Atmospheric
Measurement
Techniques

Open Access

The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparative validation of GHG-CCI SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT CO_2 and CH_4 retrieval algorithm products with measurements from the TCCON

B. Dils[1], M. Buchwitz[2], M. Reuter[2], O. Schneising[2], H. Boesch[3], R. Parker[3], S. Guerlet[4,*], I. Aben[4], T. Blumenstock[5], J. P. Burrows[2], A. Butz[5], N. M. Deutscher[2,7], C. Frankenberg[6], F. Hase[5], O. P. Hasekamp[4], J. Heymann[2], M. De Mazière[1], J. Notholt[2], R. Sussmann[5], T. Warneke[2], D. Griffith[7], V. Sherlock[8,**], and D. Wunch[9]

[1]Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium
[2]Institute of Environmental Physics (IUP), University of Bremen, Bremen, Germany
[3]University of Leicester, Leicester, UK
[4]SRON Netherlands Institute for Space Research, Utrecht, the Netherlands
[5]Karlsruhe Institute of Technology (KIT), Karlsruhe and Garmisch-Partenkirchen, Germany
[6]Jet Propulsion Laboratory (JPL), Pasadena, California, USA
[7]University of Wollongong, Wollongong, Australia
[8]National Institute of Water and Atmospheric Research (NIWA), Lauder, New Zealand
[9]California Institute of Technology, Pasadena, California, USA
[*]now at: Laboratoire de Météorologie Dynamique (LMD), Palaiseau, France
[**]now at: Laboratoire des Sciences du Climat et de l'Environnement (LSCE), Gif-sur-Yvette, France

Correspondence to: B. Dils (bart.dils@aeronomy.be)

Abstract. Column-averaged dry-air mole fractions of carbon dioxide and methane have been retrieved from spectra acquired by the TANSO-FTS (Thermal And Near-infrared Sensor for carbon Observations-Fourier Transform Spectrometer) and SCIAMACHY (Scanning Imaging Absorption Spectrometer for Atmospheric Cartography) instruments on board GOSAT (Greenhouse gases Observing SATellite) and ENVISAT (ENVIronmental SATellite), respectively, using a range of European retrieval algorithms. These retrievals have been compared with data from ground-based high-resolution Fourier transform spectrometers (FTSs) from the Total Carbon Column Observing Network (TCCON). The participating algorithms are the weighting function modified differential optical absorption spectroscopy (DOAS) algorithm (WFMD, University of Bremen), the Bremen optimal estimation DOAS algorithm (BESD, University of Bremen), the iterative maximum a posteriori DOAS (IMAP, Jet Propulsion Laboratory (JPL) and Netherlands Institute for Space Research algorithm (SRON)), the proxy and full-physics versions of SRON's RemoTeC algorithm (SRPR and SRFP, respectively) and the proxy and full-physics versions of the University of Leicester's adaptation of the OCO (Orbiting Carbon Observatory) algorithm (OCPR and OCFP, respectively). The goal of this algorithm inter-comparison was to identify strengths and weaknesses of the various so-called round- robin data sets generated with the various algorithms so as to determine which of the competing algorithms would proceed to the next round of the European Space Agency's (ESA) Greenhouse Gas Climate Change Initiative (GHG-CCI) project, which is the generation of the so-called Climate Research Data Package (CRDP), which is the first version of the Essential Climate Variable (ECV) "greenhouse gases" (GHGs).

For XCO_2, all algorithms reach the precision requirements for inverse modelling (< 8 ppm), with only WFMD having a lower precision (4.7 ppm) than the other algorithm products (2.4–2.5 ppm). When looking at the seasonal relative accuracy (SRA, variability of the bias in space and time), none of the algorithms have reached the demanding < 0.5 ppm threshold.

For XCH_4, the precision for both SCIAMACHY products (50.2 ppb for IMAP and 76.4 ppb for WFMD) fails to meet the < 34 ppb threshold for inverse modelling, but note that this work focusses on the period after the 2005 SCIAMACHY detector degradation. The GOSAT XCH_4 precision ranges between 18.1 and 14.0 ppb. Looking at the SRA, all GOSAT algorithm products reach the < 10 ppm threshold (values ranging between 5.4 and 6.2 ppb). For SCIAMACHY, IMAP and WFMD have a SRA of 17.2 and 10.5 ppb, respectively.

1 Introduction

According to the IPCC 2007 report (Solomon et al., 2007), based on estimates of radiative forcing between 1750 and 2005, carbon dioxide and methane combined account for over 80 % of the anthropogenic greenhouse gas warming effect. It is therefore important to understand the magnitude and distribution of the CO_2 and CH_4 sources and sinks. Despite their importance, our knowledge of the sources and sinks still has significant gaps (e.g. Stephens et al., 2007; Canadell et al., 2010). For instance it is still unclear why between ∼ 2000 and 2006 methane levels in the atmosphere were rather stable (Simpson et al., 2012), while before and after this period they were rising (currently by about 7–8 ppb year^{-1}; e.g. Rigby et al., 2008; Dlugokencky at al., 2009; Schneising et al., 2011).

Currently surface in situ trace gas concentration measurements are the primary data used to constrain inverse model estimates of surface fluxes (Baker et al., 2006), but these measurements only cover a fraction of Earth's atmosphere. Global satellite observations, sensitive to the near-surface CO_2 and CH_4 variations, are therefore important data sets to improve these flux estimations (Chevallier et al., 2007; Bergamaschi et al., 2009). However given the long atmospheric lifetimes of both gases (30–95 years for CO_2, ∼ 12 years for CH_4; e.g. Jacobson, 2005; Prather, 1994; Prather et al., 2001), the fluxes are small compared to the resident quantity in the atmosphere. Therefore the satellite accuracy requirements are very demanding, since small errors in the retrieved total column concentrations may result in significant errors in the derived fluxes (e.g. Meirink et al., 2006; Chevallier et al., 2007).

Currently only two satellite instruments, SCIAMACHY (Scanning Imaging Absorption Spectrometer for Atmospheric Cartography) on board ENVISAT (ENVironmental SATellite) (Bovensmann et al., 1999) and TANSO-FTS (Thermal And Near infrared Sensor for carbon Observations-Fourier Transform Spectrometer) on board GOSAT (Greenhouse gases Observing SATellite; Kuze et al., 2009), deliver, or have delivered (SCIAMACHY operation ended in April 2012), measurements that are sensitive to near-surface CO_2 and CH_4 concentration variations. Both make use of the near-infrared/short-wave-infrared (NIR/SWIR) spectral region, to analyse the reflected solar radiation in a nadir-looking configuration.

The aim of the European Space Agency's (ESA) Greenhouse Gas Climate Change Initiative (GHG-CCI) project is to provide a single high-quality satellite product for each trace gas retrieval (four satellite–species combinations in total): the so-called Essential Climate Variables (ECVs). In the round-robin (RR) evaluation phase of the project, a number of different algorithms are competing to proceed into the next phase of the project, which is the development of the aforementioned ECV records. Here we will present the validation results of these algorithms, using retrievals from spectra acquired by ground-based high-resolution Fourier transform spectrometers (FTSs) in the Total Carbon Column Observing Network (TCCON). All the algorithms discussed in this paper have already been validated to some extent at various stages in their development, often using the very same TCCON data. However, approaches such as the collocation area and time, averaging of data over time, etc., often vary between each study. Here we will present a comparative validation study, using a uniform strategy, focussing on the inter-algorithm differences and the significance thereof. The decision reached at the end of the round-robin analysis was based on more than this study alone. A general overview of the project's complete quality assessment results is given in Buchwitz et al. (2013).

2 Instruments

SCIAMACHY is a grating spectrometer on board the European environmental satellite ENVISAT, which was launched on 1 March 2002 into a sun-synchronous polar orbit. After a decade in orbit, contact with the satellite was finally lost on 8 April 2012. The SCIAMACHY instrument measured reflected, transmitted and backscattered solar radiation with a 0.2–1.4 nm resolution (Bovensmann et al., 1999). Its spectral band pass was divided into 8 channels. The first 6 covered the 214–1750 nm region while channels 7 and 8 covered the 1940–2040 nm and 2265–2380 nm intervals, respectively. Unfortunately NIR/SWIR channels 7 and 8 suffered from inflight ice deposition on the detector. Therefore, despite the fact that these channels featured many CO_2 and CH_4 absorption features, the retrieval algorithms discussed in this paper make use of channel 6. A problem of channel 6 is that the number of dead and bad detector pixels continued to increase in the spectral region used for methane retrieval during the instrument's lifetime.

GOSAT was launched on 23 January 2009 by the Japanese Space Agency (JAXA) as a dedicated greenhouse-gas-monitoring satellite (Kuze et al., 2009). It is equipped with two instruments: TANSO-FTS and TANSO-CAI (the latter being a Cloud and Aerosol Imager that supports the FTS measurements). The TANSO-FTS instrument has four spectral bands with a resolution of 0.3 cm^{-1}, of which three

Table 1. List of all GHG-CCI algorithms inter-compared in this study, their time coverage and references.

Molec	Algorithm	Institute	Satellite	Time covered	References
XCO2	BESD v01.00.01	IUP	SCIAMACHY	01/2006–12/2011	Reuter et al. (2010, 2011)
XCO2	WFMD v2.2	IUP	SCIAMACHY	01/2003–12/2009	Schneising et al. (2011, 2012), Heymann et al. (2012b)
XCO2	OCFC* v3.0	UoL	GOSAT	04/2009–05/2011	Cogan et al. (2012)
XCO2	SRFC* v1.1	SRON	GOSAT	04/2009–04/2011	Butz et al. (2011)
XCH4	IMAP v6.0	SRON	SCIAMACHY	01/2003–12/2010	Frankenberg et al. (2011)
XCH4	WFMD v2.3	IUP	SCIAMACHY	01/2003–12/2009	Schneising et al. (2010, 2011)
XCH4	OCFP v3.2	UoL	GOSAT	04/2009–04/2011	Parker et al. (2011)
XCH4	OCPR v3.2	UoL	GOSAT	04/2009–04/2011	Parker et al. (2011)
XCH4	SRFP v1.1	SRON	GOSAT	04/2009–05/2011	Butz et al. (2011)
XCH4	SRPR v1.1	SRON	GOSAT	04/2009–05/2011	Schepers et al. (2012)

* OCFC and SRFC are bias-corrected versions of OCFP and SRFP, respectively.

operate in the SWIR (around 760, 1600 and 2000 nm) and one (between 5500 and 14 300 nm) in the thermal infrared. The first three provide sensitivity to the entire column including good near-surface sensitivity, while the latter is sensitive to the mid-troposphere.

ENVISAT/SCIAMACHY retrieval algorithms are typically associated with the instrument (i.e. SCIAMACHY), while GOSAT/TANSO-FTS algorithms typically use the satellite (i.e. GOSAT) identifier. For the sake of consistency, we use the above-mentioned convention in this paper. Therefore, if we refer to GOSAT, we are implying the TANSO-FTS instrument on board GOSAT.

3 Retrieval algorithms

In total, 10 retrieval algorithm products (listed in Table 1 together with their version number and appropriate references) have been compared in four separate comparison pools for the four ECVs, namely SCIAMACHY XCH_4, SCIAMACHY XCO_2, GOSAT XCH_4 and GOSAT XCO_2. The data used in this study contain over-land measurements only. The features of all algorithms have already been reported in several peer-reviewed publications, and in the GHG-CCI Algorithm Theoretical Basis Document (ATBD; Reuter et al., 2012), so we will only give a very brief overview. Several algorithms come in a full-physics (typically tagged by FP in their four-letter acronym) and proxy (PR) version. The proxy method uses a "reference gas" to derive the dry-air column-averaged mole fraction (XCO_2 and XCH_4). This reference gas (in the case of CH_4, CO_2 is used as the reference; in the case of CO_2, O_2 is used) needs to have a far lower variability (in space and time) than the species of interest. This method allows for a very fast but still at least reasonably accurate retrieval in which many of the retrieval errors are cancelled in the CH_4/CO_2 or CO_2/O_2 ratio. On the downside, some error components do not cancel out and, in the case of XCH_4, one needs to correct for the remaining variability of the CO_2

reference gas, typically by using a global model (see for instance Frankenberg et al., 2005, 2011; Parker et al., 2011; Schneising et al., 2009, 2011; Schepers et al., 2012). The full-physics algorithms, on the other hand, model all relevant physical effects and derive the dry-air column-averaged mole fractions from the retrieved surface pressure or meteorological data. They are computationally more demanding than their proxy counterparts, but their dependence on models is reduced (Butz et al., 2011). All algorithms are still under continuous development, and indeed in some cases have already released an updated version (e.g. Guerlet et al., 2013; Oshchepkov et al., 2013). This paper deals with the versions submitted to the GHG-CCI round-robin data pool.

3.1 SCIAMACHY XCO₂ algorithms

Here the weighting function modified (WFM) differential optical absorption spectroscopy (DOAS) algorithm (henceforward referred to as WFMD) competes with the Bremen optimal estimation DOAS (BESD) algorithm, both developed at the University of Bremen. For WFMD we refer to Buchwitz et al. (2000, 2005, 2007), Schneising et al. (2008, 2009, 2011, 2012) and Heymann et al. (2012a). The version validated in this paper is described by Heymann et al. (2012b). For BESD, a more recent product, we refer to Reuter et al. (2010, 2011). WFMD is a proxy least-squares method based on a fast look-up table (LUT) scheme and uses a single constant atmospheric prior. BESD on the other hand is a full-physics algorithm based on optimal estimation (Rodgers, 2000) and uses on-line radiative transfer (RT) model simulations. Note that WFMD is the only XCO_2 retrieval algorithm that did not feature a bias-correction post-processing step based on TCCON (which would improve its validation parameters).

Table 2. List of all participating TCCON stations, their location, time coverage and number of data points (N).

TCCON station name	Lat (°)	Lon (°)	Alt (m)	Time	N
Bialystok (BIA)	53.23	23.03	183	03/2009–11/2011	31 256
Bremen (BRE)	53.10	8.85	7	01/2007–11/2011	10 634
Karlsruhe (KAR)	49.10	8.44	110	04/2010–11/2011	8586
Orleans (ORL)	47.96	2.11	132	08/2009–11/2011	18 169
Garmisch (GAR)	47.48	11.06	744	05/2009–11/2011	26 528
Park Falls (PAR)	45.94	−90.27	442	06/2004–08/2012	169 912
Lamont (LAM)	36.60	−97.49	320	07/2008–08/2012	207 855
Darwin (DAR)	−12.42	130.89	30	08/2005–11/2011	158 879
Wollongong (WOL)	−34.41	150.88	30	06/2008–11/2011	40 622
Lauder (LAU)	−45.05	169.68	370	06/2004–11/2011	117 349

3.2 GOSAT XCO$_2$ algorithms

Here we have two full-physics algorithms: one developed at the University of Leicester (UoL), referred to in this article as OCFP, and one at SRON, the Netherlands Institute for Space Research, referred to as SRFP. The first is UoL's implementation of the OCO (Orbiting Carbon Observatory; Crisp et al., 2004) full-physics algorithm (Cogan et al., 2012). The second is a development of SRON's RemoTeC algorithm (Butz et al., 2011). Both algorithms adjust parameters of a surface–atmosphere state vector and other parameters to the satellite observations, but differ in many other aspects such as their inversion scheme (optimal estimation versus Tikhonov–Phillips), RT models, pre- and post-processing, etc. For more information we refer to Cogan et al. (2012) and Butz et al. (2011). Note that both algorithms feature a post-processing bias-correction scheme. The algorithms are henceforward referred to as SRFC and OCFC to contrast with the non-bias-corrected SRFP and OCFP products.

3.3 SCIAMACHY XCH$_4$ algorithms

Again we have the WFMD algorithm, although this time the version described in Schneising et al. (2011) together with the IMAP (iterative maximum a posteriori) DOAS (Frankenberg et al., 2011) algorithm (in this article further referred to as IMAP). Both algorithms are fairly mature but have primarily focussed on the first three years of SCIAMACHY retrievals up until the 2005 SCIAMACHY detector degradation in the methane spectral region. Extending the time series beyond 2005 remains a challenge (see Frankenberg et al., 2011; Schneising et al., 2011, for details). Both are proxy algorithms. Apart from calibration, pre- and post-filtering differences, WFMD uses a method in which a linearized radiative transfer model (chosen from a look-up table) plus a low-order polynomial is linear least-squares fitted to the logarithm of the measured sun-normalized radiance. IMAP on the other hand uses an optimal estimation inversion method, which minimizes both the least-squares difference between

forward model and measurement as well as between the a priori and a posteriori state vector.

3.4 GOSAT XCH$_4$ algorithms

Here we have both the full-physics and proxy versions of the UoL (OCFP & OCPR) and SRON (SRFP & SRPR) algorithms mentioned above in Sect. 3.2. We refer to Parker et al. (2011) for information on OCFP and OCPR, to Butz et al. (2011) for SRFP and Schepers et al. (2012) for SRPR.

4 TCCON

The Total Carbon Column Observing Network (TCCON) (Wunch et al., 2011a) is a network of ground-based FTSs that provide long and quasi-continuous time series of precise and accurate column abundances of CO$_2$, CH$_4$, N$_2$O and CO, retrieved from NIR solar absorption spectra using a nonlinear least-squares fitting algorithm called GFIT. Rather than retrieving the entire profile, GFIT scales an a priori profile to produce a synthetic spectrum that provides the best match with the measured spectrum. TCCON also makes use of the retrieved O$_2$ columns to derive the corresponding dry-air column-averaged mole fractions.

$$XCO_2 = 0.2095(CO_2 \text{column}/O_2 \text{column}) \quad (1)$$

$$XCH_4 = 0.2095(CH_4 \text{column}/O_2 \text{column}). \quad (2)$$

Note that the TCCON O$_2$ retrieval uses the 1.27 micron band of O$_2$, not the O$_2$ A band used in satellite retrievals. An important aspect of TCCON is that aircraft measurements have been performed over many sites, which allows for an empirical scaling to calibrate the TCCON measurements to the WMO standard reference scale (Wunch et al., 2010; Deutscher et al., 2010; Geibel et al., 2012; Messerschmidt et al., 2011; Washenfelder et al., 2006). The scaling factor is uniform for all sites: $0.989 \pm 0.001(1\sigma)$ and 0.978 ± 0.002 for XCO$_2$ and XCH$_4$, respectively. The uncertainty on the TCCON / aircraft ratio also yields information on the total (station-to-station) network consistency (1σ uncertainty of
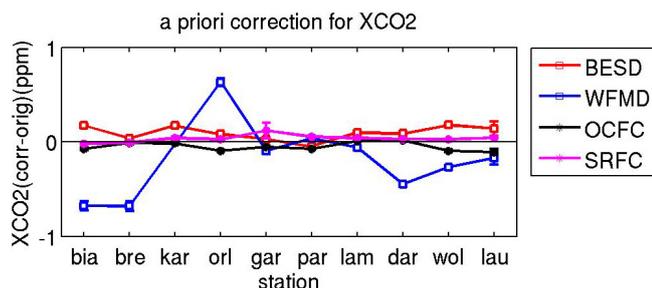
Figure 1. Mean a priori correction (XCO_2 corrected – XCO_2 original), in ppm, on the XCO_2 result, per station.

Figure 2. Mean a priori correction (XCH_4 corrected – XCH_4 original), in ppb, on the XCH_4 result, per station.

0.4 ppm for XCO_2 and 3.5 ppb for XCH_4; see Wunch et al., 2010). There is a continuous effort to decrease any station-to-station biases through improving the network-wide compatibility of the instrumental line shape (ILS) of the spectrometer (Hase et al., 2013). These are monitored by performing regular lamp measurements with a low-pressure HCl gas cell. Another issue which could contribute to the uncertainty is the bias caused by faulty laser sampling boards in the Bruker 125HR instruments (Messerschmidt at al., 2010). These have all since been replaced, but the historical data set remains somewhat compromised. Dohe et al. (2013) have devised a correction scheme, but this still needs to be implemented. In the meantime the TCCON community offer estimated bias corrections for various stations and periods in time. The strongest suggested correction is -1.2 ppm XCO_2 for pre-17 June 2009 Bremen data, while other stations were unaffected. We have not applied these corrections, since all the retrieval algorithms in this study used the equally uncorrected TCCON data for the assessment of their bias-correction procedures and the impact of any such correction on the reported network accuracy (which applies to the uncorrected data set used and should be taken into account when interpreting the validation results) is still unknown. The 10 TCCON stations employed in this study together with their coordinates and periods of operation are listed in Table 2. It is clear that not only the time at which these stations became operational differs, but also the amount of data obtained within a given time period. Because solar absorption FTS measurements can only be made under clear-sky conditions, site location, and the corresponding occurrence of clear-sky days, has a large impact on the number of available measurements.

The TCCON data used in this paper were analysed with the GGG2012 version of the standard TCCON retrieval algorithm.

5 Methodology

The scope of the round-robin algorithm–TCCON comparisons was to identify any remaining shortcomings in the data products generated with the competing algorithms and determine any inter-algor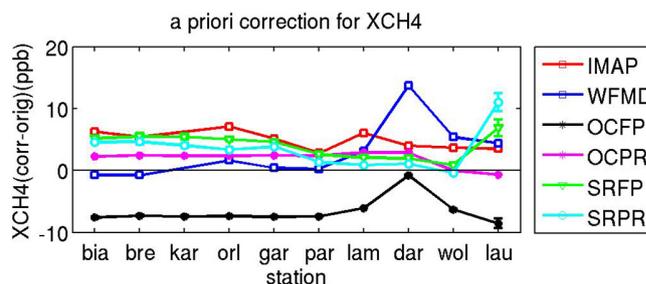ithm quality differences. Therefore the methodology has been kept straightforward and simple, but identical for all algorithms involved.

Complicating the validation is the fact that both TCCON and satellite measurements provide best estimates of the true atmospheric state, based on their own individual sensitivities and a priori information. According to Rodgers (2000), one can correct for the different a priori profiles used in the TCCON and satellite retrieval algorithms. Here we have opted to use the TCCON a priori as the common a priori profile for all measurements. Using Rodgers (2000),

$$x_{\text{cor}} = x + \frac{1}{m_0} \sum_i m^i \left(A^i - 1 \right) \left(ap_x^i - ap_T^i \right) \qquad (3)$$

in which x_{cor} and x are the a-priori-corrected and original column-averaged dry-air mole fractions; i is the vertical layer index; and m^i corresponds to the mass of dry air in layer i, which is directly derived from $\Delta p^i / g^i$. Here Δp^i is the dry-air pressure difference over layer i and g the gravitational constant. m_0 is the sum of m^i over all layers. A^i corresponds to the satellite algorithm's column-averaging kernel, while ap_x and ap_T are the algorithm and TCCON a priori dry-air mole fractions in layer i, respectively.

The impact of the a priori correction is fairly limited. For XCO_2, most algorithms exhibit a quasi-constant correction factor (a priori corrected–original) over all stations ranging between -0.68 and 0.63 ppm. Only WFMD exhibits a stronger and more erratic a priori correction, no doubt due to the single constant a priori it uses in its retrieval scheme (see Fig. 1). For XCH_4, we again notice a quasi-constant correction apart from OCFP and WFMD at Darwin and the SRON products at Lauder. OCFP uses an a priori directly from the TM3 model, while for OCPR a stratospheric adjustment is made using GEOS-Chem model simulations. As the OCPR data exhibit a far smaller correction at Darwin compared to OCFP, an offset in the TM3 stratospheric output is probably the cause. SRON on the other hand uses a XCH_4 a priori derived from the TM4 model (Meirink et al., 2006).

Also noticeable in Fig. 2 is the gradual increase in the WFMD correction factor as we move from north to south, while the SRON products show a slight decrease (apart from Lauder). All the XCH_4 a priori corrections range between -8.6 and 13.7 ppb.

Table 3. BESD and WFMD XCO$_2$ validation results for all individual stations and using all data combined (ALL). All results apart from R and N are in ppm units. The station flagged by * has been excluded from the relative accuracy calculation.

Station	BESD/SCIA XCO$_2$				WFMD/SCIA XCO$_2$			
	Bias	Scatter	R	N	Bias	Scatter	R	N
BIA	0.12	1.99	0.85	504	0.03	5.04	0.56	1714
BRE	−0.20	2.53	0.75	237	0.34	5.14	0.50	1354
ORL	0.53	2.40	0.22	166	2.17	4.07	0.19	209
GAR	1.43	2.10	0.80	144	0.22	6.34	0.11	551
PAR	0.45	2.67	0.83	738	−1.35	5.27	0.70	8206
LAM	−0.78	2.08	0.78	2338	−1.58	4.13	0.46	11 288
DAR	0.24	2.66	0.76	4890	−2.13	4.16	0.41	7250
WOL	0.26	2.64	0.67	654	−0.44	4.67	0.27	1061
LAU*	3.66	1.21	–	3	0.19	6.41	0.04	185
ALL	0.02	2.53	0.81	9674	−1.37	4.69	0.61	31 818
RA	1.28				1.29			
	0.63*				1.36*			

Table 4. BESD and WFMD XCO$_2$ seasonal mean bias, seasonality and SRA results. All results are in ppm units.

Station	BESD/SCIA XCO$_2$					WFMD/SCIA XCO$_2$				
	JFM	AMJ	JAS	OND	Seas	JFM	AMJ	JAS	OND	Seas
BIA	–	0.07	0.69	–	–	–	0.01	0.09	–	–
BRE	–	−0.19	0.34	–	–	–	0.17	0.76	–	–
ORL	–	–	0.68	−1.70	–	–	–	2.15	–	–
GAR	–	1.98	1.31	–	–	–	−1.72	1.09	–	–
PAR	–	0.33	1.04	−1.52	–	–	−1.48	−1.27	−1.62	–
LAM	0.13	−0.69	−1.07	−0.69	0.51	−2.52	−3.08	−0.88	−0.92	1.12
DAR	−0.96	0.48	0.59	−0.71	0.80	−1.28	−2.62	−2.80	0.67	1.60
WOL	3.83	–	−0.62	0.07	–	−0.25	–	1.03	−1.42	–
LAU	–	–	–	–	–	–	–	–	–	–
ALL	−0.34	0.20	0.20	−0.57	0.39	−1.88	−1.79	−1.21	−0.66	0.57
SRA	1.19					1.43				

Note that we only corrected for the a priori difference and not for the difference in vertical sensitivity. That is, even with the same a priori profile its relative contribution to the end result still depends on the averaging kernels. Considering this aspect in the TCCON-satellite comparisons, both of which yield only total column information, requires a reasonable estimate of the true atmospheric variability, which is not available on a global scale. In Wunch et al. (2011b) a detailed assessment of this issue was made, comparing ACOS-GOSAT XCO$_2$ (O'Dell et al., 2012) with TCCON measurements. The study was limited to data taken at the Lamont station only, where the real atmospheric variability could be derived from regular aircraft observations. They found that smoothing the TCCON profile with the ACOS-GOSAT averaging kernel at Lamont induced a bias of about 0.6 ppm with no significant seasonal cycle or airmass dependence. The a priori correction on the other hand did feature a seasonal and latitudinal

dependence, as expected, given that, contrary to the TCCON a priori, the ACOS a priori does not feature a seasonal cycle. No such evaluation has yet been made for a XCH$_4$ retrieval product, nor can it currently be made for any other station. With no ad hoc information on what best represents the true state for all stations, we limit ourselves to the a priori correction described above.

After the a priori correction, all available time series have been trimmed so as to work, in each given comparison round, with data that have matching temporal coverage. For SCIA-MACHY XCH$_4$ this corresponds to 2003–2009, for SCIA-MACHY XCO$_2$: 2006–2009, and both GOSAT XCO$_2$ and XCH$_4$ are limited to between April 2009 and April 2011.

As with every satellite versus FTS comparison we defined a collocation time and area in which satellite and ground-based measurements can be paired. Ideally these criteria are as strict as possible in order to minimize the impact of spatial

Table 5. OCFC and SRFC XCO$_2$ validation results for all individual stations and using all data combined (ALL). All results apart from R and N are in ppm units. Stations flagged by * have been excluded from the relative accuracy calculation (equally flagged by *).

Station	OCFC/GOSAT XCO$_2$				SRFC/GOSAT XCO$_2$			
	Bias	Scatter	R	N	Bias	Scatter	R	N
BIA	−0.52	2.73	0.87	157	−0.13	2.76	0.82	174
BRE	−0.50	2.80	0.68	92	−0.90	2.40	0.83	81
KAR	−0.61	2.87	0.72	188	−1.08	2.51	0.83	151
ORL	−0.72	2.81	0.80	247	−0.83	2.39	0.90	223
GAR	0.31	2.72	0.78	182	0.32	2.66	0.82	168
PAR	−1.31	1.91	0.92	211	−0.77	2.52	0.91	268
LAM	−1.04	1.99	0.78	1432	−0.84	2.14	0.83	1274
DAR	−0.96	2.07	0.33	117	0.67	2.56	0.21	81
WOL	0.15	2.66	0.41	239	0.88	3.27	0.33	189
LAU*	0.71	2.98	0.63	25	−1.74	3.90	0.15	24
ALL	−0.76	2.37	0.79	2890	−0.57	2.50	0.81	2633
RA	0.64				0.84			
	0.53*				0.75*			

and temporal variability on the comparison. Here we have set the collocation time to ± 2 h. The spatial collocation criterion was set at a 500 km radius around the TCCON site. Smaller collocation areas have been tested (100, 350 km) but often yielded unstable results, due to insufficient data. All FTS data points that fall within the temporal overlap criteria of a single satellite measurement (that fell in the spatial overlap area) are then averaged to obtain a unique satellite–FTS data pair.

The typical variability (1σ), including random errors and real atmospheric variability, of the FTS measurements within this 4 h overlap time frame is on average 2.5 ppb for XCH$_4$ and 0.4 ppm for XCO$_2$. Relaxing the overlap criteria does have a significant impact on the variability, and at ± 6 h the variability increases to 3.5 ppb (XCH$_4$) and 0.5 ppm (XCO$_2$).

From these data pairs we derived various statistical parameters. In the figures and tables within this article, N corresponds to the number of collocated data pairs; R is the Pearson's r correlation coefficient; Bias is the average satellite–FTS difference:

$$\text{Bias} = \text{mean}(X_{\text{sat}} - X_{\text{FTS}}); \tag{4}$$

while the scatter corresponds to the standard deviation of said difference:

$$\text{Scatter} = \text{std}(X_{\text{sat}} - X_{\text{FTS}}). \tag{5}$$

Note that the single measurement precision requirements for inverse modelling, set forward by the users, is < 8 ppm for XCO$_2$ and < 34 ppb in the case of XCH$_4$ (Buchwitz et al., 2012).

All these parameters have been calculated using the individual data pairs as well as daily and monthly means. Note that both the daily and monthly means are derived from the individual data pairs; thus the ± 2 h collocation criterion still

applies. In the analysis all data pairs are considered to have equal weight. In this article we will show the results of the individual data pairs only, except for the correlation coefficient R, which is based on the daily averages. Also the time series plots shown are daily averages.

One of the important quality criteria put forward by the users is the so-called "relative accuracy" (RA). This parameter is an indication of the variability of the bias in space and time. The relative accuracy user requirements (1σ standard deviation) put forward by the inverse modelling community are 10 ppb for XCH$_4$ and 0.5 ppm for XCO$_2$ (Buchwitz et al., 2012) based on 1000 km^2 monthly averages. For inverse modelling purposes this parameter is more important than the overall bias as this, if consistent, can be easily corrected for. While this parameter cannot be exactly replicated in our analysis, we calculate a RA, which attempts to yield some information on the station-to-station variability of the bias. We define RA as the standard deviation on the overall biases (derived from individual data) obtained at each station.

The "seasonal relative accuracy" (SRA) is the standard deviation over all seasonal bias results (40 in total: 4 seasonal biases over 10 stations). The seasonal bias results for each station are constructed from all data pairs which fall within the months of January to March (JFM), April to June (AMJ), July to September (JAS) or October to December (OND), regardless of the year the measurements are taken. Some stations feature only limited data during certain seasons, which sometimes results in erratic bias results. To avoid the inclusion of these results into the RA and SRA calculation, we do not include those bias results that are derived from fewer than 10 individual data points or have a standard error ($\sigma/\sqrt{N}$) which exceeds the user relative accuracy requirements (0.5 ppm XCO$_2$, 10 ppb XCH$_4$). RA and SRA are also derived from a common data set; thus if one algorithm in the

validation round fails to meet the quality requirements for station x and season y, the corresponding bias result is also excluded from the SRA and RA calculation of its competitor.

In the case of all four seasonal biases for a station meeting the quality requirements, we also derive the standard deviation on these four results as an indicator of their variability. This parameter is referred to as the "seasonality" (Seas).

6 Results

Shown in each section are overview figures (Figs. 4, 7, 10 and 13) and tables (Tables 3 through 12) that list the statistical parameters obtained at each station, and for all station data combined (ALL). Given the uneven distribution of data among the 10 TCCON stations, stations with high data density such as Lamont have a higher impact on the "all data" results. For practical purposes we will only show an example time series of a single European, North American and Oceanian station.

The overview Tables 13 and 14 also list the 95 % confidence interval of the overall parameters. The confidence intervals on the scatter, RA, Seas and SRA are inferred from the Chi squared (χ^2) distribution in which

$$\sqrt{\frac{(N-1)s^2}{\chi^2_{(\frac{\alpha}{2},N-1)}}} \leq \sigma \leq \sqrt{\frac{(N-1)s^2}{\chi^2_{(1-\frac{\alpha}{2},N-1)}}}, \qquad (6)$$

with σ the population standard deviation, s the sample standard deviation, N the number of data points in the sample and α determining the confidence level (here 0.05 for 95 % confidence).

We also performed a so-called F test, to quantify the probability that the statistical parameters of two competing results stem from the same population (Snedecor and Cochran, 1989). The null hypothesis (H_0) of the test states that the variances of the two populations are equal ($\sigma_1^2 = \sigma_2^2$). The result of the test is the probability that the stated hypothesis is true. Thus, the lower this number, the more likely it is that the obtained parameters such as RA and SRA of two competing algorithms are different.

Note that the F test relies on the presumption that the population exhibits a normal distribution. One could invoke the central limit theorem as the data from which the RA, Seas and SRA are drawn are sample means from the overall population. However, the sampling itself can hardly be called random. Thus to test for normality, we performed a Shapiro–Wilk normality test (Shapiro and Wilk, 1965), on a 0.05 confidence level, on all the relevant data samples. All data samples passed the test apart from the SRA samples from OCPR XCH_4 and BESD XCO_2. However an analysis of the quantile–quantile probability showed no clear departure from normality.
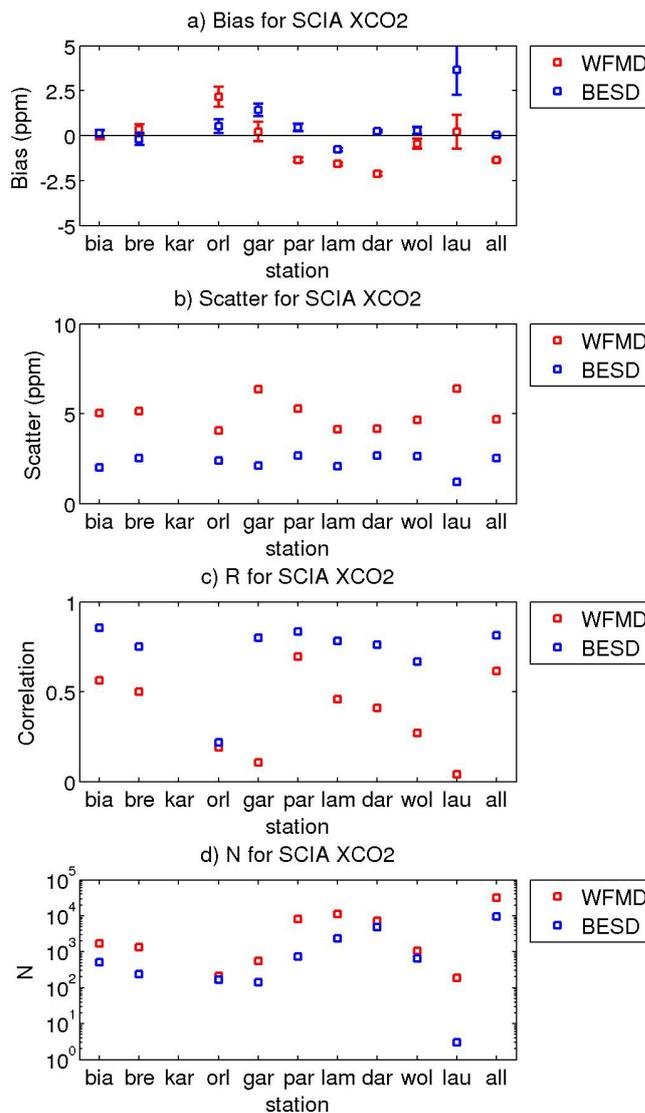

Figure 3. SCIAMACHY XCO_2 bias **(a)**, scatter **(b)**, correlation **(c)** and number of data pairs **(d)**, for all individual TCCON stations and all data combined.

6.1 SCIAMACHY XCO_2

The two competing algorithms are BESD and WFMD. Table 3 and Fig. 3a show the evolution of the bias over the different stations. The error bars in Fig. 3a correspond to the 95 % confidence bands of the bias. Note that there are no data for Karlsruhe since the TCCON measurements there commenced in 2010, while there were no post-2009 WFMD data at the time of this analysis. The overall bias is slightly smaller for BESD, but the variability of the bias (i.e. relative accuracy) is almost identical (1.28 versus 1.29 ppm).

The most significant differences between both data sets are the scatter and data density (Fig. 3b and d). While the overall scatter for BESD is significantly lower (2.5 vs. 4.7 ppm for

Table 6. OCFC and SRFC XCO_2 seasonal mean bias, seasonality and SRA results. All results are in ppm units.

Station	OCFC/GOSAT XCO_2					SRFC/GOSAT XCO_2				
	JFM	AMJ	JAS	OND	Seas	JFM	AMJ	JAS	OND	Seas
BIA	–	0.26	−0.68	–	–	–	−0.20	0.37	–	–
BRE	−1.77	0.76	−0.99	–	–	–	−0.26	−1.79	–	–
KAR	−2.02	0.21	−0.20	–	–	−1.90	−0.70	−0.83	–	–
ORL	−3.30	0.73	−0.52	–	–	−2.09	−0.37	−0.41	−1.31	0.82
GAR	0.08	–	0.96	−2.08	–	–	–	0.54	–	–
PAR	–	−0.55	−2.04	−1.40	–	0.23	0.15	−1.34	−0.74	0.75
LAM	−1.42	0.09	−0.81	−1.64	0.78	−1.03	−0.69	−0.68	−0.99	0.19
DAR	–	−1.04	−0.85	–	–	–	–	1.39	–	–
WOL	1.52	−0.04	−0.45	–	–	–	–	0.73	–	–
LAU	–	–	–	–	–	–	–	–	–	–
ALL	−1.15	0.15	−0.58	−1.56	0.74	−0.85	−0.35	−0.33	−0.97	0.33
SRA	1.08					0.89				

Table 7. IMAP and WFMD XCH_4 validation results for all individual stations and using all data combined (ALL). All results apart from R and N are in ppb units.

Station	IMAP/SCIA XCH_4				WFMD/SCIA XCH_4			
	Bias	Scatter	R	N	Bias	Scatter	R	N
BIA	14.2	42.1	0.29	1228	2.7	85.5	−0.13	2067
BRE	−1.2	54.1	0.17	946	−5.9	88.6	0.21	1516
ORL	0.2	47.9	0.26	287	−6.0	74.7	0.16	255
GAR	10.7	49.8	0.34	641	−4.7	92.6	0.34	630
PAR	2.3	48.9	0.30	22 078	3.7	75.1	0.09	13 607
LAM	11.8	46.0	0.25	9430	2.1	74.0	−0.01	10 808
DAR	−25.3	51.9	0.16	6500	−13.5	72.6	0.10	8044
WOL	−24.0	46.2	−0.23	881	−19.3	79.4	−0.10	1377
LAU	−10.6	49.6	0.23	329	−9.8	91.1	0.01	287
ALL	−0.1	50.2	0.65	42 320	−1.9	76.4	0.44	38 591
RA	14.7				7.8			

WFMD), its data density is also lower (9674 vs. 31 818 data pairs). Interestingly this makes the uncertainty on the overall bias, i.e. the standard error ($\sigma/\sqrt{N}$), very similar (0.025 versus 0.026 ppm for BESD and WFMD, respectively). The higher scatter for WFMD also reveals itself in the generally lower correlation coefficients. Note that for the Lauder station, situated in New Zealand, BESD only offers three data pairs, all of which are measured on the same day (hence the lack of a daily correlation coefficient for this station). Both algorithms fail the above-stated quality requirements at this site, and if we thus exclude the Lauder station from our analysis, the relative accuracy (RA in the Table) of BESD improves to 0.63 ppm, while that of WFMD (slightly) deteriorates to 1.36 ppm.

The time series in Figs. 4 (BESD) and 5 (WFMD) are collocated daily averaged FTS and satellite measurements from Bialystok (a), Lamont (b) and Darwin (c). Comparing these

figures, it is clear that BESD features substantially fewer data than WFMD, due to its more restrictive filtering process (particularly a very strict MERIS cloud mask). Also clearly visible is the extremely limited (if any) seasonal cycle in the Darwin data. BESD data clearly exhibit lower scatter, but some outliers can be identified. This has been identified as an issue related to the SCIAMACHY Level 1 version 7 consolidation product (L1v7u), used in the retrieval. Tests with the new L1v7w data show that these outliers are eliminated, which should further increase BESD's precision.

The seasonality of Lamont, Darwin and the overall results are slightly in favour of BESD as well as the SRA value (see Table 4). Keep in mind however that these parameters are derived from a limited data sample. Neither the seasonality nor SRA difference is significant (P value of the H_0: $\sigma_1^2 = \sigma_2^2$, or the probability that both samples are from a population with

Table 8. IMAP and WFMD XCH$_4$ seasonal mean bias, seasonality and SRA results. All results are in ppb units.

Station	IMAP/SCIA XCH$_4$					WFMD/SCIA XCH$_4$				
	JFM	AMJ	JAS	OND	Seas	JFM	AMJ	JAS	OND	Seas
BIA	–	13.9	14.4	–	–	–	5.7	−4.2	–	–
BRE	35.9	1.9	−7.9	−5.4	20.3	–	−10.6	3.1	–	–
ORL	–	–	7.4	−1.6	–	–	–	−7.1	–	–
GAR	–	1.8	10.7	22.8	–	–	−20.9	5.0	–	–
PAR	14.3	6.9	3.9	−7.3	9.0	7.1	7.8	1.4	−1.6	4.5
LAM	–	18.2	18.4	3.0	–	−8.7	6.1	15.9	−15.2	14.1
DAR	−10.6	−37.7	−23.4	−9.4	13.2	1.8	−18.3	−14.6	−10.8	8.7
WOL	–	–	−41.7	−18.6	–	−45.8	–	−3.1	−22.1	–
LAU	–	–	–	−13.3	–	−10.1	–	–	−0.0	–
ALL	5.1	−0.1	2.5	−4.4	4.0	−7.3	0.5	1.3	−12.5	6.6
SRA	17.2					10.5				

Table 9. OCPR and OCFP XCH$_4$ validation results for all individual stations and using all data combined (ALL). All results apart from R and N are in ppb units.

Station	OCPR/GOSAT XCH$_4$				OCFP/GOSAT XCH$_4$			
	Bias	Scatter	R	N	Bias	Scatter	R	N
BIA	8.3	13.8	0.54	799	3.4	18.6	0.31	213
BRE	5.4	12.6	0.54	279	2.9	14.6	0.25	128
KAR	4.8	13.7	0.52	576	−2.0	17.7	0.14	267
ORL	6.6	12.9	0.49	597	1.0	15.8	0.22	286
GAR	12.1	14.1	0.39	623	8.0	18.1	0.13	232
PAR	5.9	14.0	0.50	887	4.2	16.6	0.51	263
LAM	8.0	14.8	0.49	2757	−0.5	18.1	0.47	1603
DAR	5.8	10.1	0.47	312	−13.9	13.3	0.28	68
WOL	2.4	12.9	0.69	636	−2.9	20.8	0.38	225
LAU	3.6	8.6	0.83	203	−4.7	16.6	0.48	35
ALL	7.0	14.0	0.87	7669	0.4	18.1	0.78	3320
RA	2.7				6.0			

equal variances is 0.55 and 0.42, respectively). The P value for the RA $H_0: \sigma_1^2 = \sigma_2^2$ hypothesis on the other hand is 0.06.

6.2 GOSAT XCO$_2$

Here we have two competing algorithms, OCFC and SRFC, which are the full-physics, bias-corrected versions of University of Leicester's OCO and SRON's RemoTeC algorithms, respectively.

As one can see in Fig. 6 and Tables 5 and 6, the differences concerning all parameters are extremely small. Number of data points, scatter and correlation coefficients are never consistently in favour of one algorithm. Note that the correlation coefficients are quite low for the Southern Hemisphere stations of Darwin, Wollongong and Lauder, which is attributed to the limited seasonal XCO$_2$ variability at these sites. The RA is slightly in favour of OCFC (0.64 vs. 0.84 ppm for

SRFC). Again we have a large uncertainty on the bias values for Lauder. Excluding this station from the relative accuracy calculation yields an RA equal to 0.53 and 0.74 ppm for OCFC and SRFC, respectively. The probability that both sample RA values stem from an equal distribution is 0.32.

Looking at the time series for Orleans, Lamont and Wollongong, (Figs. 7 and 8) there is hardly any difference between the two algorithms. However, OCFC does feature several strong outliers in all three stations. Unlike the station-to-station bias variability, SRFC has a lower variability in the overall seasonal bias (see Table 6). For both algorithms the winter–autumn (October through March) biases seem to be more negative than their spring–summer counterparts. This is also the case for the BESD algorithm. While the difference in overall seasonality (0.74 for OCFC vs. 0.33 for SRFC) is somewhat distinct ($P(H_0 : \sigma_1^2 = \sigma_2^2)$ is 0.22), the difference

Table 10. SRPR and SRFP XCH$_4$ validation results for all individual stations and using all data combined (ALL). All results apart from R and N are in ppb units.

	SRPR/GOSAT XCH$_4$				SRFP/GOSAT XCH$_4$			
Station	Bias	Scatter	R	N	Bias	Scatter	R	N
BIA	9.5	14.7	0.53	423	−0.4	14.2	0.60	174
BRE	2.4	15.7	0.38	125	−4.3	14.9	0.26	81
KAR	1.5	16.9	0.39	322	−5.6	13.3	0.54	151
ORL	6.8	14.2	0.37	359	−2.6	13.4	0.18	223
GAR	7.7	19.4	0.30	345	2.4	16.3	0.28	168
PAR	0.5	15.0	0.44	679	−2.5	14.8	0.50	268
LAM	2.7	13.0	0.74	2096	−2.8	13.9	0.60	1274
DAR	−3.8	8.3	0.66	157	−3.4	14.8	0.14	81
WOL	−1.0	13.4	0.56	418	−2.1	21.4	0.27	189
LAU	5.9	11.7	0.85	82	−8.7	16.2	0.58	24
ALL	3.1	14.6	0.87	5006	−2.5	14.9	0.83	2633
RA	4.2				3.0			

Table 11. OCPR and OCFP XCH$_4$ seasonal mean bias, seasonality and SRA results. All results are in ppb units.

	OCPR/GOSAT XCH$_4$					OCFP/GOSAT XCH$_4$				
Station	JFM	AMJ	JJA	SON	Seas	JFM	AMJ	JJA	SON	Seas
BIA	4.5	10.4	15.1	0.4	6.5	−3.3	3.3	10.0	–	–
BRE	2.3	8.5	3.8	1.9	3.0	5.7	2.4	2.3	–	–
KAR	3.0	10.3	−0.2	−0.4	5.0	−2.8	−3.1	−1.3	7.2	4.9
ORL	1.7	12.0	8.5	0.0	5.7	−2.1	1.2	2.4	−1.0	2.0
GAR	13.5	15.4	11.9	4.6	4.7	8.1	5.2	11.1	0.2	4.6
PAR	10.1	10.7	3.7	1.7	4.5	4.9	5.4	4.3	3.4	0.9
LAM	1.5	13.9	13.4	2.6	6.7	−7.1	1.4	4.2	−1.8	4.8
DAR	15.3	−1.9	4.7	10.8	7.5	–	−18.7	−11.8	−13.7	–
WOL	3.2	5.3	3.3	−2.3	3.3	−5.5	4.4	−3.9	−2.1	4.3
LAU	7.2	4.3	1.0	1.9	2.8	−2.6	–	–	−6.1	–
ALL	4.8	11.0	8.8	2.4	3.9	−2.9	1.3	3.1	−1.1	2.6
SRA	5.4					6.2				

in SRA (1.08 for OCFC vs. 0.89 for SRFC) is very small ($P(H_0 : \sigma_1^2 = \sigma_2^2)$ is 0.68).

6.3 SCIAMACHY XCH$_4$

Both IMAP and WFMD are fairly mature proxy type algorithms. Note however that since November 2005 the SCIA-MACHY XCH$_4$ retrievals have suffered from a detector degradation in channel 6. Most of the TCCON stations (apart from Park Falls, Darwin and Lauder) commenced their measurements after this event. The quality assessment in this paper is therefore primarily representative of this post-decay period.

We also have to note that during the course of the validation, we detected strong biases in the January-through-March IMAP seasonal values. This turned out to be a processing error in IMAP (one of the clusters used incorrect settings from a previous IMAP run). All the data derived from that processing unit have been removed from the IMAP data set. This reduced the amount of overlapping data from 55 626 to 42 320 points (or almost 24 %).

The differences between the algorithms are fairly distinctive (see Fig. 9 and Table 7). Obvious is the far larger scatter (see Fig. 9b) in the WFMD data (overall 76 ppb compared to 50 ppb for IMAP). This also translates to an inferior correlation coefficient over all stations except for Wollongong (which features a negative correlation for both algorithms) and Bremen (by a very small amount). Unlike the BESD–WFMD XCO$_2$ comparison, WFMD's higher scatter properties are not offset by a superior data density. So on these parameters alone IMAP seems to outperform WFMD. The reason for the larger scatter of the WFMD data is likely
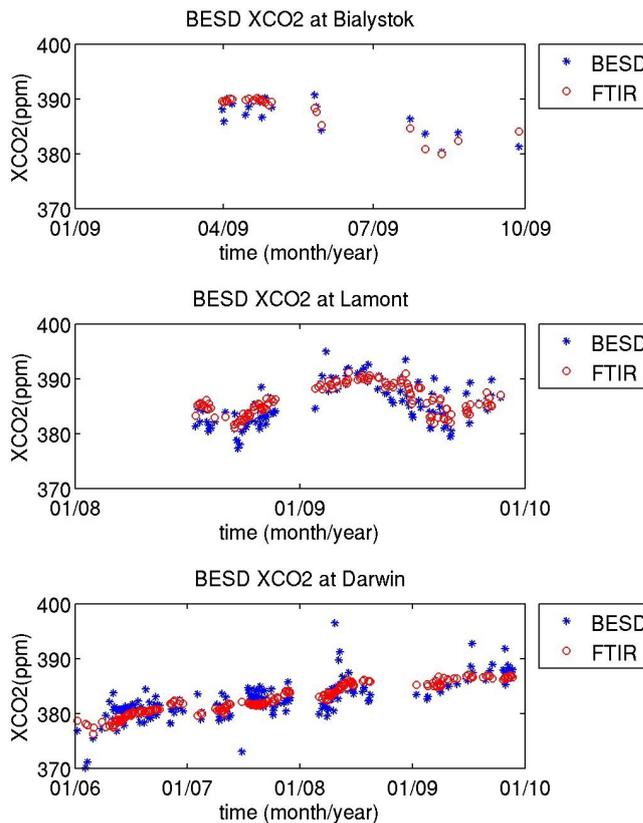
Table 12. SRPR and SRFP XCH$_4$ seasonal mean bias, seasonality and SRA results. All results are in ppb units.

SRPR	SRPR/GOSAT XCH$_4$					SRFP/GOSAT XCH$_4$				
	JFM	AMJ	JJA	SON	Seas	JFM	AMJ	JJA	SON	Seas
BIA	5.1	10.7	10.8	−0.6	5.4	–	−1.8	3.2	−7.0	–
BRE	7.8	9.3	−5.1	−4.3	7.7	–	−3.1	−4.3	–	–
KAR	1.9	8.3	−6.1	0.5	5.9	−5.1	−1.7	−7.7	−7.8	2.9
ORL	6.6	12.9	5.0	−2.9	6.5	1.1	−0.5	−4.3	−10.8	5.3
GAR	11.0	11.7	5.5	4.3	3.8	7.9	4.1	1.2	0.5	3.4
PAR	14.6	6.9	−4.2	2.7	7.9	10.2	2.1	−8.1	−0.1	7.5
LAM	0.5	5.0	1.6	3.2	2.0	−4.5	−0.3	−1.5	−4.6	2.2
DAR	–	−7.1	−2.7	−1.4	–	–	−8.8	1.6	–	–
WOL	−6.9	3.7	1.7	−4.8	5.1	−6.2	7.8	−4.2	0.2	6.2
LAU	3.3	3.8	7.9	7.3	2.3	–	–	–	−16.7	–
ALL	2.7	7.0	1.3	2.1	2.6	−2.0	−0.5	−2.6	−4.5	1.6
SRA	6.2					5.7				

Table 13. Overview table, listing all overall (ALL) results. The listed uncertainties on the bias and scatter correspond to the 0.95 confidence interval. XCO$_2$ bias and scatter in ppm, XCH$_4$ bias and scatter in ppb.

Algo	Bias	Scat	R	N
	SCIA XCO$_2$			
BESD	0.02 ± 0.05	2.53 ± 0.04	0.81	9674
WFMD	−1.37 ± 0.05	4.69 ± 0.04	0.61	31 818
	GOSAT XCO$_2$			
OCFC	−0.76 ± 0.09	2.37 ± 0.06	0.79	2890
SRFC	−0.57 ± 0.10	2.50 ± 0.07	0.81	2633
	SCIA XCH$_4$			
IMAP	−0.1 ± 0.5	50.2 ± 0.3	0.65	42 320
WFMD	−1.9 ± 0.8	76.4 ± 0.5	0.44	38 591
	GOSAT XCH$_4$			
OCFP	0.4 ± 0.6	18.1 ± 0.4	0.78	3320
OCPR	7.0 ± 0.3	14.0 ± 0.2	0.87	7669
SRFP	−2.5 ± 0.6	14.9 ± 0.4	0.83	2633
SRPR	3.1 ± 0.4	14.6 ± 0.3	0.87	5006

Figure 4. Daily averaged time series of paired BESD and TCCON FTIR XCO$_2$ at Bialystok (top), Lamont (mid) and Darwin (bottom).

due to the fact that WFMD is based on unconstrained linear least squares using a single constant methane a priori profile, whereas IMAP is based on optimal estimation using methane model data as a priori information. In addition, there are also other reasons which can explain the differences. For example, IMAP and WFMD differ greatly in their pre-processing steps, targeted at dealing with the problematic SCIAMACHY instrument degradation. IMAP, for instance, uses SRON's own specifically calibrated input spectra, while WFMD uses the official standard SCIAMACHY level 1 data. IMAP uses one single pixel filter (the so-called "Dead and

Bad detector Pixel Mask", or DBPM), while WFMD uses several masks, each one optimized for a certain time period.

However looking at the bias distribution, all three Southern Hemisphere stations (Darwin, Wollongong and Lauder) exhibit a considerable negative bias. Also for WFMD, these

Table 14. Overview table, listing the "relative accuracy" (RA), overall "seasonality" (Seas) and "seasonal relative accuracy" (SRA), together with their 95 % confidence interval (RA 95 %, Seas 95 % and SRA 95 %) and the probability that the obtained sample variances stem from the same population ($P(H_0 : \sigma_1^2 = \sigma_2^2)$). The P values for the GOSAT XCH$_4$ results correspond to the following pairs: (a) OCFP vs. OCPR, (b) OCPR vs. SRFP and (c) SRFP vs. SRPR.

Algorithm	RA	RA 95 %	P(RA)	Seas	Seas 95 %	P(Seas)	SRA	SRA 95 %	P(SRA)
				SCIA XCO$_2$					
BESD	0.63	0.42–1.28	0.06	0.39	0.22–1.45	0.55	1.19	0.91–1.72	0.42
WFMD	1.36	0.90–2.77		0.57	0.32–2.13		1.43	1.09–2.07	
				GOSAT XCO$_2$					
OCFC	0.53	0.36–0.97	0.32	0.74	0.42–2.76	0.22	1.08	0.82–1.58	0.68
SRFC	0.75	0.52–1.37		0.33	0.19–1.23		0.89	0.68–1.30	
				SCIA XCH$_4$					
IMAP	14.7	9.9–28.2	0.09	4.0	2.3–14.9	0.43	17.2	13.2–24.8	0.28
WFMD	7.8	5.3–14.9		6.6	3.7–24.6		10.5	8.0–15.2	
				GOSAT XCH$_4$					
OCFP	6.0	4.1–11.0	0.03[a]	2.6	1.5–9.7	0.52[a]	6.2	5.0–8.3	0.45[a]
OCPR	2.7	1.9–4.9	0.76[b]	3.9	2.2–14.5	0.18[b]	5.4	4.3–7.2	0.77[b]
SRFP	3.0	2.1–5.5	0.33[c]	1.6	0.9–6.0	0.45[c]	5.7	4.6–7.6	0.65[c]
SRPR	4.2	2.9–7.7		2.6	1.5–9.7		6.2	5.0–8.3	

three stations feature a more negative bias than their Northern Hemisphere counterparts but not as distinctive as for IMAP. For IMAP the difference between the mean Southern and Northern Hemisphere bias is 26 ppb, while for WFMD it is 13 ppb. These values are larger than what could reasonably be inferred by the TCCON laser sampling error. This also reflects itself in the relative accuracy, which is 7.8 ppb for WFMD and 14.7 ppb for IMAP.

Given the large scatter, it is difficult to assess any systematic seasonality errors in the time series plots (see Figs. 10 and 11). The IMAP underestimation at Darwin is clear, as well as the stronger scatter in WFMD. Table 8 lists the overall seasonal biases. As with the RA, we see a higher SRA in the IMAP data, although the difference is far less distinctive. For RA $P(H_0 : \sigma_1^2 = \sigma_2^2)$ is 0.09, while for SRA $P(H_0 : \sigma_1^2 = \sigma_2^2)$ equals 0.28. The difference in overall seasonality (4.0 for IMAP vs. 6.6 for WFMD) is even less significant ($P(H_0 : \sigma_1^2 = \sigma_2^2) = 0.43$).

6.4 GOSAT XCH$_4$

Concerning the bias (see Fig. 12 and Tables 9 and 10), as with SCIAMACHY XCH$_4$, the Southern Hemisphere bias values tend to be somewhat lower (in absolute values) than their Northern Hemisphere counterparts, although only consistently so for SRPR and OCFP. The average Northern Hemisphere–Southern Hemisphere bias difference is 3.3 ppb for OCPR, 8.8 ppb for OCFP, 11.0 ppb for SRPR and 6.7 ppb for SRFP, all of which are considerably lower than that observed in IMAP (26 ppb). This is also reflected in the RA

numbers, which range from 2.7 ppb (for OCPR) to 6 ppb (OCFP) (see Tables 9 to 10). The overall bias values themselves range from −2.5 (SRFP) to 7 ppb (OCPR).

Only OCFP has a somewhat lower precision (18.1 ppb), while the overall scatter of the other algorithms ranges between 14 and 14.9 ppb. None of the algorithms is consistently better or worse across all stations involved though (see Fig. 12 and Tables 9 and 10). Similar observations can be made about the correlation coefficients where each algorithm comes out with the best R value at, at least, one station (see Fig. 12c). OCFP has the worst overall scatter, correlation and relative accuracy of all the algorithms involved, while OCPR has the best scatter, data density, relative accuracy and correlation results (the latter a tie with SRPR). The difference between the best (OCPR = 2.7 ppb) and worst (OCFP = 6 ppb) relative accuracy result is significant on a 95 % level ($P(H_0 : \sigma_1^2 = \sigma_2^2) = 0.03$). However the difference between the two best results (OCPR and SRFP at 3 ppb) clearly is not ($P(H_0 : \sigma_1^2 = \sigma_2^2) = 0.76$). The two SRON RA values have a $P(H_0 : \sigma_1^2 = \sigma_2^2) = 0.33$.

The strong difference between the Leicester algorithms can be attributed to the fact that the full-physics version of Leicester's OCO algorithm is a more recent development than its more mature proxy counterpart.

Turning to the seasonality, the full-physics algorithms outperform their respective proxy counterparts by a small margin. Interestingly OCPR, which so far featured the best overall statistics, performs worst when looking at the seasonality. This is also somewhat evident from the time series plot (Fig. 13) where OCPR seems to underestimate the XCH$_4$
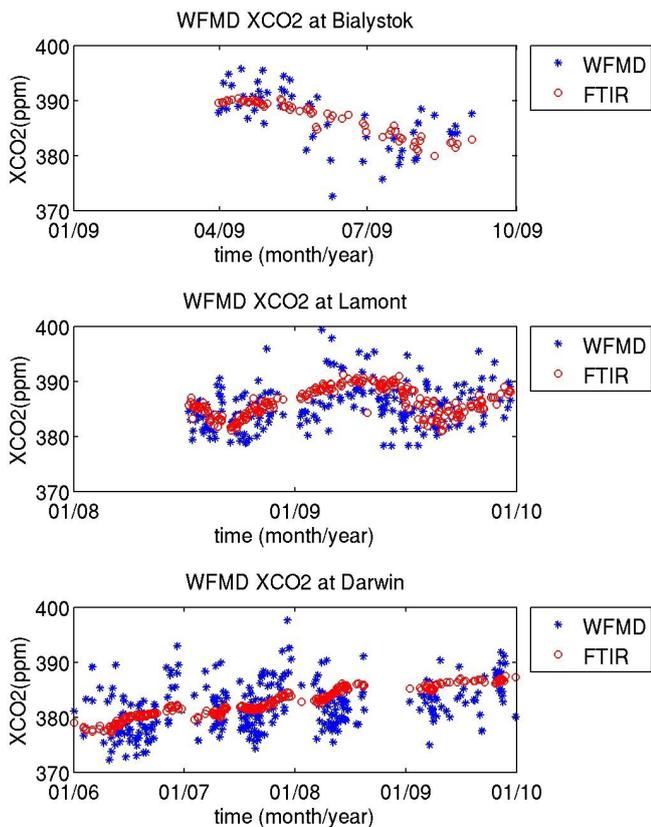
Figure 5. Daily averaged time series of paired WFMD and TCCON FTIR XCO$_2$ at Bialystok (top), Lamont (mid) and Darwin (bottom).

seasonal amplitude (obvious in Lamont). The differences between the other algorithms are less obvious (Figs. 14 to 16). Of course, being a proxy algorithm, some of the effects might come from the model used in the dry-air conversion (i.e. CarbonTracker CT2010; Peters et al., 2007). SRFP has the best seasonality, keeping in mind that difference between the OCPR and SRFP seasonality is not conclusive ($P(H_0 : \sigma_1^2 = \sigma_2^2) = 0.18$).

All SRA values range between 5.4 and 6.2, and no inter-algorithm difference is significant in this aspect (lowest $P(H_0 : \sigma_1^2 = \sigma_2^2) = 0.45$)

7 Summary

Tables 13 and 14 list the overview results using all combined data as well as their 0.95 confidence intervals and the equal variance hypothesis probabilities. The results in Table 13 correspond to the overall (ALL) results in Tables 3, 5, 7 and 9. The reported errors are also derived from this complete data set (using all available data pairs). Keep in mind that the station-to-station range in bias, scatter and correlation often far exceeds the error boundaries in Table 13. That said, looking at Table 13, we see that distinctive differences between algorithms do exist; however Table 14, which features
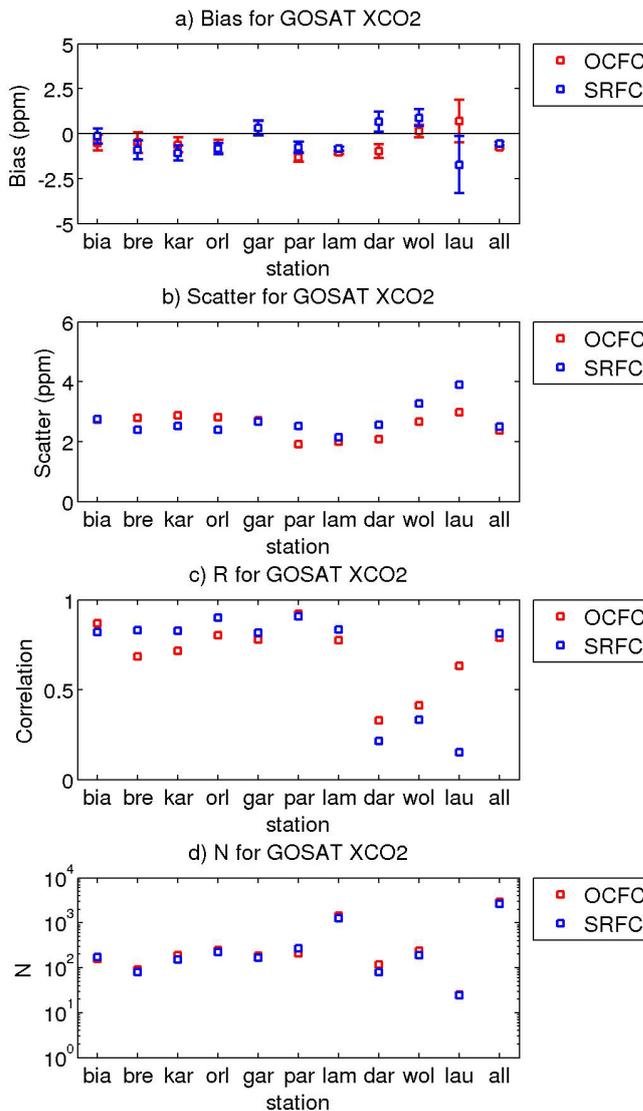
Figure 6. GOSAT XCO$_2$ bias (**a**), scatter (**b**), correlation (**c**) and number of data pairs (**d**), for all individual TCCON stations and all data combined.

the analysis results of the inter-station and seasonal variability, is far more ambiguous. This is of course a direct result of the difference in sample size from which the parameters in Tables 13 and 14 are obtained. Table 13's results are derived from the (many) individual data pairs, while Table 14's sample consists of the limited number of (seasonal) bias means only. Only one inter-algorithm difference parameter reached the 0.95 confidence level (the RA results between OCPR and OCFP XCH$_4$). Of all the H$_0$ probability values (P), only three parameters reach 0.9, four 0.8 and seven parameters a 0.68 confidence level ($\sim 1\sigma$) out of the 18 listed in Table 14. Also, the inter-algorithm differences between RA values are more significant than those of seasonality and SRA, even though the latter is probably the best quality estimator
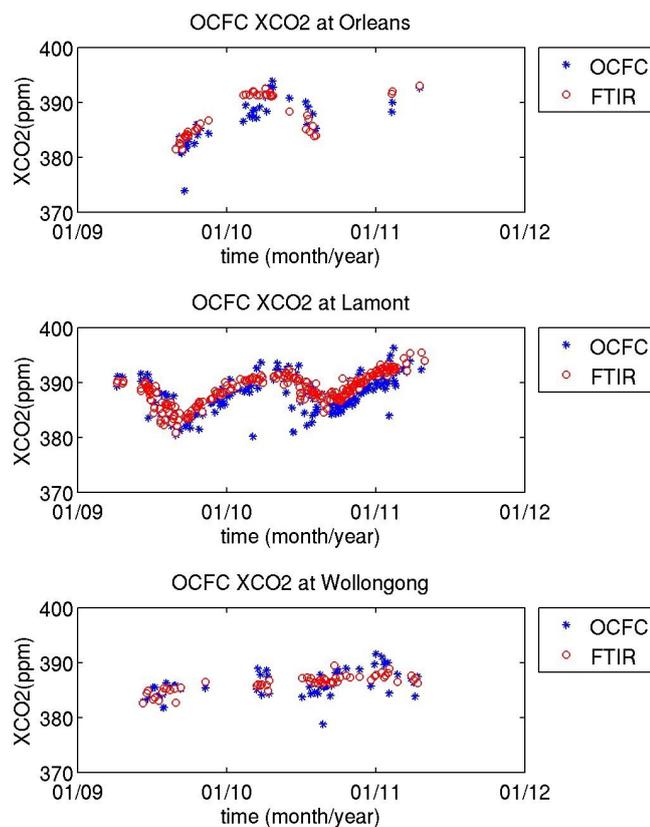
Figure 7. Daily averaged time series of paired OCFC and TCCON FTIR XCO_2 at Orleans (top), Lamont (mid) and Wollongong (bottom).
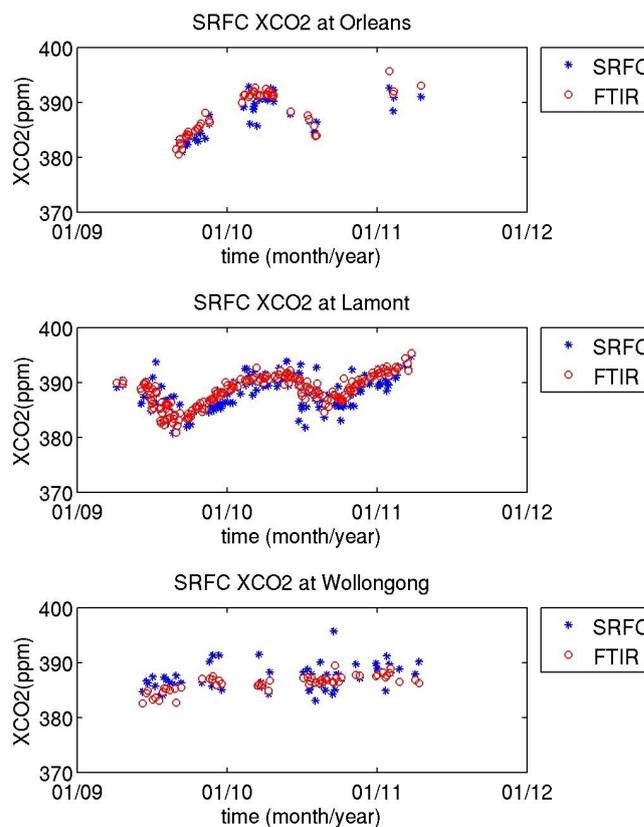
Figure 8. Daily averaged time series of paired SRFC and TCCON FTIR XCO_2 at Orleans (top), Lamont (mid) and Wollongong (bottom).

of what the users have defined as the relative accuracy. To remedy this ambiguity, one would need to increase the number of sample data. One way would be to use monthly instead of seasonal means. However this would greatly increase the number of unstable samples (due to the limited amount of correlative data from which these averages are constructed). This could also be improved by using dynamic collocation criteria, using either the free tropospheric temperature as a proxy for XCO_2 (Keppel-Aleks et al., 2011; Wunch et al., 2011b) or CTM model data (Guerlet et al., 2013). This results in more data and thus a more robust data set. To ensure the robustness of our data we need to reject results which do not meet our quality criteria. A dynamic collocation method would probably result in less rejections and a larger data set. Such an approach was not feasible due to practical considerations. The most desirable option would be to expand the TCCON. Note however that, for instance, to reach 0.95 confidence in the SRA difference between BESD and WFMD XCO_2 (1.19 vs 1.43 ppm), one would need 115 data samples (currently 21). Alternatively, with a perfect sample size of 40 (4 seasons × 10 stations), we can currently only distinguish, with 95 % accuracy, an SRA of 0.5 ppm (the threshold XCO_2 quality) from that of 0.68 ppm. With our best actual samples size (31) the latter becomes 0.72 ppm.

Note also that in the case of XCO_2 none of the algorithms' RA or SRA values reach said relative accuracy threshold value of 0.5 ppm as set forward by the users, nor do the SRA 95 % confidence bands encompass this value. What we obtain is the combined TCCON–satellite accuracy, and according to Wunch et al. (2010) the current TCCON XCO_2 network accuracy (1σ, station to station) is 0.4 ppm. Adding additional uncertainty due to collocation and smoothing errors, and the above-mentioned uncertainty on the analysis itself, leaves little room for an accurate assessment of such a demanding threshold value for inverse modelling purposes. Efforts to decrease the station-to-station biases between TCCON stations are thus desirable and ongoing (Hase et al., 2013; Dohe et al., 2013). For XCH_4, SRA reaches the 10 ppb user quality threshold for all GOSAT algorithms, while SCIAMACHY WFMD's SRA approaches this number (10.5). IMAP would probably also meet this threshold if not for the Southern Hemisphere bias.

Even taking into account these uncertainties, at least in certain comparison rounds, the differences between the algorithm products were distinct enough to draw binding conclusions as to which one would proceed into the next round of the GHG-CCI project. Again it must be stressed that the decisions reached were not based on the comparisons with TCCON alone (see Buchwitz et al., 2013).
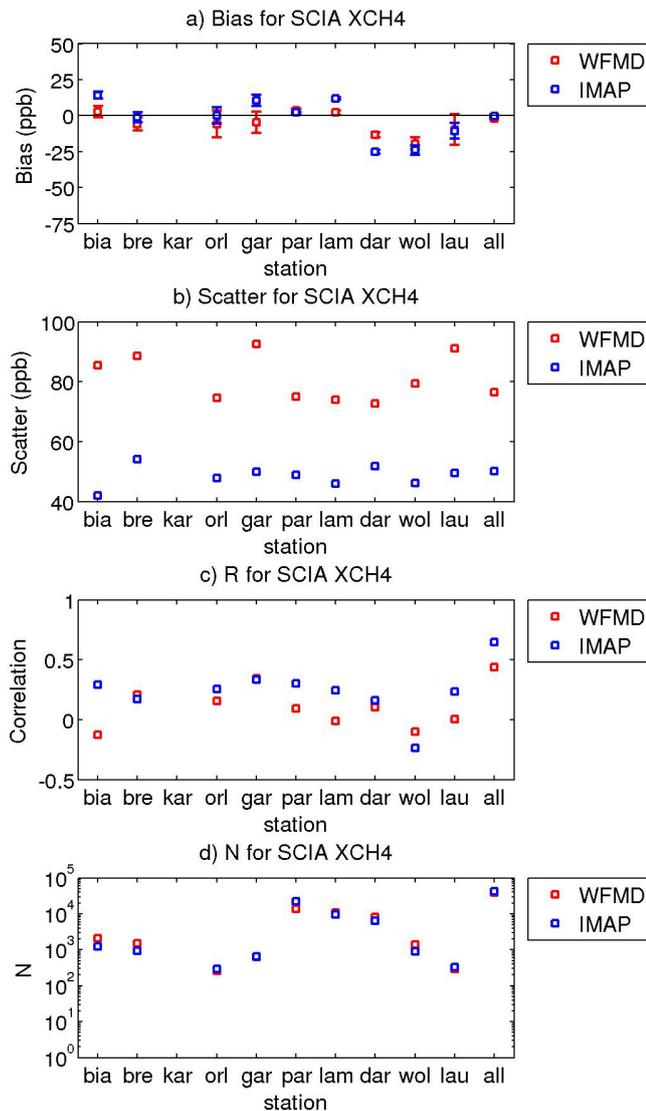
Figure 9. SCIAMACHY XCH$_4$ bias **(a)**, scatter **(b)**, correlation **(c)** and number of data pairs **(d)**, for all individual TCCON stations and all data combined.
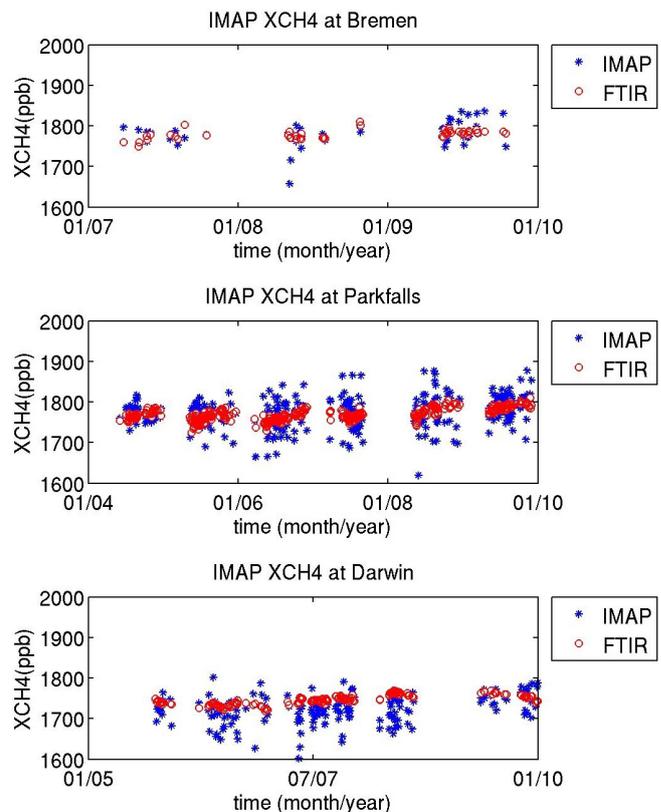

Figure 10. Daily averaged time series of paired IMAP and TCCON FTIR XCH$_4$ at Bremen (top), Park Falls (mid) and Darwin (bottom).

In the case of SCIAMACHY XCO$_2$, we see that BESD has a superior bias, scatter and correlation compared to WFMD. Its RA, Seas and SRA values are also consistently better, albeit only the RA with reasonable confidence ($P(H_0 : \sigma_1^2 = \sigma_2^2) = 0.06$). So in this round the conclusion was to proceed with BESD.

The GOSAT XCO$_2$ comparisons on the other hand yielded no clear winner. Both have comparable scatter and correlation values (in fact using different collocation criteria yielded different winners in this category). OCFC's RA value is slightly better, while its seasonality and SRA are slightly worse. Neither of these parameters is distinctive. As discussed in Buchwitz et al. (2013) global analysis of the data does yield, contrary to the TCCON locations, significant

inter-algorithm differences. Certainly in areas with high (e.g. the Sahara) or low (e.g. Amazon forest) surface albedo, which are not covered by TCCON, differences become significant (Guerlet et al., 2013). This observation triggered the development of a new algorithm which uses ensemble medians called the ensemble median algorithm (EMMA; see Reuter et al., 2012). While the EMMA algorithm might be the best solution at hand, it does not negate the pressing need for expanding the TCCON into key areas, enlarging the surface albedo range and geographical distribution of the network.

The SCIAMACHY XCH$_4$ comparisons between IMAP and WFMD, showed that in many aspects IMAP was the best-performing algorithm (scatter, data density, correlation). However the inter-station bias difference, certainly between the Northern and Southern Hemisphere, appears to be large. This results in an inferior RA and SRA value (although the statistical certainty of the latter parameter is far less distinct, and the RA difference reaches a 0.9 confidence level only). WFMD also shows an inter-hemispheric bias difference, only less significant. Also neither of the algorithms reach the threshold single observation precision (< 34 ppb), set forward by the users. Since these issues need to be resolved first,
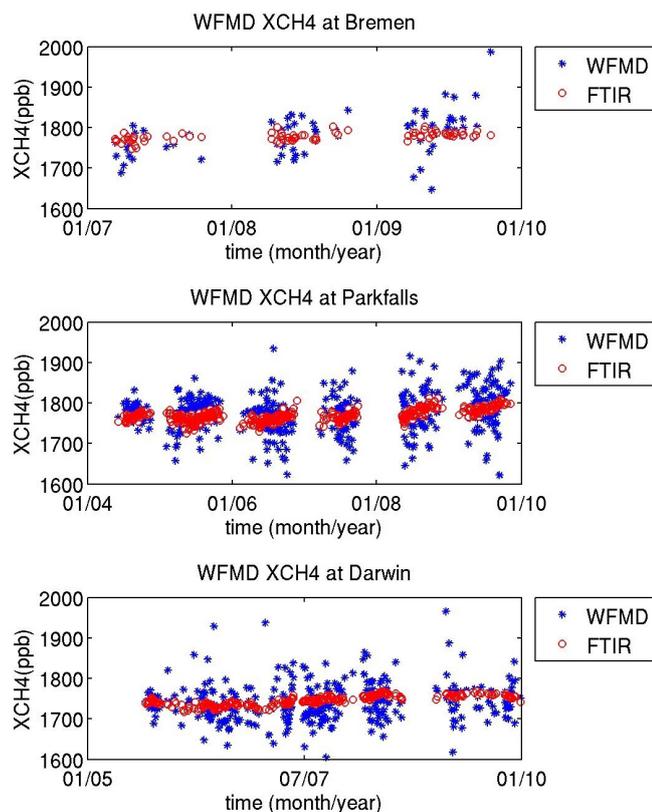
Figure 11. Daily averaged time series of paired WFMD and TC-CON FTIR XCH_4 at Bremen (top), Park Falls (mid) and Darwin (bottom).
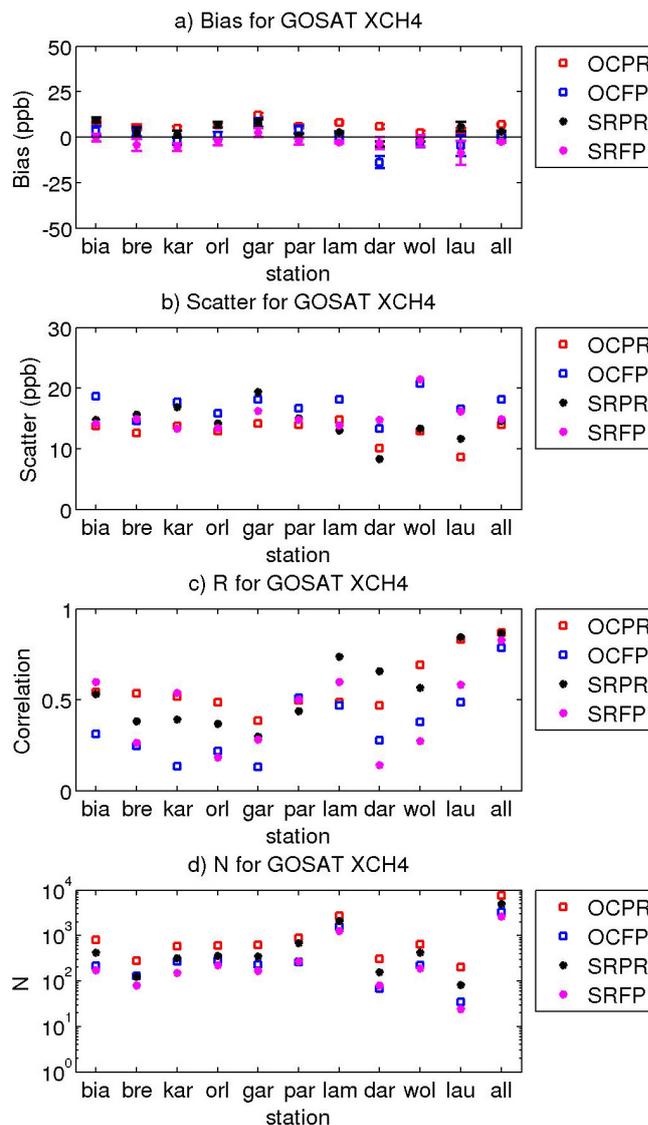

Figure 12. GOSAT XCH_4 bias (**a**), scatter (**b**), correlation (**c**) and number of data pairs (**d**), for all individual TCCON stations and all data combined.

both algorithms proceeded to the next round. Note that the results are only representative of the time period after 2005. That year featured a SCIAMACHY detector degradation in the spectral region used for methane retrievals, causing a significant deterioration of the methane retrieval quality (e.g., Buchwitz et al., 2013, and references given therein).

For GOSAT XCH_4, it is the less mature OCFP algorithm that stands out in a negative way. It has distinctively more scatter and a lower correlation coefficient, and its RA value is distinctly worse than its proxy OCPR counterpart. OCPR on the other hand has the lowest scatter and highest data density, as well as the lowest RA value (although hardly distinct from its SRFP competitor). Neither of the algorithms, including OCFP, have a distinct SRA value. The margin in which OCPR stands out from its SRON competitors is however very small, and in terms of seasonality it seems to perform worse (although with a $0.18 P(H_0 : \sigma_1^2 = \sigma_2^2)$, only with little more than 80 % certainty). Given this small margin and the fact that the comparison between the proxy and full-physics SRON products shows that the full-physics method is a viable option, it was decided to proceed with both OCPR and SRFP.

8   Conclusions

We have analysed 10 retrieval products produced by the BESD, WFM-DOAS, IMAP-DOAS, RemoTeC and Leicester OCO algorithms. We focussed specifically on the inter-product differences. It was found that for SCIAMACHY (both XCO_2 and XCH_4) the competing algorithms yielded significantly different products – especially in terms of single measurement precision (i.e. scatter). In both XCO_2 and XCH_4, WFMD featured higher scatter than its competitor, being BESD for XCO_2 and IMAP for XCH_4. The latter on the other hand seems to suffer (more) from a significant Northern vs. Southern Hemisphere bias, an issue which requires more analysis. One reason for the larger scatter of
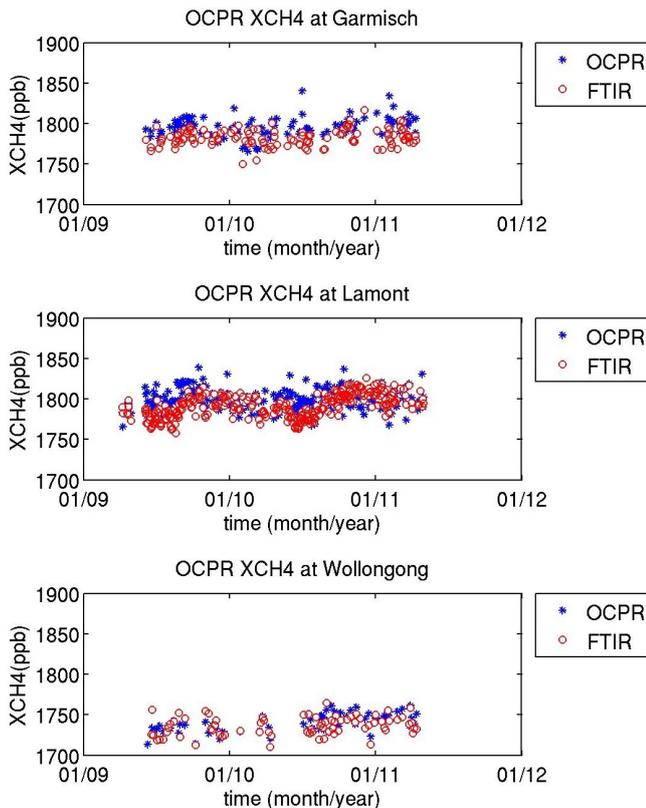
Figure 13. Daily averaged time series of paired OCPR and TCCON FTIR XCH$_4$ at Garmisch (top), Lamont (mid) and Wollongong (bottom).
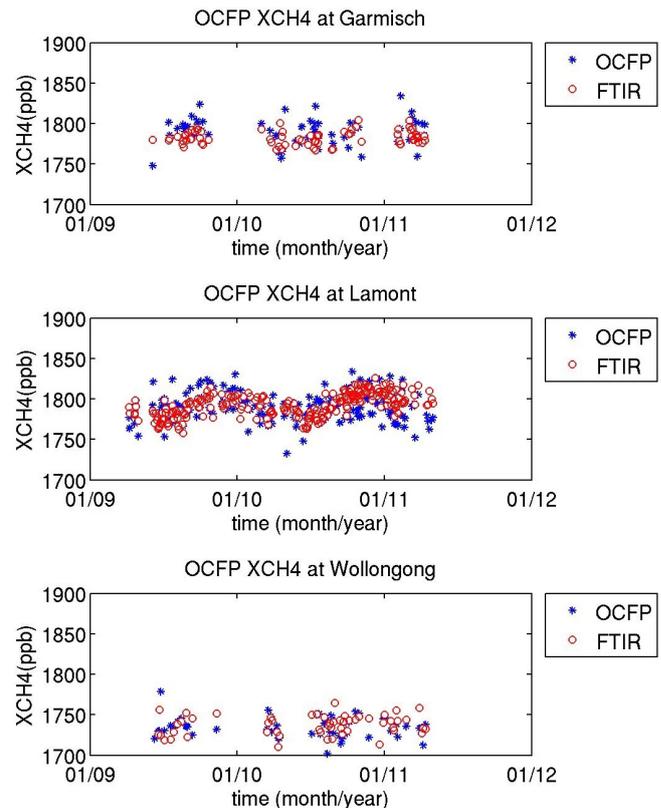

Figure 14. Daily averaged time series of paired OCFP and TCCON FTIR XCH4 at Garmisch (top), Lamont (mid) and Wollongong (bottom).

the WFMD data product is that WFMD is based on unconstrained linear least squares, whereas BESD and IMAP are based on optimal estimation. However, there are several other retrieval properties – e.g. meteorological profiles, cloud filtering and consideration of light scattering – which influence the retrieval scatter.

Differences between all the competing GOSAT products are far less striking. For XCH$_4$, apart from the full-physics version of Leicester's OCO algorithm (OCFP), the other algorithms (OCPR and SRON's RemoTeC full-physics (SRFP) and proxy (SRPR) products) are very alike. In terms of precision, the proxy versions, especially OCPR, seem to have a slight edge, but in terms of inter-station bias variability and capturing the seasonal cycle, the SRON full-physics algorithm is more than competitive. In fact there are indications that OCPR underestimates the XCH$_4$, but this might be due to the CarbonTracker CT2010 model (Peters et al., 2007) used in the dry-air conversion instead of the proxy algorithm itself. For XCO$_2$, the competing products are closer still. Differences are small for all obtained statistical parameters, and no one algorithm betters the others across the board. This does not imply that these products feature no differences at all. In some regions (e.g. South America, Africa, China) differences between algorithms can be substantial, but there are

no TCCON data available in these regions to discriminate between algorithm performance (Buchwitz et al., 2013; Reuter et al., 2012).

The relative accuracy and single precision threshold quality criteria for inverse modelling (10 ppb and 34 ppb XCH$_4$, respectively) have been reached by all GOSAT XCH$_4$ products, and if the inter-hemispheric bias difference is mitigated (in a future version of the product or by using in situ data; see Bergamachi et al., 2009), so will be, for the relative accuracy, in all likelihood, the SCIAMACHY XCH$_4$ products. However both IMAP and WFMD XCH$_4$ still do not reach the precision user requirement. Again it needs to be pointed out that the validation results presented here are dominated by data generated after 2005 when SCIAMACHY suffered from detector degradation in the spectral region used for methane retrieval. The results presented here are therefore not representative of the time period 2005 and earlier years, where the quality of the SCIAMACHY methane retrievals is much better (e.g., Buchwitz et al., 2013, and references given therein).

For XCO$_2$, all algorithms reach the single observation precision threshold (8 ppm), but none of the algorithms meet the relative accuracy user requirement (0.5 ppm). Unfortunately, given the current constellation of TCCON measurements, the assessment of whether an algorithm product has
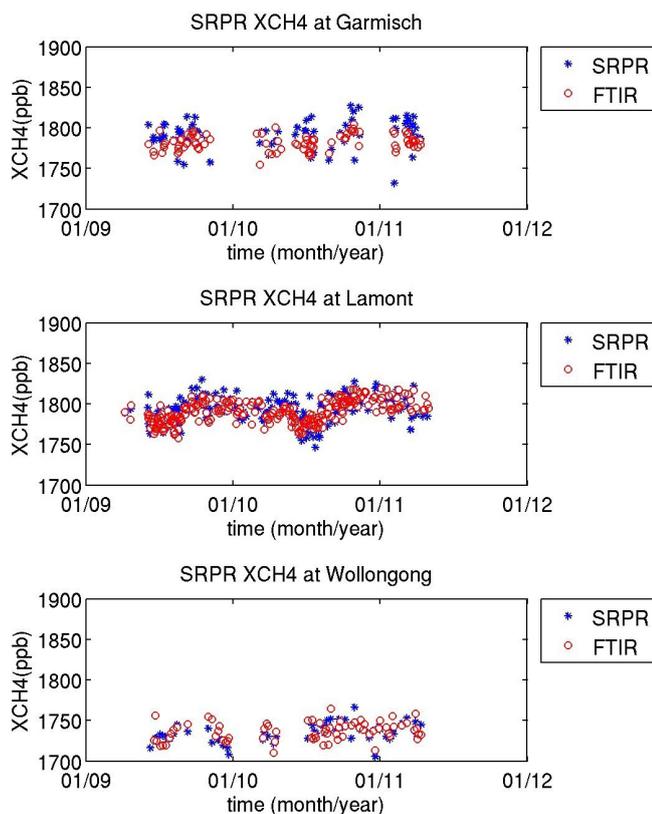
Figure 15. Daily averaged time series of paired SRPR and TC-CON FTIR XCH$_4$ at Garmisch (top), Lamont (mid) and Wollongong (bottom).
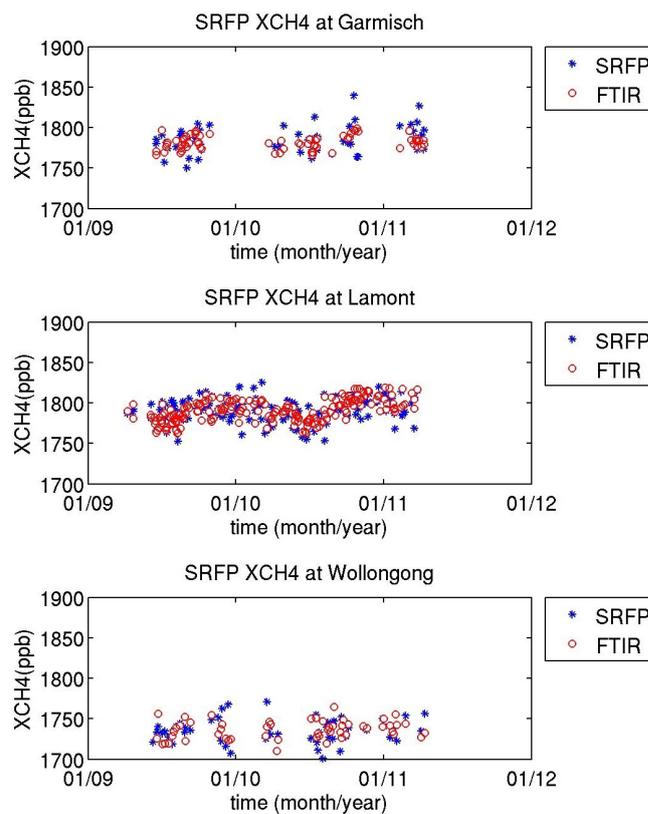

Figure 16. Daily averaged time series of paired SRFP and TC-CON FTIR XCH$_4$ at Garmisch (top), Lamont (mid) and Wollongong (bottom).

indeed reached this demanding value contains considerable uncertainty by itself. An expansion of the TCCON into key geographic areas and efforts to even further reduce the TC-CON station-to-station biases would be most welcome in this respect.

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 oftransport model errors on the interannual variability of regional CO$_2$ fluxes, 1988–2003, Global Biogeochem. Cy., 20, GB1002, doi:10.1029/2004GB002439, 2006.

Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Villani, M. G., Houweling, S., Dentener, F., Dlugokencky, E. J., Miller, J. B., Gatti, L. V., Engel, A., Levin, I.: Inverse modeling of global and regional CH$_4$ emissions using SCIA-MACHY satellite retrievals, J. Geophys. Res., 114, D22301, doi:10.1029/2009JD012287, 2009.

Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A. H. P.: SCIA-MACHY – Mission objectives and measurement modes, J. Atmos. Sci., 56, 127–150, 1999.

Buchwitz, M., Rozanov, V. V., and Burrows, J. P.: A near-infrared optimized DOAS method forthe fast global retrieval of atmospheric CH$_4$, CO, CO$_2$, H$_2$O, and N$_2$O total column amounts from SCIAMACHY Envisat-1 nadir radiances, J. Geophys. Res., 105, 15231–15245, 2000.

Buchwitz, M., de Beek, R., Burrows, J. P., Bovensmann, H., Warneke, T., Notholt, J., Meirink, J. F., Goede, A. P. H., Bergamaschi, P., Körner, S., Heimann, M., and Schulz, A.: Atmospheric methane and carbon dioxide from SCIAMACHY satellite data: initial comparison with chemistry and transport models, Atmos. Chem. Phys., 5, 941–962, doi:10.5194/acp-5-941-2005, 2005.

Buchwitz, M., Schneising, O., Burrows, J. P., Bovensmann, H., Reuter, M., and Notholt, J.: First direct observation of the atmospheric CO_2 year-to-year increase from space, Atmos. Chem. Phys., 7, 4249–4256, doi:10.5194/acp-7-4249-2007, 2007.

Buchwitz, M., Chevallier, F., Bergamaschi, P. : User Requirements Document (URD) for the GHG-CCI project of ESA's Climate Change Initiative, Technical Report, 45 pp., version 1(URDv1), 3 February 2011, available at: http://www.esa-ghg-cci.org (last access: 5 June 2014), 2012.

Buchwitz, M., Reuter, M., Schneising, O., Boesch, H., Guerlet, S., Dils, B., Aben, I., Armante, R., Bergamaschi, P., Blumenstock, T., Bovensmann, H., Brunner, D., Buchmann, B., Burrows, J. P., Butz, A.,Chédin, A., Chevallier, F., Crevoisier, C. D., Deutscher, N. M., Frankenberg, C., Hase, F., Hasekamp, O. P., Heymann, J., Kaminski, T., Laeng, A., Lichtenberg, G., De Mazière, M., Noël, S., Notholt, J. Orphal, J., Popp, C., Parker, R., Scholze, M., Sussmann, R., Stiller, G. P., Warneke, T., Zehner, C., Bril, A., Crisp, D., Griffith, D. W. T.,Kuze, A., O'Dell, C., Oshchepkov, S., Sherlock, V., Suto, H., Wennberg, P., Wunch, D., Yokota, T., and Yoshida, Y.: The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparison and quality assessment of near-surface-sensitive satellite-derived CO_2 and CH_4 global data sets, Remote Sens. Environ., doi:10.1016/j.rse.2013.04.024, 19 pp., in press, corrected proof, 2013.

Butz, A., Guerlet, S., Hasekamp, O., Schepers, D., Galli, A., Aben, I., Frankenberg, C., Hartmann, J.-M., Tran, H., Kuze, A., Keppel-Aleks, G., Toon, G., Wunch, D., Wennberg, P., Deutscher, N., Griffith, D., Macatangay, R. , Messerschmidt, J., Notholt, J., and Warneke, T. : Towards accurate CO_2 and CH_4 observations from GOSAT, Geophys. Res. Lett., 38, L14812, doi:10.1029/2011GL047888, 2011.

Canadell, J. G., Ciais, P., Dhakal, S., Dolman, H., Friedlingstein, P., Gurney. K. R,, Held, A., Jackson,R. B., Le Quéré, C., Malone, E. L., Ojima, D. S., Patwardhan, A., Peters, G. P., and Raupach, M. R.: Interactions of the carbon cycle, human activity, and the climate system: a research portfolio, Curr. Opinion Environ. Sustain., 2, 301–311, 2010.

Chevallier, F., Engelen, R. J., and Peylin, P.: The contribution of AIRS data to the estimation of CO_2 sources and sinks. Geophys. Res. Lett., 32, L23801, doi:10.1029/2005GL024229, 2005.

Chevallier, F., Bréon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework, J. Geophys. Res., 112, D09307, doi:10.1029/2006JD007375, 2007.

Cogan, A. J., Boesch, H, Parker, R. J., Feng, L., Palmer, P. I., Blavier, J.-F. L., Deutscher, N. M., Macatangay, R., Notholt, J., Roehl, C., Warneke, T., and Wunch, D. : Atmospheric carbon dioxide retrieved from the Greenhouse gases Observing SATellite (GOSAT): Comparison with ground-based TCCON observations and GEOS-Chem model calculations, J. Geophys. Res., 117, D21301, doi:10.1029/2012JD018087, 2012.

Crisp, D., Atlas, R. M., Bréon, F.-M., Brown, L. R., Burrows, J. P., Ciais, P., Connor, B. J., Doney, S.C., Fung, I. Y., Jacob, D. J., Miller, C. E., O'Brien, D., Pawson, S., Randerson, J. T., Rayner, P., Salawitch, R. S., Sander, S. P., Sen, B., Stephens, G. L., Tans, P. P., Toon, G. C.,Wennberg, P. O., Wofsy, S. C., Yung, Y. L., Kuang, Z., Chudasama, B., Sprague, G., Weiss,P., Pollock, R., Kenyon, D., and Schroll, S.: The Orbiting Carbon Observatory (OCO) mission, Adv. Space Res., 34, 700–709, 2004.

Deutscher, N. M., Griffith, D. W. T., Bryant, G. W., Wennberg, P. O., Toon, G. C., Washenfelder, R. A., Keppel-Aleks, G., Wunch, D., Yavin, Y., Allen, N. T., Blavier, J.-F., Jiménez, R., Daube, B. C., Bright, A. V., Matross, D. M., Wofsy, S. C., and Park, S.: Total column CO_2 measurements at Darwin, Australia – site description and calibration against in situ aircraft profiles, Atmos. Meas. Tech., 3, 947–958, doi:10.5194/amt-3-947-2010, 2010.

Dlugokencky, E. J., Bruhwiler, L., White, J. W. C., Emmons, L. K., Novelli, P. C., Montzka, S. A., Masarie, K. A., Lang, P. M., Crotwell, A. M., Miller, J. B., and Gatti, L. V.: Observational constraints on recent increases in the atmospheric CH_4 burden, Geophys. Res. Lett., 36, L18803, doi:10.1029/2009GL039780, 2009.

Dohe, S., Sherlock, V., Hase, F., Gisi, M., Robinson, J., Sepúlveda, E., Schneider, M., and Blumenstock, T.: A method to correct sampling ghosts in historic near-infrared Fourier transform spectrometer (FTS) measurements, Atmos. Meas. Tech., 6, 1981–1992, doi:10.5194/amt-6-1981-2013, 2013.

Frankenberg, C., Meirink, J. F., van Weele, M., Platt, U., and Wagner, T. : Assessing methane emissions from global spaceborne observations, Science, 308, 1010–1014, 2005.

Frankenberg, C., Aben, I., Bergamaschi, P., Dlugokencky, E. J., van Hees, R., Houweling, S., van derMeer, P., Snel, R., and Tol, P.: Global column-averaged methane mixing ratios from 2003 to 2009 as derived from SCIAMACHY: Trends and variability, J. Geophys. Res., 116, D04302, doi:10.1029/2010JD014849, 2011.

Geibel, M. C., Messerschmidt, J., Gerbig, C., Blumenstock, T., Chen, H., Hase, F., Kolle, O., Lavrič, J. V., Notholt, J., Palm, M., Rettinger, M., Schmidt, M., Sussmann, R., Warneke, T., and Feist, D. G.: Calibration of column-averaged CH_4 over European TCCON FTS sites with airborne in-situ measurements, Atmos. Chem. Phys., 12, 8763–8775, doi:10.5194/acp-12-8763-2012, 2012.

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., Kyro, E., Morino, I., Sherlock, V., Sussmann, R., Galli, A., and Aben, I.: Impact of aerosol and thin cirrus on retrieving and validating XCO_2 from GOSAT shortwave infrared measurements, J. Geophys. Res. Atmos., 118, 4887–4905, doi:10.1002/jgrd.50332, 2013.

Hase, F., Drouin, B. J., Roehl, C. M., Toon, G. C., Wennberg, P. O., Wunch, D., Blumenstock, T., Desmet, F., Feist, D. G., Heikkinen, P., De Mazière, M., Rettinger, M., Robinson, J., Schneider, M., Sherlock, V., Sussmann, R., Té, Y., Warneke, T., and Weinzierl, C.: Calibration of sealed HCl cells used for TCCON instrumental line shape monitoring, Atmos. Meas. Tech., 6, 3527–3537, doi:10.5194/amt-6-3527-2013, 2013.

Heymann, J., Schneising, O., Reuter, M., Buchwitz, M., Rozanov, V. V., Velazco, V. A., Bovensmann, H., and Burrows, J. P.: SCIAMACHY WFM-DOAS XCO2: comparison with CarbonTracker

XCO$_2$ focusing on aerosols and thin clouds, Atmos. Meas. Tech., 5, 1935–1952, doi:10.5194/amt-5-1935-2012, 2012a.

Heymann, J., Bovensmann, H., Buchwitz, M., Burrows, J. P., Deutscher, N. M., Notholt, J., Rettinger, M., Reuter, M., Schneising, O., Sussmann, R., and Warneke, T.: SCIAMACHY WFM-DOAS XCO$_2$: reduction of scattering related errors, Atmos. Meas. Tech., 5, 2375–2390, doi:10.5194/amt-5-2375-2012, 2012b.

Jacobson, M. Z.: Correction to "Control of fossil-fuel particulate black carbon and organic matter, possibly the most effective method of slowing global warming", J. Geophys. Res., 110, D14105, doi:10.1029/2005JD005888, 2005.

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.

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.

Meirink, J. F., Eskes, H. J., and Goede, A. P. H.: Sensitivity analysis of methane emissions derived from SCIAMACHY observations through inverse modelling, Atmos. Chem. Phys., 6, 1275–1292, doi:10.5194/acp-6-1275-2006, 2006.

Messerschmidt, J., Macatangay, R., Notholt, J., Petri, C., Warneke, T., and Weinzierl, C. : Side by side measurements of CO$_2$ by ground-based Fourier transform spectrometry (FTS). Tellus B, 62, 749–758, doi:10.1111/j.1600-0889.2010.00491.x, 2010.

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., Lavrič, 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 CO$_2$: the first aircraft campaign over European TCCON sites, Atmos. Chem. Phys., 11, 10765–10777, doi:10.5194/acp-11-10765-2011, 2011.

O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., 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.: The ACOS CO$_2$ retrieval algorithm – Part 1: Description and validation against synthetic observations, Atmos. Meas. Tech., 5, 99–121, doi:10.5194/amt-5-99-2012, 2012.

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., Kyro, E., Heikkinen, P., Feist, D. G., Nagahama, T., Kadygrow, N., Maksyutov, S., Uchino, O., and Watanabe, H.: Effects of atmospheric light scattering on spectroscopic observations of greenhouse gases from space: Validation of PPDF-based CO$_2$ retrievals from GOSAT, J. Geophys. Res. Atmos., 117, D12305, doi:10.1029/2012JD017505, 2012.

Parker, R., Boesch, H., Cogan, A., Fraser, A., Feng, L, Palmer, P., Messerschmidt, J., Deutscher, N., Griffith, D., Notholt, J., Wennberg, P. and Wunch, D.: Methane Observations from the Greenhouse gases Observing SATellite: Comparison to ground-

based TCCON data and Model Calculations, Geophys. Res. Lett., 38, L15807, doi:10.1029/2011GL047871, 2011.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Pétron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, Proc. Natl. Acad. Sci. USA, 104, 18925–18930, doi:10.1073/pnas.0708986104, 2007.

Prather, M. J.: Lifetimes and eigenstates in atmospheric chemistry, Geophys. Res. Lett., 21, 801–804, doi:10.1029/94GL00840, 1994.

Prather, M. J., Ehhalt, D., Dentener, F., Derwent, R., Dlugokencky, E., Holland, E., Isaksen, I., Katima, J., Kirchho, V., Matson, P., Midgley, P., and Wang, M.: Atmospheric Chemistry and Greenhouse Gases, in: Climate Change 2001: The Scientific Basis, Third Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Houghton, J., Ding, Y., Griggs, D., Noguer, M., van der Linden, P., Dai, X., Maskell, K., and Johnson, C., Cambridge University Press, Cambridge, UK, 2001.

Reuter, M., Buchwitz, M., Schneising, O., Heymann, J., Bovensmann, H., and Burrows, J. P.: A method for improved SCIAMACHY CO$_2$ 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., Velazco, V., Warneke, T., Wennberg, P. O., and Wunch, D.: Retrieval of atmospheric CO$_2$ with enhanced accuracy and precision from SCIAMACHY: Validation with FTS measurements and comparison with model results, J. Geophys. Res., 116, D04301, doi:10.1029/2010JD015047, 2011.

Reuter, M. and the GHG-CCI project team: ESA Climate Change Initiative (CCI) Algorithm Theoretical Basis Document Version 1 (ATBDv1) for the Essential Climate Variable (ECV) Greenhouse Gases (GHG), 15 March 2012, available at: http://www.esa-ghg-cci.org (last access: 1 June 2014), 2012.

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 CO$_2$ concentrations for surface flux inversions: the ensemble median algorithm EMMA, Atmos. Chem. Phys., 13, 17710–1780, doi:10.5194/acp-13-1771-2013, 2013.

Rigby, M., Prinn, R. G., Fraser, P. J., Simmonds, P. G., Langenfelds, R. L., Huang, J., Cunnold, D. M.,Steele, L. P., Krummel, P. B., Weiss, R. F., O'Doherty, S., Salameh, P. K., Wang, H. J., Harth,C. M., Mühle, J., and Porter, L. W. : Renewed growth of atmospheric methane, Geophys. Res. Lett., 35, L22805, doi:10.1029/2008GL036037, 2008.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific Publishing, 2000.

Schepers, D., Guerlet, S., Butz, A., Landgraf, J., Frankenberg, C., Hasekamp, O., Blavier, J.-F.,Deutscher, N. M., Griffith, D. W. T., Hase, F., Kyro, E., Morino, I., Sherlock, V., Sussmann, R., and Aben, I.: Methane retrievals from Greenhouse Gases Observ-

ing Satellite (GOSAT) shortwave infrared measurements: Performance comparison of proxy and physics retrieval algorithms, J. Geophys. Res., 117, D10307, doi:10.1029/2012JD017549, 2012.

Schneising, O., Buchwitz, M., Burrows, J. P., Bovensmann, H., Reuter, M., Notholt, J., Macatangay, R., and Warneke, T.: Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite – Part 1: Carbon dioxide, Atmos. Chem. Phys., 8, 3827–3853, doi:10.5194/acp-8-3827-2008, 2008.

Schneising, O., Buchwitz, M., Burrows, J. P., Bovensmann, H., Bergamaschi, P., and Peters, W.: Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite – Part 2: Methane, Atmos. Chem. Phys., 9, 443–465, doi:10.5194/acp-9-443-2009, 2009.

Schneising, O., Buchwitz, M., Reuter, M., Heymann, J., Bovensmann, H., and Burrows, J. P.: Long-term analysis of carbon dioxide and methane column-averaged mole fractions retrieved from SCIAMACHY, Atmos. Chem. Phys., 11, 2863–2880, doi:10.5194/acp-11-2863-2011, 2011.

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.

Shapiro, S. S. and Wilk, M. B. :An analysis of variance test for normality (complete samples), Biometrika, 52, 591–611, doi:10.1093/biomet/52.3-4.591, 1965.

Simpson, I. J., Sulbaek Andersen, M. P., Meinardi, S., Bruhwiler, L., Blake, N. J., Helmig, D., Rowland, F. S., and Blake, D. R. : Long-term decline of global atmospheric ethane concentrations and implications for methane, Nature, 488, 490–494, doi:10.1038/nature11342, 2012.

Snedecor, G. W. and Cochran, W. G.: Statistical Methods, 8th Edn., Iowa State University Press, 1989.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L. (Eds.): Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

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., and Denning, A. S.: Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO_2, Science, 316, 1732–1735, doi:10.1126/science.1137004, 2007.

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

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, Phil. Trans. 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.