PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The Total Carbon Column Observing Network

Debra Wunch, Geoffrey C. Toon, Jean-François L. Blavier, Rebecca A. Washenfelder, Justus Notholt, Brian J. Connor, David W. T. Griffith, Vanessa Sherlock and Paul O. Wennberg

Phil. Trans. R. Soc. A 2011 **369**, 2087-2112
doi: 10.1098/rsta.2010.0240

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

Phil. Trans. R. Soc. A (2011) **369**, 2087–2112
doi:10.1098/rsta.2010.0240

The Total Carbon Column Observing Network

By Debra Wunch[1,*], Geoffrey C. Toon[2], Jean-François L. Blavier[2], Rebecca A. Washenfelder[1,†], Justus Notholt[3], Brian J. Connor[4], David W. T. Griffith[5], Vanessa Sherlock[6] and Paul O. Wennberg[1]

[1]*Department of Earth Science and Engineering, California Institute of Technology, Pasadena, CA 91125, USA*
[2]*Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, CA 91109, USA*
[3]*Institute of Environmental Physics, University of Bremen, 28334 Bremen, Germany*
[4]*BC Consulting Limited, Alexandra 9320, New Zealand*
[5]*School of Chemistry, University of Wollongong, Wollongong, NSW 2522, Australia*
[6]*Department of Atmospheric Research, National Institute of Water and Atmospheric Research (NIWA) Ltd, Wellington 6021, New Zealand*

A global network of ground-based Fourier transform spectrometers has been founded to remotely measure column abundances of CO_2, CO, CH_4, N_2O and other molecules that absorb in the near-infrared. These measurements are directly comparable with the near-infrared total column measurements from space-based instruments. With stringent requirements on the instrumentation, acquisition procedures, data processing and calibration, the Total Carbon Column Observing Network (TCCON) achieves an accuracy and precision in total column measurements that is unprecedented for remote-sensing observations (better than 0.25% for CO_2). This has enabled carbon-cycle science investigations using the TCCON dataset, and allows the TCCON to provide a link between satellite measurements and the extensive ground-based *in situ* network.

Keywords: greenhouse gases; carbon cycle; Fourier transform spectrometry

1. Introduction

The global strengths and distributions of the sources and sinks of greenhouse gases such as carbon dioxide (CO_2) and methane (CH_4) are currently inferred primarily from *in situ* measurements made at a network of surface sites (e.g. GLOBALVIEW-CO_2 [1]). Because of the covariance between the exchange of carbon between the surface and atmosphere and transport on diurnal and seasonal time scales, errors in simulated transport are aliased into the inferred carbon exchange [2–6]. This makes it difficult to use surface *in situ* data to test hypotheses

*Author for correspondence (dwunch@caltech.edu).
†Present address: National Oceanic and Atmospheric Administration, Boulder, CO 80305, USA.

One contribution of 17 to a Discussion Meeting Issue 'Greenhouse gases in the Earth system: setting the agenda to 2030'.

about the exchange, especially in the interior of continents. The proximity of local sources and sinks to surface measurement sites further complicates interpretation of these data.

To help disentangle the effects of atmospheric mixing from the surface exchange, column measurements can be employed. Column-averaged dry-air mole fractions (DMFs; denoted X_G for gas G) are particularly useful for this purpose because they are insensitive to variations in surface pressure and atmospheric water vapour. Because the column vertically integrates the concentration of CO_2 above the surface, it is much less affected by vertical transport than surface *in situ* measurements. Therefore, the horizontal gradients in measured X_{CO_2} are more directly related to the underlying regional-scale fluxes than is the case for the surface *in situ* measurements of CO_2 [6]. The latitudinal gradients and temporal variations in X_{CO_2} are even smaller than those in the surface CO_2, however, and therefore column measurements must have unprecedented measurement precision and accuracy. Rayner & O'Brien [7] determined that, for global, monthly averaged data, approximately 0.25 per cent precision in X_{CO_2} is sufficient to improve upon our current knowledge of the carbon cycle. Olsen & Randerson [8] calculated that a 0.1 per cent precision in a column measurement would be needed to detect the estimated $1\,\mathrm{GtC\,yr^{-1}}$ interhemispheric flux difference.

The Total Carbon Column Observing Network (TCCON) is a ground-based network of Fourier transform spectrometers (FTSs) designed to retrieve precise and accurate column abundances of CO_2, CH_4, N_2O and CO from near-infrared (NIR) solar absorption spectra. The scientific goals of the network are to improve our understanding of the carbon cycle, to provide the primary validation dataset for retrievals of X_{CO_2} and X_{CH_4} from space-based instruments, and to provide a transfer standard between the satellite measurements and the ground-based *in situ* network.

2. Network

The TCCON was established in 2004 with a primary focus of measuring precise and accurate columns of CO_2. Currently, there are 18 sites affiliated with TCCON, 15 of which are operational. The sites and their locations are shown in figure 1.

The TCCON sites at Park Falls, Darwin, Lamont, Orléans, Białystok and Lauder are co-located with or near tall towers or surface *in situ* measurements that are part of the GLOBALVIEW *in situ* network. This facilitates intercomparisons with data assimilation products like CarbonTracker [9], and also with aircraft measurements, which often occur over tall tower sites.

TCCON sites span a wide range of atmospheric states and observing conditions, from the tropics to the polar regions, continental and maritime, polluted and clean. There are notable gaps in coverage, particularly in South America, Africa and Asia.

The TCCON aims to provide a critically maintained, long-time-scale record to identify temporal drifts and spatial biases in the calibration of space-based sensors (for example, the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY), Tropospheric Emission Spectrometer (TES), Atmospheric Infrared Sounder (AIRS), Orbiting Carbon Observatory-2 (OCO-2)

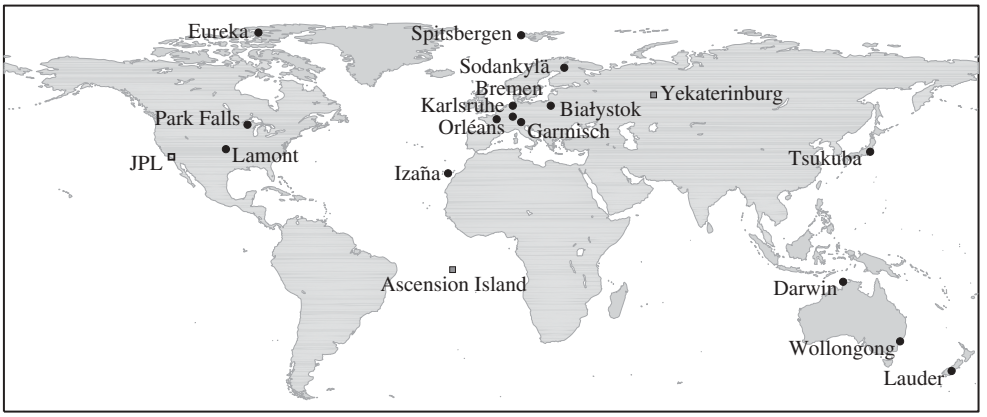

Figure 1. TCCON site locations. Currently operational sites are indicated with filled circles, and the planned future sites are indicated with squares. The JPL instrument was moved to Lamont in 2008.

and Greenhouse Gas Observing Satellite (GOSAT)). Additionally, TCCON spectra provide a means of testing and validating satellite retrieval algorithms over a broad variation in atmospheric state. In contrast to the reflected sun observations of the space-based sensors (e.g. SCIAMACHY and GOSAT), the accuracy of the retrievals from the TCCON spectra is minimally influenced by aerosol, uncertainty in airmass or variation in land surface properties. The TCCON data provide the ideal validation dataset for SCIAMACHY, GOSAT and OCO-2, since they measure in the same spectral region.

3. Technical approach and methodology

Spectrally resolved observations of the Sun obtained at the Earth's surface have been used extensively to probe the chemistry of Earth's atmosphere. Two examples are the Dobson network [10] and the Network for the Detection of Stratospheric Change [11], which was recently refocused to observe the stratosphere and troposphere and renamed the Network for the Detection of Atmospheric Composition Change (NDACC). Using the Sun as a light source allows every atmospheric molecule along the light path to be surveyed, independent of its altitude.

Although the strongest absorption bands are found in the mid-infrared spectral region, for measurements of CO_2 and CH_4, the weaker combination and overtone bands found in the NIR are more useful. Yang *et al.* [12], repeating the analysis of Wallace & Livingston [13] with updated retrieval methods and improved spectroscopy, demonstrated that FTS measurements from the Kitt Peak National Observatory could produce accurate and precise retrievals of X_{CO_2}, paving the way for the development of the TCCON. A similar analysis was performed later by Dufour *et al.* [14].

The precision and accuracy requirements for the TCCON are significantly more exacting than those required or achieved to date by ground-based FTS networks (e.g. NDACC). In fact, the precision requirement on the TCCON X_{CO_2} is likely to be stricter than for satellite instruments such as the Orbiting

Carbon Observatory, which is calculated to be approximately 0.25 per cent (less than 1 ppm) for global monthly means [15]. This is because of the reduced spatial density of the TCCON observations, and because the TCCON measurements must be more precise and accurate than the satellite measurements if they are to provide a useful validation dataset. A modelling study is needed to assess the sensitivity of the TCCON to surface exchange, and hence set the measurement precision and accuracy requirements on TCCON for improving our knowledge of the carbon cycle. As a network, however, we aim to achieve the highest possible precision, and are constantly improving our methodology to do so.

Although each site is run by its own principal investigator and team of scientists and technicians, to ensure a cohesive data product that is accurate and precise, membership in the network requires adherence to a set of standards and procedures for data acquisition and analysis, which include:

— common instrumentation, data acquisition procedures and settings between sites;
— common data processing and analysis software; and
— calibration onto the World Meteorological Organization's (WMO) gas scale.

Each requirement is described in the subsequent subsections.

(*a*) *Instrumentation*

The precision requirements on the TCCON limit the acceptable instrumentation. The IFS 125HR, made by Bruker Optics GmbH, is currently the most robust and stable high-resolution FTS commercially available, and is hence the preferred primary instrument. The TCCON instruments measure in the NIR because this region of the spectrum includes absorption bands of the gases of interest (O_2, CO_2, CH_4, CO, N_2O) with favourable characteristics (temperature-independent absorption lines that are not too saturated). Room-temperature InGaAs and Si detectors are available with the 125HR that cover the entire 3900–15 500 cm^{-1} spectral region simultaneously. These detectors are supported by custom electronics that allow for solar intensity variations during data acquisition to be corrected using the method described by Keppel-Aleks *et al.* [16], resulting in higher-quality spectra and less data loss during partly cloudy conditions (see appendix A(*b*) for further details). The spectra are recorded with 45 cm maximum optical path difference (OPD) (approx. 0.02 cm^{-1} spectral resolution).

Ancillary measurements at each TCCON site include accurate surface pressure and temperature measurements. Details of the instrumentation can be found in appendix A(*a*), and specific information about Park Falls and Darwin can be found in Washenfelder *et al.* [17] and Deutscher *et al.* [18].

(*b*) *Data processing*

The TCCON data processing scheme is designed to minimize algorithmic biases between sites, and to allow future software improvements to be easily disseminated and implemented throughout the network. To this end, a common, open-source software package is used network-wide for data processing and analysis. The software package processes the raw interferograms, corrects them

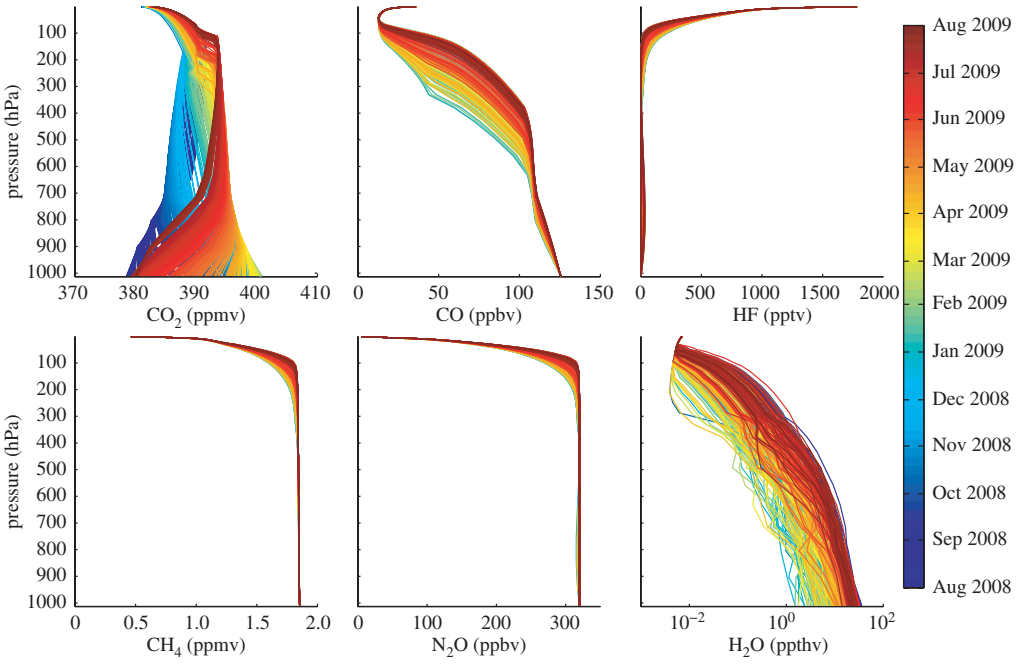

Figure 2. *A priori* volume mixing ratio profiles for Lamont, Oklahoma. The colours represent the time of year, from August 2008 in blue to August 2009 in red. Note the steep drop-off of CH_4 and N_2O in the stratosphere and the coinciding steep increase in the stratospheric HF.

for solar intensity variations and phase errors, and computes spectra using a fast Fourier transform algorithm (appendix A(*b*)). Our retrieval approach is centred around the GFIT nonlinear least-squares fitting algorithm, described in appendix A(*c*)(i), which computes column abundances from the solar absorption spectra. The GFIT algorithm scales an *a priori* profile to generate the best spectral fit, and integrates the scaled profile to compute the column abundance. The retrieved column abundances are converted to column-averaged DMFs by dividing them by the column of dry air, which we compute by retrieving the column abundance of oxygen (O_2) from the same spectra. This significantly reduces the effect of surface pressure variations on the column, avoids the need to account for H_2O in the column and improves the TCCON precision. Details of this method can be found in appendix A(*d*).

The *a priori* profiles generated by the TCCON retrieval algorithm are based on National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) analyses for temperature, pressure and humidity, coupled with empirical models for CO_2, CO, CH_4 and N_2O that were developed from MkIV FTS balloon flights [19], profiles from the satellite instrument Atmospheric Chemistry Experiment (ACE) FTS [20] and the *in situ* GLOBALVIEW data [1]. Figure 2 shows examples of the *a priori* profiles for the Lamont site. Details of how the *a priori* profiles are generated are given in appendix A(*c*)(iii).

The total columns are corrected for systematic errors introduced by spectroscopic inadequacies (e.g. spurious airmass dependences, appendix A(*e*)), and calibrated using comparisons with profiles obtained from aircraft (described

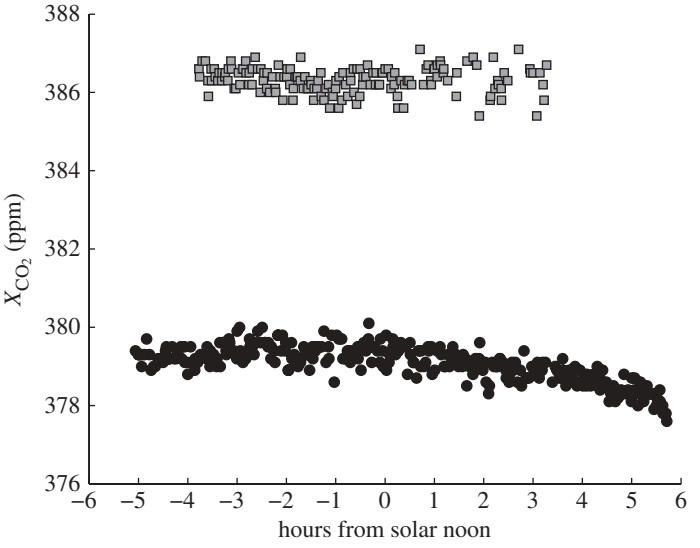

Figure 3. Changes in X_{CO_2} of as little as 1 ppm are clearly discernable on a summer's day at Park Falls, Wisconsin (25 August 2008, circles). This is contrasted by a relatively invariant autumn day (3 November 2009, squares).

in §3*c*. The final products of the data analysis are calibrated column-averaged DMFs with a sufficiently high precision to discern the small (0.2%) drawdown in the CO_2 column during summer days (figure 3 and [17]).

(*c*) *Calibration*

To use TCCON for investigating regional- and global-scale carbon science questions and for providing a transfer standard between the *in situ* network and the satellite retrievals of total column CO_2, TCCON retrievals must be placed on the same absolute calibration scale as the *in situ* network (the World Meteorological Organization (WMO) scale). Uncertainties in the spectral line parameters (currently approx. 1%) are too large to achieve the necessary accuracy using a self-calibrating approach. Instead, we have used aircraft profiling above the FTS stations to address the absolute calibration.

Washenfelder *et al.* [17] describe the first calibrations of TCCON X_{CO_2} to the WMO scale for the Park Falls site. A reanalysis of those data, using the current GFIT version and spectroscopy, determined that the FTS columns are biased approximately 1 per cent lower than the accepted WMO standard, most probably because of inaccuracies in the spectroscopic line lists for oxygen used in the retrievals. Results from Deutscher *et al.* [18] show a similar bias at the Darwin site. A recent series of aircraft profiles over Lamont, Park Falls, Lauder and Tsukuba is presented in Wunch *et al.* [21], and allowed TCCON measurements of CO_2, CO, N_2O, CH_4 and H_2O to be calibrated. The results show that a single calibration factor for each species is consistently obtained for all the sites calibrated so far. The ratios of the TCCON-retrieved columns to the aircraft columns are 0.989 ± 0.002 for CO_2, 0.98 ± 0.04 for CO, 0.978 ± 0.004 for CH_4, 0.96 ± 0.01

for N_2O and 1.0 ± 0.1 for H_2O, with 2σ errors. The method used in these calibrations is similar to that described by Macatangay *et al.* [22] to calibrate a non-TCCON FTS.

While we expect the spectroscopy to improve during the next few years, reducing the calibration biases, for now we simply scale our retrieved columns by the observed biases listed in Wunch *et al.* [21]. Until an aircraft profile can be performed at each TCCON site, we assume that the calibration value that is valid for Park Falls, Darwin, Lamont, Tsukuba and Lauder is valid everywhere. The observation that the value is the same for these five sites lends confidence that the bias is spectroscopic in origin. Additional profiles over Wollongong and several European TCCON sites [23] will be presented in forthcoming papers.

4. Use of the data

We anticipate that TCCON data will be valuable to the carbon-cycle scientific community for satellite validation, data assimilation and global and regional carbon-cycle studies. The public output of data from the TCCON is available from http://www.tccon.caltech.edu. When comparing with other data, averaging kernels and *a priori* profiles must be taken into account as described in §4*a*. When comparing with surface *in situ* networks, it may be more appropriate to extract the tropospheric component of the column for the species of interest. For CH_4, one can apply the method described by Washenfelder *et al.* [24]. For N_2O and CO, an extension to the Washenfelder *et al.* method is needed to account for the measurement averaging kernels.

(*a*) *Data comparisons*

In order to directly compare two measurements, the averaging kernels and *a priori* profiles of both measurements must be taken into account [25,26]. Column averaging kernels describe the altitude-dependent sensitivity of the retrieved total column to errors in the assumed vertical profile shape. For a perfect column measurement, the column averaging kernel would be 1.0 at all altitudes, but in practice, there is a greater sensitivity to some altitudes than others. The TCCON column averaging kernels are a function of the GFIT profile scaling retrieval method and the choice of parameters to be fitted, and these are the same for all TCCON data. The column averaging kernels also depend on the solar zenith angle (SZA) of the measurement, but little else. Gases with saturated absorption lines are insensitive to the stratosphere, because the stratospheric pressure broadening contribution to the line width is narrower than the saturated central region. Consequently, as the airmass increases and the lines become more saturated, the stratospheric part of the averaging kernel decreases. The column averaging kernels for the Lamont site are shown in figure 4. Other sites have very similar averaging kernels.

To compare with highly sampled vertical profile data, such as *in situ* profiles from balloons or aircraft, or model output at high vertical resolution, only the TCCON averaging kernels and *a priori* profiles are required. This is because the full averaging kernels of the highly sampled profiles are very close to delta

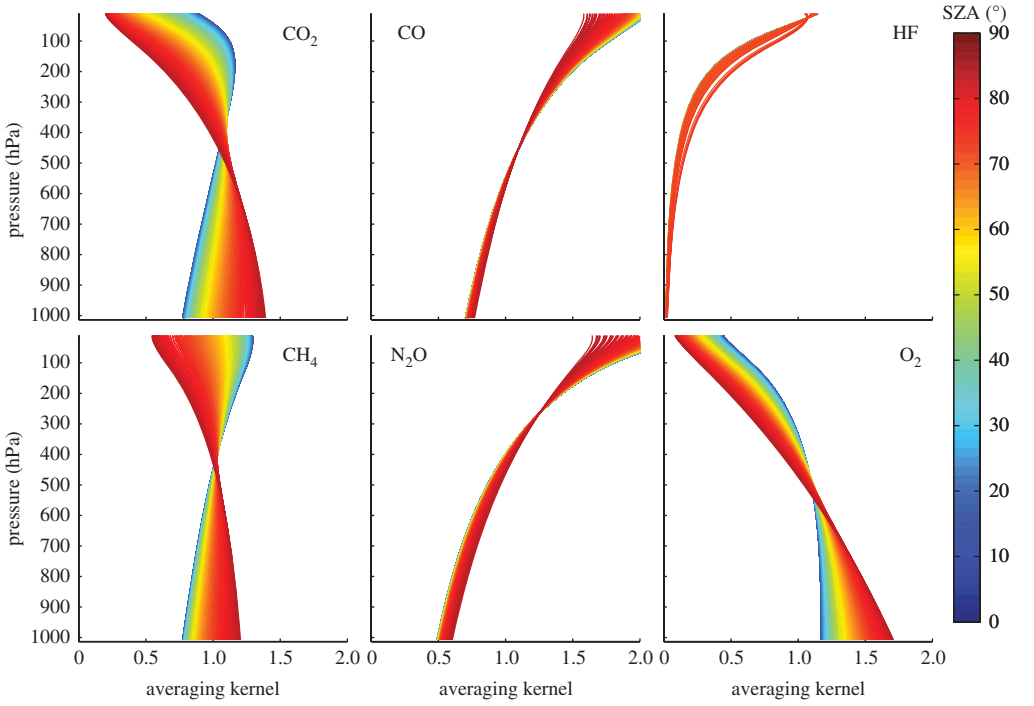

Figure 4. Examples of column averaging kernels for the Lamont TCCON site, plotted as a function of pressure. The colours represent the solar zenith angle. Column averaging kernels for other sites are similar.

functions peaked at the measurement altitudes. In these cases, the TCCON column averaging kernels are applied to the difference between the highly sampled profile and the TCCON *a priori* profiles and then vertically integrated, as described in Wunch *et al.* [21].

To use TCCON data in data assimilation, procedures like the Migliorini *et al.* [27] method (also based on the Rodgers approach [26]) can be applied.

(*b*) *Error budget*

In order to meaningfully compare TCCON measurements with others, or incorporate them into assimilation products, an error budget is required. The mean bias in our X_{CO_2} (and other gas) measurements is removed through aircraft calibration exercises. Short- and long-term precision is influenced by random noise, varying errors in the *a priori* profiles, variation in the instrument alignment and performance, and surface pressure errors.

The total error for a single measurement of X_{CO_2} at the Lamont site is 0.2 per cent or less for SZAs less than 83°. The largest errors are from instrument alignment errors, assuming a linear spectral continuum across a large spectral band, and *a priori* profile uncertainties. Details of this error calculation, including a list of potential sources of error not included in this estimate, can be found in appendix B.

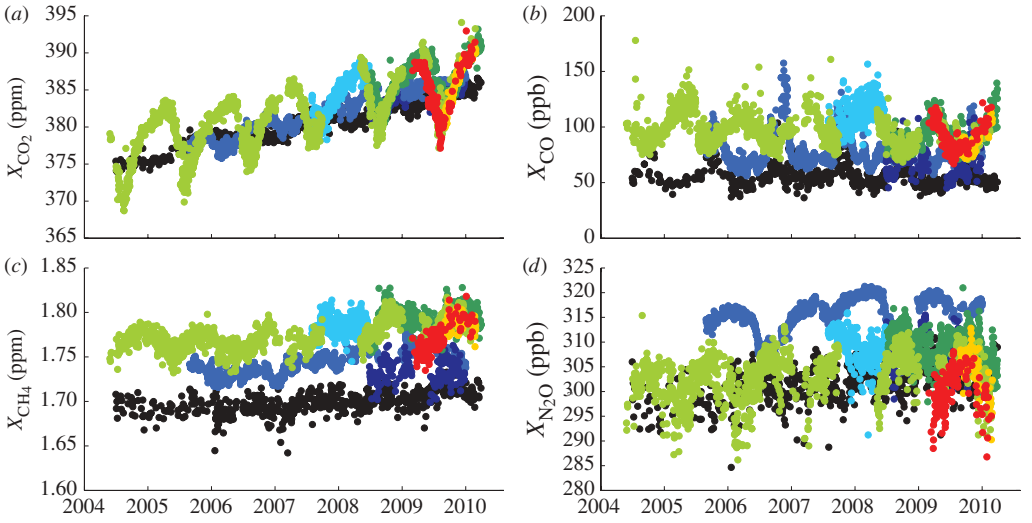

Figure 5. The daily median time series of column-averaged DMFs of (*a*) X_{CO_2}, (*b*) X_{CO}, (*c*) X_{CH_4} and (*d*) X_{N_2O} from a selection of TCCON sites. Note that the majority of the variability in X_{N_2O} and some of the variability in X_{CH_4} and X_{CO} are due to changes in the tropopause height and hence stratospheric overburden, and not to tropospheric variability. Sites: Lauder, black; Wollongong, dark blue; Darwin, mid blue; JPL, light blue; Lamont, dark green; Park Falls, light green; Orléans, yellow; Białystok, red.

5. Scientific applications

(*a*) *Surface flux and emissions estimates*

Figure 5 shows the time series of X_{CO_2}, X_{CH_4}, X_{N_2O} and X_{CO} from a selection of TCCON sites. The secular increase in X_{CO_2} since 2004 is clear, as is the large difference in the seasonal cycle amplitudes between the Northern and Southern Hemispheres. The TCCON's high precision has allowed Yang *et al.* [6] and Warneke *et al.* [28] to look closely at the seasonal cycles in X_{CO_2} from Park Falls and Spitsbergen. These studies demonstrated that the net ecosystem exchange between the atmosphere and terrestrial biosphere (i.e. the seasonal cycle in X_{CO_2}) is approximately 25 per cent larger than many models predict, revealing that vertical mixing in the models may be too weak. Keppel-Aleks *et al.* [29,30], using Park Falls, INTEX-NA aircraft data and a global general circulation model, show that the meridional gradients in X_{CO_2} in the Northern Hemisphere summer are larger than predicted, and that about one-third of the X_{CO_2} variability measured over Park Falls can be explained by synoptic (weather)-scale advection of CO_2 in the free troposphere. Keppel-Aleks *et al.* conclude that the information contained in synoptic-scale variations can constrain the large-scale spatial distribution of fluxes, and provide an additional tool for evaluating models.

TCCON methods have also been used for scientific studies of other gases. A TCCON station that was temporarily situated in the urban, polluted Los Angeles region of southern California for nearly a year was used to calculate emissions of CO and CH_4, using the known CO_2 emissions for the area [31]. The

results of this study suggest that CH_4 emissions are significant in the Los Angeles region and are higher than the reported emissions estimates. The elevated CO levels measured at the polluted JPL site are apparent in figure 5.

For gases whose stratospheric volume mixing ratio (VMR) decreases with altitude (e.g. CH_4, N_2O), much of the variability in the total column is caused by meteorological variations in tropopause pressure and stratospheric transport (e.g. figure 5, bottom panels). To separate the dynamical variations from those due to changes in the tropospheric VMRs (which are more directly related to the surface fluxes), Washenfelder *et al.* [24], using Kitt Peak data, use the linear correlation between stratospheric HF and CH_4 to infer the stratospheric column of CH_4, and hence derive a column-averaged tropospheric CH_4 VMR.

(*b*) *Satellite validation*

The Park Falls TCCON station has been used in numerous studies to validate satellite data. We anticipate that more sites will be used in future comparisons as the data become available. Barkley *et al.* [32] have used the Park Falls time series to evaluate the sensitivity of the SCIAMACHY instrument to the lowermost troposphere. They have shown that SCIAMACHY can accurately capture the seasonal cycle in X_{CO_2} due to vegetation growth on monthly time scales. Bösch *et al.* [33] use X_{CO_2} Park Falls data to reveal biases in the SCIAMACHY data. De Beek *et al.* [34], Buchwitz *et al.* [35] and Reuter *et al.* [36,37] have significantly improved the SCIAMACHY retrieval algorithms, again using TCCON data as their ground reference.

The TCCON provides an important validation dataset for GOSAT measurements of X_{CO_2} and X_{CH_4}, and will provide the primary X_{CO_2} validation dataset for OCO-2, scheduled for launch in 2013.

(*c*) *Improved spectroscopy*

Because of the high precision and spectral resolution of the TCCON data, new spectroscopic properties have been resolved and identified. Tran & Hartmann [38] have used the Park Falls spectra to verify an improved O_2 A-band absorption model. Hartmann *et al.* [39] and Tran *et al.* [40] used Park Falls spectra to illustrate their line-mixing algorithm for CO_2 and CH_4. Electric quadrupole transitions of O_2 in the 7882 cm^{-1} region were first found and identified in Park Falls TCCON data by Gordon *et al.* [41]. These enhancements to the spectroscopy will improve the precision and accuracy of both ground-based and satellite CO_2 and CH_4 retrievals in these regions. We look forward to continued collaboration with the spectroscopic community to further improve the representation of absorption of NIR radiation by atmospheric gases.

6. Conclusions

Since its inception in 2004, the TCCON network has grown to include 18 sites globally, and produces total column DMFs of CO_2, CO, CH_4, N_2O, HF and H_2O. High accuracy (approx. 0.2% in X_{CO_2}) is achieved by adhering to strict requirements on instrumentation, data processing and calibration. As the time series of the TCCON data lengthen and the data processing and analysis methods

improve, it will become an increasingly valuable dataset for studies in carbon-cycle science, for satellite validation and for linking the satellite measurements to the ground-based *in situ* network. Assimilations of satellite observations, *in situ* measurements and our ground-based total column measurements should yield new results that could not have been uncovered by any single data type alone.

NCEP reanalysis data are provided by the NOAA/OAR/ESRL PSD, Boulder, CO, USA, from their website at http://www.cdc.noaa.gov/. US funding for TCCON comes from NASA's Terrestrial Ecology Program (NNX08A186G), the Orbiting Carbon Observatory Program (NAS7-03001), the DOE/ARM Program and the Atmospheric CO_2 Observations from Space Program. Some of the research described in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. We acknowledge funding within the EU projects GEOMON and IMECC and the grant from the Deutsche Forschungsgemeinschaft (DFG) NO 404/14-1. Lauder measurements are supported by New Zealand Foundation of Research Science and Technology contracts C01X0204 and C01X0406. Australian funding is from the Australian Research Council, DP0879468 and LP0562346.

Appendix A. Instrumentation and data processing details

A glossary of terms, acronyms and abbreviations is listed in table 1.

(*a*) *Instrumentation*

The preferred instrument for TCCON measurements is the Bruker 125HR FTS, because it is currently the most stable, robust FTS commercially available. TCCON interferograms are recorded with a maximum OPD of 45 cm, giving a spectral resolution of approximately 0.02 cm^{-1}, which is the optimum resolution for a source-photon-noise-limited detector in the 6000–8000 cm^{-1} region [42]. The Bruker 125HR spectrometers use the Brault [43] method of data acquisition whereby the interferograms are sampled and digitized by a 24 bit Sigma-Delta analogue-to-digital converter (ADC), and then numerically resampled at constant intervals of OPD as defined by the average crossings (both rising and falling) of the 633 nm He : Ne reference laser signal. This provides a free spectral range of 15 798 cm^{-1}. The 24 bit ADC has sufficient dynamic range to allow the interferogram to be digitized without gain switching. The electronics also allow for DC recording of the detector signals, so that the Keppel-Aleks *et al.* [16] algorithm can be applied (appendix A(*b*)). The signal-to-noise ratio (SNR) for a single spectrum (75 s acquisition) is approximately 750 near 5000 cm^{-1}. The instrument line shape (ILS) is monitored in all instruments with a low-pressure HCl cell (at 5 hPa) that is permanently mounted in the solar beam. With proper instrument set-up and maintenance, two side-by-side instruments have been shown to retrieve X_{CO_2} with a mean bias of 0.06 per cent (0.2 ppm [44]).

A Sun tracker, which must have at least a 3 arcmin (0.05°) precision to remain centred on the Sun, directs the solar beam into the FTS. For most TCCON instruments, the Sun tracker mirrors are gold-coated to avoid the aluminium absorption at 12 000 cm^{-1}, and to reduce the visible radiation that would add photon noise and cause aliasing problems. To further attenuate the visible radiation, a red filter (with an approx. 16 000 cm^{-1} cut-off) is placed before the Si detector. A dichroic is used to separate the spectral domains of the InGaAs and

Table 1. Definition of terms.

term	definition
instrument and analysis terms	
ADC	analogue-to-digital converter
DC	direct current
DMF	dry-air mole fraction
FOV	field of view
FTS	Fourier transform spectrometer
GFIT	spectral fitting and line-by-line retrieval algorithm
HITRAN	high-resolution transmission molecular absorption database
ILS	instrument line shape
ME	modulation efficiency
(M)OPD	(maximum) optical path difference
NIR	near-infrared
SNR	signal-to-noise ratio
SZA	solar zenith angle
VMR	volume mixing ratio
VSF	volume mixing ratio scale factor
ZPD	zero path difference
networks and satellites	
ACE FTS	Atmospheric Chemistry Experiment Fourier Transform Spectrometer
AIRS	Atmospheric Infrared Sounder
GOSAT	Greenhouse Gas Observing Satellite
INTEX-NA	Intercontinental Chemical Transport Experiment–North America
NCEP/NCAR	National Centers for Environmental Prediction/National Center for Atmospheric Research
NDACC	Network for the Detection of Atmospheric Composition Change
NDSC	Network for the Detection of Stratospheric Change
OCO	Orbiting Carbon Observatory
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Chartography
TCCON	Total Carbon Column Observing Network
TES	Tropospheric Emission Spectrometer
WMO	World Meteorological Organization

Si detectors. The pointing accuracy is monitored by analysis of the spectra. Solar absorption features (so-called Fraunhofer lines) are readily observable in TCCON spectra. Because of the rotation of the Sun (with an equatorial circumferential speed of $\pm 2000\,\mathrm{m\,s^{-1}}$), lines on the east limb are significantly Doppler-shifted from those on the west limb. Thus, the difference between the calculated and measured relative position of the solar and telluric lines provides a diagnostic of the pointing performance of the solar tracking system.

Additional equipment includes a weather station containing an accurate surface pressure measurement (to better than 0.3 hPa) and an accurate surface temperature measurement (to better than 1 K). Most sites also possess auxiliary instrumentation that is useful for automation and analysis, such as humidity sensors, anemometers and wind vanes, rain gauges and pyranometers. The sites are generally autonomous, many being situated in remote locations.

(*b*) *Interferogram processing*

We choose to process our own interferograms, instead of using a packaged processing algorithm from Bruker, because the latter is neither open-source nor adequately documented for users. Our interferogram processing currently includes a solar intensity variation correction (described below), a phase correction described by Forman *et al.* [45], but with a convolution in the spectral domain as described by Mertz [46,47], and a fast Fourier transform based on the work of Bergland [48]. The InGaAs and Si detectors are sufficiently linear that no nonlinearity correction is required.

FTSs receive information about all spectral frequencies simultaneously, a feature commonly referred to as the 'Fellgett' or 'multiplex' advantage. As a consequence of this, low-spectral-resolution information (from the zero-path-difference (ZPD) region of the interferogram) is not acquired simultaneously with the high-spectral-resolution information (from the high OPD region of the interferogram). If the intensity of the received solar radiation varies during the acquisition of an interferogram, which occurs when there are clouds or aerosols in the path between the FTS and the Sun, the resulting spectrum will be distorted, since the continuum level will have a different 'gain' than the higher-resolution spectral structure. Although this distortion may be subtle, it can significantly alter the retrievals.

Keppel-Aleks *et al.* [16] developed and implemented a method of correcting for solar intensity variations. They divide the interferogram by the unmodulated DC detector signal to restore the interferogram fringes to their correct amplitudes at each OPD. In the case of TCCON data, the unmodulated DC detector signal comes from smoothing the DC-coupled detector signals. Keppel-Aleks *et al.* [16] show that, on partly cloudy days, this scheme results in a substantial reduction in the scatter of the retrieval results. For interferograms containing less than 10 per cent solar intensity variation, the standard deviation of the retrieved O_2 columns between scans is reduced from 2.5 to 0.25 per cent. This correction allows TCCON instruments to measure accurately through light cloud and aerosol, and salvages interferograms that would have previously been discarded.

(*c*) *Spectrum processing and analysis*

To compute total column abundances of the gases of interest from the spectra, an open-source software package called GGG is used. Within the GGG package, individual spectra are passed through GFIT, the main fitting and retrieval program (appendix A(*c*)(i)). Unprecedented attention to detail is required in the spectral fitting routine to accurately and precisely retrieve X_{CO_2}. Examples of this include instrument details, such as frequency shifts, air-to-vacuum correction, zero-level offsets and ILS errors, and atmospheric modelling details, such as the spectroscopic line list, *a priori* profiles and the solar model. Spectra are not averaged, as the forward and reverse scans have the opposite sensitivity to certain types of systematic errors (e.g. timing errors), which provides useful diagnostic information. It is also necessary to include the air path between the FTS and the Sun tracker, which causes a small absorption contribution that is independent of the SZA.

(i) *Retrieval algorithm*

GFIT is a nonlinear least-squares spectral fitting program designed for the analysis of FTS absorption spectra. It consists of a 'forward model', which computes an atmospheric transmittance spectrum for a prescribed set of conditions, and an 'inverse method', which compares each measured spectrum with the calculation, and decides how best to adjust the retrieved parameters to achieve a better match. The adjusted parameters are user-selectable and can include the continuum level, continuum tilt, frequency shift, zero-level offset, and volume mixing ratio scale factors (VSFs) for the various fitted gases and their isotopologues. For standard TCCON data products, these choices are prescribed to ensure consistency between sites.

In the line-by-line calculations, GFIT assumes a Voigt line shape. More complex line shapes, such as those arising from line mixing or speed dependence, are under investigation, but not yet used in the forward model. Thermal emission is negligible in the NIR. Scattering is negligible in a direct-sun viewing geometry. To increase the speed of the retrievals, a single set of absorption coefficients is used for all spectra acquired on a given day, computed for the exact pressure and temperature conditions for that day's model atmosphere. For the temperature-insensitive absorption bands used by TCCON, the error due to the neglect of temperature variations during each day is generally small (less than 0.1%).

Surface pressure measurements are used to derive the site pressure–altitude for each spectrum. In general, this differs from the true geometric altitude by up to 0.1 km mainly because of errors in the pressure–temperature model. Thus, there are two choices: (i) to have the altitude correct, but the surface pressure wrong; or (ii) to have the surface pressure correct, but the altitude wrong. We have found empirically that the latter option produces better results for gases whose VMR profile is fairly constant in the troposphere (e.g. CO_2, CH_4, O_2, N_2O, HF).

GFIT does not do a profile retrieval (i.e. the VMRs at different altitudes are not scaled independently to obtain the best fit to the spectrum). Instead, it assumes that the shapes of the gas profiles are known and simply scales these *a priori* profiles. This is faster and much simpler than a full profile retrieval. It is also less sensitive to certain types of systematic errors in the shapes of the calculated spectral lines (e.g. ILS, spectroscopic line widths, zero-level offsets, collision narrowing, line mixing). Since the VMR profile shapes of the gases most important to TCCON (e.g. O_2, CO_2) are highly predictable, little is lost by the profile-scaling approach, and because some of the TCCON sites have already acquired over 300 000 spectra, speedy analysis is essential.

The most important GFIT outputs are the VSFs and their uncertainties. The uncertainties are calculated from the diagonal elements of the inverse covariance matrix. The VSFs are multiplied by the *a priori* vertical column abundances to yield the retrieved vertical column, the number of molecules of gas G per unit area between the surface and the top of the atmosphere:

$$\text{column}_\text{G} = \text{VSF}_\text{G} \int_{z_\text{s}}^{\infty} f_\text{G}^{a\ priori} n \, \text{d}z \tag{A 1}$$

where VSF_G is the VSF for gas G, $f_\text{G}^{a\ priori}$ is the *a priori* mole fraction of gas G, n is the total number density, z is the altitude and z_s is the surface altitude.

Jacobians (weighting functions) are calculated analytically in parallel with the computation of each spectrum. This is quicker and more accurate than the finite differences method. The Jacobians are used to determine the optimal changes to the assumed state vector and to produce the averaging kernels. The optimal estimation method is used to determine the change to the state vector ($\mathbf{\Delta x}$) according to the equation

$$(\mathbf{J}^{\mathrm{T}}\mathbf{S}_{\epsilon}^{-1}\mathbf{J}+\mathbf{S}_{\mathrm{a}}^{-1})\mathbf{\Delta x}=\mathbf{J}^{\mathrm{T}}\mathbf{S}_{\epsilon}^{-1}\mathbf{r}+\mathbf{S}_{\mathrm{a}}^{-1}(\mathbf{x}-\mathbf{x}_{\mathrm{a}}), \tag{A 2}$$

where $\mathbf{J}$ is the Jacobian matrix, $\mathbf{S}_{\epsilon}$ is the measurement error covariance matrix, assumed to be a diagonal matrix consisting of the spectral noise values, $\mathbf{S}_{\mathrm{a}}$ is the *a priori* covariance matrix, which contains the estimated *a priori* errors, $\mathbf{r}$ is the spectral residuals, and $\mathbf{x}$ and $\mathbf{x}_{\mathrm{a}}$ are the current best estimates of the state vector and the *a priori* state vector, respectively.

(ii) *Instrument line-shape calculation*

The instrument line shape (ILS) of a spectrometer is the shape of the spectral line the spectrometer would produce if it measured an infinitesimally narrow spectral line. The theoretical 'perfect' FTS ILS, when the instrument is well aligned, is a convolution of sinc and rectangular functions (defined in equations (A 3) and (A 4), respectively), representing the finite length of the interferogram and the finite field of view (FOV) of the FTS, respectively. The finite FOV causes the frequency locations of the spectral intensities to spread out and shift lower (e.g. [49,50]). Thus

$$\mathrm{SINC}(\sigma,L)=2L\frac{\sin(2\pi\sigma L)}{2\pi\sigma L}, \tag{A 3}$$

$$\mathrm{RECT}(\sigma,\sigma_0,\theta)=\begin{cases}\dfrac{2}{\sigma_0\theta^2} & \text{if } -0.5\sigma_0\theta^2\le\sigma\le 0\\ 0 & \text{otherwise}\end{cases} \tag{A 4}$$

and

$$\mathrm{ILS}(\sigma,\sigma_0,L,\theta)=\mathrm{SINC}(\sigma,L)*\mathrm{RECT}(\sigma,\sigma_0,\theta), \tag{A 5}$$

where σ is the wavenumber, σ_0 is the central wavenumber, L is the maximum OPD and θ is the angular radius of the circular internal FOV of the FTS. For standard TCCON measuring conditions, $L=45$ cm and $\theta=1.2$ mrad. The ILS used in GFIT is a numerical convolution of the sinc function with the rectangular function, that is then truncated and weakly apodized using a common apodization function (e.g. Forman *et al.* [45]).

(iii) *A priori profiles*

The altitude, pressure, temperature and specific humidity profiles are provided to GFIT from the NCEP/NCAR analysis product [51]. A single profile is generated per day from the 6 hourly NCEP/NCAR data, interpolated to the latitude and longitude of the individual sites, and to local noon. These, together with the measured surface pressure at the site and *a priori* VMR profiles, are the main inputs to GFIT.

Because GFIT scales *a priori* profiles of the atmospheric gases of interest to retrieve total column abundances, *a priori* knowledge of the shape of the profile is important. To this end, global *a priori* VMR profiles, with the exception of CO_2 and H_2O, are derived from ACE FTS [20] and MkIV FTS [19] solar occultation measurements.

For CH_4 and N_2O, *a priori* profiles in the stratosphere are constrained by the well-known relationship between profiles of HF and those of CH_4 and N_2O [24,52], and the known tropopause height. In the troposphere, constant values of N_2O (320 ppb) and HF (0 ppt) are assumed. CO *a priori* profiles are generated from a climatology of ACE FTS profiles. These *a priori* profiles are not applicable to the polar winter vortices, and other *a priori* profiles must be used for polar stations.

The CO_2 *a priori* profiles are from a climatology based on the GLOBALVIEW dataset, and change based on the time of year and the latitude of the site. Stratospheric CO_2 profiles are generated from the age of air relationship derived by Andrews *et al.* [53]. The H_2O *a priori* profiles come from the NCEP/NCAR analysis for that day, interpolated to local noon, with stratospheric VMRs based on MkIV balloon profiles. A year of *a priori* profiles for Lamont are shown in figure 2.

(iv) *Spectroscopy*

Our spectroscopic line lists are updated as laboratory studies produce improved descriptions of the spectroscopy. Currently, all molecules (exceptions listed below) are fitted using the spectroscopic line lists contained in the HITRAN 2008 database [54]. The remainder of this section discusses the exceptions.

— Hundreds of supplemental weak H_2O lines are added empirically based on high-airmass Darwin spectra.

— A list of discrete O_2 lines in the band centred at 7882 cm^{-1} was created by A. J. Orr-Ewing (2006, personal communication) based on the laboratory measurements of Newman *et al.* [55], and modified in accordance with Yang *et al.* [56]. Collision-induced O_2 absorption is present in this band and is fitted using a line list based on laboratory measurements by Smith & Newnham [57] to minimize its impact on the discrete lines. Quadrupole O_2 lines have been added based on Gordon *et al.* [41].

— For ground-based instruments, it is impossible to acquire an exo-atmospheric solar spectrum with which to ratio out the instrumental and solar features. These must therefore be represented in the forward model. The solar lines are calculated from a solar line list containing over 25 000 lines extending from 600 to 20 000 cm^{-1}. This solar line list was derived from MkIV balloon spectra in the mid-infrared (600–5600 cm^{-1}) and ground-based Kitt Peak spectra in the NIR and visible. The instrument response is represented by the continuum level and tilt (and channel fringes, if necessary), which is adequate over the fairly narrow windows used by TCCON (1–2% of the full spectral bandwidth).

(*d*) O_2 *as an internal standard: calculation of dry-air mole fractions*

The column abundances of gas G computed from GFIT (i.e. equation (A 1)) are converted to column-averaged DMFs by dividing them by the total column of dry air:

$$X_{\mathrm{G}} = \frac{\mathrm{column_G}}{\mathrm{column\ dry\ air}}. \tag{A 6}$$

This division removes variations due to surface pressure changes, making results from different days or sites more directly comparable. We could compute the column of dry air using a surface pressure measurement and a water vapour measurement:

$$\mathrm{column\ dry\ air} = \frac{P_{\mathrm{s}}}{\{g\}_{\mathrm{air}} m_{\mathrm{air}}^{\mathrm{dry}}} - \mathrm{column}_{\mathrm{H_2O}} \frac{m_{\mathrm{H_2O}}}{m_{\mathrm{air}}^{\mathrm{dry}}}, \tag{A 7}$$

where $m_{\mathrm{H_2O}}$ and $m_{\mathrm{air}}^{\mathrm{dry}}$ are the mean molecular masses of water and dry air, respectively, P_{s} is the surface pressure and $\{g\}_{\mathrm{air}}$ is the column-averaged gravitational acceleration.

An alternative method of measuring the column of dry air is to measure the column of O_2 from the FTS spectrum, and divide by the known DMF of O_2 (0.2095):

$$\mathrm{column\ dry\ air} = \frac{\mathrm{column_{O_2}}}{0.2095}. \tag{A 8}$$

This approach has several advantages. Errors that are common to the target gas (e.g. CO_2) and O_2 bands (such as mis-pointing or zero-level offsets) will generally cancel in the column ratio, as shown in appendix B. Another advantage is that the water column is not required to correct the surface pressure. The O_2 column can also be used as a network-wide internal standard, in lieu of standard gas mixtures that are used for the global *in situ* network, since, to the degree required here, the mixing ratio of O_2 is constant and well known. Furthermore, Yang *et al.* [12] and Washenfelder *et al.* [24] analysed Kitt Peak FTS data using this technique for CO_2 and CH_4, respectively, and attained precisions of about 0.5 per cent. Hence, for TCCON, the column of dry air is retrieved from the O_2 column, and DMFs are computed using

$$X_{\mathrm{G}} = 0.2095 \frac{\mathrm{column_G}}{\mathrm{column_{O_2}}}. \tag{A 9}$$

There are several electronic transitions of O_2 in the NIR that can be observed and analysed; the TCCON uses the band centred at 7882 $\mathrm{cm^{-1}}$. Absorption in the O_2 A-band ($\mathrm{b^1\Sigma_g^+ - X^3\Sigma_g^-}(0,0)$ centred at 0.76 µm) is possible at sites with both Si and InGaAs detectors [56]. However, the A-band O_2 absorption lines are much stronger than the target gas absorption lines, which can lead to systematic errors. Although atmospheric 'dayglow' emission compromises the use of the weaker 7882 $\mathrm{cm^{-1}}$ (1.27 µm O_2) band ($\mathrm{a^1\Delta_g - X^3\Sigma_g^-}(0,0)$) for observing reflected sunlight from space, direct solar beam measurements using this band are possible from the ground because (i) the direct solar brightness greatly exceeds the dayglow, and (ii) the spectrally narrow mesospheric dayglow emission is absorbed by the stratospheric and tropospheric O_2.

Table 2. The TCCON error budget, list of included sources of error and their magnitudes. These data are listed below in percentages and given in graphical form in appendix B.

error source		magnitude of perturbation	20° SZA		70° SZA	
			CO_2/O_2	CO_2/P_s	CO_2/O_2	CO_2/P_s
a priori	temperature	+1 K	0.056	0.047	0.060	0.072
	pressure	−0.1%	−0.036	−0.018	−0.033	0.0075
	water vapour	−5%	−0.0039	−0.0035	−0.0090	−0.00016
	VMR shape	constant	0.062	0.062	0.058	0.058
tracker	observer–Sun Doppler shift	+2 ppm	0.052	0.082	−0.0028	0.021
	Sun tracker pointing offset	−0.05°	−0.00068	0.031	−0.00059	0.23
ILS	internal field of view	+7%	0.0037	0.0078	0.00050	0.0019
	angular misalignment	0.95 ME at MOPD	0.053	0.058	0.016	0.0099
	shear misalignment	$\alpha = -1.5$, $\beta = 0.5$	−0.098	−0.044	−0.020	0.076
	zero-level offset	+0.1%	−0.027	0.15	−0.031	0.18
	surface pressure	+1 hPa	0.042	−0.083	0.039	−0.11
	random noise	500 : 1 (CO_2); 200 : 1 (O_2)	0.036	0.017	0.033	0.012
	continuum curvature	+0.125%	−0.078	−0.070	−0.064	−0.056
total			0.18	0.23	0.13	0.34

There are several other advantages to using the O_2 column in the 7882 cm^{-1} band to compute the column of dry air. Its closer proximity to the CO_2, CH_4, N_2O, CO and HF absorption bands in the 4000–6400 cm^{-1} region means that the effect of the ILS errors will partially cancel. Using this band also minimizes the effect of errors in retrieving the airmass for a particular spectrum (most commonly due to solar tracker pointing errors or pressure measurement errors), because this will affect O_2 and the other gases in the same way (see appendix B and table 2). Furthermore, because the 1.27 µm band is observed with the same detector (InGaAs) as CO_2, CH_4 and the other gases, all TCCON sites can observe this band.

(e) *Airmass correction*

All TCCON X_{CO_2} data have an SZA or airmass-dependent artefact that causes the retrievals to be approximately 1 per cent larger at 20° SZA than at 80° SZA. The airmass dependence arises from the variation in both retrieved O_2 and CO_2. Airmass-dependent changes in X_{CO_2} occur at clean air sites (Park Falls, Darwin and Lauder) and at times of the year when there is no vegetation growth and therefore no cause for a real diurnal change of X_{CO_2}. If not corrected, the artefact would be aliased into the seasonal cycle (because the mean SZA varies seasonally),

which would lead to spurious differences between sites (because the mean SZA varies with latitude). The impact of the problem is particularly obvious in the Southern Hemisphere where the seasonal cycle is small.

There are several phenomena that can give rise to an airmass-dependent artefact. These include: spectroscopic inadequacies (e.g. widths, neglect of line mixing, inconsistencies in the relative strengths of weak and strong lines) and instrumental problems (zero-level offsets, continuum curvature, ILS uncertainties). While our preference would be to fix the spurious airmass dependence by finding physics-based solutions to each of the contributing problems, we must wait on the (promising) current effort in spectroscopic research in order to do this rigorously. In the mean time, we derive and apply a single empirical correction that represents the aggregate of all the aforementioned causes.

To develop our empirical correction, we assume that, on a given day, any variation in X_{CO_2} that is symmetrical about noon is an artefact and any variation that is antisymmetric is real. This is probably a reasonable assumption for unpolluted sites since the net effect of photosynthesis and respiration is to cause a maximum in X_{CO_2} at sunrise and a minimum at sunset. At polluted sites, however, it is necessary to first subtract the polluted CO_2 component using its correlation with CO.

(i) *Formulation*

For each day, we represent the measured X_{CO_2} by three terms: (i) a constant value ($\hat{y}$) representing the mean value of X_{CO_2} on that day, (ii) a component that varies antisymmetrically about solar noon ($\alpha A(t)$), which represents a simple model of the true variation of X_{CO_2} (i.e. the integrated drawdown of CO_2 during sunlight hours), and (iii) a component that varies symmetrically about solar noon ($\beta S(\theta)$), which represents the airmass-dependent artefact. The amplitudes of these three terms are determined by minimizing the difference between the measured column abundances (y_i) and the fitted functions $\hat{y}$, α and β:

$$y_i = \hat{y}[1 + \alpha A(t_i) + \beta S(\theta_i)] \tag{A 10}$$

The optimal functional forms of the basis functions $A(t_i)$ and $S(\theta_i)$ were determined using the X_{CO_2} results from Darwin, which are the most plentiful and cover the widest range of SZA:

$$A(t_i) = \sin(2\pi(t_i - t_{\text{noon}})) \tag{A 11}$$

and

$$S(\theta_i) = \left(\frac{\theta_i + \theta_0}{90 + \theta_0}\right)^3 - \left(\frac{45 + \theta_0}{90 + \theta_0}\right)^3, \tag{A 12}$$

where t_i and t_{noon} are in units of days and θ_i is in degrees. The optimization was performed manually, guided by the overall χ^2 of the minimization (for all days). The cubic exponent and the θ_0 value of 13° were found to minimize the root-mean-square deviation to the measured X_{CO_2} data. The choice to set $S(\theta)$ to zero at $\theta_i = 45°$ is arbitrary, and it does not affect the overall goodness of fit (χ^2) or the final X_{CO_2} values (i.e. after the *in situ* correction described in §3*c*). Choosing a value in the middle of the range makes the problem better posed

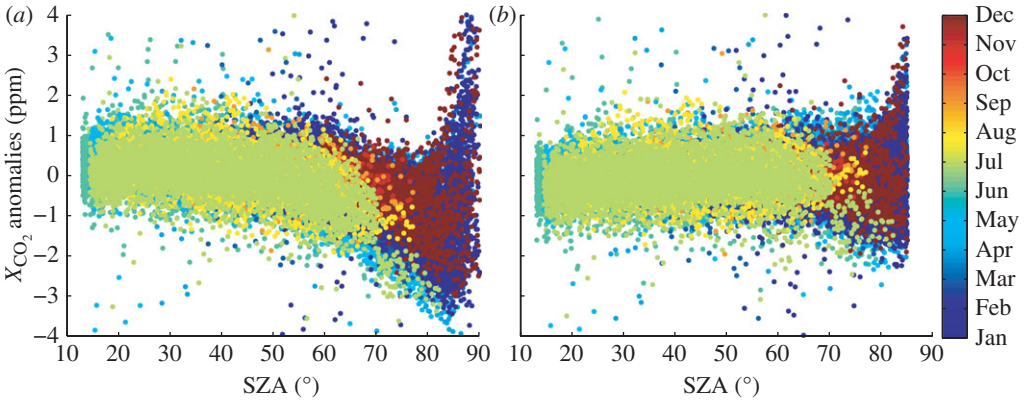

Figure 6. (*a*) The uncorrected Lamont X_{CO_2} data as a function of solar zenith angle (degrees) during the time period from July 2008 to July 2009. (*b*) The same data, after the airmass-dependent correction has been applied. The colours in both panels represent different months of the year. The data are expressed as daily anomalies: the differences between the X_{CO_2} values and the daily median values, which serve to remove the seasonal cycle. No airmass correction was performed on the SZA $> 85°$ data. Each point represents a single spectrum of 75 s observation.

mathematically, because it causes the symmetrical basis function S to become more orthogonal to $\hat{y}$. No airmass correction is performed for $\theta > 85°$, because the large errors at those airmasses are not sufficiently reproducible from site to site. These large airmass data are not currently released as part of the standard TCCON data product.

These basis functions were then applied to Park Falls, Lauder, Lamont and Darwin TCCON data, and a single set of coefficients for all sites (α, β) was computed. These coefficients will change (and hopefully diminish in magnitude) as improved spectroscopic line lists become available. The X_{CO_2} values as a function of SZA for the Lamont dataset are shown in figure 6, before (figure 6*a*) and after (figure 6*b*) the airmass correction.

The symmetric variations βS are highly consistent from day to day, season to season and site to site, verifying that this simple model is not removing real changes in X_{CO_2}. The antisymmetric variations αA are more variable, depending on the amount of photosynthesis on a given day and the proximity of X_{CO_2} gradients. They are largest over land in the Northern Hemisphere summer.

Appendix B. TCCON error budget

This section presents a sensitivity study, whereby each source of error is perturbed by a realistic amount (listed in table 2) in the GFIT forward model, and the fractional difference for each error source, relative to the unperturbed case, is computed. These sensitivities are computed for spectra from a clear day in Lamont, over a large range of SZAs. Fractional errors for each source are computed using both the O_2 column (equation (A 8)) and the surface pressure P_s (equation (A 7)) to compute the total column of dry air. The total error is the sum, in quadrature, of each individual error (figure 7 and table 2).

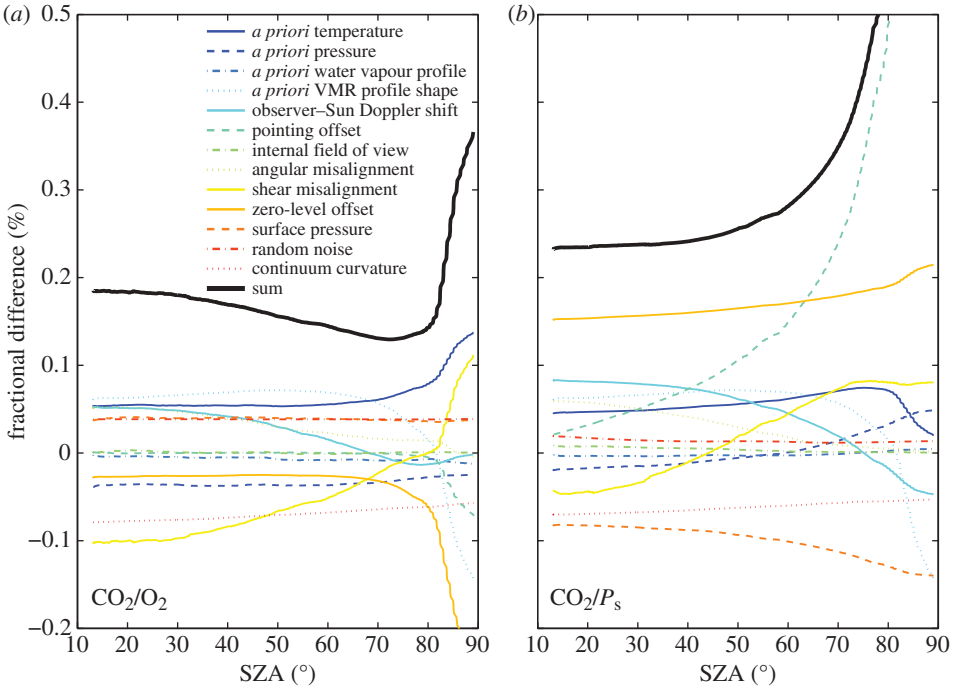

Figure 7. TCCON error budget from a clear day of spectra recorded in Lamont. (*a*) The results of the sensitivity tests for the standard TCCON X_{CO_2} retrieval using column O_2 (equation (A 9)); (*b*) the results of the sensitivity tests for X_{CO_2} computed using the surface pressure to infer the total column of dry air (equation (A 7)). Table 2 contains the definitions of the sensitivity tests and their magnitudes.

GFIT retrievals require good *a priori* knowledge of the atmospheric state (pressure, temperature and water vapour), and the overall shape of the vertical profile of the gases of interest. The *a priori* profile errors for pressure and water vapour are perturbed by a multiplicative change at each altitude. An additive perturbation is applied at all altitudes for temperature. The CO_2 *a priori* profile shape error is tested by replacing our empirically derived CO_2 *a priori* profile with one that is constant with altitude ('constant').

If the TCCON instruments have significant ILS errors, the resulting spectral fitting residuals will contain systematic artefacts (unrelated to any caused by spectroscopic errors), which may cause a miscalculation of the VSF for the retrieval. The ILS errors tested here include angular and shear misalignments, and FOV uncertainties. Angular misalignments cause a decrease in modulation efficiency as a function of OPD. A moderately well-aligned FTS may have a 5 per cent loss in modulation efficiency at maximum OPD. The internal FOV uncertainty affects the rectangular part of the ILS (equation (A 4), appendix A(*c*)(ii)), and estimates the error in the size and position of the field stops in the instrument.

Shear misalignments, first described by Murty [58] and later by Kauppinen & Saarinen [59], are caused by an OPD-independent displacement in the scanning cube corner relative to the optical axis. In Bruker 125HR instruments, this

displacement can occur from wear of the Teflon pads that support the scanning cube corner. This shear misalignment causes a decrease in modulation efficiency near ZPD relative to maximum OPD. Normalizing at ZPD, the modulation efficiency (ME) is modelled here by the approximation

$$\mathrm{ME}(x) = 1 + \alpha(1 - \mathrm{e}^{-\beta|x|}) \tag{A 13}$$

where α and β are free parameters and x is the OPD. Equation (A 13) results in $\mathrm{ME} = 1$ at ZPD and $\mathrm{ME} \to 1 + \alpha$ as $x \to \infty$. For this sensitivity study, the free parameters were chosen to represent a badly shear-misaligned Lamont spectrum.

Pointing errors can affect the retrieved columns in two ways: by causing an error in the estimated airmass, biasing the retrieved columns, and by causing the solar lines to be Doppler-shifted, causing systematic residuals to appear in the spectral fits. These two effects are tested by perturbing the pointing offset and observer–Sun Doppler shifts, respectively, for a pointing error of about one-fifth the radius of the Sun.

The other important errors considered are zero-level offsets, surface pressure measurement bias, random noise in the spectra and spectral continuum curvature. Zero-level offsets in the spectra create errors in the retrieved column because the depths of lines are incorrect (i.e. saturated lines do not have a transmission of zero at the line centre, and hence the depths of all lines are incorrectly calculated in the forward model). These offsets occur due to detector or electronics nonlinearities or an ILS shear error, and can be intensity dependent.

Errors in the surface pressure measured by the weather stations will cause an error in the line broadening of CO_2 and O_2, and will significantly affect the column of dry air when computed using the surface pressure (i.e. via equation (A 7)). Surface pressure measurement errors arise from a faulty or inaccurate pressure sensor.

The effects of random noise on X_{CO_2} are small and benign. They are small because the signal levels are high (as a result of the direct-sun viewing geometry), and the spectral absorption lines are strong for most of the target gases. Random noise has a benign effect on the retrievals because it causes no bias. For this error sensitivity test, noise consistent with the known SNRs of the spectra is added.

GFIT only fits a linear continuum level across a window, which is a good assumption for narrow spectral windows, but less valid for wider windows (i.e. for CO_2). If there is significant continuum curvature, which results from changes in the solar intensity and in detector or optical coating efficiency as a function of wavenumber within a given window, GFIT will incorrectly fit the continuum, and miscalculate the depth of the absorption lines. There is no obvious evidence of continuum curvature in the spectral fits, but a continuum curvature error of 0.125 per cent or less would not be detectable outside the noise, so this value was adopted for the sensitivity analysis.

The largest errors for the TCCON, which calculates X_{CO_2} using the O_2 column via equation (A 9), are shear misalignments, continuum curvature, *a priori* uncertainties and angular misalignments. Zero-level offsets make a significant contribution at high SZA. The total error (all sources added in quadrature) is about 0.2 per cent or less for $\mathrm{SZA} \leq 83°$.

The largest errors for X_{CO_2} computed using the surface pressure via equation (A 7) are zero-level offsets, surface pressure errors and Sun tracker pointing errors. These particular errors are larger in this case than in the CO_2/O_2

case, because they do not cancel between the CO_2 bands and the surface pressure measurement. The total error for this case is approximately 0.23 per cent at 20° SZA and approximately 0.34 per cent at 70° SZA. The SZA dependences of the errors are shown in figure 7. At all zenith angles, the CO_2/O_2 produces a smaller total error than CO_2/P_s.

Errors caused by 'ghosts' in the spectra have not been included in this error budget. If the NIR signal is sampled with a periodic error in OPD, owing to incorrectly sampling the metrology laser, antisymmetric 'parasitic' spectra are produced on each side of the desired ('parent') spectrum, offset by the sampling frequency (laser frequency). The parasitic spectra are aliased back into the region of interest by the Fourier transform. These aliased images are called 'ghosts' and were first described mathematically by Guelachvili [60] and their effect on the spectra was described later by Learner *et al.* [61]. As an example, for our TCCON FTS instruments, where the laser wavenumber is approximately 15 798 cm^{-1} and the sampling is twice per fringe, a parent line located at 8000 cm^{-1} would have ghosts at −7798 and 23 798 cm^{-1} that are both aliased back to 7798 cm^{-1}, interfering with the main O_2 spectral region. Ghosts have recently been found in some of the Bruker 125HR instruments [44] and can cause errors in X_{CO_2} of up to approximately 1 ppm in some cases. These ghosts can be eliminated with careful adjustment, using the procedure described by Messerschmidt *et al.* [44]. Bruker has developed and distributed a hardware solution to eliminate the ghosts for the affected instruments and we are currently developing an algorithm to correct older spectra and eliminate the effects of the laser mis-sampling.

References

1 GLOBALVIEW-CO_2. 2006 *Cooperative Atmospheric Data Integration Project—carbon dioxide* (CD-ROM). Boulder, CO: NOAA GMD.

2 Denning, A. S., Randall, D. A., Collatz, G. J. & Sellers, P. J. 1996 Simulations of terrestrial carbon metabolism and atmospheric CO_2 in a general circulation model. II. Simulated CO_2 concentrations. *Tellus B Chem. Phys. Meteorol.* **48**, 543–567. (doi:10.1034/j.1600-0889.1996.t01-1-00010.x)

3 Gurney, K. R. *et al.* 2004 TransCom 3 inversion intercomparison: model mean results for the estimation of seasonal carbon sources and sinks. *Global Biogeochem. Cycl.* **18**, GB1010. (doi:10.1029/2003GB002111)

4 Baker, D. F. *et al.* 2006 TransCom 3 inversion intercomparison: impact of transport model errors on the interannual variability of regional CO_2 fluxes, 1988–2003. *Global Biogeochem. Cycl.* **20**, GB1002. (doi:10.1029/2004GB002439)

5 Stephens, B. B. *et al.* 2007 Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO_2. *Science* **316**, 1732–1735. (doi:10.1126/science.1137004)

6 Yang, Z., Washenfelder, R. A., Keppel-Aleks, G., Krakauer, N. Y., Randerson, J. T., Tans, P. P., Sweeney, C. & Wennberg, P. O. 2007 New constraints on Northern Hemisphere growing season net flux. *Geophys. Res. Lett.* **34**, L12807. (doi:10.1029/2007GL029742)

7 Rayner, P. J. & O'Brien, D. M. 2001 The utility of remotely sensed CO_2 concentration data in surface source inversions. *Geophys. Res. Lett.* **28**, 175–178. (doi:10.1029/2000GL011912)

8 Olsen, S. C. & Randerson, J. T. 2004 Differences between surface and column atmospheric CO_2 and implications for carbon cycle research. *J. Geophys. Res.* **109**, D02301. (doi:10.1029/2003JD003968)

9 Peters, W. *et al.* 2007 An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *Proc. Natl Acad. Sci.* USA **104**, 18 925–18 930. (doi:10.1073/pnas.0708986104)

10 Dobson, G. M. B. 1931 A photoelectric spectrophotometer for measuring the amount of atmospheric ozone. *Proc. Phys. Soc.* **43**, 324–339. (doi:10.1088/0959-5309/43/3/308)

11 Kurylo, M. J. & Solomon, S. 1990 Network for the Detection of Stratospheric Change: a status and implementation report. Technical report, NASA, Upper Atmosphere Research Program and NOAA Climate and Global Change Program, Washington, DC.

12 Yang, Z., Toon, G. C., Margolis, J. S. & Wennberg, P. O. 2002 Atmospheric CO_2 retrieved from ground-based near IR solar spectra. *Geophys. Res. Lett.* **29**, 531. (doi:10.1029/2001GL014537)

13 Wallace, L. & Livingston, W. 1990 Spectroscopic observations of atmospheric trace gases over Kitt Peak. I. Carbon dioxide and methane from 1979 to 1985. *J. Geophys. Res.* **95**, 9823–9827. (doi:10.1029/JD095iD07p09823)

14 Dufour, E., Bréon, F. & Peylin, P. 2004 CO_2 column averaged mixing ratio from inversion of ground-based solar spectra. *J. Geophys. Res.* **109**, D09304. (doi:10.1029/2003JD004469)

15 Miller, C. E. *et al.* 2007 Precision requirements for space-based X_{CO_2} data. *J. Geophys. Res. D Atmos.* **112**, D10314. (doi:10.1029/2006JD007659)

16 Keppel-Aleks, G., Toon, G. C., Wennberg, P. O. & Deutscher, N. M. 2007 Reducing the impact of source brightness fluctuations on spectra obtained by Fourier-transform spectrometry. *Appl. Opt.* **46**, 4774–4779. (doi:10.1364/AO.46.004774)

17 Washenfelder, R. A., Toon, G. C., Blavier, J. F., Yang, Z., Allen, N. T., Wennberg, P. O., Vay, S. A., Matross, D. M. & Daube, B. C. 2006 Carbon dioxide column abundances at the Wisconsin Tall Tower site. *J. Geophys. Res.* **111**, D22305. (doi:10.1029/2006JD007154)

18 Deutscher, N. M. *et al.* 2010 Total column CO_2 measurements at Darwin, Australia—site description and calibration against *in situ* aircraft profiles. *Atmos. Meas. Tech.* **3**, 947–958. (doi:10.5194/amt-3-947-2010)

19 Toon, G. C. The JPL MkIV interferometer. *Opt. Photon. News* **2**, 19–21. (doi:10.1364/OPN.2.10.000019)

20 Bernath, P. F. *et al.* 2005 Atmospheric Chemistry Experiment (ACE): mission overview. *Geophys. Res. Lett.* **32**, L15S01. (doi:10.1029/2005GL022386)

21 Wunch, D. *et al.* 2010 Calibration of the Total Carbon Column Observing Network using aircraft profile data. *Atmos. Meas. Tech.* **3**, 1351–1362. (doi:10.5194/amt-3-1351-2010)

22 Macatangay, R., Warneke, T., Gerbig, C., Körner, S., Ahmadov, R., Heimann, M. & Notholt, J. 2008 A framework for comparing remotely sensed and *in-situ* CO_2 concentrations. *Atmos. Chem. Phys.* **8**, 2555–2568. (doi:10.5194/acp-8-2555-2008)

23 Messerschmidt, J. *et al.* 2010 IMECC campaign—the first European aircraft calibration campaign of Fourier transform infrared (FTIR) spectrometer sites. In *Geophysical Research Abstracts*, no. 12, EGU2010-2920-1. EGU General Assembly 2010.

24 Washenfelder, R. A., Wennberg, P. O. & Toon, G. C. 2003 Tropospheric methane retrieved from ground-based near-IR solar absorption spectra. *Geophys. Res. Lett.* **30**, 2226. (doi:10.1029/2003GL017969)

25 Rodgers, C. D. & Connor, B. J. 2003 Intercomparison of remote sounding instruments. *J. Geophys. Res.* **108**, 4116–4229. (doi:10.1029/2002JD002299)

26 Rodgers, C. D. 2000 *Inverse methods for atmospheric sounding: theory and practice*. Singapore: World Scientific. (doi:10.1142/9789812813718)

27 Migliorini, S., Piccolo, C. & Rodgers, C. D. 2008 Use of the information content in satellite measurements for an efficient interface to data assimilation. *Mon. Weather Rev.* **136**, 2633–2650. (doi:10.1175/2007MWR2236.1)

28 Warneke, T., Yang, Z., Olsen, S., Körner, S., Notholt, J., Toon, G. C., Velazco, V., Schulz, A. & Schrems, O. 2005 Seasonal and latitudinal variations of column averaged volume-mixing ratios of atmospheric CO_2. *Geophys. Res. Lett.* **32**, L03808. (doi:10.1029/2004GL021597)

29 Keppel-Aleks, G., Wennberg, P. O., Schneider, T., Honsowetz, N. Q. & Vay, S. A. 2008 Total column constraints on Northern Hemisphere carbon dioxide surface exchange. In *American Geophysical Union, Fall Meeting 2008*, abstract #A43F-03.

30 Keppel-Aleks, G., Wennberg, P. O. & Schneider, T. 2010 Sources of variations in total column carbon dioxide. *Atmos. Chem. Phys. Discuss.* **10**, 30 569–30 611. (doi:10.5194/acpd-10-30569-2010)

31 Wunch, D., Wennberg, P. O., Toon, G. C., Keppel-Aleks, G. & Yavin, Y. G. 2009 Emissions of greenhouse gases from a North American megacity. *Geophys. Res. Lett.* **36**, L15810. (doi:10.1029/2009GL039825)

32 Barkley, M. P. *et al.* 2007 Assessing the near surface sensitivity of SCIAMACHY atmospheric CO_2 retrieved using (FSI) WFM-DOAS. *Atmos. Chem. Phys.* **7**, 3597–3619. (doi:10.5194/acp-7-3597-2007)

33 Bösch, H. *et al.* 2006 Space-based near-infrared CO_2 measurements: testing the Orbiting Carbon Observatory retrieval algorithm and validation concept using SCIAMACHY observations over Park Falls, Wisconsin. *J. Geophys. Res.* **111**, D23302. (doi:10.1029/2006JD007080)

34 De Beek, R. *et al.* 2006 Atmospheric carbon gases retrieved from SCIAMACHY by WFM-DOAS: improved global CO and CH_4 and initial verification of CO_2 over Park Falls (46°N, 90°W). *Atmos. Chem. Phys. Discuss.* **6**, 363–399. (doi:10.5194/acpd-6-363-2006)

35 Buchwitz, M. *et al.* 2006 Atmospheric carbon gases retrieved from SCIAMACHY by WFM-DOAS: version 0.5. CO and CH_4 and impact of calibration improvements on CO_2 retrieval. *Atmos. Chem. Phys.* **6**, 2727–2751. (doi:10.5194/acp-6-2727-2006)

36 Reuter, M. *et al.* 2010 Retrieval of atmospheric CO_2 from SCIAMACHY nadir spectra considering scattering at thin ice clouds and aerosols. In *ESA Living Planet Symposium*, July 2010.

37 Reuter, M. *et al.* 2011 Retrieval of atmospheric CO_2 with enhanced accuracy and precision from SCIAMACHY: Validation with FTS measurements and comparison with model results. *J. Geophys. Res.* **116**, D04301. (doi:10.1029/2010JD015047)

38 Tran, H. & Hartmann, J. M. 2008 An improved O_2 A band absorption model and its consequences for retrievals of photon paths and surface pressures. *J. Geophys. Res.* **113**, D18104. (doi:10.1029/2008JD010011)

39 Hartmann, J.-M., Tran, H. & Toon, G. C. 2009 Influence of line mixing on the retrievals of atmospheric CO_2 from spectra in the 1.6 and 2.1 μm regions. *Atmos. Chem. Phys.* **9**, 7303–7312. (doi:10.5194/acp-9-7303-2009)

40 Tran, H., Hartmann, J.-M., Toon, G., Brown, L. R., Frankenberg, C., Warneke, T., Spietz, P. & Hase, F. 2010 The $2\nu_3$ band of CH_4 revisited with line mixing: consequences for spectroscopy and atmospheric retrievals at 1.67 μm. *J. Quant. Spectrosc. Radiat. Transf.* **111**, 1344–1356. (doi:10.1016/j.jqsrt.2010.02.015)

41 Gordon, I. E., Kassi, S., Campargue, A. & Toon, G. C. 2010 First identification of the $a^1\Delta_g - X^3\Sigma_g^-$ electric quadrupole transitions of oxygen in solar and laboratory spectra. *J. Quant. Spectrosc. Radiat. Transf.* **111**, 1174–1183. (doi:10.1016/j.jqsrt.2010.01.008)

42 Hase, F. & Delbouille, L. 2007 Retrieval of trace gas profiles from ground-based NDSC FTIR spectra: estimating the optimal spectral resolution. In *Network for the detection of stratospheric change—Infrared Working Group Meeting, Toronto.*

43 Brault, J. W. 1996 New approach to high-precision Fourier transform spectrometer design. *Appl. Opt.* **35**, 2891–2896. (doi:10.1364/AO.35.002891)

44 Messerschmidt, J., Macatangay, R., Notholt, J., Warneke, T. & Weinzierl, C. 2010 Side by side measurements of column-averaged CO_2 by ground-based Fourier transform spectrometry (FTS). *Tellus B* **62**, 749–758. (doi:10.1111/j.1600-0889.2010.00491.x)

45 Forman, M. L., Steel, W. H. & Vanasse, G. A. 1966 Correction of asymmetric interferograms obtained in Fourier spectroscopy. *J. Opt. Soc. Am.* **56**, 59–61. (doi:10.1364/JOSA.56.000059)

46 Mertz, L. 1965 *Transformations in optics.* New York, NY: Wiley.

47 Mertz, L. 1967 Auxiliary computation for Fourier spectrometry. *Infrared Phys.* **7**, 17–23. (doi:10.1016/0020-0891(67)90026-7)

48 Bergland, G. 1969 A radix-eight fast Fourier transform subroutine for real-valued series. *IEEE Trans. Audio Electroacoust.* **17**, 138–144. (doi:10.1109/TAU.1969.1162043)

49 Connes, J. 1961 Recherches sur la spectroscopie par transformation de Fourier. *Rev. Opt. Théor. Instrum.* **40**, 45–79, 116–140, 171–190, 231–265. [Transl. into English by C. A. Flanagan, Navweps Rep. 8099, NOT S TP 3157].

50 Davis, S. P., Abrams, M. C. & Brault, J. W. 2001 *Fourier transform spectrometry.* New York, NY: Academic Press.

51 Kalnay, E. C. *et al.* 1996 The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471. (doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)

52 Luo, M., Cicerone, R. J. & Russell, J. M. 1995 Analysis of halogen occultation experiment HF versus CH_4 correlation plots: chemistry and transport implications. *J. Geophys. Res.* **100**, 13 927–13 937. (doi:10.1029/95JD00621)

53 Andrews, A. E. *et al.* 2001 Mean ages of stratospheric air derived from in situ observations of CO_2, CH_4, and N_2O. *J. Geophys. Res. Atmos.* **106**, 32 295–32 314. (doi:10.1029/2001JD000465)
54 Rothman, L. S. *et al.* 2009 The HITRAN 2008 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **110**, 533–572. (doi:10.1016/j.jqsrt.2009.02.013)
55 Newman, S. M., Lane, I. C., Orr-Ewing, A. J., Newnham, D. A. & Ballard, J. 1999 Integrated absorption intensity and Einstein coefficients for the O_2 $a^1\Delta_g - X^3\Sigma_g^-(0,0)$ transition: a comparison of cavity ringdown and high resolution Fourier transform spectroscopy with a long-path absorption cell. *J. Chem. Phys.* **110**, 10 749–10 757. (doi:10.1063/1.479018)
56 Yang, Z., Wennberg, P. O., Cageao, R. P., Pongetti, T. J., Toon, G. C. & Sander, S. P. 2005 Ground-based photon path measurements from solar absorption spectra of the O_2 A-band. *J. Quant. Spectrosc. Radiat. Transf.* **90**, 309–321. (doi:10.1016/j.jqsrt.2004.03.020)
57 Smith, K. M. & Newnham, D. 2000 Near-infrared absorption cross sections and integrated absorption intensities of molecular oxygen (O_2, O_2–O_2, and O_2–N_2). *J. Geophys. Res.* **105**, 7383–7396. (doi:10.1029/1999JD901171)
58 Murty, M. 1960 Some more aspects of the Michelson interferometer with cube corners. *J. Opt. Soc. Am.* **50**, 7–9. (doi:10.1364/JOSA.50.000007)
59 Kauppinen, J. & Saarinen, P. 1992 Line-shape distortions in misaligned cube corner interferometers. *Appl. Opt.* **31**, 69–74. (doi:10.1364/AO.31.000069)
60 Guelachvili, G. 1981 *Spectrometric techniques II. Distortions in Fourier spectra and diagnosis.* New York, NY: Academic Press.
61 Learner, R. C. M., Thorne, A. P. & Brault, J. W. 1996 Ghosts and artifacts in Fourier-transform spectrometry. *Appl. Opt.* **35**, 2947–2954. (doi:10.1364/AO.35.002947)