

Supplemental Material: Revision of the WMO/GAW CO₂ Calibration Scale

S1. Notes on historical manometric CO₂ determinations and corrections applied

S1.1 Historical manometric records

Manometric CO₂ mole fractions from 1996-2017 were determined based on the maximum CO₂ mole fraction calculated after the gas in the small volume was allowed to warm to oven temperature. Using a database of manometric results as reference, we examined raw data files and flagged data records when a) the CO₂ maximum could not be clearly identified (not common), b) the data record was not sufficient to determine a loss rate (e.g. too little data), c) raw files corresponding to a result in our database could not be found, or d) data appeared abnormal, e.g. temperatures not converging or showing abnormal variability. We were able to recover and apply corrections to 93% of the original data records.

Most manometer runs show one maximum in the calculated CO₂ record. However, some manometer runs from 1998 and 2004 show two peaks (Fig. S1). This secondary peak could be related to H₂O desorbing from surfaces in the small volume. In these cases, we calculated the loss correction using the time associated with the first CO₂ peak, and the slope from the section of data after the second peak.

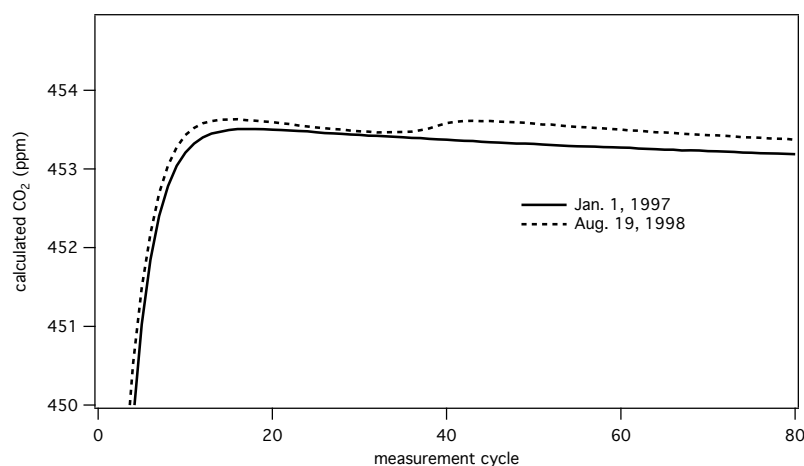


Figure S1: CO₂ mole fraction calculated from the pressure of gas in the small volume and oven temperature measured at various locations within the oven, as function of time (measurement cycle), where each cycle is ~60s. During the final extraction step, the small volume is cooled using a small dewar of liquid nitrogen. The dewar is removed after all CO₂ has been collected in the small volume. As CO₂ sublimates and equilibrates to the oven temperature, the calculated CO₂ reaches a maximum 15-20 minutes (~16 measurement cycles in this case) after the L-N₂ was removed. Shown are manometer data from two different runs of primary standard AL47-107. The January 1997 run shows the typical CO₂ maximum, followed by a slow decline. The August 1998 run shows a second peak around measurement cycle 40. For consistency, we applied corrections to existing CO₂ data based on timing of the first CO₂ peak.

S1.2 CO₂ Loss Rate

We made corrections for CO₂ loss assuming a linear loss rate. Extended data records indicate that the loss rate decreases with time (Fig. S2). However, in the short term, the initial loss rate is approximately constant and can be approximated by a linear function (Fig. S2 inset).

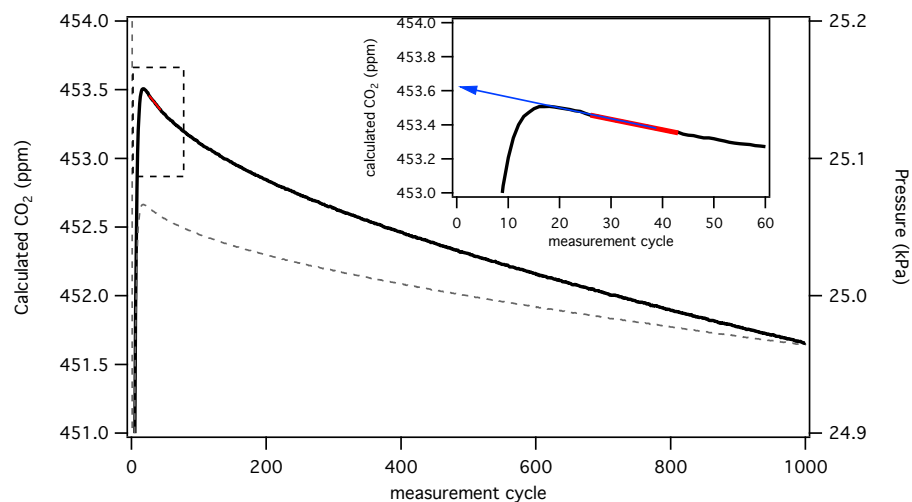


Figure S2: Extended manometric run showing the non-linear decline of CO₂ in the small volume (solid line) and the pressure in the small volume (dashed line) (temperature is nearly constant after cycle 20). The calculated CO₂ loss rate is higher at first, and then gradually slows. The inset shows a close-up view of the first 60 measurement cycles (~30 s per cycle). While the true CO₂ loss rate is non-linear, the loss rate is approximately linear in the vicinity of the maximum CO₂.

Fig. S3 shows the results of a pure CO₂ injection test, in which pure CO₂ was introduced into the small volume through an auxiliary port, without the cryogenic extraction step. The goal was to observe CO₂ loss upon immediate injection of CO₂ (without a large temperature change). This period is not observable in most data files since the small volume temperatures and pressure are changing rapidly during the initial minutes. Fig. S3 suggests that the loss of CO₂ begins immediately, and is non-linear. Linear extrapolation from around the time of the typical CO₂ peak to the beginning of the record (~100 s) results in a CO₂ result that is ~0.2 ppm lower than observed from the CO₂ injection experiment. However, the conditions of this experiment are not the same as those during an extraction procedure. With direct injection, surfaces, including O-rings, in the manometer are exposed to higher CO₂ pressures and temperatures compared to when CO₂ is extracted from air. Thus, we consider the 0.2 ppm underestimate suggested from these data (Fig. S3) an upper limit. These experiments prompted the preparation of gravimetric CO₂-in-air standards as a way to independently assess the magnitude of the proposed scale correction.

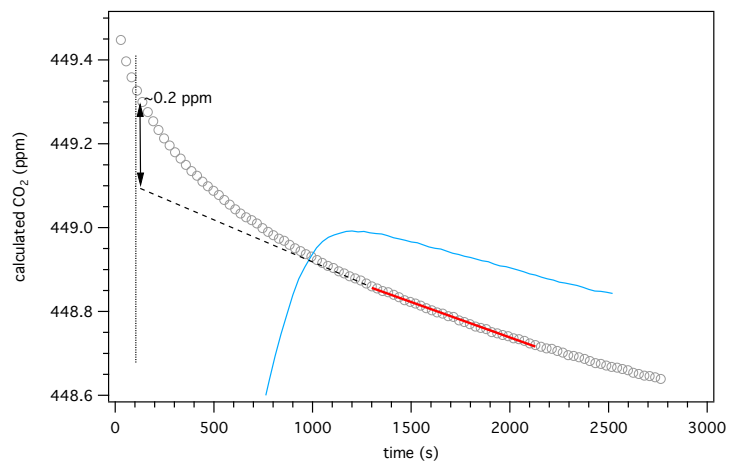
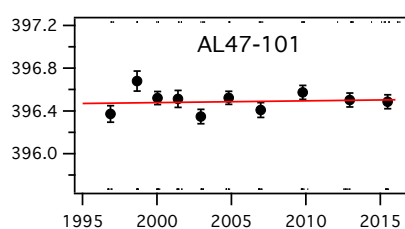
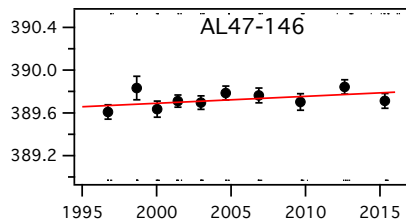
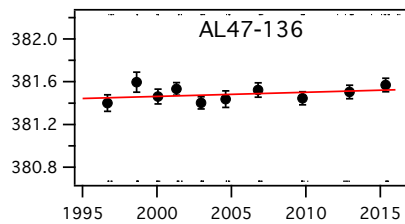
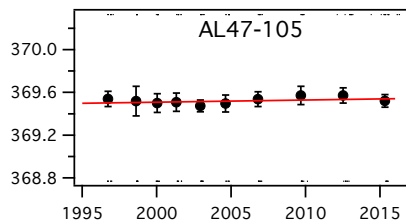
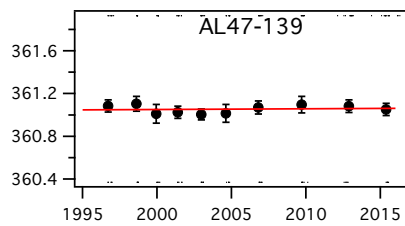
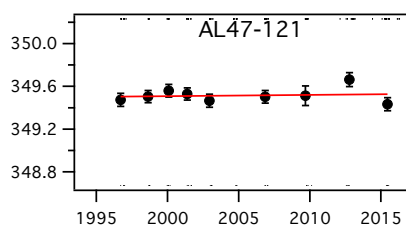
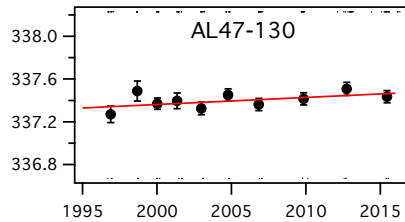
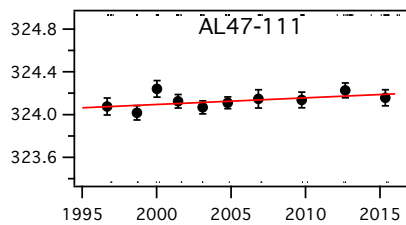
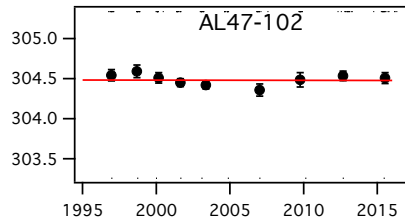
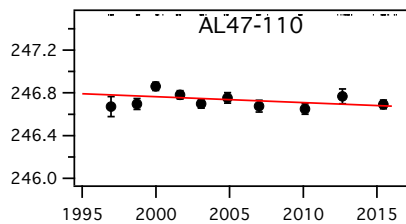


Figure S3: Test in which pure CO₂ was introduced into the small volume directly, without a cryogenic extraction step. Open symbols denote the CO₂ mole fraction that would correspond to this sample. The red line is a linear fit to the open symbols. The blue line is a representative extraction of a 450 ppm CO₂ sample, in which the purified CO₂ starts cold, and warms over time, showing the characteristic CO₂ maximum.

S1.3 Drift Assessment



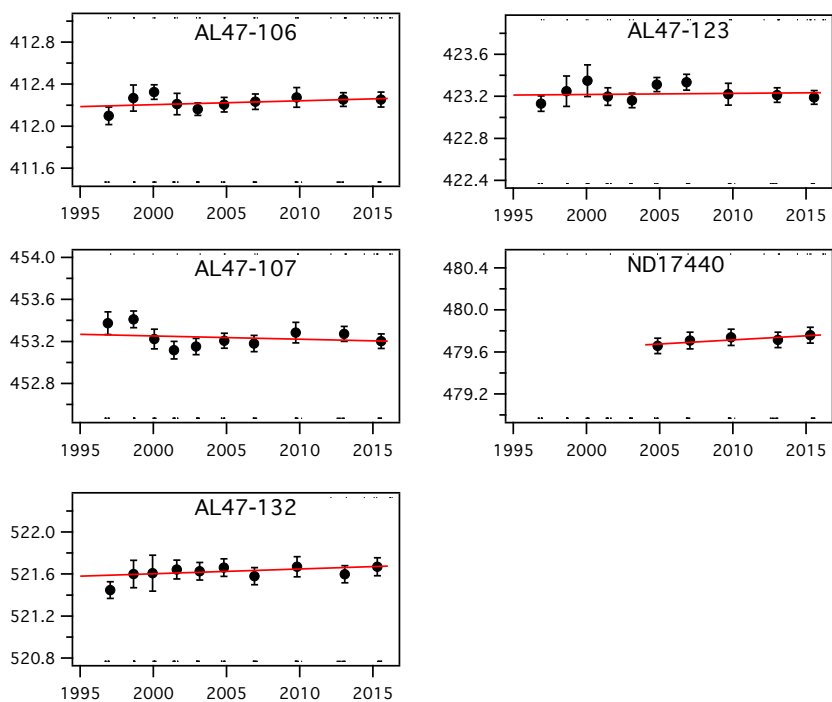


Figure S4: Average manometric results for each episode (1-sigma error bars), and associated weighted linear fits versus time (solid lines).

Table S1: Average mole fractions and drift rates (ppm decade^{-1}) for 15 primary standards with measurement histories spanning at least 10 years. Drift rates were determined from a weighted linear fit to manometer data, averaged by episode, over N_{ep} episodes. Uncertainties are 95% C.L.

Cylinder	Avg. (x2019)	Drift Rate	Uncertainty	N_{ep}
AL47-110	246.72	-0.0543	0.0623	10
AL47-102	304.50	-0.0012	0.0882	9
AL47-111	324.13	0.0616	0.0921	10
AL47-130	337.40	0.0656	0.0798	10
AL47-121	349.52	0.0106	0.0802	9
AL47-139	361.05	0.0061	0.0757	10
AL47-105	369.52	0.0203	0.0904	10
AL47-136	381.49	0.0387	0.0851	10
AL47-146	389.73	0.0642	0.0890	10
AL47-101	396.50	0.0166	0.0869	10
AL47-106	412.23	0.0368	0.0961	10
AL47-123	423.22	0.0108	0.0980	10
AL47-107	453.26	-0.0314	0.0994	10
ND17440	479.72	0.0771	0.2816	5
AL47-132	521.61	0.0681	0.1130	10

S2. Uncertainty Analysis

S2.1 Temperature Measurement during CO₂ determination

Standard uncertainties reported by accredited laboratories for calibration of three platinum resistance thermometers (PRTs) are 0.0125 °C. The device used to measure resistance (Hart 1529) adds an additional 0.007 °C. Combining these in quadrature we get 0.014 °C for a single probe. We use the average temperature determined from 2-3 probes placed in the vicinity of the large and small volumes to determine T. Combining the uncertainty from 3 probes with the typical standard deviation observed during the measurement we obtain:

$$uT_{air} = \sqrt{\left(\frac{0.014}{\sqrt{3}}\right)^2 + 0.004^2} = 0.009 \text{ °C} \quad (s1)$$

$$uT_{co2} = \sqrt{\left(\frac{0.014}{\sqrt{2}}\right)^2 + 0.014^2} = 0.017 \text{ °C} \quad (s2)$$

uT_{CO2} is larger than uT_{air} because the oven must respond to a change in temperature (oven door open briefly for cryogenic trapping of CO₂ in the small volume).

S2.2 Pressure Measurement during CO₂ determination

Pressure is measured using a Paroscientific 6000-15A pressure sensor. The 6000-15A is calibrated in-house using a Ruska 2465 piston gauge. We apply a linear calibration (span, offset) to the pressure output (Fig S5). Secondary checks on the stability of the span have been determined relative to a secondary pressure standard (Fluke, RPM4) at 10 kPa increments from 10-90 kPa since 2015. Since 2008, the span has varied between 0.99977 and 0.99999. The offset is determined by reading the pressure at vacuum following initial evacuation of the manometer. The offset recorded at vacuum usually agrees within ~0.003 kPa with that determined from the intercept using a linear fit to the piston gauge data. From 2004-2012 the offset varied between 0.001-0.005 kPa. A larger offset has been observed (0.018-0.024 kPa) since 2014 following a calibration by the manufacturer.

The estimated accuracy of the piston gauge is 5 parts-per-million (ppm) at full scale (100 kPa), 6-8 ppm at mid-scale, and 13 ppm at 10 kPa (manufacture's specifications). In addition, we observe drift in the 6000-15A "offset", or zero, of 0.0005 kPa over a few hours. The repeatability of the pressure measurement over short time periods (a few minutes) is excellent, often <0.0002 kPa, 1-sigma. Our estimates of uP are mainly functions of the accuracy of the pressure calibration and stability of the device over the course of an extraction (3 h). It is also possible that small leaks or outgassing in the system would lead us to over-estimate P_{CO2} since P_{CO2} is always less than room pressure and the measurement of P_{CO2} takes ~20 minutes to complete. System leaks are not expected to contribute significantly to other pressure measurements because they are made without significant lag-time. Recent

experiments performed at vacuum and also with 30 kPa air in the small volume indicate out-gassing and/or leaks that induce pressure changes as high as 4 Pa hr⁻¹ in the absence of CO₂. A 4 Pa hr⁻¹ leak would lead to ~0.02 ppm overestimate of X_{CO_2} . Since we do not have similar data for all episodes, we include the potential for leaks in the uncertainty associated with P_{CO_2} . Combined estimates of uncertainty in pressure are as follows:

For P_{air} , typically 83 kPa:

uP_{air} components: piston gauge ($5 \cdot 10^{-6} \cdot P_{air}$), zero drift (0.0002 kPa), repeatability (0.0001 kPa)

$uP_{air} = 0.0007$ kPa

For P_{CO_2} , at 30 kPa:

uP_{CO_2} components: piston gauge ($8 \cdot 10^{-6} \cdot P_{CO_2}$), zero drift (0.0004 kPa), leak (0.0003 kPa), repeatability (0.0001 kPa)

$uP_{CO_2} = 0.0010$ kPa

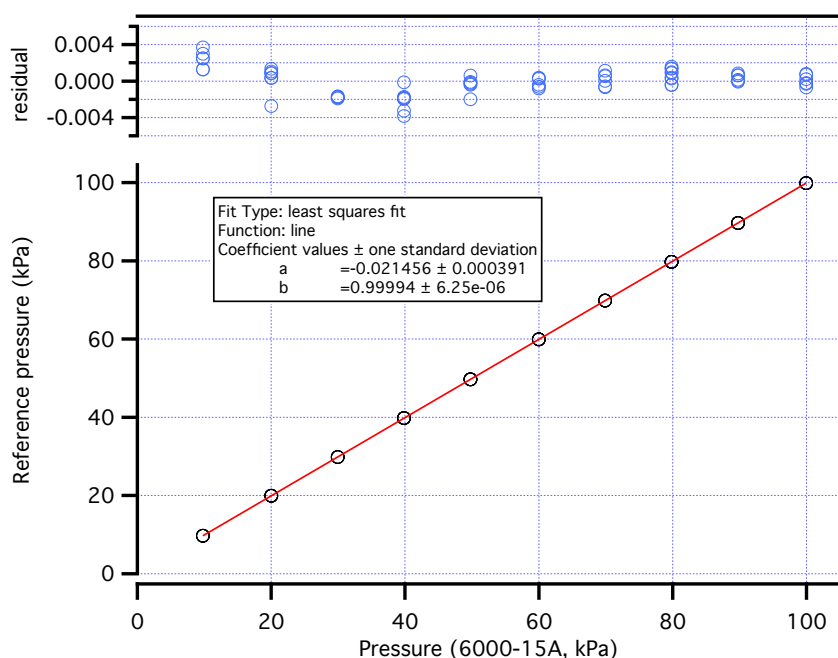


Figure S5: Example calibration of the main pressure sensor using the piston gauge, from 2014.

S2.3 Volume Ratio

The volume ratio, Φ , is calculated from a series of four gas expansions, using two auxiliary glass volumes. For each expansion, we calculate an intermediate volume ratio, r_i . Each intermediate volume ratio is determined from measurements of temperature, pressure, and second virial coefficients for the expanding gas (usually air or nitrogen) (Zhao et al., 1997).

The overall volume ratio is calculated as:

$$\Phi = r_1 * r_2 * r_3 * r_4 - r_1 * r_2 * r_3 + r_1 \quad (s3)$$

where r_i are the intermediate volume ratios from $i=1,4$ =gas expansion steps (see Zhao et al. (1997) for further description of the volume ratios r_i and the derivation of equation (s3)). The repeatability of volume ratio experiments is often very good. For example, for 24 volume ratio experiments run over 6 days in January 2017, the mean and standard deviation were 880.10 and 0.05 respectively. This repeatability is similar to that described in (Zhao et al., 1997). However, we need to consider uncertainties in T and P measurements, and compare Φ calculated using different gases in order to estimate the uncertainty (Figure S6b).

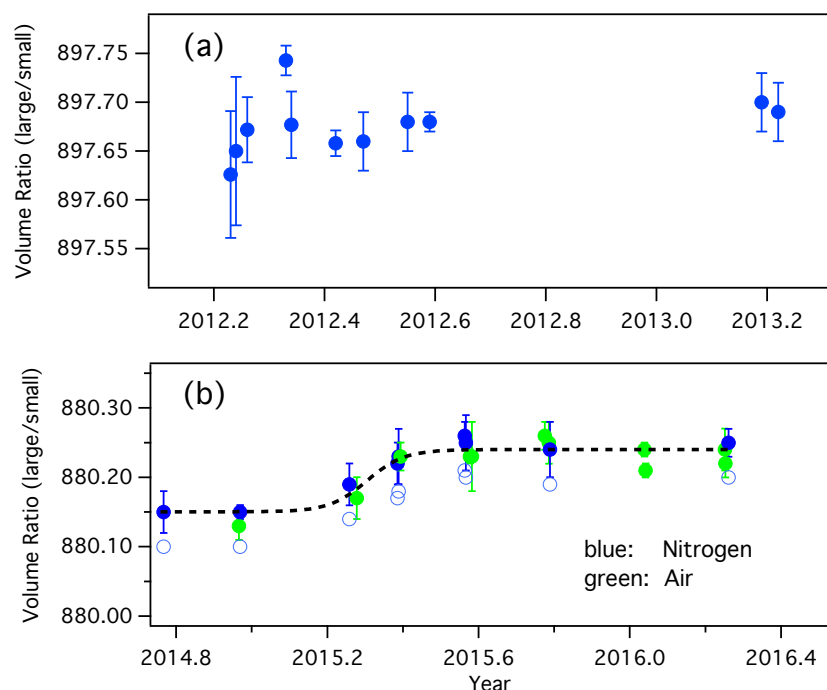


Figure S6: Results from volume ratio expansion experiments (a) 2012-2013 and (b) 2014-2016. Shown are the mean and standard deviation from four experiments performed in a given day, using different gases: CO₂-free air (green), and nitrogen (blue). Open circles in (b) show volume ratios calculated using an alternate second virial coefficient for N₂: -4.64 cm³/mol at 300 K (Dymond et al., 2002) instead of -5.18 cm³/mol (<https://web.archive.org/web/20190506031327/http://www.kayelaby.npl.co.uk/>). Note that the small volume was replaced in 2014, so the volume ratios in (a) and (b) are not expected to be the same. Volume ratios performed in 2019 using argon gas were found, on average, to be 0.018% lower than those shown in (b).

S2.3.1 Effect of uncertainties in temperature on volume ratio

We begin with the estimate for $u_{T_{\text{air}}}$ from 2.1 since the oven is not disturbed during a volume ratio experiment. On each expansion, we measure temperature as the average of two PRT probes placed in a central location in the oven, with an initial estimate for $u_T = 0.009$ °C, as before. However, during pressurization and expansion steps we

observe adiabatic heating and cooling, which could lead to temperature gradients in the glass volumes. During a normal volume ratio experiment, we allow seven minutes for equilibration prior to taking measurements. By placing 4-6 calibrated thermistors at various locations on each volume (22 in total), we observed variations in temperature ($\sigma = 0.02$ °C). Variations were largest during volume expansions 3 and 4 (corresponding to r_3 , r_4), and negligible on expansions r_1 and r_2 . Therefore, we include additional uncertainty for r_3 and r_4 .

$$uT (r_1 \text{ \& } r_2) = 0.01 \text{ } ^\circ\text{C} \quad (\text{s4})$$

$$uT (r_3 \text{ \& } r_4) = 0.02 \text{ } ^\circ\text{C} \quad (\text{s5})$$

S2.3.2 *Effect of uncertainties in pressure on volume ratio*

For pressure, we apply the same logic as in 2.2, except that we increase contributions from potentials leaks and zero drift because larger volumes are held at vacuum for ~ 1 hour on each experiment. We assume zero drift uncertainties of 0.0002 kPa, and leak contributions ranging from 0-0.0005 kPa. Repeatability is a small component (0.0001 kPa). Listed here are the components corresponding to uncertainties on each pressure measurement for the four successive volume expansions used to calculate the intermediate volume ratios in equation (s3) (respectively: manufacturer's specification, zero drift, leak potential). Each volume expansion starts at 80 kPa.

On filling to 80 kPa,	$5 \cdot 10^{-6} \cdot 80$, 0.0002, 0.0000;	$uP = 0.0006$ kPa.
r_1 , expansion to 19 kPa,	$8 \cdot 10^{-6} \cdot 19$, 0.0002, 0.0004;	$uP = 0.0006$ kPa
r_2 , expansion to 10 kPa,	$13 \cdot 10^{-6} \cdot 10$, 0.0002, 0.0004;	$uP = 0.0007$ kPa
r_3 , expansion to 18 kPa,	$8 \cdot 10^{-6} \cdot 18$, 0.0002, 0.0005;	$uP = 0.0008$ kPa
r_4 , expansion to 13 kPa,	$12 \cdot 10^{-6} \cdot 13$, 0.0002, 0.0005;	$uP = 0.0007$ kPa

S2.3.3 *Second virial coefficients*

Uncertainty in the second virial coefficient has a small impact on the calculated volume ratio (Figure S6b). We include an uncertainty component derived from testing with different gases (see 2.3.4).

S2.3.4 *Propagation of uncertainty for the Volume Ratio*

Estimated uncertainties for each expansion step are shown in Table S2. We made some assumptions about the correlations of uncertainties since the same pressure transducer and temperature sensors are used for each measurement of P and T. For pressure, we assume high correlation between P_1 and P_2 ($r=0.9$). For T_1 and T_2 , we assume a correlation coefficient of 0.5, since some of the uncertainty is related to adiabatic compression/expansion and would be uncorrelated.

Table S2: Parameters of a typical series of four sequential volume expansions ($i=1,4$), each from $P_1 \Rightarrow P_2$, used to determine the volume ratio.

i	P_1 (kPa)	uP_1 (kPa)	P_2 (kPa)	uP_2 (kPa)	r_i	ur_i
1	80	0.0006	19	0.0006	4.15	0.00019
2	80	0.0006	10	0.0007	7.95	0.00057
3	80	0.0006	18	0.0008	4.53	0.00029
4	80	0.0006	13	0.0007	6.85	0.00045

To estimate the total uncertainty in the volume ratio, we used the NIST metRology package (<https://www.nist.gov/programs-projects/metrology-software-project>) to evaluate equation s3 using parameters and uncertainties in Table S2. The uncertainty calculated from parameters in Table S2 is 0.115 for a volume ratio of 880. Adding repeatability (0.032) in quadrature we obtain 0.119. Additional terms associated with PRT placement (0.01) and different gases (0.043) are included in a further step. The oven contains three PRTs but only two are used for the calculation of VR, and we get slightly different results using different combinations of PRTs. We also calculate different volume ratios using different gases (air, nitrogen, and argon). These differences could be partly related to uncertainty in virial coefficients, but could also involve surface interactions. For this component we assume a uniform distribution over the range 0.15 ($0.15/2/\sqrt{3} = 0.043$). Summing all terms in quadrature we obtain: $u\Phi = \sqrt{0.115^2 + 0.032^2 + 0.01^2 + 0.043^2} = 0.127$ for a volume ratio of 880.1, or 0.014%. This uncertainty estimate is about 40% larger than that reported by Zhao and Tans (2006). Since the last two uncertainty components are meant to capture elements that would be common to all volume ratio experiments, these are not included in the uncertainty applied to the drift assessment.

A series of volume ratio experiments were carried out prior to and during each episode to verify the volume ratio. Since the physical configuration of the NOAA manometer has remained basically the same since 1996, although with some changes related to valves and pressure sensors, we assume that the relative uncertainty of the volume ratio is applicable to all volume ratios used since 1996.

S2.3.5 Water vapor

While we do not have measurements of the water vapor content in the extracted CO_2 , we can estimate $X_{\text{H}_2\text{O}}$ from the temperature of the traps. Both trap 1 and trap 2 are typically held at -67°C during the separation step. The vapor pressure of H_2O over ice at -67°C is 0.406 Pa (Huang, 2018). The estimated volume of trap 1 is 40 cm^3 . For a 400 ppm CO_2 sample, 0.1 mmoles of CO_2 would be present in the trap, corresponding to a pressure of 6.5 kPa. The fraction of gas in trap 1 that is water vapor is then $0.406/(6500+0.406) = 6.25 \cdot 10^{-5}$. As this gas is trapped in the small volume and then warmed, water vapor would exert a pressure equivalent to $(6.25 \cdot 10^{-5})(400) = 0.025\text{ ppm CO}_2$. There is no active control on the trap temperatures (we use an alcohol bath), and the typical range is -65°C to -70°C . For this range of temperatures, water vapor would correspond to 0.033 to 0.016 ppm CO_2 , respectively. We therefore include a conservative uncertainty of 0.03 ppm due to water vapor (Table S3).

S2.3.6 Purity Assessment

The primary function of the separation steps is to remove H₂O from the extracted CO₂. The purified CO₂ introduced into the small volume contains N₂O and trace amounts of other gases. Considering the major constituents in air and their boiling points under the conditions at which CO₂ is trapped, Zhao et al. (1997) concluded that a correction for N₂O is sufficient to account for impurities in the purified CO₂. We go one step further here by verifying, through analysis, that the purified CO₂ is, indeed, highly pure and that additional purity corrections are not needed.

We used the manometer to extract CO₂ from a 380 ppm air sample. At the end of a normal manometer run, we transferred the purified CO₂, first to a stainless steel tube (5 mL volume with stainless steel metal bellows valve) and then to a 2.3 L stainless steel flask with a stainless steel metal bellows valve. We then added ~0.24 MPa (35 psia) UHP-grade nitrogen to create a mixture with X_{CO_2} at approximately 380 ppm, the same as that of the original air. We analyzed this mixture by GC-MS, GC-FID (Dlugokencky et al., 2005) and GC-ECD (Hall et al., 2011). We confirmed that gases likely to be trapped in the extraction step (nitrous oxide, ethane, propane, some chlorofluorocarbons) are present in the purified CO₂ sample. The combined mole fraction of all gases measured in the flask, excluding CO₂, N₂O, and H₂O, was 6 parts per billion (ppb). Of this 6 ppb, we found ~3.4 ppb Xe, 0.8 ppb ethane, 0.5 ppb CCl₂F₂ (CFC-12), 0.2 ppb CFCI₃ (CFC-11) and trace amounts of other halogenated gases. Had Xe been quantitatively trapped and retained, we would have found ~87 ppb, which would then require a correction. CH₄ was not detected in the purified CO₂ sample, confirming that CH₄ is not trapped during the extraction process. We did not attempt to measure H₂O, krypton, argon, and oxygen since these would either not be trapped at the pressure of the traps (~4 kPa), or would likely be present at very low levels and would be difficult to measure.

S2.4 Manometric uncertainty

To arrive at the total uncertainty associated with a manometric measurement of CO₂, we use the NIST metRology package and equation (s6),

$$X_{CO_2} = (\phi^{-1}) \frac{P_{CO_2} T_{air}}{P_{air} T_{CO_2}} (1 + A_1 - A_2) - X_{N_2O} + X_{virial_{correction}} + X_{loss_{correction}} - X_{H_2O} \quad (s6)$$

$$A_1 = \frac{P_{air} \beta_{air}}{RT_{air}}$$

$$A_2 = \frac{P_{CO_2} \beta_{CO_2}}{RT_{CO_2}}$$

where T and P are the temperatures and pressures of the large volume (air) and small volume (CO₂), β_{air} and β_{CO_2} are second virial coefficients, R is the gas constant, Φ is the volume ratio, and X_{N_2O} is the mole fraction of N₂O in the air sample (see main document). We do not calculate an uncertainty for each individual measurement. Instead we use

typical values and uncertainties for variables in (s6), and calculate uX_{CO_2} over a range of mole fractions. This is justified because the measurement conditions do not vary significantly between measurements.

For each measurement in the database, we calculate a loss correction and a virial correction. The uncertainty associated with $X_{virial_correction}$ is ~ 0.005 ppm. For $X_{loss_correction}$, we estimate the uncertainty in loss rate at 10% for most measurements, and 20% for those exhibiting a second maximum. We assume that the time corresponding to peak CO_2 (t) is known to within one measurement cycle, and that the initial time (t_0) has an uncertainty of 2 minutes (2-4 measurement cycles). Together, the uncertainty associated with the loss correction is $\sim 12\%$ for most measurements, and 25-40% when a second CO_2 maximum was observed. Although a potential bias resulting from a non-linear adsorption at the beginning of the experiment was observed in separate tests (fig. S3), the magnitude of this potential bias could not be quantified experimentally under conditions consistent with manometric experiments.

Propagating uncertainties in Table S3 through equation s6, we estimate the uncertainty associated with a single manometric measurement to be ~ 0.070 ppm at 400 ppm. the largest component of which is uncertainty in the volume ratio, accounting for nearly half. The average standard error across all manometric episodes and all primaries is 0.044 ppm. We include this term to account for the average reproducibility of the manometric method. Summing in quadrature the average manometric reproducibility and the uncertainty associated an individual manometric measurement (0.074 ppm), we estimate the standard uncertainty of the scale, prior to scale transfer, to be 0.086 ppm at 400 ppm.

Table S3: Typical values and uncertainties for used to estimate the scale uncertainty at 400 ppm.

Variable	Typical value	Standard uncertainty	unit	Approx. Relative contribution (%)
P_{air}	83	0.0007	kPa	1
P_{CO_2}	29	0.001	kPa	3
T_{air}	310	0.01	K	3
T_{CO_2}	310	0.017	K	7
Φ (vol. ratio)	880.10	0.127	dimensionless	46
β_{air}	-5.87	0.2 ^(note 1)	cm ³ /mol	< 1
β_{CO_2}	-112.8	0.2 ^(note 1)	cm ³ /mol	< 1
$X_{virial_correction}$	0.03	0.005	ppm	1
$X_{loss_correction}$	0.14	0.02	ppm	6
X_{N_2O}	0.325	0.0005	ppm	<i>negligible</i>
X_{H_2O}	0 ^(note 2)	0.03	ppm	12
Total from above		0.074	ppm	
Reproducibility, (manometer)		0.044	ppm	21
Total manometric component		0.086	ppm	

¹ From Zhao et al. (1997); Sengers et al. (1971)

² We do not make a correction for water vapor, so its value is assumed to be zero.

S2.5 Scale transfer uncertainty

Scale transfer from primary to secondary and secondary to tertiary standards is currently performed using multiple

laser spectroscopic techniques (Tans et al., 2017). Differences among tertiary cylinders analyzed more than once over different time periods, spread across multiple suites of secondary standards as available, and multiple primary-secondary comparisons, captures the reproducibility of the scale transfer process. Tans et al. (2017) estimated this uncertainty contribution to be 0.01 ppm (1 sigma) based on a limited (approximately 1 year) evaluation time period. We now have 4+ years of measurements and provide further assessment here. While still short compared to the decadal timescales of the scale, we reassess this estimate since it is vital for the WMO GAW community.

We routinely measure two suites of target tanks (TT) at various frequencies to assess scale transfer stability at the tertiary level. The first set (short term TTs) are measured weekly to monthly, while the second set have been measured approximately annually following several initial measurements when the laser spectroscopic system was first developed. These long term TTs should last decades and provide information on the long-term stability of the scale (provided that long-term drift is either negligible or known). The long-term TTs are also measured periodically directly against the primary standards to guard against potential biases due to scale transfer effects. Table S4 shows average X_{CO_2} and standard deviations of both sets at the tertiary level on the laser spectroscopic system. Standard deviations were calculated both without correcting for drift and also by correcting for linear drift using unweighted fits to the data, even though assessing the significance of the drift requires knowledge of the reproducibility of the measurements (assuming 0.01 ppm reproducibility).

The average standard deviation of nine short term TTs is 0.009 ppm assuming none is drifting, and 0.007 ppm if we allow for linear growth in all. There is some evidence that the reproducibility of the system degrades slightly at higher mole fractions. Average standard deviations of 14 long-term TTs are 0.011 ppm and 0.010 ppm. The slightly worse performance for the long-term TTs may be related to fewer measurements and/or potential influence from regulators due to lower measurement frequency. Based on these two sets of TTs we continue to estimate the scale transfer uncertainty for the laser spectroscopic calibration system as 0.01 ppm (1-sigma).

Prior to the laser spectroscopic system, CO₂ value assignment was performed using NDIR analyzers. From 1995 to 2016 we used five NDIR analyzers from Siemens and Licor. The scale transfer uncertainty on the NDIR system(s) was previously estimated at 0.03 ppm on the X2007 scale (Tans et al., 2015). We verify this on X2019 using eight target tanks from the NDIR system that are still in service. These tanks have been measured routinely by NOAA for 15 to 32 years and continue to be measured annually or biennially on the laser spectroscopic system. We fit the full history for each TT, weighting each measurement by our estimates for scale transfer uncertainty. This is in contrast to the above section on the laser spectroscopic system where an unweighted fit to only measurements on the laser spectroscopic system were considered. Seven of the eight tanks appear stable relative to the scale over these time frames. The average standard deviation of the residuals of the NDIR measurements from a fit to the entire history for these cylinders is 0.028 ppm when allowing for linear drift in one cylinder. We therefore continue to recommend a scale transfer uncertainty of 0.03 ppm (1-sigma) for historical measurements on the NOAA NDIR calibration system.

The scale transfer uncertainty of CO₂ value assignments prior to NOAA becoming CCL in 1995 has been estimated

for internal use but is beyond the scope of this paper. We welcome interest in these earlier measurements and encourage users to contact the CCL.

One aspect of the scale transfer not represented by TT results is any impact that changing regulators would have since regulators are not typically removed from TT's to prevent damage to the cylinder valve fittings. For normal calibration services, a regulator is installed and conditioned following standard protocols (ref <https://www.esrl.noaa.gov/gmd/ccl/reg.guide.html>). Comparisons of pre- and post-deployment value assignments of standards used at NOAA sites, while complicated by drift issues during use, align with the expected reproducibility based on TT's. Regulators remain an issue requiring further investigations as the CCL attempts to improve calibration services.

Table S4: Analysis of various cylinders of air used to estimate scale transfer uncertainty (reproducibility) associated with the laser spectroscopic system.

Short-term target tanks	CO ₂ (ppm)	Std dev ¹ (ppm)	Std dev ² (ppm)	N
CC71624	356.8	0.006	0.006	122
CA05562	397.2	0.007	0.007	88
CB11127	393.2	0.011	0.007	171
CB12113	401.3	0.008	0.007	82
CC726899	407.5	0.007	0.006	15
CC302566	428.7	0.006	0.006	14
CB10826	456.0	0.011	0.008	118
CC726893	475.8	0.008	0.006	14
CB11860	604.4	0.017	0.012	79
Average		0.009	0.007	
Long-term target tanks				
CB11939	300.1	0.011	0.011	20
CB11942	351.0	0.012	0.011	13
CB11855	390.9	0.014	0.010	12
CC71563	395.0	0.006	0.006	55
CC105857	399.9	0.008	0.007	14
CB11975	400.3	0.012	0.008	14
CB11934	402.8	0.014	0.010	13
CB11951	406.4	0.014	0.012	13
CA05008	406.8	0.007	0.006	47
CB11842	409.1	0.013	0.012	12
CC71625	427.3	0.010	0.007	15
CB11943	452.4	0.006	0.006	14
CB11966	497.9	0.011	0.011	14
CB11957	574.2	0.018	0.016	15
Average		0.011	0.010	

¹ Standard deviation without drift correction.

² Standard deviation after applying linear drift correction.

Table S5: Analysis of various cylinders of air used to estimate scale transfer uncertainty (as reproducibility) associated with the historical NDIR calibration system. Note that CA05008 has an extensive history on both the NDIR and laser spectroscopic systems. The residuals here are for the NDIR system.

Long-term target tanks	CO ₂ (ppm)	year in service	Std. dev. ¹ (ppm)	N ³
CC91272	332.4	1991	0.019	48
CC71648	342.5	1994	0.044	36
CC71643	351.7	1988	0.028	55
CC71656	370.9	2000	0.026	18
CC1790	382.7	2003	0.028	21
CC71613 ²	382.9	1996	0.023	49
CA05008	406.8	2002	0.030	70
CA06136	416.9	2005	0.022	25
Average			0.028	

¹ Standard deviation for residuals from complete measurement history for NDIR measurements.

² Residuals for CC71613 are from a linear fit to account for growth of CO₂ over time. Without accounting for this growth, the standard deviations of the NDIR residuals would be 0.039 ppm.

³Total number of measurement episodes, including those on the laser spectroscopic system.

S2.6 Total uncertainty estimate for the X2019 scale

Adding scale transfer uncertainty (0.01 ppm) in quadrature to the typical manometric uncertainty we obtain an estimate of the total standard uncertainty associated with value assignment on the X2019 scale (Table S5). For expanded uncertainty (~95% Confidence Level) we multiply standard uncertainties by coverage factor k=2. The uncertainty due to differences in ¹³C-CO₂ and ¹⁸O-CO₂ between tertiary standards and reference standards is expected to be <0.005 ppm, and is ignored.

Table S6: Standard and expanded uncertainties (ppm) associated with X_{CO2} determination (scale X2019).

X _{CO2}	Standard Uncertainty	Expanded Uncertainty (k=2)
250	0.070	0.14
300	0.075	0.15
350	0.081	0.16
400	0.087	0.17
450	0.093	0.19
500	0.100	0.20
550	0.107	0.21
600	0.114	0.23

S3. Comparing primary standards (with manometric CO₂ determination) to gravimetrically-prepared standards

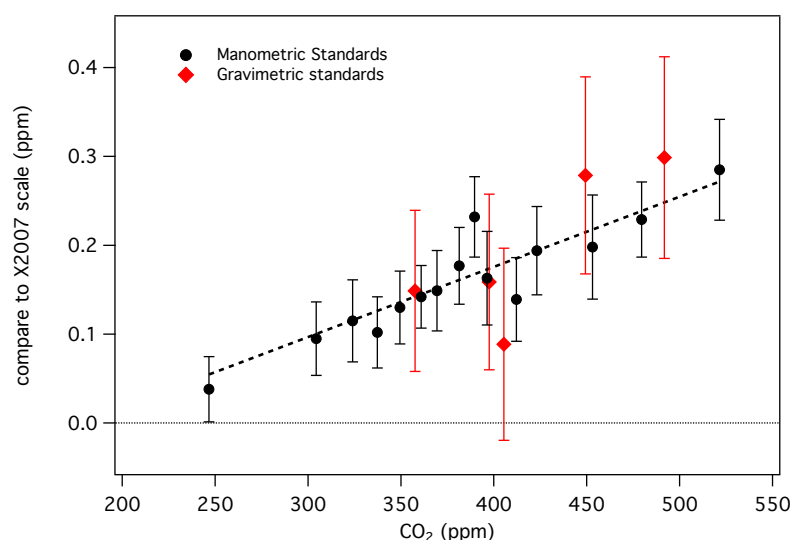


Figure S7: Differences between the X2019 scale and the X2007 scale derived from manometric standards (black symbols, dashed line), and differences between gravimetric standards and the X2007 scale (red symbols). Here, gravimetric standards were analyzed by laser spectroscopy and assigned CO₂ mole fractions on the X2007 scale. Error bars (1σ) are larger for gravimetric standards because they are independent of the manometric system and incorporate uncertainties in both methods, whereas manometric primary standards share common uncertainty components. The gravimetric standards show a mole-fraction dependent offset from X2007 similar to X2019 primary standards, suggesting that the linear correction for CO₂ loss in the manometer small volume is a reasonable approach.

References

- Dlugokencky, E. J., Myers, R. C., Lang, P. M., Masarie, K. A., Crotwell, A. M., Thoning, K. W., Hall, B. D., Elkins, J. W., and Steele, L. P.: Conversion of NOAA atmospheric dry air CH₄ mole fractions to a gravimetrically prepared standard scale, *Journal of Geophysical Research: Atmospheres*, 110, 10.1029/2005JD006035, 2005.
- Dymond, J. H., Marsh, K. N., Wilhoit, R. C., and Wong, K. C.: *Virial Coefficients of Pure Gases and Mixtures, Subvolume A: Virial Coefficients of Pure Gases, Numerical Data and Functional Relationships in Science and Technology, Group IV: Physical Chemistry*, edited by: Frenkel, M., and Marsh, K. N., Springer-Verlag, New York, 2002.
- Hall, B. D., Dutton, G. S., Mondeel, D. J., Nance, J. D., Rigby, M., Butler, J. H., Moore, F. L., Hurst, D. F., and Elkins, J. W.: Improving measurements of SF₆ for the study of atmospheric transport and emissions, *Atmos. Meas. Tech.*, 4, 2441-2451, 10.5194/amt-4-2441-2011, 2011.
- Huang, J.: A Simple Accurate Formula for Calculating Saturation Vapor Pressure of Water and Ice, *Journal of Applied Meteorology and Climatology*, 57, 1265-1272, 10.1175/JAMC-D-17-0334.1, 2018.
- Sengers, J. M. H., Klien, M., Gallagher, J. S.: *Pressure-Volume-Temperature Relationships of Gases Virial Coefficients*, 54 pp., 1971.

Tans, P. P., Crotwell, A. M., and Thoning, K. W.: Abundances of isotopologues and calibration of CO₂ greenhouse gas measurements, *Atmos. Meas. Tech.*, 10, 2669-2685, 10.5194/amt-10-2669-2017, 2017.

Tans, P. P., B. Hall., D. Kitzis, “History of WMO CO₂ X2007 scale: long-term reproducibility”, presented at the 18th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases, and Related Measurement Techniques (GGMT-2015), La Jolla, CA, 2015, <https://community.wmo.int/meetings/ggmt-2015>.

Zhao, C. L., Tans, P. P., and Thoning, K. W.: A high precision manometric system for absolute calibrations of CO₂ in dry air, *J. Geophys. Res.*, 102, 5885-5894, 1997.

Zhao, C. L., and Tans, P. P.: Estimating uncertainty of the WMO mole fraction scale for carbon dioxide in air, *J. Geophys. Res.*, 111, 10.1029/2005JD006003, 2006.