############### Internal Documentation of Details of uncertainty calculations

draft: 2026-04-09

Follows on documentation Kirk wrote before retiring.

1. Introduction

One of the primary responsibilities of the CCL is to disseminate the WMO scales consistently over time and with small enough uncertainty to prevent significant biases from developing between various monitoring sites and programs. This scale transfer uncertainty is separate from the total uncertainty of the scale and relates to how well an individual tertiary standard measurement and its subsequent value assignment is related to the defined scale. This is the relevant measure of uncertainty to use when propagating uncertainties of standards to atmospheric measurements and comparing to other measurements traceable to the same WMO scale. It is generally not the same as the SI traceable uncertainty, which is relevant when comparing measurements traceable to different scales. One exception to this is CO where due to the complexity of maintaining the CO scale there is no distinction between the total and scale transfer uncertainty.

Historically, the CCL provided rough guidance on its ability to disseminate the scales (via calibrated whole air standards). This was done as a general analytical reproducibility term based on the results from many cylinders over long time periods rather than as values specific to any individual cylinder. This type of guidance was important for the community when trying to put general limits on the ability to distinguish small spatial gradients. However, it was not responsive to changes in analytical performance at the CCL. GML has undertaken a review of how it calculates and how it disseminates information on the quality of its calibration services to better address these needs. We present here an overview of the strategy implemented to calculate the scale transfer uncertainty of individual calibration episodes and value assignments determined from multiple episodes.

2. CCL calibration strategy

Calibrations at the CCL follow a strict hierarchy where secondary standards are value assigned by calibrations verses the suite of primary standards and tertiary standards are value assigned by calibrations verses a suite of secondary standards that cover the nominal range of the defined scales. CCL calibrations are done on measurement systems that are dedicated to calibrating whole air standards for the community.

All current measurement systems use multi-point calibration curves (referred to as response curves) generated during dedicated instrument calibration episodes. Each member of the suit of standards is measured multiple times (typically 6-10) alternating with a reference cylinder

(REF-STD-REF-STD-...) used to track analyzer drift during the response curve episode and between the response curve and sample analysis episode.

Mole fractions are calculated from a polynomial instrument response curve, depicting mole fraction (y axis) as a function of the analyzer output (x axis), for example:

$$Mf = C_0 + C_1*x + C_2*x^2 \quad (1)$$

An ODR fit is made to the data to determine the response curve where:

- y = value assignment of the standard on the analysis date
- y weights = 1 / (uncertainty of assigned value)^2

- x value = analyzer response to the standard
- x weights = 1 / (uncertainty of analyzer response)^2

The uncertainty of the assigned value and the uncertainty of the analyzer response are discussed below.

From the odr fit we save *(in the reftank.response database table)*

- Coefficients of the polynomial (C_0, C_1, C_2)
- Residual standard deviation
- Covariance matrix

Response curves are typically in service for 2-4 weeks until the next scheduled instrument calibration episode generates a new response curve. The same reference cylinder used in determining the response curve is also used during sample calibration episodes to account for drift in the analyzer occurring between the instrument calibration and sample measurement episodes.

3. Uncertainty of calibration episode

We define the uncertainty of a calibration episode ($u_{episode}$) as the combined standard uncertainty of three terms

$$u_{episode} = \sqrt{u_{meas}^2 + u_{reproducibility}^2 + u_{typeB}^2} \qquad (4)$$

Where u_{meas} is the uncertainty calculated from data collected during the measurement, $u_{reprocucibility}$ is based on the long term performance of the analytical system over many years, and u_{typeB} are any additional terms that may be required.

Note, for measurements where a calculated measurement uncertainty term is not available (typically data from the older measurement systems) we use the standard deviation of the episode mean in place of u_{meas} in equation 4.

3.1. Measurement uncertainty

The first component, u_{meas}, referred to here as "measurement uncertainty", is calculated using the short-term analytical repeatability observed within the measurement episode and the uncertainty of the calibration curve used to convert instrument response to mole fraction. The calibration curve uncertainty is taken as the predictive interval of the ODR fit and incorporates the uncertainty of the value assignment of higher level standards (typically secondary standards) relative to the scale along with the analytical performance during the calibration episode. Measurement uncertainty responds to analytical problems over short time scales (days to weeks) and to longer term issues related to the ability to consistently value assign the higher level standards. We store measurement uncertainty with the episode mean results in the database and it serves as a useful diagnostic of analytical performance over time.

In equation 1 above, ‘x’ is actually the sample/reference ratio or the sample - reference difference depending on the dominate drift characteristics of the analyzer. We refer to both as the 'normalized instrument response' for simplicity.

The reference value (Ref) used in both cases is the average of the previous reference measurement and the next reference measurement bracketing the sample aliquot.

$$Ref = (Rp + Rn) / 2 \qquad (5)$$

Where

Rp = previous reference measurement before the sample
Rn = next reference measurement after the sample

The normalized instrument response, termed 'R', is then

For ratio: $R = Smp / Ref$ (6a)
For difference: $R = Smp - Ref$ (6b)

where Smp is the instrument response to the sample aliquot and Ref is the average reference measurement from equation 5.

Using propagation of error, the uncertainty on the average reference is

$$\sigma_{Ref} = \sqrt{\sigma_{Rn}^2 + \sigma_{Rp}^2} \qquad (7)$$

The uncertainty on the ratio (6a) or difference (6b) is then

$$\sigma_R = R * \sqrt{\left(\frac{\sigma_{Smp}}{Smp}\right)^2 + \left(\frac{\sigma_{Ref}}{Ref}\right)^2} \quad (8a)$$

$$\sigma_R = \sqrt{\sigma_{Smp}^2 + \sigma_{Ref}^2} \quad (8b)$$

We then have the normalized instrument response as

$$R \pm \sigma_R$$

In these equations, we'll use for σ the *experimental standard deviation of the mean, as*suming that the more measurements we make of the analyzer output, the better we will know the average value. This means that σ is the *experimental standard deviation* divided by the square root of n. (See for example, https://physics.nist.gov/cuu/Uncertainty/typea.html and section 4.2.3 of the GUM (2008))

The mole fraction of the sample aliquot is

$$Mf = C_0 + C_1 * R + C_2 * R^2 \quad (9)$$

The prediction interval of the response curve is used to determine the uncertainty of the calibration curve that is used to determine the mole fraction.

$$Var(\hat{y}) = rsd^2 + d^T \cdot Cov \cdot d \quad (10)$$

Where rsd is the standard deviation of the response curve residuals, Cov is the covariance matrix, and d = array of partial derivatives of polynomial with respect to the coefficients evaluated at x, i.e.

$$d = \left[\frac{\partial y}{\partial C_0}(x) + \frac{\partial y}{\partial C_1}(x) + \frac{\partial y}{\partial C_2}(x)\right] \quad (11)$$

(For a polynomial response curve, d = [1, x, x^2]).

So the 'uncertainty from the response curve' of a measured sample, is

$$\mu_{curve} = \sqrt{rsd^2 + d^T(x) \cdot Cov \cdot d(x)} \quad (12)$$

Where

x = measured analyzer output (normalized instrument response R)
rsd = response curve residual standard deviation

Use equation 12 to get the 'uncertainty from the response curve', μ_{curve} , using x = R.

Use the coefficients of the response curve (ignoring the intercept term) to convert the uncertainty of the normalized instrument response (σ_R) to mole fractions

$$\mu_R = C1 * \sigma_R + C2 * \sigma_R^2 \qquad (13)$$

The uncertainty on the sample aliquot is then

$$\mu = \sqrt{\mu_{curve}^2 + \mu_R^2} \qquad (14)$$

Mole fractions and uncertainty values are calculated for each aliquot of the episode. The mean mole fraction is determined and the uncertainty of the average is determined by the combined variance.

$$S = \sqrt{\frac{\sum_{i=1}^{n} n_i\left(s_i^2 + (x_i - \overline{x})^2\right)}{\sum_{i=1}^{n} n_i}} \qquad (15)$$

Where n_i is number of observations for the i^{th} measurement (usually 1), s_i is the standard deviation of i^{th} measurement (from above formula for calculating individual measurement uncertainty for each aliquot), x_i is the average value of i^{th} measurement, $\overline{x}$ is the overall average.

3.1.1. Example of measurement uncertainty calculated for a discrete air measurement which uses a similar process:

Use co2 measurement of event number 522901 analyzed on magicc-3 on 2023-09-13.

N = 10 (10 measurements of the analyzer output for each sample)

Rp = 409.0706 σ_{Rp} = 0.0388 σRp = 0.01227
Rn = 409.0575 σ_{Rn} = 0.0479 σRn = 0.01515

Smp = 415.3468 σ_{smp} = 0.0584 σsmp = 0.01847

Ref = 409.06405 σ_{ref} = 0.01949

R = 1.01536 σ_R = 0.000066176

The coefficients of the response curve are

C0 = -0.151832695463
C1 = 411.751633323,
C2 = 0
Using equation 9 we get

Mf = 417.924

The uncertainty of the response curve at the value of R is

μ_{curve} = 0.01894

And converting σ_R to mole fraction using equation 13 we get

μ_R = 411.751633323 * 0.000066176 = 0.02725

Therefore, using equation 14 we get

μ = 0.03318

3.2. Reproducibility

The long term reproducibility of the scale transfer is estimated by the variations observed in repeated analysis of target tanks over multiple years. We estimate the long term reproducibility (at 68% CL) for each instrument used for CCL calibrations. Values are stored in a lookup table and are matched with results for each calibration episode when retrieving results from the database.

We are limited in the number of target tanks and in the mole fraction range covered for the earlier calibration systems. The long term reproducibility estimates for older systems/instruments are generally applicable to most of the range of the scales, but they may be underestimated for cylinders on the extremes of the scale. Measurements made on an extension of the WMO scale are provided on a case by case basis. We extrapolate as a ratio of mole fraction for these measurements to give an approximation of the reproducibility.

3.3. TypeB uncertainty

Type B uncertainty terms (u_{typeB}) are not related to quantities measured during calibration episodes. Instead, they include uncertainties arising from other sources, such as known bias corrections. Values for these terms are stored in a lookup table and paired with measurement results upon data extraction from the database. Primarily this table is used to define uncertainty terms which are related to other GML measurement programs. For example, uncertainty estimated associated with discrete air storage corrections or sample collection biases have terms listed in this table. For CCL calibrations, currently only one entry is applicable. A repeatability term is included for single aliquot tank calibration episodes for CO measured on instrument "V1" from 2004-2006. However, uncertainty associated with any future bias corrections can be handled with this table so we include it in this description.

4. Value assignments of CCL standards

CCL value assignments for standards are based on multiple calibration episodes. Typically 3 episodes initially with recalibrations over the lifetime of the cylinder following guidance in the GGMT recommendations. Internally, the CCL uses custom software to fit the history of calibrations, weighted by the scale transfer uncertainty of each episode, and use a statistical determination of changes over time to assign a constant or time variant value (and uncertainty) for standards.

Value assignments of standards are stored in a database table (reftank.scale_assignments) as the coefficients (and coefficient uncertainties) of a 2nd order polynomial which give the value and uncertainty as a function of time. For stable or linearly drifting tanks, higher order coefficients are set to 0. Time (as decimal year) in the polynomial equation is relative to a "time zero" (T_{zero}) parameter stored with the coefficients. T_{zero} is set to the weighted mean date of the calibration history as a convenience to allow the uncertainty of the time dependent value assignment to be approximated without considering correlations between the coefficients.

The value and uncertainty of a standard on date D_a would be

$$Val(D_a) = coef0 + coef1*dt + coef2*dt^2$$

$$\mu(D_a) = \sqrt{(unc\_c0)^2 + (unc\_c1 * dt)^2 + (unc\_c2 * dt^2)^2 + (sd\_resid)^2}$$

Where $dt = D_a - T_{zero}$ (both as a decimal year notation).

Note to internal GML users: The database table Reftank.scale_assignments is "insert only" to preserve the history of value assignments and allow the CCL to document changes. This means that there can be multiple entries for individual standards. The record with the latest "assignment_date" is considered the valid assignment. For routine use, the reftank.scale_assignments_view includes additional information to make retrieving value assignments easier. It includes a bit field "current_assignment" to clearly indicate which assignment is the current valid assignment.

5. Uncertainty related to drift in cylinders

The CCL typically calibrates tertiary level standards for external partner labs 3 times over the course of 4-6 weeks for an initial calibration. A value assignment and certificate are generated using this data. However, this short time is not sufficient for determining the stability of a standard. Re-calibrations months to years later are required for this and are highly recommended.

6. Annex - Statistical test for drift

We use a statistical test to determine if replicate measurements of a cylinder over time indicate a time dependent change in mole fraction. If changes are significant within measurement uncertainty a time dependent value assignment is made for the standard. The time dependent assignments can be linear or non-linear polynomial.

Steps used in statistical evaluation of cylinder stability

1. Get mole fraction and scale transfer uncertainty for each calibration episode. Data is restricted to official calibration measurement systems and any flagged episodes are excluded.
2. For each calibration episode, determine the weight on the y axis data as $1/(unc^2)$
3. Calculate the central date as the weighted average of the dates for the calibrations
4. For each calibration date, compute the deviation from the central date
5. Start with quadratic polynomial.
6. Compute polynomial fit to the weighted data (using python scipy.curve_fit, which is equivalent to idl svdfit)
   x = deviation from central date in years,
   y = mole fraction weighted by scale transfer uncertainty
7. Degrees of freedom = (number of cals – degree of fit)
8. t = probability of two-tailed t-distribution for degrees of freedom, 95% confidence interval
9. t* = coefficient / stdv of coefficient
10. If |t*| > t then assume coefficient != 0. Exit and use polynomial fit coefficients
11. If |t*| < t then assume coefficient == 0.
12. Reduce degree of polynomial fit by 1, go to step 6
13. Repeat steps 6-11 until |t*| > t or degree of polynomial = 0 (mean)

The output of this code includes coefficients of a 2nd order polynomial which give the value as a function of time. For stable or linearly drifting tanks, higher order coefficients are set to 0.

Time (as decimal year) in the polynomial equation is relative to a stated "time zero" (T_{zero}) parameter stored with the coefficients. T_{zero} is set to the weighted mean date of the calibration history as a convenience to allow the uncertainty of the time dependent value assignment to be approximated without considering correlations between the coefficients. Uncertainties are taken as the 68% predictive interval of the fits.

Indications of non-linear drift using this process are not the same as assessing the appropriateness of the simple 2nd order polynomial model. It is an indication of a non-linear correlation with time. Indications of non-linear drift should be evaluated carefully since the simple polynomial model displayed may not represent actual tank drift well. Other functions or piecewise fitting may often represent the drift better. Users are encouraged to do their own assessments in evaluating the output from this tool