19ENV05 STELLAR

(Stable isotope metrology to enable climate action and regulation)

D5

Good practice guide for specification and application of OIRS for atmospheric measurements, including sample handling protocol, optimised analytical procedures, traceability to the international standards and target uncertainties (0.05 ‰ for $\delta^{13}C\text{-}CO_2$ and $\delta^{18}O\text{-}CO_2$)

Coordinator	Paul Brewer (NPL)
Authors:	J. Braden-Behrens (PTB), A. Emad (PTB), V. Ebert (PTB), F. Durbiano (INRiM), G. Li (PTB), H. Moosen (MPG), J. Nwaboh (PTB), S. Pavarelli (INRiM), F. Rolle (INRiM), M. Sega (INRiM), P. M. Steur (RUG)

Due Date: Feb 2023 (M30)

Submission Date: Oct 2023

Contact author: jelka.braden-behrens@ptb.de

Table of contents

1. Introduction

This document addresses scientists who work with an optical isotope ratio spectrometer (OIRS) to measure $\delta^{13}C$ and $\delta^{18}O$ in ambient CO_2. This document outlines a series of steps to be taken into account when a spectroscopic isotope analyser is to be metrologically characterized. For each characterization step, we give background information to clarify the underlying definitions and to enable the user to interpret the resulting data. Further we suggest a set of specific experiments to provide robust characterization with limited time and gas consumption to fit the needs of the isotope ratio measurement community. In this document, we discuss traceability and uncertainty evaluation compliant to the 'Guide to the Expression of Uncertainty' (BIPM et al., 2008). Further this document contains a sample handling protocol to avoid fractionation and suggestions to optimize analytical procedures. We want to emphasize that gas and cylinder handling (including the tubing used, the choice and storage of cylinders, the choice and usage of regulators and valve seals) is crucial to optimise the experiment.

2. Nomenclature

The nomenclature used throughout this document is listed in Table 1. Further, we use abbreviated AFGL code for isotopologues, indicating isotope mass numbers modulo 10, so for example 636 for isotopologue $^{16}O^{13}C^{16}O$.

Number of molecules of a species i in a test volume V	N_i
Amount fraction of a species i	$\chi_i := \dfrac{N_i}{N_{tot}}$
Isotopic ratio	$R := \dfrac{\chi_{heavy}}{\chi_{light}}$
Delta value (relative deviation from standard, often given in ‰)	$\delta := \dfrac{R - R_{std}}{R_{std}}$
Time constant	τ
Averaging time	Δt
Standard deviation	σ
Allan deviation	σ_A

Table 1 Nomenclature used throughout this document.

3. Characterization steps

3.1. Information about the instrument, data evaluation steps and the fitted spectrum

General information about the analyser and its fitting routine helps to understand and interpret the measurement results as well as the limitations of the analyser. To characterize an optical instrument, we recommend to collect information about the instrument and data evaluation steps in a table (see Table 2 as an example) and also plot and fit an example spectrum (see Figure 1).

Step 1: Information about the instrument and data evaluation steps

Collect, if available, information about technical details, operating conditions, and the spectral fit of the instrument. If it is a self-built instrument, this information is readily available, if a commercial instrument is used, this information could be provided by the manufacturer. It is further useful to find out if modifications to these instrument characteristics are possible for the user.

- Technical details
    - measurement principle (e.g., direct absorption, ringdown-time, …)
    - the light source (laser/broadband)
    - wavelength range (NIR 0.78-3 μm/MIR 3-50 μm[1])
    - effective optical path length
- Operating conditions/ranges
    - guaranteed CO_2 amount fraction range
    - operating CO_2 amount fraction range
    - cell pressure and temperature
    - flow rate (range)
- Information about spectral fitting
    - baseline fit (type of baseline fit - if polynomial: order of the fit)
    - line positions
    - line shape (e.g., Voigt lines, HTP profile
    - pressure broadening coefficients (fixed/free)
    - used fit data (e.g., HITRAN)
    - which other molecules are included in the fit?

Background information

- **Technical details**

  There are different spectroscopic measurement methods available to measure the isotope composition of CO_2-in-air. These methods differ in the measurement principle, such as direct absorption spectroscopy, cavity ringdown time spectroscopy (CRDS), integrated cavity output spectroscopy (ICOS) or Fourier-transform infrared spectroscopy (FTIR). Different spectroscopic instruments differ in used light source (laser/broadband) and wavelength range (NIR/MIR) and the effective optical path length. For an overview of the various methods that are used, please see (Brewer et al., 2019) and (Griffith, 2018).

- **Operating conditions**

  Different operating conditions are in use, for instance, continuous flow or static mode measurements, as well as various ranges of cell volumes and pressures. Those factors can also influence instrument performance, gas consumption and field applicability. Thus, operating conditions should be considered when deciding on the gas handling procedure. In particular, the operating temperatures and the cooling techniques that are used are seen to have an influence on the possibility to conduct in-field measurements with the analyser. Further characterization steps

---

[1] ISO 20473 division scheme for dividing the infrared spectrum.

should be performed spanning the expected measurement conditions.

- **Spectral fitting for lased based instruments**

  For laser-based instruments, many different fitting algorithms are in use. Those algorithms can differ in the baseline fit and in the used absorption lines of CO_2 as well as of other molecules in the spectral range. Further, different algorithms differ in the used fit data and line shapes, the way pressure broadening is treated.

    o When it comes to line shape, there are different choices, such as Voigt line shape and more advanced line shapes such as the so-called Hartmann-Tran profile recommended by the IUPAC Task Group for H_2O, based on hard-collision model (Tennyson et al., 2014) Those line shapes have advantages and disadvantages when applied to OIRS calibrated using reference materials. At first glance, HT profile retrieves the correct line area with flat fit residuals, whilst Voigt line shape typically results in "w" shape residual because it does not consider Dicke narrowing effect and speed dependence effect. However, when sample and reference gases are measured in identical conditions, particularly at the same pressure and temperature and with the same gas matrix, the bias in the retrieved line area using Voigt profile, to a large extent, cancels out. Further the Voigt profile has fewer free parameters in fit, yielding smaller statistical uncertainty.

    o If the HITRAN database is used, it can be interesting to figure out, which parameters are used only as a first guess (e.g., line positions) and which parameters are fixed to the HITRAN values (e.g., line intensity or broadening parameters). Further, it can provide some insights to evaluate if the used HITRAN data is based on measurements, interpolations, or ab/initio data, as the first of these options would be the preferred options while there might be potential for improvement for data based on interpolations and in particular ab/initio simulations.

    o To evaluate possible interferences not included in the fit, one could also check the spectral region for potential absorption lines of other molecules with a spectral simulation tool such as spectral calc(https://www.spectralcalc.com/).

- **Spectral fitting for FTIR spectrometers**

  The published FTIR isotope methods are technically different from laser-based instruments. Although the fundamental principle of isotopologue-specific absorption in the infrared is the same for both laser- and FTIR- based techniques, current low-resolution FTIR methods use the band contour instead of the ratio of individual line areas. For this reason, the method is software dependant. To our knowledge, the success of FTIR-based OIRS has been achieved so far only with the MALT software. The MALT software is a radiative transfer calculation code that can generate synthetic infrared calibration spectra of gas phase mixtures. These calibration spectra are calculated from the HITRAN database of absorption line parameters (hitran.org). MALT calculates atmospheric transmission or absorbance spectra and then modifies the ideal spectra to take in account different effects that occur in the real spectrometers. These effects include environmental (pressure, temperature, path length, …) and instrumental effects (e.g., resolution, line shape, wavenumber shift), and are considered in the calculations, to obtain spectra that closely approximate real measured spectra. Recently, there are advances in accurate determination of line intensity using high-resolution FTIR technique (Bielska et al., 2022). Such a technique

could be useful in the future for calibration based OIRS and ultimately absolute OIRS method as successfully demonstrated in (Fleisher et al., 2021).

Example Information about an OIRS instrument, its data evaluation steps and spectral fit

Information about the Aerodyne CW-IC-TILDAS to measure ^{13}C and ^{18}O in CO2	Parameter	Value	Modifiable by user (y/n)
Technical details	Measurement principle (direct/CRDS/ICOS)	Direct absorption	n
	Light source (laser/broadband)	Interband cascade laser (ICL)	n
	Wavelength range (NIR/MIR)	MIR	n
	Effective optical path length	36 m	n
	Commercial instrument (y/n)	y	n
Operating conditions	Guaranteed CO_2 amount fraction range	0-5000 ppm	n
	Operating CO_2 amount fraction range	Depending on calibration materials	n
	Cell pressure	50 mbar	y
	Cell temperature	20 °C	y
	Flow rate (range)	/ (static mode)	/
Spectral fit information	Baseline fit	Polynomial, 3rd order	y
	Dominant line positions	4.25 and 4.33 µm	y
	Line shape	Voigt	n
	Pressure broadening	fixed	n
	Is HITRAN used in fitting (y/n)	y	y
	Other molecules included	n	y

Table 2 Example for a collection of information about an OIRS instrument for an Aerodyne CW-IC-TILDAS.

> **Step 2: Spectral fit**
> Export raw and fitted spectra from the instrument at representative operating conditions (p, T, χ, Φ). Plot measured vs. fitted spectra and residuals deviations from fit.
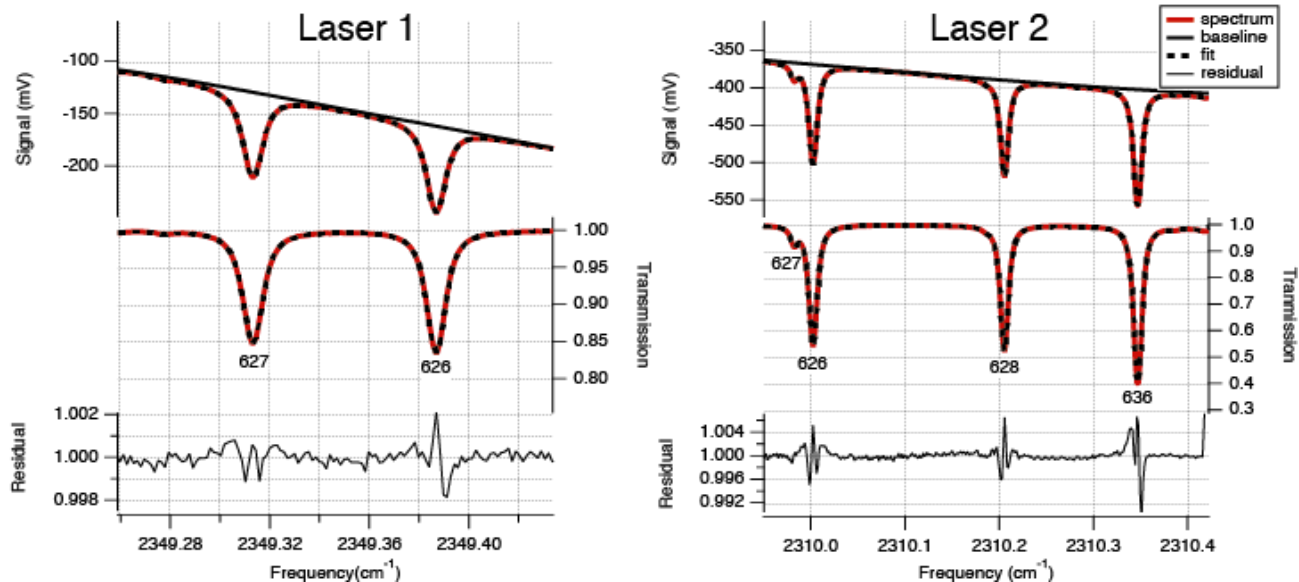
Figure 1 Typical absorption spectrum, transmission spectrum and residual for laser 1 (left) for measurement of χ^{627} and χ^{628}, and laser 2 (right) for measurement of χ^{626}, χ^{628} and χ^{636} of an Aerodyne CW-IC-TILDAS-D.

3.2. Limit of detection

The instrument detection limit (IDL) of an OIRS instrument can be defined for each isotopologue separately. Measuring and comparing the different limits of detection for the different individual isotopologues can provide insights about the instrument performance at lower concentration.

> **Step 3: Measuring the instrument detection limit**
> Measure a cylinder of CO_2 free air (e.g., synthetic air) or N_2 (preferably, 6.0 grade) and use the measured isotopologue amount fractions to determine the IDL as defined above for each isotopologue separately. Please note that the so defined IDL generally depends on Δt, because σ depends on Δt – so choose an averaging time that merges with planned experiments and always report Δt together with your measured IDL.

Background information: Defining the limit of detection

There are different definitions for the limit of detection available. A common definition is given by the IUPAC Compendium of Chemical Terminology (A. D. McNaught and A. Wilkinson, 1997). Here the instrument detection limit (IDL) is defined as the lowest amount fraction of the respective isotopologue that can be detected with a reasonable certainty (e.g., $k=3$ for 99.85 % confidence of a one-tailed gaussian distribution). The IDL can be calculated based on a time series of measurements of a blank sample – here we show an example for isotopologue 626:

$$\text{IDL}(\chi^{626})_{\Delta t} := \overline{\chi_{blank}^{626}} + 3\,\sigma_{\Delta t}\left(\chi_{blank}^{626}\right) \tag{1}$$

With the mean of the measured amount fractions $\overline{\chi_{blank}^{626}}$ and their standard deviation σ at a given data acquisition or averaging time Δt. The blank amount fraction is measured with the matrix gas of the intended application.

3.3. Response of the instrument to changing input

The understanding of the response of an instrument to changing input is essential to design an experiment and interpret measured data. For example, the instrument's response must be considered to choose appropriate flushing times and to determine the instrument's measurement frequency, i.e., the maximum frequency that an instrument can resolve. This can be limited by either gas exchange in the cell or by the instruments maximum data acquisition frequency. Typically, gas exchange in cell frequency is lower than the instruments data acquisition frequency. Further, if air samples are introduced continuously to OIRS instruments, previous samples yield sample cross contamination on later measurements due to continuous mixing and surface interaction. However, these effects will decay with time and after sufficiently long intervals will disappear.

> **Step 4: Measuring the response time**
> Measure the instrument's response to a switching input between two gas mixtures with different amount fractions (and if possible different delta values) within the application range. The switch should be located as close as possible to the analysers inlet to avoid additional mixing in the tube before the air enters the optical cell.
> - Measure a suitable response time t_p (as dfined below) for the different isotopologues separately. Comparing the time constants of the different isotopologues to each other might provide insight about fractionation.
> - Test if the instrument response is exponential e.g., by evaluating if a semi-log plot of $\chi(t)$ shows a straight line. If this is the case, evaluate the time constant τ of the different isotopologues (for the time that shows a clear exponential decay) and compare it to the ideal time constant $\tau_\mathrm{ideal} = \frac{V_\mathrm{cell}}{\Phi}$. This might provide insights into memory effects.
> - If available: Compare the measured time constant to the available manufacturer information about response times (e.g., using the time constant related to the 10-90% rise-fall time given by the manufacturer).

Background information:

- **Different definitions of turnover times**
  The instrument's response to a changing input can be characterized by the time the instrument needs to respond to this change. Different definitions of response times are possible – most of those definitions are defined as the time required after a step change of magnitude $\Delta\chi := \chi_{start} - \chi_{end}$ to reduce deviations from the asymptotic value χ_{end} by a certain factor. Here we define the instrument's response time t_p as the time required to reduce $\Delta\chi$ to a certain percentage of the steps magnitude. For example, $t_{0.01}$, is the time it takes to reduce the deviation from the final value $\delta(t)$ - δ_{end} to 1% of its initial value $\Delta\delta$. A common choice for p is 1/e =app. 0.368, yielding the time constant τ of an assumed exponential decay. Further, the commonly used 10-90% rise-fall time differentiates between $\chi_{start} < \chi_{end} \ (rise) \ and \ \chi_{start} > \chi_{end} (fall)$ can be calculated as $t_{10-90 \ \mathrm{rise/fall}} = t_{\mathrm{p}=0.1} - t_{\mathrm{p}=0.9}$.

- **Converting between different response times (applicable for exponential turnover)**

EMPIR

EURAMET

The EMPIR initiative is co-funded by the European Union's Horizon 2020
research and innovation programme and the EMPIR Participating States

If perfect mixing in the optical cell dominates the instruments response, the signal (here we take χ as an example) can approximated as an exponential decay from χ_{start} to χ_{end}.

$$\chi(t) = \chi + \Delta\chi\, e^{-t/\tau} \tag{2}$$

However, the 1/e threshold is arbitrary and other threshold-based response time t_p with other thresholds p can be defined accordingly. Assuming exponential decay, one can convert between any threshold-based response time t_t and the the time constant τ by:

$$\boxed{t_{0.01} = -ln(0.01)\,\tau} \tag{3}$$

An alternative definition for the instrument's response to a changing input is the 10-90% rise-fall time. For an exponential rise or decay, this is related to the time constant τ by:

$$t_{10-90\ \text{rise/fall}} := t_{0.9} - t_{0.1} = \tau * ln\left(\frac{0.1}{0.9}\right) \tag{4}$$

Further, if the air in the measurement cell would be perfectly mixed and the changing input enters the instrument as a step change, the theoretical time constant τ can be calculated from the flow rate, the cell pressure, and the cell volume by:

$$\tau_{ideal} = \frac{V_{cell}}{\Phi} = \frac{V_{cell}}{\Phi_s}\frac{p_{cell}}{p_s}\frac{T_s}{T_{cell}} \tag{5}$$

With volume flow rate Φ, pressure p, temperature T and volume V and the indices s for standard conditions and cell for cell conditions. The difference between the ideal time constant τ_{ideal} and the measured time constant τ_{meas} provides information in how far the assumptions of a) a rectangular step change and b) perfect mixing in the cell are fulfilled. Further, changes in τ_{meas} with time can provide information on lingering effects at different timescales.

3.4. Instrument stability

The instrument's stability and in particular short-term noise and drift determine the optimal usage and the limitations of a specific instrument. The analysis of the instrument's stability should involve amount fractions bracketing the measurement range and the respective cell pressures and flow rates should be noted and should match with the planned experiments. The stability of laser spectrometers is often analysed via measurements of Allan deviation and so-called Allan Plots (defined below) that can be used to analyse different noise components of a time series – e.g., $\delta(t)$. The analysers minimum short-term noise can be defined as the minimum of the Allan deviation $\sigma_{A,min}$ with its corresponding optimal averaging time Δt_{opt}. Further, the typical drift at the calibration interval Δt_{cal} can be detected as $\sigma_{A,cal}$.

Step 5: Allan deviation
Connect a stable gas source (e.g., a cylinder of compressed air) to the instrument and measure it for a longer time period resulting in an Allan deviation test. We recommend a measurement that spans the expected calibration interval multiple times, yielding e.g., a measurement for one day or longer.
Perform at least three such Allan deviation tests at different amount fractions (bracketing expected measurement range) and plot the Allan deviation σ_A of raw δ values (and/or isotopologue amount fractions χ if applicable) for the different CO_2 amount fractions.
Identify timescales that are dominated by white noise or instrument drift. Discuss deviations from white noise and discuss if individual dominant noise frequencies appear.

Extract short term noise at data acquisition frequency f_(aq.), minimum noise at optimal averaging time and noise at recommended calibration interval including the respective averaging times Δt. Optional: If the instruments flow rate and cell pressure can be modified by the user, it can make sense to vary both and plot e.g., the minimum of Allan variance as a function of flow rate/pressure.

Background information

- **Factors that influence stability**
  The instrument's stability can generally depend on amount fraction, cell pressure and flow rate in various ways: A change in amount fraction can change the signal-to noise ratio. Differences in cell pressures influence collisional broadening of the absorption lines. And even if changes in flow rate theoretically do not change the absorption signal, there might be remaining effects that are related to changes in the effective pressure in the cell, depending on the design of the cell and the position of the pressure sensor.

- **Definitions: Allan variance, Allan deviation and Allan-plot**
  The analysis of Allan deviation or Allan variance (D. W. Allan, 1966), see also (Werle et al., 1993) is a commonly used tool to classify the noise of a laser spectrometer, as well as to explore optimal averaging and calibration times. Allan variance (also known as two-sample variance), Allan deviation and a convenient way to plot Allan variance - called Allan Plot can be defined as:
  Allan variance: The Allan variance of a time series $\delta(t)$ is half the average of the squared difference between two adjacent temporal averages $\overline{\delta_n}$ and $\overline{\delta_{n+1}}$ that are averaged at averaging time Δt [see also Werle 1993]. If the number of all datapoints N at a given averaging time Δt is N, this yields:

$$\sigma_A{}^2(\Delta t) := \frac{1}{2}\sum_{n=1}^{N-1}\frac{\left(\overline{\delta_{n+1}} - \overline{\delta_n}\right)^2}{N-1} \tag{6}$$

  Allan deviation: The Allan deviation $\sigma_A(\Delta t)$ is the square root of Allan variance $\sigma_A{}^2(\Delta t)$

  Allan Plot: A log-log plot of Allan deviation vs. averaging time *Δt* provides insights into the instrument's characteristic noise and drift contributions. Starting at a sufficiently fast data acquisition-time, we observe short term noise, that decreases with increasing averaging time **Δt**.

- **Interpreting slopes in Allan plots**
  The short-term noise of OIRS instruments is typically dominated by white noise frequency modulation (i.e., the spectral density *S(f)* of the noise is independent from the frequency). This yields a -1/2 slope in a log-log plot of Allan deviation. Due to instabilities such as temperature drifts, moving fringes or changes in background spectra, OIRS instruments will reach an optimum averaging time and start drifting (Werle et al., 1993). Table 3 shows different types of noise and drift and their respective slopes in a log-log plot of $\sigma_A(\Delta t)$.

Type of noise	Frequency dependency of $S_f(f)$ (phase spectral density)	Frequency dependency of $S_\Phi(f)$ (phase spectral density)	$\sigma_A(\Delta t)$ proportional to:	Slope in log-log plot of $\sigma_A(\Delta t)$
white noise phase modulation	f^2	const.	Δt^{-1}	-1
Flicker noise phase modulation	f	$1/f$	Δt^{-1}	-1
white noise frequency modulation	**const.**	$1/f^2$	$\Delta t^{-1/2}$	**-1/2**
flicker noise frequency modulation (also: 1/f noise)	$1/f$	$1/f^3$	**const.**	0
Drift	$1/f^2$	$1/f^4$	$\Delta t^{+1/2}$	**+1/2**
Linear drift	$1/f^3$	$1/f^5$	Δt	1

Table 3 Typical noise contributions d their respective slopes [based on Werle et al 1993 and Allan 1966], most pronounced types of noise in OIRS instruments are highlighted in bold.

- **Allan deviation as a measure of precision**
  Allan deviation is often used as an indicator for precision. Without further assumptions about the type of noise, this usage is not compatible with common definitions of precision – commonly expressed as standard deviation of a sample with N>2.

Example Allan deviation plot:
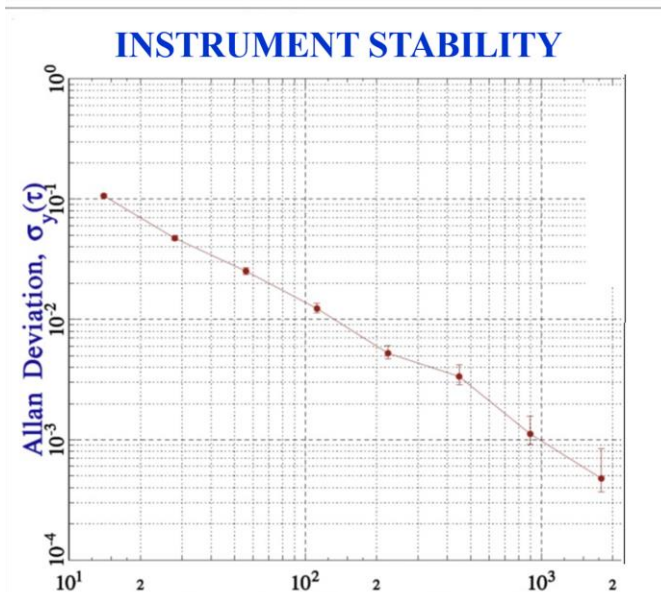
Measurement conditions		Allan deviation / ppm	Averaging time Δt /s
Amount fraction: 400 ppm Flow rate: 1200 sccm Cell pressure: app. 1 bar			
Short term noise $\sigma_A(\Delta t_{aq.})$ at data acquisition frequency $f_{aq.}$		0.107	14
Optimum noise $\sigma_{A,opt}$ at optimal averaging time Δt_{opt}		0.0053	224
Noise at recommended calibration interval $\sigma_{A,cal}$		0.0005	1792

Figure 2 Example Allan plot of an a FTIR spectrometer for isotopologue 626, based on a continues measurement for 7 hours at 400 µmol mol-1 mixture of CO₂ in air. The sample was produced by diluting a parent mixture (STELLAR 021) having amount fraction of 4992.25 (± 0.29) µmol mol-1 with CO2-free air. The data acquisition time τ was 14 s (5 scans) and the chosen averaging time was 224 s yielding an Allan deviation of app. 0.0053 µmol mol-1.

EMPIR EURAMET

The EMPIR initiative is co-funded by the European Union's Horizon 2020
research and innovation programme and the EMPIR Participating States

3.5.   Calibration

3.5.1.   Calibration approach

Different calibration approaches can be applied to calibration OIRS analysers, most common calibration approaches are the isotope ratio and isotopologue based calibration approach. Those approaches differ in assuming a certain shape (most commonly a linear shape) of the analyser's response either in delta space or in isotopologue amount fraction space. Some other calibration approaches[2] have been discussed e.g., by (Wen et al., 2013).
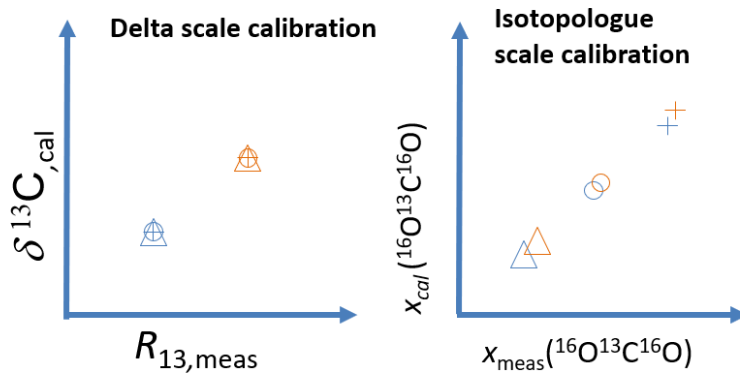
different colors: different ratios
different symbols: different amount fractions

Figure 3 Schematic drawing of the two most common calibration approaches that can be used for OIRS measurements. For delta scale calibration (left panel).

Limitations of the calibration approaches can be related to linearity (or a different assumed nonlinear calibration curve, e.g., polynomial) of the analyser in delta space and/or in isotopologue space within the application range, but also to the availability and uncertainty of reference material. For delta scale calibration, additional uncertainty is typically added by the applied correction for concentration dependency of the measured isotopologue ratio.

> **Step 6: Analysing the calibration curve**
> Test the applicability of the assumed calibration shape (e.g., linear shape) with a set of N>2 calibration materials (for calibration curves with more degrees of freedom, more points are needed). Plot the measured vs. the known (independently measured) values of the calibration material including the (propagated) uncertainty in the calibration material. Discuss if the assumed shape of the calibration curve (e.g., linear) is justified.

Background information

- **Reasons for nonlinearity**

  The applicability of the assumed calibration curve depends on the analyser. The linearity of an analyser in either space over the measurement range might be in general limited by

  a. interactions between different lines,
  b. changes in the baseline due to absorption of lines that are not captured by the fit,
  c. saturation effects,
  d. linearity of the frequency scale of lasers.

  Further, based on the definition of the δ values e.g., $\delta^{13}C \propto \chi^{636}/\chi^{626}$, linearity in both, delta space and isotopologue space for the same analyser would only be possible if χ^{626} would be linear with an offset of 0. Further, if there are offsets in isotopologue based calibration curves, they generally yield to concentration dependency of isotope ratios[3] (Griffith et al., 2012; Wen et al., 2013).

- **Details on the two most common calibration approaches**

  - **Isotope ratio-based calibration**

    For this approach, the isotope ratio response of the analyser is calibrated using reference material with known isotopic composition that brackets the intended measurement range, while the amount fraction is constant close to the concentration of the sample. This approach has been traditionally used by a variety of OIRS users and is typically used for IRMS instruments. Using this approach, the concentration dependence of isotope ratio (if existing for a specific instrument) must be corrected (see section 3.5.2). The uncertainties of this calibration approach are linked to the uncertainty of the used calibration material (typically measured with IRMS) and to the uncertainty of the correction of the amount fraction dependency.

  - **Isotopologue-based calibration**

    For this approach, the isotopologue amount fraction response of the analyser is calibrated using reference materials with different isotopologue amount fractions that are known with sufficient accuracy. It has been recommended to use calibration material with different amount fraction and identical isotopic composition. The calibration materials should cover the range of isotopologue amount fractions that will be measured in the application. This calibration approach (see e.g., Griffith, 2018) is not applicable for IRMS instruments as IRMS directly measures isotope ratios. But OIRS instruments measure the individual isotopologue absorption lines. However, the calibration material used for this approach typically uses IRMS-based ratio measurements in combination with additional amount fraction measurements. The latter might be optical, gravimetrical or manometric. Uncertainties of this conversion need to be propagated carefully and typically uncertainties in concentration measurements dominate the uncertainty of isotopologue amount fractions. Further, the conversion to isotopologue concentrations might include additional assumptions like assuming natural abundance for minor isotopologues.

---

[3] Exceptions only for very specific combinations of slopes and non-zero offsets

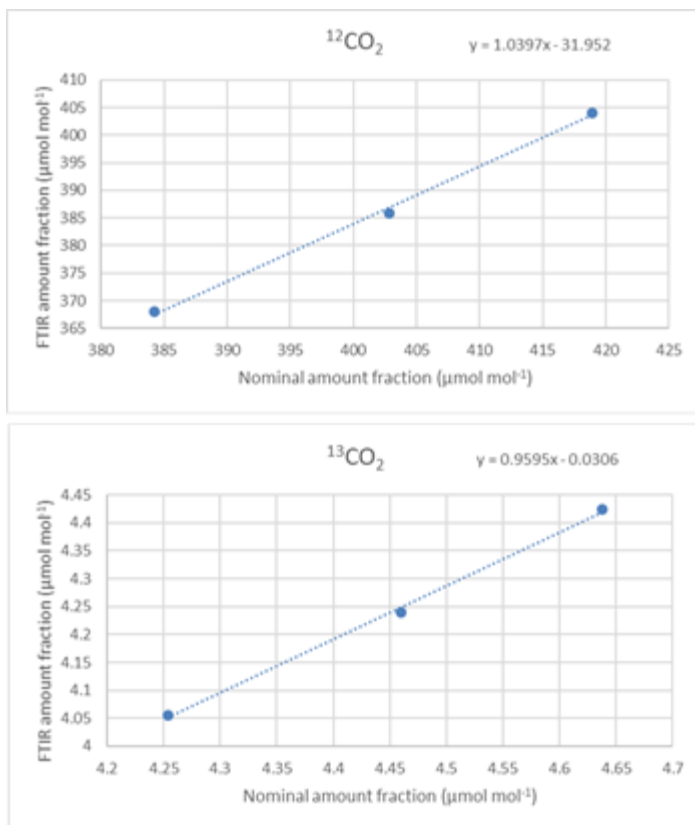**Example: Analysing the calibration curve of an OIRS analyser**

Figure 4 Example for a 3-point calibration on isotopologue scale with static mixtures with the same delta value (-9.8 ‰) and different amount fractions in the range (390.4-425.5) µmol mol^{-1}

3.5.2. Amount fraction dependency of the isotope ratio

If the calibration method is based on isotope ratios, the amount fraction dependency of the isotope ratio needs to be considered in most cases. The measured isotope ratio has been reported to depend on mole fraction for a variety of OIRS instruments. This dependence is typically different for each instrument. To quantify this dependence and its stability in time, repeated measurements of a gas with stable isotopic composition and varying amount fraction should be performed with the instrument.

> **Step 7: Amount fraction dependency (if applicable)**
> Perform repeated measurements to quantify the dependency of the uncalibrated delta value on amount fraction and its stability for at least two different delta values. This can be done by diluting a gas with constant delta value with N_2 or synthetic air.

Background information: Amount fraction dependency as a result of offsets and nonlinearities in the analysis function of individual isotopologues

A theoretical background to assume a certain shape of the amount fraction dependency was presented by (Griffith 2012, Griffith 2018). Based on the Beer-Lambert law, it can be motivated to assume a linear analysis function $\chi^{abc} = m^{abc}\chi^{abc} + b^{abc}$ for all individual isotopologues. This assumption yields[4] $\delta_{cal}(\chi^{626}_{raw})$ of the following shape (c.f. Griffith 2012):

$$\delta_{cal} = m * \delta_{raw} + b + p_1/\chi^{626}_{raw} \quad\quad\quad (7)$$

If there is no offset in the analysis function of the isotopologues (e.g. $b^{626} = b^{636} = 0$), the parameter p_1 would also be 0 and there would not be a concentration dependency. Further, if the analysis function of the individual isotopologues can be approximated as 2nd order polynomials, the following shape of the amount fraction dependency of δ values can be derived (Griffith 2018).

$$\delta_{cal} = m * \delta_{raw} + b + p_1/\chi^{626}_{raw} + p_2 \chi^{626}_{raw} \quad\quad\quad (8)$$

If we further use $\chi^{626}_{raw} = (1 - \sum_i R_i) \chi_{CO2,raw} = p_3 \chi_{CO2,raw}$ with constant p_3 for constant delta values, a similar equation holds for $\delta_{cal}(\chi_{CO2,raw})$:

$$\delta_{cal} = m * \delta_{raw} + b + p_a/\chi_{CO2,raw} + p_b * \chi_{CO2,raw} \quad\quad\quad (9)$$

However, please note, that these equations are based on a set of assumptions[5] and in general more complex amount fraction dependencies are possible, e.g. if the analysis function χ^{abc} cannot be approximated by a linear/quadratic shape.

3.6. Matrix gas effects

The composition of the gas matrix influences the absorption spectrum and can yield to changes in the measured isotopic compositions. Gas matrix effects might be minimized by optimizing the fitting and calibration routine. However, to quantify (remaining) gas matrix effects for an instrument used with a given fitting procedure, we recommend evaluating the effect of gas matrix changes by analysing samples with a constant delta value in different air matrices. This could be implemented by using gas cylinders with a subsample of the same parent CO_2 in different air matrices at the same (similar) amount fractions or by diluting pure or sufficiently high concentrated CO_2 with different matrices (e.g., N_2, binary or ternary air). No matter which of these implementations is chosen, it will be necessary to disentangle matrix gas effects from changes in amount fraction as there might be remaining amount fraction changes in the differently diluted samples.

> **Step 8: Characterize matrix effects**
> Perform a rough estimation of gas matrix effects on calibrated delta values. This can be done by diluting pure CO_2 (or sufficiently high concentrated CO_2 in N_2) with different background gases, while the CO_2 amount fraction is kept constant. Ideally those background gases would allow a variation of O_2 and Ar that spans the expected variations.

Background information

- **Physical reasons for matrix effects**
  Background gases influence the absorption spectrum in multiple ways such as I) direct interference due to absorption lines of other molecules in the spectral region II) changes in the shape of the lines

---

[4] This equation can be derived as a 1st order Taylor approximation for $1 + b^{626}/\chi^{626}$ - thus neglecting terms smaller $(b^{626}/\chi^{626})^2$. Further, assuming $(\delta_{cal} + 1) \chi^{626} > b^{636} - b^{626}$ yields the parameter p_1 shown by Griffith (2012).
[5] As discussed above the underlying assumptions are a 2nd order polynomial analysis function for χ^{abc} as well as $b^{626} << \chi^{626}$ and $(\delta_{cal} + 1) \chi^{626} >> b^{636} - b^{636}$.

due to pressure broadening. In how far these effects occur depend on the spectral region as well as in the analyser's fitting procedure. For CO_2 OIRS systems, gas matrix effects have been reported for changes in O_2, Ar. If such effects are not captured by the analyser's fitting procedure, a difference in the air matrix between calibration and sample gas can yield a significant bias in the isotope ratio.

- **Options to minimize gas matrix effects**
  Gas matrix effects could be minimized by a) choosing a calibration gas that reduces the difference between the air matrices of the sample and the calibration gas b) applying empirical correction curves determined for differences in the gas matrices and/or c) improving fit procedure to cover the matrix gas change (e.g., by adjustments to pressure broadening parameters).[6] These procedures might be very promising but can also have practical or conceptual limitations. Examples for such limitations are a) limited information and temporal changes in the sample gas matrix composition during field measurements as well as the limited availability or the high price of appropriate calibration gases b) conceptual limitations of assuming orthogonal correction coefficients for different species or c) limited user access to spectral fitting procedures or a limited number of spectral data points that does not allow more elaborate fitting procedures. Thus, there can be remaining gas matrix effects that are not captured by the fit, the calibration procedure or empirical correction factors.

3.7.   Repeatability and reproducibility

Repeatability measurements can help to evaluate uncertainty contributions that are due to multiple poorly understood factors without separating uncertainty contributions due to different input quantities.

> **Step 9: Repeatability experiment**
> Evaluate the repeatability of the instrument by applying measurements under repeatability conditions, i.e., by measuring a cylinder of natural air from a high-pressure cylinder, are measured alternately. If possible, further evaluate reproducibility.

Background information: Definitions
The repeatability and reproducibility of results can be used as a statistically uncertainty contribution in the uncertainty budget. The measurement repeatability is defined as the 'measurement precision under a set of repeatability conditions of measurement' (BIPM,VIM 2.21-3.6). The BIPM International vocabulary of Metrology (VIM) further defines repeatability conditions as follows:

> **Repeatability condition of measurement** 'condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time'
>
> (BIPM,VIM 2.20 -3.6, Notes 1 and 2)

Like repeatability, the focus of reproducibility studies is on random variation in the measurand without separating the effects associated with individual input quantities. Measurements made in particular with different instruments, in different laboratories, at different times, and by different personnel are carried out under a well-defined reproducibility condition of measurement [c.f. GUM].

> **Reproducibility condition of measurement**
>
> 'condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects'
>
> (BIPM,VIM 2.20 -2.24 -3.7, Note 2)

Such studies help to understand the uncertainties budget of the used instrument and to reveal uncertainty contributions that are not captured by the uncertainty analysis. However, such studies are recommended, but might be beyond the scope of an individual characterization of a single instrument.

Example

All SD	$n = 5$		$n = 10$	
in ‰	uncor	cor	uncor	cor
r636	0.04	0.020	0.06	0.025
r628	0.05	0.021	0.10	0.029
r627	0.06	0.018	0.18	0.03

Table 4 Repeatability of isotope ratio measurements of an Aerodyne CW-D-TILDAS in static mode. A sample and working gas, both natural air from a high-pressure cylinder, are measured alternately. Uncorrected values are standard deviations of the sample.

4. Sample handling protocol

4.1. Gas cylinders and regulators; reducing fractionation at cylinder walls and regulators

Gas stability
For calibration-based measurements, the stability of the calibration gases is crucial for obtaining accurate measurement results over longer time periods. A recent study on the stability of reference gases in high pressure cylinders has shown that the isotopic stability of pure and CO_2-in-air reference materials can be preserved over a wide range of cylinder pressures, ranging from 45 to 0 bar if the gas mixture has moisture levels below 3 µmol/mol and when the appropriate cylinder treatments are applied (Socki et al., 2020).
- We recommend to carefully treat cylinders e.g. according to Socki et al., (2020) and make sure that moisture levels are below 3 µmol/mol.

Gas cylinders for CO_2 in air mixtures
Aluminium surfaces absorb CO_2 which affects the CO_2 amount fraction of CO_2-in-air mixtures when pressure in the cylinder drops below 10 bar, as below this pressure CO_2 desorbs from the cylinder walls (Hill-Pearce et al., 2021). To realize stable and high precision results over a long time using OIRS it is therefore crucial that the pressure in the cylinder does not drop below 10 bar.
- We recommend to use CO_2 in air mixtures at pressures above 10 bar.

Gas cylinders for pure CO_2

It is known that the isotope composition of pressurized, pure CO_2 occurring in both the supercritical liquid (or supercritical fluid) and the gas phase in cylinders will change as the cylinder is emptied since the ^{13}C atoms favour the gas phase and the ^{18}O atoms favour the liquid phase, and less liquid will be present in the cylinder when the pressure drops (Socki et al., 2019). Pure CO_2 occurring both as liquid and gas in a cylinder containing a reference gas that is used over a longer time period should therefore be avoided, or a correction should be applied for the drift in $\delta\,^{13}C$ and $\delta^{18}O$ values. To avoid the formation of a liquid (or supercritical fluid) phase, the pressure in the gas cylinders needs to be below the temperature dependent saturation pressure (see Table 6). If gas cylinders are stored and used in the lab at 20°C room temperature, a cylinder pressure below 57 bar would be sufficient while for field measurements and in particular if gas cylinders are stored outside, lower pressures are recommended (see Tab. 6). Socki and Jacksier, (2021) performed experiments based on measurements with cylinders that have been exposed to low temperatures, yielding recommendations to avoid fractionation of pure CO_2 cylinders.

- Based on the recommendations by Socki and Jacksier, (2021), we recommend to:
  - use pure CO_2 cylinders with single-phase CO_2 (e. g. $p < 34.88$ bar at 0°C),
  - store the cylinders inside, if this is possible,
  - stabilize cylinders at lab temperature for at least 1-2 days (if stored in cold/hot conditions).

Temperature (°C)	Pressure (bar)
-30	14.278
-20	19.696
-10	26.487
0	34.851
10	45.022
20	57.291
30	72.137
https://webbook.nist.gov/chemistry/fluid/	

Table 5 Saturation pressure of CO_2 as a function of temperature.

Pressure regulators

If a pressure reducer is used for regulation of the pressure from the reference cylinder to the OIRS device, the choice of the type of reducer can be of importance for potential fractionation effects occurring as a result of the pressure reduction. The study from Socki et al., (2020) demonstrated that a two-stage, low dead volume pressure reducer showed no fractionation effects due to the use of the pressure reducer over a sampling range of 0.1 to 0.8 bar. More research can be conducted on single-stage pressure reducers, as well as the effects of pressure reducers over a wider sampling range.

- We recommend to use two-stage, low dead volume pressure reducers.

4.2. Gas sample conditions (pressure, temperature, and flow rate)

Fluctuations of pressure and temperature that are not captured by fitting and calibration routines might add uncertainty. For field applications that typically take place at less stable or less controlled environments, these uncertainties might get more pronounced.

- We recommend to analyse fluctuations of p, T and Φ for both, laboratory, and field conditions.

4.3. Avoiding memory effects

Memory effects occurring in OIRS systems depend on the following factors:
1. Gas inlet system and cell material
2. Occurring changes in CO_2 isotopologue amount fractions and isotope ratios

(including changes due to flushing the system with nitrogen in between measurements)
3. Static or continuous mode

Choosing appropriate flushing times can reduce those effects.

- We recommend to carefully evaluate the instrument's response to a step change (cf. section 3.3) and choose appropriate cell flushing time (e.g. $t_{flush} > t_{p=0.999}$) to avoid memory effects. The size of the step change should capture the maximum step changes occurring in the experiment.
- We additionally recommend to further evaluate the response of the whole setup including tubing and filters to a changing input.

Background information: Memory effects for analysers operating in different operational modes

- **Static mode**

  Devices operating in static mode will be most affected by surface effects, occurring when the cell is evacuated, and a sample gas is let in. This effect, dominated by absorption of CO_2 at the cylinder walls was clearly observed in an aluminium cell with a volume of 0.16 L when it was flushed with CO_2-free air, thereby stripping the surface from its CO_2 molecules, before letting in dry atmospheric air. Such high effects will not occur when measuring atmospheric samples without flushing with CO_2 free air in between measurements, since the CO_2 amount fractions in the gases are of the same order of magnitude. This effect will however occur to a lesser degree when switching from a low to a high CO_2 amount fraction gas. For the Aerodyne CW-IC-TILDAS-D system at the University of Groningen (RUG) it was determined that ~0.01% of the preceding sample in the cell affected a sample measurement. A sensitivity analysis was performed showing that this is such a small amount that scale effects due to cross-contamination are well below the measurement uncertainties (Steur et al., 2021).

- **Continuous mode**

  For devices operating in continuous mode, surface effects of the cell will probably be of less importance as the cell is not evacuated between measurements. It would be helpful to get more insight in the process of cross-contamination due to surface effects, for instance the differences of this effects as a result of cell material choice and cell treatment. A comparison of static mode measurement procedures versus continuous flow measurement procedures can also help to gain insight into the actual effects of surface processes in the cell and the consequences on the measurement stability and precision.

5. GUM compliant uncertainty evaluation for OIRS analysers

5.1. The 'Guide to the Expression of Uncertainty' (GUM)

The 'Guide to the Expression of Uncertainty' (BIPM et al., 2008) provides recommendations on the evaluation, propagation, and expression of uncertainty in measurements. We recommend to explicitly perform a GUM compliant uncertainty evaluation to characterize a specific instrument. The basis of this uncertainty estimation is the formulation of a measurement model that is used to gain information about the measurand. The measurement model $\delta = f(X_1, X_2, ... X_N)$ describes mathematically how the measurand (here theisotopic δ value) is estimated by different input quantities X_i.

EMPIR · EURAMET

The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States

- **Different measurement models**
  The measurement model can be based on scientific law, an empirical relation, or a combination of laws that link the input quantities to the measurand. For OIRS instruments, a measurement model could be in principle based on scientific theory or based on calibration (most common).
  - **A theoretical measurement model**
    In principle, a theoretical measurement model for OIRS measurements would link the measured isotope ratios to measured laser absorption parameters in a similar way as the Beer-Lamberts law. Such a model is possible, and would involve measurements of line strength, temperature, and pressure as input quantities *X*. However, many of the needed line parameters are not well known and so far, most OIRS instruments use calibration-based measurements models.
  - **A calibration-based measurement model**
    Currently, most OIRS measurements are performed in comparator mode with (frequent) calibration. Thus, the calibration function provides the basic measurement model, assuming a certain relationship between the instrument's response and the corresponding values of the calibration standards.

- **Including well and poorly understood effects**
  This basic measurement model can be extended to account for well/poorly understood effects such as instrument drift[7], concentration dependency or matrix effects, nonlinearity of the calibration slope or sample preparation.
  - **Well understood effects** can be corrected for based on physical understanding of the effect or empirically. The remaining uncertainty related to well understood effects, is the residual doubt about the how well the correction factors are described.
  - **Poorly understood effects** on the other hand are known to exist, but little is known about their magnitude, thus they cannot be corrected for. However, they can be included in the uncertainty estimation by adding a random variable to the model. The random variable does not change the expectation of the output but contributes to increase the uncertainty. When random variation is included in the model, considerable care should be taken in the treatment of uncertainties associated with individual input quantities to ensure that effects are neither counted twice nor omitted (BIPM et al., 2008).

- **Assigning uncertainties**
  When it comes to assigning and estimating uncertainties, it can be helpful to distinguish two types of uncertainty: Type A uncertainty which can be estimated statistically and Type B uncertainty which is obtained from other sources of information (e.g., supporting measurements, literature values, certificates of reference materials). Based on the model function, standard uncertainties can be assigned to all input variables and if all input quantities are distributed gaussian, propagated to the measurand, by using the law of propagation of uncertainty using a first- or higher-order Taylor series approximation to the model see GUM for details.

$$\sigma^2 \left(M(x_i) \right) = \sum_{i=1}^{N} \left(\sigma^2(x_i) * \left(\frac{\partial M}{\partial x_i} \right)^2 \right)$$

For the measurand M and the input quantities x_i. The so called combined standard uncertainty $u_c := \sqrt{\sigma^2 \left(\delta_{cal} \right)}$, is defined as the standard deviation associated with the output estimate δ_{cal}.

---

[7] E.g., if raw δ values are interpolated between measurement times, this can be directly incorporated in the measurement model.

The expanded uncertainty U is useful to provide an interval y – U to y+U that is expected to encompass a large fraction of the distribution of values that could reasonable attributed to the measurement Y. A coverage factor k, typically in the range 2 to 3 is used to obtain U = k* u_c (y).

5.2.   An example calibration-based measurement model for a laser-based OIRS analyser

Depending on the used analyser and the available calibration material, different measurement models are possible to provide robust estimates of isotopic compositions. Those methods differ in the calibration approach and in the way well/poorly understood effects are included in the model. As an example, here we present a measurement model based on a two-point isotope ratio calibration. In this example model, we include empirical corrections for amount fraction and matrix gas effects as well understood effects. Poorly understood effects are included in this model by an additive random variable. This model might be appropriate for some OIRS analysers, but the underlying assumptions and in particular empirical correction functions must be properly tested for individual analysers.

- **A commonly used basic measurement model: Linear calibration in delta scale**
  The basic measurement model for a two-point linear calibration in delta scale is simply the analysis function[8] between the instrument realization, also called raw value δ_{raw} and the value assigned to the measurand, also called calibrated value δ_{cal}:
  $$\delta_{cal} = m * \delta_{raw} + b \qquad (10)$$
  With the slope $m = \frac{\delta_{cal,2}-\delta_{cal,1}}{\delta_{raw,2}-\delta_{raw,1}} =$ and the intercept $b = \frac{\delta_{cal1}\delta_{raw,2} - \delta_{cal2}\delta_{raw,1}}{\delta_{raw,2}-\delta_{raw,1}}$ defined by the calibrated/raw δ values of the calibration material.

- **Including amount fraction dependencies as well understood effects**
  For some instruments, the amount fraction dependency can be corrected for by measuring and applying an empirical amount-fraction correction function *f* to the instrument realization of the measurand δ_{raw}.
  $$\delta_{raw,\chi-cor} = \delta_{raw} + f\left(\chi_{CO2,raw}\right)$$
  Here we apply the concentration dependency correction as an additive function $f(\chi_{CO2,raw})$. The choice of this correction as an additive term might seem arbitrary at a first glance, however, it can be related to the Beer-Lambert law by assuming a certain analysis function for individual isotopologues, as discussed in section 3.5.2, see also (Griffith 2012, Griffith 2018). Thus an additive correction function $f\left(\chi_{CO2,raw}\right) = p_a/\chi_{CO2,raw} + p_b\,\chi_{CO2,raw}$ might be an appropriate choice for many analysers. Here, we allow even more complicated additive correction functions to provide a more general measurement model.

- **Including gas matrix effects as well understood effects**
  If rigorously tested, matrix gas dependencies might be incorporated into the measurement model via empirical correction terms. that depend. Here we include matrix effects as additive correction terms $g\left(\Delta\chi_{O2}\right)$ and $h\left(\Delta\chi_{Ar}\right)$ that depend only on the difference $\Delta\chi_j := \chi_{j,ref} - \chi_{j,sample}$ for species j.
  $$\delta_{raw,cor} = \delta_{raw} + f(\chi_{CO2,raw}) + g\left(\Delta\chi_{O2}\right) + h\left(\Delta\chi_{Ar}\right) \qquad (11)$$
  Please note that the applicability of such orthogonal additive correction functions that only depend on $\Delta\chi_j$ and are in particular independent on each other and do not depend on the δ value and the

---

[8] The analysis function is defined as the inverse of the calibration curve.

CO2 amount fraction $\chi_{CO2,raw}$ can be questioned and needs to be tested for an individual instrument. However, assuming an additive correction function might be a good choice, because in a first order approximation, an additive correction function for δ values corresponds to a relative correction factor for isotopic ratios.

One possible correction function, that might be appropriate for small differences $\Delta\chi_j$ would be a linear relationship e.g., $g\left(\Delta\chi_{O2}\right) = \Delta\chi_{O2}S_{O2}$ with constant S_{O2}.

However, for any empirically based correction function would need to be rigorously tested. Further, depending on the instrument and its fitting procedures, there might be options to minimize the magnitude of the matrix gas dependency by improving the fit and/or reducing $\Delta\chi_j$ throughout the measurement series might reduce uncertainties.

- **Including poorly understood effects**
  Poorly understood effects can yield variability and bias of the measurand, that is not captured by the uncertainty estimation of used model function. Such effects might be related to sample preparation, instrument instability (in flow rate, temperature, or pressure) that is not captured by the calibration strategy or other unknown effects that increase the uncertainty of the data. The statistical part of poorly understood effects can be included into the measurement model by including a random variable that does not change the expectation of the output but contributes to the uncertainty. Including a random variable can be done additive (with an expectation value of 0) or multiplicative (with an expectation value of 1). An additive random variable is appropriate when the uncertainty contribution of the random effect can be assumed to be independent of the value of the measurand while a multiplicative term is appropriate when the standard deviation can be assumed to be proportional to the measurand.
  For our example model, we include poorly understood effects via an additive random variable P with an expectation value of 0. This choice is not the only possible choice and depending on the used instrument and fitting routine, alternative ways to include poorly understood effects can be more appropriate. However, including poorly understood effects as an additive random variable is motivated by the fact that a multiplicative random variable would assume a linear relationship between the δ value and the magnitude of the effect. This assumption might not be applicable in many cases, because the δ value defined as the relative deviation from an arbitrary reference standard is an artificial quantity and we do not see a physical reason for the random effect to scale with this artificial quantity.

- **Resulting extended measurement model**
  Including the above mentioned well and poorly understood effects to the basic measurement model as shown above, yields the following extended measurement model:

$$\delta_{cal} = m_{exp} * (\underbrace{\delta_{raw} + f\left(\chi_{CO2}\right) + g\left(\Delta\chi_{O2}\right) + h\left(\Delta\chi_{Ar}\right)}_{\delta_{raw,cor}}) + b_{exp} + P \quad (12)$$

With slope $m_{exp} =: \frac{\delta_{cal,2}-\delta_{cal,1}}{\delta_{raw,cor,2}-\delta_{raw,cor,1}} =$ and intercept $b_{exp} = \frac{\delta_{cal1}\delta_{raw,cor,2}-\delta_{cal2}\delta_{raw,cor,1}}{\delta_{raw,cor,2}-\delta_{raw,cor,1}}$ of the expanded analysis function based on $\delta_{raw,cor,i}$, the amount fraction and matrix-effect corrected instrument realization of a calibration material i and the value assigned to the calibration material $\delta_{cal,i}$.

5.3. Uncertainty assignment and propagation for the example measurement model

For our example measurement model, we assume non-correlated gaussian distribution for the different input quantities. The expectation values and standard uncertainties (defined via standard deviation) for the

---

Used model function:

$$\delta_{cal} = m_{exp} * (\underbrace{\delta_{raw} + f(\chi_{CO2}) + g(\Delta\chi_{O2}) + h(\Delta\chi_{Ar})}_{\delta_{raw,cor}}) + b_{exp} + P$$

With $m_{exp} := \frac{\delta_{cal,2} - \delta_{cal,1}}{\delta_{raw,cor,2} - \delta_{raw,cor,1}}$ and $b_{exp} := \frac{\delta_{cal1}\delta_{raw,cor,2} - \delta_{cal2}\delta_{raw,cor,1}}{\delta_{raw,cor,2} - \delta_{raw,cor,1}}$

Operating conditions:

Cell pressure: _____
Flow rate: _____
Averaging time: _____
Fit settings: (baseline fit, pressure broadening, …)

Input variable	Expectation value	Standard uncertainty	Source of information	Type A/B
Instrument realization of the measurand δ_{raw} (also for calibration material $\delta_{raw,i}$)	$x_1 = \delta_{raw} := \overline{\delta_{raw}(t)}$	$\sigma^2(\overline{\delta_{raw}(t)}) = \frac{\sigma^2(\delta_{raw}(t))}{\sqrt{N}}$	Instrument data	A
Instrument realization of χ_{CO2}	$x_2 = \chi_{CO2} := \overline{\chi_{CO2}(t)}$	$\sigma^2(\overline{\chi_{CO2}(t)}) = \frac{\sigma^2(\chi_{CO2}(t))}{\sqrt{N}}$	Instrument data	A
Amount fraction correction factor f	$x_3 = f$	$\sigma^2(f)$	Experiment on variability of f	B
Value assigned to reference material $\delta_{cal,i}$	$x_4 = \delta_{cal,i}$	$\sigma^2(\delta_{cal,i})$	Certificate of reference material	B
Amount fraction of reference gas for species j (e.g., O$_2$ or Ar)	$x_6 = \chi_{j,ref}$	$\sigma^2(\chi_{j,ref})$	Certificate of reference air	B
Amount fraction of the sample gas of species j (e.g. O$_2$ or Ar)	$x_7 = \chi_{j,sample}$	$\sigma^2(\chi_{j,sample})$	Certificate (if applicable), measurement/ literature data	B
Poorly understood effects P	0	$\sigma^2(P)$	Estimated by repeatability & other uncertainty contributions	A

Table 6 Assignment of uncertainties to input parameters of the example model function

different input variables are summarized in Table 6.[9] Further, the analysers uncertainty can generally depend on the instruments operating conditions such as CO$_2$ amount fraction, cell pressure, flow rate. Thus, the uncertainty analysis needs to include these conditions. The uncertainties of the different input parameters can be propagated to the output parameter by the law of propagation of uncertainty using a first- or higher-order Taylor series approximation to the model. For the used measurement model, the uncertainty propagation yields:

$$\sigma^2(\delta_{cal}) = \sum_{i=1}^{7}\left(\sigma^2(x_i) * \left(\frac{\partial\delta_{cal}}{\partial x_i}\right)^2\right) + \sigma^2(P) \qquad (13)$$

---

[9] Please note, that the slope m and the intercept b are no uncorrelated input variables. They are functions of their uncorrelated input variables (e. g. $\delta_{cal,i}$ and $\delta_{ref,i}$) as shown in their definitions.

With x_1 to x_7 defined in Table 6. An individual analysis of the different addends in Eq. $\sigma^2 \left(\delta_{cal} \right) = \sum_{i=1}^{7} \left(\sigma^2(x_i) * \left(\frac{\partial \delta_{cal}}{\partial x_i} \right)^2 \right) + \sigma^2(P)$ (13) provides insights into the dominance of different uncertainty contributions.

Estimating the poorly understood effects P is a challenging task and it would be beyond the scope of this good practice guide to provide a detailed discussion on how this should be done. As a first approach, we suggest estimating the magnitude of P by using a repeatability experiment and removing those sources of uncertainty that were not kept constant during the repeatability experiment:

$$\sigma^2(P) = \sigma^2 \left(\delta_{rep} \right) - \sum_{i=1,2,3} \left(\sigma^2(x_i) * \left(\frac{\partial \delta_{cal}}{\partial x_i} \right)^2 \right) \qquad (14)$$

6. Traceability

When interpreting measurement data, it is important to provide information about the traceability chain. This is necessary for comparability when different datasets are to be compared as they need to be traced back to a common origin. Thus, if datasets are used in an internal laboratory comparison, it might be appropriate to provide traceability to internal working standards. However, if the results are meant to provide information about the isotopic composition on a global scale, it is recommended to provide full traceability to a scale used by the atmospheric measurement community – such as NBS19 (but be aware that this reference material might be replaced in future). Most preferable would be SI traceability – providing SI traceability for measurements of isotopic CO_2 composition is the subject of current metrological research and will hopefully be possible in future.

Figure 5 Example of a traceability chain for a set of experiments performed at PTB. These experiments aim at measuring the isotopic composition of CO₂ and interpreting it on a global scale. The internal working standards are measured at the BGL-IsoLab of the Max Planck Institute for Biogeochemistry in Jena, that provides traceability to NBS19, the scale-defining anchor for the VPDB scale.

7. Literature

A. D. McNaught and A. Wilkinson, 1997. IUPAC Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). IUPAC, Blackwell Scientific Publications, Oxford.

Bielska, K., Kyuberis, A.A., Reed, Z.D., Li, G., Cygan, A., Ciury\lo, R., Adkins, E.M., Lodi, L., Zobov, N.F., Ebert, V., Lisak, D., Hodges, J.T., Tennyson, J., Polyansky, O.L., 2022. Subpromille Measurements and Calculations of CO (3–0) Overtone Line Intensities. Phys. Rev. Lett. 129, 043002. https://doi.org/10.1103/PhysRevLett.129.043002

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML, 2008. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. JCGM 100:2008 GUM 1995 with minor corrections Evaluation of measurement data — Guide to the expression of uncertainty in measurement

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML, 2012 International vocabulary of metrology - Basic and general concepts and associated terms (VIM). JCGM 200:2012 International vocabulary of metrology – Basic and general concepts and associated terms (VIM) 3rd edition 2008 version with minor corrections

Brewer, P.J., Kim, J.S., Lee, S., Tarasova, O.A., Viallon, J., Flores, E., Wielgosz, R.I., Shimosaka, T., Assonov, S., Allison, C.E., Veen, A.M.H. van der, Hall, B., Crotwell, A.M., Rhoderick, G.C., Hodges, J.T., Mohn, J., Zellweger, C., Moossen, H., Ebert, V., Griffith, D.W.T., 2019. Advances in reference materials and measurement techniques for greenhouse gas atmospheric observations. Metrologia 56. https://doi.org/10.1088/1681-7575/ab1506

D. W. Allan, 1966. Statistics of atomic frequency standards. Proceedings of the IEEE 54, 221–230. https://doi.org/10.1109/PROC.1966.4634

Fleisher, A.J., Yi, H., Srivastava, A., Polyansky, O.L., Zobov, N.F., Hodges, J.T., 2021. Absolute 13C/12C isotope amount ratio for Vienna PeeDee Belemnite from infrared absorption spectroscopy. Nature Physics 17, 889–893. https://doi.org/10.1038/s41567-021-01226-y

Griffith, D.W.T., 2018. Calibration of isotopologue-specific optical trace gas analysers:\hack\break a practical guide. Atmospheric Measurement Techniques 11, 6189–6201. https://doi.org/10.5194/amt-11-6189-2018

Griffith, D.W.T., Deutscher, N.M., Caldow, C., Kettlewell, G., Riggenbach, M., Hammer, S., 2012. A Fourier transform infrared trace gas and isotope analyser for atmospheric applications. Atmospheric Measurement Techniques 5, 2481–2498. https://doi.org/10.5194/amt-5-2481-2012

Hill-Pearce, R.E., Hillier, A., Mussell Webber, E., Charoenpornpukdee, K., O'Doherty, S., Mohn, J., Zellweger, C., Worton, D.R., Brewer, P.J., 2021. Characterisation of gas reference materials for underpinning atmospheric measurements of stable isotopes of nitrous oxide. Atmospheric Measurement Techniques 14, 5447–5458. https://doi.org/10.5194/amt-14-5447-2021

Socki, R., Jacksier, T., VanPlantinga, A., 2019. Carbon and Oxygen Isotope Fractionation of Pressurized CO_2 in Dual and Single Phase Cylinders, in: AGU Fall Meeting Abstracts. pp. V11D-0105.

Socki, R., Matthew, M., McHale, J., Sonobe, J., Isaji, M., Jacksier, T., 2020. Enhanced Stability of Stable Isotopic Gases. ACS Omega 5, 17926–17930. https://doi.org/10.1021/acsomega.0c00839

Socki, R.A., Jacksier, T., 2021. Isotope Fractionation of Gaseous-, Supercritical-, and Dual-Phase Carbon Dioxide in Pressurized Cylinders Exposed to Variable Temperatures. ACS Omega 6, 8625–8629. https://doi.org/10.1021/acsomega.1c00474

Steur, P.M., Scheeren, H.A., Nelson, D.D., McManus, J.B., Meijer, H.A.J., 2021. Simultaneous measurement of \chem\delta^13C, \chem\delta^18O and \chem\delta^17O of atmospheric \chemCO_2 – performance assessment of a dual-laser absorption spectrometer. Atmospheric Measurement Techniques 14, 4279–4304. https://doi.org/10.5194/amt-14-4279-2021

Wen, X.-F., Meng, Y., Zhang, X.-Y., Sun, X.-M., Lee, X., 2013. Evaluating calibration strategies for isotope ratio infrared spectroscopy for atmospheric $^{13}CO_2$ / $^{12}CO_2$ measurement. Atmospheric Measurement Techniques 6, 1491–1501. https://doi.org/10.5194/amt-6-1491-2013

Werle, P., Mücke, R., Slemr, F., 1993. The limits of signal averaging in atmospheric trace-gas monitoring by tunable diode-laser absorption spectroscopy (TDLAS). Applied Physics B 57, 131–139. https://doi.org/10.1007/BF00425997