low statistical uncertainty.
- Will often be combining with other (potentially non-GRUAN) data products e.g. turbulence measurements, boundary layer height, heat flux etc.
- Co-location uncertainties may be an issue.

Discussion points

- Team Membership
  - Team is small and struggling to deliver in this important area. Would be good to extend team membership including site representation.
- Uncertainty Definitions
  - Need to try and ensure common terminology through GRUAN documentation, particularly with regards to uncertainties. What basis should we use for this (the GUM) ?
  - How to deal with the issue of combined uncertainties for different applications ?
- Instrument Specific Scheduling
  - There are instrument specific issues, but where should these be dealt with ? (GRUAN Operations Guide or technical documents)
- Future activities. Initial focus will be on the Manual and Guide, but given limited resource, what are the priorities for the Task Team next year ?
  - Complete similar review to water vapour for temperature and/or priority 2 variables.
  - Assess benefit of weighting sampling according to natural variability.
  - Focus on one of the application areas, and if so which would be of most use to the sites ?

Formal definition in GUM

Expression of experimental uncertainties

- 1) The uncertainty in the result of a measurement generally consists of several components which may be grouped into two categories according to the way in which their numerical value is estimated:

  A. those which are evaluated by statistical methods,

  B. those which are evaluated by other means.

- There is not always a simple correspondence between the classification into categories A or B and the previously used classification into “random” and “systematic” uncertainties. The term “systematic uncertainty” can be misleading and should be avoided.
- Any detailed report of the uncertainty should consist of a complete list of the components, specifying for each the method used to obtain its numerical value.

Formal definition in GUM – cont.

National Physical Laboratory

Expression of experimental uncertainties

- 2) The components in category A are characterized by the estimated variances (or the estimated “standard deviations”) and the number of degrees of freedom. Where appropriate, the covariances should be given.
- 3) The components in category B should be characterized by quantities which may be considered as approximations to the corresponding variances (or standard deviations), the existence of which is assumed. Where appropriate, the covariances should be treated in a similar way.
- 4) The combined uncertainty should be characterized by the numerical value obtained by applying the usual method for the combination of variances. The combined uncertainty and its components should be expressed in the form of “standard deviations”.
- 5) If, for particular applications, it is necessary to multiply the combined uncertainty by a factor to obtain an overall uncertainty, the multiplying factor used must always be stated.