Data Processing / Data Management

GML Technical Procedure

Contents

Approved by

VIMONT.ISAAC.JOSEF.1512650232 Digitally signed by VIMONT.ISAAC.JOSEF.1512650232 Date: 2026.06.16 15:18:09 -06'00'

Isaac Vimont Date
CCL QMS Lead
Long-term Observation of Greenhouse Gases and Ozone-depleting Substances Division
NOAA Global Monitoring Laboratory

1. Purpose

The World Meteorological Organization (WMO) Central Calibration Laboratory (CCL) is tasked with defining and disseminating a common reference scale to ensure comparable data across contributing monitoring programs. The CCL defines reference scales by value assigning sets of primary standards using SI traceable methods. Gravimetry for CH_4, CO, N_2O, and SF_6 (Hall et. al, 2007, Hall et. al, 2011) and manometry for CO_2 (Hall et. al, 2021). The sets of primary standards are typically harmonized through intercomparison measurements to establish the scale

One of the primary responsibilities of the CCL is to disseminate the World Meteorological Organization (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, described in publications and supplemental documentation, 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.

This document provides the technical procedures for mole fraction and scale transfer uncertainty calculations for CCL analysis systems and includes an overview of the CCL data management strategy.

2. Scope

NOAA/GML provides compressed Reference Materials (as gas standards) to the World Meteorological Organization/Global Atmosphere Watch (WMO/GAW) community. Natural air or modified natural air gas standards are analyzed, relative to the WMO mole fraction scales. This technical procedure (TP) describes calibration strategies and calculations used to determine mole fractions and uncertainty relative to the scales.

The procedures described here pertain to analysis for which a certificate of analysis is issued, and are also widely used for other GML measurement programs.

3. Informative References

Hall, B.D., G.S. Dutton, and J.W. Elkins (2007), The NOAA nitrous oxide standard scale for atmospheric observations*, J. Geophys. Res*., 112, D09305, doi:10.1029/2006JD007954.

Hall, B.D., G.S. Dutton, D.J. Mondeel, J. D. Nance, M. Rigby, J.H. Butler1, F.L. Moore, D.F. Hurst, and J.W. Elkins (2011), Improving measurements of SF_6 for the study of atmospheric transport and emissions, *Atmos. Meas. Tech.*, doi:10.5194/amt-4-2441-2011.

Hall, B. D., Crotwell, A. M., Kitzis, D. R., Mefford, T., Miller, B. R., Schibig, M. F., and Tans, P. P. (2021), Revision of the World Meteorological Organization Global Atmosphere Watch (WMO/GAW) CO2 calibration scale, Atmos. Meas. Tech., 14, 3015-3032, https://doi.org/10.5194/amt-14-3015-2021.

JCGM (2008), International vocabulary of metrology — Basic and general concepts and associated terms (VIM), JCGM 200:2008.

JCGM 100:2008 Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement (ISO GUM 1995 with minor corrections), Joint Committee for Guides in Metrology (2008); http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

Tans, P. P., Crotwell, A. M., and Thoning, K. W. (2017), Abundances of isotopologues and calibration of CO_2 greenhouse gas measurements, Atmos. Meas. Tech., 10, 2669-2685, doi:10.5194/amt-10-2669-2017.

Barry N. Taylor and Chris E. Kuyatt (2001). *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, [Online]. Available: http://physics.nist.gov/TN1297 [access date *2026, 05 01*]. National Institute of Standards and Technology, Gaithersburg, MD.

Originally published as *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, Barry N. Taylor and Chris E. Kuyatt, NIST Technical Note 1297 (1994 Edition).

Also see, https://physics.nist.gov/cuu/Uncertainty/index.html

4. Terms and Definitions

- **Analysis System**: Includes the laser spectroscopic instruments, associated hardware, and computer used to analyze trace gases in compressed gas cylinders (synonymous with measuring system**)**.

- **Gas Standard**: A cylinder of compressed gas with mole fractions assigned by metrological methods or by comparison to higher-level standards, used to characterize the response of an instrument for calibration or quality control purposes. For the purposes of this TP, primary, secondary, and tertiary standards

are gas standards.

- **Mole Fraction**: The ratio of the number of moles of analyte to the total number of moles. Dry air mole fraction is the ratio of the number of moles of analyte to the total number of moles in dry air. Within the scope of this TP, all samples are analyzed for dry air mole fraction.

- **Primary Standard**: A measurement standard established using a primary reference measurement procedure, or created as an artifact, chosen by convention. Primary standards are typically aluminum cylinders containing dry natural air or modified natural air. The CO_2 mole fraction has been determined by manometric determinations (see TP_primary_manometer), while other species are produced using gravimetric techniques (see TP_Gravimetry).

- **Reference Cylinder**: Cylinder of dry, natural air with near-ambient trace gas mole fractions used to normalize variations in analyzer response during analysis sequences.

- **Reference Material**: A material, sufficiently homogeneous and stable with respect to one or more specified properties, which has been established to be fit for its intended use in a measurement process. (JCGM 200:2008, 3.6)

- **Response Curve**: A function that relates the instrument response to the amount of substance (mole fraction).

- **Scale Transfer Uncertainty**: Estimated uncertainty of calibration episode or standard value assignments relative to the defined scale.

- **Secondary Standard**: A standard whose value is determined through analysis relative to primary standards, for a quantity of the same kind. These standards are used to calibrate the instrument response. Use of secondary standards for routine calibration prolongs the life of primary standards.

- **Target Tank**: A tertiary standard used for routine monitoring of system performance. The system should be capable of reproducing the assigned value of the target tank (within expected uncertainties).

- **Tertiary Standard:** A standard whose value is determined through analysis relative to secondary standards, for a quantity of the same kind.

- **WMO/GAW**: World Meteorological Organization, Global Atmosphere Watch.

5. CCL Calibration Strategy

Calibrations at the CCL follow a strict hierarchy where secondary standards are value assigned by calibrations versus the suite of primary standards which define the scale. Subsequently, tertiary standards are value assigned by calibrations versus 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 times), alternating with a reference cylinder (in an A-B-A-B-A sequence) used to track analyzer drift during the response curve episode, and between the response curve and sample analysis episode.

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 = the normalized instrument response to the standard (see description of normalized instrument response in section 6.1 below)
- x weights = 1 / (uncertainty of analyzer response)^2

The uncertainty of the assigned value and 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

Mole fractions for sample measurements are calculated from the response curve, which depicts mole fraction (y-axis) as a function of the analyzer output (x-axis), for example:

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

Where x is the analyzer output for the sample and C_0, C_1, and C_2 are coefficients of the polynomial fit to the response of the analyzer to the suite of standards. C_2 is set to 0 when a linear response curve is prescribed.

Response curves are typically in service for two to four 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.

6.0 Uncertainty of Calibration Episode

The scale transfer uncertainty of the calibration episode ($u_{episode}$) is the combined standard uncertainty of three terms described below:

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

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 2.

6.1 Measurement Uncertainty

The first component, u_{meas}, referred to here as, "measurement uncertainty," is calculated along with the mole fraction, by using both 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, the analyzer output, '*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.

$$Ref = (Rp + Rn) / 2 \quad (3)$$

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: $R = Smp / Ref$ (4a)
For difference: $R = Smp - Ref$ (4b)

Where *Smp* is the instrument response to the sample aliquot and *Ref* is the average reference measurement from Equation 3.

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

$$\sigma_{Ref} = \sqrt{\sigma_{Rn}^2 + \sigma_{Rp}^2} \qquad (5)$$

The uncertainty on the ratio (4a) or difference (4b) is then

$$\sigma_R = R * \sqrt{\left(\frac{\sigma_{Smp}}{Smp}\right)^2 + \left(\frac{\sigma_{Ref}}{Ref}\right)^2} \qquad (6a)$$

$$\sigma_R = \sqrt{\sigma_{Smp}^2 + \sigma_{Ref}^2} \qquad (6b)$$

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,* assuming 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

$$Mf = C_0 + C_1 * R + C_2 * R^2 \qquad (7)$$

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.

$$Var(\hat{y}) = rsd^2 + d^T \cdot Cov \cdot d \qquad (8)$$

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.

$$d = \left[\frac{\partial y}{\partial C_0}(x) + \frac{\partial y}{\partial C_1}(x) + \frac{\partial y}{\partial C_2}(x)\right] \qquad (9)$$

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

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

$$\mu_{curve} = \sqrt{rsd^2 + d^T(x) \cdot Cov \cdot d(x)} \qquad (10)$$

Where

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

Use Equation 10 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

$$\mu_R = C1 * \sigma_R + C2 * \sigma_R^2 \qquad (11)$$

The uncertainty on the sample aliquot is then

$$\mu = \sqrt{\mu_{curve}^2 + \mu_R^2} \qquad (12)$$

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 calculated as the combined variance.

$$S = \sqrt{\frac{\sum_{i=1}^{n} n_i(s_i^2 + (x_i - \overline{x})^2)}{\sum_{i=1}^{n} n_i}} \qquad (13)$$

Where n_i is number of observations for the ith 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.

See appendix 1 for an example calculation of measurement uncertainty.

6.2 Reproducibility

The long-term reproducibility of the analytical systems 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

(/ccg/flask/unc_master.conf) 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 expected reproducibility.

6.3 Type B 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 (/ccg/flask/unc_master.conf) 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, estimated uncertainty associated with discrete air storage corrections or sample collection biases have terms listed in this table. For CCL calibrations, Type B uncertainty terms are rarely used. However, uncertainty associated with any future bias corrections can be handled with this table so we include it in this description.

7.0 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 assess the stability of the standard using a statistical test (see section 8.0 below). Constant or time variant value (and uncertainty) assignments are made for each standard used by the CCL.

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

$$Val(D_a) = coef0 + coef1*dt + coef2*dt\text{^}2 \qquad (14)$$

$$\mu(D_a) = \sqrt{(unc\_c0)^2 + (unc\_c1 * dt)^2 + (unc\_c2 * dt^2)^2 + (sd\_resid)^2} \qquad (15)$$

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

Note: The database table Reftank.scale_assignments is "insert only" to preserve the history of all value assignments used 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 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. Similarly, in the reftank.response table, the 'current' column is set to 1 for the most recent/active response curve for a datetime, 0 otherwise. The reftank.calibrations table also records modification history, but stores it in a separate table. All processing software has been programmed to be aware of these settings.

8.0 Statistical Test for Drift

We use a statistical test to help 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 ($u_{episode}$). 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 $1/(u_{episode}^2)$
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 (and uncertainty) as a function of time (see equations 14 and 15). For stable or linearly drifting tanks, higher order coefficients are set to 0. Time (as decimal year) in the polynomial equation is relative to the 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.

In the specific case where $n=2$, the two tailed t-test is replaced by a simple comparison of the difference in the two results with the combined uncertainty of the difference as an expanded uncertainty term ($k=2$). If the difference is larger than the expanded uncertainty of the difference then caldrift.py fits a linear function to the two data points. Otherwise, the weighted mean is returned. Note that the weights used for the weighted mean are 1/unc^2.

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.

9. Data Reporting

Certificates issued by the CCL conform to relevant ISO guidelines. For a given trace gas, the assigned value and expanded total uncertainty, including traceability to the SI, are reported. In addition, the CCL reports scale transfer uncertainty separately. This is the expected uncertainty of the calibration results relative to the defined scale.

Values reported on certificates are derived as the weighted average from multiple calibration episodes (typically 3), with weights based on the calibration episode uncertainties relative to the scale. Values reported on certificates are limited to the date range of the calibration request. For example, an initial order for a tertiary standard will result in a certificate based on a weighted mean of the initial calibration episodes. When that cylinder is returned for re-calibration, any additional certificate issued by the CCL would only include the measurement episodes corresponding to the re-calibration request. In general, the CCL does not issue certificates for recalibrations unless specifically requested.

To meet the needs of the atmospheric monitoring community, the CCL provides on its website further information beyond what is included on the certificates. On the website, the CCL reports all measurement results for a given filling of a cylinder. This would include initial calibrations and all re-calibrations requested by the user. We report each calibration episode mean along with the scale transfer uncertainty of the episode. Additional information including the standard deviation of the mean, the measurement uncertainty term, and reproducibility (combined with any additional Type B terms and listed as *TypeB*), are available by downloading the associated json file. This information is provided to allow users to assess stability over time and value assign standards using protocols implemented within their own programs.

As a convenience for the community, the CCL also provides an assessment of the stability of standards if they have been re-calibrated. The CCL provides access to the output of a data analysis software package (caldrift.py) developed by GML which the CCL and GML more broadly use to value assign standards. This code uses the full calibration history of a standard, with the associated scale transfer uncertainties of each episode, to assess the cylinder for stability and suggest either a constant or a time dependent value and scale transfer uncertainty for the standard. Value assignments are represented as a 2nd order polynomial. The code outputs include:

T_{zero} - Time zero is the weighted mean date of the calibrations.
Coef0 (unc_co) - coefficient 0 of a 2nd order polynomial fit (and uncertainty of the coefficient).
Coef1 (unc_c1) - coefficient 1 of a 2nd order polynomial fit (and uncertainty of the coefficient). Equals 0 for stable cylinders.
Coef2 (unc_c2) - coefficient 2 of a 2nd order polynomial fit (and uncertainty of the coefficient). Equals 0 for stable or linearly drifting cylinders.
sd_resid - standard deviation of the residuals to the fit.

In cases where drift is indicated, the code output is included on the website for informational purposes. The value and uncertainty of the standard on date D_a (as a decimal year) are derived from equations 14 and 15 above.

The weighted mean output of this software package is used by the CCL to value assign cylinders for the issuance of certificates for calibration requests. Output from this software package on the website to assess the stability of standards is informational with the hope that it helps users quantify potential drift. However, users are encouraged to also do their own assessments. Indications of non-linear drift should be evaluated carefully since the simple polynomial model displayed may not represent actual tank drift well.

10. Code

The CCL uses custom software tools developed by GML to process data and extract results from the database. These tools are widely used across GML measurement systems for common data processing needs. These are described only briefly here since they have extensive help menus available by passing in the '--help' option from the command line call.

nlpro.py -

Program used to process a response curve raw file to determine a relationship between mole fraction of a set of standards and instrument responses using an ODR fit as described above. Output from nlpro.py is stored in DB table reftank.response.

Note that reprocessing response curves and updating reftank.response does not automatically get applied to flask and calibration data. For the new coefficients to be implemented flpro.py and calpro.py would need to be called separately after the curve coefficients are updated.

Current version: /ccg/bin/nlpro.py

calpro.py -

Code used to process tank calibration measurements on all CCGG systems and calculate the mole fraction for a tank calibration episode. This code has the ability to process all tank calibration measurements made by CCGG since 1979 by keying off various instruments, methods,& time periods for configuration information.

Current version: /ccg/bin/calpro.py

caldrift.py -

Procedure to value assign standards and test for drift. To standardize the value assignment of standards, code was developed to allow users to easily access calibration histories and determine the significance of differences in calibration results over time given our best understanding of the uncertainty of the measurements. This tool links calibration measurement results with TypeB uncertainty terms stored in the lookup table.

Full help menu is given by passing in the --help keyword. However, one notable item to mention here is the '--official' keyword. This limits data extracted to results from the instruments making official CCL calibration measurements. It excludes results from test instruments or cross comparisons with other analytical systems.

Current version: /ccg/bin/caldrift
Note, this is a shell script to setup the correct python environment and then call caldrift.py

reftank.py -

Command line tool for extracting tank calibration data from the DB. This uses similar python modules for extracting the data as used in caldrift so it also pairs measurement results with TypeB uncertainty terms. However, it is faster than caldrift since it does not do any processing. Useful for simple data extractions.

Current version: /ccg/bin/reftank
Note, this is a shell script to set up the correct python environment and then call reftank.py

get_standard_value.py -

Command line tool for extracting value assignments for standards from the DB. Useful for quick access to value assignments and histories.

Current version: /ccg/bin/get_standard_value.py

11. Data Management

The CCL uses a hybrid data management strategy where data collected during analysis is stored in text files on the /ccg virtual server and processed mole fraction results are stored in a relational database. CCL tank calibration systems follow a common strategy with other GML programs. Older systems may not have collected as much information as current system. However, the basic directory structure and the formatting of the files given here is consistent for all systems beginning with the initial flask measurements in 1967.

The database is backed up once a day and the server which hosts the text files has a full backup monthly with daily incremental backups in between. Full backups are kept for one year.

11.1 Text Files

There are 3 types of text files recorded by the processing systems; data, raw and qc. The base text file naming scheme for all three uses the date / time of the beginning of the analysis episode, the instrument code of the analyzer used, and the species measured:
YYYY-MM-DD.HHMM.inst.sp

Raw files are named using the base filename, qc files are named using the base filename plus a '.qc' extension, and data files are named using the base filename with a '.dat' extension for optical analyzers and a '.zip' extension for GC data files.

The files are stored on the /ccg drive with the path set by the species measured, sample type (flask, cals, response curve (abbreviated nl)), and system name. With file type (described below) and year subdirectories:

/ccg/[species]/[sample_type]/[system]/[file_type]/YYYY/

For example, CO_2 measurements of target tank CB12113 started at 14:34 on Jan 15th, 2025 using instrument 'PC1' on system 'co2cal-2' would use 2025-01-15.1434.pc1.co2 as the base filename and files would be located:
Rawfile: /ccg/co2/cals/co2cal-2/raw/2025/2025-01-15.1434.pc1.co2
Qc file: /ccg/co2/cals/co2cal-2/qc/2025/2025-01-15.1434.pc1.co2.qc
Data file: /ccg/co2/cals/co2cal-2/data/2025/2025-01-15.1434.pc1.co2.dat

--- Raw files:
Raw files contain information needed to calculate sample mole fractions or to provide input for deriving calibration curves. Calibration aliquots are identified by serial number listed in the file header. Raw files are named using the base filename.

ex. directory/file:

/ccg/co2/cals/co2cal-2/raw/2025/2025-01-15.1434.pc1.co2

There are two formats for rawfiles (for all system/times), one for optical analyzers and the second for gas chromatography instruments.

Optical Analyzer File Format
Format: type gas yr mo dy hr mn sc sig sig_sd sig_n flag
REF R0 2025 01 15 14 38 24 416.8995 0.0196 10 .
SMP W 2025 01 15 14 41 35 399.1819 0.0148 10 .

Where *sig*, *sig_sd*, and *sig_n* are the average, standard deviation, and number of measurements recorded for the aliquot analysis. Flag is a data quality indicator where anything other than '.' indicates a fatal problem with the measurement and that the aliquot should not be used during processing. Mole fractions can be generated if one of the bracketing reference aliquots is missing or flagged but not if both are missing / flagged.

Gas Chromatography File Format
Format: type gas yr mo dy hr mn sc pH pA Tr flag bc
REF R0 2020 01 02 12 53 27 2.296530e+05 1.757642e+06 60.8 . BB

SMP 468840 2020 01 02 13 00 43 2.164365e+05 1.650708e+06 60.8 . BB

Where pH, pA, and Tr are the peak height, peak area, and retention time recorded from processing the chromatograms. Flag is a data quality indicator where anything other than '.' indicates a fatal problem with the measurement and that the aliquot should not be used during processing. 'bc' is the baseline code reported by the chromatogram integration software. Acceptable baseline codes are identified in the rawfile header for gas chromatography analysis and vary between species and analytical instruments.

--- QC files:
Contain QC data particular to an instrument. When the data applies to multiple species (for example CO_2 and CH_4 measurements from a CRDS instrument) the files are duplicated and accessible from both specie's directories. Information in the QC files varies by instrument and by time if changes to the system are made. The "Format" line in the file header identifies the parameter in each column with associated units. QC files are named using the base filename plus a '.qc' extension.

ex. directory/file:

/ccg/co2/cals/co2cal-2/qc/2025/2025-01-15.1434.pc1.co2.qc

Format: type gas yr mo dy hr mn sc analysis_time_delta(secs) h2o(percent) cell_press(Torr) cell_temp(C) das_temp(C) etl_temp(C) wb_temp(C) roomT(C) chillerT(C) flow(ml/min) inlet_press(Torr)
;+ Start of Data - Do not write below this line!!
REF R0 2025 01 15 14 38 24 153 -0.0003 139.9999 45.0000 43.8695 45.0090 45.0002 21.6 -67.0 158.5 760.7
SMP W 2025 01 15 14 41 35 154 -0.0005 140.0000 45.0000 43.8490 45.0088 45.0000 22.0 -67.9 149.7 760.3

--- Data files:
Contain the data strings recorded from optical instruments or the chromatogram data for GC's. For Picarro and Aerodyne analyzers these files contain the actual data lines reported by the analyzer that went into the averages recorded in the rawfiles. Content is instrument dependent. Data files are named using the base filename with a '.dat' extension for optical instruments. For GC's the data file is a zip file of all the chromatograms collected during the analysis episode so is named using the base filename with a '.zip' extension.

ex. directory/file:

/ccg/co2/cals/co2cal-2/data/2025/2025-01-15.1434.pc1.co2.dat

11.2 Description of relevant database tables

REFTANK.SCALE_ASSIGNMENTS
Table used to store value assignments for all standards used for CCGG flask, calibration, and observatory in-situ systems. Note that the history of assignments is preserved so a single cylinder/fill/start_date may have multiple assignments on any given scale. Only the latest record (most recent "assign_date") for any given scale is considered current, older records are for documenting historical assignments and establishing traceability paths for certificates issued by the CCL. Assignments on different scales are considered unique to preserve the historical records so cylinders will have "current" value assignments on multiple scales.

This table is important for all aspects of GML data processing. It holds value assignments for standards used for all GML programs (flask, tank calibrations, surface in-situ observations, etc.) All value assignments should be made using the caldrift software (described above) to ensure additions or changes are done correctly.

Value assignments can be time dependent to account for measured drift occurring in the cylinders:

$$X = coef0 + coef1 * (dt\text{-}Tzero) + coef2 * (dt\text{-}Tzero)^2$$

where *X* is the mole fraction assignment of the standard at time dt. Coef 0-2 are the coefficients of a polynomial fit to the calibration history of the cylinder. Where no drift is occurring coef1 and coeff2 are set to 0. Tzero is time zero for time dependent assignments. T zero is set to the weighted mean date of calibration history used in the fit where the weights are based on individual calibration episode uncertainties. Setting Tzero to the weighted mean date allows linear approximation of time dependent uncertainty as

$$Unc = unc\_c0 + unc\_c1 * (dt\text{-}Tzero) + unc\_c2 * (dt\text{-}Tzero)^2$$

This is a close approximation for linear functions and is adequate for non-linear drift modeled by a quadratic function.

Start_date is the date that the value assignment goes into service and the value assignment is valid until the cylinder is refilled or another value assignment for the same tank/fill is made. Typically, start_date is the same as the cylinder fill date and there is only one value assignment per cylinder/filling. However, we allow multiple assignments for cases where there is a step change in the cylinder or the polynomial functions do not adequately model the measured drift and we need piecewise assignments.

The assignment date is independent of the start_date. This is the date that the assignment was made. Only the most recent assignment for a cylinder /start_date is considered current and used. Older value assignments based on the assignment dates are preserved for historical records and are used to document traceability for certificates issued by the CCL.

Field,	Description
num,	Unique record identifier
scale_num,	Unique scale number to relate assignment to reftank.scales
serial_number,	Cylinder serial number
start_date,	Date assignment goes into service. Typically equals cylinder fill date unless a split assignment is needed
tzero,	Time zero for time dependent assignments. Set to weighted mean date of calibration history where weights are based on calibration episode uncertainties. Allows approximation of time dependent uncertainty
coef0,	Coef 0 of time dependent fit
coef1,	Coef 1 of time dependent fit
coef2,	Coef 2 of time dependent fit
unc_c0,	Uncertainty of coef0
unc_c1,	Uncertainty of coef1
unc_c2,	Uncertainty of coef2
sd_resid,	Standard deviation of residuals
standard_unc,	Not used
level,	Indicator of level in hierarchy. This is for information only. Other designates a cylinder outside of the NOAA calibration hierarchy.
assign_date,	Date assignment was made
comment,	Comments / notes on assignment
n,	Number of data points used in fit.
certificate_use_only	1/0 to indicate whether this assignment was created during certificate creation. Only assignments with a 0 should be used internally for value assignments.

REFTANK.SCALE_ASSIGNMENTS_CALIBRATIONS

This table records the date and the unique identifiers of the specific calibrations table rows used to determine the assigned value for a scale_assignments row. The date recorded is the same as the "assign_date" in reftank.scale_assignments and links the assignment to the calibration results in the reftank.calibrations_history table.

REFTANK.SCALE_ASSIGNMENTS_VIEW
View to allow easier accessibility to value assignments of standards based on fill records, filtering of the history of the value assignments, etc. Note the "current_assignment" field provides a simple way to determine the current value assignment for a cylinder/fill/start_date. Historical value assignments that have been superseded have a 0 in this field.

Additional fields included in view:

Field,	Description
fill_code,	Fill code for this serial number / filling
end_date,	End date for value assignment. Either the next start date for serial number/ fill or the next fill date for serial number
next_fill_date,	Next recorded fill date for serial number, serves as end date for value assignment if no additional assignments have been made
current_assignment,	Indicator of current value assignment for cylinder / start_date, based on the most recent assign_date field

REFTANK.FILL
DB table to record the history of repeated filling of the same cylinder. Fill codes are linked to the fill date and provide an easy way to distinguish records from multiple fillings.

Field,	Description
idx,	Unique record identifier
serial_number,	Serial number of the cylinder
date,	Date of filling
code,	Fill code
location,	Location of filling
method,	Fill method
type,	Undefined use
h2o,	Water content if measured
notes,	Fill notes

REFTANK.RESPONSE
Database table used to store coefficients and associated information for response curves. Coefficients are determined by nlpro.py and used by flpro.py and calpro.py to generate mole fraction results.

The record with the most recent start date for the system, instrument & species relative to the analysis date is used during the mole fraction calculation.

Coef 0 - 3 are the coefficients of the equation defined by the function field to use to calculate mole fraction of samples. Uncertainty of the curve is taken as the predictive interval of the fit. The covar field is used to store the covariance matrix of the fit which is used to generate the confidence interval which is then added to the standard deviation of the residuals (rsd field) in quadrature to give predictive interval.

Reference operation field defines the reference normalization procedure used to produce the response curve and to calculate mole fractions for samples. Either subtraction, division, or no reference normalization (none).

Field,	Description
id,	Unique record identifier
site,	BLD or field site code
parameter_num,	Unique parameter number for species
scale_num,	Unique scale number
system,	Analysis system name
inst_id,	Instrument code
start_date,	Date curve should go into service
start_date_id,	Indicator used when multiple start dates are required
analysis_date,	Date of response curve analysis
coef0,	Coefficients of fit
coef1,	Coefficients of fit
coef2,	Coefficients of fit
coef3,	Coefficients of fit
rsd,	Standard deviation of residuals
n,	number of data points
flag,	quality control indicator. Anything other than '.' indicates a fatal problem and the response curve should not be used.
function,	Function (polynomial or power). Typically only polynomial functions are used.
ref_op,	Reference operation (subtraction, division, or none)
ref_sernum,	Serial number of the reference tank used
standard_set,	Standard set used
filename,	Name of rawfile
covar,	Covariance matrix
comment,	comments

REFTANK.RESPONSE_SCALE_ASSIGNMENTS

This table records the specific versioned reftank.scale_assignment rows used to create the

response curve.

REFTANK.SCALES

Database table used to record scales in use by GML or by external partners. The scale number is unique and is used to directly indicate traceability of data to the scales. "Current" indicates the current scale used by NOAA for measurements, others are historical or test versions. "Published" field indicates if the scale was ever used for published data. This allows test scales to be distinguished from those actually used.

Field,	Description
idx,	Unique identifier of the record. Used in other tables to link measurement results to a unique scale. Listed as scale_num in other tables.
parameter_num,	Unique parameter number
species,	Species
name,	Name of scale
current,	1 if it is the current scale for this parameter, 0 if it is an archived or test scale
start_date,	Date of scale release
end_date,	Date scale was archived
scale_min,	Min scale range
scale_max,	Max scale range
comment,	Notes
published,	1 if data on this scale was ever released, 0 if test scale only

REFTANK.CALIBRATIONS

Database table used to hold episode means for tank calibrations.

Note that the reproducibility and other typeB uncertainty terms needed for tank calibration results is not included in this DB table. These are stored in a lookup table and paired with the results on extraction from the DB.

Field,	Description
idx,	Unique index number
serial_number,	Serial number of sample cylinder
date,	Analysis date

time,	Analysis time
species,	Parameter measured
mixratio,	Episode average
stddev,	Standard deviation of aliquots
num,	Number of aliquots
method,	Analysis method
inst,	Id of instrument used for analysis
system,	Analysis system
pressure,	Recorded cylinder pressure
flag,	Quality control flag. Anything other than '.' indicates fatal measurement problem and data should not be used for value assignment
location,	Location of measurement
regulator,	Recorded id of regulator used on sample tank
notes,	Notes on analysis episode
mod_date,	Last date the DB record was modified
meas_unc,	Episode uncertainty relative to the scale
scale_num,	Unique identifier of the reference scale used for the measurement
parameter_num,	Unique identifier of measurement parameter
run_number,	not used

REFTANK.CALIBRATIONS_HISTORY

Database table used to record all modifications to the calibrations table.

Text file for TypeB uncertainties:

/ccg/flask/unc_master.conf

Appendix 1 - Example of Measurement Uncertainty Calculated for a Discrete Air Measurement Which Uses a Similar Process:

Use CO_2 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 7, 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 11, we get

μ_R = 411.751633323 * 0.000066176 = 0.02725

Therefore, using Equation 12 we get

μ = 0.03318