ESA Climate Change Initiative (CCI)

Product Validation and Algorithm Selection Report (PVASR)

for the Essential Climate Variable (ECV)

Greenhouse Gases (GHG)

Written by:

GHG-CCI VALidation Team (VALT):

Justus Notholt (lead author), Thomas Blumenstock, Dominik Brunner, Brigitte Buchmann, Bart Dils, Martine De Mazière, Christoph Popp, Ralf Sussmann

Change log:

Version Nr.	Date	Status	Reason for change
Draft 0.1	9. Aug. 2011	Initial Draft Author: J. Notholt	New document. Main purpose: Input for discussion with GHG-CCI project team.
Draft 0.2	19. August 2011	Improved draft	Comments from validation team considered.
Draft 0.3	30. July 2012	Draft with initial Round Robin (RR) results.	Results from RR exercise included
Draft 0.4	31. July 2012	Draft with final Round Robin validation results	Comments from VALT team considered
Draft 1.0	2. August 2012	Final draft for final iteration with entire GHG-CCI team	Minor improvements
Final version	22 August 2012	Final version	Final version after comments from TCCON partners

Table of Contents

1 Executive Summary

This document is the Product Validation and Algorithm Selection Report (PVASR), which is a deliverable of the ESA project GHG-CCI (http://www.esa-ghg-cci.org/).

The GHG-CCI project started on 1st September 2010. The GHG-CCI project is one of several projects of ESA's Climate Change Initiative (CCI). The GHG-CCI project aims to deliver the Essential Climate Variable (ECV) Greenhouse Gases (GHG) in line with the "Systematic observation requirements for satellite-based products for climate" as defined by GCOS (Global Climate Observing System): "Product A.9: Distribution of greenhouse gases, such as CO_2 and CH_4, of sufficient quality to estimate regional sources and sinks".

This document has been written by the GHG-CCI VALidation Team (VALT). VALT is led by J. Notholt of the IUP, University of Bremen. VALT members are Thomas Blumenstock (KIT, Karlsruhe, Germany), Bart Dils and Martine De Mazière, (BIRA, Belgium), Dominik Brunner, Brigitte Buchmann and Christoph Popp (EMPA, Switzerland) and Ralf Sussmann (KIT, Garmisch, Germany).

The PVASR describes the analysis done by VALT during the Round Robin (RR) phase of the project. It presents and discusses validation results obtained by comparing the satellite retrievals with ground-based observations. The main purpose of this document is to estimate the quality of the satellite data products contained in the Round Robin Data Package (RRDP) and to provide the VALT recommendations which algorithms to use to generate the GHG-CCI Climate Research Data Package (CRDP).

The final decision will however not only be based on this document primarily because of the sparseness of the TCCON sites. To mitigate this limitation, the GHG-CCI satellite retrieval team (EOST) is performing an analysis of the global data sets, e.g., by generating global maps and regional time series including comparisons with global models. The results of this independent analysis will be documented in the Algorithm Inter-comparison and Error Characterization and Analysis Report (AIECAR version 1.0 (AIECARv1)). The final decision will be presented in a short separate synthesis document approved by the GHG-CCI Climate Research Group (CRG) called Algorithm Selection Report (ASR). ASR will focus on the decision results and justification referring to PVASR (i.e., this document) and AIECARv1 for details.

The VALT has compared the four core GHG-CCI ECV data products XCO_2 and XCH_4 from SCIAMACHY/ENVISAT and TANSO/GOSAT, generated with ECV Core Algorithms (ECAs), with ground-based reference data obtained from the Total Carbon Column Observing Network (TCCON). For each of the four GHG-CCI core data products several algorithms have been further developed and used within GHG-CCI in competition to generate global XCO_2 and/or XCH_4 data sets. This document describes the validation of these data sets using TCCON retrievals and gives recommendations on which algorithms perform best and shall be used to generate the CRDP.

In order to give clear and well justified recommendations, the following activities have been carried out by VALT:

All candidate ECA algorithms (see section 3) have been compared with data from ground-based FTS data from the TCCON network (see section 4.2, Table 1). Only satellite data that fell within 2 hours and 100, 350 or 500 km of a FTS measurement are used in this comparison. Reduced sample sizes often limited the statistical validity of results for the 100 and 350 km collocations. Therefore, the validation results presented here correspond to the 2 hours/500 km collocation criteria.

All FTS measurements that fall within ±2 hour of a satellite measurement are first averaged. The averaged values are then used in the validation process. Since the focus of this validation exercise was to intercompare various retrieval algorithms, a common time period was used in the four intercomparison rounds, RR1 to RR4 (RR1: SCIAMACHY/XCH_4, RR2: TANSO/XCH_4, RR3: SCIAMACHY/XCO_2, and RR4: TANSO/XCO_2). For example, the 2010 IMAP XCH_4 results have been ignored, since WFMD XCH_4 only extended up to 2009. Yearly analyses have been performed as well but are not shown in this report for the sake of brevity, nor did the results impact the VALT recommendations. The statistical parameters calculated, referred to as Figures of Merit (FoM), include the bias, scatter, number of data, and correlation coefficients. These numbers have been calculated for individual, daily and monthly averaged data, for each TCCON station and for all data combined. The same FoM are also calculated per season. All in all, 75 parameters are calculated for each TCCON station. These parameters and the plotted timeseries were sufficient for a well-grounded recommendation on the ECV selection process.

Results for XCH_4 from SCIAMACHY/ENVISAT:

- The algorithms used are WFMD (IUP, Univ. Bremen) and IMAP (SRON/JPL).
- VALT recommendation incl. justification:
  - Recommendation: VALT recommends proceeding with both algorithms, WFMD and IMAP.
  - Justification: The two SCIAMACHY methane data products do not yet fulfil the relative accuracy threshold requirement (< 10 ppb) under all conditions and both algorithms appear to have different strength and weaknesses. Both data sets show similar biases and this criterion cannot be used to recommend only one of the two algorithms. At certain TCCON sites the biases are significantly larger (~30 ppb) than the required relative accuracy. These biases are not yet well understood and more studies are needed to identify the reasons for that. Both algorithms need further development (including how to better/optimally use the Level 1 data and/or to select the successful retrievals, i.e., including pre- and post-processing procedures and filtering) to ensure that the relative accuracy requirement is met under all conditions. If this can be achieved with only one or even with both algorithms is not clear and therefore it is recommended to proceed with both algorithms. The IMAP data product shows less scatter and therefore also higher correlation with the TCCON reference data. This is an advantage for IMAP and a disadvantage for WFMD. However, IMAP uses global CH_4 model data as a priori information whereas WFMD is independent of any global CH_4 model. Certain users may prefer a data product independent of any global CH_4 model. This is a pro for WFMD and a con for IMAP.

XCH_4 from TANSO/GOSAT:

- Algorithms / data products: Both algorithms come in a full physics and proxy version. The algorithms are OCPR & OCFP (Univ. Leicester) and SRPR & SRFP (SRON).
- VALT recommendation incl. justification:
  - Recommendation: VALT recommends to proceed with two of the four algorithms, namely with SRFP and OCPR.
  - Justification: From the four algorithms used in this round, only OCFP seems to perform significantly worse than its competitors on several key statistics. On the other hand its proxy method counterpart edges out its SRON competitors by a small margin and could be considered to offer the best currently available dataset. However at this point it is still unclear whether the full physics or proxy method will offer the best results for the future. OCFP was still in an early development phase. If we regard the SRON algorithms only, the results are very similar with a lower scatter and higher data density for the proxy version and better relative accuracy for the full physics version. We deem it therefore prudent to advance at least one proxy and one full physics algorithm. The obvious winners in each category are OCPR and SRFP.

XCO_2 from SCIAMACHY/ENVISAT:

- Algorithms / data products: The algorithms are WFMD and BESD (both IUP, Univ. Bremen).
- VALT recommendation incl. justification:
  - Recommendation: VALT recommends proceeding with BESD. However VALT notices that the BESD data set is very sparse compared to WFMD. To what extent the accuracy and precision of WFMD can be further improved cannot be judged by VALT. This is also true concerning the usefulness of WFMD retrievals within the context of the Ensemble Median Algorithm (EMMA). It may therefore be wise to proceed with both algorithms. Therefore, VALT does not recommend stopping the further development of WFMD.
  - Justification: BESD and WFMD are two sides of a coin (high data density and high scatter versus low data density with low scatter). There is a small advantage of BESD (relative accuracy is lower) but the low data density causes problems at certain stations (only a few points at Lauder).

    BESD has higher accuracy and precision compared to WFMD. If only one algorithm can be used in the future within GHG-CCI, VALT recommends BESD. However, BESD also has its weaknesses, especially the low amount of retrievals compared to WFMD. It therefore cannot be excluded that WFMD is more useful than BESD at least for certain applications. It is also not clear to what extent the accuracy and precision of WFMD can be further improved. WFMD may also be useful in the context of an ensemble of data products, e.g., EMMA.

XCO_2 from TANSO/GOSAT:

- Algorithms / data products: The algorithms used are OCFP (Univ. Leicester) and SRFP (SRON). Both algorithms feature an additional bias correction scheme and are henceforward identified as OCFC and SRFC
- VALT recommendation incl. justification:
    - Recommendation: VALT recommends to proceed with both algorithms, namely with SRFP and OCFP.
    - Justification: For the whole RR validation, this comparison features the least distinctive differences between two algorithms. An example for the overall usefulness of both algorithms is the fact that, using a 100 km collocation domain, SRFC has the highest overall correlation coefficient, while at 500 km OCFC performs better (in both cases by a very small margin).

      OCFC has a slightly better relative accuracy, but still fails to reach the 0.5 ppm threshold. Also its scatter and data density is slightly better, but its seasonal bias range is slightly worse. However when interpreting this data one needs to consider that this (as any) validation method has limits. For instance no a-priori or averaging kernel correction was taken into account. VALT feels that the observed differences lack the significance to indicate a clear winner. OCFC can only be tentatively assigned as the better algorithm. Furthermore none of the algorithms have reached the requirements for inverse modelling and still need to be developed. To what extent the accuracy and precision of both algorithms can be further improved cannot be judged by VALT. This is also true concerning the usefulness of both retrievals within the context of the Ensemble Median Algorithm (EMMA). It is therefore wise to proceed with both algorithms.

The following **Tables S-1** and **S-2** give an overview about the main results:

VALT estimates of achieved data quality: XCO_2 (in ppm)					
Sensor	**Algorithm**	**Precision Single obs.**	**Precision Regional / monthly**	**Relative accuracy**	**Method / comments**
SCIAMACHY	BESD v01.00.01	2.5	2.4	0.7	Details see Section 4.4
SCIAMACHY	WFMD v2.2	4.7	3.2	1.3	Details see Section 4.4
GOSAT	OCFP v3.0	2.4	2.0	0.6	Details see Section 4.4
GOSAT	SRFP v1.1	2.6	1.7	1.0	Details see Section 4.4
Required:		< 8.0	< 1.3	< 0.5	**/URD GHG-CCI v1/**

Table S-1: Estimated quality of the satellite XCO_2 data products obtained from comparisons with ground-based TCCON XCO_2 retrievals. Precision: The precision is based on single observations, and corresponds with the standard deviation on the SAT-FTS difference (using all individual data pairs). Precision regional/monthly: Like Precision Single observation, corresponds with the standard deviation on the SAT-FTS differences. Only this time the individual data pairs are first averaged per month. Note that in this case the ±2 hour temporal overlap criteria still stands! This avoids pairing data with potentially large mismatched timescales (a data cluster at the beginning of the month vs one at the end of the month), but at the cost of data density. Relative accuracy: This corresponds to the standard deviation on the 10 station individual data sat-fts bias results.

VALT estimates of achieved data quality: XCH_4 (in ppb)					
Sensor	**Algorithm**	**Precision Single observation**	**Precision Regional / monthly**	**Relative accuracy**	**Method / comments**
SCIAMACHY	WFMD v2.3	76.9	46.3	11.8	Details see Section 4.4
SCIAMACHY	IMAP v.6.0	50.6	26.5	15.2	Details see Section 4.4
TANSO	OCPR v3.2	14.1	8.2	2.6	Details see Section 4.4
TANSO	OCFP v3.2	18.3	12.3	8.4	Details see Section 4.4
TANSO	SRPR v1.1	14.7	8.1	3.8	Details see Section 4.4
TANSO	SRFP v1.1	15.1	8.6	3.5	Details see Section 4.4
Required:		< 34	< 11	< 10	**/URD GHG-CCI v1/**

Table S-2: Estimated quality of the satellite XCH_4 data products obtained from comparisons with ground-based TCCON XCH_4 retrievals. For details, see caption of Table S-1.

Final conclusions

XCH_4: We feel that meaningful conclusions on the selection can be drawn from the validation results. WFMD and IMAP (due to ENVISAT end time) and OCPR and SRFP are our suggestions to proceed further.

XCO_2: The different algorithms exhibit limited differences at the TCCON sites. The TCCON network is too sparse to determine the optimal algorithm given the large differences on a global scale. Although the number of TCCON sites is increasing, it can be expected that a meaningful coverage will not be achieved. Without knowledge of the "truth" an ensemble method looks to be the best avenue to proceed. All algorithms need further development in order to reach the stringent quality requirements

Given that there is no clear winner regarding the current XCO_2 products, another intermediate solution needs to be found. A prime candidate would be the EnseMble Median Algorithm **(EMMA)**, currently developed at the University of Bremen **/Reuter et al., 2012/**.

The motivation for EMMA is largely based on the observation that, while the differences at TCCON stations are small, global maps differ greatly (see **/AIECARv1/** report) and we have no objective way to tell which one is right. It is important to note that currently none of the CO_2 algorithms (including EMMA) reach the 0.5 ppm relative accuracy mark as noted in the user requirement document.

2 Selection Protocol

The GHG-CCI selection protocol is described in the Round Robin Evaluation Protocol (RREP) **/RREP GHG-CCI v2/** covering data products generated with ECAs and ACAs. The purpose of this document PVASR is to validate the competing ECAs and to give recommendations on which ECAs to be used in future phases of GHG-CCI. This document describes the selection procedure for ECAs but not for ACAs. The following is largely based on the requirements and procedures given in the RREP.

2.1 General approach

The GHG-CCI project aims at generating global satellite derived data sets of atmospheric CO_2 and CH_4 information useful for constraining regional surface fluxes (emission and uptake) via inverse modelling of these two important anthropogenic greenhouse gases (GHG) (see GHG-CCI User Requirements Document **/URD GHG-CCI v1/**).

The four core GHG-CCI ECV data products generated with the GHG-CCI “ECV Core Algorithms” (ECAs) are XCO_2 (in ppm) and XCH_4 (in ppb) from SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT. Within GHG-CCI several ECAs are being further improved and the corresponding data products evaluated. Several algorithms are being further developed within this project – at least two for each of the four core products. It is planned to select the best algorithm for each of the four core products.

Within the GHG-CCI project two types of algorithms to retrieve CO_2 and CH_4 information from satellite data are distinguished:

- “ECV Core Algorithms” (ECAs): These are algorithms for retrieving near-surface sensitive column-averaged mixing ratios of CO_2 and CH_4, denoted XCO_2 and XCH_4, from SCIAMACHY/ENVISAT and TANSO/GOSAT.
- “Additional Constraints Algorithms” (ACAs): These are algorithms for retrieving CO_2 and CH_4 information from satellite data with no or limited near-surface sensitivity. They have the potential to deliver important additional constraints when used in a (inverse modelling) framework that exploits satellite data to infer information on surface fluxes. This is because they can constrain CO_2 and CH_4 in upper layers, i.e., the atmosphere above the Planetary Boundary Layer (PBL).

The selection of ECAs is based on (but not limited to) comparisons with the highly precise and accurate ground-based TCCON FTS XCO_2 and XCH_4 retrievals. For each product and each FTS site several of Figures of Merit (FoM) - bias, standard deviation, etc. - will be computed to characterize the quality of a given data product. In addition, minimum requirements for each FoM have been defined. The FoMs have been computed by the independent GHG-CCI VALidation Team (VALT) and are reported in this document.

This information will ultimately be evaluated by the independent GHG-CCI Climate Research Group (CRG) and by the satellite retrieval team (EOST).

Due to limitations of this approach (e.g., spatial sparseness of the FTS reference data), the GHG-CCI CRG also needs to base the selection on their expert knowledge / judgement. For this purpose, additional information will be made available (in document AIECAR). This comprises global and regional maps and time series of the satellite data products including comparisons with model data and comparisons with the data products generated with the competing ECAs and other satellite derived data products of the same quantity (if available) generated elsewhere (e.g., at NIES and NASA). Last but not least, the CRG will to some extent also use and analyze the GHG-CCI data products. It was expected that for each product a clear “winner” can be identified. Where several ECAs have an identical performance, the CRG will decide, together with the retrieval team, which algorithm will be used to generate the corresponding ECV. In case several ECAs have a different performance and if it cannot be decided unambiguously which ECA is the best an option is to use a similar approach as also used for climate models, namely an “ensemble approach”, i.e., to use several retrievals when confronting the satellite retrievals with models e.g. via direct comparison or inverse modelling. This is expected to provide important additional error information, namely an estimate of algorithm dependent errors (taking this into account the ensemble approach is likely superior compared to using a single data product for reliably estimating surface fluxes and their uncertainties).

The final GHG-CCI RR decision will be based on this document, i.e., PVASR, written by VALT, and an independent document (AIECARv1) written by EOST presenting an analysis of the global data products. The final decision will be reported and documented in the Algorithm Selection Report (ASR), approved by the GHG-CCI CRG, focusing on the final decision and its justification referring to PVASR and AIRCARv1 for details.

Within GHG-CCI several ACAs will also be further improved and their quality will be assessed. ACAs are not in competition within GHG-CCI. They deliver sub-columns and vertical profiles of CO_2 and CH_4. Instruments used are AIRS, IASI, MIPAS, SCIAMACHY in solar occultation mode, and ACE-FTS. The ACAs are not in competition in contrast to ECAs and can typically not be (directly) compared with TCCON XCO_2 and XCH_4. ACAs are therefore not discussed in this document.

2.2 Spatio-temporal requirements for TCCON comparisons

At least one year of data need to be used to obtain robust conclusions, e.g., 2010. Especially for SCIAMACHY, but also for GOSAT, it is important to also analyze longer time periods. For example, 2010 is not a representative year for the quality of the SCIAMACHY methane retrievals due to severe detector degradation after October 2005. It is therefore important to also include the years before 2010, i.e., ideally 2003-2010 for SCIAMACHY and 2009 (partially) to (at least) 2010 for GOSAT.

Within the comparison period FTS, GOSAT, and SCIAMACHY data are available. The overlap between SCIAMACHY with TCCON and GOSAT with TCCON is a requirement for a suitable validation. In addition, a time overlap between SCIAMACHY and GOSAT is important to assess the consistency of the data products of the two sensors.

The period should cover at least one year in order to draw conclusions on, e.g., the seasonal cycle representation (in case a seasonal cycle is expected). In most cases the Round Robin Exercise has been performed for more than one year. Only the RR validation for XCO_2 using the SRFP algorithm for TANSO was limited to one year.

SCIAMACHY was operational between 2002 and 2012. TANSO on GOSAT was launched in January 2009 and calibrated Level 1 data are available since mid 2009. Many FTS sites are operational longer than GOSAT data are available; however, there are some sites which got operational not before August 2009.

Note that due to detector degradation issues, the quality of the SCIAMACHY methane retrievals is highest before November 2005. Comparison of data products with TCCON for 2003-2005 is limited to a few stations only and only after mid 2004. Therefore the list of FTS sites to be used for comparison has to be extended, using also NDACC FTS sites/retrievals. Currently the comparison between the near infrared observations (TCCON) and the mid infrared ones (NDACC) is performed to study the inclusion of the NDACC CO_2 and CH_4 data products. However, the RR validation discussed here is based only on the TCCON data.

For the comparison, all satellite retrievals within a given distance (radius or latitude/longitude range) around the FTS sites shall be used. A smaller distance (e.g., 100 km) is preferred for better representativeness but may result in too few data points for comparison. Results for the 100 and 350 km collocation criteria often yielded unstable results for many of the algorithms. Therefore, while the 100 and 350 km analysis has been performed, the data presented in this document focuses on the 500km collocation criteria. The temporal co-registration criterion is 2 hours, i.e., only FTS data shall be used for comparison, which have been obtained within +/- 2 hours compared to the time of the satellite retrievals.

2.3 Figures of Merit for TCCON comparisons

Each individual satellite retrieval fulfilling the spatial and temporal collocation criteria for the TCCON comparisons shall be used for the comparison as long as it is flagged “good” (as indicated by the quality flag of the individual data product or by using pre-filtered data products delivered by the retrieval team).

The RR exercise aims at identifying the best SCIAMACHY XCO_2, SCIAMACHY XCH_4, GOSAT XCO_2, and GOSAT XCH_4 algorithm.

Several different important selection criteria (“Figures of Merit”) have been defined to characterize the performance of a given algorithm, as will be discussed below.

One can expect that different algorithms will have different strengths and weaknesses. Therefore, ideally a procedure ("benchmark formula") would be good which combines all criteria into one single score value. This aspect has been extensively discussed within GHG-CCI. There is however consensus that at present no reliable procedure exists which guarantees the selection of the best algorithm.

In addition, there are issues related to the FTS data to be used for comparisons, in particular the potentially severe limitations due to the spatial sparseness of the TCCON network.

Therefore it has been decided not to define a benchmark formula now. Instead a number of Figures of Merit (FoM) shall be computed. They are listed in Table 4.

For each FoM a threshold value is given indicating the minimum required performance (by defining a potential "rejection interval"). An algorithm and its corresponding data product is considered to be of very low quality (e.g., large bias relative to FTS) and/or hard or not to be evaluated (e.g., too few retrievals) if one or more of the FoM do not meet the corresponding minimum requirement.

In addition, auxiliary criteria ("Aux") have been defined to provide potentially important additional information.

For each (algorithm and its corresponding) data product and for each FTS site (and for each of the 3 spatial co-location radii) and each year, a table has to be generated to provide the information described in Table 4. Based on these FoM summary tables have been prepared to enable a quantitative analysis of the results.

3 Retrieval Algorithms

The ECV Core Algorithms (ECAs) used to generate the satellite-derived XCO_2 and XCH_4 data products validated in this document are described in the Algorithm Theoretical Baseline Document (ATBD) **/ATBD GHG-CCI v1/**. Table 1 presents an overview of the validated satellite data products.

Algorithms used in the RR exercise					
No.	Molecule	Algorithm	Institute	Satellite	Time covered
RR1	XCH_4	IMAP	SRON	SCIAMACHY	01/2003 -12/2010
RR1	XCH_4	WFMD	IUP	SCIAMACHY	01/2003 -12/2009
RR2	XCH_4	OCFP	UoL	TANSO	04/2009 -04/2011
RR2	XCH_4	OCPR	UoL	TANSO	04/2009 - 04/2011
RR2	XCH_4	SRFP	SRON	TANSO	04/2009 - 04/2011
RR2	XCH_4	SRPR	SRON	TANSO	04/2009 – 05/2011
RR3	XCO_2	WFMD	IUP	SCIAMACHY	01/2003 – 12/2009
RR3	XCO_2	BESD	IUP	SCIAMACHY	01/2006 – 12/2011
RR4	XCO_2	OCFC*	UoL	TANSO	04/2009 – 05/2011
RR4	XCO_2	SRFC*	SRON	TANSO	04/2009 - 04/2011

Table 1: Algorithms used for the RR exercise together with the time period covered.
* The two algorithms OCFP and SRFP feature an additional bias correction scheme and are henceforward identified as OCFC and SRFC.

4 Validation

4.1 Introduction

This document reports the validation of the GHG-CCI satellite-derived data products. The validation has been carried out by the independent validation team (VALT). VALT is also responsible to compute various "Figures of Merit" (FoM), as defined in RREP, which are key quantities defined to characterize the quality of the various data products as needed for the final selection. The FoMs are reported in Section 4.3.

4.2 Ground-based validation sites

The validation has been performed using the TCCON sites, as listed below.

TCCON ground base sites used for validation					
Site	Latitude (°)	Longitude (°)	Altitude (m)	Coverage	N
Bialystock	53.231	23.025	183	03/2009-03/2011	26576
Bremen	53.104	8.850	7	01/2009-12/2010	3338
Karlsruhe	49.102	8.440	110	04/2010-05/2011	4944
Orleans	47.965	2.113	132	08/2009-11/2010	14014
Garmisch	47.476	11.063	744	05/2009-12/2010	15623
ParkFalls	45.945	-90.273	442	06/2004-04/2011	134176
Lamont	36.604	-97.486	320	07/2008-05/2011	144537
Darwin	-12.425	130.891	30	08/2005-02/2011	149846
Wollongong	-34.406	150.879	30	06/2008-03/2011	29715
Lauder	-45.050	169.680	370	06/2004-06/2011	139003

Table 1: TCCON sites that have been used for the RR validation. N gives the number of observations

The retrieved quantities have not been corrected for the averaging kernels or assumptions on the a-priori profiles. The accuracy of the TCCON observations are listed in Table 2.

Precision and systematic bias of the TCCON sites			
Product	Precision	Systematic error (bias)	Calibration factor
XCO_2	0.25% (1.0 ppm)	0.2% (0.8 ppm)	0.989
XCH_4	0.40% (7 ppb)	0.4% (7 ppb)	0.978

Table 2: Precision and systematic errors (all 2 σ) of a single TCCON measurement. The calibration factors in the last column have been derived by comparison with aircraft profiles. The deviation mainly results from the O_2 line list. The aircraft comparisons demonstrate the TCCON data can be calibrated to the WMO in situ trace gas measurement scales using a single site-independent scale factor for each gas **/Wunch et al., 2010/.**

4.3 Methodology

The general RR validation has been performed for three different spatial collocation criteria, as defined in the Project Validation Plan, version 1 (PVPv1). The temporal collocation was set to ±2 hours, the spatial collocation to ±100, ±350 and ±500 km. A smaller distance (e.g., 100 km) is preferred for better representativeness but may result in too few data points for comparison. Results for the 100 and 350 km collocation criteria often yielded unstable results for many of the algorithms. Therefore, while the 100 and 350 km analysis has been performed, the data presented in this document focuses on the 500km collocation criteria. The temporal co-registration criterion is 2 hours, i.e., only FTS data shall be used for comparison, which have been obtained within +/- 2 hours compared to the time of the satellite retrievals. Prior to the comparison the FTIR data have been averaged for 2 hours according to the satellite observation over flight time. The a-priori and averaging kernel differences have not been taken into account in the validation.

The results for each site, year, and algorithm have been summarized in various ‘Figures of Merit’ (FoM), as defined in RREP. They are the key quantities defined to characterize the quality of the various data products as needed for the final selection. These FoM are listed in table 4. Several additional statistical parameters have been calculated next to the RREP defined FoM.

- The total bias computes the mean difference for the entire (common) timeperiod (Sat-FTS). When we refer to a bias in this document it refers to the total bias. The overall bias or overall total bias refers to the total bias, using all individual data over all stations
- The relative accuracy corresponds with the standard deviation of the total bias results for each individual station
- The seasonal bias range takes the difference between the maximal and minimal seasonal bias values (JFM,AMJ,JAS & OND),

Unless stated otherwise all listed results in this document refer to those obtained by the analysis of individual data, except the listed Correlation Coefficient which is derived from daily averaged data.

All FoM have been calculated for individual, daily and monthly means, as well as for the entire common dataset as per year (not discussed in this document). Common refers to the time period covered by all competing algorithms in a round. The daily and monthly means are derived from the individual 2h collocation data, not the entire dataset.

Figures 1 to 3 show various FoM results, namely Number of datapoints, total bias and scatter, for all individual TCCON sites and all data combined.. Several key FoM results are also listed in tables 5 and 6.

GHG-CCI Round Robin Figures of Merit (FoM) for satellite – TCCON comparisons

FoM ID	FoM	Unit	Threshold (rejection range; baseline; all numbers are to be confirmed)	Explanation
FoM_B	Bias [1)]	XCO_2: ppm, XCH_4: ppb	XCO_2: > 4 XCH_4: > 40	Annual bias = Average difference Sat-FTS for entire year
FoM_B1	Bias JFM	XCO_2: ppm, XCH_4: ppb	"	Seasonal bias
FoM_B2	Bias AMJ	XCO_2: ppm, XCH_4: ppb	"	Seasonal bias
FoM_B3	Bias JAS	XCO_2: ppm, XCH_4: ppb	"	Seasonal bias
FoM_B4	Bias OND	XCO_2: ppm, XCH_4: ppb	"	Seasonal bias
FoM_SD	StdDev	XCO_2: ppm, XCH_4: ppb	XCO2: > 12 XCH4: > 90	Standard deviation Sat-FTS
FoM_R	Linear correlation coefficient	dimensionless	> -0.2 and < 0.2	Pearson's R (Sat., FTS)
FoM_NR	Number of sat. retrievals	dimensionless	< 10	Entire year
Aux_FA	Fraction of a-priori information	dimensionless	> 0.05	Range: 0-1. 1: data product fully determined by a-priori; 0: no influence of a-priori; Mathematical definition: TBD
Aux_PS	Processing speed (to process 1 year of data)	dimensionless (time in units of realtime)	< 1	Time needed to process 1 year of global data on the existing infrastructure; 2 means that 1 year of data requires 2 years of processing time
Aux_ND	Number of days processed	dimensionless	< 10	Number of days used for comparison for entire year

Table 4: Figures of Merit and additional auxiliary information to be used to characterize a given data product at a certain FTS site (for a certain spatial co-location radius) during a certain year.

4.4 Discussion of the validation results

The results for XCO_2 and XCH_4 are separately discussed for the four groups, RR1 to RR4. The main numbers are summarized in Table 5. Details on the the individual sites are summarized in Table 6.

Summary of results for all algorithms from the FoM for the RR validation						
Molecule algorithm	N	Bias	Std	R	Rel Acc	Seasonal Bias range
XCH_4/IMAP	39489	-6.21	50.583	0.69	14.36	15.20
XCH_4/WFMD	37628	-6.33	76.898	0.46	11.10	18.03
XCH_4/OCFP	3176	5.61	18.296	0.80	7.99	6.14
XCH_4/OCPR	7323	3.44	14.080	0.88	2.49	10.82
XCH_4/SRFP	2558	-6.61	15.057	0.84	3.31	3.15
XCH_4/SRPR	4900	0.04	14.684	0.80	3.64	7.24
XCO_2/WFMD	30752	-0.73	4.701	0.64	1.26	1.00
XCO_2/BESD	9471	0.52	2.533	0.76	1.37	0.92
XCO_2/OCFC	2830	-0.40	2.438	0.76	0.58	2.01
XCO_2/SRFC	2558	-0.23	2.565	0.73	0.84	0.93

Table 5: Results for the RR validation for the 500 km collocation criteria. The numbers given refer to the validation considering all TCCON sites. N: number of collocated satellite data points between the satellite and TCCON. Bias: overall mean total bias (Sat-FTS) in ppm for XCO_2 and in ppb for XCH_4. Std: standard deviation (Sat-FTS) in ppm for XCO_2, in ppb for XCH_4. R: Pearson correlation coefficient using daily averaged collocated data. Rel. Acc.: Relative Accuracy, standard deviation (bias results for individual stations). Seasonal Bias range: Derived from the overall seasonal bias results (Bias_JFM: Bias for January, February and March, Bias_AMJ etc.). The range corresponds to the difference between the maximal and minimal seasonal bias value.

Results from the FoM for the RR validation for all TCCON sites

Site	XCH_4						XCO_2			
	IMAP	WFMD	OCFP	OCPR	SRFP	SRPR	WFMD	BESD	OCFC	SRFC
	Bias Std N	Bias Std N	Bias Std N	Bias Std N	Bias Std N	Bias Std N	Bias Std N	Bias Std N	Bias Std N	Bias Std N
Bialystock	5.2 46.7 2170	5.4 86.7 2065	13.8 18.4 198	7.6 14.4 771	-3.9 14.6 174	7.0 15.5 404	1.3 5.2 1712	0.6 2.0 504	0.1 2.8 155	0.4 2.8 174
Bremen	5.1 51.2 664	-8.7 95.4 511	13.3 15.5 95	5.6 14.0 214	-7.6 16.3 73	-1.7 16.3 103	2.2 5.4 390	1.5 2.0 60	0.4 3.0 73	-0.2 2.5 73
Karlsruhe	- - -	- - -	5.6 17.3 247	1.9 14.3 540	-10.9 12.8 143	-1.6 16.7 331	- - -	- - -	-0.2 2.9 180	-0.7 2.6 143
Orleans	4.2 45.8 812	-6.7 74.1 267	10.4 16.0 279	5.8 13.0 564	-5.2 13.4 193	4.7 14.3 339	2.1 4.2 218	1.4 2.5 166	0.0 2.8 243	-0.1 2.4 193
Garmisch	4.1 52.4 930	-6.5 92.4 628	13.7 16.8 167	6.8 14.1 465	-3.7 16.6 145	3.5 18.1 281	0.6 6.4 549	2.1 2.1 143	0.7 2.8 149	0.7 2.8 145
ParkFalls	-2.7 48.7 18717	3.1 75.2 13652	11.7 16.4 268	3.3 14.3 893	-5.2 14.5 271	-0.8 14.9 686	-0.8 5.4 8208	1.2 2.7 736	-0.7 1.9 214	-0.3 2.5 271
Lamont	-0.5 47.2 8273	-1.4 73.8 10545	3.0 17.8 1612	2.5 14.4 2780	-7.3 14.1 1279	-0.7 13.4 2125	-1.0 4.1 11030	-0.3 2.1 2285	-0.9 2.1 1453	-0.7 2.2 1279
Darwin	-26.6 53.4 5625	-27.1 79.5 8116	-12.8 13.3 65	3.3 10.0 305	-4.2 14.3 80	-4.1 8.2 153	-1.2 4.2 7337	0.7 2.7 4920	-0.6 2.1 116	1.2 2.5 80
Wollongong	-31.3 49.9 1252	-27.2 79.5 1483	0.4 21.4 216	-1.0 13.2 605	-6.4 21.9 181	-4.0 14.6 396	0.3 4.7 1089	0.7 2.6 653	0.8 2.7 225	1.4 3.3 181
Lauder	-23.6 61.4 1046	-14.3 94.6 361	1.9 16.6 29	2.7 9.4 186	-14.6 17.1 19	-3.2 10.7 82	1.4 6.4 219	4.8 1.9 4	0.6 4.0 22	-1.4 4.4 19
All sites	-6.21 50.58 39489	-6.32 76.90 37628	5.61 18.30 3176	3.44 14.08 7323	-6.61 15.06 2558	0.04 14.68 4900	-0.73 4.7 30752	0.52 2.53 9471	-0.40 2.44 2830	-0.23 2.57 2558
Rel. Acc	14.36	11.10	7.98	2.49	3.31	3.64	1.26	1.37	0.58	0.84

Table 6: Results for the total bias and the standard deviation for all individual TCCON sites. The values are given for XCO_2 in ppm and for XCH_4 in ppb. In addition the number of collocated data pairs is given. The last two lines give the values as average for all sites and the relative accuracy, as also shown in Table 5.

The validation has been performed by comparing the total columns of XCO_2 and XCH_4 from the TCCON ground-based sites with SCIAMACHY and GOSAT satellite data, using the different algorithms. The analysis has been done by comparing the scatter, bias, correlation, relative accuracy and seasonal bias range of the different algorithms and by considering the amount of data for each algorithm.

Based on the careful validation of the satellite data products as describe in this document, the VALT recommendations concerning which algorithms to be used to generate the CRDP are:

Results for XCH_4 from SCIAMACHY/ENVISAT:

- The algorithms used are WFMD (IUP, Univ. Bremen) and IMAP (SRON/JPL).
- VALT findings:
  - Systematic errors (biases): See table 5,6 and figure 2. Both data products have very similar biases, -6.21 ppb for IMAP, and -6.33 ppb for WFMD, both relative to the TCCON reference data. But the biases can be as large as 30 ppb at certain sites (southern hemisphere sites Darwin, Wollongong and Lauder). The reason for the quite large biases at the southern hemispheric TCCON sites is unclear. The relative accuracy (the standard deviation of the biases for all months) is also quite similar with slightly larger values for IMAP (14.36 ppb for IMAP and 11.10 ppb for WFMD).

    The biases may be related to SCIAMACHY detector degradation after October 2005. However modifying our validation to a post and pre-2005 period does not show a noticeable change. Note that the amount of pre-2005 TCCON data is very limited, so findings from these validation runs are tentative only. Additional studies still need to confirm these results. As both algorithms suffer from similar biases a recommendation which of the two algorithms to use to generate the CRDP cannot be based on systematic errors.
  - Seasonal bias range: See table 5. The seasonal bias range (max. bias minus min bias) is quite high. 15.20 ppb and 18.03 ppb for IMAP and WFMD, respectively. In JFM we get -17.48 ppb for IMAP and -10.48 ppb for WFMD compared to JAS, where we get -2.28 ppb and -3.93 ppb, respectively.
  - Random errors (scatter): See table 5, 6 and figure 3. The WFMD data set exhibits a larger scatter (~80 ppb) compared to IMAP (~50 ppb). The likely reason is that WFMD is based on unconstrained least-squares fitting using a constant CH_4 profile as a priori information whereas IMAP is based on Optimal Estimation and uses latitude and time dependent CH_4 profiles from the a global model.
  - Amount of data: See table 5, 6 and figure 1. The number of data points at the TCCON sites is very similar for both data sets, 39489 for IMAP and 37628 for WFMD.
  - Correlation coefficients: See table 5. the correlation coefficient for IMAP is quite high (0.69), but varies to a large extent from site to site (-0.176 for Wollongong to 0.393 for Orleans). For WFMD the coefficient is 0.46, in part to

the increased scatter, and also shows a large inter-station variability between -0.132 for Bialystock and 0.337 for Garmisch.

- VALT recommendation incl. justification:
    - Recommendation: VALT recommends proceeding with both algorithms, WFMD and IMAP.
    - Justification: The two SCIAMACHY methane data products do not yet fulfil the relative accuracy threshold requirement (< 10 ppb) under all conditions and both algorithms appear to have different strength and weaknesses. Both data sets show similar biases and this criterion cannot be used to recommend only one of the two algorithms. At certain TCCON sites the biases are significantly larger (~30 ppb) than the required relative accuracy. These biases are not yet well understood and more studies are needed to identify the reasons for that. Both algorithms need further development (including how to better/optimally use the Level 1 data and/or to select the successful retrievals, i.e., including pre- and post-processing procedures and filtering) to ensure that the relative accuracy requirement is met under all conditions. If this can be achieved with only one or even with both algorithms is not clear and therefore it is recommended to proceed with both algorithms. The IMAP data product shows less scatter and therefore also higher correlation with the TCCON reference data. This is a pro for IMAP and a con for WFMD. However, IMAP uses global CH_4 model data as a priori information whereas WFMD is independent of any global CH_4 model. Certain users may prefer a data product independent of any global CH_4 model. This is a pro for WFMD and a con for IMAP.

XCH$_4$ from TANSO/GOSAT:

- Algorithms / data products: Both algorithms come in a full physics and proxy version. The algorithms are OCPR & OCFP (Univ. Leicester) and SRPR & SRFP (SRON).
- VALT findings:
    - Systematic errors (biases): See table 5, 6 and figure 2. The overall total biases range between 0.04 ppb (SRPR) and 5.6 ppb (OCFP). The relative accuracy is 7.99 ppb for OCFP, 2.49 ppb for OCPR, 3.31 ppb for SRFP and 3.64 ppb for SRPR.

      The bias remains pretty consistent over all stations, thus unlike the SCIAMACHY data, no southern hemisphere jump is observed. More telling then the absolute bias is the variability between the bias results of individual stations. The relative accuracy, the standard deviation of these biases, acting as an indicator of the relative accuracy, ranges between 2.49 ppb (OCPR) and 7.99 ppb (OCFP). SRPR and SRFP score 3.31 ppb and 3.64 ppb, respectively. Thus, OCFP thus seems to stand out in a negative way.
    - Seasonal bias range: See table 5.The seasonal bias is relatively low, 6.14 ppb for OCFP, 10.82 ppb for OCPR, 3.15 for SRFP and 7.24 for SRFP. The seasonal bias change between JFM and JAS vary between -0.42 ppb for JAS for SRPR to 1.70 ppb for JFM for OCFP.

   - Random errors (scatter): See table 5, 6 and figure 3. Also on the scatter, OCFP (18.3 ppb) performs slightly worse than the other contenders (14.1 to 15.1 ppb). The lowest scatter, by a small margin, is obtained by the OCPR algorithm.
   - Amount of data: See table 5, 6 and figure 1. Both proxy versions have a higher data density than their full physics counterparts. Again OCPR performs best, with about 50% more data than SRPR. Of the full physics versions, OCFP performs slightly better than SRFP.
   - Correlation coefficient: See table 5.The correlation coefficients are around 0.8 for all four algorithms. The values vary from 0.076 for Darwin and OCFP to 0.803 for Lauder and SRFP and SRPR. It must be mentioned that the low correlation coefficients at Darwin are partly due to the low seasonal variability at that site.

- VALT recommendation incl. justification:
   - Recommendation: VALT recommends to proceed with two of the four algorithms, namely with SRFP and OCPR.
   - Justification: From the four algorithms used in this round, only OCFP seems to perform significantly worse than its competitors on several key statistics. On the other hand its proxy method counterpart edges out its SRON competitors by a small margin and could be considered to offer the best currently available dataset. However at this point it is still unclear whether the full physics or proxy method will offer the best results for the future. OCFP was still in an early development phase. If we regard the SRON algorithms only, the results are very similar with a lower scatter and higher data density for the proxy version and better relative accuracy for the full physics version. We deem it therefore prudent to advance at least one proxy and one full physics algorithm. The obvious winners in each category are OCPR and SRFP.

XCO_2 from SCIAMACHY/ENVISAT:

- Algorithms / data products: The algorithms are WFMD and BESD (both IUP, Univ. Bremen).
- VALT findings:
   - Systematic errors (biases): See table 5, 6 and figure 2. The BESD algorithm gives a smaller biases of 0.52 ppm compared to WFMD, which results in -0.73 ppm, relative to TCCON. But both results do not achieve the required relative accuracy of < 0.5 ppm[1]. respectively. The relative accuracy on BESD

[1] It has been identified (not shown here) that the BESD Round Robin data set suffers from biases caused by problems related to the Level 1 version 7 consolidation level U (L1v7u) data product used for retrieval and that biases will be much smaller when using the new L1v7w data set. It is expected that the BESD biases will be significantly smaller after re-processing using the improved SCIAMACHY

is better when we exclude Lauder (very limited amount of data). No deviations are seen in the seasonality plots. Accordingly, the relative accuracy is similar with 1.26 ppm for WFMD and 1.37 for BESD.

- Seasonal bias range: See table 5.The seasonal bias range is quite similar for both algorithms, 1.0 ppm for WMFD and 0.92 for BESD. The variabilities between JFM and JAS between both algorithms are -0.75 ppm and 0.05 ppm for WFMD and BESD in JFM and in JAS we get -0.99 and 0.80 ppm..
- Random errors (scatter): See table 5, 6 and figure 3.The BESD algorithm shows less scatter (~ 2.5 ppm) compared to WFMD (~ 4.7 ppm), i.e., BESD has a higher precision compared to WFMD.
- Amount of data: See table 5, 6 and figure 1. Overall WFMD has 3 times more data compared to BESD (30752 to 9471).
- Correlation coefficient: See table 5. The correlation coefficients are 0.64 and 0.76 for WFMD and BESD, respectively. They vary from 0.191 for Orleans and BESD to 0.919 for Bremen and BESD. Both sites featured only 9 daily averaged datapoints.

- VALT recommendation incl. justification:
  - Recommendation: VALT recommends proceeding with BESD. However VALT notices that the BESD data set is very sparse compared to WFMD. To what extent the accuracy and precision of WFMD can be further improved cannot be judged by VALT. This is also true concerning the usefulness of WFMD retrievals within the context of the Ensemble Median Algorithm (EMMA). It may therefore be wise to proceed with both algorithms. Therefore, VALT does not recommend stopping the further development of WFMD.
  - Justification: BESD and WFMD are two sides of a coin (high data density and high scatter versus low data density with low scatter). There is a small advantage of BESD (relative accuracy is lower) but the low data density causes problems at certain stations (only a few points at Lauder).

    BESD has higher accuracy and precision compared to WFMD. If only one algorithm can be used in the future within GHG-CCI, VALT recommends BESD. However, BESD also has its weaknesses, especially the low amount of retrievals compared to WFMD. It therefore cannot be excluded that WFMD is more useful than BESD at least for certain applications. It is also not clear to what extent the accuracy and precision of WFMD can be further improved. WFMD may also be useful in the context of an ensemble of data products, e.g., EMMA.

Level 1 data, i.e., L1v7w. Future validation of the reprocessed SCIAMACHY data will show what the ultimate accuracy is that can be achieved with BESD.

<u>XCO_2 from TANSO/GOSAT:</u>

- Algorithms / data products: The algorithms used are OCFP (Univ. Leicester) and SRFP (SRON). Both algorithms feature an additional bias correction scheme and are henceforward identified as OCFC and SRFC

- VALT findings:
    - Systematic errors (biases): See table 5, 6 and figure 2. The overall total bias for OCFC is –0.4 ppm, while SRFC reaches -0.2 ppm. The relative accuracy is slightly larger for SRFC (0.58 ppm for OCFC and 0.84 for SRFC). The variability of the bias between stations is slightly higher for SRFC, but it becomes smaller for SRFC when excluding Lauder due to the limited amount of data.
    - Seasonal bias range: See table 5. The seasonal bias range is 2.01 for OCFC and 0.93 for SRFC. The seasonality varies from -0.99 ppm (OCFC) and -0.73 ppm (SRFC) for JFM to -0.07 ppm and 0.12 ppm for JAS.
    - Random errors (scatter): See table 5, 6 and figure 3. Overall scatter is very similar: 2.4 vs 2.6 ppm, for OCFC and SRFC respectively. Also the scatter for all sites is similar, ranging from 2.1 ppm at Lamont for OCFC to 4.4 ppm at Lauder for SRFC.
    - Amount of data: See table 5, 6 and figure 1. Again no strong difference between the algorithms is observed. OCFC only has a slightly higher data density (2830 versus 2558).
    - Correlation coefficient: See table 5. The correlation coefficients are quite similar, 0.76 and 0.73. They vary from 0.072 for Darwin and OCFC to 0.915 for Park Falls. It must be mentioned that the low correlation coefficients at Darwin are partly due to the low seasonal variability at that site.

- VALT recommendation incl. justification:
    - Recommendation: VALT recommends to proceed with both algorithms, namely with SRFP and OCFP.
    - Justification: For the whole RR validation, this comparison features the least distinctive differences between two algorithms. An example for the overall usefulness of both algorithms is the fact that, using a 100 km collocation domain, SRFC has the highest overall correlation coefficient, while at 500 km OCFC performs better (in both cases by a very small margin).

      OCFC has a slightly better relative accuracy, but still fails to reach the 0.5 threshold. Also its scatter and data density is slightly better, on the other hand its seasonal bias range is worse. However when interpreting this data one needs to consider that this (as any) validation method has limits. For instance no a-priori or averaging kernel correction was taken into account. VALT feels that the observed differences lack the significance to indicate a clear winner.

OCFC can only be tentatively assigned as the better algorithm. Furthermore none of the algorithms have reached the requirements for inverse modelling and still need to be developed. To what extent the accuracy and precision of both algorithms can be further improved cannot be judged by VALT. This is also true concerning the usefulness of both retrievals within the context of the Ensemble Median Algorithm (EMMA). It is therefore be wise to proceed with both algorithms.

Tables 7 and 8 give an overview about the main results: The Figures 1 to 4 give examples for the number of data points, the bias, the scatter and the time series for a few algorithms and sites. Due to the large amount of data a representative selection has been chosen for Figure 4, showing the seasonal time-series.

VALT estimates of achieved data quality: XCO_2 (in ppm)					
Sensor	**Algorithm**	**Precision** **Single obs.**	**Precision** **Regional / monthly**	**Relative accuracy**	**Method / comments**
SCIAMACHY	BESD v01.00.01	2.5	2.4	0.7	Details see Section 4.4
SCIAMACHY	WFMD v2.2	4.7	3.2	1.3	Details see Section 4.4
GOSAT	OCFP v3.0	2.4	2.0	0.6	Details see Section 4.4
GOSAT	SRFP v1.1	2.6	1.7	1.0	Details see Section 4.4
Required:		< 8.0	< 1.3	< 0.5	**/URD GHG-CCI v1/**

Table 7: Estimated quality of the satellite XCO_2 data products obtained from comparisons with ground-based TCCON XCO_2 retrievals. Precision: The precision is based on single observations, and corresponds with the standard deviation on the SAT-FTS difference (using all individual data pairs). Precision regional/monthly: Like Precision Single observation, corresponds with the standard deviation on the SAT-FTS differences. Only this time the individual data pairs are first averaged per month. Note that in this case the ±2 hour temporal overlap criteria still stands! This avoids pairing data with potentially large mismatched timescales (a data cluster at the beginning of the month vs. one at the end of the month), but at the cost of data density. Relative accuracy: This corresponds to the standard

VALT estimates of achieved data quality: XCH_4 (in ppb)					
Sensor	**Algorithm**	**Precision** **Single observation**	**Precision** **Regional / monthly**	**Relative accuracy**	**Method / comments**
SCIAMACHY	WFMD v2.3	76.9	46.3	11.8	Details see Section 4.4
SCIAMACHY	IMAP v.6.0	50.6	26.5	15.2	Details see Section 4.4
TANSO	OCPR v3.2	14.1	8.2	2.6	Details see Section 4.4
TANSO	OCFP v3.2	18.3	12.3	8.4	Details see Section 4.4
TANSO	SRPR v1.1	14.7	8.1	3.8	Details see Section 4.4
TANSO	SRFP v1.1	15.1	8.6	3.5	Details see Section 4.4
Required:		< 34	< 11	< 10	**/URD GHG-CCI v1/**

Table 8: Estimated quality of the satellite XCH_4 data products obtained from comparisons with ground-based TCCON XCH_4 retrievals. For details, see caption of Table 7.

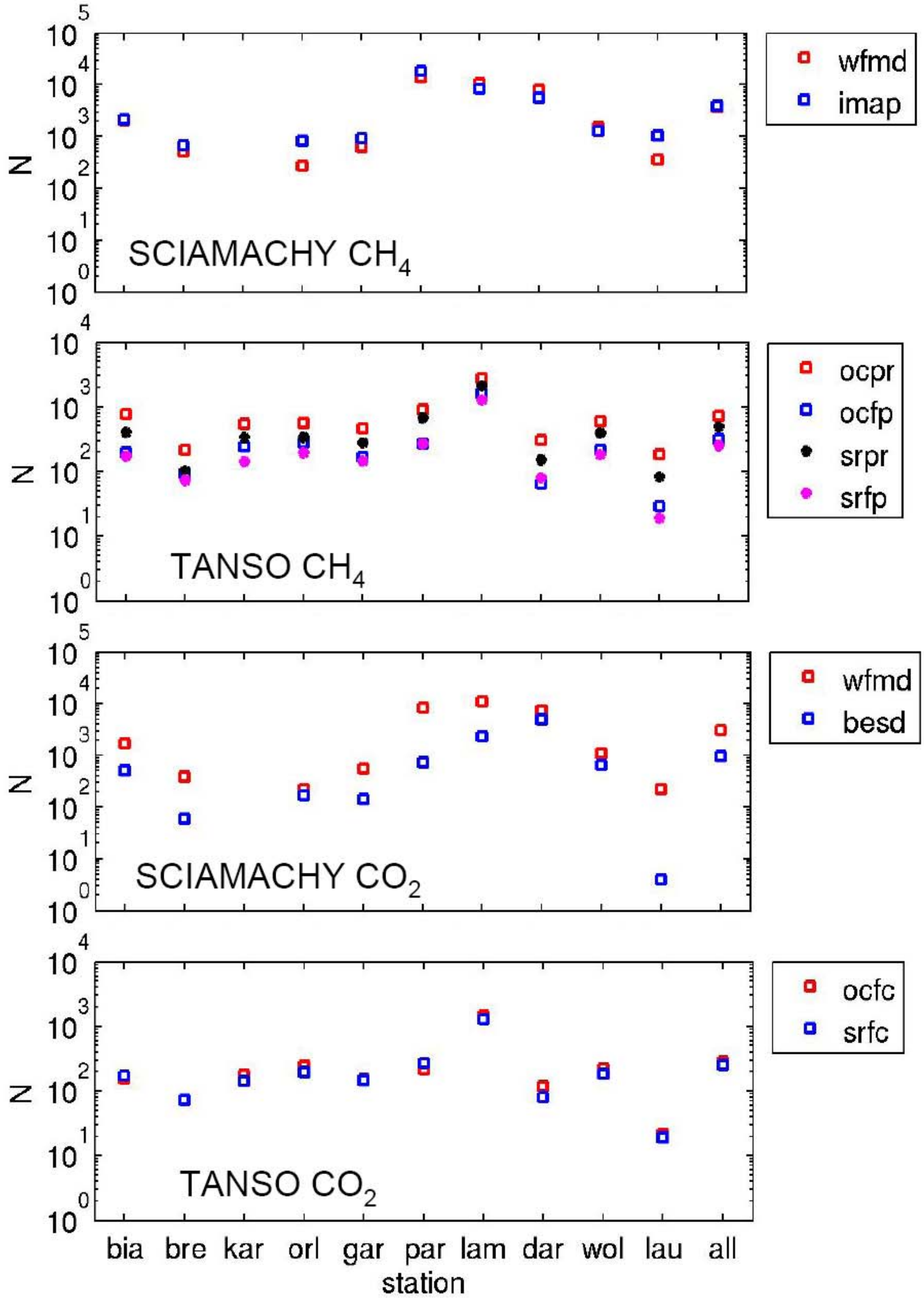

Figure 1: Number of data points for all sites and algorithms together with the overall value (divided by 10).

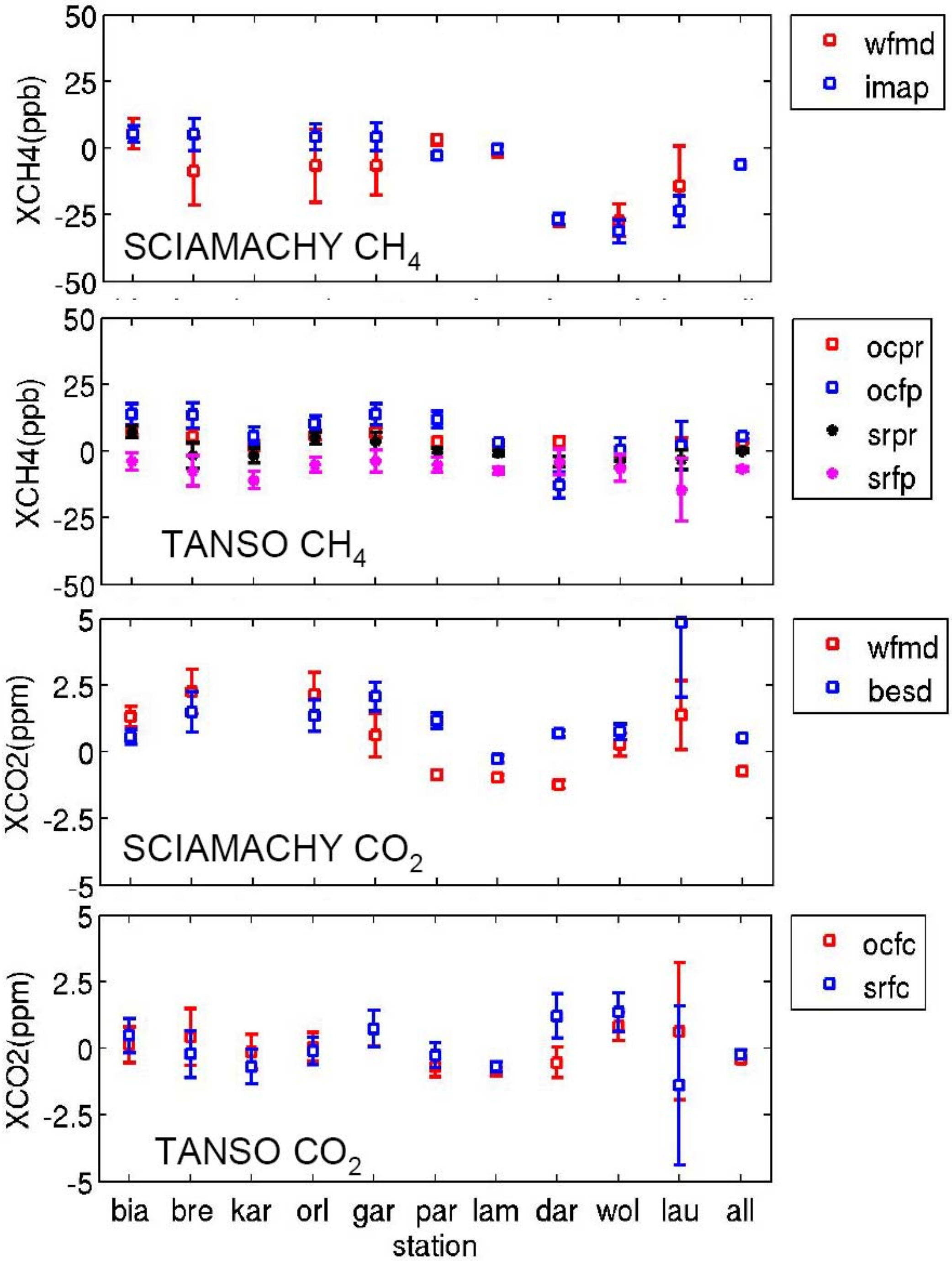

Figure 2: Bias for all sites and algorithms together with the overall bias. For the WFMD/IMAP comparisons was no data at Karlsruhe.

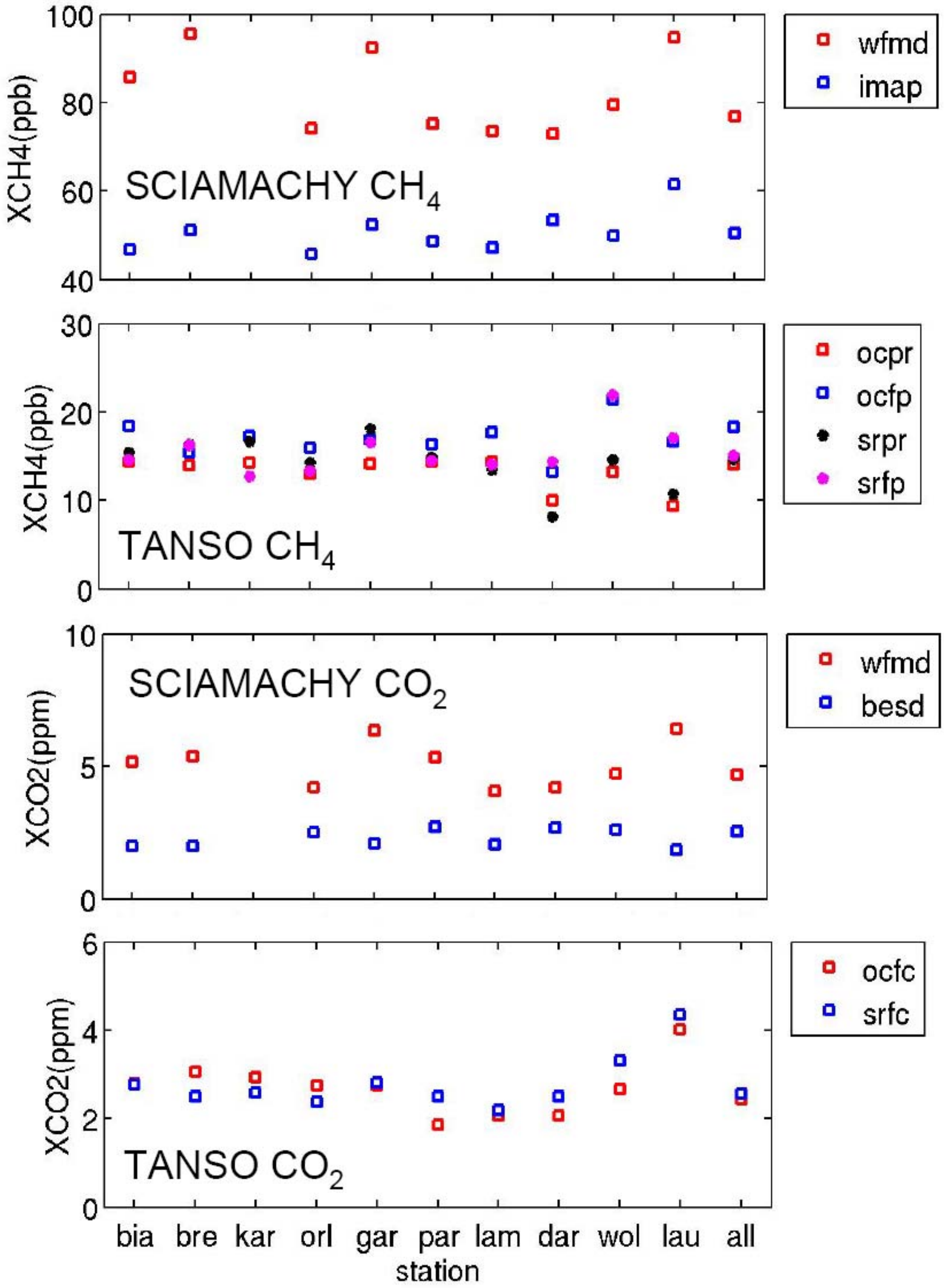

Figure 3: Scatter for all sites and algorithms together with the overall scatter.

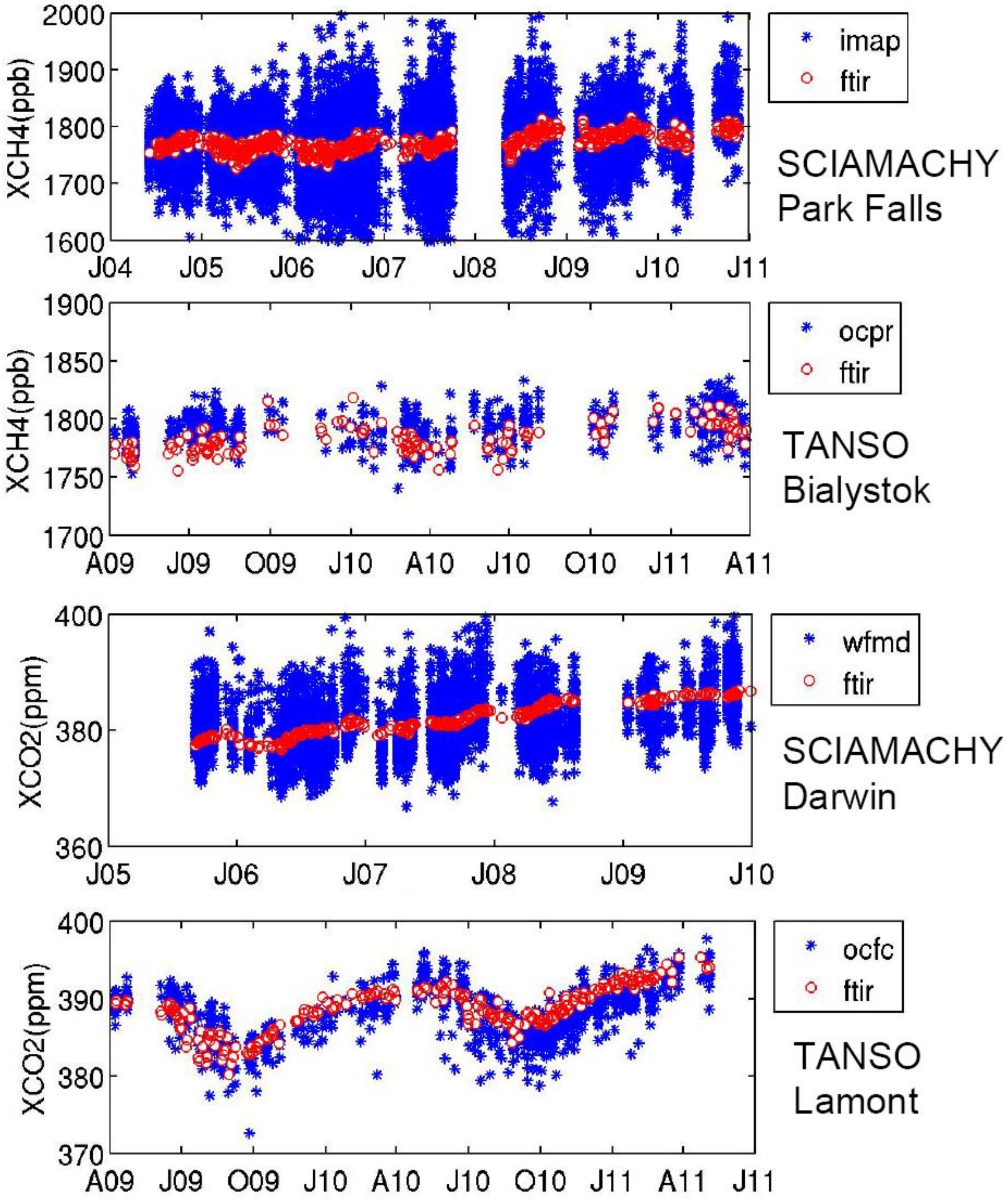

Figure 4: Examples for the time-series at a few sites for the FTS TCCON data together with selected algorithm results.

5 VALT recommendation for algorithm selection

XCH_4: We feel that meaningful conclusions on the selection can be drawn from the validation results. WFMD and IMAP (due to ENVISAT end time) and OCPR and SRFP are our recommendations to proceed further.

XCO_2: The different algorithms exhibit limited differences at the TCCON sites. The TCCON network is too sparse to determine the optimal algorithm given the large differences on a global scale. Although the number of TCCON sites is increasing, it can be expected that a meaningful coverage will not be achieved. Without knowledge of the "truth" an ensemble method (see below) looks to be the best avenue to proceed. All algorithms need further development in order to reach the stringent quality requirements.

The ensemble algorithm EMMA for XCO_2

Given that there is no clear winner regarding the current XCO_2 products, another intermediate solution needs to be found. A prime candidate would be the EnseMble Median Algorithm (EMMA), currently developed at the University of Bremen.

The motivation for an ensemble product (EMMA) is based on the observation that, while the differences at TCCON stations are relatively small, global maps differ greatly and we currently have no objective way of telling which algorithm corresponds with the truth. The largest differences in global maps typically appear where no TCCON observations are available (South America, Africa, China-South East Asia). While the TCCON network acknowledges that certain key areas are still left uncovered, TCCON measurements require a significant amount of manpower and infrastructure and are therefore not easily deployed in remote regions. It is also important to note that none of the CO_2 algorithms (including EMMA) reach the 0.5 ppm relative accuracy mark as noted in the user requirement document.

6 References

/AIECARv1/ Algorithm Inter-comparison and Error Characterization & Analysis Report Version 1 (AIECARv1) of the GHG-CCI project, Aug. 2012, available from GHG-CCI website: http://www.esa-ghg-cci.org (end of August 2012), 2012.

/ASR/ Algorithm Selection Report (ASR), ESA Climate Change Initiative (CCI) GHG-CCI project, August 2012, available from http://www.esa-ghg-cci.org (end of August 2012), 2012.

/ATBD GHG-CCI v1/ Algorithm Theoretical Baseline Document (ATBD), ESA Climate Change Initiative (CCI) GHG-CCI project, Version 1, 15 March 2012, available from http://www.esa-ghg-cci.org, 2012.

/GCOS-107/ Global Climate Observing System (GCOS), SYSTEMATIC OBSERVATION REQUIREMENTS FOR SATELLITE-BASED PRODUCTS FOR CLIMATE, Supplemental details to the satellite-based component of the “Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC”, Prepared by World Meteorological Organization (WMO), Intergovernmental Oceanographic Commission, United Nations Environment Programme (UNEP), International Council for Science, Doc.: GCOS 107 (WMO/TD No. 1338), Sept 2006, 2006.

/Reuter et al., 2012/ Reuter, M. H. Bösch, H. Bovensmann, A. Bril, M. Buchwitz, A. Butz, J. P. Burrows, C. W. O’Dell, S. Guerlet, O. Hasekamp, J. Heymann, N. Kikuchi, S. Oshchepkov, R. Parker, S. Pfeifer, O. Schneising, T. Yokota, and Y. Yoshida, A joint effort to deliver satellite retrieved atmospheric CO_2 concentrations for surface flux inversions: The ensemble median algorithm EMMA, Atmos. Chem. And Phys. Disscussions, submitted, 2012.

/RREP GHG-CCI v2/ Round Robin Evaluation Protocol (RREP), ESA Climate Change Initiative (CCI) GHG-CCI project, Version 2, 17 Aug 2011, available from http://www.esa-ghg-cci.org/sites/default/files/documents/public/documents/RREPv2_GHG-CCI_final.pdf, 2011.

/URD GHG-CCI v1/ User Requirements Document (URD), ESA Climate Change Initiative (CCI) GHG-CCI project, Version1, 3 Feb 2011, available from: , http://www.esa-ghg-cci.org/sites/default/files/documents/public/documents/URDv1_GHG-CCI_final.pdf, 2011.

/Wunch et al., 2010/ Wunch, D, G. C. Toon, J.-F. L. Blavier, R. A. Washenfelder, J. Notholt, B. J. Connor, D. W. T. Griffith, V. Sherlock and P. O. Wennberg. The Total Carbon Column Observing Network. Phil. Trans. R. Soc. A, 2011, link: https://tccon-wiki.caltech.edu/@api/deki/files/885/=ProofTCCON.pdf

7 Acronyms

Abbreviation	Meaning
ACE-FTS	Atmospheric Chemistry Experiment-Fourier Transform Spectrometer
ACA	Additional Constraints Algorithm
ATBD	Algorithm Theoretical Basis Document
BESD	Bremen optimal EStimation DOAS
CCI	Climate Change Initiative
CDR	Climate Data Record
CMUG	Climate Modelling User Group (of ESA's CCI)
CRDP	Climate Research Data Package
CRG	Climate Research Group
DOAS	Differential Optical Absorption Spectroscopy
ECA	ECV Core Algorithm
ECMWF	European Centre for Medium Range Weather Forecasting
ECV	Essential Climate Variable
EO	Earth Observation
ESA	European Space Agency
FCDR	Fundamental Climate Data Record
FP	Full Physics
FTIR	Fourier Transform InfraRed
FTS	Fourier Transform Spectrometer
GCOS	Global Climate Observing System
GHG	GreenHouse Gas
GMES	Global Monitoring for Environment and Security
GOSAT	Greenhouse Gases Observing Satellite
IASI	Infrared Atmospheric Sounding Interferometer
IMAP-DOAS	Iterative Maximum A posteriori DOAS
IMLM	Iterative Maximum Likelihood Method
IPCC	International Panel in Climate Change

IUP	Institute of Environmental Physics (IUP) of the University of Bremen, Germany
LMD	Laboratoire de Météorologie Dynamique
LUT	Look-up table
MACC	Monitoring Atmospheric Composition and Climate, EU GMES project
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MODIS	Moderate Resolution Imaging Spectrometer
NA	Not applicable
NDACC	Network for the Detection of Atmospheric Composition Change
NASA	National Aeronautics and Space Administration
NIES	National Institute for Environmental Studies
NOAA	National Oceanic and Atmospheric Administration
OCO	Orbiting Carbon Observatory
OE	Optimal Estimation
PBL	Planetary Boundary Layer
PVP	Product Validation Plan
PVR	Product Validation Report
RMS	Root-Mean-Square
RTM	Radiative transfer model
SCIATRAN	RTM for SCIAMACHY
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric ChartographY
TANSO	Thermal And Near infrared Sensor for carbon Observation
TBC	To be confirmed
TBD	To be defined / to be determined
TCCON	Total Carbon Column Observing Network
WFM-DOAS (or WFMD)	Weighting Function Modified DOAS

End of document