ESA Climate Change Initiative (CCI)

Product Validation Plan (PVP)

Version 1 (PVPv1)

for the Essential Climate Variable (ECV)

Greenhouse Gases (GHG)

Written by:

GHG-CCI validation team:

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
Version 1 – Draft 1	25. Mar. 2011	Initial Draft for validation team Author: J. Notholt	New document. Main purpose: Input for initial discussion.
Version 1 – Draft 2	9. April 2011	Author: J. Notholt and validation team	Comments from GHG-CCI Progress Meeting 2 (PM2) at EGU considered
Version 1 – Draft 3	13. April 2011	Author: J. Notholt and validation team	Modified after comments from validation team
Version 1 – Draft 4	11. May 2011	Author: J. Notholt and validation team	Modified after comments from GHG-CCI team
Version 1 – Draft 5	12. May 2011	Author: J. Notholt and validation team	Version with comments from M. Buchwitz
Version 1 – Draft 6	13. May 2011	Author: J. Notholt and validation team	Last check of final version
Version1 – final	20.May 2011	Author: J. Notholt and validation team	Final version

Table of contents

1 Summary

This document is the Product Validation Plan (PVP), which is a deliverable of the ESA project GHG-CCI. The validation group is lead by J. Notholt of the IUP. The contributing 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).

In this document the validation plan for the ECV Greenhouse Gases is formulated. The validation will be performed for satellite observations of XCO_2 and XCH_4 from SCIAMACHY on ENVISAT and TANSO onboard GOSAT. XCO_2 and XCH_4 are the core GHG ECV data products generated within GHG-CCI. These data products will be validated by the most suitable independent observations. The most suitable observations are measurements by the ground-based solar absorption spectrometry in the near-infrared spectral region,

(i) because they have similar observations geometry as the satellites,

(ii) yield total columns data, and

(iii) use similar spectral regions for the analysis.

These near infrared observations are performed and organized within the TCCON. The time period covered can be enlarged by including additional FTIR mid-infrared observations measured by the NDACC. In addition to the total column data of TCCON and NDACC, in-situ data from the GAW network will be used for the comparison with the satellite observations to construct total column data.

In a first step an initial validation as part of the Round Robin (RR) exercise will be performed during the first year using available satellite retrievals. This will be done using mainly the TCCON data. In parallel, NDACC retrievals will be evaluated and in-situ data will be used to build total columns, to be used in this first step for the comparison. Prior to their inclusion in the validation dataset, both the NDACC data of CO_2 and CH_4 data as well as the in-situ constructed columns need and will be compared to the TCCON dataset at those sites where both observations are available. While the TCCON data are calibrated against the WMO standard, the NDACC data are not and thus need the connection to the TCCON data. The total columns based on in-situ data, will be compared to the TCCON data to ensure their suitability.

In a second step a comprehensive validation will be performed at the end of the second year following the Round Robin Evaluation Protocol (RREP) to guide the decision which algorithms will be used to generate the first version of the ECP product.

In the final step the ECV data products will be validated in the third year for the whole time period (SCIAMACHY/ENVISAT: 2003-2010, TANSO/GOSAT: 2009-2010) using TCCON, NDACC and the total column data based on the in-situ results.

2 Requirements for the validation

The independence of the validation process should follow three requirements, as given in the ***/CCI project guidelines/***. The main parts of the document together with comments from the validation group are given here below:

> *<u>Requirement 1:</u> CCI project teams shall use, for validation, in situ or other suitable reference datasets that have not been used during the production of their CCI products.*
>
> The validation will be performed using the TCCON and NDACC total column observations. In addition in-situ observations combined with appropriate models to get total column data will be used.
>
> *<u>Requirement 2:</u> CCI project teams shall consider the independence of the geophysical process and ensure that if a particular auxiliary dataset is used in the production of their CCI products then the same dataset is not used in the validation and, if required, alternative auxiliary data are used.*
>
> The independency of the validation results will be considered throughout the validation as far as possible. The use of the data mentioned above ensures the maximum independency.
>
> *<u>Requirement 3:</u> CCI project teams shall ensure that the validation is carried out (or at least verified) by staff not involved in the final algorithm selection; ideally the validation of the CCI products should be carried out by external parties, i.e. by staff / institutions not involved in the production of the ECV products.*
>
> The validation team will ensure that its members are not involved in the production of the ECV products. The validation team is not involved in the satellite retrieval.
>
> It is clear that unique independent reference data may not always be available. Data for validation should be selected to ensure complete coverage of the various spatial and temporal scales in each CCI product. Therefore, the selection of validation data sets should follow different levels of rigour depending on the level of independence of each data set, thus making sure that some level of confidence can be given to every output product. Each CCI product should contain an indication of the level (or confidence) in the data quality resulting from the validation process. Possible levels may include validation with:
>
> 1. Independent data (the 'true' reference dataset).
>
> As independent data set we use the TCCON total column data because they are most similar to nadir satellite data. In addition NDACC data will be used, as they cover a longer time scale.
>
> 2. Other in situ data.
>
> The in-situ data will be used in conjunction with appropriate models in a way that they are suitable for comparison with the total column observations
>
> 3. Impact studies using other CCI products

Specific spatial or temporal cases, that are scientifically interesting, will be studied in detail.

Currently, each CCI project has its own specificities in terms of ECVs and thus of validation datasets and protocols. In the first year, each team will re-use their existing validation infrastructure based on their expertise and needs. It is anticipated, as progress is made by each consortium (delivery of the PVP, definition of archived data to collect, etc.), that new validation requirements emerge and/or overlap across other ECVs. The need for a common centralized infrastructure (possibly hosted or elaborated by CMUG) to store validation (including common reference data available from other space agencies and previous ESA activities) and ancillary datasets should be carefully examined. Each CCI team is asked to keep in mind requirements of openness and possible sharing when further developing its validation infrastructure and to specifically examine common opportunities / synergies at the next co-location meeting.

To assure the quality of an ECV data product, and that the product specifications are reached, a validation process shall be performed. The validation is an iterative process that shall take into account requirements and responses from users. The validation process is unique to each CCI project and must be fully documented in the PVP. The validation process should use approved community protocols where they exist and must be fully traceable and subject to scrutiny by peer review.

The validation should follow the following requirements as given in **Table 1** of the *ESA Climate Change Initiative* ***/CCI project guidelines/***:

Validation	
V-1	*All CCI projects should use the definition of validation approved by the CEOS-WGCV.*
V-2	*All CCI project Product Validation Plans (PVP) shall adhere to the above three requirements (see section 5.4) regarding independence.*
V-3	*The CCI teams shall use established, community accepted, traceable validation protocols where they exist. If such protocols do not exist then CCI projects may adapt existing protocols if appropriate and in any event shall offer their final protocol for future community acceptance.*
V-4	*Each CCI project shall select appropriate validation data to ensure that an adequate level of validation (confidence) is applied to all output products. The level of validation (confidence) should be indicated in the output product.*
V-5	*The CCI programme should hold a dedicated session (or workshop) on common validation infrastructure during (or prior to) the next co-location meeting.*
V-6	*The PVP shall fully describe the validation process for each CCI project. An independent international review board of experts should be invited to review the PVP of each project team. Each CCI project should involve experts from the CMUG throughout their validation activities. A CCI product will be deemed to be validated once all steps of the validation process documented in the PVP have been completed and documented accordingly.*

Table 1: Validation requirements (from ***/CCI project guidelines/***).

3. Satellite data to be validated

The validation will be performed for XCO_2 and XCH_4 from SCIAMACHY/ENVISAT and TANSO/GOSAT. SCIAMACHY started measurements in 2002. TANSO started measurements in 2009. The data sets used for the validation, named the ECV Core Algorithms (ECAs), are given in **Table 2 /GHG-CCI PSDv1/**. The validation period for SCIAMACHY/ENVISAT XCO_2 and XCH_4 is January 2003 to December 2010. The validation period for TANSO/GOSAT XCO_2 and XCH_4 is April 2009 to December 2010.

ECV Core Algorithm Identifier	Full Name	Institute	Technique	Reference	Appendix For Full Product Description
CO2_SCI_BESD	IUP, Univ. Bremen, Bremen optimal ESimation DOAS (BESD) SCIAMACHY XCO_2 algorithm	IUP	Full Physics, Optimal Estimation	Reuter et al., 2010, 2011	Appendix A
CO2_SCI_WFMD	IUP, Univ. Bremen, Weighting Function Modified DOAS (WFMD) SCIAMACHY XCO_2 algorithm	IUP	Least-squares DOAS (proxy method)	Schneising et al., 2010	Appendix B
CH4_SCI_WFMD	IUP, Univ. Bremen, Weighting Function Modified DOAS (WFMD) SCIAMACHY XCH_4 algorithm	IUP	Least-squares DOAS (proxy method)	Schneising et al., 2010	Appendix C
CO2_GOS_OCFP	University of Leicester OCO Full Physics GOSAT CO2	ULE	Full Physics, Optimal Estimation	Boesch et al., 2006, Connor et al., 2008	Appendix D
CH4_GOS_OCFP	University of Leicester OCO Full Physics GOSAT CH4	ULE	Full Physics, Optimal Estimation	Boesch et al., 2006, Connor et al., 2008	Appendix D
CH4_SCI_IMAP	SRON Proxy SCIAMACHY CH4	SRON	Proxy method, Iterative maximum a posteriori solution	Frankenberg et al., 2005	Appendix E
CH4_GOS_SRPR	SRON Proxy GOSAT CH4	SRON	Proxy Method, Truncated SVD	Frankenberg et al., 2005	Appendix F
CO2_GOS_SRFP	SRON Full Physics GOSAT CO2	SRON	Full Physics, Phillips-Tikhonov regularization	Butz et al., 2009, 2010	Appendix F
CH4_GOS_SRFP	SRON Full Physics GOSAT CH4	SRON	Full Physics, Phillips-Tikhonov regularization	Butz et al., 2009, 2010	Appendix F

Table 2: Overview ECV Core Algorithms (ECAs) **/GHG_cci PSDv1/.**

4 Data sets for validation

Several data sets will be used for the validation of the satellite retrievals. The main work horse for the validation source will be the TCCON data. These total column observations have the most similar observation geometry compared to the satellites and have the required precision. TCCON measurements are performed in the near-infrared spectral region. TCCON was founded in 2004, but several sites have been established within the last two years.

In order to lengthen the validation period the NDACC observations will be included. Within the NDACC the total columns are measured in the mid-infrared region. The use of the NDACC data requires a comparison between the TCCON and NDACC products for CO_2 and CH_4, which will be performed in this validation project.

In addition to the total column observations in-situ surface data will be used for the validation. The advantage of the in-situ surface data is their long data record, their high accuracy and their good global coverage. The disadvantage is the measurement geometry, sampling the air at the surface. The application of these data requires the use of appropriate models to create total columns from the surface in-situ data that can be compared with the satellite data.

4.1 The ground-based solar absorption measurements networks TCCON and NDACC

The ground-based observations within TCCON are performed by high resolution Michelson interferometers. The Bruker instruments IFS 120 HR and 125 HR have the required short and long-term stability for these observations. These observations are organised in the international Total Carbon Column Observing network (TCCON). The network was established in 2004 in the frame of the planned OCO satellite mission by NASA. Since 2004 TCCON is working, and the number of sites is increasing. **Figure 1** gives an overview of the TCCON sites in the current situation. A few more stations (indicated as 'future sites' in the map) may become operational in the timeframe of this validation effort. Currently the TCCON observations serve for the validation of SCIAMACHY, TANSO, and AIRS. Many TCCON instruments are also part of the international Network for the Detection of Atmospheric Composition Change (NDACC). NDACC was founded in 1992 in the frame of the stratospheric ozone depletion research. Its former name was Network for Detection of Stratospheric Change (NDSC). For the first 15 years the observation of stratospheric species, like HCl, $ClONO_2$, O_3, HNO_3, was its main topic. In the last years timeseries of tropospheric species, like CH_4, CO, C_2H_6, HCN, have been included in the reported datasets, and in 2006 the network's name was changed from NDSC to NDACC.

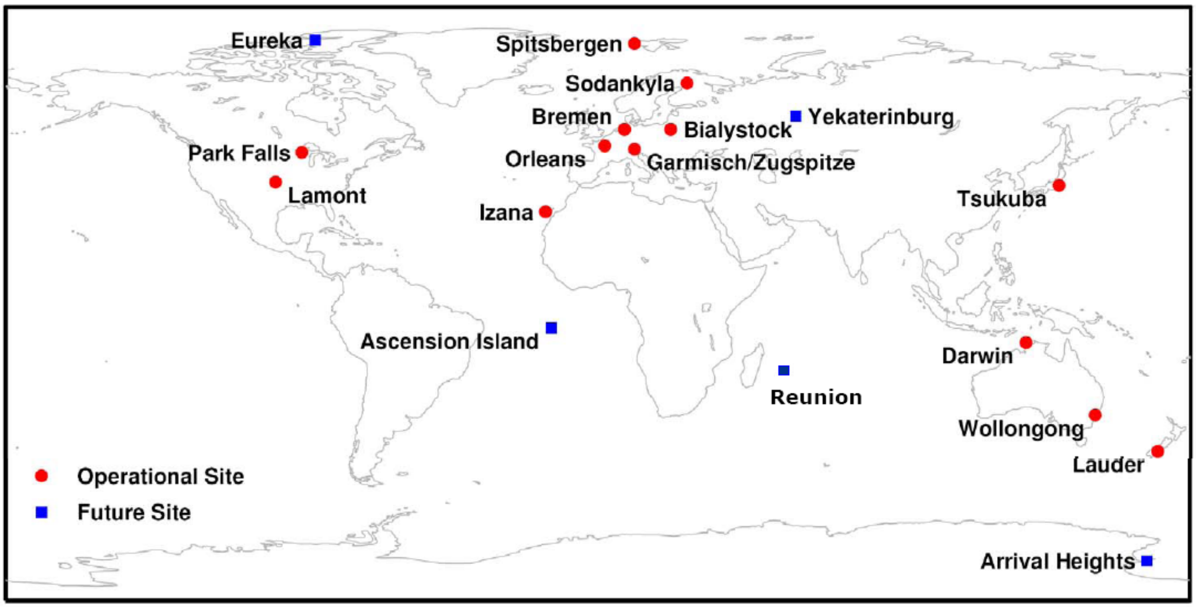

Figure 1: Operational TCCON sites (March 2011). The NDACC sites coincide often with the TCCON sites.

4.1.1 Specifics of the TCCON and NDACC remote sensing measurements

The TCCON observations are performed in the near infrared (NIR) spectral region between 4000 and 13000 cm^{-1} by pointing the solar tracker directly on the sun. The spectra are recorded at a resolution better than 0.02 cm^{-1}, where the resolution is defined as 0.9/(optical path difference, in cm). The required pointing accuracy of the solar tracker has to be better than 0.2 arc. The instrumental line shape (ILS) of the TCCON instruments is regularly measured by cell measurements, where a cell filled with HCl of known pressure is placed in the solar beam. The analysis of the ILS is performed using LINEFIT, a program developed at KIT ***/Hase et al.,1999/***.

The total column observations need to be calibrated to the WMO standard and the observations by the individual sites have to be compared to each other. This has been done throughout the last years by high flying aircraft campaigns, where CO_2 and CH_4 are measured by *in-situ* instruments between about 200 m and 12 km altitude. The calibration flights were performed between 2004 and 2009 in the US in the US, Japan, New Zealand, and Australia ***/Wunch et al., 2009/***, and as part of the European IMECC campaign in September and October 2010 in Germany, Poland and France ***/Messerschmidt et al.,2011/***. It was found that a scaling factor is required to get the agreement of the TCCON data with the WMO standard. This scaling factor can be assigned to errors in the spectroscopic data of O_2 and CO_2. It is important to note that all flight campaigns resulted in a single, constant scaling factor. This demonstrates that the TCCON observations have the required comparability between the sites and the necessary long and short term stability. It can therefore be concluded that the data are suitable for the satellite validation.

The NDACC measurements are performed in the same observation geometry as TCCON in the infrared region between about 700 cm^{-1} and 4500 cm^{-1}, resulting in a small overlap with the NIR observations of TCCON. The NDACC observations are typically recorded at a higher resolution between 0.005 and 0.0035 cm^{-1}. The observations are also irregularly monitored by cell measurements. Within NDACC several comparisons of the retrieval codes have been performed, resulting in an agreement of about 1%, sufficient for the requirements of NDACC. In order to be used for GHG_cci a further analysis of the accuracy and precision is required, as discussed later on.

The precision and accuracy of the TCCON and NDACC data are given in the table below

	XCO_2 precision (%)	XCO_2 sys. bias (%)	XCH_4 precision (%)	XCH_4 sys. bias (%)	Reference
TCCON	0.25	0.20	0.40	0.40	*
NDACC	Tbd**	Tbd**	Tbd**	Tbd**	Tbd**

Table 3: precision and error for the systhematic bias of the two ground-based FTS networks. The results are always given as 2 σ for single spectra. Tbd means to be done, as part of this project.

* **/Messerschmidt et al., 2011/**; **/Wunch et al., 2011/**; **/Wunch et al., 2010/**

** Tbd: To be done in this ESA project

4.1.2 Differences between satellite and ground-based remote observations

When comparing ground-based and satellite observations several important issues need to be considered, (1) the temporal and spatial match of satellite and validation data (2) spatial representativeness of the sites (3) the number of sites to be used for the validation (4) the spectral resolution, (5) the sensitivity of the observations to the trace gas of interest as a function of altitude, (6) the airpath through the atmosphere, (7) the spectral regions analysed, (8) the a-priori profile assumed for the retrieval, and (9) the spectral data (linelists) used for the retrieval.

(1) The validation will be performed for satellite data that are within 3 radii of the ground sites: 100 km, 350 km and 500 km. The temporal co-registration criterion is 2 hours, i.e., only data shall be used for comparison, which have been obtained within +/- 2 hours compared to the time of the satellite retrievals. Both spatial and temporal collocation conditions will be evaluated during the course of the validation and, if need be, adjusted to obtain the optimal statistically representative comparison dataset. Any deviation from the above mentioned preset conditions will be motivated and documented and consistently applied in the validation of all satellite products.

(2) Due to the large pixel size of satellite data the validation can best be performed for flat sites with only minor local sources and sinks. For sites inside a city with local emissions, or for mountain sites, the specifics of the sites need to be considered in the validation. E.g., a concept for altitude-correction of XCH_4 data from ground-based sites on top of mountains that makes use of the correlation between satellite derived XCH_4 and ground-pixel altitude will be employed.

(3) The number of ground sites is limited. While a fairly good coverage is available for Western Europe there are no stations across Siberia. The station Yekaterinburg just started operation and is currently not integrated in the TCCON quality control scheme. In the Southern hemisphere the ground-truth validation is provided by only three stations, located in Australia and New Zealand. The greatest problem of the ground-based network is the lack of stations in the tropics. The tropics play a key region in the global carbon cycle. E and especially in the tropics validation would be required due to the high water content and frequently occurring convective clouds, which have an impact on the satellite retrievals.

(4) The spectral resolution of the GOSAT satellite instrument is about 10 times worse compared to the ground-based observations that work at 0.02 cm^{-1} or better. Theoretically differences in the spectral resolution should not influence the results. However, the choice of the a-priori used in the analysis, the effect of interfering gases, and errors in the assumed temperature profiles will influence the spectra recorded at different resolutions in a different way. Most probably this discrepancy is not constant but variable as function of SZA, latitude, etc, so systematic errors will vary from site to site.

(5) The sensitivity of the measurements to the trace gas of interest as a function of altitude is expressed by the averaging kernels (AK). The averaging kernels depend on the observation geometry, the retrieval algorithm, the absorption depth, the spectral line shape, and the assumptions on the concentration profiles. Since satellite and ground-based observations of CO_2 and CH_4 use the same spectral region and similar observation geometry the AKs are similar. For O_2 the spectral regions differ (see below at (7)), but the O_2 concentration can be assumed to be constant and well known, therefore the influence of the AKs is negligible.

(6) The satellite instruments work with sunlight scattered at the earth surface, while the ground-based observations use the direct sunlight. For the satellite observations the photons pass the atmosphere twice in different atmospheric regions. The determination of the effective total airmass observed by the satellite sensor is not determined by the geometry of the observation in a simple manner (scattering from cirrus, aerosols, partial cloud coverage, etc). This introduces additional errors in the data of the satellite sensor which do not show up in the ground-based solar absorption measurement.

The second satellite airpath, from the earth to the instrument does not agree with the ground-based geometry. Thus, a 100% agreement can never be achieved. This means that a temporal and apparent agreement of 100% between satellite and ground-based pixels should not be over interpreted.

(7) The spectral regions used by satellites and the ground-based observations differ, especially for O_2. While the satellites observe O_2 at 0.76 µm the TCCON routine O_2 retrieval works at 1.27 µm (centered at 7882 cm^{-1}). For ground-based observations there are several advantages to use this band arising from its closer proximity to the CO_2 and CH_4 bands (e.g. **/Wunch et al., 2010/**). However, this region cannot be used for satellite spectra due to the strong airglow, which cannot be considered sufficiently well in the retrieval.

(8) The retrieval requires as input the volume mixing ratios (VMRs) of all atmospheric trace gases (a-priori profiles) absorbing in the spectral window that is analysed. Of special importance is the a-priori mixing ratio profile of the target gas. Due to the limited vertical resolution of the retrieved solution, any analysis scheme will be affected by the choice of the a-priori. Even if the same a-priori is used for the satellite and the ground-based retrieval, systematic differences in the deduced columns still will result due to different vertical resolutions / sensitivities of the two kinds of observations.

(9) The results for the total columns depend on the spectral linelist used for the retrieval. The linelists used differ between the different satellite retrievals as well as between ground-based and satellite retrievals.

4.2 Comparison of mid-infrared and near infrared observations

Most TCCON observations started within the last years **/Wunch et al., 2011/**. Since SCIAMACHY was launched in 2002 it would be desirable to increase the length of the ground-based data set for the validation. The NDACC, working in the mid-infrared, has been established in 1990 **/Kurylo and Solomon, 1990/**. Many TCCON sites work also in the mid-infrared spectral region and are part of the NDACC. The NDACC mid-infrared observations of CO_2 and CH_4 can also be used for the satellite validation. This requires a careful comparison between the NDACC and TCCON observations.

4.2.1 NDACC to TCCON comparison for CO_2

Responsibility: Thomas Blumenstock, Thorsten Warneke

Within this task a retrieval strategy for CO_2 in the mid-IR spectral region will be developed and a comparison will be performed with the retrievals in the near-IR spectral region. The comparison will be performed at one site, where NDACC as well as TCCON measurements are available.

Within TCCON the retrieval of XCO_2 is performed in the near-IR spectral region. The retrieval in the near-IR has several advantages compared to the mid-IR. The main advantage is the small number of interfering gases, which allows the retrieval in broad spectral windows reducing the temperature sensitivity. This holds also for O_2, which can be retrieved in near infrared close to the CO_2 and CH_2 bands. The O_2 is used to convert the slant columns into a mean dry-air mole fraction in the following way: The slant column of the CO_2 is divided by the O_2 slant column and multiplied by the dry-air mole fraction of O_2 (0.2095). Due to the division with O_2 several systematic errors that affect CO_2 and O_2 equally, like pointing errors, cancel. Last but not least the TCCON retrievals are performed in the same spectral region as the satellite retrievals. Due to the advantages of the TCCON retrievals mentioned above it cannot be expected that the results from the mid-IR spectral region meet the TCCON standards. Nevertheless it is useful to develop a mid-IR retrieval strategy and compare it to the TCCON retrievals. The retrieval should include a determination of the precision and the bias, where the bias might have a seasonal dependency.

Mid-IR spectra are available from the NDACC. If XCO_2 could be retrieved from these mid-IR spectra with a precision useful for satellite validation, all NDACC stations would potentially be available as ground-based validation sites. This would provide a better spatial coverage. In addition the NDACC observations are performed since more than 15 years at many sites and such long records of column observations would be highly valuable for carbon cycle studies.

4.2.2 NDACC to TCCON comparison for CH_4

Responsibility: Ralf Sussmann, Frank Forster

This task comprises two steps, i) update of the NDACC retrieval strategy for CH_4 and ii) NDACC-TCCON comparison and inter-calibration.

During the last year it has become obvious that the hitherto used NDACC retrieval strategy for CH_4 requires updates. First, an empirically tuned line-parameters compilation had been used in the community and it was favored meanwhile within the NDACC infrared working group that it should be replaced by an official version of the HITRAN line list. Second, it was felt but not proven, that for wet sites H_2O/HDO-CH_4 interference errors may occur in case of using non-optimum line lists, and/or using a non-optimum strategy for taking interfering species into account, and/or using a non-optimum micro-window selection. An intense study on this complex problem is under way that investigates multi-annual data from 3 representative test sites, i.e., Wollongong (high humidity), Garmisch (medium humidity) and Zugspitze (low humidity) and 3 different official version HITRAN line lists. The outcome of this study is a recommendation for an updated new mid-IR retrieval strategy for methane with optimum precision and minimum H_2O/HDO interference errors.

NDACC-TCCON comparisons will be performed with the updated mid-IR retrieval strategy. Data from stations with coincident mid-IR (NDACC) and near-IR (TCCON) measurements are exploited for this purpose (Ny-Ålesund, Bremen, Karlsruhe, Garmisch, Izana, Wollongong). The overall goal of the study is to quantify the NDACC-TCCON intercalibration. One question to be addressed on this way is whether or not one overall intercalibration factor for all stations can be found. Furthermore, it shall be investigated whether the time series of intercalibration factors show a significant seasonal component which could arise due to airmass-dependent artifacts and could be parameterized.

Results for 4.2.1 (CO_2) and 4.2.2 (CH_4):

i) Recommendation on the updated NDACC retrieval strategy including micro window selection, interfering species treatment, spectroscopic line list, regularization approach, final quality selection of retrievals, harmonized procedure to calculate dry-air mole fractions. Characterization of information content and limiting errors, station-to-station bias.

ii) TCCON/NDACC-intercalibration factor for XCO_2 and XCH_4 determined from the stations with coincident NDACC and TCCON measurements. The outcome will ideally be one number, but also could well be a seasonally and/or latitudinally varying intercalibration function.

4.3 Use of in-situ data for satellite validation

Responsibility: B. Buchmann

A large number of in-situ ground-based sites exist where greenhouse gases have been measured routinely for many years, in some cases decades, by established networks such as the Global Atmospheric Watch (GAW) program of WMO. The protocol of the 4th WMO Round Robin Reference Gas Intercomparison (2002-2007) indicates an accuracy of about 2 ppb for CH_4 and 0.1 ppm for CO_2 (14th WMO/IAEA Meeting of Experts on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurement Techniques. Helsinki, Finland, 10-13 September 2007, GAW report No. 186)). CH_4 has been measured in the past with Gas Chromatography - Flame Ionization Detectors (GC-FID) at both GAW and AGAGE sites, but since 2009 more and more sites are being equipped with laser spectroscopic instruments (e.g. from Picarrro Inc. USA). CO_2 has been measured using non- dispersive infrared sensors (NDIR) and similarly as for CH_4, are currently being replaced with cavity enhanced laser-spectrometers at GAW sites. All measurements are traceable to the WMO CO_2 mole fraction scale and NOAA04 CH_4 mol fraction scale (except for a small number of sites for which Empa will apply the respective scaling to NOAA04), respectively.

The following data sources are planed to be used in order to compare in-situ with satellite (SCIAMACHY and TANSO) derived trace gas products.

Ground-based:

- Global Atmosphere Watch (GAW) and Advanced Global Atmospheric Gases Experiment (AGAGE) sites with continuous (hourly) observations of CO_2 and/or CH_4, see **Figure 2**.
- Regional networks e.g. CarboEurope, NOAA ESRL GMD Tall Tower Network

Aircraft measurements:

- NOAA ESRL aircraft program (flask samples) and selected aircraft campaigns such as Hippo. The aircraft measurements will primarily be used to test different methods to construct vertical columns from ground-based in situ measurements.

Auxiliary data

- Global model CO_2 profiles for each site from
  - Carbon Tracker: 3°x2° (**/Peters et al., 2007/**)
  - Transcom models (**/Gurney et al., 2000/**), still to be specified
- Global model CH_4 profiles for each site from
  - TM5 4°x6° (**/Bergamaschi et al., 2009/**)
  - FLEXPART 1°x1° (**/Henne et al., 2011/**)
  - Transcom models (**/Gurney et al., 2000/**), still to be specified
- Meteorological data measured at the in-situ site (temperature, surface pressure, wind speed and direction) and/or from ECMWF (tropopause height, boundary layer height, specific humidity profiles, temperature profiles, wind speed and direction (profiles) and surface pressure from ECMWF operational analyses) for each site

- Information supporting the description, classification, and assessment of representativeness of the individual sites such as DEM (GTOPO30), land use type (GLOBCOVER), emission database (EDGAR) or population density.

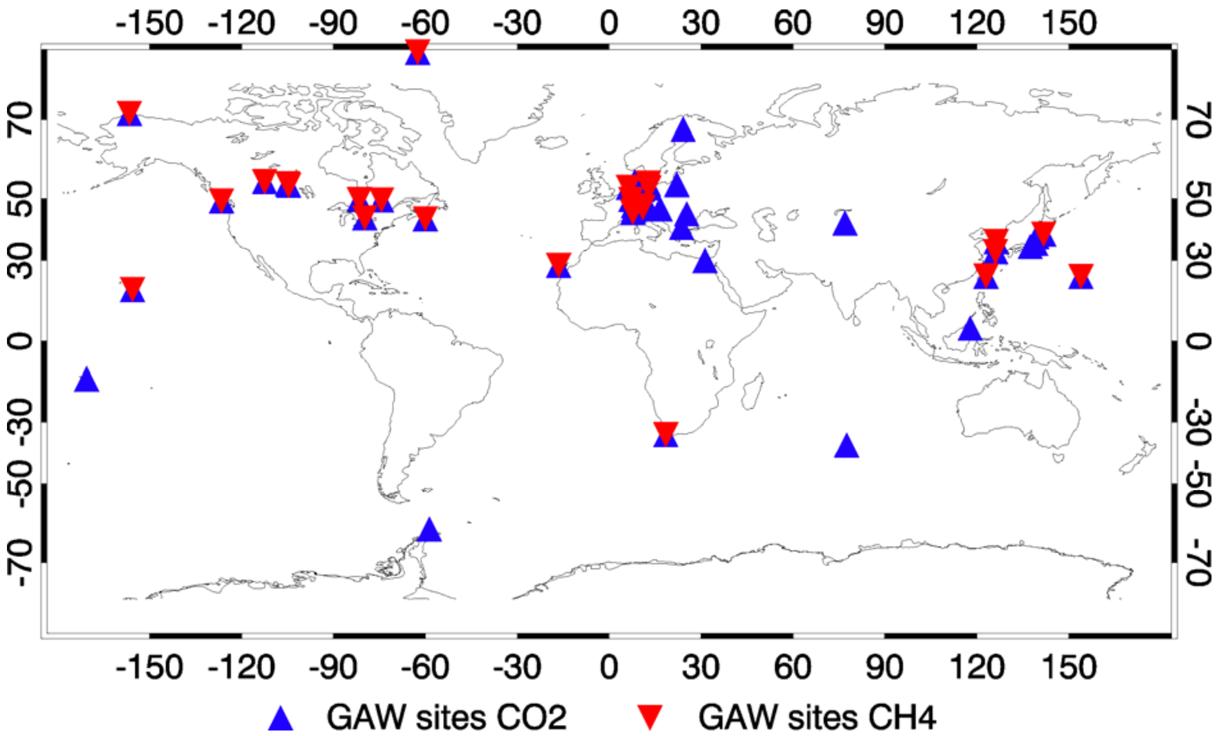

Figure 2: GAW sites with continuous observations of CH_4 and CO_2 since 2003.

4.3.1 Comparison plan

Surface mixing ratios are not necessarily representative for columnar quantities. Surface values might be strongly influenced by local sources or, in contrary, columnar quantities might contain large signals from long-range transport aloft, not detected by in-situ instruments. For example, **/Keppel-Aleks et al., 2010/** found that column averaged CO_2 are largely determined by large-scale phenomena. Further, seasonal variations can vary strongly at different atmospheric heights, e.g. the stratospheric (relative) contribution of CH_4 to the columnar quantity correlates with tropopause (**/Frankenberg et al., 2010/**). A second difficulty arises from comparing a point measurement to spatially integrated data (satellite) which requires a proper collocation of the two in both space and time. Chemical transport models (or if available aircraft profiles) could provide an indispensible link to better understand these differences and to relate and compare in-situ trace gas measurements to those retrieved from spaceborne remote sensing data.

The outline described in the sections below should be understood as an iterative process: selection of suitable sites, collocation of the different data sets, converting in-situ to columnar quantity, statistical analysis between the different data sets (especially between in-situ and satellite retrieved trace gas columns). The statistical analysis might propose to alternate the selected sites or the way different data sets should be collocated. In addition, the conversion of in-situ to columnar averaged mol fractions might depend on the collocation approach. Furthermore, it is planned to begin the testing and analysis at a few "supersites" with good data coverage from different instruments (e.g. in-situ (GC-FID, flask, etc.), FTS (TCCON or NDACC), tall towers, aircraft profiles, etc.) and subsequently extend the analysis to other sites.

4.3.1.1 Description, selection, and classification of measurement sites

In a first step, the sites providing in-situ data will be classified and suitable sites will be selected for comparison with satellite derived trace gas quantities. The selected sites should ideally cover a wide range of different characteristics in order to assess the capability of different retrieval algorithms and their limitations. In this regard, the following criteria will be considered:

1.) Global coverage: e.g. in the Northern and Southern Hemisphere, tropics.

2.) Site type: background or influenced by local sources and sinks, influenced by anthropogenic sources or biogenic sources and sinks, altitude (mountainous sites might introduce a representativeness error and relative contribution of stratospheric column increases), meteorological conditions (e.g. cloud coverage, aerosol abundance, precipitation, snow cover, water vapour), temporal, spatial, and vertical variability of trace gas.

3.) Sites with at least 1 year of continuous measurements after 2002.

4.) Availability of additional data: e.g. FTS (TCCON or NDACC), aircraft profiles,

5) Sites not assimilated into models will be preferred to ensure independence of information provided by models and in situ measurements

4.3.1.2 Collocation

Comparing point measurements to spatially integrated data (satellite, model output) requires a proper collocation of these variables in both space and time. We will follow the same approach as for the TCCON and satellite observations comparison, the temporal co-registration is +/- 2 hours (c.f. 4.1.2).

4.3.1.3 Accounting for vertical trace gas distribution

A crucial step is to relate the in-situ measurements to columnar averaged dry air mole fractions (or total columns) because surface mixing ratios are necessarily representative for columnar averages. Additional knowledge about the profile shape of the trace gas mixing ratios is required. In addition, trace gas concentrations at different layer heights can exhibit different or even opposite seasonal variations and the relative contribution of the stratosphere to the total column seasonally varies and is correlated to the tropopause height (as outlined by **/Frankenberg et al., 2010/** for CH_4).

One possibility for comparing surface and satellite observations is to assimilate the surface observations into a model and then to compare satellite versus model columns. Model values at the surface do not need to match the observations exactly but, depending on representativeness errors or model deficiencies, can differ more or less strongly from the observations. Another problem is that current global model assimilation systems are operating at coarse resolution not able to resolve the specific situation encountered at many surface observation sites. Note that the resolution of such models is typically much coarser than the observations of SCIAMACHY and GOSAT.

Here we will develop and apply an approach that allows for a more direct comparison. It integrates in-situ measurement data with model profile information, yet putting more weight on the measurements. The focus will be placed on observation sites not assimilated into the model to ensure independence of the two sources of information. The basic idea is to compile profile shapes from an ensemble of available chemical transport model results (e.g. TM5, FLEXPART, Transcom) and to adjust these profiles to match the in-situ observations at the surface based on auxiliary information such as boundary layer and tropopause heights. Different options for profile adjustment will be evaluated for sites where additional information such as total column (TCCON) or vertical profiles (aircraft) are available. Using the output of different chemical transport models allows us to describe and analyze the vertical variability in the trace gas distribution based on a larger statistical sample. This variability will provide a measure for the uncertainty of the vertical profile shape which, in combination with the uncertainty in the in-situ data, can be used to estimate the overall uncertainty of the constructed vertical column. The total column can then be compared in a straightforward manner with the satellite products taking into account the respective uncertainties. The different data sets (SCIAMACHY and TCCON or NDACC trace gas columns) could be adjusted for a common a priori profile using their averaging kernels although it is expected that this has a comparatively small impact (c.f. **/Reuter et al., 2010/**).

5 Round Robin Inter-comparisons and Evaluation

The validation will follow closely the ***/CCI project guidelines/*** and the ***/GHG-CCI Round Robin Evaluation Protocol/***. The most relevant table from the ***/CCI project guidelines/*** is given here:

Round-Robin algorithm inter-comparison and selection	
RR-1	*The meaning of the 'best' algorithm and of how to select it (evaluation protocol) has to be defined before the start of the Round Robin exercise. The definition of 'best' and the scope of the Round Robin exercise have to be specified in the Product Validation Plan (PVP).*
RR-2	*The Round Robin should be made at the beginning of the project based on objective criteria. There should be one or more iterations to show algorithm improvement throughout the project. The most objective algorithm selection would be based on blind testing to avoid any bias.*
RR-3	*Every CCI project has to perform a Round Robin exercise. In the exceptional case that a final algorithm has been pre-selected, component modules need to be tested also for this pre-selected algorithm. Furthermore, the pre-selection criteria should be in line with the CCI objectives.*
RR-4	*The same auxiliary and Level 1 data should be used in the processing, as well as the same reference data.*
RR-5	*The round robin results need to be open and the algorithm must be well-documented and public, but the actual code does not need to be public.*
RR-6	*The algorithm selection should be made by an independent team that is not directly involved in the algorithm development, although of course the members of that team should be experts. The selection shall be made based on a Round Robin evaluation protocol developed beforehand and providing objective criteria.*
RR-7	*The development of new tools should only be considered when really needed and no good tools for the purpose are available.*

Table 4: Requirements for the Round Robin Exercise (from: **/CCI project guidelines/**).

5.1 Overview of the Evaluation protocol

Our evaluation protocol follows the **/GHG-CCI Round Robin Evaluation Protocol/** (RREP). 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 modeling of these two important anthropogenic greenhouse gases (GHG). 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 (i.e., four algorithms or less).

The Evaluation protocol describes how the algorithm selection will be done at the end of the 2 year "Round Robin" (RR) phase of the GHG-CCI project.

The selection will be based on (but not limited to) comparisons with the highly precise and accurate ground-based TCCON FTS XCO_2 and XCH_4 retrievals. In addition the ground-based NDACC remote sensing total column observations of CO_2 and CH_4 will be used to cover a larger time span. The remote sensing observations are optimally suitable for the comparison but due to their spatial and temporal sparseness the amount of data is limited. Therefore *in-situ* data from well established surface networks, together with model studies to convert the surface data to a column, will be used in addition to the remote sensing data.

For each product and each site several 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 will be computed by the independent GHG-CCI VALidation Team (VALT). In addition, the FoMs will also be determined by the retrieval team to cross-check and to find out to what extent the results depend of the details of the analysis method chosen.

This information will be evaluated by the independent GHG-CCI Climate Research Group (CRG). 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 experts knowledge / judgment. For this purpose, additional information will be made available. 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 is expected that for each product a clear "winner" can be identified. In case 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 modeling. 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 approach proposed in this document is the baseline defined at project start. If it turns out that refinements are needed or that even an entirely different approach is required, this baseline may change.

5.2 Spatio-temporal requirements for TCCON comparisons

Within the comparison period FTS, GOSAT, and SCIAMACHY data should all be available. This is important to assess the consistency of the SCIAMACHY and GOSAT time series which requires overlap in time.

Additionally, 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). Ideally, the period should cover full years only, so that no seasons are over-weighted within the comparison statistics.

SCIAMACHY is operational since 2002. GOSAT was launched in January 2009 and calibrated Level 1 data are available since April 2009. Many FTS sites are operational longer than GOSAT data are available; however, there are some sites which got operational not before August 2009 (see **Table 1**).

Resulting from the limitations of data availability, 01.01.2010 – 31.12.2010 (baseline period) has been chosen as the minimum comparison time period of the Round-Robin (RR) exercise for SCIAMACHY and GOSAT. This is however not sufficient.

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.

2010 is therefore the mandatory year which needs to be analyzed for both SCIAMACHY and GOSAT. This is also important in order to assess the consistency of the data products of the two sensors. In addition, all the other years where data of the corresponding satellite are available shall also if possible be analyzed (per year, i.e., independently for each year, and for those FTS sites where data are available).

Table 5 lists all TCCON sites (https://tccon-wiki.caltech.edu/) which have been chosen for the RR comparison with TCCON. This means nearly all TCCON sites which fully cover the comparison period are used. Only the sites Eureka (Canada), Ny-Ålesund (Spitsbergen), and Izana (Tenerife) have been excluded because of orography, jagged coastline, ocean, and/or persistently snow/ice covered surfaces.

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 may have to be extended, e.g., using also NDACC FTS sites/retrievals. This will be decided later in the project, when the quality of the NDACC retrievals for this purpose has been assessed in (more) detail.

For all chosen TCCON sites all “good” satellite retrievals shall be used for the comparison (note that either the data product contains only “good” data as determined by the retrieval team or the data product contains a corresponding quality flag).

For the comparison, all satellite retrievals within a given distance (radius or latitude/longitude range) around the FTS sites shall be used. The optimal distance still needs to be defined. A smaller distance (e.g., 100 km) is preferred for better representativeness but may result in too few data points for comparison (at present 500 km is typically used for GOSAT). It is therefore suggested to deliver all retrievals within a 500 km radius around the FTS sites and to perform comparisons using 3 radii: 100 km, 350 km and 500 km.

In addition, the comparison shall be limited to satellite retrievals over land.

The temporal co-registration criterion is currently defined as 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. This is believed to be the pre-validation optimal temporal overlap criterion, which offers the best representativeness-overlap balance. This criterion (as the spatial one) will be evaluated during the initial validation phase and, if need be, adjusted on a station to station basis.

Site	Established	Lat	Lon	Alt. [km]	PI
Bialystok, Poland	03/2009	53.23N	23.025E	0.18	Notholt, Warneke
Bremen, Germany	07/2004	53.10N	8.85E	0.04	Notholt, Warneke
Darwin, Australia	08/2005	12.424S	130.892E	0.03	Griffith
Garmisch, Germany	07/2007	47.476N	11.063E	0.74	Sussmann
Karlsruhe, Germany	09/2009	49.100N	8.438E	0.11	Hase
Lamont, USA	07/2008	36.604N	97.486W	0.32	Wennberg
Lauder, New Zealand	06/2004	45.038S	169.684E	0.37	Sherlock, Connor
Orleans, France	08/2009	47.97N	2.113E	0.13	Notholt, Warneke
Park Falls, USA	05/2004	45.945N	90.273W	0.44	Wennberg
Sodankyla, Finland *)	01/2009	67.368N	26.633E	0.18	Kyro
Tsukuba, Japan *)	12/2008	36.0513N	140.1215E	0.03	Morino
Wollongong, Australia	05/2008	34.406S	150.879E	0.03	Griffith

Table 5: List of all TCCON sites that are used in the validation.

5.3 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. Furthermore it has been decided to also provide the users with additional information such as maps and time series and comparisons with models and other (corresponding) data products. The GHG-CCI CRG is confident that using this information and their expert knowledge (as well as additional analysis carried out by the CRG) it is possible to make a well justified objective decision on which algorithm to use for the ECV generation (in year 3 of the GHG-CCI project).

In **Table 6** the Figures of Merit (FoM), which have been identified, are presented.

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 6**.

To enable a quantitative analysis of the results, files in a format that has to be specified, have to be generated and used (baseline: ASCII tables).

The FoM and auxiliary information will be computed by the retrieval team and by the validation team to double check the correctness of the computation and/or to check the sensitivity of the numerical values on the details of the computational method.

Note: The FoM may be computed after bias correction. If a bias correction has been applied this needs to be properly documented and this information has to be made publicly available.

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	"	"	Seasonal bias
FoM_B2	Bias AMJ	"	"	Seasonal bias
FoM_B3	Bias JAS	"	"	Seasonal bias
FoM_B4	Bias OND	"	"	Seasonal bias
FoM_SD	StdDev	"	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 6: Figures of Merit (FoM) for the validation.

6 Validation schedule

Responsible: Bart Dils, Martine De Mazière, Frank Hase

The validation of the satellite data of SCIAMACHY and TANSO will be performed using the data sets described above. This includes TCCON, NDACC, and in-situ data. The in-situ data will be used in the way that columns will be built, where the in-situ data have been assimilated. The results in terms of the FoM will be summarized for each site and data product.

The TCCON data, NDACC data and in situ data will be delivered to the validation data bases DBT2 and DBT3 where they can be extracted by the validation team.

The TCCON data for CO_2 and CH_4 will be obtained from the CalTech TCCON data base (additional data will be provided by TCCON PIs). The NDACC data for CO_2 and CH_4 will be taken from the NDACC database. Bias results from the TCCON to NDACC comparison will be taken into account. The total columns based on in situ data will be provided by EMPA. The satellite data have to be extracted from the data base generated by the retrieval group.

The validation consists of 3 major validation phases:

In the first validation phase, the validation will be performed using TCCON data only. This initial validation will cover the year 2010. The validation data set for the second phase will include NDACC and model assimilated in situ data, if their comparisons with TCCON have shown them to be sufficiently reliable. The final ECV comparison will be expanded in order to cover the entire satellite measurement period (2003-2010 (SCIAMACHY) and 2009-2010 (TANSO)).

The validation will be performed on nine different satellite data sets (three data sets from IUP, four from SRON, two from Leicester; see **Table 2**), following the Round Robin Exercise.

Phase 1: "Intermediate Round Robin validation":

Month 10 – Month 21 (June 2011 – May 2012)

At month 10 DBT2 will be ready to deliver the TCCON data and the RRDP will contain initial satellite retrievals. Using these data a preliminary validation of the existing intermediate satellite data products will be performed. The results shall help the retrieval teams to identify issues for algorithm improvements. This phase will know 2 validation and algorithm improvement rounds. The first interim validation round is scheduled for June-August 2011, while the second is scheduled for month 17 (January 2012), while in between and after the algorithm developers will update and reprocess their data. Of course the exact timing can be adjusted (to a certain degree) to tailor for the specific needs of the algorithm development teams and the availability of intermediate satellite data products. In any case all algorithms should at minimum pass two validation-development cycles within this validation.

- Validation data: Mainly TCCON.

- NDACC activities: Perform the comparison TCCON to NDACC.

- In situ activities: Develop scheme and test it.

- Other: Study representativeness of all site

Phase 2: "RR validation for final ECV algorithm selection":

Month 22 - Month 24 (June 2012 – August 2012).

Main goal: In this phase the optimized (in phase 1) final Round Robin algorithms will be validated. Based on the results of this validation a final selection on the GHG ECV products will be made. By this time the NDACC-TCCON comparison and In situ development studies should be finished and their data, covering the 2003-2010 time period, stored in DBT3 (by month 21, May 2012), together with the TCCON data.

- Validation data: Mainly TCCON.

- NDACC data: contribute to validation, if possible.

- in situ data: contribute to validation if possible.

Phase 3: "ECV validation":

Month 31 - Month 33 (March 2013 – May 2013)

Main goal: Validate the first GHG ECV data products.

Val data: TCCON + NDACC + in situ.

Output: PVIR end of Month 33 (July 2013).

6.1 Representativeness of the sites

The representativeness of each site (TCCON, NDACC, in-situ) will be assessed based on a collection of auxiliary datasets and parameters (c.f. 4.3. and 4.3.1.1). The surroundings of each site will be characterized with respect to topography (based on digital elevation model data (DEM, GTOPO30)), land use (GLOBCOVER, ESA), population density (CIESIN, Columbia University) and local CO_2 and CH_4 emissions (EDGAR or other high-resolution inventory). Analysing the variability of the measurements at the different sites (also in combination with wind speed and direction) should provide additional information on the individual site’s representativeness and can be used to define or refine the collocation criteria.

6.2 Validation for XCH_4 and XCO_2

Prior to the actual comparison, the satellite data need to be filtered according to the spatial and temporal validation criteria as set forth in Chapter 5.2 and assessed as in Chapter 6.1. The actual comparisons will be performed with the individual single spectra or averages for up to 30 minutes, depending on the availability of comparative data. Note that whatever option is chosen, all satellite algorithms will be subjected to the same validation process, and the choice will be made based on the algorithm with the smallest available comparative dataset.

Using the obtained collocated satellite-validation set data pairs, all the validation parameters, or so-called Figures of Merit and Auxiliary criteria as listed in **Table 2** (see Chapter 5.3), will be determined. These figures, together with the observed time series, will grant us insight into the quality of each algorithm, whether it has a significant bias, accurately measures the seasonal cycle and whether the uncertainty as put forth by the satellite retrieval teams are realistic.

Documentation of the results from the Round Robin Exercise:

Two main reports will be generated:

- The Product Validation and Algorithm Selection Reports (PVASR), due month 24 (Aug 2012), i.e., at the end of the 2-year Round Robin phase.
- The Product Validation and Intercomparison Report (PVIR), due 33, i.e. May 2013.

Both reports will be made publicly available via the GHG-CCI web site.

7 References

/Bergamaschi et al., 2009/ Bergamaschi, P., et al.: Inverse modeling of global and regional CH_4 emissions using SCIAMACHY satellite retrievals, J. Geophys. Res., 114, D22301, 2009.

/CCI Project Guidelines/ CCI project guidelines of the ESA Climate Change Initiative, EOP-DTEX-EOPS-SW-10-0002, 5 Nov 2010.

/GHG-CCI PSDv1/ Project Specification Document Version 1 (PSDv1) for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)

/GHG-CCI Round Robin Evaluation Protocol/ GHG-CCI Round Robin Evaluation Protocol (RREP), Version 1, 8 Mar 2011. http://www.iup.uni-bremen.de/~buch/ghgcci_public/RREPv1_GHG-CCI.pdf

/GHG-CCI URDv1/ GHG-CCI User Requirements Document Version 1 (URDv1), Version 1, 3 February 2011.

/Gurney et al., 2000/ Gurney, K., Law, R., Rayner, P., and A.S. Denning: "TransCom 3 Experimental Protocol," Department of Atmospheric Science, Colorado State University, USA, Paper No. 707, 2000.

/Hase et al., 1999/ Hase, F., Blumenstock and, and Paton-Walsh, T. C.: Analysis of the instrumental line shape of high-resolution Fourier transform IR spectrometers with gas cell measurements and new retrieval software, Appl Optics, 38, 3417-3422, doi:10.1364/AO.38.003417, 1999.

/Henne et al., 2011/ Henne, S., D. Brunner, J. Klausen, and B. Buchmann: A global Lagrangian chemistry transport model for CO and CH_4: Concept and validation, Atmos. Chem. Phys., 2011, in preparation.

/Kurylo and Solomon, 1990/ Kurylo, M.J. and Solomon S., Network for the Detection of Stratospheric Change, NASA Rep., Code EEU, 1990.

/Messerschmidt et al., 2011/ J. Messerschmidt, M. C. Geibel, T. Blumenstock, H. Chen, N. M. Deutscher, A. Engel, D. G. Feist, C. Gerbig, M. Gisi, F. Hase, K. Katrynski, O. Kolle, J. V. Lavric, J. Notholt1, M. Palm1, M. Ramonet, M. Rettinger, M. Schmidt, R. Sussmann, G. C. Toon, F. Truong, T.Warneke, P. O.Wennberg, D. Wunch, and I. Xueref-Remy, TCCON column-averaged CO_2: the first aircraft campaign over European TCCON sites, submitted to Atmos. Chem. Phys., 2011.

/Peters et al., 2007/ Peters, W. et al.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. PNAS, 104(48), 18925–18930, 2007.

/Reuter et al., 2010/ M. Reuter, H. Bovensmann, M. Buchwitz, J. P. Burrows, B. J. Connor, N.M. Deutscher, D. W. T. Griffith, J. Heymann, G. Keppel-Aleks, J. Messerschmidt, J. Notholt, C. Petri, J. Robinson, O. Schneising, V. Sherlock, V. Velazco, T. Warneke, P. O. Wennberg, and D. Wunch, 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, 2010.

/Wunch et al., 2009/ Wunch, D.; Toon, G. C.; Wofsy, S. C.; Stephens, B. B.; Blavier, J. L.; Keppel-Aleks, G.; Sherlock, V.; Connor, B. J.; Deutscher, N. M.; Griffith, D. W.; Macatangay, R.; Wennberg, P. O., A Calibration for the Total Carbon Column Observing Network using HIPPO Aircraft Profiles, American Geophysical Union, Fall Meeting 2009, abstract #A44A-02

/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, 2010, in press.

End of document