



CO₂, CO, and CH₄ measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts

A. E. Andrews¹, J. D. Kofler^{1,2}, M. E. Trudeau^{1,2,3}, J. C. Williams^{1,4}, D. H. Neff^{1,2}, K. A. Masarie¹, D. Y. Chao^{1,2}, D. R. Kitzis^{1,2}, P. C. Novelli¹, C. L. Zhao^{1,2}, E. J. Dlugokencky¹, P. M. Lang¹, M. J. Croswell^{1,2}, M. L. Fischer⁵, M. J. Parker^{6,7}, J. T. Lee⁸, D. D. Baumann⁹, A. R. Desai¹⁰, C. O. Stanier¹¹, S. F. J. De Wekker¹², D. E. Wolfe¹, J. W. Munger¹³, and P. P. Tans¹

¹NOAA Earth System Research Laboratory, Boulder, CO, USA

²Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

³Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA

⁴Science and Technology Corporation, Boulder, CO, USA

⁵Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

⁶Savannah River National Laboratory, Aiken, SC, USA

⁷Savannah River Nuclear Solutions, LLC, Aiken, SC, USA

⁸Center for Research on Sustainable Forestry, School of Forestry, University of Maine, Orono, ME, USA

⁹Institute for Applied Ecosystem Studies, US Forest Service Northern Research Station, Rhinelander, WI, USA

¹⁰Atmospheric & Oceanic Sciences Department, University of Wisconsin, Madison, WI, USA

¹¹Department of Chemical and Biochemical Engineering, University of Iowa, Iowa City, IA, USA

¹²Department of Environmental Sciences, University of Virginia, Charlottesville, VA, USA

¹³School of Engineering and Applied Science, Harvard University, Cambridge, MA, USA

Correspondence to: A. Andrews (arlyn.andrews@noaa.gov)

Received: 5 December 2012 – Published in Atmos. Meas. Tech. Discuss.: 8 February 2013

Revised: 27 November 2013 – Accepted: 27 November 2013 – Published: 25 February 2014

Abstract. A reliable and precise in situ CO₂ and CO analysis system has been developed and deployed at eight sites in the NOAA Earth System Research Laboratory's (ESRL) Global Greenhouse Gas Reference Network. The network uses very tall (> 300 m) television and radio transmitter towers that provide a convenient platform for mid-boundary-layer trace-gas sampling. Each analyzer has three sample inlets for profile sampling, and a complete vertical profile is obtained every 15 min. The instrument suite at one site has been augmented with a cavity ring-down spectrometer for measuring CO₂ and CH₄. The long-term stability of the systems in the field is typically better than 0.1 ppm for CO₂, 6 ppb for CO, and 0.5 ppb for CH₄, as determined from repeated standard

gas measurements. The instrumentation is fully automated and includes sensors for measuring a variety of status parameters, such as temperatures, pressures, and flow rates, that are inputs for automated alerts and quality control algorithms. Detailed and time-dependent uncertainty estimates have been constructed for all of the gases, and the uncertainty framework could be readily adapted to other species or analysis systems. The design emphasizes use of off-the-shelf parts and modularity to facilitate network operations and ease of maintenance. The systems report high-quality data with > 93 % uptime. Recurrent problems and limitations of the current system are discussed along with general recommendations for high-accuracy trace-gas monitoring. The

network is a key component of the North American Carbon Program and a useful model for future research-grade operational greenhouse gas monitoring efforts.

1 Introduction

Increased concern about rising greenhouse gas concentrations has already motivated many nations to begin regulating carbon emissions. Accurate measurements of atmospheric carbon dioxide and other species can provide an objective basis for evaluating reported emissions at regional to continental scales (10^4 – 10^6 km²) (Committee on Methods for Estimating Greenhouse Gas Emissions and National Research Council, 2010). A variety of modeling approaches with a wide range of complexity can be used to estimate fluxes from atmospheric data (e.g., Bakwin et al., 2004; Peters et al., 2007; Crevoisier et al., 2010; Gourdji et al., 2012). Accumulation of CO₂ in the atmosphere is the result of anthropogenic emissions, but only about half of the emitted CO₂ remains in the atmosphere. The remainder is absorbed by the oceans and the terrestrial biosphere in roughly equal amounts (e.g., Le Quéré et al., 2009). Net carbon uptake by ecosystems results from the small difference between large uptake fluxes driven by photosynthesis and large emission fluxes from heterotrophic and autotrophic respiration. Small biases can substantially impact annual net flux estimates at the continental scale, even if monthly fluxes are fairly well constrained. Data records with very high precision and long-term stability are therefore needed to resolve the net annual flux.

Here, we describe an automated, reliable, and high-precision analysis system for routine unattended monitoring of atmospheric CO₂, CO, and CH₄ from tall towers and a framework for estimating detailed time-dependent uncertainties for data from these systems. CO₂ is the principal anthropogenic greenhouse gas, and mixing ratio measurements of its abundance are sensitive to upwind fluxes, including fossil fuel emissions as well as uptake by and emissions from vegetation and soils. CO measurements contribute to the interpretation of CO₂ data by helping to identify and quantify pollution episodes and biomass burning. CH₄ is a potent greenhouse gas, with important anthropogenic sources from agriculture, fossil fuel exploitation, landfills, wastewater treatment, and natural biological sources from wetlands. Atmospheric data records of sufficient quality, density, and duration have the potential to greatly advance understanding of the processes and reservoirs that dominate the budgets of these and other greenhouse gases on timescales of decades to centuries.

Observations from tall towers are unique because measurements at several heights along the tower describe the vertical gradient, which reflects the relative influence of remote and local sources (Bakwin et al., 1998). Measurements obtained from sampling levels above ~100 m are minimally impacted by nearby vegetation and other local

emissions. Tall towers frequently penetrate the shallow nighttime boundary layer, in which case measurements from the highest levels are decoupled from the surface. Seasonal, day-to-day, and diurnal variability of CO₂ observed at a tall tower site can be very large. For example, Miles et al. (2012) analyzed data from a temporary installation of ~100 m towers in an agricultural region and showed that short-term variations of 10 ppm (parts per million dry air mole fraction) or more are common. Even though daily and seasonal variations may be large, high-precision stable measurements of CO₂ are needed to quantify year-to-year changes in carbon fluxes.

The North American Carbon Program (NACP) Plan (Wofsy and Harriss, 2002) calls for an observing network that would enable ongoing carbon flux estimates with coast-to-coast coverage at the regional scale. The proposed network would resolve spatial differences among regions roughly the size of New England, the Midwest corn belt, the mid-Atlantic, or the southeast US at temporal scales of months to seasons. The plan calls for 30 sites with surface monitoring from tall towers and biweekly aircraft sampling. A substantially larger network would be needed in order to monitor carbon emissions on a state-by-state or city-by-city basis.

NOAA's Earth System Research Laboratory (ESRL) has been working to build a network of tall tower CO₂ measurement sites since the early 1990s (Bakwin et al., 1998; Zhao et al., 1997). Under the NACP, a new in situ CO₂/CO analysis system was developed for tall tower sites in the NOAA Global Greenhouse Gas Reference Network and the network expanded from three sites to seven that are equipped with in situ analyzers. The towers are distributed across the United States and are typically television or FM radio transmitter towers that are > 300 m in height and enable trace-gas measurements that are representative of the planetary boundary layer (one site uses a 107 m cellular telephone tower). There is also one short-tower complex terrain site located on a mountain ridge in Shenandoah National Park that was established in collaboration with the University of Virginia (Lee et al., 2012). Complex terrain sites are needed to fill gaps in the monitoring network over mountainous regions, where tall broadcast towers are uncommon, but the representativeness of these sites can be difficult to determine due to complicated meteorological conditions (Brooks et al., 2012; Pillai et al., 2011; Lee et al., 2012). The instrument suite at one site has been augmented with a CO₂/CH₄ cavity ring-down spectrometer (CRDS).

In addition to NOAA's efforts, Environment Canada operates 12 greenhouse gas monitoring sites with towers that range in height from 20 to 105 m (Worthy et al., 2003), and an eight-site European tall tower network was recently constructed under the CHIOTTO project (e.g., Vermeulen et al., 2011; Popa et al., 2010; Thompson et al., 2009). These and other data provide the basis for prototype CO₂ data assimilation systems like NOAA's CarbonTracker (Peters et al., 2007; carbontracker.noaa.gov), which provides regularly updated

estimates of carbon fluxes for a variety of ecosystems and oceans.

CarbonTracker and other models are able to capture much of the synoptic-scale variability observed at continental sites (e.g., Law et al., 2008; Peters et al., 2010; Gourdjji et al., 2012; Schuh et al., 2010; Lauvaux et al., 2012; Meesters et al., 2012), but the spatial resolution for which carbon fluxes can be determined depends on the density of the measurement network. Many regions remain under-constrained, and the current North American network falls short of the NACP recommended sampling density. Further expansion of the North American and European greenhouse gas monitoring networks is needed and could be accomplished by a variety of government, university, and private sector institutions. However, care must be taken to ensure that data from various independently operated networks are compatible, and measurement protocols must be clearly defined.

The purpose of this paper is to describe the NOAA Tall Tower CO₂/CO/CH₄ analytical system with enough detail so that other researchers seeking to make high-precision measurements of CO₂ and related gases can replicate relevant components. The evaluation of the measurement system based on laboratory tests, field calibrations, and comparisons with independent measurements is also documented here. Although CO₂ analyzers have evolved over the past several years, the gas handling and temperature control techniques described here are broadly relevant, as is the novel methodology for estimating time-varying uncertainties. Typical uncertainty for CO₂ is < 0.1 ppm, CO < 6 ppb, and CH₄ < 0.5 ppb. Recurrent problems and limitations of the systems are discussed, along with potential improvements and recommendations for future greenhouse gas monitoring efforts.

Furthermore, this paper serves as a reference for the data collected from tall towers in the NOAA Global Greenhouse Gas Reference Network from 2006 to present, which are available at <ftp://ftp.cmdl.noaa.gov/ccg/towers/>. The CO₂ data have been used in several recent continental-scale and regional-scale studies of the North American carbon budget (Gourdjji et al., 2012; Schuh et al., 2010; Lauvaux et al., 2012; Miles et al., 2012; Lee et al., 2012). The WGC CH₄ data set was the primary record for two regional-scale analyses of CH₄ emissions in California (Zhao et al., 2009; Jeong et al., 2012), and the CO record has been used to evaluate new retrievals from the MOPITT satellite (Deeter et al., 2012).

2 Instrumentation

Starting in 2004, we developed and deployed an updated system for monitoring CO₂ and CO at NOAA tall tower sampling locations. The design is similar to the original CO₂ sampling equipment that was deployed at the NOAA Grifton, North Carolina (ITN, now discontinued), and Park Falls, Wisconsin (LEF), tall tower sites (Bakwin et al., 1998; Zhao et al., 1997), but with modifications to minimize sensitivity

to environmental conditions (such as room temperature) and to simplify maintenance of a larger network. The system is modular, so that a module with a component in need of repair can be quickly replaced with a spare, minimizing downtime and data gaps. Component-level repairs can be done in the laboratory, rather than on-site, which keeps costs down and facilitates quality control. Temperature stabilization enables high-precision measurements to be made with reduced use of expensive calibration gases. All of the major components are easily replaceable commercial off-the-shelf parts, and the modularity allows for new technology to be easily incorporated.

The CO₂/CO/CH₄ trace-gas analysis system was developed according to the following design objectives: (1) ability to deliver high-quality CO₂, CO, and CH₄ data; target long-term (year-to-year) and site-to-site comparability was 0.1 ppm for CO₂, 10 ppb for CO, and 1 ppb for CH₄. (2) Ease of maintenance. (3) Comprehensive monitoring of system parameters for quality control purposes. (4) Insensitivity to environment (e.g., room temperature, humidity, and atmospheric pressure). To date, we have deployed eight of these systems to field sites. The locations are listed in Table 1 along with installation dates. All of the systems have been in the field for ≥ 4 yr. The LEF, WKT, and AMT sites were originally equipped with older measurement systems based on the design described by Bakwin et al. (1998).

We designed the CO₂/CO analysis system during 2004–2005, and at that time nondispersive infrared (NDIR) absorption sensors (e.g., Li-cor Li-6200 series and Li-7000 analyzers) were the most commonly employed commercially available high-precision CO₂ sensors. NDIR CO₂ analyzers are low cost and have high sensitivity, but require sample drying and frequent calibration. Since then, new CO₂ and multispecies analyzers using cavity-enhanced absorption spectroscopy techniques such as cavity ring-down spectroscopy (CRDS) (Crosson, 2008) and off-axis integrated cavity output spectroscopy (O’Keefe et al., 1999) have become commercially available. These new analyzers have demonstrated improved off-the-shelf stability compared to the Li-7000 CO₂ analyzer, which is the core of the tall tower system described here (e.g., Winderlich et al., 2010; Richardson et al., 2012; Welp et al., 2013). In 2007, we integrated a Picarro G-1301 CRDS CO₂/CH₄/H₂O analyzer into the system for deployment at the WGC tall tower site (Zhao et al., 2009), as described in Sect. 2.8. The precision and accuracy of the Li-cor and Picarro CO₂ measurements at WGC are comparable. However, the Li-cor requires more frequent calibration than the Picarro analyzer, as discussed in Sect. 6.1.3.

Figure 1 is a schematic diagram of the analysis system, which occupies a standard instrument rack (48.3 cm \times 59.7 cm \times 198.1 cm), not including the Picarro analyzer. The CO₂ and CO analyzers and gas handling components are described below. Additional instrumentation details are provided in Appendix A. Many quality control parameters, such as pressures, flow rates, and temperatures, are

Table 1. Site information.

Site	Start Date	Location	Lat.	Long.	Surface Elev. (m a.s.l.)	Intake Heights (m a.g.l.)	Partners
LEF	Oct 1994 Upgrade May 2009	Park Falls, WI	45.9451	−90.2732	472	30, 122, 396 11 ^a , 76 ^a , 244 ^a	Penn State U of WI US Forest Service
WKT	Feb 2001 Upgrade May 2006	Moody, TX	31.3149	−97.3269	251	30, 122, 457 9 ^a , 61 ^a , 244 ^a	Blackland Research and Extension Center
BAO	May 2007	Erie, CO	40.0500	−105.0040	1584	22, 100, 300	
AMT	Sep 2003 Upgrade Feb 2009	Argyle, ME	45.0345	−68.6821	53	12, 30 ^b , 107	Harvard U of ME US Forest Service
WBI	Jul 2007	West Branch, IA	41.7248	−91.3529	241.7	31, 99, 379	U of IA
WGC	Sep 2007	Walnut Grove, CA	38.2650	−121.4911	0	30, 91, 483	Lawrence Berkeley National Laboratory
SCT	Aug 2008	Beech Island, SC	33.4057	−81.8334	115	30, 61, 305	Savannah River National Laboratory
SNP	Aug 2008	Shenandoah National Park, VA	38.6170	−78.3500	1008	5,10,17	U of VA

^a Sampling at these heights was discontinued at time of upgrade. ^b Additional sampling level added at time of upgrade.

recorded in addition to the CO₂ and CO data (Table A1). Photographs of the equipment and installations are provided in the supplementary material. Most of the towers are equipped with meteorological sensors, but discussion of the meteorological measurement system is beyond the scope of this paper. The hardware has generally been very reliable, and most sites have reported valid CO₂ and CO data for > 93 % of days since installation. However, certain recurrent problems have been encountered and are described in Appendix B.

2.1 Sample tubing

At each site, the CO₂/CO analyzer is housed at the base of the tower in a building or portable laboratory built in a trailer or modified sea container. Air is drawn down the tower through three sampling lines (1.27 cm/0.5 in. OD tubing, wall thickness = 1.57 mm, Synflex 1300; Eaton, USA). Three sample inlets are nominally positioned at 30, 100, and ≥ 300 m (as high as practical on a particular tower). Tubing is affixed to the tower using long UV-resistant plastic cable ties or stainless steel hose clamps at 1 m intervals. Tubes are run along tower legs and protected whenever possible to minimize wind-related vibration and stress as well as exposure to falling ice. Long horizontal runs and low points in the tubing are undesirable. However, at some sites these features cannot be entirely avoided. When possible, we install three tubes to each level so that a separate automated flask-sampling unit and the in situ system can be installed on separate lines with one spare line. Each line (including the spare) has a high-surface-area PTFE 0.2 μm filter capsule (6711-7502; Whatman, USA) on the inlet. Inlet filters occasionally become encased in ice or saturated with water during foggy

conditions or following heavy rain. Under these conditions, flow through the tubes is impeded or even entirely prevented. Flow gradually returns to previous levels, typically within a few days.

Each in situ sampling line has a dedicated pump and is continuously flushed at a typical flow rate of 5 to 9 standard liters per minute (slm; equivalent to the flow rate at $T = 0^\circ\text{C}$, $P = 1013\text{ hPa}$), which corresponds to a residence time of 4 to 7 min in a 500 m synflex tube. The pressure drop in a 500 m sample tube is estimated to be ~44 (65) hPa with a Reynolds number of 889 (1333) for a flow rate of 5 (9) slm and depends strongly on tubing diameter. The actual pressure drop is likely larger owing to other components such as the inlet filter and fittings along the line. The tubes are checked for leaks at the time of installation by capping the inlet and pulling a vacuum on the tube, and the test is repeated whenever the inlet filters are replaced, ideally once per year or when climbers are on the tower for another repair. The final pressure achieved during the pump-down is typically < 200 hPa. We use a shut-off valve to isolate the evacuated tube from the test pump and monitor the extent to which the capped line will hold the vacuum.

2.2 Pumps

Pumps are located upstream of the analyzers so that air is pushed rather than pulled through the analyzers. Some advantages of this design are that (1) the condenser works more effectively at higher pressure, (2) the ambient air is delivered to the analyzers at a pressure similar to the calibration gases, (3) the reduced likelihood that leaks will affect the measurements, and (4) the higher signal-to-noise ratio for the NDIR CO₂ analyzer and for the gas filter correlation CO analyzer

due to the presence of additional molecules in the light path. Obvious disadvantages are that water is more likely to condense in the sampling lines, and that the sample air is exposed to pumps and associated components, which are not included in the calibration path.

Air from the sampling lines enters the pump enclosure through a set of 7 μm filters (S-4F-7; Swagelok, USA). The filters are intended to protect the pumps and downstream components from particulates in the event that the sample tubing is breached. Each of the three sampling lines has a dedicated pump (MPU1763-N828-6.05 115 V/60 Hz; KNF Neuberger, Germany) that compresses the air. Pump outlet pressures are set to 69 kPa (10 psi) above ambient using a back-pressure regulator (GH30XTHMXXXB; ITT Conoflow, USA) and monitored using inexpensive electronic pressure transducers (68075-44, 0–25 psig; Cole Parmer, USA). Excess flow is vented through the back-pressure regulator and measured with an electronic mass airflow sensor (AWM5102VN; Honeywell, USA). A fourth (exhaust) pump pulls a vacuum (~ 250 hPa) on the combined output from the CO₂ and CO analyzers. The exhaust pump enlarges the pressure gradients across the CO and CO₂ analyzers to provide improved pressure and flow control and also improves performance of the Nafion[®] (registered trade name of E.I. DuPont de Nemours) dryers as described in Sect. 2.3. The pressure upstream of the exhaust pump inlet is measured with a ± 103.4 kPa (15 psi) transducer (68075-32; Cole Parmer, USA), which is mounted in the Nafion dryer assembly for convenience. The inlet pumps are equipped with Viton[®] (also known as FPM) diaphragms, and more durable EPDM diaphragms are used in the exhaust pumps. The pump enclosure is cooled by a fan, and pumps are factory-equipped with automatic shut-off to prevent overheating (maximum recommended ambient temperature is 40 °C). Air temperature monitored in the interior of the pump enclosure does not typically exceed 35 °C.

We aim to refurbish each pump assembly approximately once per year. Pump diaphragms are replaced, and pumps are tested for compression and vacuum. The “bypass” flow is the portion of the flow that is vented through the back-pressure regulator (i.e., equivalent to the total flow minus that portion which is periodically delivered to the analyzers), and provides an indicator of pump performance. In addition to scheduled maintenance, pump units are recalled anytime the flow delivered by the sample pumps drops suddenly, if the total flow (sample plus bypass) drops below 4 slm, or if the exhaust line pressure rises unexpectedly or rises above 400 hPa. We have found that pumps with torn diaphragms often deliver adequate flow rates and back pressure, but will not generate a vacuum. Any leakage of air across a torn sample pump diaphragm will contaminate the sample airstream, and must be avoided. Pumps in the field can be tested for leaks by simply capping the inlet and checking whether the flow drops to zero. Future modifications will include electronic shut-off

valves upstream of the pumps so that pump leak checks can be automated. All connections to the pump box are made with Quick-Connect fittings (SS-QC4-B1-400 and SS-QC4-D-400; Swagelok, USA) that do not require wrenches or other tools so that the entire unit can be easily replaced on-site by a minimally trained technician.

2.3 Dryers

Liquid water can damage system components, and even low levels of water vapor can interfere with measurements. A humidity difference of 100 ppm of water corresponds to a so-called “dilution offset” of 0.04 ppm CO₂ if not corrected. (The dilution offset is the difference in mole fraction when computed relative to dry versus wet air; see Sect. 4.2.2 for more details.) Water vapor differences among samples and standards can also cause spectral artifacts related to line interference or pressure broadening. Water vapor artifacts can be reduced or eliminated by drying the sample airstream to a dew point ≤ -50 °C or by drying the sample and humidifying the calibration gases to minimize differences in water content (WMO, 2012, Sect. 12.1). We adopt the latter strategy. All of the CO₂, CO, and CH₄ measurements are reported as dry air mole fractions (e.g., χ_{CO_2}).

Air exiting each sample line pump is passed through one channel of a four-channel compressor chiller (02K1044A EC-4-G; M&C Products, Germany) to remove the bulk of the water vapor. The chiller is configured with four separate glass traps (one for each sample intake line plus one spare). Each channel has a dedicated peristaltic pump to remove liquid effluent from the trap. The peristaltic pumps require routine maintenance, so we reconfigure the set of four pumps as a single modular unit that can easily be removed by a nonskilled technician and returned to our laboratory for service. The temperature of the cooling element is maintained at a setpoint of 1.6 °C. The sample air pressure in the condensers is ~ 689 hPa (10 psi) above ambient, which enables drying to a 1013 hPa dew point that is lower than the cooling element temperature. The temperature of the airstream exiting the chillers is a function of the flow rate and therefore varies with pump performance. We tested the chiller performance with a flow of 6 slm using a mixture of dry and saturated water to vary the input water content from approximately 0.8 to 2.8 % (mole fraction), corresponding to a dew point range of 3.9 to 23 °C at 1013 hPa. The moisture content of the output airstream over this range was nearly invariant at 4400 ± 180 ppm, equivalent to a dew point of -3.8 °C at 1013 hPa and ~ 1.4 °C at 1530 hPa (the approximate chiller pressure in the Boulder, CO, laboratory, ~ 1700 m elevation). Liquid alarm sensors (03E4100 KS2; M&C Products, Germany) on each intake line close relays (FA1.4; M&C Products, Germany) to disable the pumps if liquid water breaks through the chiller.

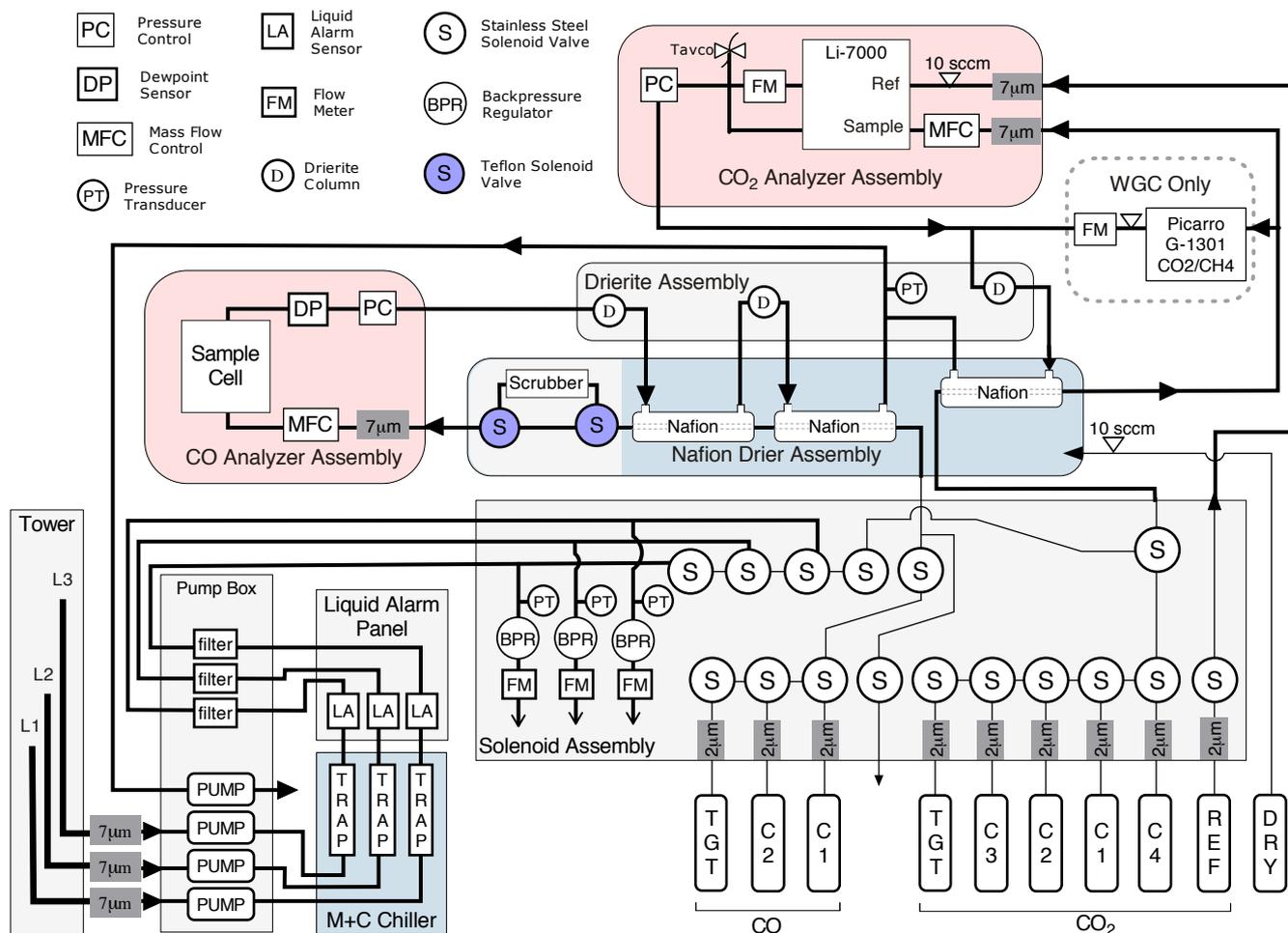


Fig. 1. Schematic diagram of the NOAA ESRL Tall Tower CO₂/CO analysis system. The Picarro G-1301 analyzer is only included at the WGC tall tower site (Walnut Grove, CA). Line thickness indicates tubing diameters of 0.125, 0.25, and 0.375 in. (0.3175, 0.635, 0.9525 cm). Pink and blue shading indicate heated and cooled enclosures. Photos are provided in the supplementary material.

A further level of protection against liquid water infiltration is provided by PTFE filter membranes that are relatively impermeable to water (TF-200 PTFE 0.2 micron filters and model 1235 47 mm filter holders; Pall Life Sciences, USA). Laboratory tests determined that an upstream pressure of 69 kPa (10 psi) is required to push liquid water through the filters. Saturated PTFE filters can block all airflow, and as a result they do not dry. Thus, we expect that the filter membranes would need to be replaced after coming into contact with liquid water, although this has not happened. Early units used polycarbonate filter holders, which are substantially less expensive, but we switched to aluminum filter holders after several of the polycarbonate units cracked during shipping. The PTFE filters are installed downstream of the liquid alarm sensors, but they are housed in the pump box for easy access and so that they can be routinely replaced when pump maintenance is performed. We originally used 2 μ m filters but have since found that fine black particles were present on the downstream face of the filters and in many downstream

components, including in the sample manifold and the bypass flow meters. We suspect that the pump diaphragms are shedding fine particles that are smaller than 2 μ m, and we are now evaluating whether the same large-capacity 0.2 μ m PTFE filter capsules (6711-7502; Whatman, USA) that are used on the tower inlets can be used in the pump box to capture these particles.

Single-strand Nafion membrane dryers (MD-110-144P-4; Perma Pure, Halma, UK) are used in self-purge configuration to further reduce the sample dew point. One 3.66 m dryer is used on the CO₂ channel, which has a flow rate of 250 standard cubic centimeters per minute (sccm). Two 3.66 m Nafion dryers are used in series for the CO channel, which has a higher flow rate of 600 sccm. The effectiveness of Nafion membrane dryers depends on the relative flow rates and partial pressures through the sample and purge tubing. The exhaust pump reduces the pressure on the purge side of the Nafion dryers, resulting in a faster volume flow rate and improved drying. A nonhazardous desiccant (Drierite;

WA Hammond, USA – part number 27070 includes a canister with Swagelok fittings) is used to remove residual water from the analyzer exhaust before it enters the purge housing. The lifetime of the desiccant is several years given the extremely low water content of the analyzer exhaust. Nafion is more effective at cooler temperatures, and rapid temperature changes can produce large changes in the water content of the sample airstream. We therefore house the Nafion dryers in an insulated enclosure equipped with a thermoelectric cooler (SD6C-HCAF-AARG; Watlow, USA). The box temperature is maintained at $\sim 20^\circ\text{C}$, and the enclosure is continuously flushed with 5–10 sccm of dry air from a cylinder to prevent condensation.

The sample line pressure in the Nafion dryers is not actively controlled. Instead, all sample pumps and calibration gas regulators are manually adjusted to deliver approximately the same pressure. At some sites, we have begun monitoring the line pressure at the exit of the Nafion dryer on the CO₂ channel and have noted that calibration curve residuals are smaller when the pressures are carefully adjusted. Future versions of the system may therefore include active pressure regulation upstream of the Nafion membrane dryers (Welp et al., 2013).

We achieve a sample dew point of approximately -36°C (at 1013 hPa) for the CO₂ channel as indicated by laboratory tests using an accurate dew point sensor (DMT 142; Vaisala, Finland) and by the WGC Picarro analyzer. Note that the Li-7000 analyzers do not provide a reliable measure of absolute humidity without routine user calibration, which we have not implemented. The sample dew point for the CO channel is approximately -34°C (at 1013 hPa), as indicated by a separate dew point sensor (DMT 142; Vaisala, Finland) immediately downstream of the sample cell. Calibration gases are introduced upstream of the Nafion membrane dryers. The Nafion membrane acts as a reservoir for water and is normally equilibrated with the chilled sample air. The dry calibration gases are humidified as they pass through the Nafion dryers and emerge with a dew point that is indistinguishable from that for dried atmospheric sample air. Differences between atmospheric samples and calibration standards are < 10 ppm H₂O.

Nafion is slightly permeable to CO₂, and thus CO₂ can be lost from the sample airstream when there is a large partial pressure gradient across the membrane (Ma and Skou, 2007). In our setup, the pressure inside the membrane is 6–8 times higher than on the purge side. Loss across the membrane is problematic only if different between samples and calibration standards, which might occur if CO₂ permeability is strongly dependent on the moisture content of air entering the dryer. We measured CO₂ loss across the Nafion in the laboratory using a Picarro CRDS (model 2401-m) with a recently calibrated water correction (after Chen et al., 2010). We found that CO₂ loss across the Nafion membrane is nearly identical for calibration gases and sample gas in our system. Calibration gases were either sent

through the Nafion dryer or routed directly into the Picarro. Field conditions were closely matched: H₂O exiting the Nafion dryer was ~ 180 ppm, flow ~ 250 sccm, internal Nafion pressure ~ 1700 hPa, external pressure ~ 265 hPa. We measured CO₂ loss of 0.125 ± 0.05 ppm. We also simulated atmospheric sampling by routing gas from a cylinder through a bubbler and into the chillers (bypassing the pumps). H₂O exiting the chiller was $\sim 0.57\%$ measured by the Picarro (chiller temperature $\sim 3.5^\circ\text{C}$, chiller pressure ~ 1700 hPa), and CO₂ loss across the Nafion membrane was 0.10 ± 0.03 ppm. Chiller temperatures are normally set at $\sim 1.6^\circ\text{C}$, so the test corresponds to a worst-case scenario. CO₂ loss from the sampled ambient air is nearly identical to the loss from the calibration standards with no detectable dependence on initial humidity. Thus, there is no bias resulting from Nafion membrane.

2.4 Sample/calibration selection manifolds

Atmospheric samples from the three inlet lines are selected through a solenoid valve manifold. Two three-way valves are plumbed in series to minimize dead volumes, and a two-way valve at the end of the chain is used as a shut-off valve for the third inlet channel (009-0933-900 (three-way) and 009-0631-900 (two-way); Parker Hannifin Pneutronics Division, USA). The solenoid valves are stainless steel and are rated to 689.5 kPa (100 psi) inlet pressure. Calibration gases are selected using a similar manifold. Solenoid valves were chosen instead of a multiposition (stream-selection) valve to increase reliability. We tested one system with a multiposition valve (10-position ECMT; VICI Valco, USA) with the expectation that the multiposition valve would have less dead volume than the solenoid valve manifolds, but the response time after transitions between calibration gases was not improved, suggesting that other components dominate flushing and equilibration in this system.

2.5 CO₂ analyzer

CO₂ is measured using a nondispersive infrared gas analyzer (Li-7000 CO₂/H₂O; Li-cor, USA). The analyzer is housed in a temperature-controlled enclosure since the analyzer baseline is sensitive to temperature variations. CO₂ mole fractions reported by the Li-cor analyzer are temperature-compensated, but some sensitivity remains that can cause errors as large as several tenths of a ppm of CO₂ if temperatures are not strictly controlled. The setpoint for the temperature controller is chosen to be 10–15 °C above the site-dependent typical maximum room temperature, and Li-7000 internal temperatures typically fall in the range from 37 to 40 °C. The cell temperature is normally controlled to within 0.1 °C (see Appendix A1 for more details about temperature control).

Flow through the sample cell of the Li-7000 is actively regulated upstream of the cell (1179A52CS1BV; MKS Instruments, USA). The reference flow is regulated by a needle valve (4171-1505; Matheson, USA) upstream of the analyzer, and sample and reference flows are joined downstream of the analyzer to ensure that the sample and reference cell pressures are nearly equal. A pressure controller (640A13TS1V22V; MKS Instruments, USA) downstream of the junction actively regulates the pressure to 1066 hPa (800 Torr). Typical sample and reference flow settings are 250 and 10 sccm, respectively. The reference flow is measured downstream of the analyzer (AWM3150V; Honeywell, USA). The difference in sample and reference flows does result in a small but invariant difference in pressure across the cells. The flow and pressure controllers are sensitive to ambient temperature and are therefore housed inside the temperature-controlled area.

The H₂O channel of the Li-7000 analyzer is used to continuously monitor the performance of the drying system. The absolute H₂O measurement from the Li-7000 is not accurate at the very low humidity levels achieved by the drying system (e.g., the analyzers may be offset by 500 ppm or more and frequently report negative H₂O mixing ratios for our dried sample airstream; see Sect. 4.2.2 for more details). However, the gain of the H₂O channel is sufficiently reliable to indicate differences in water content among calibration and sample gases. In-line filters (SS-4F-7; Swagelok, USA) are mounted on the sample and reference inlets of the Li-7000. A pressure relief valve (2391243-26-9; Tavco, USA) between the sample outlet of the Li7000 and the pressure controller protects the Li-cor analyzer from accidental over-pressurization. We occasionally have problems with the Tavco valve releasing unintentionally. It can be remotely reseated by sending a command to open the downstream pressure control valve so that the exhaust pump pulls a vacuum on the Tavco valve.

2.6 CO analyzer

CO is measured using gas filter correlation (48C Trace Level; Thermo Electron Corporation, USA). Cell pressure is maintained at 1066 hPa (800 Torr). Sample flow is controlled at 600 sccm. A dew point sensor (DMT 142; Vaisala, Finland) is located downstream of the sample cell and mounted inside the analyzer. The factory-installed internal pump is removed from the analyzer, and flow (1179A23CS1BV; MKS Instruments, USA) and pressure (640A13TS1V22V; MKS Instruments, USA) controllers are installed in that space, upstream and downstream of the sample cell, respectively. The factory-installed internal pressure and flow sensors and the heaters on the sample cell are disconnected. We do not use the optional zero and span solenoids available from the manufacturer. Instead, calibration and sample gases are introduced using an external gas selection manifold as described above.

Frequent checks of the baseline drift are needed to achieve high precision (~ 3 ppb for a 2 min average) with this analyzer. Scrubbed ambient air is measured at least twice per hour to track the analyzer baseline. The scrubber is a stainless steel tube (0.5 in. OD \times 12 in. long) filled with a catalytic reagent (Sofnocat 423, O. C. Lugo) and with a glass wool plug and stainless steel mesh at each end. The tube is mounted at a slight angle from horizontal to prevent unfilled spaces that might develop as a result of gravitational settling of the catalyst. The sample flow is periodically diverted through the scrubber by simultaneously switching two three-way solenoid valves on either end of the scrubber (203-3414-215 (three-way) and 203-1414-215 (two-way) Galtek; Entegris, USA). For convenience, the scrubber and solenoid valves are mounted in the enclosure with the Nafion dryers, outside of the temperature-controlled region. We do not routinely monitor the scrubber performance, but we can evaluate whether the measured baseline is consistent with CO = 0 ppb by evaluating the calibration residuals or by calculating the intercept of a linear fit that includes COC1, COC2, and COTGT. Lab tests indicate that a much smaller CO scrubber volume would perform equally well. However, there is no penalty for conservatively sizing the scrubber (other than an incremental cost difference). Response time, for example, is not affected, since all air exiting the scrubber is free from CO.

2.7 Standard gases and related components

A total of nine calibration gases are used for the CO₂/CO analysis system, as shown in Fig. 1 and described below in Sect. 3.2. Regulators are high-purity, two-stage nickel-plated brass with low internal volume (51-14C-CGA-590; Scott Specialty Gases, USA). The tank pressure gauge on each of these regulators is replaced with an electronic pressure transducer (68075-56; Cole-Parmer, USA). The transducer is protected from rapid pressure changes by a flow restrictor (SS-4-SRA-2-EG; Swagelok, USA). The tank pressure signals provide a measure of gas use that is used to estimate optimal replacement dates and identify tanks with slow leaks, which most often occur at the CGA connection between the tank and the regulator.

Quick-Connect fittings with automatic shut-off (SS-QC4-D-200 and SSQC4-B2PM; Swagelok, USA) are installed on the outlet of each regulator so that on-site technicians can easily purge the regulator when a new cylinder is attached. Purging the regulator consists of opening the tank valve, quickly shutting it, allowing the fresh gas to sit in the regulator for a few minutes, and then draining the gas through the regulator outlet. This process is repeated three times whenever a new cylinder is attached. Purging the regulator minimizes the introduction of room air into the calibration lines and protects the gas in the new cylinder from backward diffusion of room air or residual air from the previous cylinder.

Table 2. System modes.

	SYSMODE	Description	Approximate Concentration	Interval (Hours)
CO ₂ Calibration	C1	Standard Gas	350 ppm	12
	C2	Standard Gas	380 ppm	1–2
	C3	Standard Gas	410 ppm	12
	C4	Standard Gas	460 ppm	12
	TGT	Standard Gas	400 ppm	6
CO Calibration	C1	Standard Gas	100 ppb	12
	C2	Standard Gas	350 ppb	12
	TGT	Standard Gas	220 ppb	6
	ZER	Scrubbed Ambient Air	0 ppb	0.5–0.75
CH ₄ Calibration	C1	Standard Gas	1680 ppb	12
	C2	Standard Gas	1840 ppb	1–2
	C3	Standard Gas	1960 ppb	12
	C4	Standard Gas	2160 ppb	12
	TGT	Standard Gas	1900 ppb	6
Nominal Height (m a.g.l.)				
Sampling	L1	Lowest Inlet	30	0.25
	L2	Middle Inlet	100	0.25
	L3	Highest Inlet	> 300	0.25

Clean stainless steel tubing is used for the calibration lines (3.18 mm/0.125 in. OD, wall thickness 0.07 mm; SS-T2-S-028-20; Swagelok, USA), which are typically a few meters long. The tube specifications reflect tradeoffs between minimizing volume and providing dependable connections (i.e., connections to 3.18 mm/0.125 in. tubing are generally more durable than connections to 1.59 mm/0.0625 in. tubing). An in-line 2 µm filter (SS-2F-2; Swagelok, USA) is installed at the point where the calibration line enters the manifold to protect against introduction of particulate matter such as metal fragments from the plumbing connections or dust from the room.

2.8 Integration of CRDS CO₂/CH₄ analyzer

The installation at the WGC site includes a Picarro G-1301 CRDS for measuring CO₂, CH₄ and water vapor. The Picarro analyzer is plumbed in parallel with the Li-cor CO₂ analyzer. A 2 µm stainless steel filter (SS-2F-2; Swagelok, USA) is installed on the inlet and a needle valve is used to restrict flow through the analyzer to approximately 80 sccm, which is adequate for flushing the sample cell during the 5 min sampling interval while minimizing calibration gas consumption. Note that a higher flow rate would be desirable for a stand-alone installation in order to flush upstream tubing and regulators in a reasonable time when switching among calibration and sample modes.

The pressure of the sample airstream exiting the Nafion dryer assembly is ~68.9 kPa (10 psi) above ambient. The Picarro cell pressure is controlled at 186 hPa and the cell

temperature is maintained at 45 °C. The Picarro H₂O channel reports the humidity of the dried sample airstream. The typical value is ~0.013 % (mole fraction), corresponding to a dew point of –39.8 °C at 1013 hPa. There is no discernable difference in water content between dried ambient air and the humidified standards. The Picarro analyzer does not have a built-in a flow measurement, so an external sensor (AWM 3100V; Honeywell, USA) is installed on the outlet. Exhaust from the Picarro is captured and combined with the exhaust from the Li-7000 and used to purge the Nafion dryer. The Picarro analyzer has a dedicated computer for data acquisition and control. However, to simplify post-processing, we use the Campbell Scientific serial communications data acquisition system to integrate key Picarro output fields into our primary data stream.

The Picarro analyzer was deployed in fall 2007 and was among the first commercially available CO₂/CH₄ Picarro units to be installed at a field site. The stability of the analyzer and the reliability of the H₂O corrections to CO₂ and CH₄ were initially unknown. Our configuration was largely driven by convenience, so that standards and gas-handling could be shared between the Picarro and the Li-7000. The analyzer has demonstrated remarkable stability over nearly five years of operation, as will be described in more detail in Sect. 6.1.3.

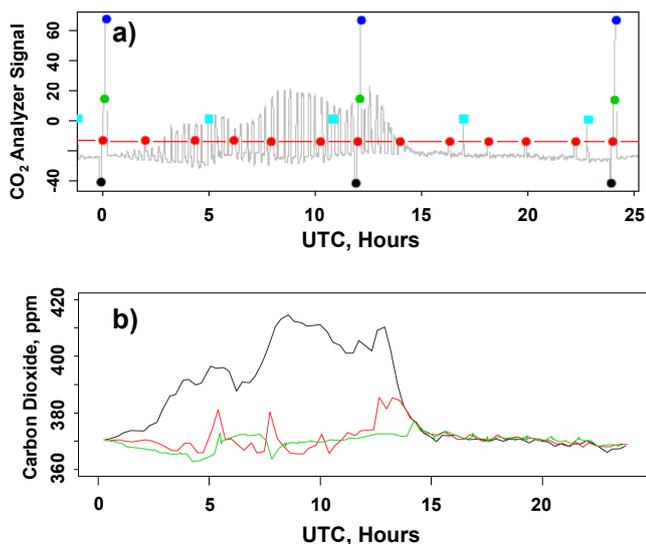


Fig. 2. (a) CO₂ analyzer signal for a typical summertime daily measurement cycle at the LEF tall tower site (Park Falls, WI; 6 August 2009). The analyzer signal is the estimated mole fraction difference from the reference gas based on the factory calibration. Filled circles correspond to calibration standards (black = C1, red = C2, green = C3, dark blue = C4). Red connecting lines show the continuous estimate of the analyzer baseline. Target (TGT) measurements (cyan squares) are treated as unknowns and used to monitor system performance. (b) Corresponding χ_{CO_2} time series for the 30 m (black), 122 m (red), and 396 m (green) sampling heights showing well-mixed conditions during afternoon and stratified conditions at night. Local standard time for the LEF site is 6 h behind GMT.

3 Sampling and calibration

3.1 Sampling sequence

Figures 2 and 3 show CO₂ and CO data from a typical daily measurement cycle under summertime conditions and illustrate the typical variability and range of values for which this system was optimized. Vertical gradients are small during the daytime, when heating of the surface causes vigorous turbulent mixing within the planetary boundary layer. During the night, any CO₂ and CO emitted at the surface will accumulate in the shallow stable layer. At the WBI site in Iowa, where vegetation and soils have high nighttime respiration rates, we have observed nighttime differences of > 60 ppm CO₂ between the 31 and 379 m inlets. We also observe strong vertical gradients at some sites during winter, when temperature inversions are present and pollutants accumulate near the surface. Calibration and atmospheric sampling modes are described in Table 2, and each mode is typically run for a 5 min interval. The CO and CO₂ analyzers can be calibrated independently, but during atmospheric sampling, they draw air from the same sampling inlet. The following discussion is centered on CO₂, which is the primary species of interest, but generally applies to CO and CH₄, except as noted.

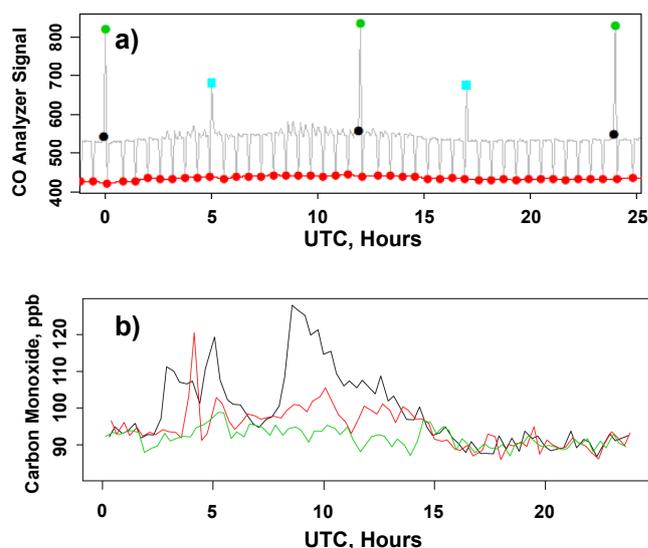


Fig. 3. Same as Fig. 2 but for CO. The baseline measurements (red = COZER) correspond to scrubbed ambient air. The analyzer signal is the estimated mole fraction in ppb based on the factory calibration. Note that it is normal for the signal from this type of analyzer to drift upward, which is the reason for the high values in (a). Since we do not rely on the factory calibration, we do not need to adjust the analyzer zero value.

The 5 min sampling interval allows for nearly complete equilibration even for system mode transitions corresponding to large χ_{CO_2} differences ($\Delta\chi_{\text{CO}_2}$) that can occur during calibration or during sampling when there is a large vertical gradient across the measurement heights. When switching between system modes, we allow three minutes for the system to flush and report data corresponding to the final 2 min of each 5 min sampling interval. Several times per day, the CO₂ analyzer dwells on the highest sampling height while the CO analyzer baseline is measured using scrubbed ambient air from that inlet. In such cases, where the CO₂ mode does not change between successive sampling intervals, there is no need to discard the first 3 min of the second 5 min interval. Field calibrations and laboratory tests show that errors associated with incomplete equilibration are < 0.05 ppm for a 3 min flushing interval for $\Delta\chi_{\text{CO}_2} < 60$ ppm. Although these errors are small, extra scrutiny is warranted because they are systematic. The Li-cor and Picarro analyzers are sufficiently precise such that field calibration data can be used to estimate the equilibrium value and adjust the analyzer signal (see Sect. 4.1.1 and Appendix C1). However, data from the CO analyzer are too noisy to reliably derive an equilibrium correction within our standard 3-day data processing window.

The 5 min sampling interval limits the temporal resolution of our analyzer to no more than four three-height profiles per hour. With a 3 min flushing time, we therefore report data corresponding to eight minutes out of every hour for each

intake height, which limits our ability to confidently compute, for example, hourly or afternoon averages. We actually get slightly fewer than four profiles per hour, since the sampling cycle is periodically interrupted for calibrations. At certain times of day, the temporal variability of CO₂ at a single height may be considerably larger than the uncertainty resulting from incomplete equilibration within the 5 min sampling interval. Clearly, it would be advantageous to reduce the flushing time in order to increase the temporal resolution of the measurements, but doing so would adversely impact the CO measurements, which are noisy and benefit from 2 min averaging. Winderlich et al. (2010) use integrating volumes on the sampling lines to achieve temporally smoothed sampling, and we may eventually incorporate something similar. Another option is to modify the sampling sequence to spend most of the time dwelling on the highest intake, since we typically use the vertical gradient information primarily to identify periods with vigorous vertical mixing.

3.2 Calibration

3.2.1 CO₂

Four standards (CO2C1, CO2C2, CO2C3, and CO2C4) are used to calibrate the response curve for the Li-7000. The approximate CO₂ values for the standards are given in Table 2. A fifth cylinder (CO2REF) supplies gas to the Li-7000 reference cell, and the concentration is chosen to approximately match the CO2C2 standard, which is used to monitor the Li-7000 baseline. A sixth calibration standard, the “target” (CO2TGT), is measured independently to monitor the stability of the instrument.

We initially used a 24 h CO₂ calibration sequence consisting of a full response curve calibration four times per day and a baseline check approximately once per hour. The CO2TGT tank was measured four times per day with two of the measurements adjacent to full calibrations, and the other two measurements temporally distant from both full calibrations and baseline checks. After ~2 yr of operation, we reduced the frequency of full calibrations to twice per day with baseline checks every two hours. The CO2TGT tank is still measured four times per day, but now all target measurements are temporally distant from calibrations and baseline checks. The sequence now runs on a 23 h interval so that the timing of the calibrations and target measurements drifts throughout the day and covers a full diurnal cycle over the course of approximately 10 days.

During experiments with a prototype system, we ran true differential zero measurements to monitor the Li-7000 baseline, where the CO2REF gas was routed through a “T” fitting so that we could send the gas simultaneously through the Li-7000’s sample and reference cells. We found that this setup, which has been widely used (e.g., Daube et al., 2002; Bakwin et al., 1998), apparently disrupts internal regulator and/or calibration line pressures and disturbs the measured

CO₂ value. Full recovery from this perturbation exceeded 10 minutes. We now use a “pseudo-differential zero” measurement technique using the CO2C2 standard to monitor the Li-7000 baseline. The concentration of the CO2REF standard should be within a few ppm of the CO2C2 standard so that baseline drift can be reliably distinguished from gain changes.

NDIR analyzer signals are not inherently linear, but the Li-7000 provides estimated CO₂ output that has been linearized according to a fifth-order polynomial with unit-specific calibration coefficients determined by the manufacturer. The linearization algorithm relies on a user-specified reference concentration, and this value must be accurate to within a few ppm to avoid significant deviations from linearity. CO2REF cylinders must be replaced several times per year, and they are not generally calibrated but are targeted during filling to within a tolerance of a few ppm. We specify a reference value of 380 ppm for the Li-7000 software, while actual values may range from 377 to 383 ppm. Raw detector signals from the Li-7000 sample and reference cells (i.e., Li-7000 “CO2A W” and “CO2B W” signals and corresponding values for H₂O) are archived so that in case of a problem, we can recover a signal that is comparable to the analog output from earlier Li-cor CO₂ analyzer models (such as the Li-6251). The linearized output is subject to additional internal signal averaging and consequently demonstrates improved precision compared to the raw signals.

3.2.2 CO

For the CO calibration, we use two standards (COC1 and COC2) and a target standard (COTGT). The baseline is tracked by measuring ambient air from which CO has been scrubbed (COZER). The baseline is measured every 30–40 min, and the other calibration standards are measured approximately every 23 h. The COTGT is measured four times per day and COTGT measurements are not typically adjacent to baseline checks or calibrations.

Note that for the Thermo Electron 48C TL CO analyzer, it is important to use CO calibration standards that are made with a balance of whole air. In particular, the absence of CO₂ in the standard gases will artificially raise the baseline of the analyzer due to spectral interference. (In gas filter correlation, spectral interference can produce either a positive or negative response.) We determined empirically that a change in the CO₂ content from 380 to 0 ppm corresponds to a change in the CO analyzer baseline that is equivalent to +20 ppb CO. The artifact was present and had consistent magnitude in all units tested (> 5 units tested to date) and is independent of the CO concentration. Variations in ambient CO₂ can cause CO measurement errors as large as 5 ppb (for a 100 ppm CO₂ variation) when scrubbed ambient air is used to track the analyzer baseline. It would be possible to use the coincident ambient CO₂ measurements to correct the CO

analyzer baseline for CO₂ interference, although we have not yet implemented this correction.

4 Post-processing

4.1 Algorithms for calculating χ_{CO_2} , χ_{CO} , and χ_{CH_4}

Data are stored with 30 s temporal resolution, and the timestamp corresponds to the end of the 30 s interval. Average values and standard deviations for each 30 s interval are recorded for the CO₂ and CO analyzer signals. The data array contains a system mode indicator (SYSMODE) for each gas that is used within the datalogger program to set the position of valves in the calibration and sample manifolds. The data array also contains a counter (INTERVAL) that is used to track how many 30 s intervals have elapsed since the SYSMODE was last switched. Thus, during a typical 5 min sampling period, INTERVAL values range from 1 to 10. Higher counter values occur when the sampling sequence contains back-to-back occurrences of the same mode. This happens when either the CO or CO₂ analyzer enters a calibration mode, while the other analyzer continues to sample ambient air with no inlet height change and for variants of the sampling program that are designed to dwell on a particular intake height for longer than 5 min. SYSMODE and INTERVAL are used in post-processing to separate data from different calibration and sample modes and to filter data immediately following a sampling mode transition.

The post-processing software is written in IDL and operates on three days of data because calibration data from the previous and subsequent days are needed to compute the most accurate values and uncertainties for the central day.

4.1.1 CO₂

We use routine calibration data to adjust the linearized, pressure-, H₂O-, and temperature-corrected differential CO₂ dry air mole fraction (χ_{CO_2}) signal reported by the Li-cor. We define s to be the vector of individual 30 s average analyzer signals s_i for all times t_i (gray curve in Fig. 2a).

The Li-cor baseline drift is tracked using repeated measurements of the CO₂C2 standard, which is measured every 1–2 h. In post-processing, we extract a vector of analyzer baseline measurements, s_b at times t_b (times when SYSMODE = “CO₂C2” and INTERVAL = 10) and linearly interpolate over time to create a continuous baseline time series b (black lines in Fig. 4a). The baseline is subtracted from the raw data time series s to obtain the drift-corrected signal, s' . For cases where a significant correlation exists between analyzer temperature and the baseline signal s_b , we have the option to enable a temperature-dependent baseline algorithm as described in Appendix B2.

The Li-7000 has very low noise (< 0.01 ppm) over timescales of several minutes, and careful examination of the calibration data reveals that the signal does not reach

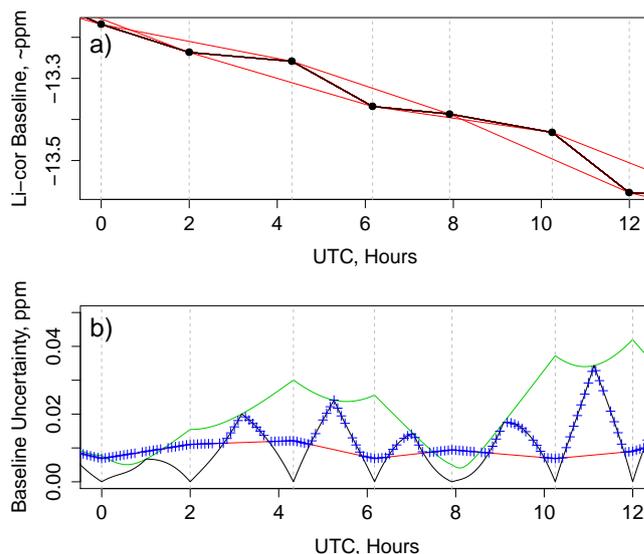


Fig. 4. (a) Measured Li-7000 baseline signal s_b (black filled circles), linearly interpolated in time (black lines), and alternate realizations of the baseline obtained by leaving out individual baseline measurements (red lines). The analyzer signal is the estimated mole fraction difference from the reference gas based on the factory calibration. (b) The analyzer short-term precision u_p (red), defined as the time-interpolated 30 s standard deviation of the individual baseline measurements; the standard deviation computed across all realizations of the analyzer baseline σ_b (green); and the baseline-drift uncertainty (black), which is the green curve weighted by a function that varies linearly from 0 at times t_b to 1 at times halfway between sequential baseline measurements. The total analyzer baseline uncertainty u_b at any time is indicated by the blue crosses and is the larger of u_p or the time-weighted drift uncertainty.

equilibrium by the end of the 5 min interval. We have applied a disequilibrium correction to s' that assumes an exponential approach to steady state where the time-constant and equilibrium value are derived from the calibration data (Appendix C1). A single set of fit coefficients is computed for the 3-day processing window. The magnitude of the correction depends on the $\Delta\chi_{\text{CO}_2}$ between successive sampling modes. For the case described in the Appendix, the largest impact corresponds to the transition between CO₂CAL3 and CO₂CAL4 measurements ($\Delta\chi_{\text{CO}_2} \cong 50$ ppm), and the correction is only 0.015 ppm. Although the correction is negligible in this case, tracking the disequilibrium is helpful when designing calibration sequences and evaluating errors in anomalous data like concentrated pollution plumes or fires. Typical equilibrium time constants (~ 70 s) are much longer than would be expected based on the volume of the Li-7000 sample cell (10.86 cm³, corresponding to a theoretical flushing time constant of 2.2 s for a flow rate of 250 sccm). The reason for the observed long flushing time is unknown, but could be caused by an inefficiently flushed volume somewhere in the sample path or perhaps by CO₂ adsorption or absorption onto surfaces such as the Bev-A-Line tubing

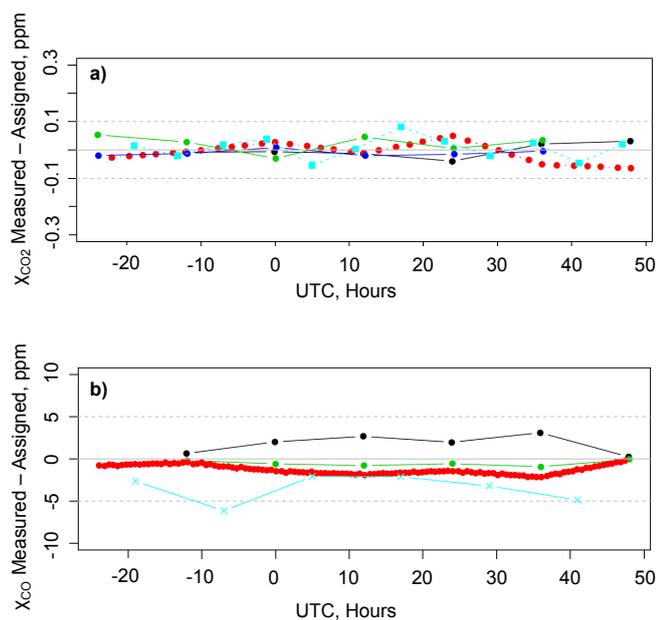


Fig. 5. Measured minus assigned values for (a) CO₂ standards (black = CO2C1, red = CO2C2, green = CO2C3, dark blue = CO2C4, and cyan = CO2TGT) and (b) CO standards (red = COZER, black = COC1, green = COC2, cyan = COGTGT) for a 3-day period from the LEF tall tower, where the central day is 6 August 2009.

inside the Li-cor. In practice, the disequilibrium correction frequently fails to provide realistic fit coefficients and sometimes worsens the calibration residuals (this may happen, for example, when a single calibration occurs during a period of larger than normal baseline drift). In such cases, the correction is not applied. In the future, we may compute disequilibrium fit coefficients over larger time ranges and implement the correction via a site- and time-dependent lookup table, since we do not expect the coefficients to vary over short timescales. Future analysis of variations of the coefficients across the network and over time may provide insight into the cause of the long equilibration times.

Drift- and disequilibrium-corrected values corresponding to standard gas measurements are extracted and interpolated to all times t_i (note that the CO2C2 standard is used to track the baseline, so $s_{c2} = s_b$ and $s'_{c2} = 0$). A first-order (linear) calibration curve is computed from the interpolated calibration values for each t_i using a simple least-squares regression algorithm. The fit coefficients are applied to s_i to compute χ_{CO_2} for all data. Baseline values, interpolated standard measurements, and fit coefficients and their uncertainties are archived. Calibration and target residuals are shown for a 3-day case in Fig. 5a.

In the case of the Picarro analyzer at WGC, the disequilibrium correction is applied as for the Li-cor, but no baseline is subtracted from the raw data since no baseline drift is detectable (see Sect. 6.1.3). Average linear calibration

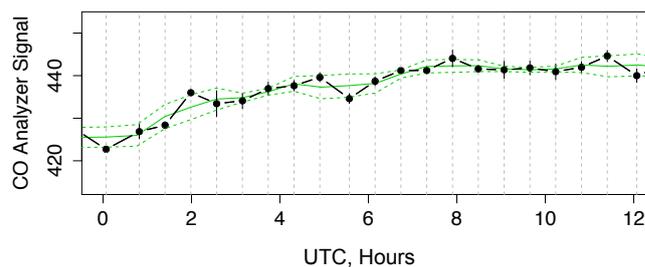


Fig. 6. Measured CO baseline signal s_b (black filled circles) linearly interpolated in time (black lines) and smoothed with a running average ($n = 3$). The baseline uncertainty u_b is the running standard deviation of the residuals (green solid curve, $n = 3$) represented by the green dashed curves. Note that it is normal for the signal from this type of analyzer to drift upward, which is the reason for the high values seen here.

coefficients are computed for each day using all calibration data within the 3-day post-processing window. The SYS-MODEs are the same as for the Li-cor.

4.1.2 CO

Because the CO analyzer signal is relatively noisy, a 2 min mean is applied to the CO analyzer output before computing the baseline. We use a moving average ($n = 3$) applied to the analyzer baseline measurements s_b to create a smoothed baseline time series b (Fig. 6). As for CO₂, b is subtracted from the raw analyzer signal s to obtain the drift-corrected signal s' . The CO analyzer gain is quite stable, so s' values for each standard are averaged across the 3-day post-processing window to minimize the impact of analyzer noise. Linear calibration coefficients are computed from a regression that includes the averaged baseline measurements ($s'_{bav} = 0$; $\chi_{CO} = 0$ ppb) and 3-day average measurements from the COC1 and COC2 standards, s'_{1av} and s'_{2av} . The resulting single set of fit coefficients is applied to s' at its native 30 s resolution. If either of the calibration gas cylinders (COC1 or COC2) are replaced during the 3-day post-processing window, then separate average s' values are computed for each cylinder, and values for all cylinders are included in the regression. Calibration and target residuals are shown for a 3-day case in Fig. 5b.

4.1.3 CH₄

The WGC Picarro is plumbed in parallel with the Li-cor, so all calibration gases are common for the two analyzers. Initially, each of the CO₂ calibration gases was also calibrated for CH₄. However, after ~ 1.5 yr, we gained confidence in the stability of the Picarro and stopped calibrating the CO2C2 standard for CH₄ since that cylinder is replaced frequently and CH₄ cylinder calibrations are time-consuming. All of the other calibration standards, including the target standard, are calibrated for CH₄, and the linear calibration coefficients are

determined using three standards instead of four. Otherwise, the post-processing is the same as described above for WGC Picarro CO₂.

4.2 Estimated uncertainty

We have developed a method to provide plausible time-varying uncertainty estimates for individual CO₂, CO, and CH₄ measurements. The uncertainty algorithms quantitatively track the major sources of error affecting the measurements. For applications like inverse modeling to estimate CO₂ surface fluxes, the most important considerations are long-term reproducibility and compatibility of measurements. That is, we need to understand the extent to which we can confidently interpret differences among measurements made at the same site and within and across networks from hour to hour, month to month, and year to year.

We separately report three aspects of the measurement error: (1) reproducibility of the values assigned to the calibration standards; (2) time-dependent analytical uncertainty for each measurement; and (3) the standard deviation of each 30 s measurement, which reflects both short-term instrument noise and atmospheric variability. These uncertainty estimates are not independent, and cannot be simply combined into a single value. For example, errors in assigned values for the calibration standards contribute to calibration curve fit residuals, and in some cases instrument precision is the dominant contributor to the standard deviation, while in other cases, real atmospheric variability dominates. The relevance of these various metrics depends on the nature of the application.

For many studies, the most important source of error when interpreting data is model-representation error, i.e., the extent to which a model with finite resolution can be expected to simulate point measurements. Inverse modeling studies often use hourly or afternoon-average data (e.g., Peters et al., 2007; Gourdjil et al., 2010; and many others). Since the system switches between different sampling heights, the data are only quasi-continuous, with valid measurements from one of three sampling heights < 13 % of the time. The standard deviation of the available measurements (typically 2 min of data four times per hour per intake height) gives an indication of the variability. Although the sampling at any one height is rather sparse, atmospheric conditions tend to persist for several hours.

4.2.1 Uncertainty of calibration standards

NOAA ESRL is responsible for maintaining the World Meteorological Organization's (WMO) mole fraction calibration scales for CO₂, CO, and CH₄. Details for each of these gases are described in Table 3. NOAA ESRL participates in ongoing standard gas and real air comparisons with a number of laboratories (Masarie et al., 2001; WMO, 2011, p. 207–211). If measurements from different laboratories or programs are

combined for a particular analysis, then any calibration-scale differences must be taken into account.

The uncertainty column in Table 3 is for χ values assigned to the primary calibration standards and encompasses absolute accuracy and precision of the scale. A more relevant metric for measurement comparability over time and across sites using the same calibration scale is reproducibility. Reproducibility was computed from repeated calibrations of cylinders separated by at least one year over the period from 2004 to present. For cases when more than two calibrations are available for a particular cylinder, all pairings are considered. We report the 68th percentile of the absolute values of the differences among all the pairs divided by the square root of two, based on the assumption that both members of a pair contribute equally to the errors in the difference. For CO, this method produces a conservative estimate because χ_{CO} has been observed to drift in many cylinders. If multiple calibrations are available for a particular cylinder, then drift can be quantified and a time- or pressure-dependent correction is applied to the assigned value, but so far the analysis has been done on the uncorrected assigned values. χ_{CO_2} has also been observed to drift in cylinders, but only rarely, and χ_{CH_4} standards are very stable. The stability of field calibration standards is discussed below.

For CH₄ in particular, reproducibility is much smaller than uncertainty. Reproducibility is not a perfect measure of the uncertainty of the standards, since systematic artifacts across the range of the calibration scale or that affect individual cylinders are not explicitly taken into account. For example, if isotopic composition were systematically different across the range of calibration standards or anomalous in a particular cylinder, then this could be a source of error that has been neglected here. Unresolved curvature of the calibration polynomial is another potential error source. Calibration curve residuals and use of multiple analytical techniques (e.g., NDIR, CRDS, GC-FID) within the NOAA calibration laboratory and across laboratories provide insight into the magnitude of possible systematic errors that may affect the values assigned to individual cylinders, but a comprehensive time-dependent analysis of these data is not yet available. The uncertainty given in Table 3 represents a conservative upper bound for these types of errors.

Field standards are calibrated in the laboratory relative to WMO standards before and after deployment. For CO₂, pre- and post-deployment calibrations are available for 177 tanks since 2004. The mean difference was 0.02 ± 0.05 ppm (post minus pre) and 14 cylinders had absolute differences > 0.1 ppm, 7 had differences > 0.15 ppm, and none had differences > 0.2 ppm. For the 59 CO standards with pre- and post-deployment calibration data, the mean difference was 3.2 ± 2.6 ppb. The distribution is strongly skewed toward positive values, with six cylinders drifting up by more than 5 ppb and 2 drifting more than 10 ppb over their lifetime. Unfortunately, post-deployment calibrations were not performed for ~ 29 CO standards prior to 2010 due to a shortage

Table 3. Uncertainty of calibration scales for CO₂, CO, and CH₄.

	Scale	Range	Uncertainty (1 σ)	Reproducibility, $\sigma_{sc}(1\sigma)$	Reference
CO ₂	WMO X2007	250–520 ppm ^a	0.069 ppm	0.03 ppm	Zhao and Tans (2006)
CO	WMO X2004	30–500 ppb ^c	0.7 %	~ 1 ppb	WMO (2010)
CH ₄	WMO X2004	300–2600 ppb ^b	0.2 %	0.31 ppb	Dlugokencky et al. (2005)

^a CO₂ standards in the range 520–3000 ppm are calibrated manometrically with an uncertainty of ~0.1 ppm. ^b Work is underway to extend the CH₄ scale to 5700 ppb. ^c The CO scale was extended 500 ppb starting in 2009. For the period 2004–2008, the upper limit was 400 ppb.

of cylinders and recurring instrument problems in the calibration laboratory. When there is a significant difference between pre- and post-deployment calibrations or when post-deployment calibration data are missing, we can use field calibration data to evaluate whether the pre- or post-deployment calibration (or perhaps a time-dependent drift correction) provides better residuals. CH₄ standards are generally stable, and field calibration residuals for CH₄ have not indicated any drift. To date, CH₄ standards have not received post-deployment calibrations.

Any errors in the assigned values of the field standards or real drift in the field standard's mole fraction will affect the reported values and should be included in uncertainty estimates. So far we have accounted for assigned value errors based on the average reproducibility of the calibrations in the ambient range (i.e., we have not assigned higher uncertainty in the case of drifting cylinders). However, for cases where pre- and post-deployment calibrations indicate drift, field calibration fit residuals are generally higher than normal. For measured atmospheric values outside the ambient range, an “extrapolation uncertainty” is assigned as described in Sect. 4.2.2.

4.2.2 Time-dependent analytical uncertainty estimates

The analytical uncertainty represents the extent to which year-to-year and site-to-site differences can be confidently interpreted. Guidelines for reporting uncertainty can be found in the “Guide to the expression of uncertainty in measurement” (GUM, 1995) and the “International vocabulary of basic and general terms in metrology” (VIM, 2008). A distinction is made between type A and type B uncertainties, where type A uncertainties can be evaluated using statistical methods, and type B uncertainties may be based on laboratory data or other information. One widely used method for tracking measurement uncertainties is to use one or more target standards that are treated as unknown samples. Variations in the measured target values and deviations from the assigned values are used to track the performance of the analyzer over time. Limitations of this approach are that information is available only for one or a few discrete χ values corresponding to the target standards, and that target standards, like calibration standards, are not exposed to all system components (e.g., inlet tubing, condensers, pumps). We have estimated measurement uncertainty using a target

standard method in combination with a statistical model that represents individual uncertainty components that affect the measurements. Algorithms for the component uncertainties use field calibration data, system data (e.g., flows, pressures), environmental data (e.g., room temperature, pressure, humidity), and laboratory test results. We have developed algorithms for the Li-cor CO₂, Picarro CO₂ and CH₄, and the Thermo Electron CO analyzer. These analyzers have diverse noise and drift characteristics. We have attempted to develop a rigorous statistical framework for uncertainty reporting, but our knowledge of the measurement uncertainty remains incomplete. The component uncertainty algorithms may be modified or new components may be added as our understanding evolves. Whenever the target method indicates that our modeled measurement uncertainty is too low, we report the larger value. The reported measurement uncertainties therefore represent our best conservative estimate.

We follow the convention of defining the independent variable x as the mole fraction values of the standards and samples and the dependent variable y as the analyzer response. More details about the statistical basis for the uncertainty estimates are provided in Appendix D, and Table D1 is a glossary of uncertainty symbols and terms. The discussion below assumes a linear analyzer response but could be generalized for nonlinear cases.

The uncertainty of a regression model may be expressed as a prediction interval, which accounts for uncertainty in the fit, uncertainty in the unknown samples, and the number of values used to compute the fit. The prediction interval can be expressed as

$$PI = \mu \pm z_{(\alpha, f)} \sqrt{se_{fit}^2 + \sigma_y^2}, \quad (1)$$

where μ is the estimated value, se_{fit} is the standard error of the fit (see Appendix D), σ_y represents the uncertainty of an individual measurement as determined by the standard deviation of the residuals of the fit, and z is a scaling factor taken from the Student t distribution that depends on the degrees of freedom f of the regression and the desired level of confidence α . For our case of χ_{CO_2} , $f = 2$, and if the desired level of confidence is 67.5 % (1 σ), then $z = 1.2938$ and for 95 % confidence $z = 4.303$. For χ_{CO} $f = 1$, and z is equal to 1.786 and 12.706 for 67.5 and 95 %, respectively. The confidence interval,

$$CI = \mu \pm z_{(\alpha, f)} se_{fit}, \quad (2)$$

represents only the uncertainty of the fit coefficients, so that if the experiment were to be run repeatedly the specified percentage of the resulting curves would fall within the confidence interval. The prediction interval describes the range of values encompassing a specified percentage of individual measurements, provided that the measurements have the same statistical uncertainty as the calibration standards (represented by σ_y). Our uncertainty model is a prediction interval, but we account for certain differences between the atmospheric measurements and the calibration standards as described below.

Equation (1) is predicated on the assumption that values of x have no error, but in our application assigned values of the calibration standards also contribute to the curve fit residuals σ_y such that

$$\sigma_y = \sqrt{\sigma'_y{}^2 + (m\sigma_x)^2}, \quad (3)$$

where σ'_y corresponds to the uncertainty of the analyzer signal, m is the slope coefficient of the fit, and σ_x represents the uncertainty of the assigned values of the calibration standards, for which a reasonable estimate is the reproducibility σ_{sc} of the cylinder calibrations (Table 3). If $\sigma_y < m\sigma_{sc}$, then we set $\sigma'_y = 0$. Note that errors in x do contribute to σ_{fit} , and this is where errors in assigned calibration standard values are taken into account.

In general, we expect that errors in unknown samples may differ from σ'_y , especially under certain anomalous conditions, such as when sampling fires or other pollution plumes where data fall outside the calibrated range or when the analyzer baseline drift is unusually large (e.g., if room temperature control is lost). We therefore attempt to model the sample uncertainty σ_u as the quadrature sum of five terms:

$$\sigma_u^2 = u_p^2 + u_b^2 + u_{ex}^2 + u_{eq}^2 + u_{wv}^2, \quad (4)$$

where σ_u and the individual components have units of mole fraction. The individual uncertainty components are described below, and typical values are given in Table 4. In the following discussion, we refer to χ_{CO_2} for simplicity, but except where otherwise noted, the same algorithms apply to χ_{CO} and χ_{CH_4} . Bold font indicates vectors containing all values within the 3-day processing window, e.g., when regression coefficients and residuals have been interpolated to all times t_i .

Analyzer precision, u_p

The analyzer precision, u_p , is estimated by interpolating the 30 s standard error of baseline measurements, s_b , to all times t_i (red curve Fig. 4b). The short-term signal-to-noise ratio for the Li-7000 CO₂ analyzer is extremely high, and the 30 s standard error for the calibration standards is typically better than 0.02 ppm, while the variability during atmospheric sampling is rarely < 0.2 ppm and often > 1 ppm. In contrast, short-term analyzer noise is large for the CO measurements,

with typical 30 s standard errors as large as 5 ppb during baseline measurements and calibrations, which is comparable to the variability observed for ambient air. The analyzer short-term precision reflects random analyzer errors and is dependent on the averaging interval.

Analyzer baseline-drift uncertainty, u_b

Unresolved temporal variations in the analyzer baseline, u_b , are estimated for Li-7000 measurements of χ_{CO_2} as follows:

1. A set of alternate realizations of the continuous baseline are created where individual baseline measurements $s_{b,i}$ have been omitted (red lines in Fig. 4a). This results in three realizations of the baseline for each time t_i (i.e., the original baseline including all available s_b and the cases where the bracketing $s_{b,i\pm 1}$ values have been excluded).
2. The standard deviation across the three unique realizations of the baseline, σ_b , is calculated for each time in t (green curve in Fig. 4b).
3. A time-varying weighting function is applied to σ_b that is zero at t_b and increases linearly to 1 halfway between successive baseline measurements, resulting in the black curve shown in Fig. 4b.
4. A threshold corresponding to the analyzer precision is applied such that $u_b \geq u_p$ (blue crosses in Fig. 4b).

This approach provides a reasonable measure of baseline-drift uncertainty in the absence of high-frequency baseline variations that are not captured by s_b . We do not expect high-frequency baseline variations when analyzer temperatures and pressures are well controlled or slowly varying. Data are screened for the presence of strong correlation between analyzer temperature and s_b . For cases where a temperature-dependent baseline correction is enabled, an additional term is included to represent the uncertainty in the baseline-temperature regression as described in Appendix C2. The target standard measurements also help to detect unresolved baseline variations as discussed in Sect. 5.1.

For χ_{CO} , u_b is the running standard deviation ($n = 3$) of the residuals from the smooth curve (Fig. 6). For Picarro χ_{CO_2} and χ_{CH_4} no baseline is subtracted, and in that case

$$u_b = \sqrt{\sigma_{C2}^2 - u_p^2}, \quad (5)$$

where σ_{C2} is the standard deviation computed over all of the CO₂C₂ or CH₄C₂ measurements within the 3-day processing window. Thus u_b accounts for imprecision on timescales of hours to days that is not already accounted for by u_p (the precision on 30 s timescales). For cases where $\sigma_{C2} < u_p$, $u_{b,i} = 0$.

Table 4. Typical values for uncertainty terms (WGC July 2011).

		CO ₂		CO	CH ₄
SENSOR		Li-cor	Picarro	Thermo Electron	Picarro
		ppm	ppm	ppb	ppb
u_p	Median	0.004	0.016	3.2	0.11
	95th %ile	0.006	0.027	5.8	0.19
u_b	Median	0.006	0.031	1.8	0.19
	95th %ile	0.017	0.036	4.2	0.23
u_{eq}	Median	0.000	0.000	N/A	N/A
	95th %ile	0.001	0.005	N/A	N/A
u_{ex}	Median	0.000	0.000	0.0	0.08
	95th %ile	0.000	0.000	0.5	0.80
u_{wv}	Median	0.001	0.002	0.0	0.01
	95th %ile	0.002	0.005	0.0	0.02
se_{fit}/m	Median	0.038	0.035	0.4	0.07
	95th %ile	0.048	0.039	1.5	0.17
σ_u	Median	0.007	0.035	3.8	0.24
	95th %ile	0.018	0.039	6.6	0.83
σ'_y/m	Median	0.067	0.060	0.0	0.00
	95th %ile	0.087	0.067	2.3	0.00
$u_M (1\sigma)$	Median	0.109	0.098	5.2	0.31
	95th %ile	0.135	0.107	9.3	0.90
$u_R (1\sigma)$	Median	0.006	0.016	3.3	0.11
	95th %ile	0.009	0.027	6.1	0.19
u_{TGT}	Median	0.060	0.070	5.0	0.31
	95th %ile	0.072	0.090	7.6	0.50
$SD_M (30\text{ s})$	Median	0.056	0.039	2.8	0.23
	95th %ile	0.420	0.270	6.7	1.51

Extrapolation error, u_{ex}

An empirically determined (type B) extrapolation uncertainty u_{ex} is applied for values outside of the calibrated range. For the Li-7000, laboratory measurements of cylinders with assigned values of 550 and 660 ppm were performed using three different systems prior to deployment. The GUM guidelines dictate that when systematic errors are present, a correction should be applied to the data. However, test results from individual analyzers varied widely, and not all analyzers have been tested. Furthermore, only a small percentage of data fall outside the calibrated range. We therefore do not apply a correction, and we define u_{ex} :

$$u_{ex} = \varepsilon \left| \chi_{CO_2} - \chi_{CO_2} [CO_2C4] \right|, \quad \text{case } \chi_{CO_2,i} > \max(\chi_{CO_2CAL}) \quad (6a)$$

$$u_{ex} = \varepsilon \left| \chi_{CO_2} - \chi_{CO_2} [CO_2C1] \right|, \quad \text{case } \chi_{CO_2,i} < \min(\chi_{CO_2CAL}). \quad (6b)$$

The maximum measured minus assigned value was 0.0075 ppm per ppm above 460 ppm. We assume that value represents the semi-range of likely extrapolation errors and divide by $\sqrt{3}$ to estimate a corresponding standard deviation (valid for a uniform distribution). This gives $\varepsilon = 0.004$ for

Li-7000 measurements of χ_{CO_2} . For CO, we use $\varepsilon = 0.02$ (based on linearity specifications), and for the WGC Picarro CO₂ and CH₄, we use ε values of 0.001 and 0.005, respectively (based on a single lab test as described above for the Li-7000).

Equilibration uncertainty, u_{eq}

When switching between sampling modes, the time required before equilibration errors become negligible is proportional to the χ_{CO_2} difference between successive modes. We use the routine field calibration data to correct for differences from the equilibrium value (see Sect. 3.1 and Appendix C1). Errors in the disequilibrium correction are described by

$$u_{eq} = \frac{1}{m} \sigma_{eq} (\Delta \chi_{CO_2}), \quad (7)$$

where $\Delta \chi_{CO_2}$ is a vector of the differences between successive sampling modes, m is the analyzer gain interpolated to all times t_i , and σ_{eq} is the standard deviation of the residuals of the disequilibrium correction. For Li-cor CO₂ and Picarro CO₂ and CH₄ measurements, u_{eq} is negligibly small under normal operating conditions (Table 4). No disequilibrium correction or equilibration uncertainty algorithms have

yet been implemented for CO, since the signal is too noisy to reliably estimate these quantities using calibration data with a 3-day processing window.

Uncertainty associated with water vapor, u_{wv}

In the case of absorption measurements, the presence of water vapor can cause spectral interference and can change the pressure broadening of the absorption lines. Water vapor also “dilutes” the mole fraction. The Li-7000 and Picarro analyzers report dilution-corrected output according to

$$\delta_{wv} = \chi_{CO_2} \left(\frac{1}{1 - \Delta\chi_{H_2O}} \right) - \chi_{CO_2}, \quad (8)$$

where δ_{wv} is the dilution correction and $\Delta\chi_{H_2O}$ is a vector of the humidity differences between samples and standards expressed as a unitless mole fraction. A value of $\Delta\chi_{H_2O} = 1 \times 10^{-4}$ (100 ppm) corresponds to $\delta_{wv} = 0.04$ ppm dilution correction for $\chi_{CO_2} = 400$ ppm. Our system minimizes humidity differences between the sample airstream and standards by passing sample and standard gas through Nafion membrane dryers as discussed in Sect. 2.3. The sample air is dried, while standard gases are humidified. We track the difference between the instantaneous measured χ_{H_2O} and χ_{H_2O} measured during calibration modes and interpolated to all times t_i and use this $\Delta\chi_{H_2O}$ in Eq. (8). Typical $\Delta\chi_{H_2O}$ values are $< 1 \times 10^{-5}$ (10 ppm). The CO analyzer does not report dilution corrected values, but typical values for δ_{wv} are < 0.01 ppb.

We have not evaluated the fidelity of the Li-7000 or Picarro water vapor corrections and instead have relied on maintaining negligible humidity differences between the sample and standard air streams. We set $u_{wv} = \delta_{wv}$, which is equivalent to assuming a 100 % error in the dilution correction (and no errors from pressure broadening or spectral interference). Values are insignificant for all gases when the system is functioning normally (see Table 4), and this term serves as a convenient metric to gauge when humidity variations are large enough to potentially cause significant errors for χ_{CO_2} and other gases.

Figure 7a shows a high-humidity case when counterflow to the purge side of the Nafion dryers was lost due to the unintentional release of a pressure relief valve. Residuals for CO₂ standards were < 0.06 ppm for this period, similar to adjacent periods where the drying system was operating normally. That the Li-7000 water calibration is unreliable at low humidity is indicated by the negative values of χ_{H_2O} in Figs. 7a and 8b. Frequent H₂O calibration would be required in order to achieve accurate water vapor measurements from the Li-7000, which would require substantial effort and additional complexity without obvious benefit. The upstream chiller temperature is controlled at 1.7 °C and pressure in the chiller is ~ 1700 hPa, corresponding to a saturation water vapor mole fraction of ~ 4100 ppm, a reasonable upper limit for χ_{H_2O} during this time. Tests to evaluate possible

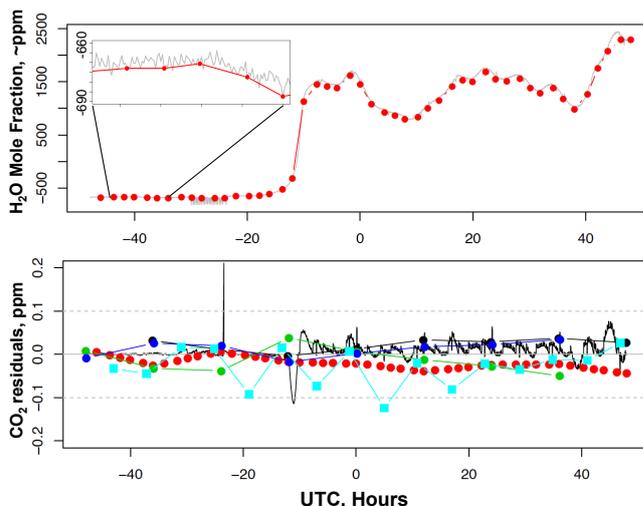


Fig. 7. (a) χ_{H_2O} reported by the Li-7000 at the SNP site for 13–15 February 2010, when the Nafion dryer counterflow was lost. Red symbols correspond to baseline calibration measurements. (b) The corresponding values of δ_{wv} (black curve) along with the CO₂ calibration and target residuals (CO2C1 = black circles, CO2C2 = red circles, CO2C3 = green circles, CO2C4 = blue circles, CO2TGT = cyan squares). Negative χ_{H_2O} values result from inaccurate (manufacturer-specified) zero-offset values for the Li-7000, but relative changes can be interpreted with some confidence.

differences in CO₂ permeation across the Nafion membrane for samples versus standards are described in Sect. 2.3 and showed negligible differences under normal operation, but we have not yet tested for differences in CO₂ permeability of Nafion between samples and reference gases under anomalous high-humidity conditions like in Fig. 9. Figure 9 does not show systematic χ_{H_2O} differences between samples and standards, but rather nonnegligible δ_{wv} values arise from unresolved temporal variations that follow room temperature.

Other sources of error and uncertainty

There are some potential sources of error that cannot be reliably detected from our available engineering data in an automated way. Two examples are (1) contamination related to the long sampling lines, pumps, and chillers that are upstream of where the calibration gases are injected and that are exposed to ambient humidity, temperature, and pressure, and (2) undetected leaks of room air or ambient air from a lower altitude into the sample airstream. We have relied on laboratory tests, field diagnostics, and comparison with independent measurements to assess the likely impact of these errors. Many independent tests over a wide range of conditions have been performed and are described in Sect. 5.

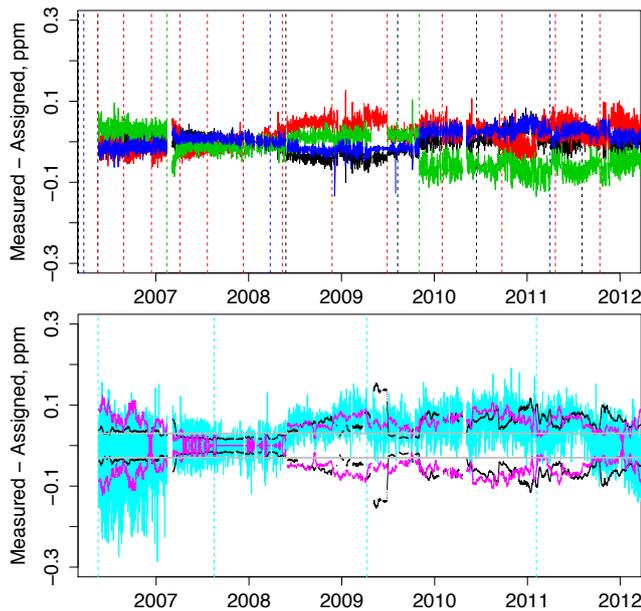


Fig. 8. Standard gas residuals (i.e., assigned minus measured χ_{CO_2} values) for (a) CO2C1 (black), CO2C2 (red), CO2C3 (green), CO2C4 (blue) calibration and (b) CO2TGT (cyan) standards at WKT for the period 2006–2011. Dashed vertical lines in both panels correspond to dates when standards were replaced. The black and magenta curves in (b) correspond to the 10-day running means of u_{M} and u_{TGT} , respectively, where u_{M} corresponds to the larger of the values produced by Eq. (9a) and (9b). From April to July 2009, u_{M} is higher than normal, because during that time no CO2C3 calibrations were run, so z in Eq. (9) corresponds to 1 degree of freedom instead of 2 during that time.

4.2.3 Uncertainty derived from target cylinder measurements

The target cylinder measurements provide an independent check on our estimated uncertainty values, and we define u_{TGT} to be the 67th percentile of the absolute difference between target measured and assigned values within the 3-day processing window. This is based on the idea that the difference between the measured and assigned target values should fall within the 1σ measurement uncertainty 67% of the time. Errors in the assigned value of the target standard affect u_{TGT} , and as for calibration cylinders, we use σ_{sc} as a measure of the assigned value errors. The assigned value is constant over the lifetime of the cylinder, and therefore is a bias rather than a random error. Assigned value errors may cause u_{TGT} to be either too large or too small. Since we do not know the sign of the error, we do not make a correction. A more conservative approach would be to use $(u_{\text{TGT}}^2 + \sigma_{\text{sc}}^2)^{1/2}$, but this would consistently overestimate measurement uncertainty (though typically by < 0.01 ppm in the case of CO₂). Note that although u_{TGT} is represented as a vector, a single value is computed for each 3-day processing window.

4.2.4 Uncertainty reporting

We aim to provide uncertainty information to data users that is complete but not overly complicated. Along with the measured values of χ_{CO_2} , χ_{CO} , and χ_{CH_4} , we report the estimated measurement uncertainty u_{M} , which is the largest among

$$u_{\text{M}} = \sqrt{(z_{(\alpha, f)})^2 \left(\frac{se_{\text{fit}}}{m}\right)^2 + \sigma_{\text{u}}^2}, \quad (9a)$$

$$u_{\text{M}} = z_{(\alpha, f)} \sqrt{\left(\frac{se_{\text{fit}}}{m}\right)^2 + \left(\frac{\sigma'_{\text{y}}}{m}\right)^2}, \quad (9b)$$

$$u_{\text{M}} = u_{\text{TGT}}, \quad (9c)$$

$$u_{\text{M}} = \sigma_{\text{sc}}. \quad (9d)$$

Here, se_{fit} is the standard error of the calibration regression given by Eq. (D5), m is the time-interpolated analyzer gain (i.e., the slope calibration coefficient), σ_{u} is the modeled uncertainty of the atmospheric data from Eq. (4), σ'_{y} is the standard deviation of the calibration fit residuals less the contribution from the assigned values of the calibration standards (see Eq. 3), u_{TGT} is measurement uncertainty estimated from the target standards described in Sect. 4.2.3, and σ_{sc} is the time-invariant reproducibility of the calibration scale given in Table 3. Note that u_{M} given by Eqs. (9a) or (9b) is equivalent to the uncertainty term for the prediction interval described by Eq. (1) but with units of mole fraction. The factor z is chosen to give a 67.5% prediction interval, corresponding to 1σ for normally distributed errors. In Eq. (9a), σ_{u} is not multiplied by z , since the individual modeled uncertainty components do not depend on the degrees of freedom of the regression. In contrast, σ'_{y} in Eq. (9b) does depend on the degrees of freedom (see Eqs. 3 and D4). In general, we expect that the value of σ_{u} given by Eq. (4) will be comparable to or greater than $z\sigma'_{\text{y}}/m$. However, for the Li-cor and Picarro CO₂ measurements, our model of σ_{u} produces values that are too small to account for the calibration gas residuals, as can be seen in Table 4. This discrepancy is discussed further in Sect. 5.1. For CO and CH₄, σ_{u} and $z\sigma'_{\text{y}}/m$ are generally comparable.

Data and uncertainties are archived and made available at their native 30 s resolution. When estimating uncertainties for aggregated data (e.g., hourly or afternoon averages), random components should be weighted according to the number of available measurements, but systematic errors should not. We therefore separately report the random component of the uncertainty, u_{R} , which includes u_{p} and any portion of u_{b} that is random on timescales of seconds to minutes. For the Li-cor, where the floor for the baseline uncertainty is u_{p} , the random uncertainty is given by

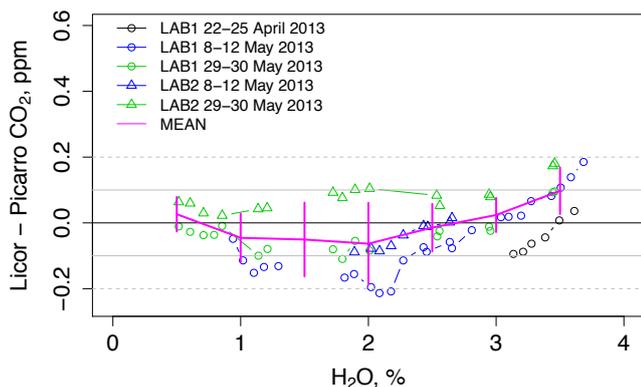


Fig. 9. Summary of recent dried Li-cor versus undried Picarro comparisons using wetted ambient air that was passed through a mixing volume. Comparisons used two separate Picarro systems and two tall tower Li-7000 systems (LAB1 and LAB2). Mean results are shown in magenta, where the error bars represent one standard deviation computed over two to four experiments.

$$u_R = \sqrt{2u_p^2}. \quad (10a)$$

For CO,

$$u_R = \sqrt{u_p^2 + \frac{\sigma_z^2}{3}}, \quad (10b)$$

where σ_z is the standard error corresponding to the 2 min averaged COZER measurements (interpolated to all times t_i), since the smoothed baseline is an $n = 3$ running mean. For the Picarro CO₂ and CH₄ where no baseline is subtracted,

$$u_R = u_p. \quad (11)$$

We also report the standard deviation SD_M of the measured value, which reflects both analyzer precision and real atmospheric variability during the measurement period. Atmospheric variability AV is given by

$$AV = \sqrt{SD_M^2 - u_R^2}. \quad (12)$$

If the atmospheric variability is not detectable above the random uncertainty (i.e., if $SD_{M,i} < u_{R,i}$), then AV_i is undefined.

5 Evaluation

5.1 Calibration residuals and target measurements

The calibration curve residuals and target tank measurements from WKT for nearly six years are shown in Fig. 8. WKT has the longest time series for the new analyzer, and the results are typical of other sites.

The residuals for individual tanks are obviously not randomly distributed around zero, and time-dependent biases approaching 0.1 ppm are seen for some cylinders. Errors in the assigned values for the reference gases are characterized by σ_{sc} , which is 0.03 ppm for CO₂, and so expect that 98 % of the calibration residuals should fall within ± 0.06 ppm. We use the linearized, temperature- and water-corrected CO₂ signal from the Li-cor and apply a first-order calibration polynomial as described in Sect. 3.2. The residuals are not significantly improved by adding a quadratic coefficient. Our uncertainty model described in Sect. 4.2.2 indicates that baseline-drift or inadequate correction for disequilibrium of the calibration signals is not to blame. We suspect that imperfect delivery of standard gases to the analyzer may contribute, such as small leaks or artifacts caused by pressure variations in the Nafion dryer or other components. Errors in linearization could also play a role (e.g., inaccurate specification of the reference concentration used by the Li-7000 internal linearization algorithm). The pattern of residuals may or may not change when a tank is replaced, and sometimes the residuals vary slowly, in a manner that suggests that the CO₂ concentration of one or more of the cylinders might be unstable, which can occur, for example, in the case of a slow and/or temperature-dependent leak. Final calibration data are not available for the CO2C3 standard that was installed in fall 2009, and the residuals near the end of the record may improve when post-deployment calibration data become available. The modeled measurement uncertainty σ_u for CO₂ is often too small to account for the calibration residuals, and in such cases the reported measurement uncertainty u_M corresponds to Eq. (9b). In the future, we may include an additional term in σ_u in order to explicitly account for uncertainty related to imperfect delivery of standard gases or errors in linearization.

The frequency of calibrations was reduced in early 2009 as described in Sect. 3.2. Figure 8a shows that the CO2C2 standard was replaced as often as three times per year during 2007 and 2008. The typical lifetime of a CO2C2 cylinder is 6 months. Target tanks last 1.2 yr, and CO2C1, CO2C3, and CO2C4 last 2.3 yr. The stability of the residuals over timescales of days to weeks suggests that the Li-cor calibration frequency could be further reduced to something on the order of once per 23 h, which would double the lifetime of CO2C1, CO2C3, and CO2C4. Li-cor baseline-drift monitoring should continue at the current frequency of once per 2 h but could be done with air from an uncalibrated cylinder or a zero-air generator. Whenever possible, we try to avoid replacing more than one standard at a time, so that any substantial change in residuals can be unambiguously attributed to a particular cylinder.

Figure 8b shows individual measured minus assigned target values along with 10-day running mean values for u_{TGT} and u_{M} , where here u_{M} corresponds to the larger of the values produced from Eqs. (9a) or (9b). The target measurements provide a separate measure of the uncertainty excluding any errors resulting from inlet components that are upstream of the calibration manifold (e.g., intake filters, sampling lines, pumps, and condensers). Target residuals were unacceptably noisy during the first six months of operation and were greatly improved after a site visit in March 2007, when the output flow from the air conditioner in the trailer was directed away from the analyzer. After the adjustment to the air conditioner vent, u_{M} and u_{TGT} have similar magnitude and temporal variability. The estimated uncertainty u_{M} is too low during this initial period to account for the target residuals early in the record. The Li-cor CO₂C2 measurements were not strongly correlated with room temperature or analyzer temperature, so the only indication of a problem was the target measurements that were not adjacent to baseline checks or full calibrations. This example illustrates the utility of having multiple target measurements distributed throughout the day that are temporally separated from other reference gas measurements for detecting problems that are not otherwise apparent and that may depend on the diurnal variation of room temperature.

5.2 Laboratory and field tests

We have evaluated the uncalibrated system components (e.g., inlet components, pumps, and chillers) in the laboratory and with field studies at the BAO tower. Results from selected tests are described below.

5.2.1 Laboratory tests of wetted air

The lab is equipped with a high-volume dynamic dilution system that provides a large volume (> 20 slm) of air with well-mixed and slowly varying χ_{CO_2} at super-ambient pressure (134.5 kPa). During our design- and early deployment phase, we developed a setup for testing the sample inlets under wet conditions, where air from the dilution system was routed through a bubbler and then split into three separate airstreams, two of which were routed to sample inlet ports and the third was passed through an MgClO₄ trap and into the CO₂TGT port on the calibration manifold. The setup included a bypass for the bubbler so that we could test for differences using either dry or wetted air. When using a system equipped with a stream-selection valve (ECMT; VICI Valco, USA) valve in place of the solenoid manifold, we found that (1) air sampled through the two sample inlet ports consistently agreed to within 0.01 ppm whether dry or wetted, (2) dry air sampled through the MgClO₄/CO₂TGT port was 0.04 ppm higher than dry air sampled through the sample inlets, and (3) wetted air sampled through the MgClO₄/CO₂TGT port was 0.04 ppm lower than wetted air

sampled through the sample inlets. When using a system with an aluminum solenoid valve manifold, we found that wetted air sampled through the MgClO₄/CO₂TGT port was 0.15 ppm lower than when sampled through the sample inlets. Informed by these test results, we subsequently switched to steel solenoid valves that perform similarly to the Valco manifold. In another test on a system with an aluminum solenoid valve manifold, we provided wetted air to all three sample channels, and one channel had a dry-ice trap installed immediately before the inlet. The wet channels measured 0.05 to 0.13 ppm higher than the dried channel.

More recently, we have performed wetted air comparisons in the lab using two independent Picarro analyzers (model G2401-m) and two nearly independent tall tower systems (C1, C3 and C4 standards were shared between the two Li-7000 systems, but they had separate C2 and TGT standards and independent inlet systems). We sampled ambient air through a mixing volume and used a bubbler to vary the humidity from 0.5 to 3.5 %. Laboratory temperatures ranged from 22 to 26 °C. Data from the Picarro analyzers were water-corrected using analyzer-specific coefficients (following Chen et al., 2010; Rella et al., 2013), and linear CO₂ calibration coefficients were applied according to lab calibrations. Results are shown in Fig. 9 and in Table 8. Individual Li-cor minus Picarro CO₂ differences were within ± 0.2 ppm over the range 0.5–3.5 %, and mean differences were within ± 0.1 ppm. Errors in the χ_{CO_2} measurements from the analyzers, any errors in the Picarro water corrections, and potential problems delivering air to the analyzers are all sources of uncertainty in these comparisons. The difference between the two Li-7000 systems of approximately 0.2 ppm during the tests on 29–30 May suggests that errors resulting from the test setup were nonnegligible, since we routinely achieve agreement better than 0.1 ppm between these systems when sampling ambient air.

Although laboratory tests using artificially wetted airstreams are susceptible to artifacts, we include these results because of the lack of satisfactory high-humidity real air comparisons. Most of the lab comparisons show differences < 0.1 ppm. None of the tests indicate water-related artifacts larger than 0.2 ppm (arising from either the tall tower analysis system or the test setup), so we conclude that 0.2 ppm is a conservative upper limit for CO₂ sampling errors under high-humidity conditions.

5.2.2 Laboratory tests of humid ambient air sampled through long and short inlet tubes

We evaluated the impact of sampling through a long inlet tube under ambient conditions with moderate humidity levels ($\chi_{\text{H}_2\text{O}} \sim 0.75$ %, 30 % RH, 18 °C, 833 hPa). Outdoor air was sampled through an integrating volume so that variability on timescales of minutes would be damped. The resulting 30 s standard deviations were ~ 0.02 ppm, but nighttime CO₂ changes were as large as 10 ppm over 15 min. Two of the

Table 5. Annual summary of flask minus in situ χ_{CO_2} .

	Median \pm Standard Deviation (Number of Samples) Dry Air Mole Fraction, ppm					
	2006	2007	2008	2009	2010	2011
LEF*	0.11 \pm 0.5 (90)	-0.10 \pm 0.4 (393)	-0.08 \pm 0.3 (401)	0.05 \pm 0.3 (175) 0.12 \pm 0.3 (169)	0.18 \pm 0.4 (309)	0.28 \pm 0.5 (235)
LEF Manual	-0.15 \pm 0.4 (46)	-0.10 \pm 0.3 (53)	-0.04 \pm 0.3 (66)	-0.02 \pm 0.2 (58)	0.04 \pm 0.2 (55)	0.00 \pm 0.3 (56)
WKT*	0.06 \pm 0.3 (107)	-0.38 \pm 0.8 (63) 0.04 \pm 0.1 (23)	0.02 \pm 0.3 (181)	0.10 \pm 0.3 (175)	0.08 \pm 0.4 (236)	0.15 \pm 0.4 (249)
AMT*			0.39 \pm 0.3 (23)	-0.00 \pm 0.3 (94)	0.20 \pm 0.4 (226)	0.16 \pm 0.3 (220)
BAO		-0.00 \pm 0.7 (31)	-0.04 \pm 0.4 (215)	-0.06 \pm 0.5 (211)	0.19 \pm 0.5 (235)	0.27 \pm 0.5 (217)
WBI		0.15 \pm 0.4 (115)	0.12 \pm 0.5 (192)	0.27 \pm 0.5 (151)	0.29 \pm 0.6 (260)	0.31 \pm 0.5 (231)
WGC		0.17 \pm 0.5 (31)	0.11 \pm 0.4 (112)	0.18 \pm 0.5 (147)	0.14 \pm 0.5 (130)	0.28 \pm 0.6 (58)
SCT				0.16 \pm 0.4 (223)	0.32 \pm 0.6 (252)	0.20 \pm 0.4 (201)

* New rows within a site entry correspond to significant configuration changes as described in the text. Note: PFP samples have not been collected at SNP site because of logistical challenges.

three sample inlets were connected to the integrating volume through short inlet tubes, and the third was sampled through a 76 m coil of 1.27 cm (0.5 in.) OD Synflex tubing. The three inlets were sampled consecutively as in normal operations, dwelling on each intake for 5 min. Measured CO₂ differences among the tubes were < 0.02 ppm during well-mixed midday conditions. The mean χ_{CO_2} over a 5 h sampling window was 395.59 \pm 0.25 (SD) ppm. During that period, there were sixteen or more independent measurements from each inlet, so the standard error of the mean value was \sim 0.06 ppm. Adjacent nighttime periods had much higher variability and also showed no significant differences across the sampling inlets. Winderlich et al. (2010) also evaluated possible sampling artifacts related to long Synflex inlet tubes and reported negligible differences for CO₂ and CH₄ for ambient air sampled through 2 and 200 m tubes under wintertime conditions (H₂O < 1 %).

5.2.3 Tank air sampled through BAO inlet tubes

The Boulder Atmospheric Observatory tall tower is a research platform equipped with two elevators. Two reference gas cylinders (assigned values χ_{CO_2} = 371.59 and

401.89 ppm and χ_{CO} = 115.7 and 133.5 ppb) were taken to the top of the tower (300 m) and sampled through the tower inlet tubes. The measured minus assigned values were -0.13 and -0.11 ppm for χ_{CO_2} and were -3.2 and +0.86 ppb for χ_{CO} .

6 Comparison with independent measurements

6.1 Ongoing colocated flask sampling

The tall tower sites are equipped with automated flask-sampling systems that are known as programmable flask packages (PFPs) that normally collect daily or alternate day midafternoon air samples. One site, LEF, is also equipped with a manual flask-sampling system that uses flasks from our laboratory's global Cooperative Air Sampling Network (<http://www.esrl.noaa.gov/gmd/ccgg/flask.html>; Conway et al., 1994). LEF manual flask samples are collected in pairs once per week. Routine comparison of flask and in situ sampling tracks the level of measurement compatibility within our own laboratory and is useful for identifying experimental problems in either system. A very useful feature of the flask-sampling strategy is that we analyze each PFP flask for about

50 compounds, including greenhouse gases, isotopic composition of CO₂, hydrocarbons, and halocarbons. This results in a wealth of data that can be used in the interpretation of observed patterns in the major greenhouse gases, enabling at least partial attribution to specific sources/processes.

The tower PFP sampling strategy has evolved over time, particularly during 2006–2008, and is subject to logistical constraints particular to individual sites. At most sites, PFP samples are drawn from the highest sampling level on the tower through a dedicated inlet and sample tube. To provide a truly independent measurement, the PFP does not share a sampling tube with the in situ system, except when only one suitable tube is available. More information about the PFP sampler and tower installation is provided in Appendix E1. A modified version of the PFP compressor package has recently been developed that includes an integrating volume and uses a variable flow rate to provide integrated sampling over ~1 h (Turnbull et al., 2012). Winderlich et al. (2010) have successfully deployed an integrated sampler at the Zotino Tall Tower Observatory. Whether integrated versus grab sampling is appropriate for a particular application depends on several factors, especially proximity to emissions sources. Flask versus in situ CO₂ and CO comparisons for quality control would likely be simplified by integrated sampling, but quality control is just one aspect of the PFP sampling objectives for our network. We plan to continue with grab sampling until we are able to thoroughly evaluate an integrating sampler and have considered the implications for data analysis on a site-by-site basis.

In situ and PFP flow rates vary from site to site and depend on pump performance, which may change over the time and with temperature. The in situ systems switch among three sampling heights, so only quasi-continuous data are available for a particular level. Accurate measurement of all flow rates would be needed to ensure synchronous sampling of the PFP and in situ systems, but so far limited flow information is available for the PFP samples. Before July 2012, we simply triggered the flask samples at a fixed time of day and compared the closest temporal match within a specified window. Comparisons can be filtered to select periods with low atmospheric variability, but CO₂ variability is rarely lower than our target precision of 0.1 ppm. The minimum time difference between flask and in situ measurements is generally < 8 min, and midafternoon atmospheric variations are mainly random on that timescale. If atmospheric variability is the main source of difference between flask and in situ measurements, then a variability threshold of 0.5 ppm should yield standard errors ≤ 0.1 ppm when comparison data are aggregated to monthly means. However, comparisons for individual samples are confounded by atmospheric variability and have limited utility. A new sampling sequence instituted in 2012 dwells for approximately 20 min on the appropriate intake height whenever a PFP sample is triggered and enables more informative comparisons.

Annual flask versus in situ comparisons for CO₂ and CO are summarized in Tables 5 and 6 for PFP samples with hourly in situ standard deviations of < 0.5 ppm for CO₂ and 10 ppb for CO. Table 7 shows PFP comparisons with the Picarro CO₂ and CH₄ data at WGC. Prior to October 2007, samples at WKT were collected from the 122 m inlet, and have been collected from 457 m since that time from a line shared with the in situ system. The LEF in situ system was upgraded in May 2009, and a separate PFP intake was installed to 396 m. Prior to that time, PFP samples were collected from a shared intake at 244 m. The AMT in situ system was replaced in February 2009, and only a handful of samples are available for comparison with the old system, which was performing poorly near the end of its lifetime.

Through 2008, CO₂ PFP versus in situ agreement throughout the network was close to the 0.1 ppm compatibility target recommended by the WMO. However, agreement has worsened during more recent years and PFP CO₂ measurements are now systematically higher than the in situ values throughout the network, with some sites approaching offsets of 0.3 ppm during 2011. Manually sampled flasks at LEF show consistently good agreement with the in situ system before and after the May 2009 upgrade. Additional statistics for LEF manual flask samples are given in Table 8 and show differences < 0.1 ppm during both summer and winter. CO comparisons are generally satisfactory, with most annual mean differences ≤ 3 ppb. The CO standard deviations at AMT since 2009 are higher than for most sites because that CO analyzer is very noisy (typical $u_p > 8$ ppb). Agreement between in situ and PFP CH₄ measurements at WGC is ≤ 1 ppb for all years except for 2007.

Karion et al. (2013) also evaluate PFP versus in situ measurements for routine aircraft flights over Alaska from 2009 to 2011. They report PFP minus in situ values for CO₂ of 0.20 ± 0.37 , when data are filtered to exclude periods of high variability, which is consistent with our results for those years. Stephens et al. (2011) compared PFP versus in situ results from a high-altitude site (Niwot Ridge, CO) for August 2005–early 2011 and reported differences with comparable magnitude but opposite sign ($-0.17 \text{ ppm} \pm 0.38 \text{ ppm}$, $n = 745$). However, Niwot Ridge PFP versus in situ comparisons since 2010 do show a trend with increasingly positive PFP minus in situ values that is consistent with our results (B. Stephens, personal communication, 2012).

Figure 10a shows the time series of PFP minus in situ CO₂ differences from BAO for samples collected when the standard deviation of in situ data within a 1.25 h window was < 0.5 ppm. The mean (median) of the individual differences is 0.12 (0.07) ± 0.49 ppm (1σ), with 67 % of the absolute monthly mean differences < 0.19 ppm and 95 % < 0.47 ppm. Months with fewer than five individual comparisons are excluded. In late 2009, the monthly mean bias shifted from negative to positive. The mean value for December 2010 was 0.79 ppm.

Table 6. Annual summary of flask minus in situ χ_{CO} .

		Mean \pm Standard Deviation (Number of Samples)				
		Dry Air Mole Fraction, ppb				
	2006	2007	2008	2009	2010	2011
LEF				1.8 \pm 2.9 (271)	2.2 \pm 2.8 (395)	2.4 \pm 2.9 (264)
LEF Manual				0.3 \pm 9.5 (44)	2.8 \pm 12.2 (80)	4.7 \pm 8.0 (89)
WKT*	-1.3 \pm 4.9 (162)	-1.1 \pm 4.8 (263) 1.3 \pm 4.1 (34)	1.8 \pm 4.4 (287)	0.8 \pm 6.8 (263)	2.7 \pm 5.2 (338)	1.6 \pm 5.3 (273)
AMT*			1.7 \pm 2.9 (31)	2.3 \pm 8.1 (111)	2.7 \pm 7.5 (276)	2.0 \pm 9.3 (157)
BAO		-1.2 \pm 13.1 (64)	-0.4 \pm 6.3 (326)	-2.7 \pm 7.1 (305)	-2.8 \pm 6.4 (330)	-3.0 \pm 7.1 (282)
WBI		-0.2 \pm 4.0 (262)	1.1 \pm 5.2 (303)	0.2 \pm 4.8 (336)	0.9 \pm 5.5 (448)	1.4 \pm 5.7 (343)
WGC		0.8 \pm 4.6 (62)	1.9 \pm 4.6 (263)	2.3 \pm 5.0 (262)	2.6 \pm 8.5 (275)	2.8 \pm 4.4 (126)
SCT			0.4 \pm 7.3 (18)	0.3 \pm 5.3 (389)	1.3 \pm 5.2 (393)	1.6 \pm 5.0 (265)

* New rows within a site entry correspond to significant configuration changes as described in the text.

Table 7. WGC Picarro comparisons.

		Mean \pm Standard Deviation (Number of Samples)				
		Dry Air Mole Fraction, ppb				
		2007	2008	2009	2010	2011
CO ₂ (ppm)	Picarro minus PFP	0.21 \pm 0.4 (32)	0.15 \pm 0.4 (93)	0.15 \pm 0.4 (150)	0.10 \pm 0.5 (136)	0.20 \pm 0.6 (72)
CH ₄ (ppb)	Picarro minus PFP	3.1 \pm 3.2 (31)	-1.1 \pm 3.4 (134)	-0.7 \pm 3.1 (188)	0.2 \pm 2.8 (203)	-1.3 \pm 3.1 (89)

Several lines of evidence, including laboratory tests and a Picarro–Li-cor–PFP comparison at BAO described below, point toward biases in an increasing number of the PFP samples as the driver of increasingly positive PFP versus in situ differences. Ongoing laboratory experiments show enhanced CO₂ in some PFP flasks when water vapor is present. The CO₂ enhancements measured in the laboratory span an order of magnitude from approximately 0.1 ppm to nearly 2 ppm. It appears that modifying the sampling protocol so that the PFPs are pressurized with ambient air prior to collecting the sample may eliminate the biases. We are testing a simple strategy at BAO where flasks are flushed and pressurized with ambient air approximately two hours prior to the desired sampling time. That air is vented when the sample is

otherwise collected as usual. More details and preliminary results using the new protocol are provided in Appendix E2.

6.2 Picarro–Li-cor comparison and intensive flask sampling at BAO

Starting in September 2011, we configured the BAO in situ system to dwell on the 300 m intake and commenced a series of experiments to investigate strategies for improving in situ versus flask agreement. Leak checks had been performed on both the PFP and in situ sampling lines in June 2011. A Picarro analyzer (Model G1301) was installed on the PFP intake from 9 September until 28 October 2011 and reported 2 s data. For the first several days, no PFP samples were collected in order to enable an uncomplicated comparison

Table 8. Summary of comparisons between NOAA ESRL Tall Tower Li-cor CO₂ and other CO₂ measurements.

Site	Date	Instruments	CO ₂ Difference, ppm (Other– NOAA TT)	Comparison Type
BAO	9–12 Sep 2011	Picarro 30 s	0.04 ± 0.06 (<i>n</i> = 6982)	Independent intake to 300 m (Rella et al., 2013 their Fig. 26)
	30 Sep– 28 Oct 2011	hourly, SD < 0.3	0.0 ± 0.03 (<i>n</i> = 193)	
BAO	15–18 Nov 2007	P-3 Li-cor 30-sec, SD < 0.5 (inter- mittent)	0.16 ± 0.20 (<i>n</i> = 118)	P-3 instrument on elevator
BAO	29 Jul–1 Aug 2008	P-3 Li-cor 30-sec, SD < 0.5	0.04 ± 0.06 (<i>n</i> = 3130)	Shared intake line to 300 m
WBI	Jan–Oct 2010	PSU CRDS		Separate intake to 99 m (Richardson et al., 2012)
		5-min	–0.12 ± 1.37	
		afternoon average	–0.13 ± 0.63	
		Jul/Aug, 16:00–17:00 LST	–0.33 ± 0.83	
WKT	13 Sep 2006	P-3 Li-cor 10 min*	0.02 ± 0.17	Aircraft Spiral (Peischl et al., 2010)
	25 Sep 2006	10 min*	–0.03 ± 0.23	
BAO	1 Apr 2008	P-3 Li-cor 20 min*	0.01 ± 0.27	Aircraft Spiral (Peischl et al., 2010)
LEF	Jun 2009–May 2013	Manual Flasks		Shared line to 300 m Hours where in situ SD < 0.5 ppm 3-sigma outliers removed (7 samples)
		All months	0.03 ± 0.27 (<i>n</i> = 199)	
		JJA	0.05 ± 0.36 (<i>n</i> = 57)	
		DJF	–0.07 ± 0.26 (<i>n</i> = 116)	
LAB	Apr–May 2013	Picarro		Laboratory tests of ambient air with added humidity (see Sect. 5.2)
		χ _{H₂O} = 1 %	–0.05 ± 0.07 (<i>n</i> = 3)	
		χ _{H₂O} = 2 %	–0.06 ± 0.12 (<i>n</i> = 4)	
		χ _{H₂O} = 3.5 %	0.10 ± 0.07 (<i>n</i> = 4)	

* Duration of spiral.

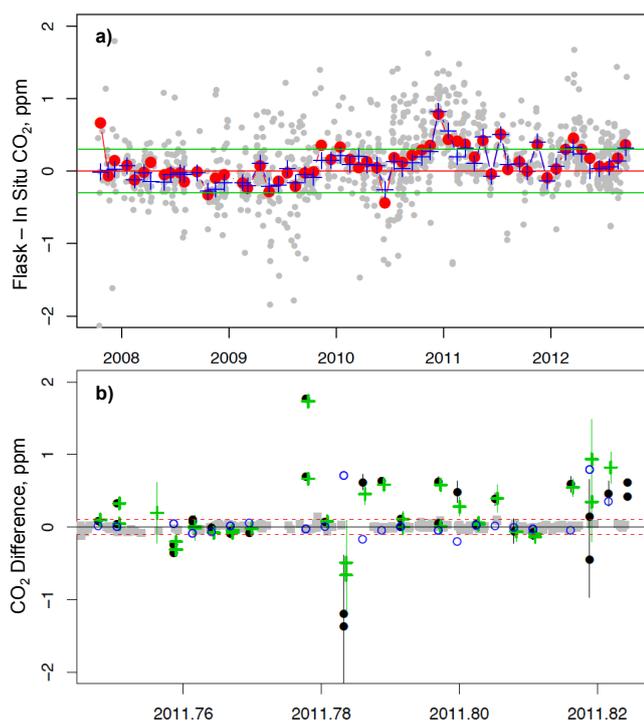


Fig. 10. (a) Comparison of individual (gray filled circles), monthly mean (red filled circles) and monthly median (blue crosses) PFP flask and in situ CO₂ measurements from the BAO tower for samples collected when the standard deviation of the in situ data within a 1.25 h window < 0.5 ppm. True pair samples were collected starting in January 2011. Horizontal lines correspond to ± 0.3 ppm. (b) PFP minus in situ Li-cor (black filled circles), PFP minus in situ Picarro (green crosses), Li-cor minus Picarro corresponding to the PFP sample times (blue open circles), and Li-cor minus Picarro hourly averages for hours with standard deviations < 0.3 ppm (gray squares, $N = 193$).

of the Li-cor and Picarro CO₂ measurements. A laboratory calibration and water correction were applied to the Picarro CO₂ data, but no field calibrations were performed. The Li-cor and Picarro measurements were completely independent (i.e., separate sample air streams and no shared calibration gases or other components). For comparison with the Li-cor, the Picarro data were smoothed using a 30 s running average and the time shifted by -71 s to account for differing flows in the separate intake lines. The median difference between the Li-cor and Picarro measurements was 0.04 ± 0.06 ppm for 9–12 September, as shown in Rella et al. (2013), and that level of agreement was typical of the entire period when the Picarro analyzer was online, during which the atmospheric water vapor mole fraction varied from 0.30 to 1.34 %.

PFP versus in situ agreement for 30 September–28 October 2011 is shown in Fig. 10b for the Li-cor and the Picarro analyzer. During this period, the PFP and the Picarro analyzer shared a common intake line in order to test for sampling artifacts that might result from perturbing the pressures

in the PFP sampling line. We found optimal agreement with the Li-cor 30 s measurements when the PFP time was shifted by -180 s to account for different flow rates in the separate sample inlets. The PFP time was not shifted for comparison with the Picarro analyzer, since they shared a common sampling line. Picarro and Li-cor data within 60 s of the flask-fill end time were averaged for comparison with PFP data. Differences were relatively insensitive to the width of the averaging window applied to the Li-cor data up to at least 2 min. We did not apply a sophisticated weighting function because it was apparent from consideration of the time series that agreement would not substantially improve. The in situ standard deviation within the averaging window was used to filter periods with atmospheric variability. Of 35 comparisons, only 4 had 2 min standard deviations > 0.2 ppm and were excluded from the statistics.

The mean (median) CO₂ difference between the Li-cor and the Picarro values corresponding to the PFP samples was 0.00 (0.00) \pm 0.07 ppm (1σ), the PFP minus Li-cor difference was 0.16 (0.02) \pm 0.4 ppm, and the PFP minus Picarro difference was 0.19 (0.05) \pm 0.4 ppm. Hourly average Li-cor minus Picarro differences are also shown in Fig. 10b for hours where the standard deviations of both in situ analyzers were < 0.3 ppm. For the hourly data, the mean (median) CO₂ difference was 0.00 (0.00) \pm 0.03 ppm ($N = 193$). PFP minus Picarro CH₄ differences are not shown, but exhibit good agreement with a mean (median) difference of -0.84 (-1.37) \pm 0.12 ppb for the same subset of samples. The consistency between the undried, minimally calibrated Picarro and the well-calibrated Li-cor measurements unambiguously shows that PFP versus in situ offsets are attributable to collection, storage, or analysis problems with the PFPs. Differences originating in the PFP intake line that might result from pressure fluctuations would also affect the Picarro data and are negligible during this test. These PFP samples were collected with flush settings corresponding to > 7 volumes of the 300 m intake line.

6.3 Long-term Picarro–Li-cor comparison at WGC

The agreement between the WGC Li-cor and Picarro CO₂ measurements is shown in Fig. 11 for 1–31 July 2011. This was a period where the room temperature was reasonably well controlled, but the level of agreement is representative of the entire 5 yr record. Differences between the analyzers during calibration measurements show no detectable bias and are normally distributed (Fig. 11b) with a standard deviation of 0.04 ppm. For the ambient air comparison, the data were filtered to exclude periods of high variability. Data with 30 s standard deviations > 0.3 ppm were excluded, corresponding to 32 % of the available observations. Since the analyzers share standard gases that span a wide range of CO₂ concentrations, it is not surprising that the bias is negligible. However, the post-processing for the two-analyzers differs significantly in that a time- and/or temperature-dependent baseline

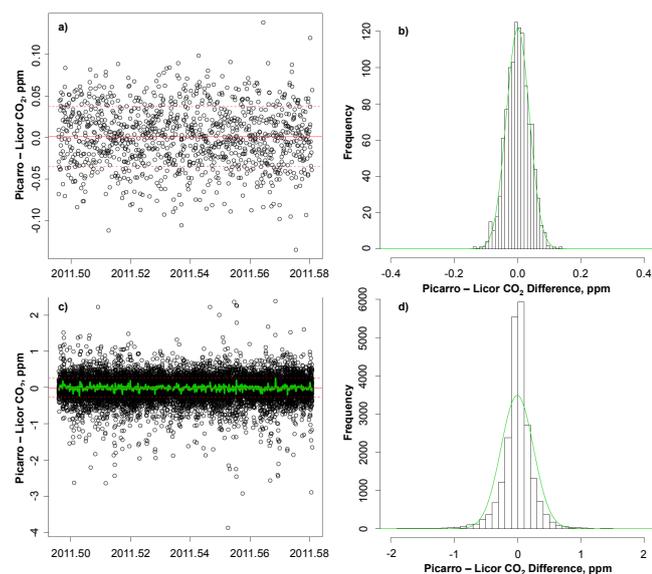


Fig. 11. Comparison of Li-cor and Picarro CO₂ analyzers at WGC. (a) Time series and (b) histogram of measurements of standard gases for the period 1–31 July 2011 with a mean difference of 0.00 ± 0.04 ppm (1σ). (c) Time series and (d) histogram for ambient air samples with 30 s standard deviations < 0.3 ppm. The mean difference is -0.01 ± 0.26 ppm (1σ).

is subtracted from the Li-cor data, and the first-order calibration coefficients are temporally interpolated between 6-hourly calibration cycles, whereas no baseline is subtracted and a 3-day average first-order calibration curve is used for the Picarro.

The lifetime of the reference gases at WGC is shorter than at other sites because of the increased frequency of calibrations to compensate for poor temperature control (Appendix C2) and because the additional gas is used to calibrate the Picarro. The CO₂C3, CO₂C1, and CO₂C4 cylinders are the longest lived and typically last ~ 18 months. Figure 12 shows the uncalibrated Li-cor and Picarro data corresponding to repeated measurements of a single CO₂C3 cylinder over 16.3 months. The standard deviation of the Picarro measurements is 0.05 ppm, whereas the Li-cor signal varies by ~ 20 ppm with discontinuities that correspond to Li-cor reference gas replacements and a power outage. We are able to reliably correct for variations in the Li-cor signal with hourly baseline checks, as evidenced by the reproducibility of our target tank measurements (see, for example, Fig. 9) and by the excellent agreement between the post-processed data from the Picarro and the Li-cor. However, the effort and expense associated with frequent calibrations and gas cylinder replacements is substantial. The short-term precision of the Picarro analyzer (i.e., 30 s standard deviation) is 0.04 ppm, which is consistent with the range of values observed in Fig. 12a. There is a step change in the Picarro signal of ~ 0.1 ppm that corresponds to a period in August 2011

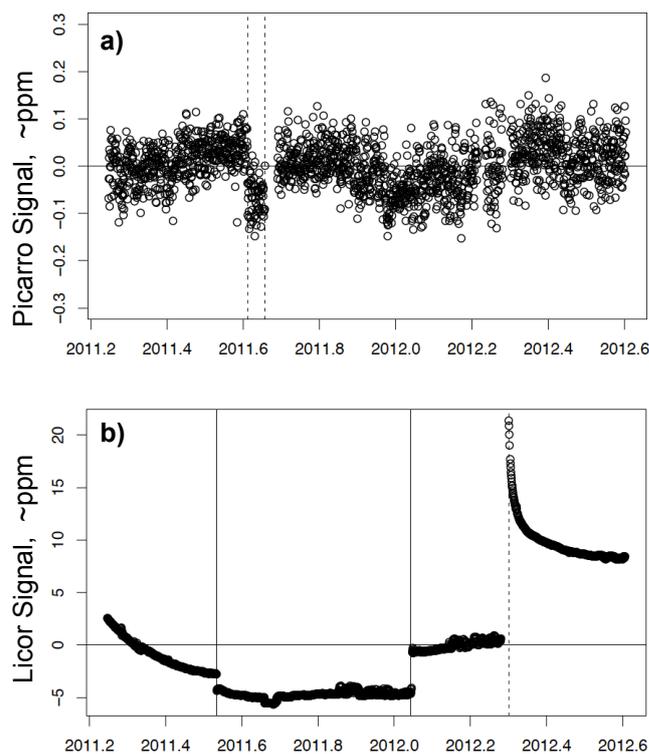


Fig. 12. Uncorrected signal minus the mean value for all CO₂C3 measurements (assigned $\chi_{\text{CO}_2} = 407.77$ ppm) for the period 1 April 2011 to 9 August 2012 for the WGC (a) Li-cor and (b) Picarro analyzers. Dashed vertical lines in (a) correspond to 12 and 29 August 2011, a period when the CO₂C2 reference tank was offline, which caused flow and pressure disruptions in the Picarro sample cell. Solid lines in (b) correspond to dates when the CO₂ reference gas was changed. The dashed vertical line in (b) corresponds to 20 April 2012, when the analyzer was restarted after a power supply failure.

when the CO₂C2 standard was offline, resulting in no flow through the analyzer for the 5 min intervals immediately preceding the CO₂C3 measurements. The values returned to their previous mean when CO₂C2 flow was restored.

Based on the stability of the Picarro response, we could reduce calibration frequency to once per 19 h, which would resolve the diurnal cycle over 5 days. A reasonable calibration strategy would be to use two calibration standards spanning the expected range of ambient values (350–650 ppm for CO₂, 1700–5000 ppb for CH₄), with a single mid-ambient target standard (390 ppm CO₂, 1800 ppb CH₄). Residuals could be evaluated for a regression using all three gases, and mean calibration coefficients could be computed using several days of data. In order to completely flush the current regulators (51-15C-CGA-590, Scott Specialty), it is necessary to use > 1 L of gas per calibration. Since we use 1.25 L for the concurrent Li-7000 calibrations, this has not been a concern, but it would become important if we removed the Li-cor. A Picarro-only approach to minimize gas use would be to run infrequent

long calibrations (e.g., once per 65 h) and to sample the target standard more often (e.g., once per 15 h). The primary driver of calibration frequency for this analyzer is the ability to detect a problem within a day or two of occurrence.

6.4 Additional comparisons

We have had several opportunities to compare our CO₂ measurements with other sensors. Results are summarized in Table 8, some of which have been published previously. The experiments varied in duration, site, season, time averaging, and filtering strategies to remove periods with high variability. Agreement is within the combined uncertainties of the measurements and close to the WMO recommendation for compatibility of independent measurements of 0.1 ppm (WMO, 2012), with the exception of the summertime WBI comparison with the Penn State University cavity ring-down spectrometer during 2009. Details of that comparison are presented in Richardson et al. (2012), but the source of the 0.3 ppm difference is unknown and underscores the difficulty of achieving the WMO goal. Possible contributors are small leaks in the sample tubing or artifacts related to uncalibrated inlet components or errors in the Picarro water correction under conditions of high humidity. The BAO tower is a unique resource, where we have easy access to the sample tubing on the tower for frequent leak checks and the ability to install additional sensors any time. We have shown repeatedly at BAO that comparability of 0.1 ppm can be achieved with well-characterized, independent analysis systems. The only drawback of testing at BAO is that high humidity is rare, whereas sites like LEF, WBI, WKT, and SCT routinely experience humidity levels up to 3 % in summer. The two WKT aircraft spiral comparisons with a well-calibrated analyzer on the NOAA P-3 occurred on days with $\chi_{\text{H}_2\text{O}}$ of $\sim 1\%$, which is relatively low for that area and season. Laboratory tests with wetted ambient air (described earlier in Sect. 5.2) showed that artifacts under controlled conditions are < 0.2 ppm for individual tests and < 0.1 ppm on average. Manually filled flask samples from LEF tower do not show significant seasonal biases that might result from humidity-related errors. However, more work is needed to unequivocally demonstrate < 0.1 comparability in the field under conditions of high humidity.

7 Recommendations

We have learned many lessons over the course of this work and have attempted here to summarize the most critical in the form of recommendations. Many of these recommendations are already documented elsewhere (e.g., WMO, 2011, 2012) or are simply practical, and our experience further underscores their importance.

7.1 Modularity and automation

The modular design of the analytical system has greatly simplified maintenance and repair. Component-level repairs are rarely if ever performed in the field. For eight field systems, we maintain one working system in the laboratory for testing components or proposed design changes, evaluating new gas analyzers, and for other diagnostic testing (e.g., attempting to replicate anomalies or suspected problems under controlled conditions). At least one complete set of spare modules is also needed. We have a few extra pump modules, since they require routine maintenance. The system should be entirely automated with minimal need for human attention and on-site diagnosis. Use of Quick-Connect fittings on reference gases and between modules minimizes or eliminates the need for trained technicians in the field. The control software should have a user-friendly interactive mode to enable remote troubleshooting, e.g., switching valves and power switching for certain components (e.g., pumps, heaters). It is convenient to have a separate system mode for troubleshooting remotely or during site visits so that affected data can be automatically filtered.

7.2 Calibrations

Although modern CO₂, CO, and CH₄ spectrometers are extremely stable compared to the previous generation of analyzers, field calibrations are still needed to establish continuity and comparability within and across networks. Long-term stability of the analytical system is critical, since day-to-day, year-to-year, and site-to-site comparability is the relevant measure of uncertainty for data analysis. We recommend deploying any analyzer with at least one and preferably two calibration gas cylinders beyond the minimum required to generate a calibration curve. For example, an analyzer with a linear response should be deployed with three or four calibration standards, and an analyzer that requires only an offset correction should be deployed with two or three standards. This approach provides meaningful residuals from the calibration polynomial, and one standard can be treated as a target that is not included in the regression. We recommend a minimum of two standards for any analyzer in order to protect against leaks and drifting or erroneous cylinder concentrations. The standards should span the range of expected ambient mole fractions and must be sampled frequently enough to resolve temporal drift in the analyzer baseline or response. Ideally the calibration cycle will have a period not equal to 24 h, so that gaps in the sampling do not always occur at the same time of day. Target standards should be measured so that they are not temporally adjacent to full calibrations in order to maximize sensitivity to unresolved analyzer drift. Some analyzers may still require frequent baseline correction, which can be performed using inexpensive uncalibrated cylinders or, in some cases, a source of zero air. Standards should have the same composition of

interferents as the sample air, and the isotopic composition of the calibrated species should be close to that of ambient air. Whenever possible, multiple standards should not be replaced on the same day so that any problems related to improper installation or altered concentration can be unambiguously attributed to a particular cylinder.

7.3 Drying the sample airstream

There has been much debate about whether sample drying is necessary for CO₂ and CH₄ measurement systems using CRDS or other cavity-enhanced spectroscopic techniques, since those methods potentially enable reliable correction for water vapor interference and dilution. Several groups have described implementations of CRDS systems that do not rely on sample drying (e.g., Winderlich et al., 2010; Chen et al., 2010; Karion et al., 2013; Richardson et al., 2012; Rella et al., 2013), but more work is needed to characterize water corrections at high ambient humidity. The two lines of argument against drying are that it requires additional hardware that increases expense and complexity and that accurate water vapor measurements are intrinsically valuable. Sample drying is a requirement for our system because the water vapor correction intrinsic to the Li-7000 analyzers is not sufficiently accurate or stable to meet our target precision for CO₂. Our experience demonstrates that, if needed, sample drying can be accomplished at a remote site with modest initial expense and minimal need for maintenance. By routing calibration gases through the Nafion dryer, we render negligible any biases associated with CO₂ permeation across the membrane, as demonstrated by small calibration and target gas residuals, and laboratory tests show that loss of CO₂ across the membrane is the same for samples and calibration gases. The upstream chiller and liquid alarm sensors ensure that the gas analyzers and the Nafion dryer are not exposed to liquid water, which can cause swelling of the membrane and flow restriction, or to very high humidity, which may exacerbate cross-membrane transport of CO₂. The humidification of standard gases to the same level as sample gas avoids abrupt transitions between dry standards and potentially humid ambient air that could result in long equilibration times or artifacts. Desiccant is consumed extremely slowly during normal operation, and replacement is needed only after many years (note that one site, WKT, has been operating for > 7 yr and desiccant has not been replaced). The only routine maintenance required is annual replacement of the peristaltic pump module, which involves a single Quick-Connect plumbing connection and a simple electronic connection. The pump module is returned to the laboratory for refurbishment, which simply involves replacing the compressible tubing and a few springs in the roller assembly.

7.4 Sample integrity and redundancy

Reproducibility of target gas measurements is a key measure of long-term analytical stability but is not sufficient to ensure the integrity of the data record. Comparison with totally independent data of comparable quality is the best measure of overall data uncertainty and provides redundancy to protect against gaps in the data record that can cause significant uncertainty in mean data, inferred trends, and estimated fluxes. Care must be taken to ensure that any components upstream of the point where calibration gases enter the sample line do not cause artifacts. For our system, this includes inlet filters, sample tubing, condensers, and pumps. Testing should be done under a wide range of representative conditions and should be performed on aged as well as new components. Routine and preferably automated checks that inlet tubing is intact are necessary and could be simply achieved by installing remotely actuated valves at each intake and pressurizing or evacuating the lines.

7.5 Post-processing

Comprehensive status data for critical pressures, flow rates, and temperatures are necessary for detecting insidious problems such as cross-port leaks in the sample gas manifold. Automated alerts based on these data can provide near-real-time notification of a failure. Prior to the development of automated alerts for our system, problems sometimes went unnoticed for several days or occasionally much longer. Many times a problem can be solved remotely, such as when a pump fails to restart after a power outage. Other failures have been detected simply by monitoring the number and size of data files. The time-dependent uncertainty algorithms that we have described adequately represent the main sources of error. It is inevitable that the analyzers deployed for long-term monitoring will experience periods of suboptimal performance. Our algorithms facilitate quality control and enable automatic filtering of data depending on the error tolerance for a particular application. The algorithms perform well for a variety of sensors with a wide range of performance specifications and characteristics and could be adapted for other modes of operation (e.g., undried CRDS) or for other analyzer types.

7.6 Tower height

It is difficult to justify the expense and complication associated with operating solely on very tall towers. During well-mixed periods, vertical gradients of CO₂ between 100 and 400 m are typically < 0.1 ppm. At night, levels higher than 200 m are frequently decoupled from the surface, and vertical gradients frequently exceed 10 ppm. Under these conditions, the highest levels often sample remnants of the previous afternoon's boundary layer from some distance upwind. Night-time tower data are especially difficult to model because of

the steep vertical gradients near the surface and wind shear associated with nocturnal jets. Tower lease, installation, and maintenance costs are largely driven by height, and shorter towers are more abundant than very tall towers. Data from ~ 100 m above ground level would likely suffice for most carbon-budgeting applications with current models. Many studies rely primarily on afternoon data, and model residuals are generally much larger than 0.1 ppm. However, tall tower observations are extremely useful for evaluating the fidelity of boundary layer processes in models, especially when a full complement of meteorological measurements and additional trace-gas data are available. An effective strategy for carbon monitoring would be to maintain a small number of tall tower “super-sites” representative of a variety of environmental conditions, and a larger network of shorter tower installations with a simpler instrument suite.

7.7 Complementary measurements

Whenever possible, tower greenhouse gas measurements should be colocated with other observations that are useful for evaluating atmospheric transport models and that provide additional constraints on flux estimates. Other trace gases and isotope measurements can aid in source attribution. Measurements of meteorological parameters such as wind speed and direction, as well as temperature and humidity, should be included at two or more heights on the tower to enable gradient-method flux estimation. Meteorological measurements should be made using high-quality and routinely calibrated sensors, and radiation shields for temperature and humidity probes should be mechanically aspirated and include flow or Hall effect sensors to verify adequate ventilation (French and May, 2004). The North American Carbon Program Plan (Wofsy and Harriss, 2002) recommends biweekly aircraft profile measurements of greenhouse gases and tracers over surface monitoring sites. Commercially available remote sensors such as microwave temperature profilers, pulsed Doppler light detection and ranging (lidar) wind profilers, and laser ceilometers can provide detailed information about atmospheric structure and/or estimates of mixed layer height that are useful for evaluating model boundary layer parameterizations, especially when combined with other data that describe the surface energy budget, e.g., radiation and eddy covariance measurements. Solar occultation measurements from ground-based spectrometers such as those used in the TCCON network (Wunch et al., 2011) along with colocated tower measurements and boundary layer height data would place strong constraints on estimates of surface fluxes. Tall tower eddy covariance measurements of CO₂ and H₂O fluxes (Berger et al., 2001) can potentially help to separate near- and far-field contributions to observed CO₂.

8 Conclusions

In situ measurement and communications technologies have improved dramatically over the last decade. For the first time, research-grade operational monitoring is feasible for CO₂, CH₄, and a growing suite of other important trace gases, but measurement requirements for future greenhouse gas monitoring efforts need to be carefully defined. Data records with high precision and long-term stability are needed to resolve annual mean gradients and trends of CO₂ and other greenhouse gases. Many processes that drive net biological and oceanic fluxes operate on timescales of decades to centuries, so multidecade records are needed to diagnose the underlying mechanisms. If atmospheric data are to be used for emissions verification, or to inform policy more generally, then the data must be fully disclosed and documented with minimal delay. For both research and regulatory applications, the uncertainties must be well understood and thoroughly documented.

We have designed a robust system for quasi-continuous measurements of χ_{CO_2} , χ_{CO} , and χ_{CH_4} at unattended tall tower monitoring stations. Eight systems have been deployed, all of which have been operational for > 5 yr. The system reports extensive engineering data so that most problems can be diagnosed remotely. The modular design facilitates maintenance and repairs. Faulty modules can be quickly replaced and returned to the laboratory for component-level repairs. Certain recurring or otherwise notable failure modes are documented here, and we have taken steps to reduce or prevent future occurrences. We have developed algorithms for computing calibrated χ_{CO_2} , χ_{CO} , and χ_{CH_4} and for estimating statistically rigorous time-dependent uncertainties. The algorithms are flexible and return credible uncertainties from three gas analyzers with diverse noise characteristics. We report detailed uncertainty information in our data files, including total measurement uncertainty, random measurement uncertainty, atmospheric variability, and calibration scale uncertainty. The analyzers have been thoroughly evaluated in the laboratory and compared with independent data from our own and other laboratories. Lab tests and comparisons with independent data show that we are meeting the WMO recommended target of 0.1 ppm comparability for CO₂ under conditions of low to moderate humidity. More evaluation under high-humidity conditions is needed, but lab tests and limited comparison data suggest that 0.2 ppm is a conservative upper limit for errors for $\chi_{\text{H}_2\text{O}} \leq 3.5\%$. We have identified an apparent bias affecting CO₂ measurements from our automated flask samplers, and we are continuing to characterize the bias and evaluate strategies for mitigating the impact.

Flask versus in situ comparisons for CO and CH₄ do not exhibit biases and show that on monthly to annual timescales we are achieving long-term comparability for these gases that is in accordance with the WMO recommended targets of 2 ppb for both gases. For CO, our most significant problem has been drift in calibration gas standards. The CO data

have estimated uncertainty of order 10 ppb on timescales of minutes, but this is mainly random, and hourly average values generally have uncertainties < 2 ppb. For CH₄, we meet the WMO recommendations on short and long timescales. The only noteworthy complication is that ambient values frequently exceed the current upper limit of the WMO calibration scale for CH₄, and that will soon be resolved since work is underway to extend the scale to 5700 ppb.

Several of the measurement comparisons described here meet the WMO recommended goal for compatibility of 0.1 ppm, but others fall short. Agreement better than 0.3 ppm is relatively easy to achieve but is insufficient for emissions verification on continental to global scales. Other research groups also have demonstrated robust detector calibration strategies that account for analyzer drifts and deliver records with long-term stability of calibration standard residuals and target measurements. Remaining challenges relate to sample integrity: are the sampling lines intact? Is the sample being modified en route to the detector? Is the sampling strategy adequate for capturing mean values over relevant timescales in the presence of typical variability? We have outlined tractable solutions to address these issues and have shown that a network of high-quality sensors can be efficiently maintained. The analytical system and post-processing methods described here provide one model to inform future expanded monitoring efforts. The time-dependent uncertainty algorithms are flexible and readily adaptable to other species and analytical systems.

Appendix A

Additional system components

A1 Power

DC power for the instrument components is provided by a power supply with 12 V (75 W), ± 15 V (75 W each), and two 24 V (200 W each) output modules (Mini-Megapak MM5-15699; Vicor, USA). This power supply was selected for its compact size, robustness, and low noise (ripple). The pumps and some of the temperature control equipment are powered through relays (SDM-CD16AC; Campbell Scientific, USA) so that they can be shut down remotely or automatically restarted if necessary. An uninterruptible power supply (UPS) protects against short-duration power outages and power surges (9130, 1.5 KVA rackmount; Eaton, USA).

Table A1. Signal list.

Signal
Timestamp
Analyzer signals (CO ₂ , CO, CH ₄)
Water content of sample flow through each analyzer
Sample flow through each analyzer
Analyzer pressures
Li-cor CO ₂ analyzer reference flow
Gas cylinder pressures
Bypass flow from each sampling height
Bypass back pressure for each sampling height
Analyzer enclosure temperature
Analyzer internal temperature
Room temperature
Pump box temperature
Chiller element temperature
Nafion box purge flow
Combined analyzer exhaust pressure
Manifold/valve position (SYS.MODE)
Liquid alarm status for each sampling height

A2 Data acquisition and control

A datalogger (CR-10X-ST-MA-NC; Campbell Scientific, USA) with accessories is used for all data acquisition and control functions. All engineering and trace-gas data are recorded every 30 s. We wanted a simple, commercially available, reliable operating system, as well as the ability to take advantage of evolving technology for communications and data storage. In addition to the datalogger, other Campbell Scientific components include two multiplexer boards (AM16/32A-ST-SW), relay modules (SDMCD16AC), an analog output module (SDM-AO4-SW), and a serial communications module (SDM-SIO4). Custom-printed circuit boards simplify connections to the datalogger's wiring panel. The datalogger memory can store approximately two days' worth of data, which provides some protection against communication interruptions or PC failures. The CR-10X datalogger has been discontinued, and we are transitioning to the replacement CR-1000, which has improved serial communications and larger data storage capacity.

Most of the engineering data are differential analog signals, but serial communications are used to retrieve data from the Li-7000 CO₂ analyzer and from the Thermo Electron 48C TL CO analyzer. Serial communications with the datalogger are inefficient and limit the speed at which we are able to interrogate the sensors. The datalogger program runs on a 5 s interval to allow adequate time for serial polling and response. To compensate for the low sampling frequency, we rely on the built-in averaging capabilities of the CO₂ and CO analyzers. The Li-7000 CO₂ analyzer reports a 5 s average. The 48C TL CO analyzer, which is noisier, is set to report a 30 s average, and thus the 5 s samples recorded by

the datalogger are not independent. The 5 s measurements are then aggregated to 30 s averages and stored in the datalogger's memory along with the corresponding standard deviations.

An onsite PC laptop is used for data storage and remote access by cellular modem or digital subscriber line (DSL). The PC runs a Windows operating system, software to communicate with the datalogger (Loggernet; Campbell Scientific, USA), and remote administrator software (Radmin; Famatech, Russia). The data are downloaded to the PC every minute and a program (Baler; Campbell Scientific, USA) running on the PC bins the data into hourly average files. The PC time is synchronized to a time server every 15 min using commercially available software (Dimension 4, Thinking Man Software), which also logs differences due to PC clock drift. The PC time is uploaded to the datalogger daily. PC clock drifts are of the order of seconds per day, and become significant if uncorrected over periods of weeks or more.

A3 Temperature control

Both the CO₂ and CO analyzers are carefully temperature-controlled to a setpoint that is chosen to be 10–15 °C above typical maximum room temperature for each site. The Li-7000 CO₂ analyzer is specified to operate at temperatures up to 50 °C (although we have observed that serial communications may be unreliable above 45 °C), and the 48C TL CO analyzer has a specified operating temperature up to 45 °C. The CO₂ analyzer is housed in a rack-mounted aluminum chassis box (48.3 cm × 17.8 cm × 55.9 cm) along with its pressure and flow controllers. The CO analyzer is rack-mountable, so no separate enclosure is required. A small temperature controller unit (CT325PD2C1; Minco, USA) is mounted inside each enclosure that drives six Kapton[®] (registered trade name of E.I. DuPont and Nemours) tape heaters (HK5340R58.9L36B; Minco, USA), which are distributed evenly over the interior surface of the boxes, including the lid. The control temperature is measured with a four-wire platinum RTD (S665PDZ40AC; Minco, USA), and the sensor element is suspended in the air near the center of the enclosure. The temperature controllers are inexpensive and easy to use. However, we had several unexplained failures where the Minco temperature controller unit overheated and melted. Reliability improved when used with a solid-state relay driver (e.g., MPDCD-3; Crydom, USA), but with some degradation of temperature stability. Each temperature-controlled box is wrapped with a single layer of Aramid fabric insulation (MC8-4596B 48"; Tex Tech, USA). A small fan mounted inside each enclosure provides air circulation. The CO₂ enclosure is mounted above the CO analyzer in a standard instrument rack, with a gap of approximately 1 cm between the boxes. A scroll fan is used to circulate air between the boxes to prevent overheating. The variability in the CO₂ assembly is typically < 0.2 °C (1 σ), and the CO₂ analyzer temperature is typically stable to ~0.05 °C.

Room temperature at some sites exhibits strong seasonality and is outside of our control at sites where the equipment is located in the tower's transmitter building. There is no single setpoint for the temperature controllers that will work at all sites under all conditions. Unfortunately, the setpoint potentiometers for the temperature controllers are located inside the CO₂ and CO analyzer assemblies and are difficult to access. Ideally, we would be able to adjust setpoint temperatures remotely, or at least install an external adjustment dial.

Appendix B

Reliability

B1 Automated alerts

An important feature of the post-processing software is that it provides daily summaries of errors and anomalies that are emailed to lab personnel. Alerts are generated if fewer than expected data files are transferred, if the file sizes are smaller than normal, or if signals are outside of the expected range. One data record per hour containing all instrument signals is uploaded from the site computer to a server so that certain signals can be monitored on an hourly basis. Errors such as pump failures, power outages, and losses of communication are typically detected within one or two hours. Approximately 50 plots are created nightly for each site that display measured χ_{CO_2} , χ_{CO} , χ_{CH_4} , and detailed uncertainty information, along with important engineering signals and other diagnostics. The plots are accessible via Internet browser. Plots for all sites are reviewed at least twice per week and whenever an automated alert is generated.

Certain failures result in automatic flagging of the data. For example, fatal flags are assigned when flow through one or more of the analyzers is lost. Loss of flow may occur for all levels if there is a systematic problem or for a single intake when a pump fails or a liquid alarm sensor is triggered. New automatic flagging algorithms and alerts are developed whenever a new failure mode is discovered or for cases where manual flagging would be overly tedious. Automated flagging reduces the likelihood of human error associated with data entry; however some manual flagging is unavoidable, for example, if work is being done while the system is running.

B2 Notable or recurrent problems

B2.1 Cross-port leaks and relay failures

Within a month after deployment at WGC in fall 2007, unusual patterns appeared in the sample line bypass flow signals. Investigation revealed that air was leaking across the ports of PTFE solenoid valves (Galtek, 203-3414-215) originally used in the sampling manifold so that air reaching

the analyzers was a mixture from different intake heights. Mounting screws securing the valves to the floor of the enclosure had been over-tightened, distorting the valve base. We subsequently replaced the sample valves in all of our systems with steel solenoid valves identical to those used in the calibration manifold.

The original intent of the bypass flow sensors was to monitor pump performance, but after WGC sample solenoid cross-port leakage was detected, we implemented a “flow accounting” algorithm that has detected subsequent valve-switching failures. Recurring problems with valve switching have affected at least four sites and in all cases worsened over time. At AMT sample solenoids have intermittently failed to switch, and at AMT, SNP, and WGC, a similar problem has affected the CO zeroing solenoids. At WKT, two of the CO₂ calibration solenoids intermittently failed. Evidence suggests problems internal to the Campbell Scientific relay module used to drive the valves (SDM-CD16) or perhaps faulty electrical connections elsewhere rather than defective valves. Calibration and CO zeroing valve failures were easy to detect based on calibration residuals, since assigned standard values did not correspond to the air that was being sampled. However, flow accounting based on analyzer and sample-line bypass flows is needed to detect valve-switching failures in the sampling manifold and to flag the affected data.

B2.2 Sampling line leaks

Contamination resulting from leaks in the sampling lines may be difficult to detect. Leaks within the building rarely develop spontaneously, but we have occasionally lost data because of failure to properly tighten one or more connections during an installation or maintenance/repair visit. When possible, we test for leaking fittings by placing a few pounds of dry ice near the system for several hours or overnight while monitoring the measured CO₂ signal. Care must be taken to avoid exposing personnel to dangerous levels of CO₂. It is useful to have an inexpensive handheld CO₂ monitor when performing these tests to ensure that ambient CO₂ levels are safe (< 5000 ppm for 8 h time-weighted average exposure, US Occupational Safety and Health Administration Permissible Exposure Limit). We have also developed a leak-checking apparatus consisting of a hand pump with an electronic pressure gauge that can be used to check whether a section of plumbing holds a vacuum. After pumping down the line, a valve between the pump and the gauge is closed, and the pressure is tracked for several minutes or longer.

Tubing on the tower can be damaged by falling ice, high winds, or fatigue at the points where it is secured to the tower. Once a leak has developed, rainwater can infiltrate the tubing and freeze–thaw cycles may cause additional damage. The first indication of a leak is often a liquid alarm signal after heavy rain. Other times, severe tubing damage was visible from the ground. The start date of a leak on the tower is often difficult or impossible to determine, and unless the tube

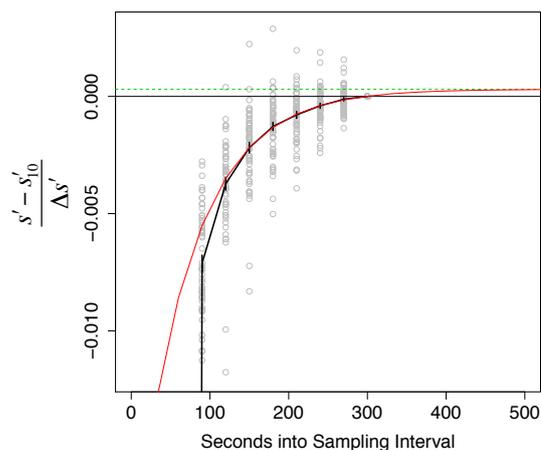


Fig. C1. Approach to equilibration for CO₂ calibration standards for WBI 5–7 August 2009 normalized by the signal difference $\Delta s'$ from the final value for the previous sampling interval. Values represent the signal difference from the INTERVAL = 10 (300 s) value per unit difference from previous sampling interval (the analyzer gain is ~ 1 , since the Li-7000 output is estimated χ_{CO_2} with units of ppm). Gray circles are for individual calibrations, and the heavy solid line corresponds to the mean response computed over all the calibration and target modes. The red curve is a fit to the final three minutes of the mean response given by $y = 0.00030 - 0.020815 \times \exp(-(x - 150)/70.55)$. For $\Delta \chi_{\text{CO}_2} \sim 100$ ppm between successive sampling modes, the equilibration correction would be +0.03 ppm at INTERVAL = 10.

was severed, the sampled air would have been a mixture from two or more heights. The impact on the data depends on the vertical gradient of the gas being measured, so data collected during well-mixed periods may be minimally affected. We have successfully worked with tower climbers to repair damaged tubing, and vacuum leak checks are performed whenever we have climbers on a tower. One effective method for finding leaks is to pressurize the line with a pump so that climbers can hear the air hissing out. We are developing an automated system to enable routine leak checking, where a large-orifice remotely actuated and normally open valve is installed on the tower at the sample inlet, and the valve is periodically closed to check the vacuum created by the sample pump. Modern radio modems provide extremely reliable communication with equipment mounted high on the towers, and AC power is generally available on the towers.

Appendix C

Additional data processing details

C1 Disequilibrium correction

Calibration data must be combined and averaged in order to precisely derive the disequilibrium correction. The amplitude of the correction is proportional to the χ_{CO_2} difference

Table C1. Terms used in the disequilibrium correction for CO₂.

SYSCODE	INTERVAL	s'	$s' - s'_{t=10}$	$\Delta s'$	Time, UTC	$t - t_0$
L2	10	-27.18167	0	NA	11:49:55	300
C1	7	-31.82166	0.00667	-4.64666	11:53:25	210
C1	8	-31.81333	0.015	-4.64666	11:53:55	240
C1	9	-31.80833	0.02	-4.64666	11:54:25	270
C1	10	-31.82833	0	-4.64666	11:54:55	300
C2	7	-5.206667	-0.016667	26.63833	11:58:25	210
C2	8	-5.211666	-0.021666	26.63833	11:58:55	240
C2	9	-5.208333	-0.018333	26.63833	11:59:25	270
C2	10	-5.19	0	26.63833	11:59:55	300
C3	7	29.27167	-0.04166	34.50333	12:03:25	210
C3	8	29.29667	-0.01666	34.50333	12:03:55	240
C3	9	29.31	-0.00333	34.50333	12:04:25	270
C3	10	29.31333	0	34.50333	12:04:55	300
C4	7	79.29833	-0.03001	50.01501	12:08:25	210
C4	8	79.31333	-0.01501	50.01501	12:08:55	240
C4	9	79.32666	-0.00168	50.01501	12:09:25	270
C4	10	79.32834	0	50.01501	12:09:55	300

* Data records with INTERVAL < 7 have been omitted for brevity.

** Data from WBI, 5–8 August 2009.

between consecutive sampling intervals. We therefore normalize the drift-corrected analyzer signal s' from individual calibrations by first computing $s' - s'_{10}$, where s'_{10} is the final value (INTERVAL = 10) from the 5 min sampling interval, and then dividing by $\Delta s'$, the difference from the previous sampling mode's final value. An example is shown in Fig. C1, and a subset of corresponding data are given in Table C1. An exponential function of the form

$$\frac{s' - s'_{10}}{\Delta s'} = \alpha_1 + \alpha_2 e^{\frac{t-t_0}{\tau_{\text{eq}}}} \quad (\text{C1})$$

is fitted to calibration data with INTERVAL ≥ 5 and $\Delta s' > 5$ ppm where

$$\alpha_1 = \frac{s'_{\text{eq}} - s'_{10}}{\Delta s'}. \quad (\text{C2})$$

Solving for s'_{eq} gives

$$s'_{\text{eq}} = s' - \Delta s' \alpha_2 e^{\frac{t-t_0}{\tau_{\text{eq}}}}. \quad (\text{C3})$$

C2 Temperature-dependent analyzer baseline

We have had persistent problems maintaining the temperature control at WGC due to wide extremes in room temperature, which cannot be accommodated with a seasonally invariant temperature setpoint. At that site, the equipment is housed in the antenna's transmitter building and we do not have direct control of the room temperature. Rather than repeatedly adjust the setpoint temperature, we let the analyzer

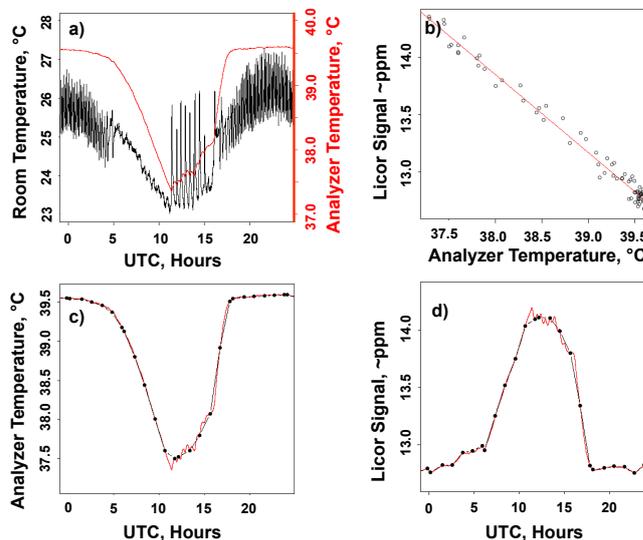


Fig. C2. Poor temperature control performance for the CO₂ analyzer at WGC on 7 August 2010. (a) Room temperature (black curve, left axis) and Li-cor cell temperature (red curve, right axis). (b) Li-cor CO₂C₂ (baseline) signal versus analyzer temperature. (c) Li-cor cell temperature (red curve) and the same temperature signal sampled corresponding to CO₂C₂ measurements (black filled circles) and interpolated to all times (black connecting lines). (d) CO₂C₂ measurements (black filled circles) interpolated linearly in time (black connecting lines) and estimated for all times using the slope from (b) multiplied by the difference between the black and red lines in (c).

temperatures float (Fig. C2a). The insulation causes the analyzer temperatures to vary slowly enough that we can effectively correct for instrument drift using frequent baseline measurements, along with an empirically determined relationship between the internal analyzer temperature and analyzer baseline described below. For CO₂, our implementation is a relatively expensive solution, in that it requires frequent use of standard gases. The calibration frequency at WGC is approximately twice that now used at other sites. Fortunately, the WGC Picarro CO₂/CH₄ CRDS (described in Sect. 2.8) is insensitive even to large room temperature variations, and we plan to rely primarily on that sensor going forward so that we can reduce gas use.

For cases where a significant correlation exists between analyzer temperature and the baseline signal, we have the option to enable a temperature-dependent baseline algorithm. The slope from the baseline–temperature relationship (Fig. C2b) is applied to the difference between the measured analyzer temperature (red curve in Fig. C2c) and the analyzer temperature extracted at t_b and interpolated to all times t_i (black symbols and connecting lines in Fig. C2c). The resulting temperature-dependent baseline correction is added to the usual time-interpolated baseline (black symbols and connecting lines in Fig. C2d) to generate a continuous representation of the analyzer baseline (red curve in Fig. C2d).

Appendix D

Statistical basis for the uncertainty framework

Consider the case of an analyzer with a linear response such that

$$y = mx + b, \quad (D1)$$

where the dependent variable y corresponds to the analyzer signal and the independent variable x represents the assigned values of the calibration standards. For cases where the coefficient of determination $R^2 \approx 1$, then the choice of dependent versus independent variable is not critical, and conversions between x and y units are accomplished via the regression slope, m . Calibration regressions for CO₂, CO, and CH₄ typically have $R^2 > 0.99$. Table D1 contains a list of symbols used below and in Sect. 4.

Following Skoog and Leary (1992), for the case where there is no error in the x values, the standard deviation of the slope m is given by

$$\sigma_m = \frac{\sigma_y}{\sqrt{\sum (x_i - \bar{x})^2}}, \quad (D2)$$

where \bar{x} is the mean assigned value of the calibration standards and σ_y is the standard deviation of the residuals:

$$\sigma_y = \sqrt{\frac{\sum (y - y_{\text{fit}})^2}{N - 2}}. \quad (D3)$$

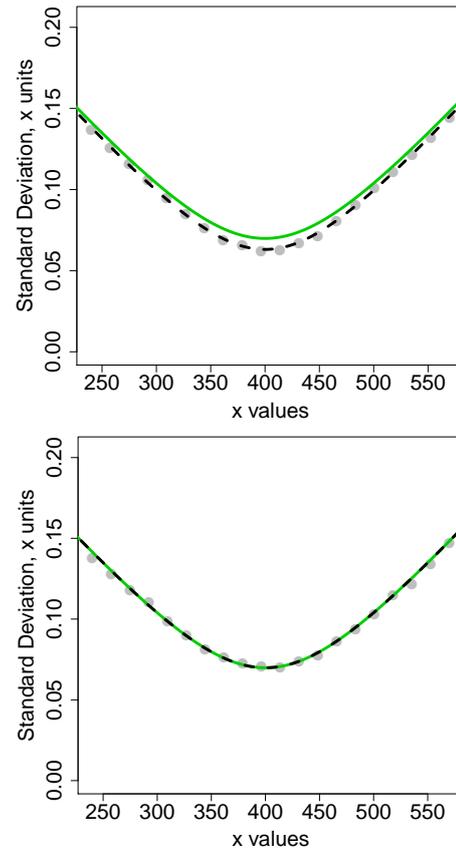


Fig. D1. Regression uncertainty for a typical CO₂ calibration. (a) Results for case A with $\sigma_y = 0.05$, $\sigma_x = 0.03$, and $y = 0.8x - 370$. (b) Results for case B where $\sigma_x = 0$. The green curves correspond to SD_u described by Eqs. (D6) and (D7). The dashed curves correspond to SD'_u where σ'_y replaces σ_y in Eq. (D7). The filled circles represent the standard deviations across 2500 realizations for the simulated unknown samples from the Monte Carlo analysis.

The standard deviation of the intercept b is given by

$$\sigma_b = \sigma_y \sqrt{\frac{\sum x_i^2}{N \sum x_i^2 - (\sum x_i)^2}}, \quad (D4)$$

where N is the number of calibration standards. Note that the value of σ_b changes when x is shifted by a constant value such that the minimum value $\sigma_{b\text{min}}$ occurs when x_i is replaced by $x_i - \bar{x}$ in Eq. (D4). The standard error of the fit for any x can be computed by propagating the error in the coefficients:

$$s_{\text{fit}} = \sqrt{(\sigma_m(x - \bar{x}))^2 + \sigma_{b\text{min}}^2}. \quad (D5)$$

Skoog and Leary (1992) give the following equation for the standard deviation SD_u for analytical results obtained with the calibration curve:

$$SD_u = \frac{\sigma_y}{m} \sqrt{\frac{1}{L} + \frac{1}{N} + \frac{(\bar{y}_u - \bar{y})^2}{m^2 \sum (x_i - \bar{x})^2}}, \quad (D6)$$

Table D1. Glossary of uncertainty terms.

Symbol	Description	Units
y	Analyzer response	Analyzer-specific “raw” units
x	Mole fraction	ppm or ppb
m	Slope of calibration curve	(units of y)/(units of x)
se_{fit}	Standard error of the fit	Same as y
σ_y	Standard error of the fit residuals	Same as y
σ_u	Uncertainty of the unknown samples (i.e., of the atmospheric and target measurements treated as unknowns)	Same as x
σ_{sc}	Uncertainty of the assigned values of the standards (also known as σ_x)	Same as x
σ'_y	Portion of σ_y not attributable to uncertainty in x (not directly observed)	Same as y
u_M	Total measurement uncertainty	Same as x
u_{TGT}	67th percentile of the absolute difference between measured and assigned target values	Same as x
SD_M	Standard deviation of the mean measured value	Same as x
u_R	Random component of measurement uncertainty	Same as x
AV	Atmospheric variability = $(SE_M^2 - u_R^2)$	Same as x
SD_u	Standard deviation of analytical results obtained with the calibration curve	Same as x

Throughout the text uncertainty terms are represented as vectors whenever the quantity is inherently time-varying or has been interpolated to all times t_i in the 3-day processing window.

where \bar{y} is the mean of the analyzer signals for the calibration standards, N is the number of calibration standards, \bar{y}_u is the mean of L replicate analyses of an unknown sample, and SD_u is the corresponding standard deviation. Although it is not obvious, it can be shown numerically that for the case where $L = 1$ (i.e., no replicate samples), this is equivalent to

$$SD_u = \sqrt{\left(\frac{se_{\text{fit}}}{m}\right)^2 + \left(\frac{\sigma_y}{m}\right)^2}. \quad (\text{D7})$$

Notice that u_M defined by Eq. (9b) is equivalent to zSD'_u where σ'_y replaces σ_y in Eq. (D7) and z is a factor taken from the Student's t distribution depending on the degrees of freedom of the regression and the desired level of confidence. Thus our uncertainty framework is a generalization of the textbook treatment presented in Skoog and Leary (1992), such that errors in x are taken into account and sample errors σ_u are allowed to differ from σ'_y/m . We attempt to model sample errors σ_u as described in Sect. 4. It is straightforward to extend this framework to include analyzers with nonlinear response or for situations when orthogonal distance regression is preferred (e.g., when R^2 is significantly < 1).

We performed a Monte Carlo analysis to ensure that x errors are treated properly in our framework. We generated 2500 realizations of the calibration curve such that x values were perturbed by an amount selected randomly from a normal distribution with zero mean and a standard deviation of σ_{sc} . The y values were similarly perturbed using a distribution with standard deviation, σ'_y . Calibration curves

were fit to each realization and x values were computed for a set of twenty “unknown” samples, y_u . We considered two cases: case A is similar to a typical Li-7000 field calibration, where $\sigma_y = 0.05$, $\sigma_x = 0.03$ ppm, and $m = 0.8$ (generally $m \approx 1$, but we used 0.8 to more clearly show dependence on m); case B is identical except $\sigma_x = 0$. Results are shown in Fig. D1. The R script used to create Fig. D1 is included with the Supplement.

Appendix E

Additional flask-sampling details

E1 Programmable flask packages

Each PFP unit contains 12 individual 0.7 L borosilicate glass flasks, each with a valve on both ends. The valve manifold is stainless steel, and valves are glass with Teflon O-rings. Pumps for the PFP sampling are housed in a separate programmable compressor package (PCP). The PFP/PCP system was originally designed to operate on an airplane over a wide range of altitudes. A detailed description of the PFP and associated components will appear in a separate publication. A schematic diagram of the tower PFP sampling system is shown in Fig. E1, and a photograph showing the interior of a PFP is included in the Supplement. The protocol for preparing PFPs for field sampling has been to flush residual sample air from the PFP with dried and filtered air from

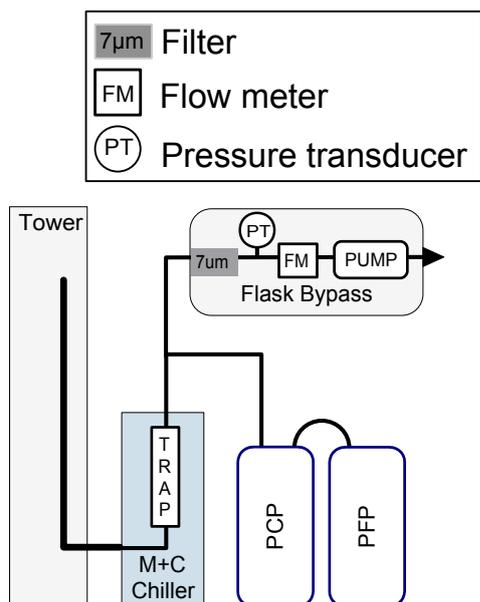


Fig. E1. Diagram of the PFP flask-sampling system.

the building's compressed air system. Flasks are then flushed and pressurized to 2200 hPa (~ 20 psig in Boulder, CO) with cylinder air from which CH₄ has been scrubbed (in order to serve as a tracer for insufficiently flushed samples).

To date, we have stayed close to the original PFP/PCP design in order to maximize consistency of data and logistics with our laboratory's aircraft-sampling program (e.g., Karion et al., 2013), but we hope to eventually modify the flask-sampling apparatus for optimal performance at tower sites. Tower-specific modifications that have already been implemented include (1) a separate pump assembly to continuously flush the long sample tubes at ~ 4 slm, (2) a pressure sensor and flow meter on the sampling line, (3) an optional A/C power supply to replace the batteries and trickle charger that normally provide power for aircraft sampling, and (4) a datalogger and cellular modem to trigger samples and record line pressure and flow through the flush pump. The PFP sample airstream is routed through the spare channel of the M&C four-channel chiller for the in situ system. The chiller temperature is 1.6 °C, and because of the high flow rates (> 10 slm) during sampling, the pressure is subambient (~ 600 hPa) and the residence time in the condenser is very short. The current drying configuration is inadequate to prevent condensation in the pressurized samples, but the cost of a more capable drying system has been prohibitive so far.

Generally, flasks are sampled in pairs at approximately 14:00 LST. Paired sampling enables radiocarbon analysis in a subset of the samples, and pair agreement provides a measure of repeatability of other gases. The PFP units were not originally designed for parallel sampling, so until recently paired flasks were filled sequentially, with a typical time difference of 3 to 5 min. True-paired sampling began at BAO in January 2011 and throughout the network in January 2012.

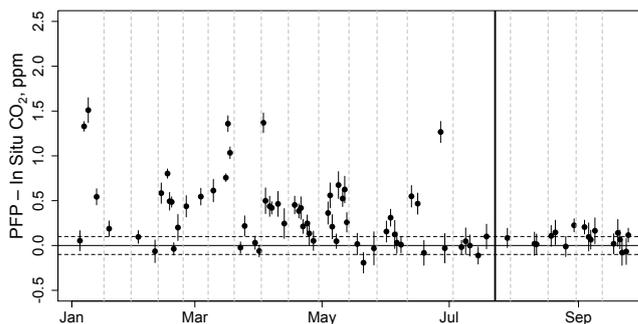


Fig. E2. PFP flask minus in situ differences at BAO for 2013 corresponding to periods when the in situ standard deviation over 5 min was < 0.15 ppm. The solid black vertical line corresponds to when the new conditioning protocol for PFP sampling was implemented. Dashed gray lines correspond to when PFP units were switched. Error bars show the 5 min standard deviation of the in situ values.

When a flask sample is triggered and the PCP pumps are enabled, the flow rate in the tube increases from the standby rate of about 2–4 slm to 12–16 slm, the combined flow from the PCP and flush pump. The increased flow causes a length-dependent pressure drop in the tube (140–250 hPa) relative to the standby pressure. Pressure fluctuations perturb the equilibrium between sample air and the walls of the tubing, and flush times of 10 min or more are needed to adequately flush the longest sample lines after a new equilibrium is reached. In the current configuration, the inlet tube and PFP manifold are flushed with 80 ± 5 L of air, and then flasks are opened and flushed with another 70 L. A 500 m length of 1.17 cm (0.5 in.) OD Synflex 1300 tubing has a volume of 35 L. The total time for flushing and filling is ~ 15 min, of which the fill time is less than one minute.

The PFP flasks are sealed with Teflon O-rings. Tests have shown that CO₂ preferentially diffuses across the Teflon O-rings compared to O₂ and N₂, so a storage correction of 0.007 ppm per day is applied to the data. Some tower PFP samples have been stored for 20 days or more between collection and analysis, so corrections can be 0.15 ppm or more.

E2 Apparent bias affecting PFP samples

Through testing at BAO and regular deployment in the network, we have identified several PFP units with multiple individual flasks repeatedly showing CO₂ that is 0.5–3 ppm higher than corresponding in situ data. Spurious CO₂ enhancements have also been measured in the laboratory for two of these units. The laboratory measurements of the enhancements under carefully controlled conditions are of similar magnitude to the apparent biases seen at tower sites. Enhancements have been observed in the laboratory for sample humidity as low as 0.075 % (~ -23 °C dew point at 1013 hPa), but our routine procedure for testing of PFPs using very dry air from a cylinder shows smaller biases and with opposite sign.

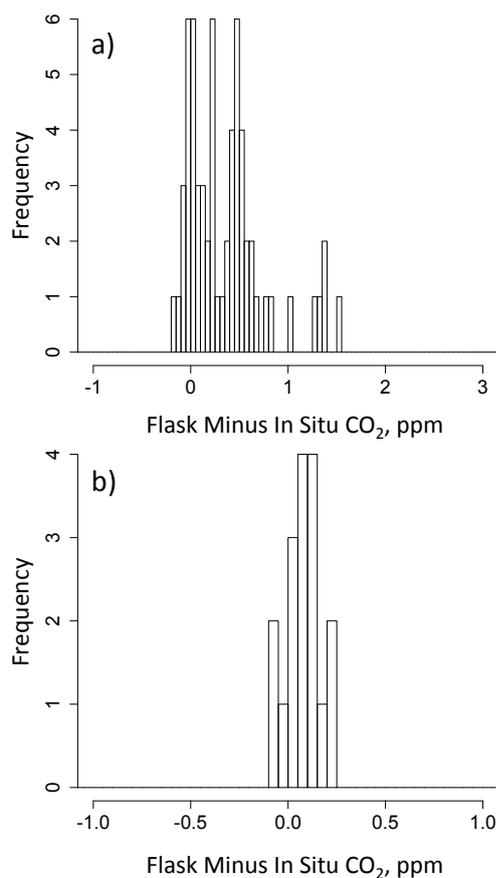


Fig. E3. Histograms of BAO PFP flask minus in situ differences for conditions with atmospheric variability < 0.15 ppm for (a) 1 January–19 July 2013 with PFP samples collected using the standard sampling protocol (mean = 0.36 ± 0.40 SD, $N = 62$), and (b) for 20 July–30 September 2013, during which the PFP samples were collected using the modified conditioning protocol (mean = 0.08 ± 0.09 SD, $N = 17$).

Tests show that if a bias-prone PFP flask is not prepared with dry air before refilling (i.e., if the new sample is taken without flushing the residual air from the previous sample), then the bias is eliminated in the second sample. A possible explanation is that CO₂ is adsorbed onto contaminated surfaces when the flasks are pressurized with dry air, and that water from ambient or partially dried sampled air displaces contaminant-bound CO₂. The PFP samples are analyzed for many species, and we are reluctant to change the preparation procedure without additional testing. Flushing with clean air between samples is meant to guard against hysteresis that may occur when a PFP is sampled at a polluted site and then at a clean site. We are now testing a sampling strategy at BAO where flasks are sampled and then vented and re-sampled a few hours later. Preliminary results are encouraging (Figs. E2 and E3), but to date only 17 samples have been collected with the new protocol under low-variability conditions. BAO CO₂ data exhibit high variability even on timescales of minutes,

so filtering for variability is necessary even when continuous in situ data are available for comparison.

Flask versus in situ comparisons throughout the tower network indicate that a growing number of flasks are producing biased data. The mean difference at BAO for 1 January–18 July 2013 was $+0.36$ ppm and the majority of PFPs appear to have one or more positively biased flasks (Fig. E2). If the bias is caused by contamination, the increasing frequency may be the result of routine use throughout the network or use under polluted conditions (e.g., oil and gas fields, urban areas). In the past, PFP units that failed dry-air testing have improved after disassembling and then cleaning and baking the individual flasks at temperatures > 500 °C – a time-consuming and expensive process. We hope to identify any contaminants so that steps can be taken to prevent future occurrences, such as additional filtering, improved sample drying to prevent condensation, and/or dedicating certain PFPs for use in polluted conditions.

Supplementary material related to this article is available online at <http://www.atmos-meas-tech.net/7/647/2014/amt-7-647-2014-supplement.zip>.

Acknowledgements. We thank Peter Bakwin for valuable advice based on his experiences as lead scientist for the NOAA tall tower network. Aris Legoretta and David (Zim) Sherman contributed to design and testing of the CO₂/CO system. Ron Teclaw helped to establish the LEF site and assisted with operations for more than a decade. Ranjan Muttiah, Joaquin Sanabria, and Onyango Okello provided support at WKT with cooperation from the Blackland Research and Extension Center. Local support at WBI has been provided by students from the University of Iowa, including Alicia Pettibone, Adam Beranek-Collins, Jameson Schoenfelder, Alex Bender, Robert Bullard, Ben Behrendt, and Andrew Hesselink, several of whom were sponsored by the NASA Iowa Space Grant Consortium. Site support at SNP has been provided by Temple Lee. Sonja Wolter, Jack Higgs, Doug Guenther, and Kelly Sours have contributed substantially to tower PFP sampling efforts. Colm Sweeney, Anna Karion, Huilin Chen, and Tim Newberger loaned Picarro CRDS analyzers for testing and provided advice particularly regarding installation and calibration of Picarro analyzers, uncertainty estimation, and PFP comparisons. Jeff Peischl and Tom Ryerson contributed data and results from comparisons with the NOAA P-3 CO₂ analyzer. Scott Richardson and Tasha Miles carried out and analyzed data from the WBI CRDS comparison. Installation of equipment at AMT was made possible by a grant from the National Science Foundation Biocomplexity in the Environment Program (ATM-0221850). We thank Verizon Wireless for continuing support at AMT and Steve Wofsy and David Hollinger for assistance with establishing the AMT site. Installation of equipment at SCT was accomplished with funding provided by the DOE Office of Science – Terrestrial Carbon Processes program and performed under contract no. DE-AC09-08SR22470. The Savannah River National Laboratory is operated by Savannah River Nuclear Solutions, LLC under Contract No. DE-AC09-08SR22470 with the US Department of Energy. Roger Strand and others at Wisconsin Public Television have facilitated installation and

maintenance of equipment at LEF. Deployment and operation of the Picarro analyzer at WGC is collaboration between NOAA ESRL and LBNL and has been supported by the California Energy Commission (CEC) Public Interest Environmental Research Program and the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH11231. Ankur Desai's work at LEF has been supported by NSF grant no. DEB-0845166. Anna Karion, Huilin Chen, Colm Sweeney, Andrew Crotwell, and two anonymous reviewers provided many helpful suggestions for improving the paper. Referee #1 provided an especially thorough review that motivated additional testing of our analytical system and more careful description of the uncertainty framework.

Edited by: D. Brunner

References

- Bakwin, P. S., Tans, P. P., Hurst, D. F., and Zhao, C. L.: Measurements of carbon dioxide on very tall towers: results of the NOAA/CMDL program, *Tellus Ser. B-Chemical and Physical Meteorology*, 50, 401–415, 1998.
- Bakwin, P. S., Davis, K. J., Yi, C., Wofsy, S. C., Munger, J. W., Haszpra, L., and Barcza, Z.: Regional carbon dioxide fluxes from mixing ratio data, *Tellus*, 56, 301–311, doi:10.1111/j.1600-0889.2004.00111.x, 2004.
- Berger, B. W., Davis, K. J., Yi, C. X., Bakwin, P. S., and Zhao, C. L.: Long-term carbon dioxide fluxes from a very tall tower in a northern forest: Flux measurement methodology, *J. Atmos. Ocean. Technol.*, 18, 529–542, 2001.
- Brooks, B.-G. J., Desai, A. R., Stephens, B. B., Bowling, D. R., Burns, S. P., Watt, A. S., Heck, S. L., and Sweeney, C.: Assessing filtering of mountaintop CO₂ mole fractions for application to inverse models of biosphere-atmosphere carbon exchange, *Atmos. Chem. Phys.*, 12, 2099–2115, doi:10.5194/acp-12-2099-2012, 2012.
- Chen, H., Winderlich, J., Gerbig, C., Hofer, A., Rella, C. W., Crosson, E. R., Van Pelt, A. D., Steinbach, J., Kolle, O., Beck, V., Daube, B. C., Gottlieb, E. W., Chow, V. Y., Santoni, G. W., and Wofsy, S. C.: High-accuracy continuous airborne measurements of greenhouse gases (CO₂ and CH₄) using the cavity ring-down spectroscopy (CRDS) technique, *Atmos. Meas. Tech.*, 3, 375–386, doi:10.5194/amt-3-375-2010, 2010.
- Committee on Methods for Estimating Greenhouse Gas Emissions, and National Research Council: *Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements*. The National Academies Press, 2010.
- Crevoisier, C., Sweeney, C., Gloor, M., Sarmiento, J. L., and Tans, P. P.: Regional US carbon sinks from three dimensional atmospheric CO₂ sampling, *Proc. Natl. Aca. Sci. USA*, 107, 18348–18353, doi:10.1073/pnas.0900062010, 2010.
- Crosson, E. R.: A cavity ring-down analyzer for measuring atmospheric levels of methane, carbon dioxide, and water vapor, *Appl. Phys. B-Lasers and Optics*, 92, 403–408, doi:10.1007/s00340-008-3135-y, 2008.
- Daube, B. C., Boering, K. A., Andrews, A. E., and Wofsy, S. C.: A high-precision fast-response airborne CO₂ analyzer for in situ sampling from the surface to the middle stratosphere, *J. Atmos. Ocean. Technol.*, 19, 1532–1543, 2002.
- Deeter, M. N., Worden, H. M., Edwards, D. P., Gille, J. C., and Andrews, A. E.: Evaluation of MOPITT retrievals of lower-tropospheric carbon monoxide over the United States, *J. Geophys. Res.-Atmos.*, 117, D13306, doi:10.1029/2012jd017553, 2012.
- 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, *J. Geophys. Res.*, 110, D18306, doi:10.1029/2005JD006035, 2005.
- French, B. and May, E.: USCRN Temperature Aspirated Shield Modifications, NOAA Technical Note NCDC No. USCRN-04-06, available at: <http://www1.ncdc.noaa.gov/pub/data/uscrn/documentation/program/technotes/TN04006AspiratedShieldMods.pdf> (last access: 28 January 2014), 2004.
- Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: North American CO₂ exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, *Biogeosciences*, 9, 457–475, doi:10.5194/bg-9-457-2012, 2012.
- Working Group 1 of the Joint Committee for Guides in Metrology, Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM), GUM 1995 with minor corrections, Bureau International des Poids et Mesures (BIPM) (Sèvres, France) and International Organization for Standardization (Geneva, Switzerland), available at: http://www.bipm.org/utills/common/documents/jcgm/JCGM_100_2008_E.pdf (last access: 28 January 2014), 132 pp., 2008.
- Jeong, S., Zhao, C. F., Andrews, A. E., Bianco, L., Wilczak, J. M., and Fischer, M. L.: Seasonal variation of CH₄ emissions from central California, *J. Geophys. Res.-Atmos.*, 117, D11306, doi:10.1029/2011jd016896, 2012.
- Karion, A., Sweeney, C., Wolter, S., Newberger, T., Chen, H., Andrews, A., Kofler, J., Neff, D., and Tans, P.: Long-term greenhouse gas measurements from aircraft, *Atmos. Meas. Tech.*, 6, 511–526, doi:10.5194/amt-6-511-2013, 2013.
- Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., Sweeney, C., Diaz, L. I., Martins, D., Shepson, P. B., and Davis, K. J.: Constraining the CO₂ budget of the corn belt: exploring uncertainties from the assumptions in a mesoscale inverse system, *Atmos. Chem. Phys.*, 12, 337–354, doi:10.5194/acp-12-337-2012, 2012.
- Law, R. M., Peters, W., Rodenbeck, C., Aulagnier, C., Baker, I., Bergmann, D. J., Bousquet, P., Brandt, J., Bruhwiler, L., Cameron-Smith, P. J., Christensen, J. H., Delage, F., Denning, A. S., Fan, S., Geels, C., Houweling, S., Imasu, R., Karstens, U., Kawa, S. R., Kleist, J., Krol, M. C., Lin, S. J., Lokupitiya, R., Maki, T., Maksyutov, S., Niwa, Y., Onishi, R., Parazoo, N., Patra, P. K., Pieterse, G., Rivier, L., Satoh, M., Serrar, S., Taguchi, S., Takigawa, M., Vautard, R., Vermeulen, A. T., and Zhu, Z.: TransCom model simulations of hourly atmospheric CO₂: Experimental overview and diurnal cycle results for 2002, *Global Biogeochem. Cy.*, 22, Gb3009, doi:10.1029/2007gb003050, 2008.
- Lee, T. R., De Wekker, S. F. J., Andrews, A. E., Kofler, J., and Williams, J.: Carbon dioxide variability dur-

- ing cold front passages and fair weather days at a forested mountaintop site, *Atmos. Environ.*, 46, 405–416, doi:10.1016/j.atmosenv.2011.09.068, 2012.
- Le Quééré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., Conway, T. J., Doney, S. C., Feely, R. A., Foster, P., Friedlingstein, P., Gurney, K., Houghton, R. A., House, J. I., Huntingford, C., Levy, P. E., Lomas, M. R., Majkut, J., Metzl, N., Ometto, J. P., Peters, G. P., Prentice, I. C., Randerson, J. T., Running, S. W., Sarmiento, J. L., Schuster, U., Sitch, S., Takahashi, T., Viovy, N., van der Werf, G. R., and Woodward, F. I.: Trends in the sources and sinks of carbon dioxide, *Nat. Geosci.*, 2, 831–836, doi:10.1038/ngeo689, 2009.
- Ma, S. and Skou, E.: CO₂ permeability in Nation@5 EW1100 at elevated temperature, *Solid State Ionics*, 178, 615–619, doi:10.1016/j.ssi.2007.01.030, 2007.
- Masarie, K. A., Langenfelds, R. L., Allison, C. E., Conway, T. J., Dlugokencky, E. J., Francey, R. J., Novelli, P. C., Steele, L. P., Tans, P. P., Vaughn, B., and White, J. W. C.: NOAA/CSIRO Flask Air Intercomparison Experiment: A strategy for directly assessing consistency among atmospheric measurements made by independent laboratories, *J. Geophys. Res.-Atmos.*, 106, 20445–20464, doi:10.1029/2000jd000023, 2001.
- Meesters, A. G. C. A., Tol, L. F., Peters, W., Hutjes, R. W. A., Vellinga, O. S., Elbers, J. A., Vermeulen, A. T., van der Laan, S., Neubert, R. E. M., Meijer, H. A., and Dolman, A. J.: Inverse carbon dioxide flux estimates for the Netherlands, *J. Geophys. Res.*, 117, D20306, doi:10.1029/2012JD017797, 2012.
- Miles, N. L., Richardson, S. J., Davis, K. J., Lauvaux, T., Andrews, A. E., West, T. O., Bandaru, V., and Crosson, E. R.: Large amplitude spatial and temporal gradients in atmospheric boundary layer CO₂ mole fractions detected with a tower-based network in the US upper Midwest, *J. Geophys. Res.-Biogeosci.*, 117, 01019, doi:10.1029/2011jg001781, 2012.
- O’Keefe, A., Scherer, J. J., and Paul, J. B.: cw Integrated cavity output spectroscopy, *Chem. Phys. Lett.*, 307, 343–349, 1999.
- Peischl, J., Ryerson, T. B., Holloway, J. S., Parrish, D. D., Trainer, M., Frost, G. J., Aikin, K. C., Brown, S. S., Dube, W. P., Stark, H., and Fehsenfeld, F. C.: A top-down analysis of emissions from selected Texas power plants during TexAQS 2000 and 2006, *J. Geophys. Res.-Atmos.*, 115, D16303, doi:10.1029/2009jd013527, 2010.
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *P. Natl. Acad. Sci. USA*, 104, 18925–18930, 2007.
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Pétron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, *Glob. Change Biology*, 16, 1317–1337, doi:10.1111/j.1365-2486.2009.02078.x, 2010.
- Pillai, D., Gerbig, C., Ahmadov, R., Rödenbeck, C., Kretschmer, R., Koch, T., Thompson, R., Neininger, B., and Lavrié, J. V.: High-resolution simulations of atmospheric CO₂ over complex terrain – representing the Ochsenkopf mountain tall tower, *Atmos. Chem. Phys.*, 11, 7445–7464, doi:10.5194/acp-11-7445-2011, 2011.
- Popa, M. E., Gloor, M., Manning, A. C., Jordan, A., Schultz, U., Haensel, F., Seifert, T., and Heimann, M.: Measurements of greenhouse gases and related tracers at Bialystok tall tower station in Poland, *Atmos. Meas. Tech.*, 3, 407–427, doi:10.5194/amt-3-407-2010, 2010.
- Rella, C. W., Chen, H., Andrews, A. E., Filges, A., Gerbig, C., Hatakka, J., Karion, A., Miles, N. L., Richardson, S. J., Steinbacher, M., Sweeney, C., Wastine, B., and Zellweger, C.: High accuracy measurements of dry mole fractions of carbon dioxide and methane in humid air, *Atmos. Meas. Tech.*, 6, 837–860, doi:10.5194/amt-6-837-2013, 2013.
- Richardson, S. J., Miles, N. L., Davis, K. J., Crosson, E. R., Rella, C. W., and Andrews, A. E.: Field Testing of Cavity Ring-Down Spectroscopy Analyzers Measuring Carbon Dioxide and Water Vapor, *J. Atmos. Ocean. Technol.*, 29, 397–406, doi:10.1175/jtech-d-11-00063.1, 2012.
- Schuh, A. E., Denning, A. S., Corbin, K. D., Baker, I. T., Uliasz, M., Parazoo, N., Andrews, A. E., and Worthy, D. E. J.: A regional high-resolution carbon flux inversion of North America for 2004, *Biogeosciences*, 7, 1625–1644, doi:10.5194/bg-7-1625-2010, 2010.
- Skoog, D. A. and Leary, J. J.: Principles of Instrumental Analysis, 4th Edn., Saunders College Publishing, Fort Worth, TX, 1992.
- Stephens, B. B., Miles, N. L., Richardson, S. J., Watt, A. S., and Davis, K. J.: Atmospheric CO₂ monitoring with single-cell NDIR-based analyzers, *Atmos. Meas. Tech.*, 4, 2737–2748, doi:10.5194/amt-4-2737-2011, 2011.
- Thompson, R. L., Manning, A. C., Gloor, E., Schultz, U., Seifert, T., Hänsel, F., Jordan, A., and Heimann, M.: In-situ measurements of oxygen, carbon monoxide and greenhouse gases from Ochsenkopf tall tower in Germany, *Atmos. Meas. Tech.*, 2, 573–591, doi:10.5194/amt-2-573-2009, 2009.
- Turnbull, J., Guenther, D., Karion, A., Sweeney, C., Anderson, E., Andrews, A., Kofler, J., Miles, N., Newberger, T., Richardson, S., and Tans, P.: An integrated flask sample collection system for greenhouse gas measurements, *Atmos. Meas. Tech.*, 5, 2321–2327, doi:10.5194/amt-5-2321-2012, 2012.
- Vermeulen, A. T., Hensen, A., Popa, M. E., van den Bulk, W. C. M., and Jongejan, P. A. C.: Greenhouse gas observations from Cabauw Tall Tower (1992–2010), *Atmos. Meas. Tech.*, 4, 617–644, doi:10.5194/amt-4-617-2011, 2011.
- VIM: International vocabulary of metrology. Basic and general concepts and associated terms, 3rd Edn., Joint Committee for Guides in Metrology (JCGM), 2008.
- Working Group 2 of the Joint Committee for Guides in Metrology, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), 3rd Edn., Bureau International des Poids et Mesures (BIPM) (Sèvres, France) and International Organization for Standardization (ISO) (Geneva, Switzerland), available at: http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2008.pdf (last access: 28 January 2014), 104 pp., 2008.
- Welp, L. R., Keeling, R. F., Weiss, R. F., Paplawsky, W., and Heckman, S.: Design and performance of a Nafion dryer for continuous operation at CO₂ and CH₄ air monitoring sites, *Atmos. Meas. Tech.*, 6, 1217–1226, doi:10.5194/amt-6-1217-2013, 2013.

- Winderlich, J., Chen, H., Gerbig, C., Seifert, T., Kolle, O., Lavrič, J. V., Kaiser, C., Höfer, A., and Heimann, M.: Continuous low-maintenance CO₂/CH₄/H₂O measurements at the Zotino Tall Tower Observatory (ZOTTO) in Central Siberia, *Atmos. Meas. Tech.*, 3, 1113–1128, doi:10.5194/amt-3-1113-2010, 2010.
- WMO: Guidelines for the measurement of atmospheric carbon monoxide, GAW 192, available at: <http://www.wmo.int/pages/prog/arep/gaw/gaw-reports.html> (last access: 28 January 2014), World Meteorological Organization, Geneva, Switzerland, 2010.
- WMO: Report of the 15th WMO/IAEA Meeting of Experts on Carbon Dioxide, Other Greenhouse Gases, and Related Tracers Measurement Techniques, Jena, Germany, 7–10 September 2009, GAW 194, available at: <http://www.wmo.int/pages/prog/arep/gaw/gaw-reports.html> (last access: 28 January 2014), World Meteorological Organization, Geneva, Switzerland, 2011.
- WMO: Report of the 16th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases, and Related Measurement Techniques (GGMT-2011), Wellington, New Zealand, 25–28 October 2011, GAW 206, available at: <http://www.wmo.int/pages/prog/arep/gaw/gaw-reports.html> (last access: 28 January 2014), World Meteorological Organization, Geneva, Switzerland, 2012.
- Wofsy, S. C. and Harriss, R. C.: The North American Carbon Program Plan (NACP), Report of the NACP Committee of the US Carbon Cycle Science Program, US Global Change Research Program, Washington, DC, 2002.
- Worthy, D. E. J., Platt, A., Kessler, R., Ernst, M., Braga, R., and Racki, S.: The Canadian atmospheric carbon dioxide measurement program: Measurement procedures, data quality and accuracy, In: Report of the 11th WMO/IAEA Meeting of Experts on Carbon Dioxide Concentration and Related Tracer Measurement Techniques, Tokyo, Japan, September 2001, edited by: Toru, S. and Kazuto, S., World Meteorological Organization Global Atmosphere Watch, 112–128, 2003.
- Wunch, D., Toon, G. C., Blavier, J. F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network, *Philos. Trans. Roy. Soc. a-Mathematical Physical and Engineering Sciences*, 369, 2087–2112, doi:10.1098/rsta.2010.0240, 2011.
- Zhao, C. F., Andrews, A. E., Bianco, L., Eluszkiewicz, J., Hirsch, A., MacDonald, C., Nehrkorn, T., and Fischer, M. L.: Atmospheric inverse estimates of methane emissions from Central California, *J. Geophys. Res.-Atmos.*, 114, D16302, doi:10.1029/2008jd011671, 2009.
- Zhao, C. L. and Tans, P. P.: Estimating uncertainty of the WMO mole fraction scale for carbon dioxide in air, *J. Geophys. Res.-Atmos.*, 111, D08s09, doi:10.1029/2005jd006003, 2006.
- Zhao, C. L., Bakwin, P. S., and Tans, P. P.: A design for unattended monitoring of carbon dioxide on a very tall tower, *J. Atmos. Ocean. Technol.*, 14, 1139–1145, 1997.