ECMWF COPERNICUS REPORT

Copernicus Climate Change Service

Algorithm Theoretical Basis Document (ATBD) – ANNEX E for IASI CO_2 and CH_4 and AIRS CO_2 mid-tropospheric products

C3S_312a_Lot6_IUP-UB – Greenhouse Gases

Issued by: C. Crevoisier, LMD/CNRS, France

Date: 04/10/2018

Ref: C3S_D312a_Lot6.2.1.2-v2_ATBD_ANNEX-E_v2.0

Official reference number service contract: 2016/C3S_312a_Lot6_IUP-UB/SC1

Contributors

INSTITUTE OF ENVIRONMENTAL PHYSICS (IUP), UNIVERSITY OF BREMEN, BREMEN, GERMANY
(IUP)
M. Buchwitz

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS), LABORATOIRE DE METEOROLOGIE DYNAMIQUE (LMD), PALAISEAU, FRANCE
(LMD/CNRS)
C. Crevoisier
R. Armante

Table of Contents

History of modifications

Version	Date	Description of modification	Chapters / Sections
1.1	20-October-2017	New document for data set CDR1 (until 2016)	All
2.0	4-October-2018	Update for CDR2 (until 2017)	All

Related documents

Reference ID	Document
D1	Main ATBD: Buchwitz, M., et al., Algorithm Theoretical Basis Document (ATBD) – Main document, C3S project C3S_312a_Lot6_IUP-UB – Greenhouse Gases, v2.0, 2018. *(this document is an ANNEX to the Main ATBD)*

Acronyms

Acronym	Definition
AIRS	Atmospheric Infrared Sounder
AMSU	Advanced Microwave Sounding Unit
ATBD	Algorithm Theoretical Basis Document
BESD	Bremen optimal EStimation DOAS
CAR	Climate Assessment Report
C3S	Copernicus Climate Change Service
CCDAS	Carbon Cycle Data Assimilation System
CCI	Climate Change Initiative
CDR	Climate Data Record
CDS	(Copernicus) Climate Data Store
CMUG	Climate Modelling User Group (of ESA's CCI)
CRG	Climate Research Group
D/B	Data base
DOAS	Differential Optical Absorption Spectroscopy
EC	European Commission
ECMWF	European Centre for Medium Range Weather Forecasting
ECV	Essential Climate Variable
EMMA	Ensemble Median Algorithm
ENVISAT	Environmental Satellite (of ESA)
EO	Earth Observation
ESA	European Space Agency
EU	European Union
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FCDR	Fundamental Climate Data Record
FoM	Figure of Merit
FP	Full Physics retrieval method
FTIR	Fourier Transform InfraRed
FTS	Fourier Transform Spectrometer
GCOS	Global Climate Observing System
GEO	Group on Earth Observation
GEOSS	Global Earth Observation System of Systems
GHG	GreenHouse Gas
GOME	Global Ozone Monitoring Experiment
GMES	Global Monitoring for Environment and Security
GOSAT	Greenhouse Gases Observing Satellite
IASI	Infrared Atmospheric Sounding Interferometer

IMAP-DOAS (or IMAP)	Iterative Maximum A posteriori DOAS
IPCC	International Panel in Climate Change
IUP	Institute of Environmental Physics (IUP) of the University of Bremen, Germany
JAXA	Japan Aerospace Exploration Agency
JCGM	Joint Committee for Guides in Metrology
L1	Level 1
L2	Level 2
L3	Level 3
L4	Level 4
LMD	Laboratoire de Météorologie Dynamique
MACC	Monitoring Atmospheric Composition and Climate, EU GMES project
NA	Not applicable
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Format
NDACC	Network for the Detection of Atmospheric Composition Change
NIES	National Institute for Environmental Studies
NIR	Near Infra Red
NLIS	LMD/CNRS *neuronal* network mid/upper tropospheric CO2 and CH4 retrieval algorithm
NOAA	National Oceanic and Atmospheric Administration
Obs4MIPs	Observations for Climate Model Intercomparisons
OCO	Orbiting Carbon Observatory
OE	Optimal Estimation
PBL	Planetary Boundary Layer
ppb	Parts per billion
ppm	Parts per million
PR	(light path) PRoxy retrieval method
PVIR	Product Validation and Intercomparison Report
QA	Quality Assurance
QC	Quality Control
REQ	Requirement
RMS	Root-Mean-Square
RTM	Radiative transfer model
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric ChartographY
SCIATRAN	SCIAMACHY radiative transfer model
SRON	SRON Netherlands Institute for Space Research
SWIR	Short Wava Infra Red
TANSO	Thermal And Near infrared Sensor for carbon Observation
TANSO-FTS	Fourier Transform Spectrometer on GOSAT
TBC	To be confirmed

TBD	To be defined / to be determined
TCCON	Total Carbon Column Observing Network
TIR	Thermal Infra Red
TR	Target Requirements
TRD	Target Requirements Document
WFM-DOAS (or WFMD)	Weighting Function Modified DOAS
UoL	University of Leicester, United Kingdom
URD	User Requirements Document
WMO	World Meteorological Organization
Y2Y	Year-to-year (bias variability)

General definitions

Table 1 lists some general definitions relevant for this document.

Table 1: General definitions.

Item	Definition
XCO_2	Column-averaged dry-air mixing ratios (mole fractions) of CO_2
XCH_4	Column-averaged dry-air mixing ratios (mole fractions) of CH_4
L1	Level 1 satellite data product: geolocated radiance (spectra)
L2	Level 2 satellite-derived data product: Here: CO_2 and CH_4 information for each ground-pixel
L3	Level 3 satellite-derived data product: Here: Gridded CO_2 and CH_4 information, e.g., 5 deg times 5 deg, monthly
L4	Level 4 satellite-derived data product: Here: Surface fluxes (emission and/or uptake) of CO_2 and CH_4

Scope of document

This document is an Algorithm Theoretical Basis Document (ATBD) for the Copernicus Climate Change Service (C3S, https://climate.copernicus.eu/) component as covered by project C3S_312a_Lot6 led by University of Bremen, Germany.

Within project C3S_312a_Lot6 satellite-derived atmospheric carbon dioxide (CO_2) and methane (CH_4) Essential Climate Variable (ECV) data products will be generated and delivered to ECMWF for inclusion into the Copernicus Climate Data Store (CDS) from which users can access these data products and the corresponding documentation.

The C3S_312a_Lot 6 satellite-derived data products are:

- Column-average dry-air mixing ratios (mole fractions) of CO_2 and CH_4, denoted XCO_2 (in parts per million, ppm) and XCH_4 (in parts per billion, ppb), respectively.
- Mid/upper tropospheric mixing ratios of CO_2 (in ppm) and CH_4 (in ppb).

This document describes the retrieval algorithms ("NLIS") to generate the C3S products CO2_IASA_NLIS, CH4_IASA_NLIS, CO2_IASB_NLIS, CH4_IASB_NLIS.

These products are mid-tropospheric CO_2 and CH_4 Level 2 products as retrieved from the IASI sensors on Metop-A and Metop-B using algorithms developed at CNRS-LMD, France.

The NLIS algorithm is also used to retrieve mid-tropospheric CO_2 from AIRS (product CO2_AIR_NLIS) and a description of NLIS as applied to AIRS is also given in this document.

Executive summary

This document describes the retrieval algorithms developed at CNRS-LMD to retrieve mid-tropospheric column of CO_2 and CH_4 from the IASI and AMSU instruments flying onboard the European Metop satellites. The algorithms are based on the non-linear regression inverse radiative transfer model using Multi-Layer Perceptrons (*Crevoisier et al., 2009a,b*), that was first designed to retrieve CO_2 from the first generation TOVS instruments flying onboard the NOAA polar platforms (*Chédin et al., 2003*). This document details the various input data required for retrievals, the physical theory, and the mathematical background underlying retrieval assumptions, and also outlines the retrieval implementation and the limitations of the approach used.

1. Data product overview

Four Level 2 products derived from combined IASI/AMSU spectral data are delivered by CNRS-LMD:

- CO2_IASA_NLIS.
- CH4_IASA_NLIS.
- CO2_IASB_NLIS.
- CH4_IASB_NLIS.

Each product is delivered for both Metop-A and Metop-B.
These products are mid-troposheric columns data retrieved by the CNRS-LMD non-linear inference scheme (NLIS) algorithm, which is discussed in Section 3.

Another product (a brokered product from ESA's GHG-CCI project) is mid-troposheric CO_2 from AIRS and the corresponding algorithm is also described in this document.

2. Input and auxiliary data

2.1 Satellite instruments

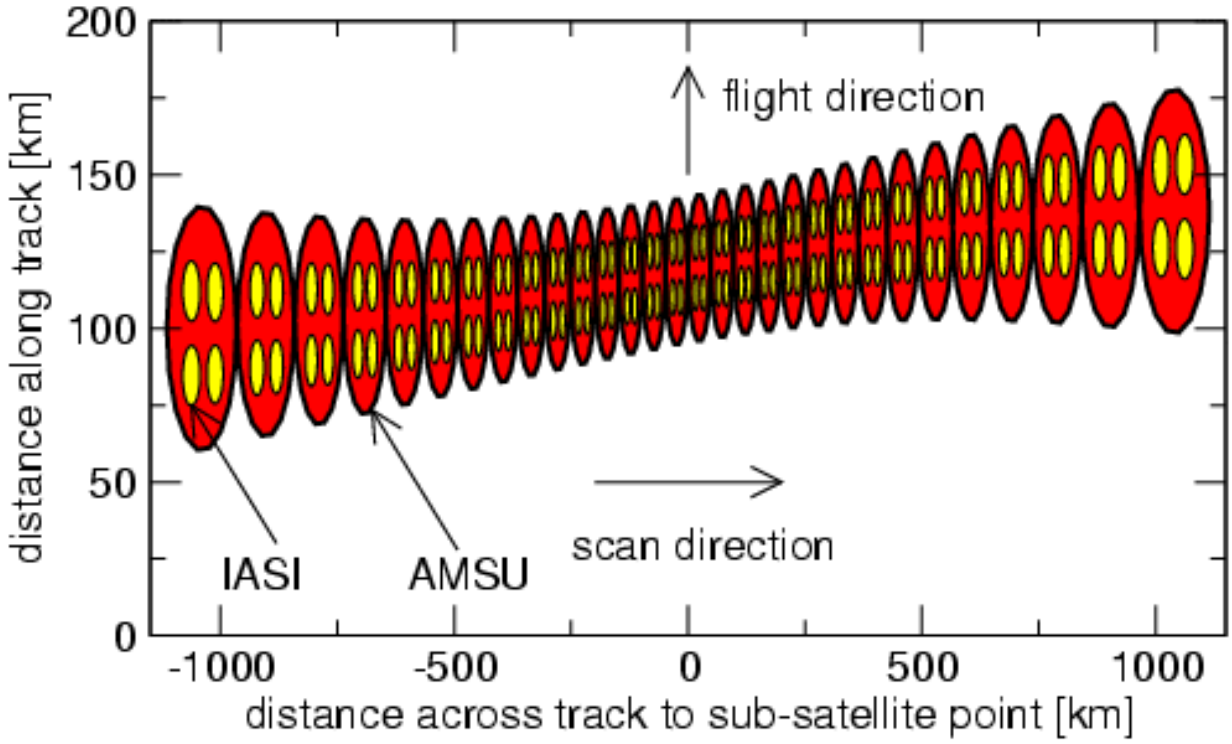
2.1.1 The IASI instrument onboard the Metop satellites

The Infrared Atmospheric Sounding Interferometer (IASI) is a high resolution Fourier Transform Spectrometer based on a Michelson Interferometer coupled to an integrated imaging system that measures infrared radiation emitted from the Earth (https://iasi.cnes.fr/en/IASI/index.htm). Developed by the Center National d'Etudes Spatiales (CNES) in collaboration with the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), IASI was launched in October 2006 onboard the polar orbiting Meteorological Operational Platform (Metop-A), and in September 2012 onboard Metop-B. A third IASI will be launched onboard Metop-C in October 2018.

IASI provides 8461 spectral samples, ranging from 645 cm^{-1} to 2760 cm^{-1} (15.5 µm and 3.6 µm), with a spectral sampling of 0.25 cm^{-1}, and a spectral resolution of 0.5 cm^{-1} after apodization. IASI is an across track scanning system, whose swath width is of 2200 km, allowing global coverage twice a day. The IFOV is sampled by 2×2 circular pixels whose ground resolution is 12 km at nadir at 9:30 am/pm local time.

The combined use of both Metop satellites, which are flying on the same orbit but with nearly half an orbit out of phase, yields a complete coverage of the Earth in one day. With the future launch of Metop-C, currently planned in 2018, these time series will cover about 20 years. In order to be useful for climate studies, it is mandatory that the time series derived from the 3 successive platforms are consistent in order to allow studying trends and growth rates.

Figure 1: IASI and AMSU scanning geometry (from IASI Level 1 Product Guide available at http://oiswww.eumetsat.org/WEBOPS/eps-pg/IASI-L1/IASIL1-PG-0TOC.htm)



2.1.2 The AMSU instrument onboard the Metop satellites

Also flying onboard Metop satellites is the AMSU-A (Advanced Microwave Sounding Unit) instrument, which is a 15-channel microwave radiometer, which measure scene radiances in 15 discreet frequency channels spanning 23-90 GHz. Thirty consecutive field of views of 48 km diameter at nadir are sampled, yielding a 2,074 km swath width. AMSU-A uses oxygen absorption bands/lines for atmospheric temperature sounding, while window channels provide information on surface temperature and emissivity.

Scanning of both sounders are synchronized, with 4 IASI fields of view (FOV) embedded in 1 AMSU FOV, allowing the same atmospheric situation to be simultaneously observed by both instruments.

2.1.3 Data product delivery

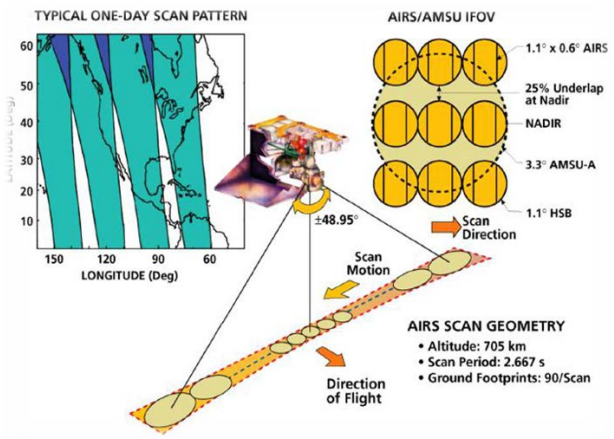
IASI and AMSU input data are Level 1c and Level 1b data respectively, disseminated in near-real time through the EUMETCast system of EUMETSAT. Metop-A data are available since July 2007 and Metop-B data since February 2013.

2.1.4 The AIRS and AMSU instruments onboard the Aqua satellite

The Atmospheric Infrared Sounder (AIRS) is a polar orbiting nadir-viewing high-resolution infrared sounder operating in a cross-track-scanning mode. It was launched onboard the EOS Aqua satellite in May 2002, with two operational microwave sounders, AMSU and HSB, and is operational since September 2002. It is a high-spectral resolution, grating multispectral infrared sounder with 2378 channels. Its spectral domain ranges from 650 cm^{-1} to 2665 cm^{-1} (15.4 µm and 3.8 µm), with a spectral resolving power of 1200 (i.e., a spectral resolution ranging from 0.5 cm^{-1} to 2 cm^{-1}). This domain is divided into three spectral bands, from 650 to 1135 cm^{-1}, from 1215 to 1615 cm^{-1} and from 2180 to 2665 cm^{-1}. AIRS cross-track scanning is 1650 km and covers 70% of the earth every day. The instantaneous field of view (IFOV) is sampled by 3×3 circular pixels whose ground resolution is 13 km at nadir. Measurements from AIRS and AMSU are analyzed jointly to filter out the effects of clouds from the IR data in order to derive various atmospheric and surface variable in clear conditions.

Scanning of both sounders are synchronized, with 9 AIRS fields of view (FOV) embedded in 1 AMSU FOV, allowing the same atmospheric situation to be simultaneously observed by both instruments (Figure 2).

Figure 2: AIRS and AMSU scan geometrics (*Jason, 2008*).



2.2 Other

2.2.1 The ARSA database

The computation of the radiative biases, which plays a critical role in the retrieval process, is based on the ARSA (Analyzed RadioSoundings Archive) database, which is available at http://ara.abct.lmd.polytechnique.fr/index.php?page=arsa. ARSA builds on observations made by worldwide distributed radiosonde stations and combines them with surface and other auxiliary observations. Physically coherent quality control tests have been developed to detect and eliminate gross errors: format problems, redundant radiosounding and levels, unrealistic jumps, physically

implausible values, temporal and vertical inconsistencies in temperature and dew point temperatures. The current ARSA database (about 6 million elements) starts in January 1979, and is extended onwards, on a monthly basis.

3. Algorithms

3.1 IASI algorithm

3.1.1 General description

Mid-tropospheric columns of methane (CH_4) and carbon dioxide (CO_2) are retrieved from simultaneous observations of the IASI and AMSU instruments flying together onboard the Metop satellites using a non-linear inference scheme using Multi-Layer Perceptrons with 2 hidden layers. IASI hyperspectral observations in the thermal infrared at 7.7 µm (resp. 15 µm), which are sensitive to both temperature and gas concentrations of CH_4 (resp. CO_2) are used in conjunction with microwave observations form the AMSU instruments, only sensitive to temperature, to decorrelate both signals.

Only a subset of channels presenting the best properties with regards to the retrieval performances are used. The neural networks are trained on a learning dataset with couples of known inputs-outputs (TIGR) and evaluated on an evaluation dataset (ARSA). The retrievals are performed during day and night-time (9:30 am/pm local time), both over land and over sea. The CO_2 retrievals are limited to the tropical region (30N:30S).

Through comparisons with regular aircraft (*Machida et al., 2008)* or balloon *(Membrive et al., 2017*) measurements as well as observations made at the surface, it has been shown that, once the radiometric characterization of the instruments is performed, IASI and AMSU capture well the trend and interannual variation of CH_4, with an excellent agreement with the rate of increase measured at the surface, giving confidence in the ability of IASI to follow its evolution over the 20 years of the Metop program.

3.1.2 Forward model and spectroscopic database

The radiative simulations in the thermal infrared performed in this study are based on the fast and accurate line-by-line radiative transfer model 4A (Automatized Atmospheric Absorption Atlas) (*Scott and Chédin, 1981*). 4A is an advanced version of the nominal line-by-line STRANSAC model (Scott, 1974) and is basically a compressed look-up-table of optical depths calculated once and for all. It can be coupled to any spectroscopic databases and can simulate any instrumental configurations (ground, airborne, satellite). In addition to the simulation of atmospheric transmissions and radiance (or equivalently brightness temperature) spectra, 4A analytically computes Jacobians for all relevant atmospheric variables. Jacobians are defined as the partial derivative of the channel brightness temperature with respect to a layer physical variable such as a gas mixing ratio, a temperature or the emissivity. Since the beginning of 2001, an operational version denoted 4A/OP

has been developed by the society Noveltis (http://www.noveltis.com/4AOP/) in collaboration with CNES and LMD. 4A is the official code chosen by CNES for calibration/validation and preparation activities of several space missions, including IASI and IASI-NG. For this study, the spectrometric parameters used as inputs to 4A, are taken form the GEISA-2011 database (*Jacquinet-Husson et al., 2011*).

3.1.3 Channel selection

IASI presents 8461 channels covering most of the infrared spectrum. Only a hundred of them are sensitive to methane, and a few hundreds to CO_2.

3.1.3.1 Channels for CH_4

IASI channels sensitive to methane are either located in band ν_4 of methane, around 7.7 µm (1306 cm^{-1}), or in band ν_3, around 3.8 µm (2630 cm^{-1}), and present various sensitivities to methane and other atmospheric or surface components. The sensitivity to methane variations of channels located in the 3.8 µm band is much lower than that of channels located in the 7.7 µm band due to weaker absorption lines: they won't be considered here. In the 7.7 µm band, channels are sensitive to water vapor (H_2O), nitrous oxide (N_2O) and surface characteristics.

The main interference, as far as CH_4 is concerned, comes from H_2O, which dominates infrared spectrum in methane absorption bands. Since water vapor variability is quite high, especially in the tropics, and knowledge of its tropospheric distribution still limited, separating CH_4 signal from water vapor is quite challenging and precludes using most of the channels. Due to much lower water vapour content in the mid-latitude regions as opposed to the tropics, the interferences between H_2O and CH_4 is less pronounced in the extra-tropical regions, giving access to more channels with a high signal-to-interference ratio. Altogether, only a few successive channels located in the 1301-1303 cm^{-1} interval present a low-enough sensitivity to water vapor to be used to retrieve methane.

24 channels have been selected to optimize the signal-to-interference ratio. They are not sensitive to variations of methane in two parts of the atmosphere: the lower troposphere (roughly below 500 hPa) and the tropopause (*Crevoisier et al., 2003*). The Jacobians of the selected channels have very similar shapes and all peak around 260 hPa. Hence, IASI only allows the retrieval of a mid-tropospheric column of methane. It is equivalent to say that IASI observations are characterized by only one degree of freedom on the vertical for CH_4.

3.1.3.2 Channels for CO_2

IASI channels sensitive to carbon dioxide are either located in band ν_2 of CO_2, around 15 µm (670 cm^{-1}), or in band ν_3, around 4.3 µm (2260 cm^{-1}), and present various sensitivities to methane and other atmospheric or surface components. The 4.3 µm band is characterized by a very high

radiometric noise that precludes using this channel for retrieving CO_2. The main interference, as far as CO_2 is concerned, comes from H_2O and ozone. Altogether, use is made of a few successive channels located in the 670 cm^{-1} interval.

These channels are not sensitive to variations of CO_2 in two parts of the atmosphere: the lower troposphere (roughly below 500 hPa) and the tropopause (*Crevoisier et al., 2003*). The Jacobians of the selected channels have very similar shapes and all peak around 200 hPa. Hence, IASI only allows the retrieval of a mid-tropospheric column of carbon dioxide. It is equivalent to say that IASI observations are characterized by only one degree of freedom on the vertical for CO_2.

3.1.4 Neural architecture

The weakness of the signal induced on IASI brightness temperature (BT) by CH_4 or CO_2 variations, associated with the complexity (in particular its non-Gaussianity) of the relationship between the gas concentration and observed BT, makes it difficult to solve this inverse problem. To tackle this problem, use is made of a non-linear inference method, based on the Multilayer Perceptron (MLP) neural network (*Rumelhart et al., 1986*) with two hidden layers. Following the selection of IASI and AMSU channels described previously, the chosen neural architectures are the following.

3.1.4.1 Architecture for CH_4

The input layer is composed of: (i) 24 IASI BT. Among them, the first 5 channels are not sensitive to CH_4 but to stratospheric temperature only. They have been included in order to deal with the slight sensitivity of the selected CH_4 IASI channels to stratospheric temperature; (ii) 2 AMSU BT of channels 6 and 8; (iii) 10 differences between IASI and AMSU BT, to help constraining the convergence process : 6-2497, 8-2497, 6-2553, 8-2553, 6-2634, 8-2634, 6- 2637, 8-2637, 6-2809, 8-2809. All together, there are 36 predictors.

The output layer of the network is composed of: (i) the difference between the true value of CH_4 concentration (associated with inputs) and the TIGR reference one (1860 ppb) ; (ii) 24 differences between the ‘‘true’’ IASI BT (associated with the true CH_4 concentration value) and the ‘‘reference’’ one (associated with the reference CH_4 concentration value), once again to constrain the solution. All together, there are 25 predictands.

Our past experience and several trials have led us to choose 70 neurons for the first hidden layer and 40 for the second one.

3.1.4.2 Architecture for CO_2

The input layer is composed of: (i) 89 IASI BT. Among them, the first 5 channels are not sensitive to CO_2 but to stratospheric temperature only. They have been included in order to deal with the slight sensitivity of the selected CO_2 IASI channels to stratospheric temperature; (ii) 2 AMSU BT of

channels 6 and 8 ; (iii) 10 differences between IASI and AMSU BT, to help constraining the convergence process. All together, there are 101 predictors.

The output layer of the network is composed of: (i) the difference between the true value of CO_2 concentration (associated with inputs) and the TIGR reference one (372 ppm) ; (ii) 24 differences between the ‘‘true’’ IASI BT (associated with the true CO_2 concentration value) and the ‘‘reference’’ one (associated with the reference CO_2 concentration value), once again to constrain the solution. All together, there are 25 predictands.

Our past experience and several trials have led us to choose 70 neurons for the first hidden layer and 40 for the second one.

3.1.5 Training of the networks

The learning algorithm is the optimization technique that estimates the optimal network parameters by minimizing a positive-definite cost function which measures, for a set of representative situations for which inputs (here the brightness temperatures) and outputs (gas) are known (the learning set), the mismatch between the neural network outputs and the desired outputs. Here, the Error Back-Propagation algorithm (*Rumelhart et al., 1986*) is used to minimize the cost function. It is a gradient descent algorithm well adapted to the MLP hierarchical architecture because the computational cost is linearly related to the number of parameters. To avoid being trapped in local minima during the minimization of the cost function, stochastic steepest descent is used. The learning step is made sample by sample, chosen iteratively and stochastically in the learning data set.

The training database from which the networks learn the relationship existing between inputs and outputs is based on the TIGR database. For all TIGR atmospheric situations, for all scan angles, and for the whole IASI channels used in the retrieval process, clear-sky brightness temperatures (BT), transmittances and Jacobians have been computed using the 4A/OP radiative transfer model with the spectroscopic database GEISA-2011 as input. The required AMSU BTs are computed using the STRANSAC microwave forward model. Network input BTs correspond to randomly drawn values of CH_4 (resp. CO_2) concentration in the range 1760-1960 ppb (resp. 362-382 ppm), centered on the TIGR reference value of 1860 ppb (resp. 372 ppm); they are computed using the stored CH_4 (resp. CO_2) Jacobians. It is worth noting that no prior information is thus given to the networks in terms of seasonality, trend, or geographical patterns of the gases.

Neural networks are trained for each of the 15 AMSU scan angles and for 2 air-masses (tropical or mid-latitude) independently. Surface elevation is also taken into account. All together, for a given neural architecture, 240 networks are trained. For each network corresponding to one air-mass, one scan angle and one surface type, the learning steps are the following:

1) One atmosphere is randomly chosen among the TIGR atmospheres of the considered air-mass.

2) A CH_4 (resp. CO_2) mixing ratio is drawn randomly (uniform distribution) in the range 1760-1960 ppb (resp. 362-382 ppm), which is centered on the reference CH_4 (resp. CO_2) mixing ratio of TIGR.
3) A perturbation of the surface temperature is randomly chosen according to the normal distribution, with a null mean value and a standard deviation of 4 K.
4) The input BTs at the drawn CH_4 (resp. CO_2) mixing ratio are computed using BTs and CH_4 (resp. CO_2) Jacobians from TIGR for the considered atmosphere.
5) For IASI channels, noise equivalent temperatures are computed at the BT according to Eq. 1:

$$NE\Delta T\left[T_B(\nu),\nu\right] = NE\Delta T\left[T_{ref},\nu\right]\frac{\frac{\partial B}{\partial T}\left(T_{ref},\nu\right)}{\frac{\partial B}{\partial T}\left[T_B(\nu),\nu\right]} \quad (1)$$

where $NE\Delta T$ is the equivalent noise temperature taken at the brightness temperature T_B, of the channel located at frequency ν, and B is the radiance. The reference noise corresponding to a reference temperature T_{ref} of 280 K is taken from the in-flight noise measurement (CNES, priv. comm.). To increase the signal to noise ratio, and speed the learning phase, these noises have been divided by 2. Since 4 IASI spots are localized within one AMSU spot, the average of IASI BT contained in a single AMSU field-of-view are therefore used as inputs to the networks.
6) The quadratic sum of the instrument noise and the forward radiative transfer model noise, are computed and added to the BT.
7) The inputs and outputs are normalized in order to homogenise the input values between 0 and 1.
8) The Error Back-Propagation algorithm (*Rumelhart et al., 1986*) is used to minimize the cost function.
9) The parameters of the networks are updated.
10) The networks are applied to the whole ARSA atmospheres following the same procedure as TIGR (steps 4 to 7) and the root mean square (RMS) error of the output is computed.
11) Come back to step 1, until the predefined number of iterations has been reached.

3.1.6 Application to observations

3.1.6.1 General description

Once the learning phase is completed, observations of IASI and AMSU can be used to infer mid-tropospheric columns of CH_4 or CO_2. The retrieval is performed at the AMSU resolution: when 4 IASI FOVs included in 1 AMSU FOV are declared clear (meaning that no cloud nor aerosol has been detected), the BTs of the channels are averaged over the 4 IASI FOVs and used together with AMSU BTs, to perform the retrieval.

Since the networks are trained with simulated data, potential systematic radiative biases existing between simulations used in the learning phase and observations must be removed before using these BTs as inputs to the network corresponding to the situation according to the scan angle, surface elevation and air-mass type. These systematic radiative biases are computed with the

calibration/validation chain that has been developed for many years at LMD (*Armante et al., 2016*). For each channel, the differences between simulations and collocated (in time and space) satellite observations are averaged over several full years of operation. These differences are called 'calc-obs' residuals. The simulations are performed using the 4A/OP forward model and radiosonde measurements from ARSA as inputs. One key element is that, during this computation, the CH_4 and CO_2 mixing ratios are kept at the reference value of the TIGR databse (1860 ppb and 372 ppm respectively) to avoid making the CH_4 nor CO_2 signals disappearing in the BT used as input to the networks. Every month, about 100 collocations are available, giving access to robust statistics.

By averaging thousands of situations together, it is possible to derive the evolution of the averaged biases with the scan angle. In order to avoid potential biases due to the incorrect modelling of the radiative effect of scan angles close to the edges of the orbit, which is particularly the case for microwave observation, specific radiative biases have to be taken into account for each scan angle.

3.1.6.2 Case of Metop-B

On August, 2nd 2017, the on-board processing of IASI/Metop-B changed, due to the update of the correction of the non-linearity of detectors. This change had a noticeable impact on IASI spectral band 1 where the channels used to retrieve CO_2 are located. Typically, it induced a 0.2 K bias in the measured brightness temperatures, which correspond to a bias of about 5 ppm. To deal with this problem, a new set of radiative biases have been computed for IASI/Metop-B. The radiative change after – before August 2nd has been computed by comparing the 'calc-obs' residuals computed over 1 month after and before the change. Atmospheric situations were taken from ECMWF analyses collocated in time and space with IASI/Metop-B. Depending on the date, a different set of radiative biases must now be considered before and after August 2nd for the processing of IASI/Metop-B.

3.1.7 Vertical characterization of the retrieval

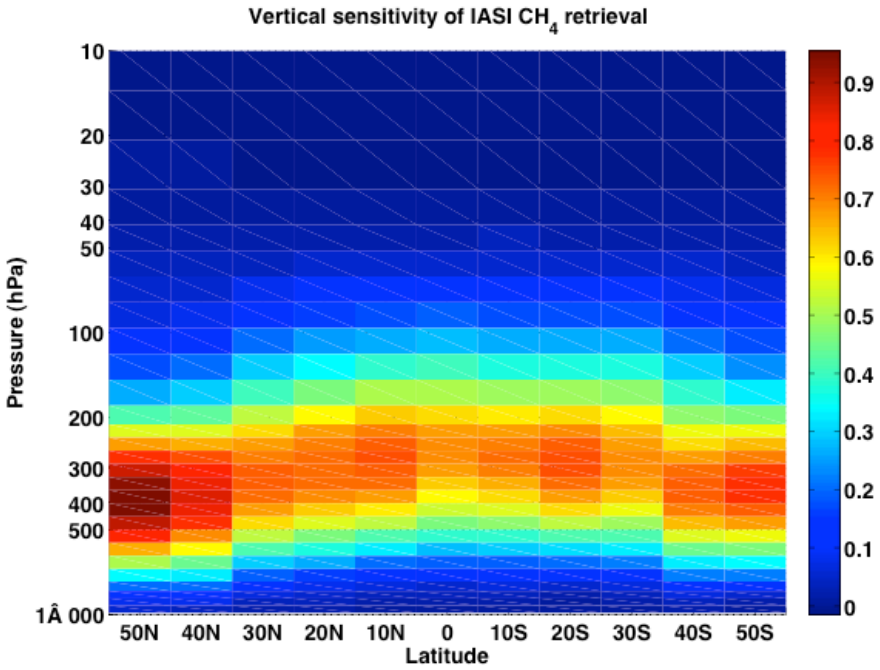
As stated before, IASI channels located in the 7.7 µm and 15 µm bands are mostly sensitive to tropospheric variations of gases. The averaging kernels, which indicate which part of the atmosphere the retrievals are representative of, are determined through radiative transfer simulations. A uniform perturbation of CH_4 or CO_2 mixing ratio is applied sequentially to each of the 40 pressure layers used in ARSA to characterize atmospheric profiles. IASI and AMSU brightness temperatures are then computed for each of the perturbed atmospheric profiles and used as inputs to the neural networks. The theoretical change F_i in ppbv/ppbv of the column mean apparent mixing ratio ($\hat{q}$) given a mixing ratio perturbation of dq^{ref} at level i, is then given by

$$F_i = \frac{\hat{q}\left(\Delta q_i = dq^{ref}\right) - \hat{q}\ \left(\Delta q_i = 0\right)}{dq^{ref}} \qquad (2)$$

The mean of the averaging kernel for CH_4 computed over the ARSA dataset is plotted in Fig. 3. In the tropics, the height of the tropopause is approximately 17 km, whereas it is closer to 8km in the mid-latitude. The non linear inference scheme gives access to a mid-to-upper tropospheric integrated

content of CH_4 covering: (i) the range 100-500 hPa (roughly 9-15 km), with the highest sensitivity around 230 hPa in the tropics ; (ii) the range 250-700 hPa (roughly 6-12 km), with the highest sensitivity around 400 hPa in the mid-latitudes.

Figure 3: Vertical sensitivity of IASI CH_4 retrievals as a function of latitude.



3.2 Algorithm for AIRS mid-tropospheric CO_2 retrieval

3.2.1 General description

Mid-tropospheric columns of carbon dioxide (CO_2) are retrieved from simultaneous observations of the AIRS and AMSU instruments flying together onboard the Aqua satellite using a first version of the non-linear inference scheme used for processing IASI and AMSU data as described in Section 3.1. This inference scheme is based on Multi-Layer Perceptrons with 2 hidden layers. AIRS hyperspectral observations in the thermal infrared at 15 and 4.3 µm, which are sensitive to both temperature and gas concentrations of CO_2, are used in conjunction with microwave observations form the AMSU instruments, only sensitive to temperature, to decorrelate both signals.

Only a subset of channels presenting the best properties with regards to the retrieval performances are used. The neural networks are trained on a learning dataset with couples of known inputs-outputs (TIGR). As opposed to the processing of IASI and AMSU observations from the Metop series, no continuous evaluation of the neural networks performance throughout the training has been performed. The retrievals are performed during day and night-time (1:30 am/pm local time), both over land and over sea. The CO_2 retrievals are limited to the tropical region (25N:25S).

Through comparisons with regular aircraft (*Machida et al., 2008)* measurements, it has been shown that, once the radiometric characterization of the instruments is performed, AIRS and AMSU capture well the trend and interannual variation of CO_2, until the loss of some of channels used in the retrieval process in July 2007. The retrieval of CO_2 from AIRS has been stopped ever since.

3.2.2 Forward model and spectroscopic database

The radiative simulations in the thermal infrared performed in the retrieval process are based on the fast and accurate line-by-line radiative transfer model 4A (Automatized Atmospheric Absorption Atlas) (*Scott and Chédin, 1981*). 4A is an advanced version of the nominal line-by-line STRANSAC model (Scott, 1974) and is basically a compressed look-up-table of optical depths calculated once and for all. It can be coupled to any spectroscopic databases and can simulate any instrumental configurations (ground, airborne, satellite). In addition to the simulation of atmospheric transmissions and radiance (or equivalently brightness temperature) spectra, 4A analytically computes Jacobians for all relevant atmospheric variables. Jacobians are defined as the partial derivative of the channel brightness temperature with respect to a layer physical variable such as a gas mixing ratio, a temperature or the emissivity. Since the beginning of 2001, an operational version denoted 4A/OP has been developed by the society Noveltis (http://www.noveltis.com/4AOP/) in collaboration with CNES and LMD. 4A is the official code chosen by CNES for calibration/validation and preparation activities of several space missions, including IASI and IASI-NG. For generating AIRS CO_2, the spectrometric parameters used as inputs to 4A, are taken form the GEISA-2008 database (*Jacquinet-Husson et al., 2009*).

3.2.3 Channel selection

AIRS presents 2378 channels covering most of the infrared spectrum. Only a few hundred of them are sensitive to CO_2. They are either located in band ν_2 of CO_2, around 15 µm (670 cm^{-1}), or in band ν_3, around 4.3 µm (2260 cm^{-1}), and present various sensitivities to CO_2 and other atmospheric or surface components. As opposed to IASI, the AIRS channels located in the 4.3 µm band are characterized by a radiometric noise of the same order as for the channels located at 15µm, which means that both CO_2 absorption bands are used to perform the retrieval. As will be detailed in Section 3.2.7, the use of channels located at 4.3 µm yields a sensitivity to CO_2 at lower altitudes than for IASI. The main interference, as far as CO_2 is concerned, comes from H_2O and ozone.

A set of AIRS and AMSU channels presenting optimal characteristics to estimate CO_2 has been selected in *Crevoisier et al. (2003)* prior to the launch of the Aqua satellite. Based on increasing experience with the instruments, the channel selection has been refined, based on three criteria: (1) their sensitivity to CO_2; (2) their sensitivity to other atmospheric components, as well as to surface characteristics; and (3) to their sensitivity to variation of CO_2 along the vertical. A set of 15 channels has been selected; 9 of them are located in the 15 µm band (channels 173, 175, 180, 185, 193, 213, 218 and 250), the other being located in the 4.3 µm band.

These channels are not sensitive to variations of CO_2 in two parts of the atmosphere: the lower troposphere (roughly below 500 hPa) and the tropopause (*Crevoisier et al., 2003*). The Jacobians of the selected channels have very similar shapes and all peak around 200-300 hPa. Hence, AIRS only allows the retrieval of a mid-tropospheric column of carbon dioxide. It is equivalent to say that AIRS observations are characterized by only one degree of freedom on the vertical for CO_2.

3.2.4 Neural architecture

The weakness of the signal induced on AIRS brightness temperature (BT) by CO_2 variations, associated with the complexity (in particular its non-Gaussianity) of the relationship between the gas concentration and observed BT, makes it difficult to solve this inverse problem. To tackle this problem, use is made of a non-linear inference method, based on the Multilayer Perceptron (MLP) neural network (*Rumelhart et al., 1986*) with two hidden layers. Following the selection of AIRS and AMSU channels described previously, the chosen neural architectures are the following.

The input layer is composed of: (i) 15 AIRS BT; (ii) 2 AMSU BT of channels 6 and 8 ; (iii) 15 differences between the 15 AIRS channels and AMSU 6 BT, to help constraining the convergence process. All together, there are 32 predictors.

The output layer of the network is composed of: (i) the difference between the true value of CO_2 concentration (associated with inputs) and the TIGR reference one (372 ppm) ; (ii) 24 differences between the ‘‘true’’ AIRS BT (associated with the true CO_2 concentration value) and the ‘‘reference’’ one (associated with the reference CO_2 concentration value), once again to constrain the solution.

All together, there are 16 predictands. The first and second hidden layers are made of 70 and 40 neurons, respectively.

3.2.5 Training of the networks

The learning algorithm is the optimization technique that estimates the optimal network parameters by minimizing a positive-definite cost function which measures, for a set of representative situations for which inputs (here the brightness temperatures) and outputs (gas) are known (the learning set), the mismatch between the neural network outputs and the desired outputs. Here, the Error Back-Propagation algorithm (*Rumelhart et al., 1986*) is used to minimize the cost function. It is a gradient descent algorithm well adapted to the MLP hierarchical architecture because the computational cost is linearly related to the number of parameters. To avoid being trapped in local minima during the minimization of the cost function, stochastic steepest descent is used. The learning step is made sample by sample, chosen iteratively and stochastically in the learning data set.

The training database from which the networks learn the relationship existing between inputs and outputs is based on the TIGR database. For all TIGR atmospheric situations, for all scan angles, and for the whole IASI channels used in the retrieval process, clear-sky brightness temperatures (BT), transmittances and Jacobians have been computed using the 4A/OP radiative transfer model with the spectroscopic database GEISA-2011 as input. The required AMSU BTs are computed using the STRANSAC microwave forward model. Network input BTs correspond to randomly drawn values of CO_2 concentration among the values [362:4:382] ppm, centered on the TIGR reference value of 372 ppm; they are computed using the stored CO_2 Jacobians. It is worth noting that no prior information is thus given to the networks in terms of seasonality, trend, or geographical patterns of the gases.

Neural networks are trained for each of the 10 first AMSU scan angles and for tropical situations, independently. Surface elevation is also taken into account. All together, for a given neural architecture, 80 networks are trained. For each network corresponding to one air-mass, one scan angle and one surface type, the learning steps are the following:

1) One atmosphere is randomly chosen among the TIGR atmospheres of the considered air-mass.
2) A CO_2 mixing ratio is drawn randomly among the values [362:4:382] ppm, centered on the reference CO_2 mixing ratio of TIGR.
3) A perturbation of the surface temperature is randomly chosen according to the normal distribution, with a null mean value and a standard deviation of 4 K.
4) The input BTs at the drawn CO_2 mixing ratio are computed using BTs and CO_2 Jacobians from TIGR for the considered atmosphere.
5) For AIRS channels, noise equivalent temperatures are computed at the BT according to Eq. 1:

$$NE\Delta T\left[T_B(\nu),\nu\right] = NE\Delta T\left[T_{ref},\nu\right]\frac{\frac{\partial B}{\partial T}\left(T_{ref},\nu\right)}{\frac{\partial B}{\partial T}\left[T_B(\nu),\nu\right]} \quad (1)$$

   where $NE\Delta T$ is the equivalent noise temperature taken at the brightness temperature T_B, of the channel located at frequency ν, and B is the radiance. The reference noise corresponding to a reference temperature T_{ref} of 280 K is taken from the in-flight noise measurement. To increase the signal to noise ratio, and speed the learning phase, these noises have been divided by 3. Since 9 IASI spots are localized within one AMSU spot, the average of IASI BT contained in a single AMSU field-of-view are therefore used as inputs to the networks.
6) The quadratic sum of the instrument noise and the forward radiative transfer model noise, are computed and added to the BT.
7) The inputs and outputs are normalized in order to homogenise the input values between 0 and 1.
8) The Error Back-Propagation algorithm (*Rumelhart et al., 1986*) is used to minimize the cost function.
9) The parameters of the networks are updated.
10) Come back to step 1, until the predefined number of iterations has been reached.

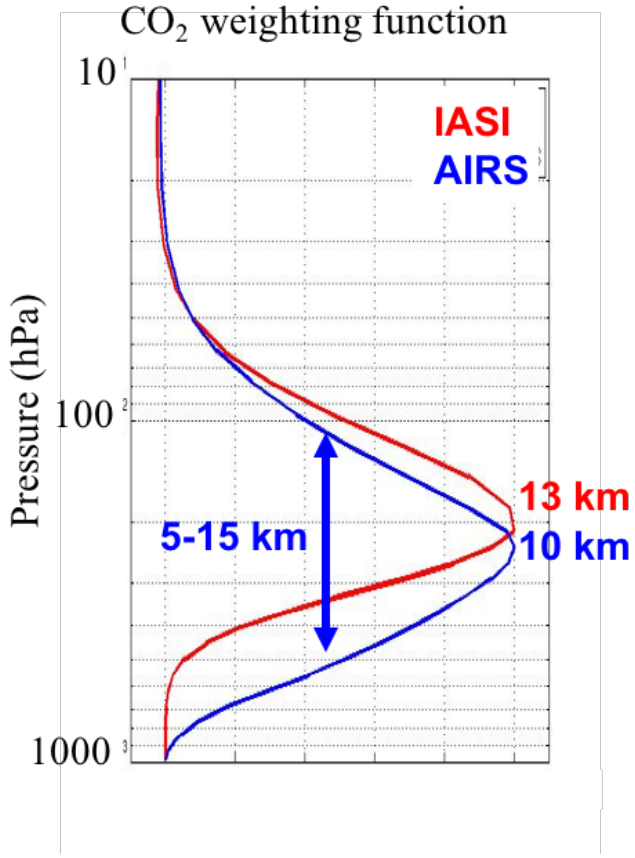
3.2.6 Application to observations

Once the learning phase is completed, observations of AIRS and AMSU can be used to infer mid-tropospheric columns of CO_2. The retrieval is performed at the AMSU resolution: when 9 AIRS FOVs included in 1 AMSU FOV are declared clear (meaning that no cloud nor aerosol has been detected), the BTs of the channels are averaged over the 9 AIRS FOVs and used together with AMSU BTs, to perform the retrieval.

Since the networks are trained with simulated data, potential systematic radiative biases existing between simulations used in the learning phase and observations must be removed before using these BTs as inputs to the network corresponding to the situation according to the scan angle,

surface elevation and air-mass type. These systematic radiative biases are computed with one of the first version of the calibration/validation chain that has been developed for many years at LMD. For each channel, the differences between simulations and collocated (in time and space) satellite observations were averaged over 5 years of operation. These differences are called 'calc-obs' residuals. The simulations were performed using the 4A/OP forward model and radiosonde measurements from the ECMWF database as inputs. One key element is that, during this computation, CO_2 mixing ratio is kept at the reference value of the TIGR databse (372 ppm) to avoid making the CO_2 signal disappear in the BT used as input to the networks. Every month, about 100 collocations were available, giving access to robust statistics. No scan-angle was applied to the biases, restricting the retrieval to the 10 smallest AMSU scan angle.

Figure 4: AIRS and IASI CO_2 averaging kernel computed over the TIGR tropical atmospheric situations.



3.2.7 Vertical characterization of the retrieval

As stated before, AIRS channels located in the 15 and 4.3 µm bands are mostly sensitive to tropospheric variations of gases. The averaging kernels, which indicate which part of the atmosphere the retrievals are representative of, are determined through radiative transfer simulations. A uniform perturbation of CO_2 mixing ratio is applied sequentially to each of the 40 pressure layers used in TIGR to characterize atmospheric profiles. AIRS and AMSU brightness temperatures are then computed for each of the perturbed atmospheric profiles and used as inputs to the neural networks. The theoretical change F_i in ppm/ppm of the column mean apparent mixing ratio ($\hat{q}$) given a mixing ratio perturbation of dq^{ref} at level *i*, is then given (*Crevoisier, 2004*) by

$$F_i = \frac{\hat{q}\left(\Delta q_i = dq^{ref}\right) - \hat{q}\ \left(\Delta q_i = 0\right)}{dq^{ref}} \qquad (2)$$

The mean of the averaging kernel for CO_2 computed over the TIGR dataset is plotted in Fig. 4. The non linear inference scheme gives access to a mid-to-upper tropospheric integrated content of CO_2 covering the range 100-600 hPa (roughly 5-15 km), with the highest sensitivity around 250 hPa in the tropics. The sensitivity to lower atmospheric layers for AIRS than for IASI comes from the use of AIRS channels located at 4.3. µm, which are sensitive to higher pressure than channels located at 15 µm, the only ones used for IASI.

4. Output data

The retrieval outputs are provided as daily netCDF files. Additional information, such as quality flags, averaging kernels, and geolocation information are also recorded in these files.

Note that the format of the main output data, which are the Level 2 data products, is described in the separate Product User Guide and Specification (PUGS) document.

References

Buchwitz et al., 2015: Buchwitz, M., Reuter, M., Schneising, O., Boesch, H., Guerlet, S., Dils, B., Aben, I., Armante, R., Bergamaschi, P., Blumenstock, T., Bovensmann, H., Brunner, D., Buchmann, B., Burrows, J.P., Butz, A., Chédin, A., Chevallier, F., Crevoisier, C.D., Deutscher, N.M., Frankenberg, C., Hase, F., Hasekamp, O.P., Heymann, J., Kaminski, T., Laeng, A., Lichtenberg, G., De Mazière, M., Noël, S., Notholt, J., Orphal, J., Popp, C., Parker, R., Scholze, M., Sussmann, R., Stiller, G.P., Warneke, T., Zehner, C., Bril, A., Crisp, D., Griffith, D.W.T., Kuze, A., O'Dell, C., Oshchepkov, S., Sherlock, V., Suto, H., Wennberg, P., Wunch, D., Yokota, T., Yoshida, Y., The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparison and quality assessment of near-surface-sensitive satellite-derived CO2 and CH4 global data sets. Remote Sens. Environ. 162:344–362, http://dx.doi.org/10.1016/j.rse.2013.04.024, 2015.

Chédin et al. 2003: Chédin, A., Saunders, R., Hollingsworth, A., Scott, N. A., Matricardi, M., Etcheto, J., Clerbaux, C., Armante, R. and Crevoisier, C.: The feasibility of monitoring CO_2 from high resolution infrared sounders. J. Geophys. Res., 108, ACH 6-1–6-19, doi: 10.1029/2001JD001443, 2003.

Cressot et al., 2014: Cressot, C., F. Chevallier, P. Bousquet, et al., On the consistency between global and regional methane emissions inferred from SCIAMACHY, TANSO-FTS, IASI and surface measurements, Atmos. Chem. Phys., 14, 577-592, 2014.

Crevoisier et al., 2003: Crevoisier, C., Chedin, A. and Scott, N.A., AIRS channel selection for CO2 and other trace-gas retrievals 129, 2719–2740. doi:10.1256/qj.02.180, 2003.

Crevoisier, 2004: Crevoisier, C., Etude de la distribution du CO_2 atmosphérique à partir des observations infrarouges à haute résolution spectrale de l'instrument Aqua/AIRS, Thèse de doctorat de l'Université Denis Diderot (Paris VII), 15 octobre 2004.

Crevoisier et al., 2004: Crevoisier, C., S. Heilliette, A. Chédin, S. Serrar, R. Armante, and N. A. Scott, Midtropospheric CO_2 concentration retrieval from AIRS observations in the tropics, Geophys. Res. Lett., 31, L17106, doi:10.1029/2004GL020141, 2004.

Crevoisier et al., 2009: Crevoisier, C., Chédin, A., Matsueda, H., et al., First year of upper tropospheric integrated content of CO_2 from IASI hyperspectral infrared observations, *Atmos. Chem. Phys.*, 9, 4797-4810, 2009.

Crevoisier et al. 2009b: Crevoisier, C., Nobileau, D., Fiore, A., Armante, R., Chédin, A., and Scott, N. A.: Tropospheric methane in the tropics – first year from IASI hyperspectral infrared observations, Atmos. Chem. Phys., 9, 6337–6350, doi:10.5194/acp-9-6337-2009, 2009b.

Crevoisier et al., 2013: Crevoisier, C., Nobileau, D., Armante, R., et al., The 2007–2011 evolution of tropical methane in the mid-troposphere as seen from space by MetOp-A/IASI, *Atmos. Chem. Phys.*, 13, 4279-4289, 2013.

ESA-CCI-GHG-URDv2.1: Chevallier, F., et al., User Requirements Document (URD), ESA Climate Change Initiative (CCI) GHG-CCI project, Version 2.1, 19 Oct 2016, link: http://www.esa-ghg-cci.org/?q=webfm_send/344, 2016.

GCOS-154: 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 (2010 update)", Prepared by World Meteorological Organization (WMO), Intergovernmental Oceanographic Commission, United Nations Environment Programme (UNEP), International Council for Science, Doc.: GCOS 154, ink: https://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf, 2010.

GCOS-200: The Global Observing System for Climate: Implementation Needs, World Meteorological Organization (WMO), GCOS-200 (GOOS-214), pp. 325, link: http://unfccc.int/files/science/workstreams/systematic_observation/application/pdf/gcos_ip_10oct2016.pdf, 2016.

Jacquinet-Husson et al., 2008: Jacquinet-Husson, N., Scott, N. A., Chédin, A., Crépeau, L., Armante, R., Capelle V., and 47 co-authors, The GEISA spectroscopic database: Current and future archive for Earth and planetary atmosphere studies, J. Quant. Spectrosc. Radiat. Transfer, 109, 1043–1059, 2008.

Jacquinet-Husson et al., 2011: Jacquinet-Husson, N., Crepeau, L., Armante, R., et al. (2011). The 2009 edition of the GEISA spectroscopic database. J. Quant. Spectrosc. Radiat. Transf. 112, 2395–2445. doi:10.1016/j.jqsrt.2011.06.004.

Jason, 2008: Jason, L., README Document for AIRS Level-2 Version 005 Standard Products. In Goddard Earth Sciences Data And Information Services Center (Ed., National Aeronautics and Space Administration (NASA), 2008.

Machida et al. 2008: Machida, T., Matsueda, H., Sawa, Y., Nakagawa, Y., Hirotani, K., Kondo, N., Goto, K., Nakazawa, T., Ishikawa, K., and Ogawa, T.: Worldwide measurements of atmospheric CO_2 and other trace gas species using commercial airlines, J. Atmos. Ocean. Tech., 25(10), 1744–1754, doi:10.1175/2008JTECHA1082.1, 2008.

Matsueda et al. 2008: Matsueda, H., Machida, T., Sawa, Y., Nakagawa, Y., Hirotani, K., Ikeda, H., Kondo, N., and Goto, K.: Evaluation of atmospheric CO_2 measurements from new flask air sampling of JAL airliner observation, Pap. Meteorol. Geophys., 59, 1–17, 2008.

Membrive et al. 2016: Membrive, O., Crevoisier, C., Sweeney, C., Danis, F., Hertzog, A., Engel, A., Bönisch, H., and Picon, L.: AirCore-HR: A high resolution column sampling to enhance the vertical description of CH_4 and CO_2, Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-236, 2016.

Rumelhart et al., 1986: Rumelhart, D.E. and McClelland, J.L. (1986) University of California, S.D.P.R.G. Parallel distributed processing : explorations in the microstructure of cognition, Parallel distributed processing: explorations in the microstructure of cognition, vol. 1. MIT Press.

Scott, 1974: Scott, N.A., A direct method of computation of the transmission function of an inhomogeneous gaseous medium— I: Description of the method. J. Quant. Spectrosc. Radiat. Transf. 14, 691–704. doi:10.1016/0022-4073(74)90116-2, 1974.

Scott and Chédin, 1981: Scott, N.A. and Chédin, A., A Fast Line-by-Line Method for Atmospheric Absorption Computations: The Automatized Atmospheric Absorption Atlas. J. Appl. Meteorol. 20, 802–812. doi:10.1175/1520-0450(1981)020<0802:AFLBLM>2.0.CO;2, 1981.

TRD GHG, 2017: Buchwitz, M., Aben, I., Anand, J., Armante, R., Boesch, H., Crevoisier, C., Detmers, R. G., Hasekamp, O. P., Reuter, M., Schneising-Weigel, O., Target Requirement Document, Copernicus Climate Change Service (C3S) project on satellite-derived Essential Climate Variable (ECV) Greenhouse Gases (CO_2 and CH_4) data products (project C3S_312a_Lot6), Version 1, 28-March-2017, pp. 52, 2017.

Copernicus Climate Change Service

ECMWF - Shinfield Park, Reading RG2 9AX, UK

Contact: info@copernicus-climate.eu

climate.copernicus.eu copernicus.eu ecmwf.int