



Consistent evaluation of ACOS-GOSAT, BESD-SCIAMACHY, CarbonTracker, and MACC through comparisons to TCCON

Susan Kulawik¹, Debra Wunch², Christopher O'Dell³, Christian Frankenberg^{2,4}, Maximilian Reuter⁵, Tomohiro Oda^{6,7}, Frederic Chevallier⁸, Vanessa Sherlock^{9,8}, Michael Buchwitz⁵, Greg Osterman⁴, Charles E. Miller⁴, Paul O. Wennberg², David Griffith¹⁰, Isamu Morino¹¹, Manvendra K. Dubey¹², Nicholas M. Deutscher^{5,10}, Justus Notholt⁵, Frank Hase¹³, Thorsten Warneke⁵, Ralf Sussmann¹³, John Robinson⁹, Kimberly Strong¹⁴, Matthias Schneider¹², Martine De Mazière¹⁵, Kei Shiomi¹⁶, Dietrich G. Feist¹⁷, Laura T. Iraci¹⁸, and Joyce Wolf^{4,*}

¹Bay Area Environm. Res. Inst., Sonoma, CA 95476, USA

²California Institute of Technology, Pasadena, CA, USA

³Colorado State University, Fort Collins, CO, USA

⁴Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

⁵University of Bremen, Institute of Environmental Physics, Bremen, Germany

⁶Goddard Earth Sciences Technology and Research, Universities Space Research Association, Columbia, MD, USA

⁷Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA

⁸Laboratoire de Meteorologie Dynamique, Palaiseau, France

⁹The National Institute of Water and Atmospheric Research, Wellington and Lauder, New Zealand

¹⁰University of Wollongong, Wollongong, New South Wales, Australia

¹¹Center for Global Environmental Research, National Institute for Environmental Studies (NIES), Tsukuba, Ibaraki, Japan

¹²Earth and Environmental Science, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

¹³Karlsruhe Institute of Technology, Institute for Meteorology and Climate Research (IMK-ASF), Karlsruhe, Germany

¹⁴Department of Physics, University of Toronto, Toronto, Ontario, Canada

¹⁵Royal Belgian Institute for Space Aeronomy, Brussels, Belgium

¹⁶Earth Observation Research Center, Japan Aerospace Exploration Agency, Tsukuba, Ibaraki, Japan

¹⁷Max Planck Institute for Biogeochemistry, Jena, Germany

¹⁸NASA Ames Research Center, Atmospheric Science Branch, Moffett Field, CA, USA

*retired

Correspondence to: Susan Kulawik (susan.s.kulawik@nasa.gov)

Received: 19 May 2015 – Published in Atmos. Meas. Tech. Discuss.: 22 June 2015

Revised: 24 January 2016 – Accepted: 29 January 2016 – Published: 29 February 2016

Abstract. Consistent validation of satellite CO₂ estimates is a prerequisite for using multiple satellite CO₂ measurements for joint flux inversion, and for establishing an accurate long-term atmospheric CO₂ data record. Harmonizing satellite CO₂ measurements is particularly important since the differences in instruments, observing geometries, sampling strategies, etc. imbue different measurement characteristics in the various satellite CO₂ data products. We focus on validating model and satellite observation attributes that impact flux estimates and CO₂ assimilation, including accurate error estimates, correlated and random errors, overall

biases, biases by season and latitude, the impact of coincidence criteria, validation of seasonal cycle phase and amplitude, yearly growth, and daily variability. We evaluate dry-air mole fraction (X_{CO₂}) for Greenhouse gases Observing SATellite (GOSAT) (Atmospheric CO₂ Observations from Space, ACOS b3.5) and SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) (Bremen Optimal Estimation DOAS, BESD v2.00.08) as well as the CarbonTracker (CT2013b) simulated CO₂ mole fraction fields and the Monitoring Atmospheric Composition and Climate (MACC) CO₂ inversion system (v13.1) and

compare these to Total Carbon Column Observing Network (TCCON) observations (GGG2012/2014). We find standard deviations of 0.9, 0.9, 1.7, and 2.1 ppm vs. TCCON for CT2013b, MACC, GOSAT, and SCIAMACHY, respectively, with the single observation errors 1.9 and 0.9 times the predicted errors for GOSAT and SCIAMACHY, respectively. We quantify how satellite error drops with data averaging by interpreting according to $\text{error}^2 = a^2 + b^2/n$ (with n being the number of observations averaged, a the systematic (correlated) errors, and b the random (uncorrelated) errors). a and b are estimated by satellites, coincidence criteria, and hemisphere. Biases at individual stations have year-to-year variability of ~ 0.3 ppm, with biases larger than the TCCON-predicted bias uncertainty of 0.4 ppm at many stations. We find that GOSAT and CT2013b underpredict the seasonal cycle amplitude in the Northern Hemisphere (NH) between 46 and 53° N, MACC overpredicts between 26 and 37° N, and CT2013b underpredicts the seasonal cycle amplitude in the Southern Hemisphere (SH). The seasonal cycle phase indicates whether a data set or model lags another data set in time. We find that the GOSAT measurements improve the seasonal cycle phase substantially over the prior while SCIAMACHY measurements improve the phase significantly for just two of seven sites. The models reproduce the measured seasonal cycle phase well except for at Lauder_125HR (CT2013b) and Darwin (MACC). We compare the variability within 1 day between TCCON and models in JJA; there is correlation between 0.2 and 0.8 in the NH, with models showing 10–50 % the variability of TCCON at different stations and CT2013b showing more variability than MACC. This paper highlights findings that provide inputs to estimate flux errors in model assimilations, and places where models and satellites need further investigation, e.g., the SH for models and 45–67° N for GOSAT and CT2013b.

1 Introduction

Carbon–climate feedbacks are a major uncertainty in predicting the climate response to anthropogenic forcing (Friedlingstein et al., 2006). Currently, about 10 Gigatons (Gt) of carbon are emitted per year from human activity (e.g., fossil fuel burning, deforestation), of which about 5 Gt stays in the atmosphere, causing an annual CO₂ increase of approximately 2 ppm yr⁻¹. The yearly increase is quite variable, estimated at 1.99 ± 0.43 ppm yr⁻¹ (<http://www.esrl.noaa.gov/gmd/ccgg/trends/global.html>), however always positive (Houghton et al., 2007). The remaining 5 Gt of carbon is taken up by the ocean and the terrestrial biosphere; however, there are uncertainties in the location and mechanism of these sinks, e.g., the distribution of land sinks between the Northern Hemisphere and the tropics (e.g., Stephens et al., 2007), and the localization of sources and sinks on regional scales (Canadell et al., 2011; Baker et al., 2006). The un-

certainties in top-down source and sink estimates are a consequence of uncertainties in model transport and dynamics (e.g., Prather et al., 2008; Stephens et al., 2007) and sparseness of available surface-based CO₂ observations (Hungerhoefer et al., 2010; Chevallier et al., 2010). Satellites offer a much denser and spatially contiguous data set for top-down estimates, but are much more susceptible to biases as compared to ground-based measurements (e.g., see summary in Sect. 3.3.2 of Ciais et al., 2014).

This paper tests different characteristics of model and satellite CO₂ (e.g., seasonal cycle amplitude and phase, regional and seasonal biases, effects of averaging, and diurnal variations) through a series of specialized comparisons to the Total Carbon Column Observing Network (TCCON). The findings from this work can be propagated into assimilation systems to determine the influence of various findings on top-down flux estimates (e.g., see Miller et al., 2007; Deng et al., 2014; Chevallier and O'Dell, 2013; Chevallier et al., 2014). For example, this paper characterizes biases by latitude and season; these biases can be assimilated to determine their effects on flux estimates (e.g., Kulawik et al., 2013). This paper also shows a set of comparisons and tests that may be useful for evaluating bottom-up flux estimates or transport schemes in models.

The remainder of the paper is organized as follows: Sect. 2 describes the satellite X_{CO₂} data and models used. The various satellite and model X_{CO₂} data are compared against TCCON observations in Sect. 3. Temporal characteristics of the different X_{CO₂} data and models, including the seasonal cycle amplitude and phase, are compared in Sect. 4. Diurnal behavior is evaluated in Sect. 5. Section 6 summarizes our findings and discusses their impact on evaluating terrestrial carbon fluxes.

2 Data and models used

The characteristics of the sets of carbon dioxide that will be compared to TCCON are summarized in Table 1. The following sections contain detailed descriptions of the data set versions and characteristics.

2.1 The TCCON

The TCCON consists of ground-based Fourier transform spectrometers (FTSs) that measure high spectral (0.02 cm⁻¹) and temporal (~ 90 s) resolution spectra of the direct sun in the near infrared spectrum (Wunch et al., 2011a). Column abundances of CO₂, O₂, and other atmospheric gases are determined from their absorption signatures in the solar spectra using the GGG software package, which employs a nonlinear least squares spectral fitting algorithm to scale an a priori volume mixing ratio profile. Absorption of CO₂ is measured in the weak CO₂ band centered on 6220 and 6339 cm⁻¹, and of O₂ in the band centered on 7885 cm⁻¹.

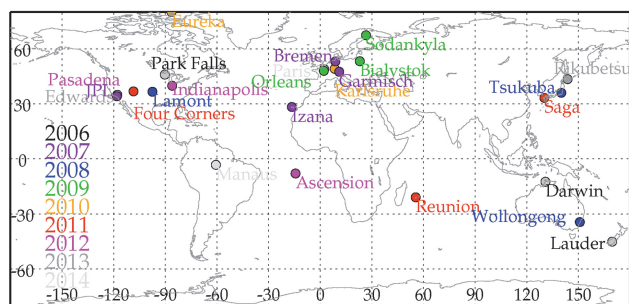


Figure 1. TCCON site locations used for this work. The color indicates the year when each station started collecting data.

The total column dry-air mole fractions of CO_2 (X_{CO_2}) are computed by ratioing the column abundances of CO_2 and O_2 . The resulting dry-air mole fractions have been calibrated against profiles of CO_2 measured by WMO-scale instrumentation aboard aircraft (Wunch et al., 2010; Messerschmidt et al., 2011). The precision and accuracy of the TCCON X_{CO_2} product is ~ 0.8 ppm (2σ) after calibration (Wunch et al., 2010). The TCCON data used in this paper are from the GGG2012 and GGG2014 releases, available from <http://tcccon.ornl.gov/>. In this paper we use 90 min average TCCON values which have 1σ precision of 0.4 ppm (see Appendix B).

We use 20 TCCON stations, distributed globally (see Fig. 1), and these data have been used extensively for satellite validation and bias correction (e.g., Butz et al., 2011; Morino et al., 2011; Wunch et al., 2011b; Reuter et al., 2011; Schneising et al., 2012; Oshchepkov et al., 2012), in flux inversions (Chevallier et al., 2011), and in model comparisons (Basu et al., 2011; Saito et al., 2012). We use the GGG2014 data when available, and the GGG2012 data from sites Four Corners and Tsukuba_120HR. Note that the overall bias between the GGG2012 and GGG2014 X_{CO_2} is 0.3 ppm (with the GGG2014 results smaller) (Wunch et al., 2015). The GGG2012 sites have corrections based on the instructions from the TCCON partners, listed on the TCCON website (https://tcccon-wiki.caltech.edu/Network_Policy/Data_Use_Policy/Data_Description_GGG2012#Laser_Sampling_Errors). We also apply a 0.9972 factor to Four Corners, as indicated here: https://tcccon-wiki.caltech.edu/Network_Policy/Data_Use_Policy/Data_Description_GGG2012. Two instruments have been operated at the Lauder site. We identify them using 120HR (for the period of 20 June 2004 through 28 February 2011) and 125HR (for 2 February 2010 through to the present) when results are instrument-specific. We find biases in the Tsukuba_125HR for the time ranges in this paper (possibly from high humidity combined with solar tracker issues during this time period that were later corrected, and/or bias updates that are needed) and do not use Tsukuba_125HR in this paper.

Stations which have special circumstances regarding validation and are considered locally influenced are Garmisch which is in the midst of complicated terrain, for which local atmospheric transport is difficult to model and to measure from space; Four Corners, which is located in the vicinity of two power plants with large CO_2 emissions (Lindenmaier et al., 2014). The meteorology is such that Four Corners regularly samples large localized plumes with column CO_2 increases of several ppm that last hours in the late morning. Therefore, the low bias in models and satellite data relative to the Four Corners TCCON is attributed to the smaller scale enhancements from the power plants measured in TCCON which are significantly diluted in the model and satellite results; Bremen is also affected by local urban sources, and satellites and models would be expected to be biased low, though it is similar to adjacent stations; and JPL, Pasadena, and Edwards are in or adjacent to a megacity with complex adjacent terrain. The Izaña TCCON station is on Tenerife, a small island (about 50×90 km) with complex topography located about 300 km west of southern Morocco. The TCCON station is located at 2.37 km above sea level (about 770 hPa), whereas models and satellite observations are of the surrounding ocean at about 1013 hPa. These stations are not included in averages (e.g., average bias, average seasonal cycle amplitude differences). In the future, either targeted observations or more spatially resolved models could make better use of these TCCON stations. Although we do not use these stations in averaging, we show results from all TCCON stations in this paper.

2.2 GOSAT CO_2

The Greenhouse gases Observing SATellite (GOSAT) takes measurements of reflected sunlight in three shortwave bands with a circular footprint of approximately 10.5 km diameter at nadir (Kuze et al., 2009; Yokota et al., 2009; Crisp et al., 2012). The first useable science measurements were made in April 2009, but due to changing observational modes in the early months, we use data beginning in July 2009. In this work, we use column-averaged dry-air mole fraction (X_{CO_2}) retrievals produced by NASA's Atmospheric CO_2 Observations from Space (ACOS) project, version 3.5 (see O'Dell et al. (2012) for retrieval details). For each sounding, the retrieval produces an estimate of X_{CO_2} , the vertical sensitivity of the measurement (i.e., the averaging kernel), and the posterior uncertainty in X_{CO_2} . It also produces a number of other retrieval variables, such as surface pressure and aerosols, which are used in both filtering and bias correction. Post-retrieval filter is employed based on a number of variables associated with the retrieval. In addition to filtering, a revised bias correction scheme has been developed for the v3.5 retrievals. This scheme is similar to the approach described in Wunch et al. (2011b), which characterized the errors in earlier versions of the ACOS retrieval using a simple spatial uniformity assumption of X_{CO_2} in the Southern Hemisphere

Table 1. Summary of the CO₂ data sets and models we used, showing the coverage for several different CO₂ products. The obs/day are the approximate number of CO₂ observations which passed quality screening.

Satellite/TCCON	Dates available	Day/ night	Land/ ocean	Latitude	Obs/ day	footprint
SCIAMACHY v2.00.08	Jan 2003–Apr 2012	day	land	80° S–80° N	~ 800	60 × 30 km
GOSAT ACOS-v3.5	Jun 2009–Apr 2014	day	both	80° S–80° N	~ 360	10.5 km cir.
TCCON GGG2012	see Fig. 1	day	both	45° S–80° N	–	–
Model	Dates available			Time res.	Spatial res. (lat × lon)	
CT2013b	2000–2012			3 h	2 × 3° global, 1 × 1° US	
MACC 13.1	1979–2013 (used 2007–2013)			3 h	1.9 × 3.75°	

(sometimes referred to as the Southern Hemisphere Approximation) to assess errors and biases in the retrievals. V3.5 has corrections of GOSAT high (*H*) and medium (*M*) gain data over land, as well as glint-mode data over the ocean, by using not only the Southern Hemisphere Approximation, but also TCCON observations, and comparisons to an ensemble mean of multiple transport model output. Details of the post-retrieval filter and the bias correction scheme can be found in the ACOS v3.5 user's guide which will soon be at <https://co2.jpl.nasa.gov/>.

2.3 SCIAMACHY CO₂

The following description of the SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) CO₂ retrieval algorithm summarizes important aspects of Reuter et al. (2010 and 2011) and is adopted in parts from the algorithm theoretical basis document (Reuter et al., 2012b).

The Bremen Optimal Estimation DOAS (BESD) algorithm is designed to analyze SCIAMACHY sun normalized radiance measurements to retrieve the column-averaged dry-air mole fraction of atmospheric carbon dioxide (X_{CO_2}). BESD is a so-called full physics algorithm, which uses a two-band retrieval, with the O₂ A absorption band used to retrieve scattering information of clouds and aerosols, while the 1580 nm band additionally contains CO₂ information. Similar to the ACOS three-band retrieval for GOSAT, the explicit consideration of scattering by this approach reduces potential systematic biases due to clouds or aerosols.

The retrieved 26-elements state vector consists of a second-order polynomial of the surface spectral albedo in both fit windows, two instrument parameters (spectral shift and slit functions full width at half maximum in both fit windows, described in Reuter et al., 2010), a temperature profile shift, a scaling of the H₂O profile and a default aerosol profile, cloud water/ice path, cloud top height, surface pressure, and a 10-layer CO₂ mixing ratio profile. Even though the number of state vector elements (26) is smaller than the number of measurement vector elements (134), the inversion problem is generally underdetermined, especially for the

CO₂ profile. For this reason BESD uses a priori knowledge as a side constraint. However, for most of the state vector elements the a priori knowledge gives only a weak constraint and therefore does not dominate the retrieval results. The degree of freedom for X_{CO_2} typically lies within an interval between 0.9 and 1.1.

A post-processor adjusts the retrieved X_{CO_2} to a priori CO₂ profiles generated with the simple empirical CO₂ model described by Reuter et al. (2012a). Additionally the post-processor performs quality filtering and bias correction. The bias correction is based on, e.g., convergence, fit residuals, and error reduction. The bias correction follows the idea of Wunch et al. (2011b) using TCCON as a reference to derive an empirical bias model depending on solar zenith angle, retrieved albedo, etc. The theoretical predicted errors have been scaled to agree with the errors vs. TCCON (Reuter et al., 2011). More details can be found in BESD's algorithm theoretical basis document (Reuter et al., 2012b).

2.4 CarbonTracker

CarbonTracker (CT) is an annually updated analysis of atmospheric carbon dioxide distributions and the surface fluxes that create them (Peters et al., 2007). CarbonTracker uses the Transport Model 5 (TM5) offline atmospheric tracer transport model (Krol et al., 2005) driven by meteorology from the European Centre for Medium-Range Weather Forecasts (ECMWF) operational forecast model and from the ERA-Interim reanalysis (Dee et al., 2011) to propagate surface emissions. TM5 runs at a global 3° × 2° resolution and at a 1° × 1° resolution over North America. CarbonTracker separately propagates signals from fossil fuel emissions, air–sea CO₂ exchange, and terrestrial fluxes from wildfire emissions and non-fire net ecosystem exchange. Similar to other existing CO₂ inverse models, oceanic and terrestrial biosphere surface fluxes are optimized to agree with atmospheric CO₂ observations, while fossil fuel and wildfire emissions are specified. First-guess fluxes from terrestrial biosphere models and surface ocean carbon analyses are modified by applying weekly multiplicative scaling factors estimated for 126 land and 30 ocean regions using an ensemble Kalman fil-

ter optimization method. The CT2013b release of CarbonTracker assimilates in situ observations between 2000 and 2012 from 103 data sets around the world, including time series from NOAA observatories, tall towers, and flasks sampled by the NOAA Cooperative Air Sampling Network, and flask and continuous measurements from partners including Environment Canada, the Australian Commonwealth Scientific and Industrial Research Organization, the National Center for Atmospheric Research, the Lawrence Berkeley National Laboratory, and the Brazilian Instituto de Pesquisas Energéticas e Nucleares.

In order to explicitly quantify the impact of transport uncertainty and prior flux model bias on inverse flux estimates from CarbonTracker, the CT2013b release is composed of a suite of inversions, each using a different combination of prior flux models and parent meteorological model. Sixteen independent inversions were conducted, using two terrestrial biosphere flux priors, two air–sea CO₂ exchange flux priors, two estimates of imposed fossil fuel emissions, and two transport estimates in a factorial design. CT2013b results are presented as the performance-weighted mean of the inversion suite, with uncertainties including a component of across-model differences. All CarbonTracker results and complete documentation can be accessed online at <http://carbontracker.noaa.gov>.

2.5 MACC

Monitoring Atmospheric Composition and Climate (MACC, <http://www.copernicus-atmosphere.eu/>) was the European Union-funded project responsible for the development of the pre-operational Copernicus atmosphere monitoring service. MACC monitors the global distributions of greenhouse gases, aerosols, and reactive gases, and estimates some of their sources and sinks. Since 2010, it has been delivering an analysis of the carbon dioxide in the atmosphere and of its surface fluxes every year, based on the assimilation of air sample mole fraction measurements (Chevallier et al., 2010). It relies on a variational inversion formulation, developed by LSCE, that estimates 8-day grid point daytime/nighttime CO₂ fluxes and the grid point total columns of CO₂ at the initial time step of the inversion window. The Bayesian error statistics of the estimate are computed by a robust randomization approach. The MACC inversion scheme relies on the global tracer transport model Laboratoire de Météorologie Dynamique (LMDZ; Hourdin et al., 2006), driven by the wind analyses from the ECMWF. For release v13.1 of the MACC inversion, used here, LMDZ was run at a horizontal resolution 3.75° longitude × 1.9° latitude with 39 vertical layers. The other elements of the inversion configuration follow Chevallier et al. (2011), with climatological (i.e., not interannually varying), terrestrial, and ocean prior fluxes and interannually varying fossil fuel and biomass burning emissions. The variational formulation of the inversion allowed the 1979–2013 period to be processed in a single assimilation

window, therefore ensuring the physical and statistical consistency of the inversion over the full 35-year period. Mole fraction records from 131 measurement sites have been used from the NOAA Earth System Research Laboratory archive, the World Data Centre for Greenhouse Gases archive, and the Réseau Atmosphérique de Mesure des Composés à Effet de Serre (RAMCES) database (see the list in the Supplement of Peylin et al., 2013).

3 Direct comparisons to TCCON

We show comparisons between satellite X_{CO₂}, model-simulated mole fraction fields, and TCCON X_{CO₂} at 25 different TCCON sites, shown in Fig. 1. These sites span the northern and southern hemispheres and cover a wide range of latitudes and longitudes. In our analysis we mainly use stations with start dates in or before 2012 which have coverage in all seasons so that seasonal cycle fits, which require 2 years of data, can be made. Newer sites are not used for SCIAMACHY comparisons, as this data set ends in mid-2012 and our version of CarbonTracker (CT2013b) ends at the end of 2013. We show representative time series for CT2013b, MACC, GOSAT, and SCIAMACHY for a Northern Hemisphere site (Lamont, OK, US at 37° N) and two Southern Hemisphere sites (Lauder, NZ at 45° S or Wollongong, New South Wales, Australia at 35° S) in Fig. 2. Lauder is shown for the models to show the phase lag for CT2013b seen at this site in Table 5. Since there are not enough coincidences for the satellites at Lauder, we show Wollongong for satellites. These plots show matches using the geometric coincidence criteria described in Table 2 below (for satellites) and give an idea of the number of coincidences for each data set using these criterion. As only TCCON/satellite-matched pairs are shown, different subsets of TCCON are included for models and the two satellites. These sites were chosen as they have the most coincidences in the northern and southern hemispheres, respectively, for satellites. All sets compare well; the 30-day moving averages show differences most easily, such as a repeating blip in CT2013b comparisons at the summer drawdown at Lamont and a seasonal mismatch in CT2013b comparisons to Lauder, which will be discussed later in the paper.

3.1 Coincidence criteria and other matching details

The SCIAMACHY and GOSAT comparisons in this paper are based on two different definitions of coincidence criteria between TCCON and satellite data. Satellite measurements, which satisfy the so-called geometric criteria, are within ±1 h, ±5° latitude and longitude of an unaveraged TCCON observation. Following the match-ups, all TCCON observations matching one satellite observation which are within 90 min are averaged, reducing the TCCON random error. The dynamical criteria (Wunch et al., 2011b; Keppel-Aleks et al., 2011, 2012) are designed to exploit informa-

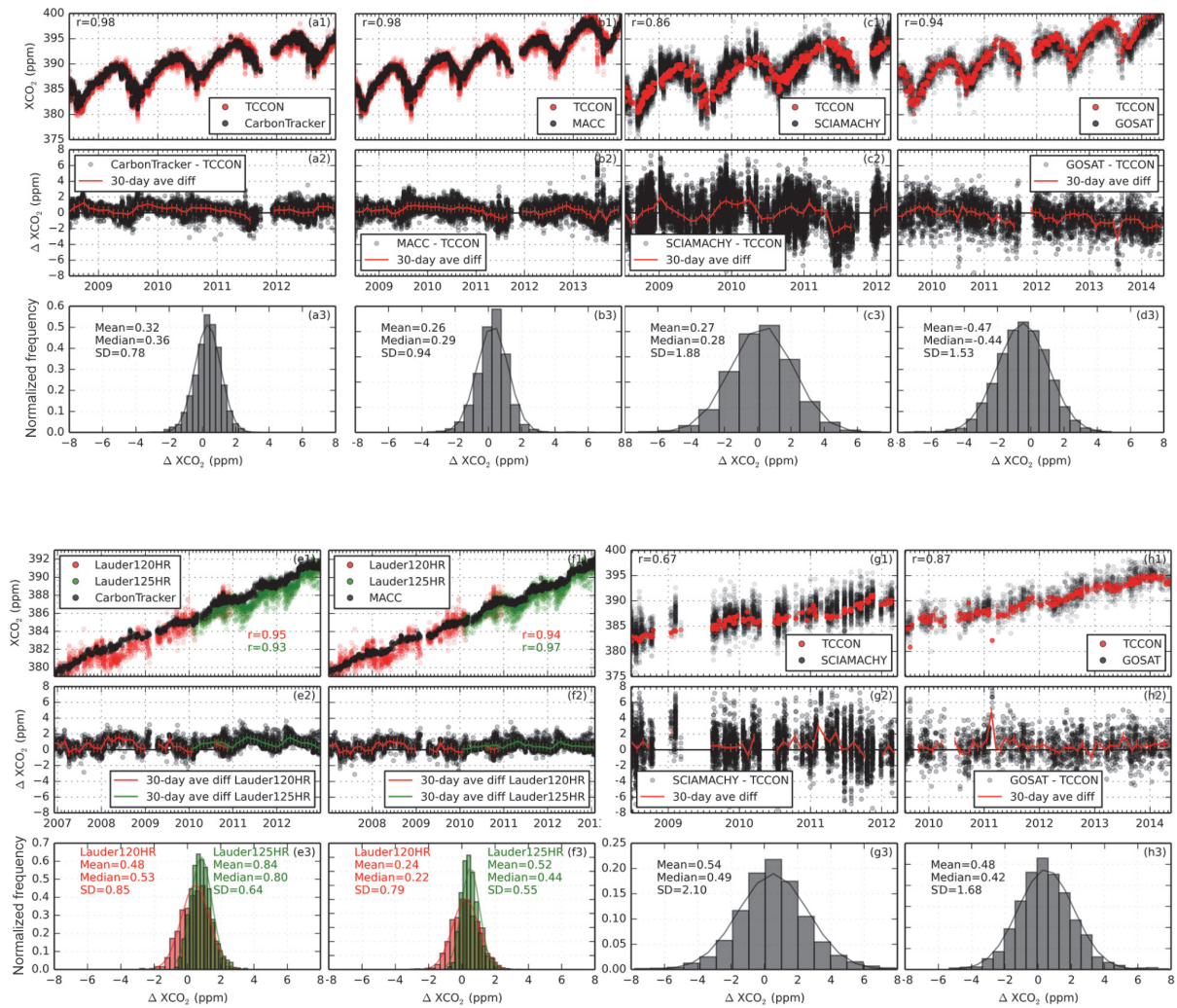


Figure 2. Time series for matches of CT2013b, MACC, SCIAMACHY, and GOSAT vs. TCCON at Lamont (**a–d**) and Lauder (for models) or Wollongong (for satellites) (**e–h**). The top plot of each set (**a1, b1, c1, ...**) shows a time series of all geometric matching pairs. The middle panel of each set (**a2, b2, c2, ...**) shows the difference vs. TCCON, with the blue line denoting the 30-day average difference. The bottom panel of each set (**a3, b3, c3, ...**) shows a histogram of the differences, indicating an approximate error and bias.

tion about the dynamical origin of an air parcel through a constraint on the free-tropospheric temperature. This allows us to relax the geometric constraints and find more coincident satellite soundings per TCCON measurement. Briefly, a match is found when the measurements are within 5 days and the following is satisfied:

$$\left(\left(\frac{\Delta \text{Latitude}}{10} \right)^2 + \left(\frac{\Delta \text{Longitude}}{30} \right)^2 + \left(\frac{\Delta \text{Temperature}}{2} \right)^2 \right) < 1, \quad (1)$$

where $\Delta \text{Temperature}$ is the co-located NCEP temperature difference at 700 hPa (Kalnay et al., 1996). Matches are found with unaveraged TCCON data; TCCON observations matching a single satellite observation are averaged within 90 min intervals. Table A2 summarizes the coincidence criteria and data versions that are used. Other matching schemes

not included in this paper include a method implemented by S. Basu described in Guerlet et al. (2013), which utilizes model CO_2 fields to determine coincidences, and Nguyen et al. (2014) which uses a weighted average of distance, time, and mid-tropospheric temperature. Dynamic and geometric coincidence criteria are compared in Sect. 3.3 and geometric coincidence criteria are used to spot-check dynamic coincidence criteria results.

The choices used in this paper regarding model/TCCON match-ups are linearly interpolating to the TCCON latitude, longitude, and time for the models, and using the TCCON surface pressure for calculating X_{CO_2} . In the cases where the TCCON surface pressure is greater than the model surface pressure, the model surface CO_2 value is replicated to the missing pressure values. When comparing models vs. TC-

Table 2. Fit of a and b in Eq. (2), with stations having data out to at least $n = 50$ for dynamic coincidence criteria, $n = 10$ (GOSAT) or $n = 40$ (SCIAMACHY) for geometric coincidence criteria. The CT-CT column describes the CT2013b coincidence error, providing a lower bound on the satellite/TCCON differences. The subtr. co-location error row estimates the correlated error for satellites only, subtracting the quadrature TCCON error (0.35 ppm) and co-location error from the CT-CT column.

Station	Dynamic			Geometric			Dynamic			Geometric		
	SCIAMACHY	CT-CT		SCIAMACHY	CT-CT		GOSAT	CT-CT		GOSAT	CT-CT	
	a	b		a	b		a	b		a	b	
Sodankylä	1.0	2.3	0.5				1.0	2.0	0.7			
Bialystok	1.1	2.1	0.6	1.5	1.6	0.3	0.7	1.9	0.7	0.6	1.6	0.2
Bremen ^a	1.2	2.0	1.0	1.4	1.6	0.3	1.0	1.9	0.6			
Karlsruhe	1.4	1.9	0.8	1.6	1.8	0.5	1.0	1.8	0.7	1.3	2.0	0.6
Orléans	1.2	2.1	0.5	1.5	1.8	0.2	0.9	1.9	0.6	1.2	1.7	0.4
Garmisch ^a	1.3	2.1	1.0				1.0	2.0	1.0	0.9	2.3	0.6
Park Falls	1.2	2.0	0.7	1.4	1.6	0.2	0.9	1.9	0.7	0.8	1.5	0.3
Four Corners ^a	1.3	1.9	0.4				1.0	1.8	0.4	1.0	1.6	0.2
Lamont	1.3	1.9	0.6	1.2	1.5	0.3	1.0	1.7	0.6	0.8	1.3	0.3
JPL2007 ^a	1.3	2.0	0.6									
JPL2011 ^a	1.5	1.8	1.0				1.0	1.7	0.7	1.1	1.5	0.4
Tsukuba	1.3	2.0	0.5									
Izaña ^a	1.0	1.7	0.4				0.6	1.3	0.5	0.8	1.0	0.4
Saga	0.9	2.2	0.5				0.9	2.0	0.7			
Mean NH	1.2 ± 0.2	2.1 ± 0.1	0.6 ± 0.1	1.4 ± 0.2	1.7 ± 0.2	0.3 ± 0.1	0.9 ± 0.1	1.9 ± 0.1	0.7 ± 0.1	0.9 ± 0.1	1.6 ± 0.1	0.4 ± 0.2
Mean NH: subtr. co-loc. error	1.0 ± 0.3			1.4 ± 0.3			0.5 ± 0.2			0.8 ± 0.2		
Darwin	1.0	1.5	0.2	1.2	1.2	0.1	0.7	1.2	0.3	0.8	1.0	0.1
Ascension							0.7	1.1	^b	0.8	0.8	^b
Reunion							0.7	1.2	0.2			
Wollongong	1.0	1.7	0.3	1.1	1.8	0.2	0.7	1.3	0.2	0.8	1.5	0.2
Mean SH	1.0 ± 0.1	1.6 ± 0.1	0.3 ± 0.1	1.2 ± 0.1	1.5 ± 0.4	0.2 ± 0.1	0.7 ± 0.1	1.2 ± 0.1	0.2 ± 0.1	0.8 ± 0.1	1.1 ± 0.4	0.2 ± 0.1
Mean SH: subtr. co-loc. error	0.9 ± 0.2			1.1 ± 0.2			0.6 ± 0.2			0.7 ± 0.1		

^a Bremen, Garmisch, Four Corners, JPL, and Izaña are influenced by local effects or complex terrain and are not included in averages. ^b Ascension data start in 2013 after CT2013 stops.

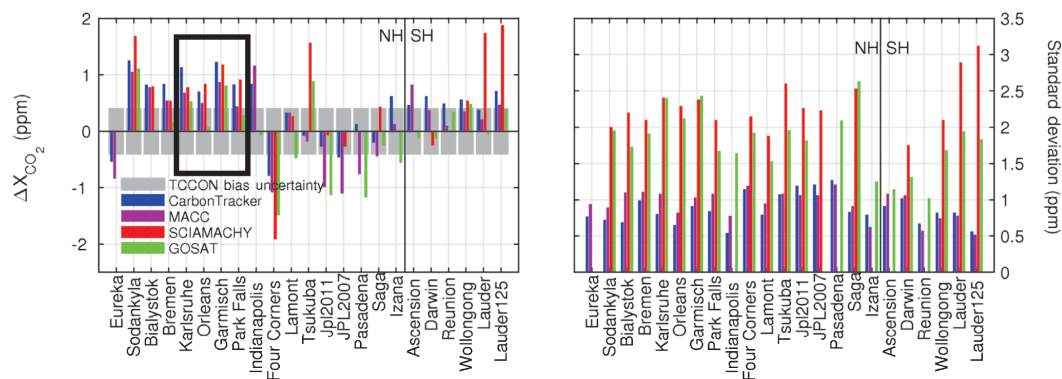


Figure 3. Bias (left panel) and standard deviation (right panel) for CT2013b, MACC, BESD-SCIAMACHY, and ACOS-GOSAT vs. TCCON stations, arranged from high to low latitude. Comparisons which have a particularly low number of matches are TSUKUBA and Lauder for SCIAMACHY and Lauder for GOSAT.

CON, the TCCON averaging kernel is applied to the model. When using the model to assess satellite coincidence error, the satellite averaging kernel is applied to the model at the satellite and TCCON coincidences. Note that the TCCON averaging kernel cannot be applied to satellite data and vice versa because a profile of higher resolution than the comparison observation is needed to apply an averaging kernel. In the satellite/TCCON comparison, both products have $\sim 1^\circ$ of freedom. (The satellites do initially retrieve a profile with

$\sim 1.6^\circ$ of freedom for GOSAT, but the profile is not what we are validating.)

Because of the earth's curvature, high-latitude sites could have relaxed coincidence in longitude, particularly for geometric coincidence. However stations north of 60° N have gaps in the winter months in the satellite record such that relaxed criteria do not add additional stations to the analysis.

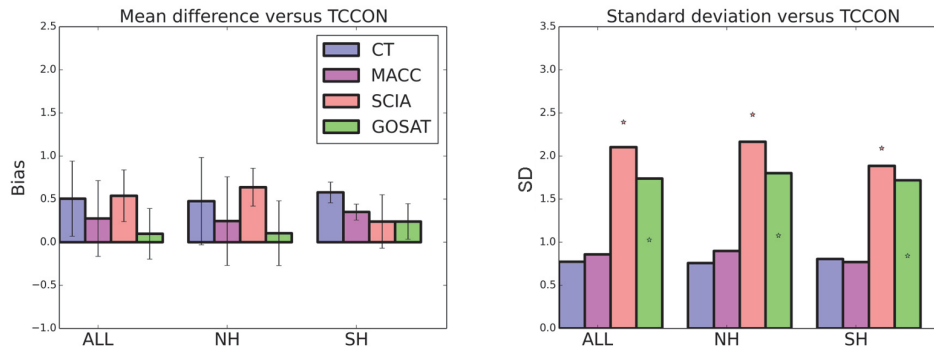


Figure 4. Overall bias (left panel, with error bars showing the standard deviation of the bias) and standard deviation (right panel, with stars showing the predicted error for satellites) for most stations (some stations removed, see text).

3.2 Bias and standard deviation for individual matches

Figure 3 shows a summary of the comparisons for geometric criteria where satellite matches are not averaged. Averaging and the effects of coincidence criteria and satellite averaging are addressed in Sect. 3.4. The black box shows five European stations which are very close, geographically, yet have different biases. The gray bars labeled TCCON bias uncertainty in Fig. 3 signify the overall calibration uncertainty in TCCON which is estimated to be 0.4 ppm (Wunch et al., 2010, 2011a). The significance of the bias vs. TCCON is estimated by the 5% t test (looking up bias/standard deviation times the square root of the number of comparisons in a t test table). The gray box is the uncertainty in the TCCON station-to-station bias and all comparisons with $|\text{bias}| > 0.4$ were found by the t test to be significant. Therefore, from the above two pieces of information, all biases differing from TCCON more than 0.4 ppm are significantly different than TCCON. For GOSAT, biases larger than the TCCON bias uncertainty occur at Sodankylä(+), Karlsruhe(+), Lamont(-), Tsukuba(+), and Wollongong(+). SCIAMACHY has the same outliers as GOSAT with additional biases at Bialystok(+), Orléans(+), Park Falls(+), and Lauder(+). The sites with identified local influences and the general bias with respect to these sites are Bremen(+), Four Corners(-), JPL2007(-), and JPL2010(-). The stations which have complex conditions are Garmisch and Izaña, which show similar biases to surrounding stations. For unaveraged results, model standard deviations are lower than either satellite as satellite differences result from both systematic and random measurement error, the latter which does not occur for models. The standard deviations show some variability from station to station which are investigated below. The effects of averaging and coincident criteria are investigated in Sect. 3.3.

Figure 4 shows the biases and standard deviations grouped globally and over the northern and southern hemispheres. To estimate the overall bias and standard deviations for single observations, we take out the outliers as follows.

As described in Sect. 2.1, we take out JPL, Four Corners, Bremen, Garmisch, and Izaña for averaging. For satellites, we remove the above plus Tsukuba and Lauder due to limited numbers of comparisons for SCIAMACHY. For the mean NH bias, we take out stations poleward of 60° N, which show a large positive bias for GOSAT and SCIAMACHY. There is an overall bias vs. TCCON on the order of 0.7 ppm for CT2013b, and 0.2–0.3 ppm for the other three sets. The overall bias is less of a concern than the bias variability in satellite data which indicates regional errors that will translate to regional errors in flux estimates. The bias variability is 0.4, 0.4, 0.5, and 0.3 ppm for CT2013b, MACC, SCIAMACHY non-polar, and GOSAT non-polar respectively. Note SCIAMACHY data are corrected to have an average zero bias with respect to TCCON GGG2012 which is 0.3 ppm higher than GGG2014 (https://tcccon-wiki.caltech.edu/Network_Policy/Data_Use_Policy/Data_Description#CORRECTIONS_AND_CALIBRATIONS). The overall standard deviations are 0.9 ppm for CT2013b, 0.9 ppm for MACC, 2.1 ppm for SCIAMACHY (all latitudes), and 1.7 ppm for GOSAT (all latitudes). These values represent the overall performance of CT2013b, MACC, and single soundings from SCIAMACHY and GOSAT. The standard deviations are somewhat lower for the SH stations used, with values of 0.8, 0.8, 2.0, and 1.6 for CT2013b, MACC, SCIAMACHY, and GOSAT, respectively. Reuter et al. (2013) validated earlier retrieval versions of BESD-SCIAMACHY and ACOS-GOSAT with TCCON and found 2.1 ppm (BESD) and 2.3 ppm (ACOS) for the single sounding precision and 0.9 ppm for the station-to-station biases. Their findings for BESD are consistent with the findings of Dils et al. (2014). The station-to-station biases are lower in our analysis due to corrections in TCCON, improvements in satellite estimates, and removal of several stations from the estimates. The polar station Sodankylä north of 67° N is compared to GOSAT and SCIAMACHY. A similar standard deviation is found vs. other TCCON stations, but a higher bias, 1.6 ppm for GOSAT and SCIAMACHY. Most of the remaining analysis in the paper does not have enough

coincidences at high latitudes so it is important to note this result.

We test whether the biases seen in Figs. 3 and 4 are persistent from year to year. When at least two full-year averages exist for a station, the standard deviation of the yearly bias is calculated. The average over all stations of the yearly bias standard deviation is 0.3 ppm for all sets (CT2013b, MACC, SCIAMACHY, GOSAT).

Another important comparison is of the predicted and actual errors. The predicted error (also referred to as the a posteriori error) is reported for each satellite product and the actual error we take to be the standard deviation of the satellite observation vs. TCCON. These two quantities should agree if the TCCON error is much smaller than the a posteriori error and the coincidence criteria do not degrade the agreement. The predicted and actual errors vary from site to site, e.g., from variations in albedo, aerosol composition, and solar zenith angle. We calculate the correlation between the standard deviation vs. TCCON and the predicted error for each site as follows: the standard deviation of the satellite vs. TCCON is calculated at each TCCON station. The correlation of the standard deviation and predicted errors by station are calculated. ACOS-GOSAT has a 0.6 correlation and BESD-SCIAMACHY has a 0.5 correlation. This indicates that the predicted error should be utilized, e.g., when assimilating ACOS-GOSAT, as the variability in the predicted error represents variability in the actual error, though not perfectly. A scale factor should also be applied to the predicted errors. For ACOS-GOSAT the predicted error averaged over all TCCON sites is 0.9 ppm, as compared to the actual error of 1.7 ppm and can be corrected by applying a factor 1.9 to the reported GOSAT errors. For BESD-SCIAMACHY, the prediction error of 2.3 ppm multiplied by 0.9 agrees with the 2.1 ppm actual error.

3.3 Errors as a function of coincidence criteria and averaging

We now directly compare performance of geometric and dynamic coincidence criteria and averaging in terms of error. Figure 5 shows SCIAMACHY and GOSAT standard deviations vs. TCCON for geometric and dynamical coincidence criteria in the Northern Hemisphere. The stations used were those that had entries for all comparisons, listed in the Fig. 5 caption. For $n = 1$ no averaging is done and the dynamic coincidence criteria perform similarly to the geometric criteria, though the dynamic error is ~ 0.1 ppm higher for both satellites. For $n = 2$, exactly two satellite observations were averaged for each coincidence. The error drops substantially, but not as $1/\sqrt{2}$, which would be expected if the error were uncorrelated. For $n = 4$, the error again drops but it is not half the $n = 1$ error, which is shown by the dotted line. At $n = 4$, the dynamic coincidence error is the same as the geometric error for SCIAMACHY, likely because dynamic coincidence involves averaging observations farther apart in location and

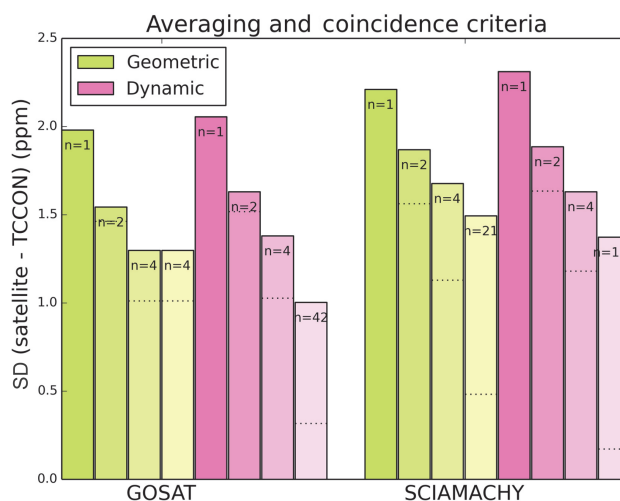


Figure 5. Standard deviation of SCIAMACHY and GOSAT minus TCCON for different coincidence criteria and number of satellite observations averaged, n , in the Northern Hemisphere. The following consistent set of stations was used for all comparisons: Bialystok, Karlsruhe, Orléans, Garmisch, Park Falls, and Lamont. The dotted line shows the error if it scaled as the inverse square root of the number of averaged observations. The far right case for each of the categories contains the maximum n that has results for all stations.

time, which are less likely to have correlated errors. The last bar is the maximum n , which has results for all stations included. The dynamic criteria allow far more coincidences, resulting in significantly lower average errors. Note that all averaged satellite observations match one particular TCCON observation.

3.4 Errors vs. averaging: random and correlated error

To test the effects of spatial averaging, we calculate station by station standard deviations of satellite–TCCON matched pairs as a function of n , where n is the number of satellite observations that are averaged, which are chosen randomly from available matches (so there should be no difference in the characteristics of chosen points for larger vs. smaller n). Figure 6 shows plots from Lamont for SCIAMACHY and GOSAT for standard deviation difference to TCCON vs. n . Initially the error drops down rapidly with n , however the decrease slows with larger n . The curve fits well to the theoretically expected form:

$$\text{error}^2 = a^2 + b^2/n, \quad (2)$$

where a represents correlated errors which do not decrease with averaging for similar cases (including smoothing errors, errors from interferences such as aerosols, TCCON error, and co-location error), b represents uncorrelated errors which decrease with averaging, and n represents the number of satellite observations that are averaged. The purple

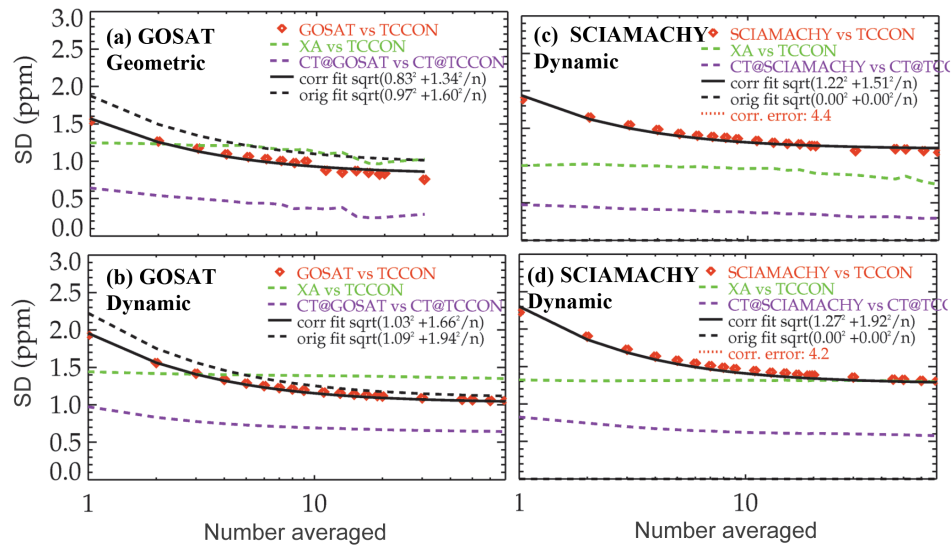


Figure 6. Averaging matches of satellite data vs. TCCON at Lamont. As the number averaged increases, the standard deviation vs. TCCON decreases. CT2013 at the satellite vs. at CT2013 at TCCON (purple) is used to quantitate spatiotemporal mismatch error. The points are fit to Eq. (2) (black). For GOSAT the uncorrected data are also fit (black dashed). The initial guess minus TCCON standard deviation is shown as a green dashed line. We see that in this case, for GOSAT at Lamont, averaging more than about four observations improves over the initial guess.

dashed line represents the standard deviation of CT2013b at the satellite time and location vs. CT2013b at the TCCON time and location and represents a lower bound of the spatiotemporal mismatch error (co-location error). As expected, this value is much smaller for geometric than for our dynamic coincidence criteria (other dynamic coincidence criteria may do better; in our case the dynamic criteria consider points $\pm 30^\circ$ longitude, $\pm 10^\circ$ latitude, ± 5 days, and ± 2 K temperature vs. 1 h, 5° for geometric criteria). The co-location error for large n is shown in the CT-CT columns of Table 2, with Northern Hemisphere averages of 0.3, 0.6, 0.4, and 0.7 ppm, for SCIAMACHY geometric, SCIAMACHY dynamic, GOSAT geometric, and GOSAT dynamic coincidences, respectively. There is not much difference between the estimated co-location error at North American sites of Lamont and Park Falls where the CT2013b model is at $1 \times 1^\circ$ vs. other Northern Hemisphere sites where the CT2013b model is at $3 \times 2^\circ$. The co-location error is subtracted from the a value in quadrature to estimate results without co-location error.

We calculate a and b by station in Table 2, splitting out geometric and dynamic coincidence criteria. The a term, which does not reduce with averaging, is the correlated error, and the b term, which reduces with averaging, is the uncorrelated error. There is more correlated error, a , for SCIAMACHY geometric vs. dynamic matches in 5/6 stations in the Northern Hemisphere, indicating that averaging is more effective when it is over a larger spatial/temporal area, probably due to variability in the source of the correlated errors. GOSAT geometric vs. dynamic averages show larger cor-

related error for 4/8 stations. The GOSAT dynamic correlated error of 0.9 ppm is significantly reduced by subtraction of the 0.7 ppm co-location error, whereas the 0.9 ppm geometric error is not strongly changed by the subtraction of 0.4 ppm co-location error. Subtracting (in quadrature) the co-location error (CT-CT column in Table 2) and TCCON error of 0.35 ppm (Appendix B) results in the corrected correlated error, a , shown in the mean NH: subtr. co-location error row of Table 2.

The northern hemispheric average values, corrected by co-location and TCCON error, are $a = 1.4 \pm 0.3$ ppm, $b = 1.7 \pm 0.2$ ppm for SCIAMACHY geometric, $a = 1.0 \pm 0.3$ ppm, $b = 2.1 \pm 0.1$ ppm for SCIAMACHY dynamic, $a = 0.8 \pm 0.2$ ppm, $b = 1.6 \pm 0.1$ ppm for GOSAT geometric, $a = 0.5 \pm 0.2$ ppm, $b = 1.9 \pm 0.1$ ppm for GOSAT dynamic. These values indicate the expected error when averaging GOSAT or SCIAMACHY observations within 5° and 90 min for geometric, and for 5 day regional averages for dynamic coincidences.

The green dashed line in Fig. 6 shows the standard deviation of the satellite prior vs. TCCON. Although using an optimal constraint will result in an error lower than the prior error in the absence of systematic errors, these satellite retrievals of CO_2 have been set up to value average results over single observations; while the error increases from the prior for a single observation, average results have both less error and minimal prior influence.

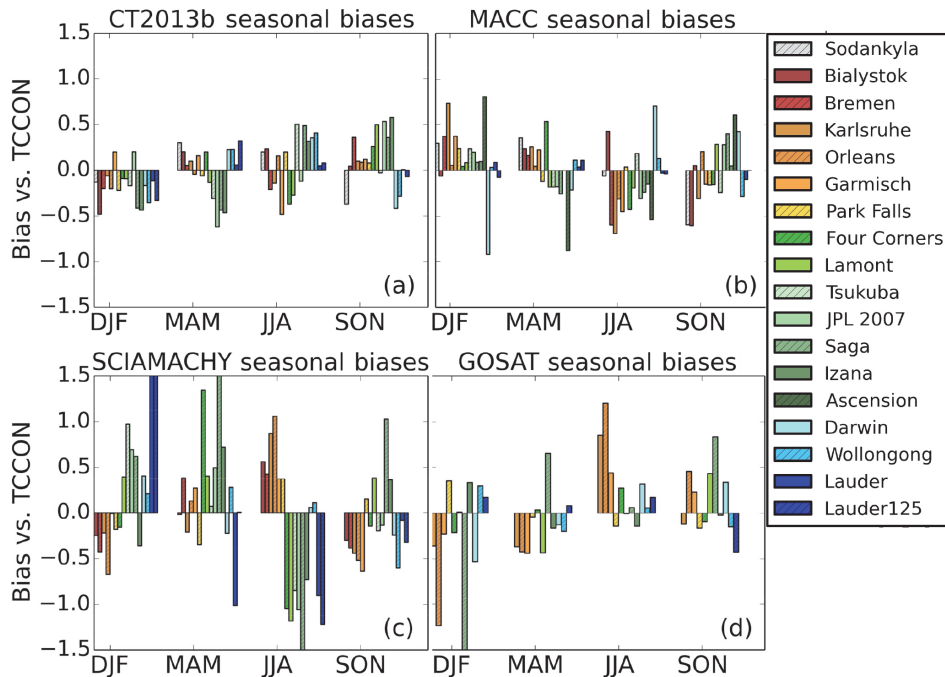


Figure 7. Bias for 3-month groups for each station, where each station is normalized to have 0 yearly bias. For satellites, stations are included when at least 20 matches are found in each season. Dynamic coincidence criteria are used. The station colors are coded by location: far NH gray, European and Park Falls red/yellow, midlatitude green, SH blue.

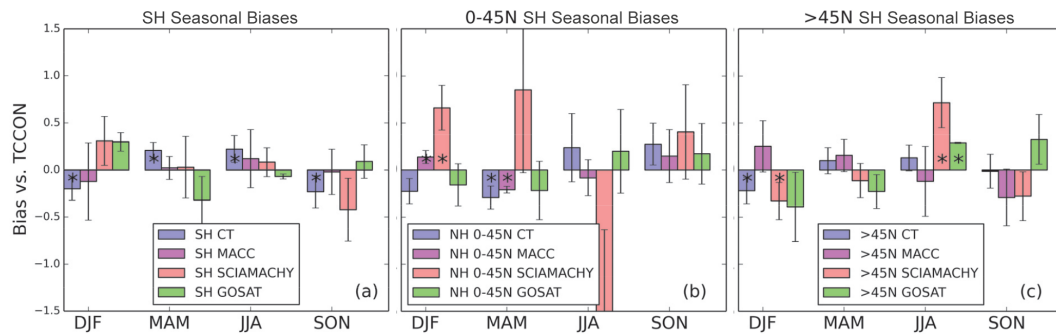


Figure 8. Bias for 3-month groups for Southern Hemisphere (left panel), 0–45° N (middle panel), and poleward of 45° N (right panel). Each station composing each group is normalized to have zero average over the year. The Southern Hemisphere (left panel, Lauder (except SCIAMACHY), Wollongong, and Darwin) has relatively small biases. 0–45° N includes Tsukuba and Lamont. >45° N includes Karlsruhe, Park Falls, and Orléans for satellites, and those plus Sodankylä and Bialystok for models. The results that are significant according to the *t* test are shown with a * in the bar plots.

3.5 Seasonally dependent biases

It is important to determine whether there are seasonally dependent biases, as these will impact flux distributions. We look at 3-month periods (DJF, MAM, JJA, SON), with the overall yearly bias at each site subtracted out to isolate the seasonal biases. To get enough comparisons, we use the dynamical criteria for satellite coincidences, as using the geometric criteria cuts down the comparisons with sufficient seasonal coverage to three stations (Park Falls, Lamont, and

Wollongong). This is a simple averaging method which will later be compared to seasonal cycle amplitude fit results.

Figure 7 shows the biases for stations that have at least 20 matches in each season, and Fig. 8 shows the results averaged by SH, 0–45° N (which includes Tsukuba and Lamont), >45° N. The error bars shown in Fig. 8 are the standard deviation of the results in each bin, once the station-dependent average biases are subtracted. Significance was tested using the *t* test with biases larger at least 0.2 ppm. For the north midlatitudes, the seasonal cycle peak is in MAM and the minimum is in JJA, so the seasonal cycle error should be

approximately bias (MAM) minus bias (JJA). For Lamont and Tsukuba in JJA, SCIAMACHY is biased low vs. TCCON by about 1.4 ppm (the low bias is 0.8 ppm for Lamont). In 45–53° N, SCIAMACHY is biased low vs. TCCON in MAM and high in JJA, with about a 0.9 ppm spread. GOSAT has the same pattern with a 0.5 ppm spread. SCIAMACHY is biased low vs. TCCON in DJF by 0.7 ppm, but this does not affect the seasonal cycle amplitude. GOSAT is similarly low but has more station-dependent variability so is not considered significant. CT2013b is biased high vs. TCCON in JJA by about 0.2 ppm and biased low vs. TCCON in DJF by about 0.3 ppm. MACC is biased low vs. TCCON in MAM by 0.2 ppm in 28–37° N. In the Southern Hemisphere, the seasonal cycle maximum is in the October to January time frame, and the minimum is in the April time frame. The significant findings in the SH are that CT2013b is biased about 0.3 ppm low vs. TCCON in DJF and biased 0.2 ppm high in MAM, which should lead to a 0.5 ppm underestimate of the SH seasonal cycle. Geometric results, which including fewer stations due to the cutoff of at least 20 observations per season, corroborate the low bias for in MAM, high bias in JJA, and low bias in DJF in 45–53° N for SCIAMACHY, and the low bias in MAM and high bias in JJA in 45–53° N for GOSAT.

4 Comparisons of seasonal cycle amplitude, phase, and yearly increase

We compare model and satellite X_{CO_2} to TCCON using the NOAA fitting software CCGCRV (Thoning et al., 1989) to calculate seasonal cycle amplitudes and yearly increases. At least 2 years are needed to distinguish the seasonal cycle from the yearly increase. The errors are calculated using the bootstrap method (Rubin, 1981; Efron, 1979), the standard deviation of differences within one bin divided by the square root of the number of stations minus 1 ($n > 1$), the variability of results when choosing different averaging (for satellites) where one, two or four satellite observations are averaged for GOSAT and SCIAMACHY, coincidence errors estimated by the mean difference between CT2013b matched to the satellite and CT2013b matched to TCCON, and the variability resulting from small changes in the time range. Two data sets at a time are matched, using the dynamic criteria, with the satellite averaging four observations for SCIAMACHY and two for GOSAT, which reduces the fit errors. Since different data sets will have different data gaps and time ranges, the TCCON results will be somewhat different for each comparison. Plots are individually examined to ensure that there are adequate data. Stations are removed when there are large errors for any of the above errors, and locally influenced stations are also removed (see Sect. 2.1). The stations that were excluded based on variability of results for changes in the time range on the order of a few months are Eureka for the models and Sodankylä, Bialystok, Saga, and Lauder_125HR for the satellites. At Bialystok, for example, the GOSAT-TCCON

seasonal cycle was 0.3 or -0.9 ppm depending on whether the time series ended in November 2013 or on 1 January 2014, respectively. The time series shows large gaps in the data in the winter and a gap around the seasonal cycle minimum in 2012. The findings show that the GOSAT seasonal cycle amplitude at Bialystok may be influenced by a small subset of the GOSAT points.

4.1 Seasonal cycle amplitude

The seasonal cycle amplitude is important for estimating source and sink estimates and global distributions. Table 3 shows the seasonal cycle amplitudes grouped by latitude. As described in Sect. 4, the error is calculated by several different methods. The error in Table 3 is root mean square of the end date choice, bootstrap error, averaging choice, and bin standard deviation. The co-location error (only relevant for satellites) is calculated as an average bias in the bin; this bias should be subtracted from the satellite–TCCON differences to remove the effects of co-location error. When $n > 1$, results are considered significant using the t test of all the results in the bin (comparison of the mean difference/standard error difference to a standard look up table). The calculated seasonal cycle amplitudes are shown in Table 3, and results are compared for consistency to the simpler seasonal cycle biases calculated in Sect. 3.5 to understand what season the discrepancy occurs.

The significant findings from Table 3 are as follows. (1) In northern latitudes (46–53° N), GOSAT underestimates the seasonal cycle by 0.4 ppm. This latitudinal range is composed of two European sites and Park Falls, all of which follows this pattern, so it is not just in Europe but also North America at this latitude. This finding is consistent with Lindqvist et al. (2015). The simple bias calculation from Sect. 3.5 estimates that GOSAT should underestimate the seasonal cycle by 0.5 ppm, with half the issue being a low bias in MAM and half the issue being a high bias in JJA. The co-location error is estimated as +0.1 ppm, so removing the co-location error should increase the GOSAT–TCCON discrepancy. The geometric criteria bias for this same bin is -0.7 ± 0.4 ppm, which is consistent with the dynamic results, though with larger error bars and significantly sparser time series plots. (2) There is an underestimate in the CT2013b seasonal cycle for 45–53° N compared to TCCON by 0.2 ppm. This is a small discrepancy but the error bars are also small. Similarly in Sect. 3.5, a 0.2 ppm low bias is predicted. (3) There is an overestimate in the MACC seasonal cycle in the 28–37° N range by 0.6 ppm. A smaller 0.2 ppm low bias is seen in MAM for MACC. (4) CT2013b underestimates the SH seasonal cycle by 0.5 ppm. This is partially from a low bias in DJF and partly from a high bias in MAM.

Findings from Sect. 3.5 which did not reach significance in Table 3 are as follows. (1) SCIAMACHY should underestimate the seasonal cycle in 45–53° N by 2.5 ppm, which it does not. There is a 0.4 ppm high bias with large error bars.

Table 3. Seasonal cycle amplitude for different latitudes. Stations included for satellites are Karlsruhe, Orléans, Park Falls, Lamont, Darwin, Reunion (GOSAT), and Wollongong. Stations included for models are Sodankylä, Bialystok, Karlsruhe, Orléans, Park Falls, Lamont, Tsukuba, Saga, Darwin, Reunion, Wollongong, Lauder_120HR, and Lauder_125HR. Bold shows entries with statistically significant differences.

Comparison	Region	Seasonal amp. (ppm)	Seasonal amp. TCCON (ppm)	Seasonal amp. difference (ppm)	End date (ppm)	Co-loc. (ppm)	Bootstrap (ppm)	Averaging choices (ppm)	Bin SD (ppm)
CT2013b	67–79° N ($n = 2$)	9.6	9.7	-0.1 ± 0.1	0.00		0.14	–	–
	46–53° N ($n = 4$)	7.9	8.1	-0.2 ± 0.1	0.04		0.13	–	0.1
	28–37° N ($n = 1$)	6.3	7.4	-0.6 ± 0.5	0.03		0.14	–	0.5
	SH ($n = 4$)	0.9	1.6	-0.5 ± 0.2	0.05		0.09	–	0.2
MACC	67–79° N ($n = 2$)	10.9	10.2	0.7 ± 0.5	0.3		0.33	–	0.33
	46–53° N ($n = 4$)	8.4	8.2	0.2 ± 0.1	0.06		0.16	–	0.13
	28–37° N ($n = 3$)	5.3	4.7	0.6 ± 0.2	0.05		0.16	–	0.16
	SH ($n = 5$)	1.1	1.3	-0.2 ± 0.2	0.05		0.08	–	0.19
GOSAT (v3.5)	46–53° N ($n = 3$)	6.7	7.1	-0.4 ± 0.1	0.05	+0.1	0.06	0.07	0.1
	28–37° N ($n = 1$)	5.1	5.5	0.4 ± 0.1	0.03	–0.1	0.01	0.02	–
	SH ($n = 3$)	2.0	2.1	-0.1 ± 0.2	0.09	+0.1	0.02	0.03	0.2
SCIAMACHY (BESD-v02.00.08)	46–53° N ($n = 3$)	7.5	7.8	0.4 ± 0.8	0.06	+0.4	0.09	0.03	0.8
	28–37° N ($n = 1$)	7.2	6.4	0.8 ± 0.1	0.11	–0.0	0.03	0.04	–
	SH ($n = 2$)	2.2	2.0	0.2 ± 0.3	0.06	–0.2	0.03	0.01	0.3

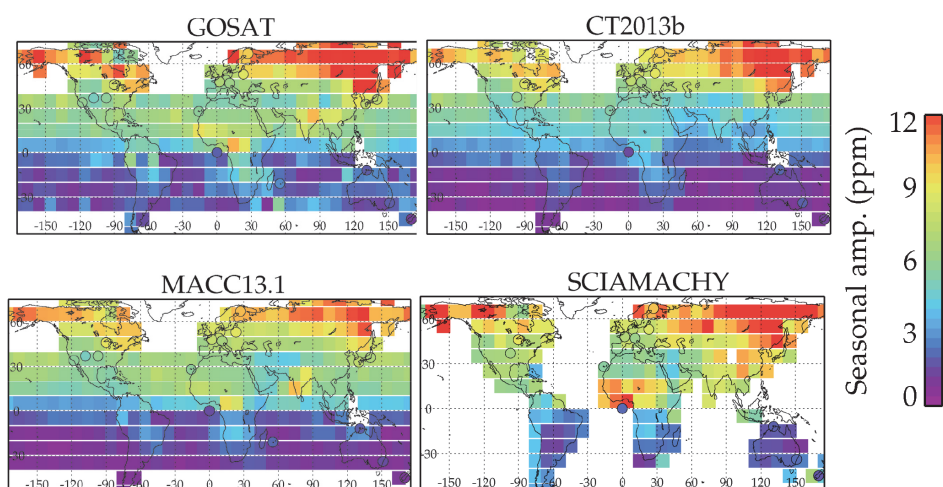


Figure 9. Seasonal cycle amplitude. The TCCON values are shown by the circles. The averaging is done over $10 \times 10^\circ$ bins every 5° . Comparisons vs. TCCON may be different than Table 4, since Table 4 has very close criteria for models and dynamic coincidence criteria for satellite data. The models are sampled at GOSAT observations.

(2) SCIAMACHY should overestimate the seasonal cycle for Lamont by 1.6 ppm. In Table 3, SCIAMACHY overestimates by 0.8 ppm but statistical significance is not calculated for one station. (3) MACC should underestimate the seasonal cycle by 0.2 ppm in 28–37° N. In Table 3 MACC overestimates the seasonal cycle by 0.6 ppm. The seasonal biases are generally consistent with the seasonal cycle amplitude differences vs. TCCON and can be used to pinpoint which months are the cause of the seasonal cycle amplitude differences vs. TCCON; however, the mean seasonal biases find more significant differences than the seasonal cycle fits.

Figure 9 shows a global map of fits of the seasonal cycle amplitude of SCIAMACHY, GOSAT, CT2013b, and MACC, with TCCON having at least 2 years of matches shown by

circles. This map shows how the results of Table 4 fit into the global pattern (with the model fields matched to GOSAT locations and times). Interestingly, the seasonal cycle amplitude varies longitudinally; this pattern is seen in both satellite data sets and both models. Since the amplitude is taken from the sampled harmonic there is no extrapolation, although the seasonal cycle will be underpredicted at high latitudes where there are significant data gaps. The model data in Fig. 9 is co-located with GOSAT, so the same gaps will occur in GOSAT and the two models, other than fit errors larger than 10% of the amplitude, were screened out. This map is consistent with Lindqvist et al. (2015), Fig. 8, which also finds high values in the 45–50° N, 120–180° E range.

Table 4. Yearly increases. Each comparison uses matched pairs with TCCON using locations which have at least 2 years of data for comparisons. See Table 3 for stations included. The start date and end date are averaged for the stations in each bin and are shown in the second to last column. Bold text shows one difference larger than predicted errors. The last column shows the average global yearly increase for the time period using Table 6.

Comparison	Region	Yearly incr. (ppm)	Yearly incr. TCCON (ppm)	Yearly incr. difference (ppm)	End date choice (ppm)	Bootstrap (ppm)	Co-loc. (ppm)	Avg. choice (ppm)	Bin SD (ppm)	Period (year fract.)	Global (ppm yr ⁻¹)
CT2013b	67–79° N (<i>n</i> = 2)	2.21	2.18	0.07 ± 0.1	0.00	0.07	–	–	–	Jun 2009–Oct 2012	2.06
	46–53° N (<i>n</i> = 4)	2.12	2.19	–0.07 ± 0.1	0.03	0.11	–	0.04	0.04	Oct 2008–Dec 2013	2.03
	28–37° N (<i>n</i> = 3)	2.17	2.45	–0.28 ± 0.5	0.02	0.11	–	0.14	0.14	Sep 2009–Feb 2012	2.01
	SH (<i>n</i> = 5)	1.94	1.92	0.02 ± 0.2	0.03	0.04	–	–	0.07	Jul 2008–Aug 2012	1.97
MACC	67–79° N (<i>n</i> = 2)	2.13	2.25	–0.12 ± 0.04	0.02	0.18	–	–	0.24	Jun 2009–Oct 2013	2.16
	46–53° N (<i>n</i> = 4)	2.27	2.29	–0.02 ± 0.1	0.04	0.09	–	0.07	0.07	Jan 2009–Dec 2013	2.14
	28–37° N (<i>n</i> = 3)	2.35	2.45	–0.10 ± 0.1	0.02	0.09	–	0.14	0.14	Sep 2009–Sep 2012	2.08
	SH (<i>n</i> = 5)	2.18	2.14	0.04 ± 0.1	0.04	0.03	–	–	0.03	Nov 2008–Jun 2013	2.07
GOSAT (v3.5)	46–53° N (<i>n</i> = 3)	2.17	2.31	–0.14 ± 0.1	0.05	0.04	–0.2	0.07	0.08	Oct 2009–Apr 2014	2.20
	28–37° N (<i>n</i> = 1)	2.25	2.55	–0.30 ± 0.02	0.01	0.01	–0.1	0.01	–	Jun 2010–Jun 2014	2.20
	SH (<i>n</i> = 4)	2.12	2.17	–0.1 ± 0.1	0.08	0.01	–0.4	0.05	0.10	Mar 2010–Jun 2014	2.21
SCIAMACHY (BESD-v02.00.08)	46–53° N (<i>n</i> = 3)	1.95	1.99	–0.04 ± 0.1	0.06	0.05	0.03	0.03	0.10	Jan 2009–Mar 2012	1.95
	28–37° N (<i>n</i> = 1)	2.03	2.16	–0.13 ± 0.1	0.11	0.02	0.09	0.04	–	Jan 2010–Mar 2012	2.08
	SH (<i>n</i> = 1)	1.76	1.96	–0.20 ± 0.1	0.06	0.01	–0.05	0.01	0.06	Oct 2007–Mar 2012	1.92

4.2 CO₂ yearly growth rate

The same fitting program in the above section, CCGCRV, also calculates a yearly increase. In Table 4 we compare the fitted yearly increase for TCCON, which ranges from 1.92 to 2.55 ppm yr⁻¹ for the different stations and time ranges, to each of the data sets. None of the TCCON/satellite or TCCON/model differences reach significance for the *t* test, which is done when there is more than one result per bin. However, the GOSAT Lamont station is low at –0.3 ppm yr⁻¹, with –0.1 of this estimated to result from co-location error. To see how much of the observed variability in the growth rate is temporal vs. spatial variability in the growth rate, we compare the satellite and model yearly X_{CO₂} increase to the global annual increase (growth rate) from surface measurements (<http://www.esrl.noaa.gov/gmd/ccgg/trends/global.html>) which are 1.74, 2.12, 1.77, 1.67, 2.39, 1.70, 2.40, 2.51, 1.89 ppm yr⁻¹ for global yearly increases 2006–2014, and 1.76, 2.22, 1.60, 1.89, 2.42, 1.86, 2.63, 2.06, 2.17 ppm yr⁻¹ for Mauna Loa yearly increases 2006–2014, with error bars 0.05–0.09 ppm yr⁻¹ for global and 0.11 ppm yr⁻¹ for Mauna Loa. The average global yearly increase predicted from the above using the time periods in Table 4 is shown in the last column of Table 4. The correlation *r* value between the yearly increase in TCCON and the above ESRL global yearly increase is 0.61, and the correlation with the Mauna Loa yearly increase is 0.67, whereas the correlation *r* value between the yearly increase in TCCON and the yearly increase in columns is 0.82. Therefore, the variability seen in Table 4 is partly from the time range used but also partly from variations due to geographic location (and sampling).

Reuter et al. (2011, JGR, Table 2) found agreement within the calculated errors at Park Falls and Darwin for BESD-SCIAMACHY and CT2009 vs. TCCON. However, older data sets were used for this result. Looking specifically at Park Falls, we see 1.80 ± 0.14 and 2.10 ± 0.22

for SCIAMACHY and TCCON, respectively and at Darwin 1.67 ± 0.08 and 2.16 ± 0.05 for SCIAMACHY and TCCON, respectively, where the errors represent the standard deviation of SCIAMACHY fits for similar latitudes.

4.3 Seasonal cycle phase

This section looks at the time offset correlation and standard deviation between the test data sets and TCCON. This checks whether, for example, a seasonal cycle is delayed or ahead of the TCCON seasonal cycle, which has important implications for flux estimates (Keppel-Aleks et al., 2012), whether there are seasonally dependent biases that are affecting the seasonal cycle, and whether the data sets are seeing the same seasonal cycle.

To compare seasonal cycle amplitudes, all data sets have 2 ppm yr⁻¹ subtracted to approximately remove secular increases (over the ±60 days' offset this has a very small effect). For a 0 day offset, the data sets are matched as usual. For a 1 day offset, TCCON is moved forward by 1 day and compared to the data set. This is repeated for all offset times. Correlations are fit to a second-order polynomial to determine the phase minimum difference. As TCCON is moved forward or backward in time, different points will match up, particularly when there are data gaps in either data set. This can cause difficulties in interpretation. The maximum correlation is limited by the ratio of the error to the variability. It follows from the definition of correlation that

$$\text{corr_max} = \text{corr}_o \frac{1}{\sqrt{1 + \left(\frac{\varepsilon_x}{\sigma_x}\right)^2} \sqrt{1 + \left(\frac{\varepsilon_y}{\sigma_y}\right)^2}}, \quad (3)$$

where corr_o is the noise-free correlation, ε_x is the error on *x* and σ_x is the true variability for *x*, ε_y is the error on *y* and σ_y is the true variability for *y*. Because we are estimating σ and ε , there is uncertainty on the correlation maximum. In our case σ_y is taken to be the TCCON variability and ε_y is esti-

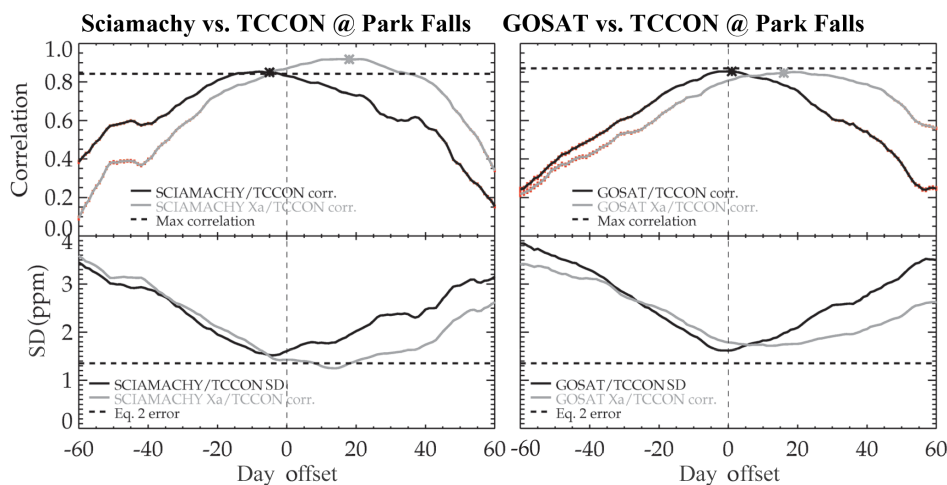


Figure 10. Top: cross correlation between TCCON and SCIAMACHY (top left) and GOSAT (top right) with matches using dynamic criteria at Park Falls. The x axis shows results when satellite data are offset in days vs. TCCON. The dashed line shows the expected maximum correlation based on the error (see Eqs. 2 and 3). The gray line is the correlation for the satellite priors, which are each out of phase by at least 10+ days. The * is the peak of a polynomial fit of the correlation between -25 and $+25$ days. Bottom: standard deviation between TCCON and SCIAMACHY (bottom left) and GOSAT (bottom right). The dashed line shows the Eq. (2) predicted error.

Table 5. The left two numerical columns show standard deviation drop within ± 2 days of zero offset. Higher values with either model indicate sites where temporal co-location is more important. The next two numerical columns show calculated phase difference between CT2013b, MACC, and TCCON in days, with entries larger than 10 in bold. A phase difference of -10 days means that the model seasonal cycle is 10 days behind TCCON. The next four columns show the same calculations for SCIAMACHY and GOSAT prior and retrieved values, with entries larger than 10 in bold. Blank values are those for which a good fit was not found.

Station	Synoptic		Phase difference (days)					
	CT2013b (ppm)	MACC (ppm)	CT2013b	MACC	SCIA prior	SCIA retrieved	GOSAT prior	GOSAT retrieved
Eureka	0.1	0.0	+4	3				
Sodankylä	0.3	0.3	-3	-4				
Bialystok	0.4	0.2	0	+1	15	17	16	4
Bremen ^a	0.2	0.2	-9	-9				
Karlsruhe	0.5	0.3	-4	-4	12	8	16	4
Orléans	0.4	0.3	+5	-2	8	10	5	-1
Garmisch ^a	0.4	0.3	-6	-7	9	9	7	0
Park Falls	0.5	0.3	-1	-5	18	-6	16	1
Four Corners ^a	0.1	0.1	-9	-10	-25	-17	-18	-10
Lamont	0.5	0.3	-6	-5	-13	-14	-10	-3
Tsukuba	0.4	0.2	+8	+2				
Edwards ^a		0.0		-5			-27	-4
JPL2007 ^a	0.1	0.0	+3	1				
JPL2011 ^a	0.1	0.0	-4	-5			-21	9
Pasadena ^a	0.0	0.0	-3	+7			-9	17
Saga	0.2	0.0	+5	-1			-1	-2
Izaña ^a	0.0	0.0	+10	-4			-22	0
Darwin	0.0	0.1	+6	-19				
Reunion	0.1	0.1	-10	3				
Wollongong	0.1	0.07		+6			-11	4
Lauder_125HR	0.1	0.1	-20	+0				

^a indicates TCCON stations which are locally influenced (see Sect. 2.1).

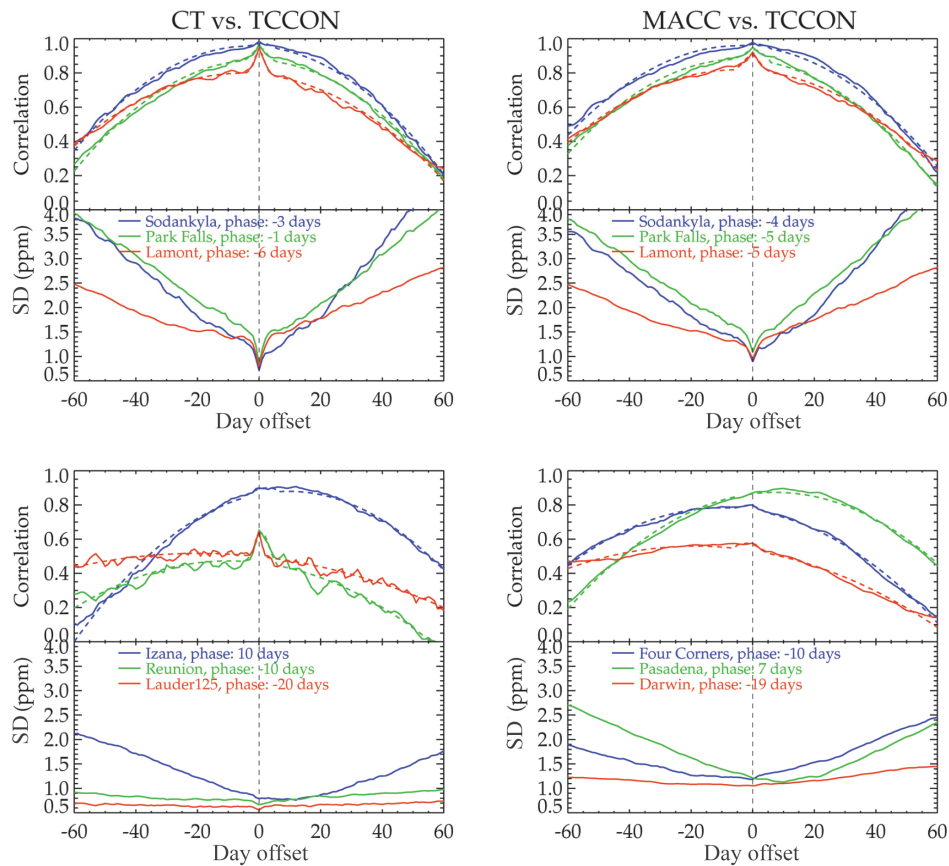


Figure 11. Top: cross correlation examples between TCCON and CT2013 (left) or MACC (right). Each panel shows the correlation and second-order polynomial fits (top) and standard deviation (bottom) vs. offset in days of TCCON vs. satellite data. The correlation should be at a maximum and standard deviation at a minimum at days offset = 0. The top panels show examples of stations with a phase difference of less than 10 days. The bottom set shows example stations which have phase differences of at least 10 days.

mated using Table 2 with 10 SCIAMACHY and 4 GOSAT averages. The error bars on the correlations are calculated from Fisher's z test. (Fisher, 1915, 1921). Figure 10 shows SCIAMACHY and GOSAT results at Park Falls. Although the prior performs well in regards to the standard deviation vs. TCCON, Fig. 10 shows the prior has a clear seasonal cycle phase error which is corrected by the satellite retrievals for both SCIAMACHY and GOSAT at Park Falls.

Results of the seasonal cycle phase error are tabulated in Table 5. Stations not shown have either too few match-ups (e.g., Sodankylä) or too little variability compared to the noise (e.g., Wollongong) to have useful comparisons. The GOSAT retrieval markedly improves over the prior seasonal cycle phase vs. TCCON at 12 out of 13 stations. For the six stations that are not locally influenced (with no ^a by the station name), GOSAT improves over the prior at all six stations, and additionally has a smaller phase difference than one or more of the models at four of the six stations, is the same at one station, and is worse at one station. SCIAMACHY improves over its prior for Park Falls and Four Corners, mildly improves at Karlsruhe, and stays the same

or gets slightly worse in four cases. Mismatches in SCIAMACHY phase could be from mismatches in vertical sensitivity (as higher altitudes have lagged seasonal cycles), effects of coincidence criteria, or seasonal-dependent biases. To check the coincidence criteria, cross-correlations were calculated for the geometric coincidence criteria which had significantly fewer match-ups. Similar results for geometric coincidence criteria were found for GOSAT and SCIAMACHY for Lamont and Park Falls; the other stations are too noisy to draw conclusions.

Table 5 also shows the phase differences for the models, which have closer spatial/temporal matches and lower single-match-up errors. Model–TCCON phase differences could result from errors in model flux distributions, seasonal timing, or transport errors. Table 5 shows the phase differences, which vary from -20 to $+10$ days. Phase differences more than 10 days are noticeable by eye when looking at time series data and occur in the NH at Bremen, Four Corners (negative), Orléans, and Izaña (positive, CT2013b only), although Bremen, Four Corners, and Izaña are locally influenced. Large phase differences also occur at some stations

Table 6. Diurnal variability of CT2013b and MACC13.1 vs. TCCON in JJA arranged by latitude. TCCON variability and maximum theoretical correlation are shown, as well as actual correlation and slope for both models. The slope is the mode vs. TCCON fit to a straight line. ^a indicates sites expected to be strongly influenced by local sources. Max correlation is calculated using Eq. (3), TCCON SD, and the lower of the model–TCCON standard deviations from Appendix B. TCCON range, SD, shows the range of the values seen for the daily variability for TCCON, as well as the standard deviation of the TCCON daily differences. The average row gives averages for all entries above it. The average rows, e.g., DJF, are for stations Bialystok, Orléans, Park Falls, and Lamont. Eureka and Karlsruhe are not included because fewer than five matches were found.

Station	Lat. (deg)	TCCON	Max corr.	CT2013b		MACC 13.1	
		Range, SD (ppm, ppm)	(Eq. 3)	Correlation	Slope	Correlation	Slope
Sodankylä	67	−1.5 to 1.5, 0.5	0.6	0.35	0.23	0.40	0.19
Bialystok	53	−2 to +1, 0.5	0.6	0.65	0.55	0.66	0.36
Bremen ^a	53	−2 to +1, 0.7	0.5	0.24	0.06	−0.20	−0.03
Orléans	48	−1 to +0.5, 0.7	0.7	0.72	0.51	0.70	0.31
Garmisch ^a	47	−3 to 0, 0.8	0.6	0.55	0.50	0.69	0.45
Park Falls	46	−3.5 to +2, 1.0	0.8	0.77	0.32	0.54	0.23
Four Corners ^a	37	−4.5 to +0.5, 1.8	0.9	0.39	0.03	0.22	0.02
Lamont	37	−3.0 to +1.0, 0.8	0.7	0.28	0.19	0.32	0.14
Tsukuba	36	−2 to +1.5, 0.8	0.6	0.79	0.47	0.64	0.24
JPL2011 ^a	34	−1 to 1, 0.5	0.5	0.18	0.06	−0.05	−0.01
Saga	33	−0.5 to 0, 0.3	0.4	−0.14	−0.07	−0.11	−0.06
Izaña ^a	28	−0.5 to +0.5, 0.1	0.2	0.73	0.11	0.45	0.04
Darwin	−12	−0.5 to +0.5, 0.2	0.3	0.04	0.03	0.10	0.03
Reunion	−21	−0.3 to 0, 0.2	0.3	−0.33	−0.04	0.72	0.10
Wollongong	−34	−1.5 to +1.0, 0.4	0.5	0.33	0.47	0.27	0.15
Lauder_120HR	−45	−0.5 to +0.5, 0.3	0.3	0.05	0.01	0.20	0.02
Lauder_125HR	−45	−0.5 to +0.5, 0.2	0.4	0.41	0.23	0.30	0.06
Average JJA for all stations			0.8	0.30	0.21	0.40	0.15
Average DJF (see caption for station list)			0.8	0.52	0.16	0.54	0.17
Average MAM (see caption for station list)			0.8	0.59	0.29	0.43	0.15
Average JJA (see caption for station list)			0.8	0.51	0.39	0.56	0.26
Average SON (see caption for station list)			0.8	0.18	0.09	0.19	0.06

in the Southern Hemisphere. Although the seasonal cycle is weaker in the Southern Hemisphere, the phase offset can be seen in the CT2013b Lauder_125HR plots in Fig. 2. The correlations vs. offset days show a phase difference of −20 days for CT2013b and +0 days for MACC at Lauder_125HR, as seen in Fig. 11. Note that the fits of the seasonal cycle in the SH display more complexity, such as multiple local maximum, than fits in the NH and “phase lag” could easily be an indication of an issue with the fit shape. Figure 11 shows correlations and standard deviations vs. day offset for three stations that have the seasonal cycle peak within ± 10 days for CT2013b and MACC (top panels), and for stations which have a larger phase lag compared to TCCON (bottom panels). There is often a small peak within ± 3 days, which indicates the models’ capability of picking up variations that occur day to day (i.e., synoptic-scale variability), which indicates both the strength of synoptic activity and matching between models and TCCON. This peak is not seen in satellite data for dynamic coincidence criteria likely due to matching, or geometric coincidence criteria likely due to the noise

– not that this synoptic peak occurs at 0, even when the seasonal cycle has a phase lag (e.g., MACC model at Bremen, in the lower right panel, or Lauder_125HR comparisons). The synoptic-scale correlation varies between 0 and 0.17, as seen in Table 5.

Izaña will be briefly discussed. As noted in Sect. 2.1, the TCCON station is located on a small island at 2.37 km above sea level (about 770 hPa), whereas the MACC and CT2013b models at $\sim 2^\circ \times 3^\circ$ resolution do not resolve topography and consequently have mean surface pressure at sea level, about 1013 hPa at this location. Our standard treatment is to interpolate the model to the TCCON pressure grid, then calculate X_{CO_2} using the TCCON pressure weighting function. At Izaña this has the effect of chopping off the lower atmosphere. The CT2013b result for this treatment has a +10 day seasonal cycle phase difference at Izaña; whereas MACC has no phase difference at Izaña. If, however, the model surface pressure is used to calculate X_{CO_2} , MACC goes from a 0 to a −10 day phase lag, and CT2013b has 0 phase difference. An argument for using the model surface pressure

would be if the upslope winds at Izaña (Sancho et al., 1991; Bergamaschi et al., 2000) shifted the profile upwards rather than chopping it off, which would occur if the air deviated around the island instead. This finding has important implications for the choice of the comparison methodology and the ideal location for validation sites. Validation sites within complex geographical terrain have to be treated as special cases as (a) the atmospheric models usually do not resolve these variations and (b) satellite measurements rarely have a perfect co-location with the ground-based site, meaning that they could sample a substantially different altitude level. This holds for both mountains (e.g., Izaña) and valleys (e.g., Garmisch). This highlights one of the many choices that are made when comparing two products (e.g., whether to apply the averaging kernel, whether to use interpolation, how to treat the surface pressure, or what coincidence criteria to use).

Another finding worth noting is the comparisons at Lauder. In 2010 the Lauder_125HR instrument began routine operation, while the Lauder_120HR instrument continued to take TCCON data through to the end of 2010. Both MACC and CT2013b show no seasonal cycle correlation with the 120HR time series at Lauder, but do show correlation with Lauder_125HR time series. We attribute this to the improved precision of the 125HR data, and an increase of the seasonal cycle amplitude in 2011 and 2012 as compared to other years (e.g., compare 2011 vs. 2007). The phasing error found in the CT2013b comparison with the Lauder_125HR may be due to CT2013b not modeling the drivers of the seasonal cycle amplification in 2011 and 2012.

At Bremen and Four Corners, local effects that are not resolved at $3 \times 2^\circ$ likely dominate, particularly since Bremen is clustered with Orléans, Garmisch, and Karlsruhe, which all compare fine, and because the correlation of daily variability, as seen in the next section, is also very low at these two stations.

5 Daily variability (models vs TCCON)

At the surface, CO_2 shows a strong diurnal cycle in areas with active vegetation, e.g., Park Falls, during summer, and synoptic trends based on regional dynamics. Even though the diurnal cycle is markedly smaller in the total column (Olsen and Randerson, 2004), it can be observed both by TCCON and also in models, in our case CT2013b and MACC, as seen in Fig. 12. Both diurnal variations and synoptic trends can be seen in Fig. 12. Validating the amplitude of the diurnal variability in the column is important as the column diurnal variability better represents the amount of CO_2 emitted or absorbed by surface processes as compared to surface measurements, which are more impacted by boundary layer height. To our knowledge this is the first comparison of model fields to TCCON to compare the diurnal cycle. As TCCON itself has not been validated at multiple times in 1 day, this is considered a comparison, not a validation. We

compare the difference between morning and afternoon in models and TCCON. To minimize potential TCCON biases that depend on the solar zenith angle (through the air mass factor), we compare at two points in each day separated by the largest time with the same solar zenith angle (SZA). The methodology is to (1) identify two points, t_1 and t_2 , from the same day with the largest time difference but with the same SZA. As the TCCON data used in this paper have been averaged over 90 min, t_1 or t_2 may be interpolated between two time points. (2) We compare TCCON at t_2 minus TCCON at t_1 and the same times for each model. We look at the variability within 1 day for one season (JJA). Looking at different seasons for the Northern Hemisphere at the bottom of Table 6, both models showed clearly higher correlations and slopes in MAM and JJA vs. DJF and SON, with SON showing the lowest correlation and slope. In the SH, correlations are highest in DJF, second highest in SON, and lowest in MAM. However in the SH, the slopes are on the order of 15 % of the slope of TCCON. For CT2013b, at Darwin, correlations are highest in DJF, less in SON, with slopes on the order of 15 % of the slope of TCCON. Table 6 shows correlations between CT2013b or MACC vs. TCCON in the daily variability. In the NH, on average, the correlations are about two thirds as large as could be expected, given the relative sizes of the variability and errors (see text around Eq. 3). Additionally, the two models have about one third to one half the daily variability of TCCON (as seen from the smaller slopes). In the far north (Eureka, Sodankylä), the correlations indicate agreement but the model daily variability is less than 1/4 TCCON. In the midlatitudes (excluding locally influenced stations) there is the highest correlation (~ 0.3 – ~ 0.7) with model daily variability 20 to 60 % of the variability of TCCON. Sites influenced by local sources, Bremen, Four Corners, and JPL2011, the models do not show the diurnal variability that is seen by TCCON, as evidenced by very small or zero slopes. Sites which have complicated terrain, Garmisch and Izaña, do show correlations in the daily variability similar to other nearby stations.

The CT2013b model in general shows more daily variability and higher correlations, which are in better agreement with TCCON. Since the satellite observations are coincident \sim once per day, the diurnal pattern will not be well constrained by satellite observations, except as preserved in transported air coincident with satellite measurements downwind. Model Observing System Simulation Experiments (OSSEs) can determine the impact of the diurnal cycle strength on flux estimates to determine the importance of independently verifying the diurnal cycle in models.

6 Discussion and conclusions

We find standard deviations of 0.9, 0.9, 1.7, and 2.1 ppm vs. TCCON for CT2013b, MACC, GOSAT, and SCIAMACHY, respectively, with the single target errors 1.9 and 0.9 times the predicted errors for GOSAT and SCIAMACHY, respec-

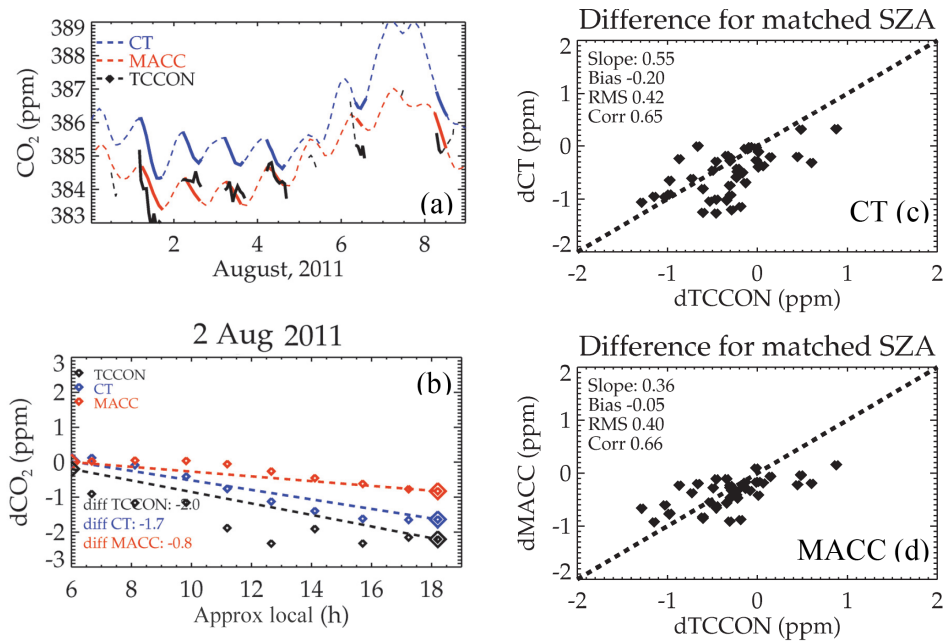


Figure 12. (a) Time series for models and TCCON from 1 to 8 August 2011 at Bialystok. The end points of the solid lines show the time points used for comparing daily variability; both diurnal and synoptic variations are seen. (b) Change throughout the day at Bialystok for 2 August 2011. The large diamonds at 06:00 and 18:00 show the two times with largest difference that have the same solar zenith angle that are used in the analysis. The difference between these times are -2 , -1.7 , and -0.7 ppm for TCCON, CT2013, and MACC, respectively. Right: CarbonTracker (c) and MACC (d) daily trends vs. TCCON daily trends for BIALYSTOK in JJA, from compiling differences as shown in (b). Correlation is seen in the daily trends as compared to TCCON with the daily amplitude for the models smaller than TCCON.

tively. There is a correlation r value of 0.5 for SCIAMACHY and 0.6 for GOSAT for the actual and predicted errors grouped by station. Equation (2) and Table 2 show how errors decrease when satellite results are averaged and estimate the magnitude of the correlated and random errors components for averaged satellite results, where random error components decrease with an increasing number of averaged observations. When satellite data are averaged and interpreted according to the model error² = $a^2 + b^2/n$ (where n is the number of observations averaged, a is the systematic (correlated) errors, and b is the random (uncorrelated) errors). The northern hemispheric average values from Table 2 are $a = 1.4 \pm 0.3$ ppm, $b = 1.7 \pm 0.2$ ppm for SCIAMACHY geometric; $a = 1.0 \pm 0.3$ ppm, $b = 2.1 \pm 0.1$ ppm for SCIAMACHY dynamic; $a = 0.8 \pm 0.2$ ppm, $b = 1.6 \pm 0.1$ ppm for GOSAT geometric; $a = 0.5 \pm 0.2$ ppm, $b = 1.9 \pm 0.1$ ppm for GOSAT dynamic. The lowest correlated errors are found when using dynamic coincidence criteria where values are averaged from a larger spatiotemporal region, but the lower value for GOSAT is only after the estimated co-location error is subtracted. The Southern Hemisphere errors are uniformly smaller, both in correlated and uncorrelated errors. These data represent averaging of satellite data which match one averaged TCCON value. The above error model should help in assigning realistic retrieval error correlations in assimilation systems in place of current ad hoc hypotheses. For

example, in Basu et al. (2013) observations within 500 km and 1 h are assumed to have 100 % correlated errors, and are inflated by a factor such that when observations are later treated as if the errors were random, the final error of the average is the same as the error of one observation. This can be improved by setting the inflation factor so that the average observation error is $a^2 + b^2/n$, with a and b set by the geometric values from Table 2, which should result in a lower error than assuming 100 % correlation.

Regarding the quality of the dynamic relative to the geometric coincidence criteria, the coincidence error estimated by models is larger for dynamic coincidences by about 0.3 ppm, as seen in Table 2. However, the coincidence error (0.3 to 0.4 ppm for geometric criteria and 0.6 to 0.7 ppm for dynamic criteria) is not the dominant error. As seen in Fig. 5, the dynamic coincidence criteria average 0.1 ppm higher error for unaveraged satellite comparisons. This is small compared to a total error of 2.0 and 2.2 ppm, respectively, for stations in the Northern Hemisphere. With maximum averaging, as seen in Fig. 5, the errors are lower for dynamic vs. geometric because the dynamic criteria finds more observations to average. Figure 6 shows that at Lamont the average difference between geometric and dynamic observations is 0.4 ppm for unaveraged satellite observations, which is higher than average. This error difference reduces to less than 0.2 ppm when all available observations are averaged, also

seen in Fig. 6. While coincidence error is an important error source, it is not the dominant error source. Although dynamic coincidence criteria allow the inclusion of more stations in analyses because of the larger number of coincidences, comparisons to geometric coincidence results are done when possible. Biases at individual stations have a year-to-year variability of ~ 0.3 ppm, with biases larger than the TCCON predicted bias uncertainty of 0.4 ppm at many stations. Using fitting software, we find that GOSAT and CT2013b underpredict the seasonal cycle amplitude in the Northern Hemisphere between 46 and 53° N, MACC overpredicts in 26–37° N, and CT2013b underpredicts the seasonal cycle amplitude in the Southern Hemisphere. The seasonal cycle phase indicates whether a data set or model lags another data set in time. We find that the GOSAT phase improves substantially over the prior and the SCIAMACHY retrieved phase improves substantially for two of seven sites. The models reproduce the measured seasonal cycle phase well except for at Lauder_125HR (CT2013b) and Darwin (MACC). We compare the variability within 1 day between TCCON and models in JJA; there is correlation between 0.2 and 0.8 in the NH, with models showing 10 to 50 % the variability of TCCON at different stations and CT2013b showing more variability than MACC. This paper highlights findings that provide inputs to estimate flux errors in model assimilations, and places where models and satellites need further investigation, e.g., the SH for models and 45–67° N for GOSAT and CT2013b.

We focus on validating aspects of model and satellite data which may be important for accurate flux estimates and CO₂ assimilation, including accurate error estimates, overall biases, biases by season and latitude, impact of coincidence criteria, validation of seasonal cycle phase and amplitude, yearly growth, and daily variability. Our methodologies can be used to correct known biases and data deficiencies (e.g., Basu et al. (2013) accounted for global land/sea biases; Nassar et al. (2011) corrected for hemispheric gradients). Alternatively, biases can be mitigated through data assimilation, such as the inversion method of Reuter et al. (2014) which is insensitive to seasonal and regional biases outside a targeted region. The bias evaluation performed in the present study has been restricted to various satellite and model CO₂ data products. To quantify the importance of each bias (seasonal biases, location-dependent biases, seasonal cycle differences, seasonal cycle phase differences, and diurnal cycle differences) on carbon flux estimates requires detailed Observing System Simulation Experiments (OSSEs). For example, Kulawik et al. (2013) tested the effect of a NH bias of 0.3–0.5 ppm in JJA, finding flux biases were not insignificant, and were comparable to flux updates from GOSAT in some regions.

Biases vary by station (See Fig. 3); these station-dependent biases have a standard deviation of ~ 0.3 ppm from year to year. Biases also vary by season as seen in Figs. 7–8. Seasonal biases affect the seasonal cycle amplitudes, which are important for biospheric flux attribution.

All sets show the same general patterns for the different latitude bands (SH, 28–37, 46–53, 67–79° N). The statistically significant differences are (1) a low bias in CT2013b seasonal cycle amplitude of 0.2 ± 0.1 ppm for 45–53° N caused by high values in JJA, (2) a high bias in the MACC seasonal cycle amplitude for 46–53° N of 0.6 ppm, (3) a low bias in the CT2013b Southern Hemisphere seasonal cycle amplitude, and (4) a low bias in the GOSAT seasonal cycle amplitude of 0.4 ± 0.1 ppm for 46–53° N, also seen in Lindqvist et al. (2015), caused by a combination of higher values in JJA and lower values in MAM vs. TCCON. A preliminary study of how a seasonal bias in JJA in GOSAT of 0.5 ppm in the NH affect fluxes using a global assimilation showed that the effect was not minor (Kulawik et al., 2013).

The seasonal cycle phase is a sensitive indicator of seasonally dependent biases in satellite data as well as issues with model fluxes or transport errors. The GOSAT root mean square (RMS) phase difference vs. TCCON across all sites is 16.1 days for the prior; this improves to 6.8 days for the GOSAT retrieved XCO₂. The SCIAMACHY RMS phase difference vs. TCCON across all sites is 16.4 days for the prior; this improves to 13.2 days for the SCIAMACHY retrieved XCO₂, reflecting the fact that SCIAMACHY data significantly improved the seasonal cycle phase at just two of the seven TCCON sites.

Model comparisons to TCCON are much less noisy as there are many more matches. Most NH stations show the expected seasonal drop-off (e.g., see Fig. 11), with the peak correlation near 0 days, and an additional spike within ± 3 days indicating the capture of synoptic variability. Stations that showed phase differences larger than 10 days are Darwin (Macc only), and Lauder_125HR (CT2013b only), as well as all of the locally influenced stations.

In comparing CO₂ diurnal variability between TCCON and models, both models show up average 0.5 correlation to the TCCON CO₂ change between afternoon and morning for select stations in the NH. This is on the order of about 2/3 of the maximum possible correlation given the error vs. variability (except the locally influenced JPL2011, Bremen, and Four Corners which had little correlation and no slope). The amplitude of the variability is higher in TCCON vs. the models, with CT2013b closer to TCCON than MACC. However, note that TCCON daily variability has not been validated. Diurnal pattern will not be constrained by satellite observations, except as preserved in transported air coincident with satellite measurements downwind, and therefore may be important to independently verify the diurnal cycle in models to ensure accurate flux attribution. The importance of the diurnal cycle on flux estimates needs to be tested.

Spatial and seasonal-dependent biases are obstacles to accurate and better resolved CO₂ flux estimates. This paper highlights findings that provide inputs to estimate flux errors in model assimilations, and places where models and satellites need additional validation or improvement. Some of the issues which need further investigation are the GOSAT

and CT2013b seasonal cycles in the 46–53° N latitude range (which are 0.4 and 0.2 ppm smaller than TCCON, respectively), the MACC seasonal cycle in 28–37° N latitude range (which is 0.6 ppm larger than TCCON), seasonal cycle amplitude and phase differences at SH stations, differences in the diurnal cycle amplitude between models and TCCON, and high biases for GOSAT and SCIAMACHY north of 67° N.

Appendix A

Table A1. Coincidence criteria, data versions, and terminology used in our analysis.

Coincidence criteria	
Geometric	5° in lat./long., ± 1 h
Dynamical	from Wunch et al. (2011b); considers free-tropospheric temperature, ± 10° lat, ± 30° long. and 5 days (see Sect. 3.1)
Averaging: all averaging is done by station first and then averaging over station results	
Model choices: MACC is interpolated to TCCON latitude, longitude, and time; CT2013b special output is interpolated to TCCON time; model X _{CO₂} uses TCCON surface pressure	
Data sets	
GOSAT	ACOS-GOSAT version 3.5 with corrections and quality flags from the user's guide
SCIAMACHY	BESD-SCIAMACHY v02.00.08
TCCON	GGG2014 when available, GGG2012 data from Four Corners, Tsukuba, and Bremen with GGG2012 bias corrections applied as described in Sect. 2.3
CarbonTracker	CT2013b
MACC	MACC v13.1

Appendix B: TCCON and model errors by station

Table B1. We estimate the 90 min average TCCON standard deviation error by calculating the standard deviation of adjacent time points and model standard deviation of CT2013b and MACC13.1 vs. TCCON by station. These values are used to estimate theoretical maximum correlations for seasonal cycle and diurnal correlations using Eq. (3). ^a denotes stations strongly influenced by local effects and ^b Tsukuba has higher TCCON error. These stations have been removed in the NH average (subset) row.

90 min average TCCON errors and model error vs. TCCON by station			
Station	TCCON adjacent standard deviation (ppm)	CT2013b standard deviation (ppm)	MACC standard deviation (ppm)
Eureka	0.2	0.8	0.9
Ny Alesund	0.8	0.8	0.8
Sodankylä	0.3	0.7	0.8
Bialystok	0.3	0.7	0.7
Bremen ^a	0.5	1.3	1.4
Karlsruhe	0.4	0.9	0.9
Orléans	0.3	0.7	0.7
Garmisch	0.3	0.9	0.9
Park Falls	0.4	0.8	0.8
Four Corners ^a	0.7	1.1	1.1
Lamont	0.3	0.8	0.8
Tsukuba ^b	0.9	1.0	1.1
JPL2007 ^a	0.6	1.3	1.1
Izaña	0.2	0.8	0.6
NH average	0.44	0.91	0.92
NH average (subset)	0.35	0.79	0.79
Darwin	0.4	0.9	1.0
Wollongong	0.4	0.8	0.7
Lauder_120HR	0.4	0.8	0.8
Lauder_125HR	0.3	0.6	0.5
SH average	0.38	0.78	0.75

Author contributions. Susan Kulawik set the direction of the research and did much of the analysis. The following authors were involved with discussions of results with specific knowledge in the listed areas: Debra Wunch, TCCON, Christopher O'Dell, ACOS-GOSAT, Christian Frankenberg, ACOS-GOSAT, Maximilian Reuter, BESD-SCIAMACHY, Tomohiro Oda, CarbonTracker, Frederic Chevallier, MACC, Vanessa Sherlock, TCCON, Michael Buchwitz, BESD-SCIAMACHY, Greg Osterman, ACOS-GOSAT, Charles Miller, CO₂ data records. The TCCON data providers, who also provide expertise regarding TCCON sites are Paul Wennberg, David Griffith, Isamu Morino, Manvendra Dubey, Nicholas M. Deutscher, Justus Notholt, Frank Hase, Thorsten Warneke, Ralf Sussmann, John Robinson, Kimberly Strong, and Matthias Schneider. Joyce Wolf is a science programmer and provided technical expertise.

Acknowledgements. Funded by NASA Roses ESDR-ERR 10/10-ESDRERR10-0031, "Estimation of biases and errors of CO₂ satellite observations from AIRS, GOSAT, SCIAMACHY, TES, and OCO-2".

Maximilian Reuter and Michael Buchwitz received funding from ESA (GHG-CCI project of ESA's Climate Change Initiative) and from the University and state of Bremen.

Information about all TCCON sites and their sources of funding can be found on the TCCON website (<https://tcccon-wiki.caltech.edu/>).

Manvendra K. Dubey is grateful for the funding for monitoring at Four Corners by LANL-LDRD, 20110081DR.

Frédéric Chevallier received funding from the EU H2020 Programme (grant agreement no. 630080, MACC III). NCEP Reanalysis data used in dynamic coincidence criteria were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at <http://www.esrl.noaa.gov/psd/>. Thanks to Andrew R. Jacobson for help with CarbonTracker.

Edited by: H. Worden

References

- Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak, B., Taguchi, S., and Zhu, Z.: TransCom 3 inversion intercomparison: impact of transport model errors on the interannual variability of regional CO₂ fluxes, 1988–2003, *Global Biogeochem. Cy.*, 20, GB1002, doi:10.1029/2004GB002439, 2006.
- Basu, S., Houweling, S., Peters, W., Sweeney, C., Machida, T., Maksyutov, S., Patra, P. K., Saito, R., Chevallier, F., Niwa, Y., Matsueda, H., and Sawa, Y.: *J. Geophys. Res.*, 116, 1–14, doi:10.1029/2011JD016124, 2011.
- Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO₂ fluxes estimated from GOSAT retrievals of total column CO₂, *Atmos. Chem. Phys.*, 13, 8695–8717, doi:10.5194/acp-13-8695-2013, 2013.
- Bergamaschi, P., Bräunlich, M., Marik, T., and Brenninkmeijer, C. A. M.: Measurements of the carbon and hydrogen isotopes of atmospheric methane at Izaña, Tenerife: Seasonal cycles and synoptic-scale variations, *J. Geophys. Res.*, 105, 14531–14546, doi:10.1029/1999JD901176, 2000.
- Butz, A., Guerlet, S., Jacob, D. J., Schepers, D., Galli, A., Aben, I., Frankenberg, C., Hartmann, J.-M., Tran, H., Kuze, A., Keppel-Aleks, G., Toon, G. C., Wunch, D., Wennberg, P. O., Deutscher, N. M., Griffith, D. W. T., Macatangay, R., Messerschmidt, J., Notholt, J., and Warneke, T.: *Geophys. Res. Lett.*, 38, 2–7, doi:10.1029/2011GL047888, 2011.
- Canadell, J. G., Ciais, P., Gurney, K., Le Quéré, C., Piao, S., Raupach, M. R., and Sabine, C. L.: An international effort to quantify regional carbon fluxes, *EOS Trans. AGU*, 92, 81–82, 2011.
- Chevallier, F. and O'Dell, C. W.: Error statistics of Bayesian CO₂ flux inversion schemes as seen from GOSAT, *Geophys. Res. Lett.*, 40, 1252–1256, doi:10.1002/grl.50228, 2013.
- Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A. J., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P., Langenfelds, R., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morguí, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO₂ surface fluxes at grid point scale estimated from a global 21-year reanalysis of atmospheric measurements, *J. Geophys. Res.*, 115, D21307, doi:10.1029/2010JD013887, 2010.
- Chevallier, F., Deutscher, N. M., Conway, T. J., Ciais, P., Ciattaglia, L., Dohe, S., Fröhlich, M., Gomez-Pelaez, A. J., Griffith, D. W. T., Hase, F., Haszpra, L., Krummel, P., Kyrö, E., Labuschagne, C., Langenfelds, R., Machida, T., Maignan, F., Matsueda, H., Morino, I., Notholt, J., Ramonet, M., Sawa, Y., Schmidt, M., Sherlock, V., Steele, P., Strong, K., Sussmann, R., Wennberg, P. O., Wofsy, S. C., Worthy, D., Wunch, D., and Zimnoch, M.: *Geophys. Res. Lett.*, 38, 1–5, doi:10.1029/2011GL049899, 2011.
- Chevallier, F., Palmer, P. I., Feng, L., Boesch, H., O'Dell, C. W., and Bousquet, P.: Toward robust and consistent regional CO₂ flux estimates from in situ and space-borne measurements of atmospheric CO₂, *Geophys. Res. Lett.*, 41, 1065–1070, doi:10.1002/2013GL058772, 2014.
- Ciais, P., Dolman, A. J., Bombelli, A., Duren, R., Peregon, A., Rayner, P. J., Miller, C., Gobron, N., Kinderman, G., Marland, G., Gruber, N., Chevallier, F., Andres, R. J., Balsamo, G., Bopp, L., Bréon, F.-M., Broquet, G., Dargaville, R., Battin, T. J., Borges, A., Bovensmann, H., Buchwitz, M., Butler, J., Canadell, J. G., Cook, R. B., DeFries, R., Engelen, R., Gurney, K. R., Heinze, C., Heimann, M., Held, A., Henry, M., Law, B., Luysaert, S., Miller, J., Moriyama, T., Moulin, C., Myneni, R. B., Nussli, C., Obersteiner, M., Ojima, D., Pan, Y., Paris, J.-D., Piao, S. L., Poulter, B., Plummer, S., Quegan, S., Raymond, P., Reichstein, M., Rivier, L., Sabine, C., Schimel, D., Tarasova, O., Valentini, R., Wang, R., van der Werf, G., Wickland, D., Williams, M., and Zehner, C.: Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system, *Biogeosciences*, 11, 3547–3602, doi:10.5194/bg-11-3547-2014, 2014.
- Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bösch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Man-

- drake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L.: The ACOS CO₂ retrieval algorithm – Part II: Global X_{CO₂} data characterization, *Atmos. Meas. Tech.*, 5, 687–707, doi:10.5194/amt-5-687-2012, 2012.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, doi:10.1002/qj.828, 2011.
- Deng, F., Jones, D. B. A., Henze, D. K., Bousserez, N., Bowman, K. W., Fisher, J. B., Nassar, R., O'Dell, C., Wunch, D., Wennberg, P. O., Kort, E. A., Wofsy, S. C., Blumenstock, T., Deutscher, N. M., Griffith, D. W. T., Hase, F., Heikkinen, P., Sherlock, V., Strong, K., Sussmann, R., and Warneke, T.: Inferring regional sources and sinks of atmospheric CO₂ from GOSAT X_{CO₂} data, *Atmos. Chem. Phys.*, 14, 3703–3727, doi:10.5194/acp-14-3703-2014, 2014.
- Dils, B., Buchwitz, M., Reuter, M., Schneising, O., Boesch, H., Parker, R., Guerlet, S., Aben, I., Blumenstock, T., Burrows, J. P., Butz, A., Deutscher, N. M., Frankenberg, C., Hase, F., Hasekamp, O. P., Heymann, J., De Mazière, M., Notholt, J., Sussmann, R., Warneke, T., Griffith, D., Sherlock, V., and Wunch, D.: The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparative validation of GHG-CCI SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT CO₂ and CH₄ retrieval algorithm products with measurements from the TCCON, *Atmos. Meas. Tech.*, 7, 1723–1744, doi:10.5194/amt-7-1723-2014, 2014.
- Efron, B.: Bootstrap Methods – Another Look at the Jackknife, *Ann. Stat.*, 7, 1–26, 1979.
- Fisher, R. A.: Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population, *Biometrika*, 10, 507–521, 1915.
- Fisher, R. A.: On the probable error of a coefficient of correlation deduced from a small sample, *Metron*, 1, 3–32, 1921.
- Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K.-G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., and Zeng, N.: Climate-Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison, *J. Climate*, 19, 3337–3353, doi:10.1175/JCLI3800.1, 2006.
- Guerlet, S., Butz, A., Schepers, D., Basu, S., Hasekamp, O. P., Kuze, A., Yokota, T., Blavier, J.-F., Deutscher, N. M., Griffith, D. W. T., Hase, F., Kuro, E., Morino, I., Sherlock, V., Sussmann, R., Galli, A., and Aben, I.: Impact of aerosols and thin cirrus on retrieving and validating XCO₂ from GOSAT shortwave infrared measurements, *J. Geophys. Res.-Atmos.*, 118, 4887–4905, doi:10.1002/jgrd.50332, 2013.
- Houghton, R. A.: Balancing the Global Carbon Budget, *Annu. Rev. Earth Pl. Sc.*, 35, 313–347, doi:10.1146/annurev.earth.35.031306.140057, 2007.
- Hourdin, F., Musat, I., Bony, S., Braconnot, P., Codron, F., Dufresne, J.-L., Fairhead, L., Filiberti, M.-A., Friedlingstein, P., Grandpeix, J., Krinner, G., LeVan, P., Li, Z.-X., and Lott, F.: The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection, *Clim. Dynam.*, 27, 787–813, doi:10.1007/s00382-006-0158-0, 2006.
- Hungershofer, K., Breon, F.-M., Peylin, P., Chevallier, F., Rayner, P., Klonecki, A., Houweling, S., and Marshall, J.: Evaluation of various observing systems for the global monitoring of CO₂ surface fluxes, *Atmos. Chem. Phys.*, 10, 10503–10520, doi:10.5194/acp-10-10503-2010, 2010.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, *B. Am. Meteorol. Soc.*, 77, 437–470, 1996.
- Keppel-Aleks, G., Wennberg, P. O., and Schneider, T.: Sources of variations in total column carbon dioxide, *Atmos. Chem. Phys.*, 11, 3581–3593, doi:10.5194/acp-11-3581-2011, 2011.
- Keppel-Aleks, G., Wennberg, P. O., Washenfelder, R. A., Wunch, D., Schneider, T., Toon, G. C., Andres, R. J., Blavier, J.-F., Connor, B., Davis, K. J., Desai, A. R., Messerschmidt, J., Notholt, J., Roehl, C. M., Sherlock, V., Stephens, B. B., Vay, S. A., and Wofsy, S. C.: The imprint of surface fluxes and transport on variations in total column carbon dioxide, *Biogeosciences*, 9, 875–891, doi:10.5194/bg-9-875-2012, 2012.
- Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener, F., and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, *Atmos. Chem. Phys.*, 5, 417–432, doi:10.5194/acp-5-417-2005, 2005.
- Kulawik, S. S., Baker, D., Wunch, D., Jacobson, A. R., Oda, T., Frankenberg, C., Reuter, M., Buchwitz, M., Osterman, G., O'Dell, C., Wennberg, P., Olsen, E., Griffith, D., Sherlock, V., Deutscher, N. M., Notholt, J., Warneke, T., Morino, I., Sussmann, R., and Wolf, J.: Spatial and Temporal Biases in ACOS-GOSAT and BESD-SCIAMACHY Carbon Dioxide and Effects on Flux Estimates, American Geophysical Union Fall Meeting, 9–13 December 2013, San Francisco, CA, USA, 2013.
- Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, *Appl. Opt.*, 48, 6716–6733, 2009.
- Lindenmaier, R., Dubeya, M. K., Henderson, B. G., Butterfield, Z. T., Hermanc, J. R., Rahna, T., and Leea, S.-H.: Multiscale observations of CO₂, ¹³CO₂, and pollutants at Four Corners for emission verification and attribution, *P. Natl. Acad. Sci. USA*, 111, 8386–8391, doi:10.1073/pnas.1321883111, 2014.
- Lindqvist, H., O'Dell, C. W., Basu, S., Boesch, H., Chevallier, F., Deutscher, N., Feng, L., Fisher, B., Hase, F., Inoue, M., Kivi, R., Morino, I., Palmer, P. I., Parker, R., Schneider, M., Sussmann, R., and Yoshida, Y.: Does GOSAT capture the true seasonal cycle of carbon dioxide?, *Atmos. Chem. Phys.*, 15, 13023–13040, doi:10.5194/acp-15-13023-2015, 2015.

- Messerschmidt, J., Geibel, M. C., Blumenstock, T., Chen, H., Deutscher, N. M., Engel, A., Feist, D. G., Gerbig, C., Gisi, M., Hase, F., Katrynski, K., Kolle, O., Lavric, J. V., Notholt, J., Palm, M., Ramonet, M., Rettinger, M., Schmidt, M., Sussmann, R., Toon, G. C., Truong, F., Warneke, T., Wennberg, P. O., Wunch, D., and Xueref-Remy, I.: Calibration of TCCON column-averaged CO₂: the first aircraft campaign over European TCCON sites, *Atmos. Chem. Phys.*, 11, 10765–10777, doi:10.5194/acp-11-10765-2011, 2011.
- Miller, C. E., Crisp, D., DeCola, P. L., Olsen, S. C., Randerson, J. T., Michalak, A. M., Alkhaled, A., Rayner, P., Jacob, D. J., Suntharalingam, P., Jones, D. B. A., Denning, A. S., Nicholls, M. E., Doney, S. C., Pawson, S., Boesch, H., Connor, B. J., Fung, I. Y., O'Brien, D., Salawitch, R. J., Sander, S. P., Sen, B., Tans, P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., and Law, R. M.: Precision requirements for space-based XCO₂ data, *J. Geophys. Res.*, 112, D10314, doi:10.1029/2006JD007659, 2007.
- Morino, I., Uchino, O., Inoue, M., Yoshida, Y., Yokota, T., Wennberg, P. O., Toon, G. C., Wunch, D., Roehl, C. M., Notholt, J., Warneke, T., Messerschmidt, J., Griffith, D. W. T., Deutscher, N. M., Sherlock, V., Connor, B., Robinson, J., Sussmann, R., and Rettinger, M.: Preliminary validation of column-averaged volume mixing ratios of carbon dioxide and methane retrieved from GOSAT short-wavelength infrared spectra, *Atmos. Meas. Tech.*, 4, 1061–1076, doi:10.5194/amt-4-1061-2011, 2011.
- Nassar, R., Jones, D. B. A., Kulawik, S. S., Worden, J. R., Bowman, K. W., Andres, R. J., Suntharalingam, P., Chen, J. M., Breninkmeijer, C. A. M., Schuck, T. J., Conway, T. J., and Worthy, D. E.: Inverse modeling of CO₂ sources and sinks using satellite observations of CO₂ from TES and surface flask measurements, *Atmos. Chem. Phys.*, 11, 6029–6047, doi:10.5194/acp-11-6029-2011, 2011.
- Nguyen, H., Osterman, G., Wunch, D., O'Dell, C., Mandrake, L., Wennberg, P., Fisher, B., and Castano, R.: A method for collocating satellite XCO₂ data to ground-based data and its application to ACOS-GOSAT and TCCON, *Atmos. Meas. Tech.*, 7, 2631–2644, doi:10.5194/amt-7-2631-2014, 2014.
- O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Crisp, D., Eldering, A., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: Corrigendum to “The ACOS CO₂ retrieval algorithm – Part 1: Description and validation against synthetic observations” published in *Atmos. Meas. Tech.*, 5, 99–121, 2012, *Atmos. Meas. Tech.*, 5, 193–193, doi:10.5194/amt-5-193-2012, 2012.
- Olsen, S. C. and Randerson, J. T.: Differences between surface and column atmospheric CO₂ and implications for carbon cycle research, *J. Geophys. Res.-Atmos.*, 109, D02301, doi:10.1029/2003JD003968, 2004.
- Oshchepkov, S., Bril, A., Yokota, T., Morino, I., Yoshida, Y., Matsunaga, T., Belikov, D., Wunch, D., Wennberg, P., Toon, G., O'Dell, C., Butz, A., Guerlet, S., Cogan, A., Boesch, H., Eguchi, N., Deutscher, N., Griffith, D., Macatangay, R., Notholt, J., Sussmann, R., Rettinger, M., Sherlock, V., Robinson, J., Kyrö, E., Heikkinen, P., Feist, D. G., Nagahama, T., Kadyrov, N., Maksyutov, S., Uchino, O., and Watanabe, H.: *J. Geophys. Res.*, 117, 1–18, doi:10.1029/2012JD017505, 2012.
- 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.
- Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO₂ inversions, *Biogeosciences*, 10, 6699–6720, doi:10.5194/bg-10-6699-2013, 2013.
- Prather, M. J., Zhu, X., Strahan, S. E., Steenrod, S. D., and Rodriguez, J. M.: Quantifying errors in trace species transport modeling, *P. Natl. Acad. Sci. USA*, 105, 19617–19621, doi:10.1073/pnas.0806541106, 2008.
- Reuter, M., Buchwitz, M., Schneising, O., Heymann, J., Bovensmann, H., and Burrows, J. P.: A method for improved SCIAMACHY CO₂ retrieval in the presence of optically thin clouds, *Atmos. Meas. Tech.*, 3, 209–232, doi:10.5194/amt-3-209-2010, 2010.
- Reuter, M., Bovensmann, H., Buchwitz, M., Burrows, J. P., Connor, B. J., Deutscher, N. M., Griffith, D. W. T., Heymann, J., Keppel-Aleks, G., Messerschmidt, J., Notholt, J., Petri, C., Robinson, J., Schneising, O., Sherlock, V., Velasco, V., Warneke, T., Wennberg, P. O., Wunch, D.: Retrieval of atmospheric CO₂ with enhanced accuracy and precision from SCIAMACHY: Validation with FTS measurements and comparison with model results, *J. Geophys. Res.-Atmos.*, 116, D04301, doi:10.1029/2010JD015047, 2011.
- Reuter, M., Buchwitz, M., Schneising, O., Hase, F., Heymann, J., Guerlet, S., Cogan, A. J., Bovensmann, H., and Burrows, J. P.: A simple empirical model estimating atmospheric CO₂ background concentrations, *Atmos. Meas. Tech.*, 5, 1349–1357, doi:10.5194/amt-5-1349-2012, 2012a.
- Reuter, M., Bovensmann, H., Buchwitz, M., Burrows, J. P., Heymann, J., Schneising, O., Schnisa, H., Velasco, V.: Algorithm Theoretical Basis Document (ATBD) Bremen Optimal Estimation DOAS (BESD) Version 1, available at: http://www.iup.uni-bremen.de/~mreuter/atbd_besd_v1.pdf (last access: 23 February 2016), 2012b.
- Reuter, M., Bösch, H., Bovensmann, H., Bril, A., Buchwitz, M., Butz, A., Burrows, J. P., O'Dell, C. W., Guerlet, S., Hasekamp, O., Heymann, J., Kikuchi, N., Oshchepkov, S., Parker, R., Pfeifer, S., Schneising, O., Yokota, T., and Yoshida, Y.: A joint effort to deliver satellite retrieved atmospheric CO₂ concentrations for surface flux inversions: the ensemble median algorithm EMMA, *Atmos. Chem. Phys.*, 13, 1771–1780, doi:10.5194/acp-13-1771-2013, 2013.
- Reuter, M., Buchwitz, M., Hilker, M., Heymann, J., Schneising, O., Pillai, D., Bovensmann, H., Burrows, J. P., Bösch, H., Parker, R., Butz, A., Hasekamp, O., O'Dell, C. W., Yoshida, Y., Gerbig, C., Nehrkorn, T., Deutscher, N. M., Warneke, T., Notholt, J., Hase, F., Kivi, R., Sussmann, R., Machida, T., Matsueda, H., and Sawa, Y.: Satellite-inferred European carbon sink larger than expected, *Atmos. Chem. Phys.*, 14, 13739–13753, doi:10.5194/acp-14-13739-2014, 2014.
- Rubin, D. B.: The Bayesian Bootstrap, *Ann. Stat.* 9, 130–134, doi:10.1214/aos/1176345338, 1981.

- Saito, R., Patra, P. K., Deutscher, N., Wunch, D., Ishijima, K., Sherlock, V., Blumenstock, T., Dohe, S., Griffith, D., Hase, F., Heikkinen, P., Kyrö, E., Macatangay, R., Mendonca, J., Messerschmidt, J., Morino, I., Notholt, J., Rettinger, M., Strong, K., Sussmann, R., and Warneke, T.: Technical Note: Latitude-time variations of atmospheric column-average dry air mole fractions of CO₂, CH₄ and N₂O, *Atmos. Chem. Phys.*, 12, 7767–7777, doi:10.5194/acp-12-7767-2012, 2012.
- Sancho, P., De La Cruz, J., Diaz, A., Martin, F., Hernandez, E., Valero, F., and Albarran, B.: A five-year climatology of back-trajectories from the Izaña baseline station, Tenerife, Canary Islands, *Atmos. Environ.*, 26, 1081–1096, 1992.
- Schneising, O., Bergamaschi, P., Bovensmann, H., Buchwitz, M., Burrows, J. P., Deutscher, N. M., Griffith, D. W. T., Heymann, J., Macatangay, R., Messerschmidt, J., Notholt, J., Rettinger, M., Reuter, M., Sussmann, R., Velazco, V. A., Warneke, T., Wennberg, P. O., and Wunch, D.: Atmospheric greenhouse gases retrieved from SCIAMACHY: comparison to ground-based FTS measurements and model results, *Atmos. Chem. Phys.*, 12, 1527–1540, doi:10.5194/acp-12-1527-2012, 2012.
- Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., Ciais, P., Ramonet, M., Bousquet, P., Nakazawa, T., Aoki, S., Machida, T., Inoue, G., Vinnichenko, N., Lloyd, J., Jordan, A., Heimann, M., Shibistova, O., Langenfelds, R. L., Steele, L. P., Francey, R. J., Denning, A. S.: Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO₂, *Science*, 22, 1732–1735, 2007.
- Thoning, K. W., Tans, P. P., and Komhyr, W. D.: Atmospheric Carbon Dioxide at Mauna Loa Observatory 2. Analysis of the NOAA GMCC Data, 1974–1985, *J. Geophys. Res.*, 94, 8549–8565, doi:10.1029/JD094iD06p08549, 1989.
- Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., Uchino, O., Abshire, J. B., Bernath, P., Biraud, S. C., Blavier, J.-F. L., Boone, C., Bowman, K. P., Browell, E. V., Campos, T., Connor, B. J., Daube, B. C., Deutscher, N. M., Diao, M., Elkins, J. W., Gerbig, C., Gottlieb, E., Griffith, D. W. T., Hurst, D. F., Jiménez, R., Keppel-Aleks, G., Kort, E. A., Macatangay, R., Machida, T., Matsueda, H., Moore, F., Morino, I., Park, S., Robinson, J., Roehl, C. M., Sawa, Y., Sherlock, V., Sweeney, C., Tanaka, T., and Zondlo, M. A.: 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, 2010.
- 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 (TCCON), *Philos. T. R. Soc. A*, 369, 2087–2112, doi:10.1098/rsta.2010.0240, 2011a.
- Wunch, D., Wennberg, P. O., Toon, G. C., Connor, B. J., Fisher, B., Osterman, G. B., Frankenberg, C., Mandrake, L., O'Dell, C., Ahonen, P., Biraud, S. C., Castano, R., Cressie, N., Crisp, D., Deutscher, N. M., Eldering, A., Fisher, M. L., Griffith, D. W. T., Gunson, M., Heikkinen, P., Keppel-Aleks, G., Kyrö, E., Lindenmaier, R., Macatangay, R., Mendonca, J., Messerschmidt, J., Miller, C. E., Morino, I., Notholt, J., Oyafuso, F. A., Rettinger, M., Robinson, J., Roehl, C. M., Salawitch, R. J., Sherlock, V., Strong, K., Sussmann, R., Tanaka, T., Thompson, D. R., Uchino, O., Warneke, T., and Wofsy, S. C.: A method for evaluating bias in global measurements of CO₂ total columns from space, *Atmos. Chem. Phys.*, 11, 12317–12337, doi:10.5194/acp-11-12317-2011, 2011b.
- Wunch, D., Toon, G. C., Sherlock, V., Deutscher, N. M., Liu, X., Feist, D. G., and Wennberg, P. O.: The Total Carbon Column Observing Network's GGG2014 Data Version. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, available at: doi:10.14291/tcccon.ggg2014.documentation.R0/1221662, 2015.
- Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H., and Maksyutov, S.: Global Concentrations of CO₂ and CH₄ Retrieved from GOSAT: First Preliminary Results, *Sola*, 5, 160–163, doi:10.2151/sola.2009-041, 2009.