



Algorithm Theoretical Basis Document (ATBD) – ANNEX D for products XCO2_EMMA, XCH4_EMMA, XCO2_OBS4MIPS, XCH4_OBS4MIPS (v4.5, CDR7, 2003-2022)

C3S2_312a_Lot2_DLR – Atmosphere

Issued by: Maximilian Reuter, Michael Buchwitz, University of Bremen,
Institute of Environmental Physics (IUP)

Date: 01/02/2024

Ref: C3S2_312a_Lot2_D-WP1_ATBD-2023-GHG_ANNEX-D_v7.1b

Official reference number service contract: 2021/C3S2_312a_Lot2_DLR/SC1





This document has been produced in the context of the Copernicus Climate Change Service (C3S). The activities leading to these results have been contracted by the European Centre for Medium-Range Weather Forecasts, operator of C3S on behalf on the European Union (Contribution Agreement signed on 22/07/2021). All information in this document is provided “as is” and no guarantee of warranty is given that the information is fit for any particular purpose.

The users thereof use the information at their sole risk and liability. For the avoidance of all doubt, the European Commission and the European Centre for Medium-Range Weather Forecasts have no liability in respect of this document, which is merely representing the author’s view.



Contributors

**INSTITUTE OF ENVIRONMENTAL PHYSICS (IUP),
UNIVERSITY OF BREMEN, BREMEN, GERMANY
(IUP)**

M. Reuter

M. Buchwitz

O. Schneising-Weigel



History of modifications

| Version | Date | Description of modification | Chapters / Sections |
|---------------|-------------------|--|---------------------|
| 1.1 | 20-October-2017 | New document for data set CDR1 (2003-2016) | All |
| 2.0 | 4-October-2018 | Update for CDR2 (2003-2017) | All |
| 3.0 | 12-August-2019 | Update for CDR3 (2003-2018) | All |
| 3.1 | 03-November-2019 | Update after review by Assimila | All |
| 4.0 | 18-August-2020 | Update for CDR4 (2003-2019) | All |
| 5.0 | 18-February-2021 | Update for CDR5 (01/2003-06/2020) | All |
| 6.0 Draft | 18-February-2022 | Update for data set CDR6: First draft | All |
| 6.0 | 4-August-2022 | Update for data set CDR6 (temporal coverage: 01.2003-12.2021) | All |
| 6.1 | 25-November-2022 | Update after review (use of new template, various mostly minor improvements at several places) | All |
| 6.2 | 31-January-2023 | Update after 2 nd review. Several improvements at various places. | All |
| 7.0 (Draft 1) | 23-February-2023 | Initial draft for data set CDR7 | All |
| 7.0 (Draft 2) | 27-April-2023 | Improvements based on review | All |
| 7.0 (Draft 3) | 29-June-2023 | Improvements based on review | All |
| 7.0 (Draft 4) | 03-August-2023 | Improvements based on review | All |
| 7.0 | 28-August-2023 | Update for data set CDR7 | All |
| 7.1 | 13-September-2023 | Minor improvements | All |
| 7.1b | 1-February-2024 | Flow diagram added after review | Sect. 3.1 |



List of datasets covered by this document

| Deliverable ID | Product title | Product type (CDR, ICDR) | Version number | Delivery date |
|-----------------|---|--------------------------|----------------|---------------|
| WP2-FDDP-GHG-v2 | XCO2_EMMA, XCH4_EMMA, XCO2_OBS4MIPS, XCH4_OBS4MIPS | CDR 7 | 4.5 | 31-Aug-2023 |

Related documents

| Reference ID | Document |
|--------------|---|
| D1 | <p>Main ATBD:</p> <p>Buchwitz et al., 2023c: Buchwitz, M., et al., Algorithm Theoretical Basis Document (ATBD) – Main document for Greenhouse Gas (GHG: CO₂ & CH₄) data set CDR 7 (2003-2022), project C3S2_312a_Lot2_DLR – Atmosphere, v7.1, 2023.</p> <p><i>(this document is an ANNEX to the Main ATBD)</i></p> |
| D2 | <p>Reuter et al., 2023a: Product Quality Assessment Report (PQAR) – ANNEX D for products XCO2_EMMA, XCH4_EMMA, XCO2_OBS4MIPS, XCH4_OBS4MIPS, project C3S2_312a_Lot2_DLR – Atmosphere, 2023.</p> |
| D3 | <p>Buchwitz, et al., 2023b: Product User Guide and Specification (PUGS) – Main document for Greenhouse Gas (GHG: CO₂ & CH₄) data set CDR 6 (2003-2021), project C3S2_312a_Lot2_DLR – Atmosphere, 2023.</p> |
| D4 | <p>Reuter et al., 2023b: Product User Guide and Specification (PUGS) – ANNEX D for products XCO2_EMMA, XCH4_EMMA, XCO2_OBS4MIPS, XCH4_OBS4MIPS (v4.4, 01/2003-12/2021) C3S_312a_Lot2_DLR – Atmosphere, 2023.</p> |



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 | Database |
| 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 |
| GOS | GOSAT |
| GO2 | GOSAT-2 |
| GOME | Global Ozone Monitoring Experiment |
| GMES | Global Monitoring for Environment and Security |
| GOSAT | Greenhouse Gases Observing Satellite |
| GOSAT-2 | Greenhouse Gases Observing Satellite 2 |
| 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 <i>neuronal</i> network mid/upper tropospheric CO ₂ and CH ₄ retrieval algorithm |
| NOAA | National Oceanic and Atmospheric Administration |



| | |
|-------------|---|
| Obs4MIPs | Observations for Climate Model Intercomparisons |
| OCFP | OCO-2 Full Physics (FP) algorithm (used by Univ. Leicester) |
| OCO | Orbiting Carbon Observatory |
| OCPR | OCO-2 Proxy (PR) algorithm (used by Univ. Leicester) |
| OE | Optimal Estimation |
| PBL | Planetary Boundary Layer |
| ppb | Parts per billion |
| ppm | Parts per million |
| PQAD | Product Quality Assurance Document |
| PQAR | Product Quality Assessment Report |
| PR | (light path) PROxy retrieval method |
| PVIR | Product Validation and Intercomparison Report |
| QA | Quality Assurance |
| QC | Quality Control |
| RemoTeC | Retrieval algorithm developed by SRON |
| REQ | Requirement |
| RMS | Root-Mean-Square |
| RTM | Radiative transfer model |
| SCIAMACHY | SCanning Imaging Absorption spectroMeter for Atmospheric ChartographY |
| SCIATRAN | SCIAMACHY radiative transfer model |
| SLIM | Simple cLimatological Model for atmospheric CO ₂ and CH ₄ |
| SRON | SRON Netherlands Institute for Space Research |
| SRFP | SRON's Full Physics (FP) algorithm (also referred to a RemoTeC FP) |
| SRPR | SRON's Proxy (PR) algorithm (also referred to a RemoTeC PR) |
| SWIR | Short Wave Infra-Red |
| TANSO | Thermal And Near infrared Sensor for carbon Observation |
| TANSO-FTS | Fourier Transform Spectrometer on GOSAT |
| TANSO-FTS-2 | Fourier Transform Spectrometer on GOSAT-2 |
| 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

Essential climate variable (ECV)

An ECV is a physical, chemical, or biological variable or a group of linked variables that critically contributes to the characterization of Earth's climate (Bojinski et al., 2014).

Climate data record (CDR)

The US National Research Council (NRC) defines a CDR as a time series of measurements of sufficient length, consistency, and continuity to determine climate variability and change (National Research Council, 2004).

Fundamental climate data record (FCDR)

A fundamental climate data record (FCDR) is a CDR of calibrated and quality-controlled data designed to allow the generation of homogeneous products that are accurate and stable enough for climate monitoring.

Thematic climate data record (TCDR)

A thematic climate data record (TCDR) is a long time series of an essential climate variable (ECV) (Werscheck, 2015).

Intermediate climate data record (ICDR)

An intermediate climate data record (ICDR) is a TCDR which undergoes regular and consistent updates (Werscheck, 2015), for example because it is being generated by a satellite sensor in operation.

Satellite data processing levels

The NASA Earth Observing System (EOS) distinguishes six processing levels of satellite data, ranging from Level 0 (L0) to Level 4 (L4) as follows (Parkinson et al., 2006).

- L0 Unprocessed instrument data
- L1A Unprocessed instrument data alongside ancillary information
- L1B Data processed to sensor units (geo-located calibrated spectral radiance and solar irradiance)
- L2 Derived geophysical variables (e.g., XCO₂) over one orbit
- L3 Geophysical variables averaged in time and mapped on a global longitude/latitude horizontal grid
- L4 Model output derived by assimilation of observations, or variables derived from multiple measurements (or both)



The C3S GHG data products include the following types of data.

Level 2 data:

- Individual sensor and multi-sensor merged CO₂ and CH₄ retrievals for each (quality flagged) satellite ground-pixel

Level 3 data:

- Gridded monthly averages (5° x 5°) covering the globe in Obs4MIPs format



Table of Contents

| | |
|---|-----------|
| History of modifications | 4 |
| List of datasets covered by this document | 5 |
| Related documents | 5 |
| Acronyms | 6 |
| General definitions | 10 |
| Scope of document | 13 |
| Executive summary | 14 |
| 1 Instruments | 16 |
| 1.1 SCIAMACHY onboard ENVISAT | 16 |
| 1.2 TANSO-FTS onboard GOSAT and TANSO-FTS-2 onboard GOSAT-2 | 17 |
| 1.3 OCO-2 | 17 |
| 2 Input and auxiliary data | 18 |
| 2.1 Satellite L2 data | 18 |
| 2.2 Other data | 20 |
| 3 Algorithms | 21 |
| 3.1 EMMA algorithm | 21 |
| 3.1.1 Overview and comparison with reference data | 21 |
| 3.1.2 EMMA ensemble spread | 29 |
| 3.1.3 EMMA ensemble median | 36 |
| 3.2 Obs4MIPs algorithm | 37 |
| 4 Output data | 38 |
| 4.1 EMMA products | 38 |
| 4.2 Obs4MIPs products | 40 |
| 5 References | 42 |



Scope of document

This document is the Algorithm Theoretical Basis Document (ATBD) ANNEX D) for the Copernicus Climate Change Service (C3S, <https://climate.copernicus.eu/>) greenhouse gas (GHG) component as covered by project C3S2_312a_Lot2.

Within this project satellite-derived atmospheric carbon dioxide (CO₂) and methane (CH₄) Essential Climate Variable (ECV) data products are being 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 satellite-derived GHG data products are:

- Column-averaged dry-air mixing ratios (mole fractions) of CO₂ and CH₄, denoted XCO₂ (in parts per million, ppm) and XCH₄ (in parts per billion, ppb), respectively.
- Mid/upper tropospheric mixing ratios of CO₂ (in ppm) and CH₄ (in ppb).

This document describes the algorithms to generate the version 4.5 C3S products XCO₂_EMMA, XCH₄_EMMA, XCO₂_OBS4MIPS and XCH₄_OBS4MIPS.

These products are merged multi-sensor XCO₂ and XCH₄ level 2 and level 3 products generated using algorithms which are described in this document and which have been developed at the University of Bremen, Germany (see also Reuter et al. (2013, 2020)).



Executive summary

In the frame of the Copernicus Climate Change Service (C3S, <https://climate.copernicus.eu/>) greenhouse gases (GHG) component (project C3S_312b_Lot2 led by DLR, Germany) several satellite-derived atmospheric carbon dioxide (CO₂) and methane (CH₄) Essential Climate Variable (ECV) data products are being generated by different institutes. The main ATBD (D1) provides only a short overview about all of these data products which are described in more detail in corresponding annexes to the main ATBD.

The annex at hand describes the algorithm theoretical basis for the algorithms used to generate the following data products: (i) XCO₂_EMMA, (ii) XCH₄_EMMA, (iii) XCO₂_OBS4MIPS and (iv) XCH₄_OBS4MIPS.

The Ensemble Median Algorithm (EMMA) products are Level 2 products obtained by merging an ensemble of individual sensor Level 2 data products. These “Ensemble Median Algorithm”, i.e., EMMA products, are used as input for the generation of the Obs4MIPs data products.

The Observations for Model Intercomparison Project (Obs4MIPs) products are gridded Level 3 products obtained via spatial (5°x5°) and temporal (monthly) averaging of the EMMA products. The Obs4MIPs products are generated in “Observations for Model Intercomparison Project” (Obs4MIPs) data format (see <https://esgf-node.llnl.gov/projects/obs4mips/>). Obs4MIPs is an activity to make observational products more accessible especially for climate model inter-comparisons (see, e.g., Gier et al., 2020).

This ATBD-annex describes the algorithms used to generate the version 4.5 data products, which cover the period January 2003 to December 2022.

The first version of the EMMA algorithm (v1.3) was described in detail using the example of CO₂ in the publication of Reuter et al. (2013). More recently, Reuter et al. (2020) described updated EMMA and Obs4MIPs CO₂ and CH₄ developments and data sets (v4.1 covering 2003-2018). These publications are the blueprint for this ATBD. The focus is on the description of the EMMA data products, which are inputs for the Obs4MIPs products.

For a long time, climate modelers have used ensemble approaches to calculate the ensemble median and to estimate uncertainties of climate projections where no ground-truth is known. Following this idea, the ensemble median algorithm EMMA brings together level 2 data of several SCIAMACHY, GOSAT, GOSAT-2, and OCO-2 XCO₂ and XCH₄ retrieval products independently developed by NASA, NIES, SRON, the University of Leicester, and the University of Bremen. EMMA determines in 10°x10° degree grid boxes monthly averages and selects the level 2 data of the median algorithm. Thresholds depending on potential information content prevent from over-weighting individual algorithms with a considerably larger amount of data.

The EMMA database consists of individual level 2 soundings retrieved by algorithms which can change from grid box to grid box and month to month. Therefore, it can be used in the same manner as any other XCO₂ or XCH₄ satellite retrieval, i.e., the EMMA database includes all information needed for inverse modeling (geo-location, time, XCO₂ or XCH₄, averaging kernels, etc.).



Additionally, it includes the inter-algorithm spread which informs about potential regional or temporal systematic uncertainties.

The following sections comprise an overview on the used satellite instruments (Section 1), the used data products and auxiliary data (Section 2), detailed descriptions of the algorithms generating the EMMA L2 database (Section 3.1) and the corresponding Obs4MIPs L3 database (Section 3.2), and a brief description of the output data (Section 4).



1 Instruments

Level 1 data for the following instruments have been used to generate the data products and corresponding algorithms as described in this document:

- SCIAMACHY onboard ENVISAT (Bovensmann et al., 1999)
- TANSO-FTS onboard GOSAT (Kuze et al., 2009, 2016)
- TANSO-FTS-2 onboard GOSAT-2 (Suto et al., 2021)
- OCO-2 (Crisp et al., 2004; Boesch et al., 2011)

Details on these instruments are given in the following sub-sections.

1.1 SCIAMACHY onboard ENVISAT

SCIAMACHY (SCanning Imaging Absorption spectroMeter for Atmospheric Chartography) was a spectrometer on ESA's ENVISAT satellite (2002-2012). SCIAMACHY (Burrows et al., 2005; Bovensmann et al., 1999) covers the spectral region from the ultra-violet to the short wave infra-red (SWIR) spectral region (240 nm - 2380 nm) at moderate spectral resolution (0.2 nm - 1.5 nm) and observes the Earth's atmosphere in various viewing geometries (nadir, limb and solar and lunar occultation).

For a SCIAMACHY overview see <https://earth.esa.int/eogateway/instruments/sciamachy> and <https://en.wikipedia.org/wiki/SCIAMACHY>. SCIAMACHY permits the retrieval of XCO₂ (e.g., Reuter et al., 2011; Schneising et al., 2011) and XCH₄ (e.g., Schneising et al., 2011; Frankenberg et al., 2011) from the appropriate spectral regions in the SWIR (around 1.6 µm) and the NIR (O₂ A-band at 760 nm used to obtain the dry-air column using the known dry-air mixing ratio of atmospheric oxygen). The ground pixel size is typically 30 km along track times 60 km across track and the swath width is about 960 km. There are no across-track gaps between the ground pixels but there are gaps along-track as SCIAMACHY operates only part of the time (approx. 50%) in nadir observation mode.



1.2 TANSO-FTS onboard GOSAT and TANSO-FTS-2 onboard GOSAT-2

TANSO-FTS is a Fourier-Transform-Spectrometer (FTS) onboard the Japanese GOSAT satellite (Kuze et al., 2009, 2014, 2016). The Greenhouse Gases Observing Satellite "IBUKI" (GOSAT) is the world's first spacecraft in orbit dedicated to measure the concentrations of carbon dioxide and methane from space. The spacecraft was launched successfully on January 23, 2009, and has been operating properly since then. GOSAT covers the relevant CO₂, CH₄ and O₂ absorption bands in the NIR and SWIR spectral region as needed for accurate XCO₂ and XCH₄ retrieval (in addition GOSAT also covers a large part of the Thermal Infrared (TIR) spectral region). The spectral resolution of TANSO-FTS is much higher compared to SCIAMACHY and the ground pixels are smaller (10 km compared to several 10 km for SCIAMACHY). However, in contrast to SCIAMACHY, the GOSAT scan pattern consists of non-consecutive individual ground pixels, i.e., the scan pattern is not gap-free.

GOSAT-2 (Suto et al., 2021) was successfully launched on 29 October 2018. GOSAT-2 XCO₂ and XCH₄ retrievals are now also included in the C3S GHG CDR.

Concerning XCO₂ and XCH₄ retrieval from GOSAT and GOSAT-2 see also Noël et al., 2021 and 2022.

1.3 OCO-2

NASA's Orbiting Carbon Observatory 2 (OCO-2) mission (Crisp et al., 2004; Boesch et al., 2011) has been successfully launched in July 2014. The OCO-2 Project's primary science objective is to collect the first space-based measurements of atmospheric carbon dioxide with the precision, resolution and coverage needed to characterize its sources and sinks and quantify their variability over the seasonal cycle. OCO-2 flies in a sun-synchronous, near-polar orbit with a group of Earth-orbiting satellites with synergistic science objectives whose ascending node crosses the equator near 13:30 hours Mean Local Time (MLT). Near-global coverage of the sunlit portion of Earth is provided in this orbit over a 16-day (233-revolution) repeat cycle. OCO-2's single instrument incorporates three high-resolution grating spectrometers, designed to measure the near-infrared absorption of reflected sunlight by carbon dioxide and molecular oxygen. OCO-2 covers 3 spectral bands (see Table 1) which are as SCIAMACHY and GOSAT but OCO-2 has much smaller ground pixels (km scale) but the swath width is much smaller (approx. 10 km) compared to SCIAMACHY. OCO-2 delivers XCO₂ but not XCH₄.

Table 1: OCO-2 spectral bands.

| Band | Wavelength range [nm] | Resolution [nm] |
|---|-----------------------|-----------------|
| Band 1 (ABO ₂ , O ₂ band) | 758 - 772 | 0.04 |
| Band 2 (WCO ₂ , weak CO ₂ band) | 1594 - 1619 | 0.08 |
| Band 3 (SCO ₂ , strong CO ₂ band) | 2042 - 2082 | 0.10 |



2 Input and auxiliary data

2.1 Satellite L2 data

Several different XCO₂ and/or XCH₄ retrieval algorithms exist for SCIAMACHY, GOSAT, GOSAT-2, and OCO-2 which are partially under active further development in order to meet the demanding user requirements for using these data products to obtain source / sinks information, i.e., for making them useful for surface flux inversions. Specifically, we make use of the algorithms and corresponding data products listed in Table 2 and Table 3.

Table 2: L2 data products used in EMMA v4.5 XCO₂. The table lists the satellite instrument, the name and version of the L2 algorithm, the institution, the algorithm ID used by EMMA, and related references.

| Satellite/Instrument | Algorithm / product and version number | Institution | ID | Reference |
|----------------------|--|-------------|----|---|
| SCIAMACHY | BESD v02.01.02 | IUP | 2 | Reuter et al. (2010, 2011, 2016) |
| GOSAT | NIES v02.9xbc (bias corrected) | NIES | 3 | Yoshida et al. (2013) |
| GOSAT | RemoTeC v2.3.8 | SRON | 5 | Butz et al. (2011), Detmers et al. (2017a) |
| GOSAT | UoL-FP v7.3 | UoL | 6 | Cogan et al (2012) Boesch and Anand (2017) |
| GOSAT | ACOS v9r | NASA | 7 | O'Dell et al. (2012), Taylor et al. (2022) |
| GOSAT | FOCAL v3.0 | IUP | 8 | Noël et al. (2022) |
| OCO-2 | NASA v11.1 | NASA | 9 | Kiel et al. (2019) |
| OCO-2 | FOCAL v10.1 | IUP | 10 | Reuter et al. (2017a, 2017b, 2021) |
| GOSAT-2 | NIES v02.00 | NIES | 11 | Yoshida and Oshio (2020) |
| GOSAT-2 | RemoTeC v2.0.0 | SRON | 12 | Krisna et al. (2021) |
| GOSAT-2 | FOCAL v3.0 | IUP | 13 | Noël et al. (2022) |



Table 3: L2 data products used in EMMA v4.5 XCH₄. The table lists the satellite instrument, the name and version of the L2 algorithm, the institution, the algorithm ID used by EMMA, and related references.

| Satellite/Instrument | Algorithm / product and version number | Institution | ID | Reference |
|----------------------|--|-------------|----|---|
| SCIAMACHY | WFMD v4.0 | IUP | 2 | Schneising et al. (2018) |
| GOSAT | FOCAL-FP v3.0 | IUP | 3 | Noël et al. (2022) |
| GOSAT | FOCAL-PR v3.0 | IUP | 4 | Noël et al. (2022) |
| GOSAT | NIES v02.9xbc (bias corrected) | NIES | 5 | Yoshida et al. (2013) |
| GOSAT | RemoTeC-FP v2.3.8 | SRON | 7 | Butz et al. (2011), Detmers et al. (2017a) |
| GOSAT | RemoTeC-PR v2.3.9 | SRON | 8 | Butz et al. (2011), Detmers et al. (2017b) |
| GOSAT | UoL-FP v7.3 | UoL | 9 | Cogan et al (2012) Boesch and Anand (2017) |
| GOSAT | UoL-PR v9.0 | UoL | 10 | Cogan et al (2012) Boesch and Anand (2017) |
| GOSAT-2 | FOCAL-FP v3.0 | IUP | 11 | Noël et al. (2022) |
| GOSAT-2 | FOCAL-PR v3.0 | IUP | 12 | Noël et al. (2022) |
| GOSAT-2 | RemoTeC-FP v2.0.0 | SRON | 13 | Krisna et al. (2021) |
| GOSAT-2 | RemoTeC-PR v2.0.1 | SRON | 14 | Krisna et al. (2021) |
| GOSAT-2 | NIES v02.00 | NIES | 15 | Yoshida and Oshio (2020) |



2.2 Other data

In order to account for different column averaging kernels, all retrieval results are adjusted to a common a priori, namely the Simple climatological Model for atmospheric CO₂ and CH₄ (SLIM) (Noël et al., 2022). Additionally, the influence of the smoothing error is reduced for the validation as described by Wunch et al. (2011).

SLIM CO₂ and CH₄ use model-based climatologies to estimate the spatial distribution and annual cycle of CO₂ and CH₄, respectively. Additionally, SLIM adjusts the used climatologies for the annual growth.

Scaling of the reported L2 uncertainties and validation is done with TCCON (Total Carbon Column Observing Network, Wunch et al., 2011) GGG2020 as reference data set. TCCON provides XCO₂ and XCH₄ ground-based retrievals which can directly be compared with the corresponding satellite-based retrievals.



3 Algorithms

3.1 EMMA algorithm

3.1.1 Overview and comparison with reference data

The EMMA and Obs4MIPs algorithms are described in detail by *Reuter et al. (2013, 2020)*. In short, the algorithms work as follows:

- The EMMA algorithm consists of collecting all required (mainly Level 2) input data, harmonization, merging/selecting and storing the results in daily EMMA Level 2 output files.
- The main input data for the Obs4MIPs products are the daily EMMA Level 2 input files. These data products are used to compute monthly $5^{\circ} \times 5^{\circ}$ grid cell averages including uncertainty and averaging kernels etc.
- Figure 1 shows a flowchart of the main steps of the EMMA algorithm and corresponding OBS4MIPS data generation.

The basic principle of all these input data Level 2 algorithms and corresponding data products as listed in Table 2 and Table 3 is the same: (i) A satellite instrument measures backscattered solar radiation in near-infrared O_2 and CO_2 or CH_4 absorption bands. (ii) A radiative transfer plus instrument model (forward model) is utilized to simulate the satellite measurement for a set of known parameters (parameter vector) and unknown parameters (state vector). (iii) An inversion method tries to find that state vector which results in best agreement of simulated and measured radiances. (iv) The retrieved state vector is assumed to represent the true (or most likely) atmospheric state.

However, when going into more detail, the algorithms have distinct conceptual differences: the algorithms are optimized for different instruments (SCIAMACHY, GOSAT, GOSAT-2, or OCO-2). They are based on different absorption bands, use different inversion methods (optimal estimation, Tikhonov-Phillips, least squares), and are based on different physical assumptions on the radiative transfer in scattering atmospheres. In order to give two examples, so-called full physics algorithms explicitly account for (multiple) scattering at molecules, aerosols, and/or clouds by having state vector elements such as cloud water path, cloud top height, and aerosol optical thickness; the light path proxy method assumes that photon path lengths are modified similarly in the CO_2 and O_2 or CH_4 absorption bands, and that scattering related effects cancel out when dividing the retrieved CO_2 and O_2 or CH_4 columns when building XCO_2 or XCH_4 . Additionally, the algorithms use different pre- and post-processing filters (e.g., cloud detection from O_2 -A band or from a cloud and aerosol imager).

Inputs for the data products and corresponding algorithms as described in this document are individual satellite sensor Level 2 data products. In this section we present an overview of these input data products.

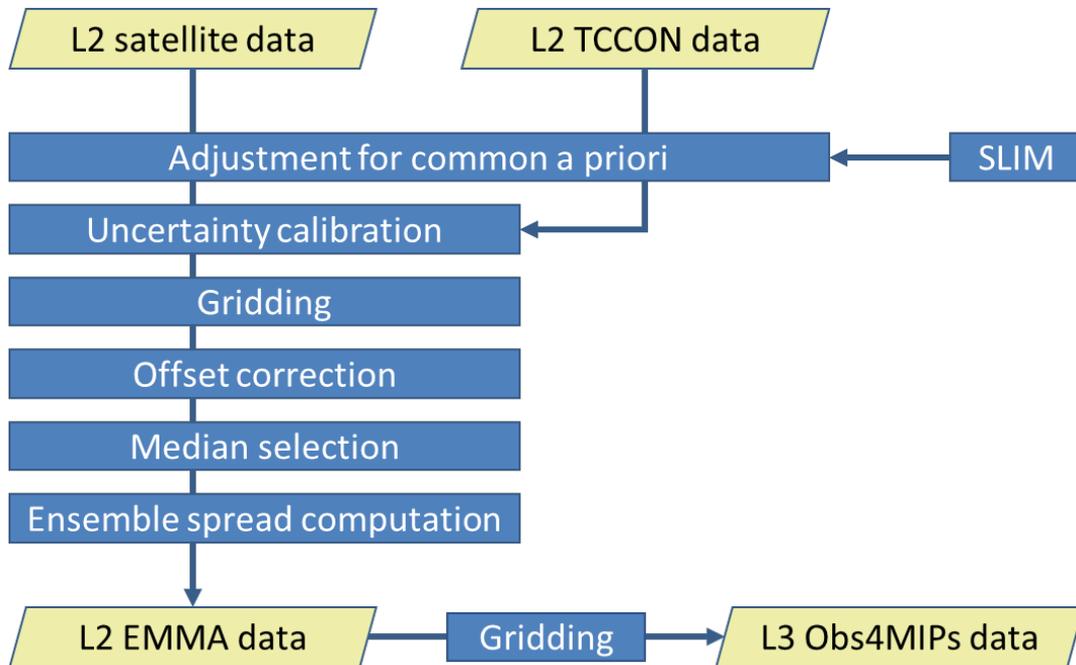


Figure 1: Flowchart showing the main steps of the EMMA algorithm and Obs4MIPs data generation.

Our current knowledge about the sources and sinks of atmospheric CO₂ and CH₄ is limited by the sparseness of highly accurate and precise measurements of these gases (e.g., Stephens *et al.*, 2007). Due to their global coverage and sensitivity down to the surface, satellite based XCO₂ and XCH₄ (column-average dry-air mole fraction of atmospheric CO₂ and CH₄) retrievals in the near infrared are promising candidates to reduce existing uncertainties if accurate and precise enough (e.g., Rayner and O’Brien, 2001; Houweling *et al.*, 2004; Miller *et al.*, 2007; Chevallier *et al.*, 2007).

At present, several independently developed XCO₂ and/or XCH₄ retrieval algorithms exist for SCIAMACHY (SCanning Imaging Absorption spectrometer of Atmospheric CHartographY; Burrows *et al.*, 1995; Bovensmann *et al.*, 1999), GOSAT (Greenhouse gases Observing SATellite; Yokota *et al.*, 2004), GOSAT-2 (Greenhouse gases Observing SATellite 2; Suto *et al.*, 2021) and OCO-2 (Orbiting Carbon Observatory-2; Crisp *et al.*, 2017); see Table 2 and Table 3 for those used for EMMA v4.5 CO₂ and EMMA v4.5 CH₄, respectively.

All retrieval teams find encouraging validation results when comparing with TCCON (total carbon column observing network, Wunch *et al.*, 2011) ground-based FTS (Fourier transform spectrometer) measurements (see references in the next section and in Table 2 and Table 3). This goes along with a good inter-algorithm agreement at TCCON sites and with the results of our unified validation study (see D2) having station-to-station biases (i.e., the standard deviation of the biases at different sites) usually below 0.6 ppm and 5.3 ppb and single measurement precisions usually below 2.1 ppm and 14 ppb for XCO₂ and XCH₄, respectively (Figure 2, Figure 3, Table 4, Table 5).



However, the inter-algorithm agreement often reduces remote from validation sites due to differing large-scale bias patterns. Such biases can be a critical issue for surface flux inversions and the user requirements are demanding; as an example, Miller et al. (2007) and Chevallier et al. (2007) found that regional biases of a few tenths of a ppm can already hamper surface flux inversions. This indicates that assessing an algorithm's quality should not be based on comparisons against current TCCON stations only. Obviously, large regions of the world possess more "complicated" retrieval conditions without the availability of ground truth measurements which could be used to judge the algorithms' performance.

Diverging model results are common to many scientific disciplines (e.g., Araujo and New, 2007; Rötter et al., 2011) and much attention and effort is devoted to this topic on the subject of weather and climate modeling. Here, the divergence of the model results arises not only from structural differences of the different models, but also from the nonlinearity of the model equations, leading to differing results of one single model when performing multiple realizations with slightly differing initial conditions (Hagedorn et al., 2005; Tebaldi and Knutti, 2007).

Especially in the case of weather forecasting or climate projections, where no ground truth is available for the verification of the forecasts and projections, it is impossible to identify the "best" model and the "perfect" initial conditions. For long-term climate projections, this problem is impaired by the unknown future greenhouse forcing.

This conceptual problem is dealt with by using multi-model, multi-realization, multi-emission-scenario ensembles of simulations, which ideally span the entire range of possible model outcomes and, thus, can be used to estimate the uncertainties of the forecast or projection.

However, interpreting the ensemble's spread as uncertainty is not the only possible application: some studies indicate that the ensemble mean, weighted mean, or median can outperform each individual model under appropriate conditions (e.g., Kharin and Zwiers, 2002; Vautard et al., 2009).

Here, we seize this idea for the ensemble median algorithm EMMA which uses data from the retrieval algorithms listed in Table 2 and Table 3 and within the next section. EMMA generates a database of individual level 2 retrievals and takes advantage of the variety of different retrieval algorithms and their independent developments.

For each month and each $10^{\circ} \times 10^{\circ}$ grid box, one algorithm is chosen to supply level 2 retrievals for the database. The algorithm is chosen on the basis that its grid box mean is the median amongst the available algorithms. This allows the reduction of occasional outliers and sometimes unrealistic bias patterns, which may be found in each individual retrieval algorithm and which may hamper surface flux inversions. EMMA relies on the assumption that it is unlikely that the majority of algorithms produce outliers in the same direction because only in this case the median is a bad choice.

Smoothing of real atmospheric variability, as it could happen when dealing with climate model ensembles, cannot be expected for EMMA because all ensemble members (XCO₂ or XCH₄ retrieval algorithms) represent the same (real) atmospheric XCO₂ or XCH₄ conditions and deviations from the real values are always due to retrieval errors (sampling issues are neglected in this context).



The EMMA database includes all information needed for inverse modeling (geo-location, time, XCO₂ or XCH₄, averaging kernels, etc.). As it consists of individual XCO₂ or XCH₄ retrievals, it can be used in the same manner as any other XCO₂ or XCH₄ satellite retrieval. Additionally, the EMMA database includes the inter-algorithm spread which gives important information about regional and/or temporal systematic uncertainties.

The Level 2 input data and the resulting EMMA data product are validated by comparison with TCCON. The results of this comparison are shown in Table 4 and Table 5 and Figure 2 and Figure 3.



Table 4: Summarizing XCO₂ TCCON validation statistics for the period and sites shown in Figure 2 with number of co-locations (#), average single measurement precision (σ) relative to TCCON, and standard deviation of station-to-station biases (Δ). The table includes only a subset of validation statistics shown in D2, where also more information on the validation method can be found.

| Algorithm | # | σ [ppm] | Δ [ppm] |
|-----------------------------|---------|----------------|----------------|
| SCIAMACHY BESD v02.01.02 | 6080 | 1.83 | 0.25 |
| GOSAT NIES v02.9xbc | 14088 | 2.21 | 0.41 |
| GOSAT RemoTeC v2.3.8 | 8490 | 2.26 | 0.62 |
| GOSAT UoL-FP v7.3 | 10009 | 1.95 | 0.45 |
| GOSAT ACOS v9r | 11239 | 1.73 | 0.28 |
| GOSAT FOCAL v3.0 | 9581 | 2.28 | 0.28 |
| OCO-2 NASA v11.1 | 1980981 | 1.29 | 0.36 |
| OCO-2 FOCAL v10.1 | 839449 | 1.76 | 0.34 |
| GOSAT-2 NIES v02.00 | 5924 | 2.52 | 0.60 |
| GOSAT-2 RemoTeC v2.0.0 | 2692 | 2.91 | 0.63 |
| GOSAT-2 FOCAL v3.0 | 2779 | 2.09 | 0.46 |
| EMMA v4.5 | 27281 | 1.78 | 0.41 |



Table 5: Summarizing XCH₄ TCCON validation statistics for the period and sites shown in Figure 3 with number of co-locations (#), average single measurement precision (σ) relative to TCCON, and standard deviation of station-to-station biases (Δ). The table includes only a subset of validation statistics shown in D2, where also more information on the validation method can be found.

| Algorithm | # | σ [ppm] | Δ [ppm] |
|--|-------|----------------|----------------|
| SCIAMACHY WFMD v4.0 | 5425 | 104.27 | 5.75 |
| GOSAT FOCAL-FP v3.0 | 11118 | 12.26 | 1.27 |
| GOSAT FOCAL-PR v3.0 | 26346 | 13.15 | 2.86 |
| GOSAT NIES v02.9xbc (bias corrected) | 13731 | 13.31 | 2.15 |
| GOSAT RemoTeC-FP v2.3.8 | 8138 | 14.29 | 1.96 |
| GOSAT RemoTeC-PR v2.3.9 | 32546 | 15.15 | 1.36 |
| GOSAT UoL-FP v7.3 | 9343 | 13.25 | 1.58 |
| GOSAT UoL-PR v9.0 | 30599 | 14.31 | 1.58 |
| GOSAT-2 FOCAL-FP v3.0 | 2527 | 11.51 | 3.36 |
| GOSAT-2 FOCAL-PR v3.0 | 6294 | 12.12 | 5.29 |
| GOSAT-2 RemoTeC-FP v2.0.0 | 2692 | 17.31 | 2.11 |
| GOSAT-2 RemoTeC-PR v2.0.1 | 9415 | 17.80 | 2.71 |
| GOSAT-2 NIES v02.00 | 5937 | 13.79 | 3.07 |
| EMMA v4.5 | 20659 | 13.51 | 1.79 |

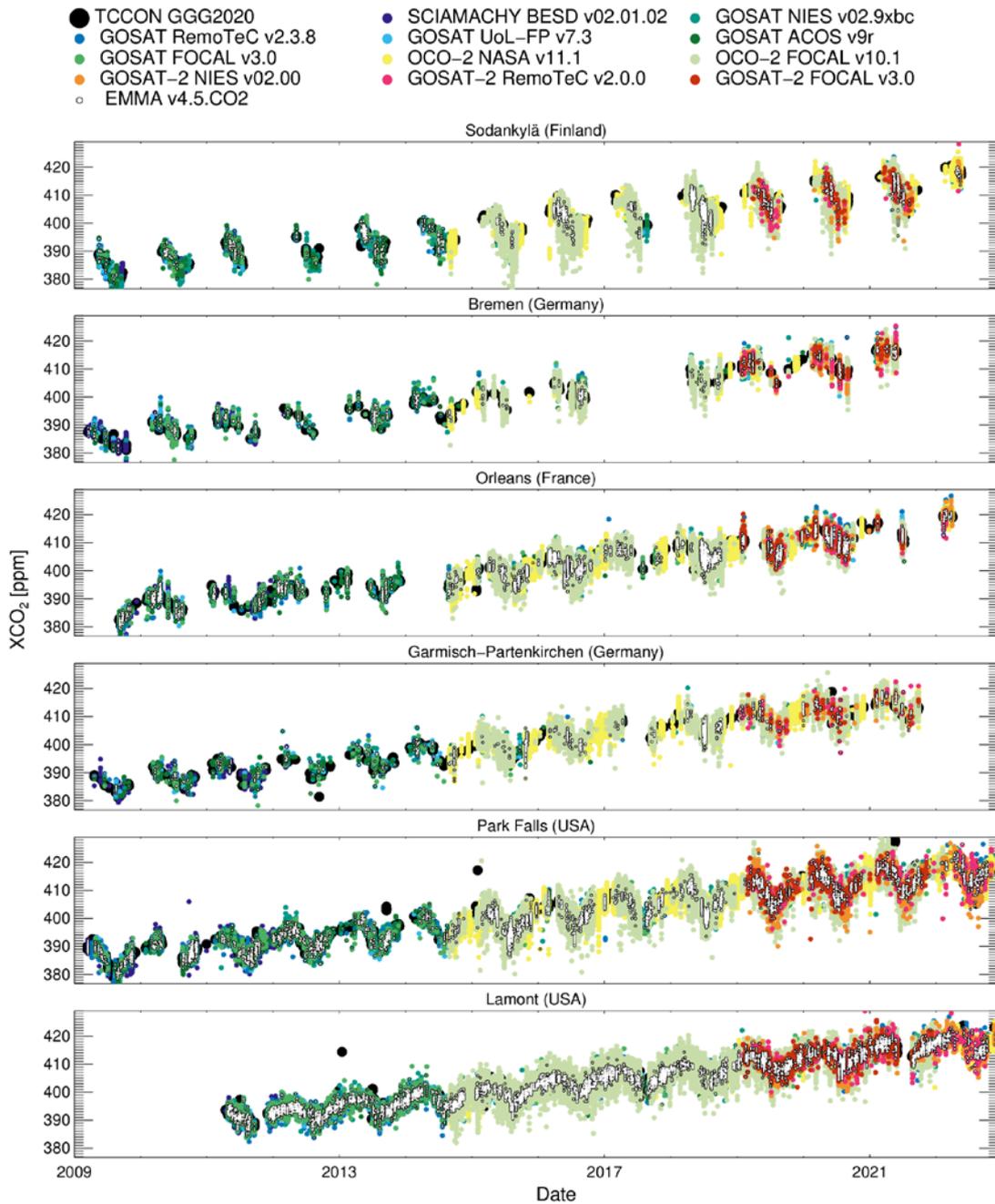


Figure 2: Validation of individual XCO₂ algorithms and EMMA’s XCO₂ with TCCON.

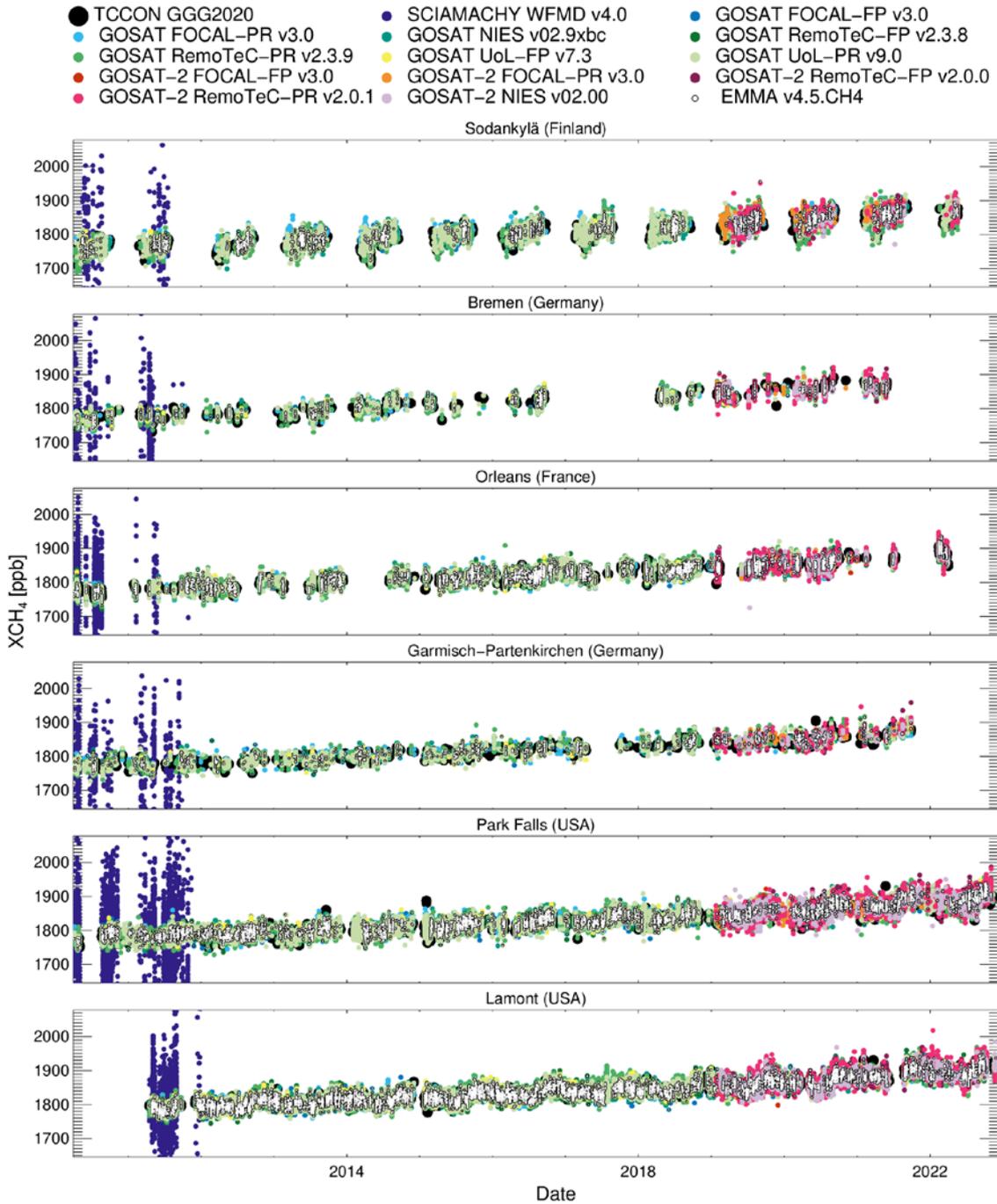


Figure 3: Validation of individual XCH₄ algorithms and EMMA's XCO₂ with TCCON.



3.1.2 EMMA ensemble spread

Due to entirely different samplings (different satellites, different filtering strategies, etc.), any algorithm inter-comparison considering the majority of individual soundings (level 2) can only be based on aggregated data (level 3), in our case monthly averages on a $10^{\circ} \times 10^{\circ}$ grid.

Before gridding, we apply the individual averaging kernels to adjust all retrieval results to a common a priori, namely the simple climatological model SLIM CO₂ and CH₄ (see Noël et al., 2022). We do this as proposed by Wunch et al. (2011) and as in the textbook of Rodgers (2000). SLIM CO₂ and CH₄ reproduce large-scale features such as the year-to-year increase, the north/south gradient, and the seasonal cycle. However, SLIM CO₂ and CH₄ are only empirically extrapolating from past/averaged modeled CO₂ and CH₄ fields. New or changing phenomena cannot be within SLIM CO₂ and CH₄, and it should also be mentioned that the adjustments are mostly minor, especially for CO₂ with typically a few tenths of a ppm.

For consistency, we remove the overall global bias of each retrieval with SLIM as reference. In order to get statistically robust results, we only use those grid boxes with more than five soundings and for which the standard error of the mean is estimated to be less than 1 ppm and 12 ppb for XCO₂ and XCH₄, respectively. This takes the individual retrieval precisions into account so that the minimum number of soundings needed to build the average of a grid box can vary from retrieval to retrieval and grid box to grid box. Additionally, only grid boxes with a maximum number of overlapping algorithms (see Figure 4) are considered for the global bias adjustment. Beforehand, we calibrate the reported retrieval uncertainty (precision) by scaling to match (on average) the precision given in Table 4 and Table 5 obtained from a unified validation with TCCON data. Figure 5 shows the influence of the global offset correction.

Figure 6 and Figure 7 show examples of the calculated monthly XCO₂ or XCH₄ averages, respectively. First of all, one can see many large-scale similarities such as the north/south gradient. However, one can also find more or less obvious outliers in the order of a percent for several algorithms. Often the observed systematic deviations (of level 3 data) are larger than expected from instrumental noise, i.e., they are dominated by specific algorithm effects. As level 3 grid boxes are always calculated from several individual level 2 soundings (ideally) sampled all over the grid box, we expect that sampling and representation errors are lower than the observed deviations. Therefore, these errors are not discussed further in this context.

Due to independent algorithm developments, different physical approaches and assumptions, different pre- and post-processing filters, and due to the different instruments, we expect relatively independent bias patterns. This is supported by Figure 6 and Figure 7, which show (uncorrelated) obvious outliers in various regions, i.e., it seems unlikely that all algorithms produce the same bias within one grid box. This implies that similar averages within one grid box can give us more confidence in the individual retrievals within this grid box.

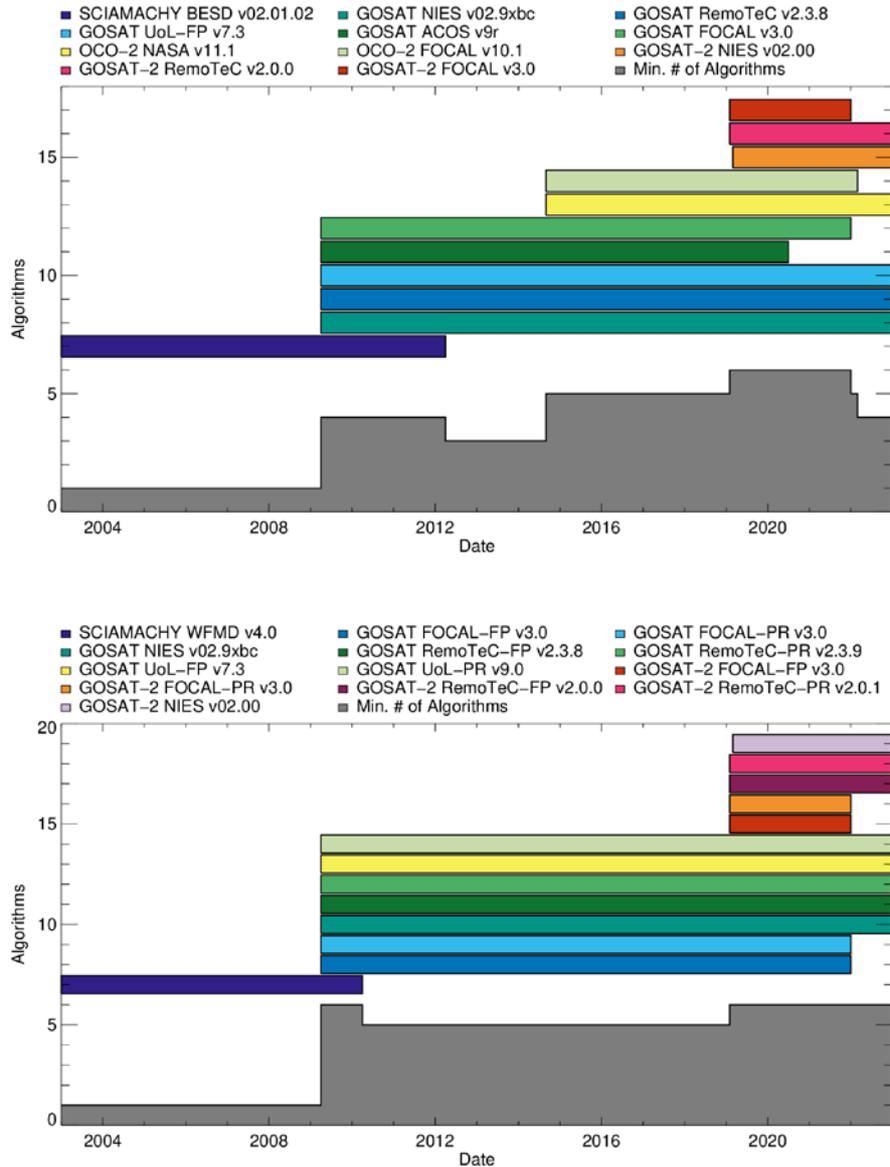


Figure 4: EMMA input data availability (colored bars) and minimum number of used algorithms (gray) for median calculation for CO₂ (top) and CH₄ (bottom).

On the other hand, large inter-algorithm spreads indicate regions with more difficult and uncertain retrieval conditions. Therefore, we interpret the ensemble spread, i.e., the standard deviation, as uncertainty due to regional retrieval biases. An example is given in Figure 8 and Figure 9 showing larger inter-algorithm spreads for XCO₂ and XCH₄ in the tropics and in Asia (mostly remote from TCCON sites). This pattern is temporally more or less stable, i.e., similar also for sub-periods.

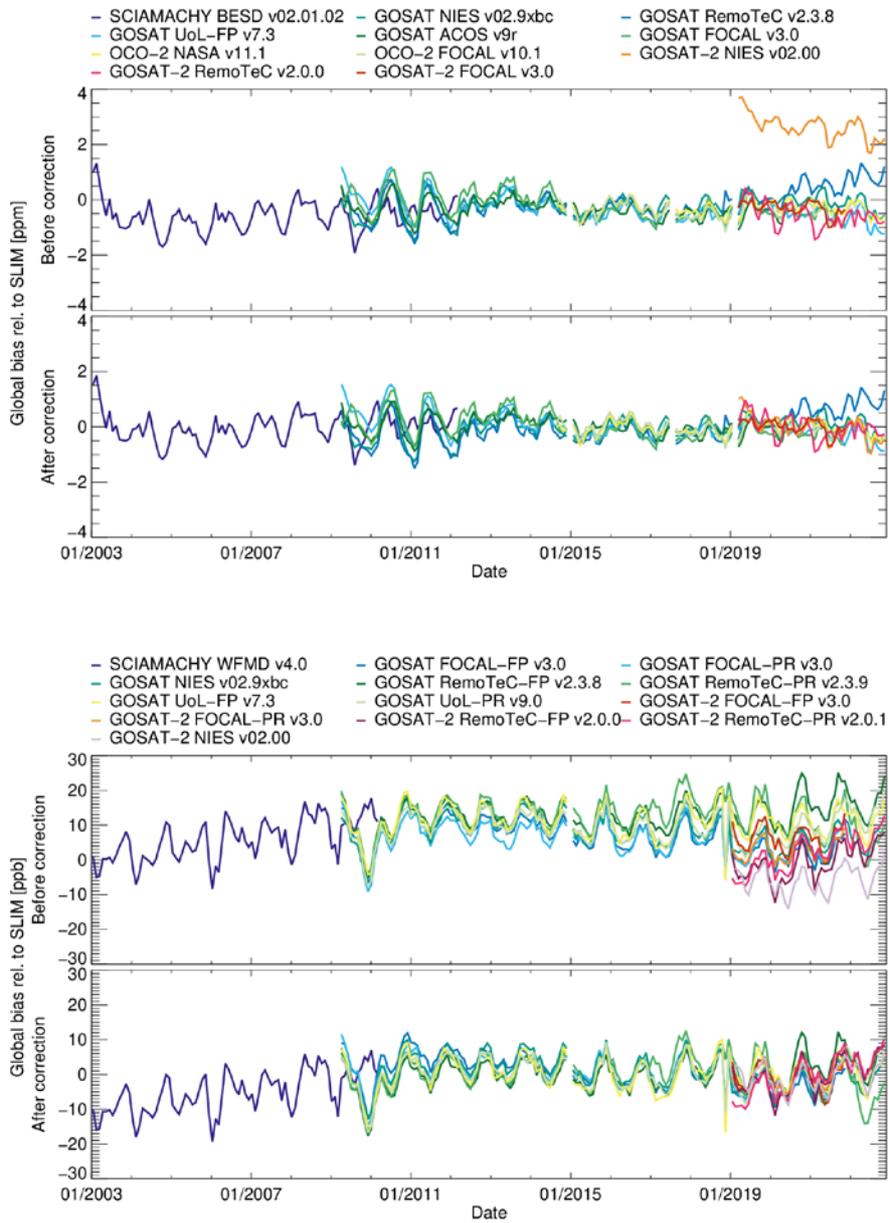


Figure 5: Global monthly average bias for XCO₂ (top) and XCH₄ (bottom) in common grid boxes relative to SLIM CO₂ and CH₄ before and after global bias correction.

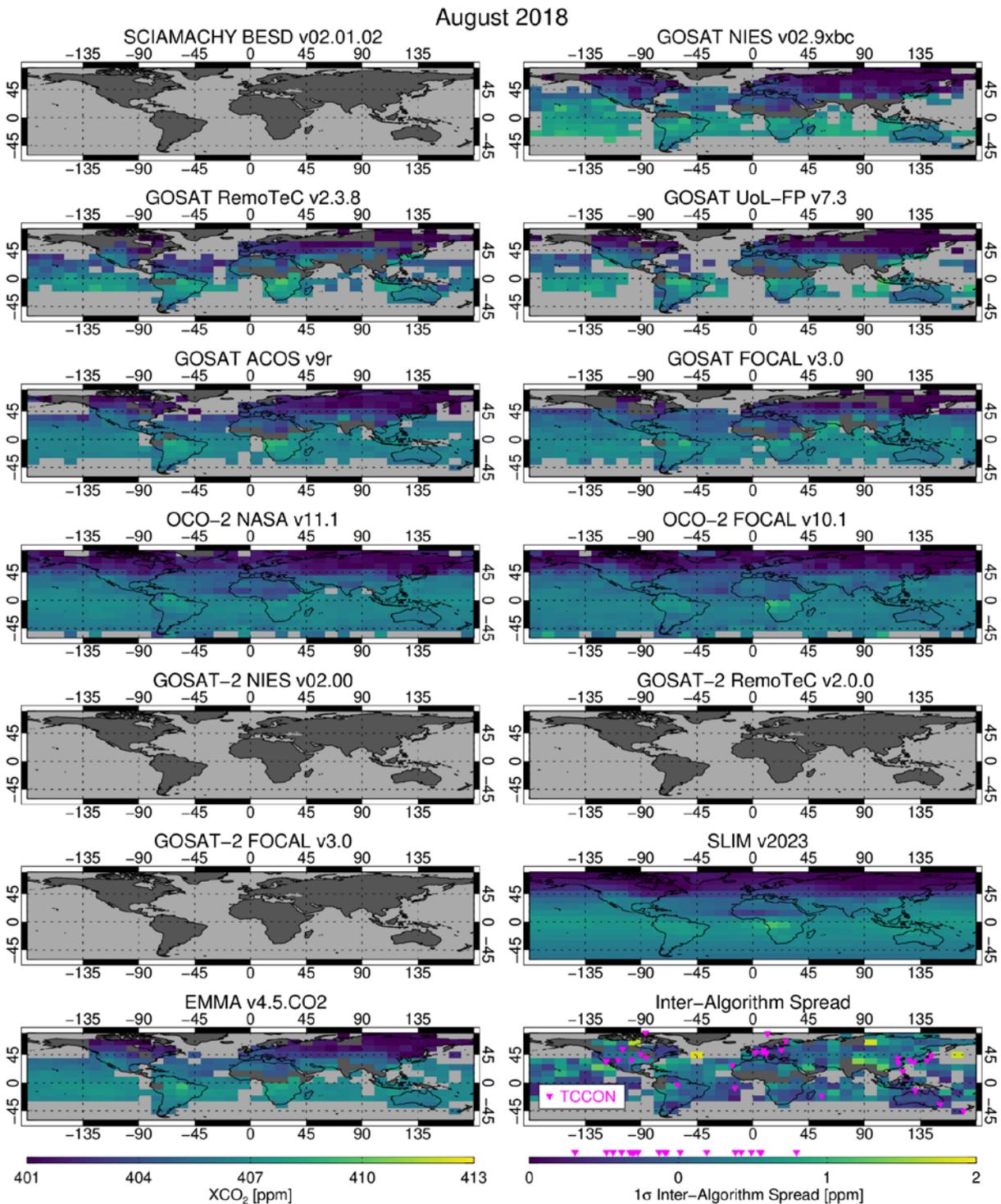


Figure 6: Monthly gridded XCO₂ averages and inter-algorithm spread for the example of August 2018 for EMMA CO₂.

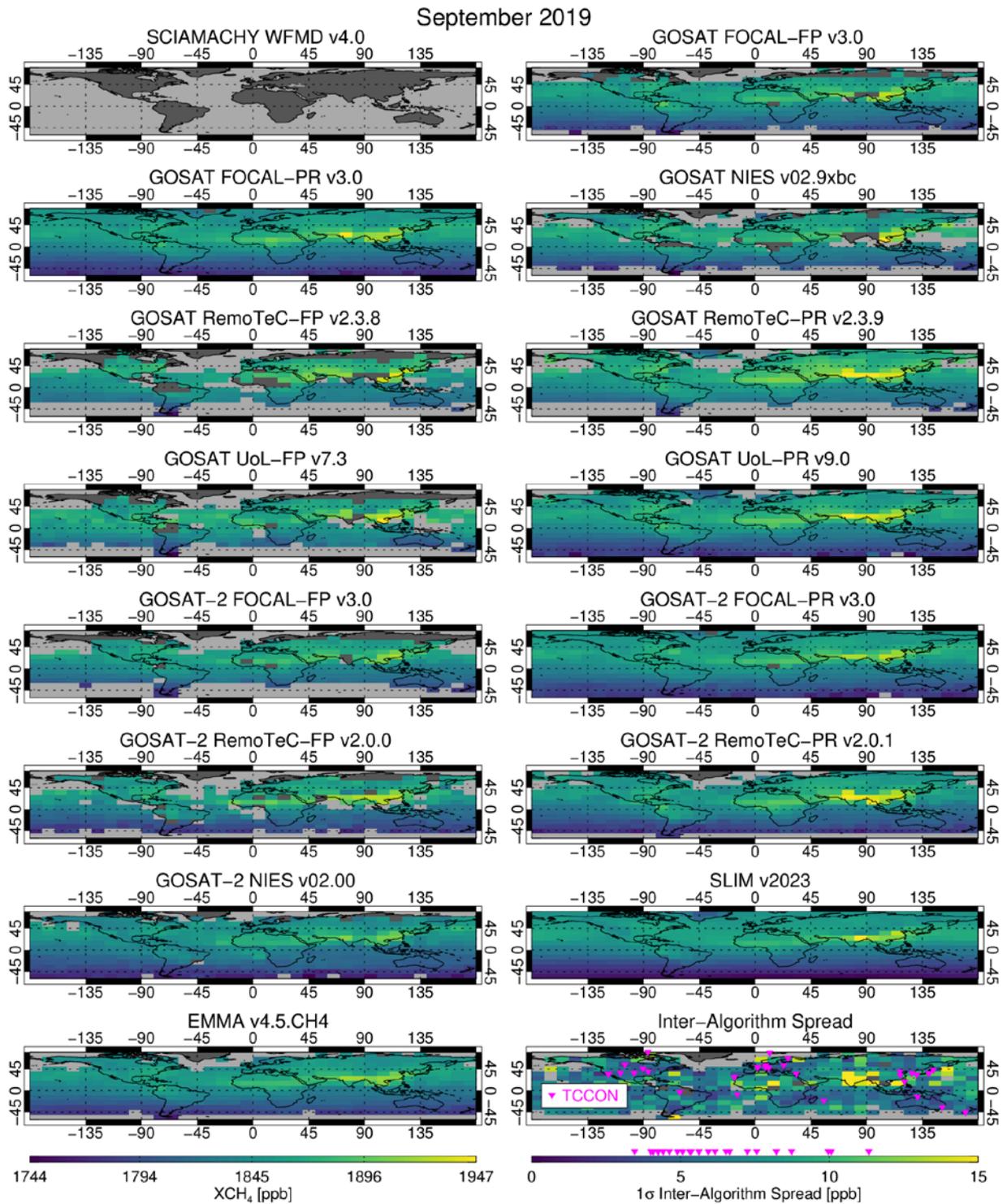


Figure 7: Monthly gridded XCH₄ averages and inter-algorithm spread for the example of September 2019 for EMMA CH₄.

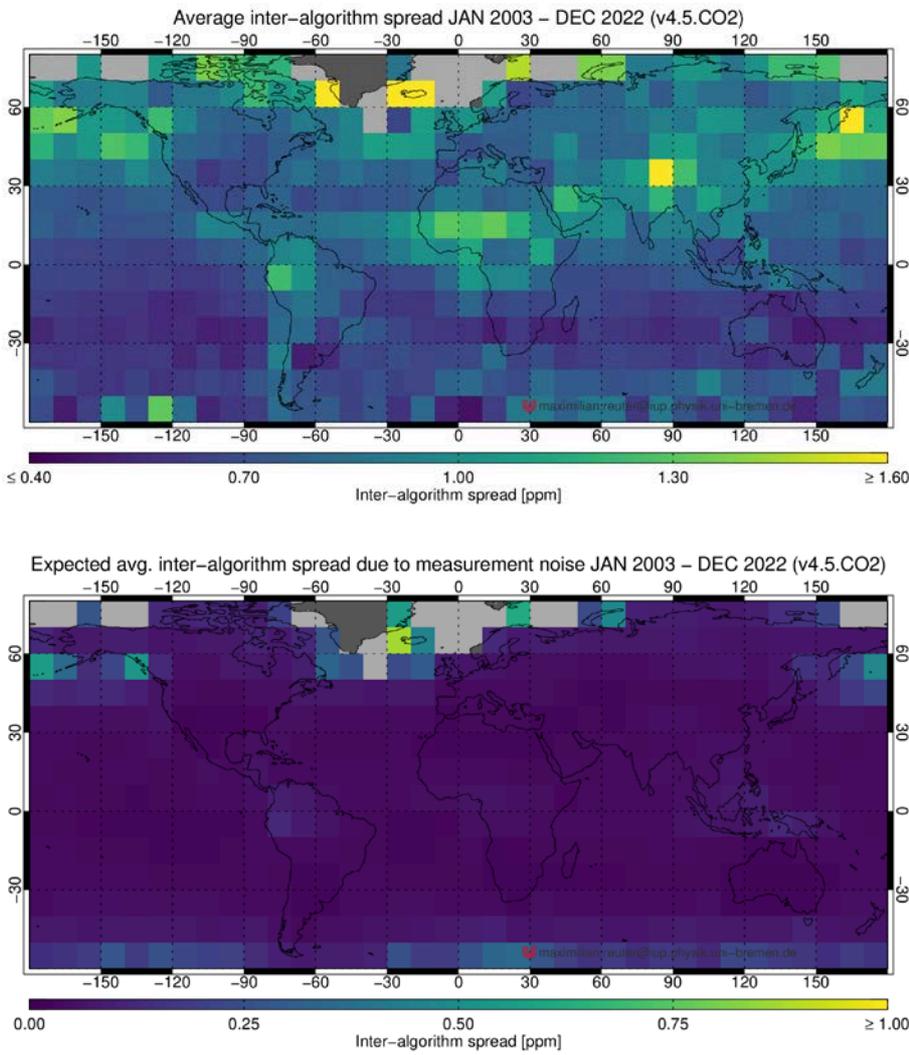


Figure 8: Average inter-algorithm spread (01/2003 – 12/2022) (top) and expected average inter-algorithm spread due to measurement noise (bottom) for EMMA CO₂.

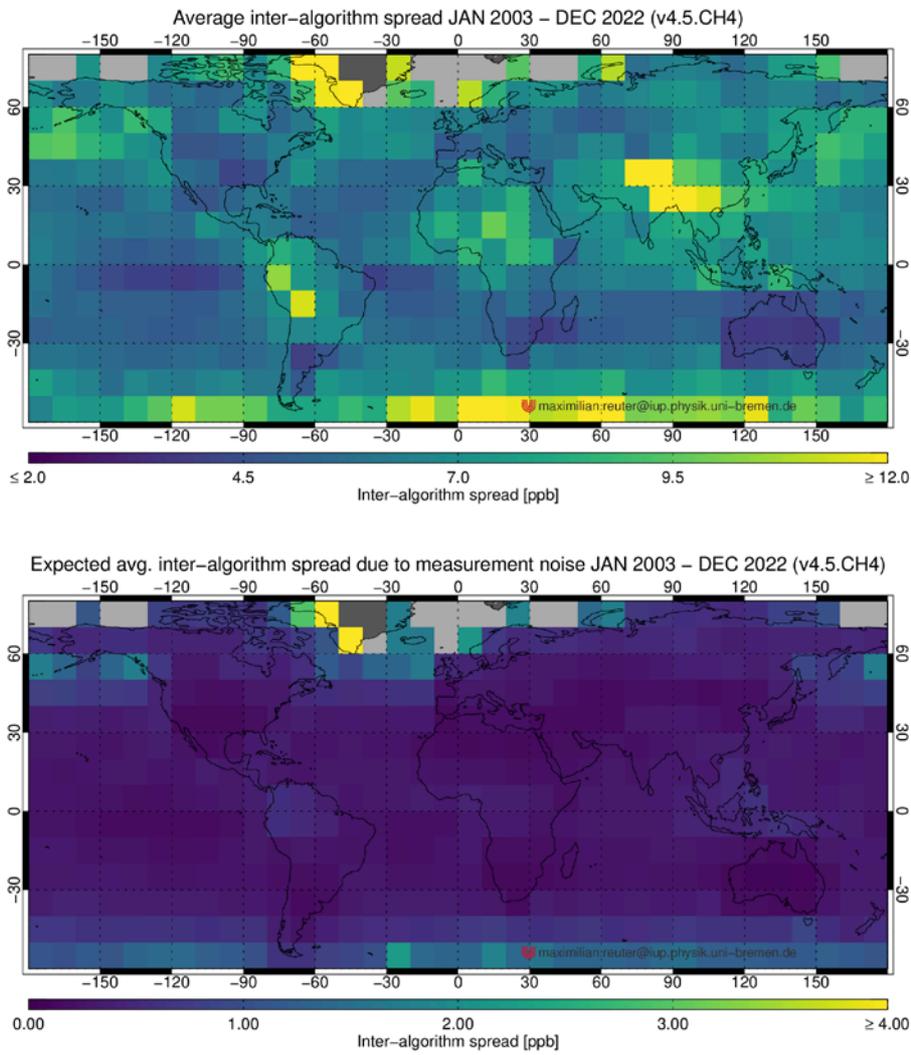


Figure 9: Average inter-algorithm spread (01/2003 – 12/2022) (top) and expected average inter-algorithm spread due to measurement noise (bottom) for EMMA CH₄.



3.1.3 EMMA ensemble median

As described in Section 3.1.2, XCO₂ or XCH₄ averages (one for each algorithm) are calculated within each grid box. However, now, we are aiming to use the ensemble not only to assess regional and temporal uncertainties but also to create a data set which is potentially less influenced by regional or temporal biases. This could be achieved, for example, by building the average, a weighted average, or the median in each grid box.

In this context, the median has some advantages: outliers are assumed to be seldom and there is a high chance that a grid box includes no or only one outlying algorithm. Therefore, cancellation of errors cannot be expected by calculating the average. The median is much less sensitive to such individual outliers. Additionally, the median calculates no new quantity from the individuals of an ensemble, it is rather a procedure to select one specific ensemble member.

This allows us to trace back from level 3 averages to individual level 2 soundings. Essentially, there are five possible scenarios for median calculation within one grid box: (i) All algorithms perform well and scatter slightly around the true XCO₂ or XCH₄ value. In this case the median will help to reduce scatter. (ii) The minority of algorithms produce outliers so that the median is influenced only marginally. (iii) The majority of algorithms produce outliers in different directions. Here, it is still likely that the median falls on a well performing algorithm in the “middle”. (iv) The majority of algorithms produce outliers in the same direction. This is the only case where the median is a bad choice, because it would select an outlier and ignore a well performing algorithm. As discussed in the previous section, we assume that the algorithms within one grid box are often realistic with uncorrelated occasional outliers, which makes this case unlikely to happen often. (v) If all algorithms are outlying, the median is not better or worse than selecting any other ensemble member.

We calculate the median only in grid boxes where reliable average XCO₂ or XCH₄ values can be computed for at least as many algorithms as specified in Figure 4 (gray area). In case of an even number of values, we define the median as that value being closer to the mean. We then trace back to the individual level 2 data, which were used to calculate that average being the median. Together, with all information needed for inverse modeling (geo-location, time, averaging kernels, etc.), these soundings are stored in the EMMA L2 database.

Some algorithms may provide considerably larger amounts of level 2 data (e.g., the NASA or FOCAL OCO-2 algorithm) than other algorithms. In order to prevent over-weighting these algorithms, we limit the maximum number of data points (per grid box). Therefore, we calculate the standard error of the mean of each successfully determined average. The idea behind this is that the lower the standard error of the mean, the larger the potential constraint on an inverse model becomes. If the standard error of the mean of the selected algorithm in a grid box is lower than $1/\sqrt{2}$ times the 25% percentile of all algorithms, the data points are randomly thinned accordingly. In this way, the number of data points can still be rather different but the potential constraint on an inverse model becomes similar.



3.2 Obs4MIPs algorithm

The L3 data products XCO₂_OBS4MIPS and XCH₄_OBS4MIPS are generated by spatial (5°x5°) and temporal (monthly) gridding of the corresponding EMMA L2 databases. The gridding is based on arithmetic unweighted averaging of all soundings falling in a grid box. For each grid box, we compute the standard error of the mean by

$$\bar{\sigma} = \frac{1}{n} \sqrt{\sum \sigma_i^2} \quad (1)$$

where n is the number of soundings within the grid box and σ_i the (corrected) reported stochastic uncertainties of the soundings. In order to reduce noise within the level 3 product, we filter out grid boxes with $n \leq 1$ and $\bar{\sigma} > 1.6\text{ppm}$ for XCO₂ or $\bar{\sigma} > 12\text{ppb}$ for XCH₄, respectively.

Beside XCO₂ or XCH₄, the final level 3 product also includes the number of soundings used for averaging, the average column averaging kernel, the average a priori profile, the standard deviation of the averaged XCO₂ or XCH₄ values, and an estimate for the total uncertainty

$$\hat{\sigma} = \sqrt{\bar{\sigma}^2 + \sigma_s^2}. \quad (2)$$

Here, σ_s represents the inter-algorithm spread computed by EMMA averaged over the soundings within a grid box. For cases including only one algorithm, σ_s is replaced by quadratically adding the estimate of the spatial and seasonal accuracy determined from the TCCON validation (see D2, Table 2). This is only the case during the SCIAMCHY-only period at the beginning of the time series (see Figure 4).



4 Output data

4.1 EMMA products

The EMMA L2 data products consist of individual soundings retrieved by algorithms which can change from grid box to grid box and month to month. They include all the common parameters specified in the Product User Guide and Specification (PUGS) main document (D3) such as XCO₂/XCH₄ (and uncertainty), column averaging kernel profile, a priori profile, pressure level profile, time, longitude, latitude, and quality flag. The EMMA L2 products are in line with the file format (NetCDF-4 classic) and file name convention specified in the PUGS main document (D3).

Therefore, EMMA L2 data products can be used in the same manner as any other XCO₂ or XCH₄ satellite retrieval data product. Especially, the EMMA databases include all information needed for the purpose of inverse modeling (geo-location, time, XCO₂ or XCH₄, averaging kernels, etc.).

In addition to the common parameters, the EMMA data products also include for each sounding information on the selected L2 algorithm, the inter algorithm spread which can serve as an estimate for potential regional uncertainties, the standard error of the median uncertainty, the inter-algorithm spread as expected from measurement noise, potential spatio/temporal biases estimated from TCCON co-locations, and the number of L2 algorithms contributing to the median calculation. More information can be found in the EMMA related PUGS annex (D4).

In order to illustrate how much the individual L2 algorithms contribute to the EMMA databases, Figure 10 shows the relative data weight of each algorithm (defined as $\sum 1/\sigma_i^2$ normalized to one) and the number of soundings within the EMMA database per month.

For the example of April 2019, Figure 11 shows the EMMA XCO₂ and XCH₄ values as well as the corresponding selected median algorithm.

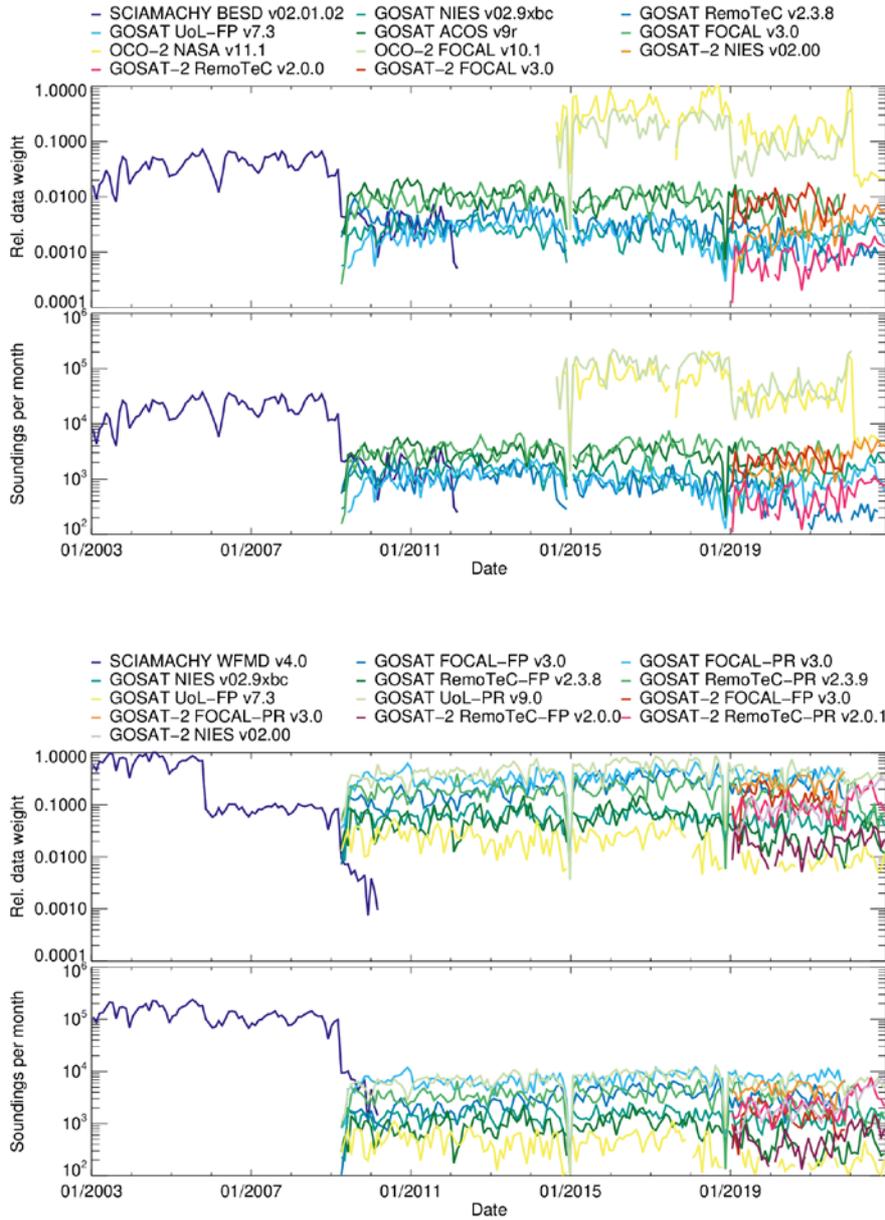


Figure 10: EMMA normalized relative data weight proportional to $\sum \mathbf{1}/\sigma_i^2$ and number of soundings per algorithm and month for CO₂ (top) and CH₄ (bottom).

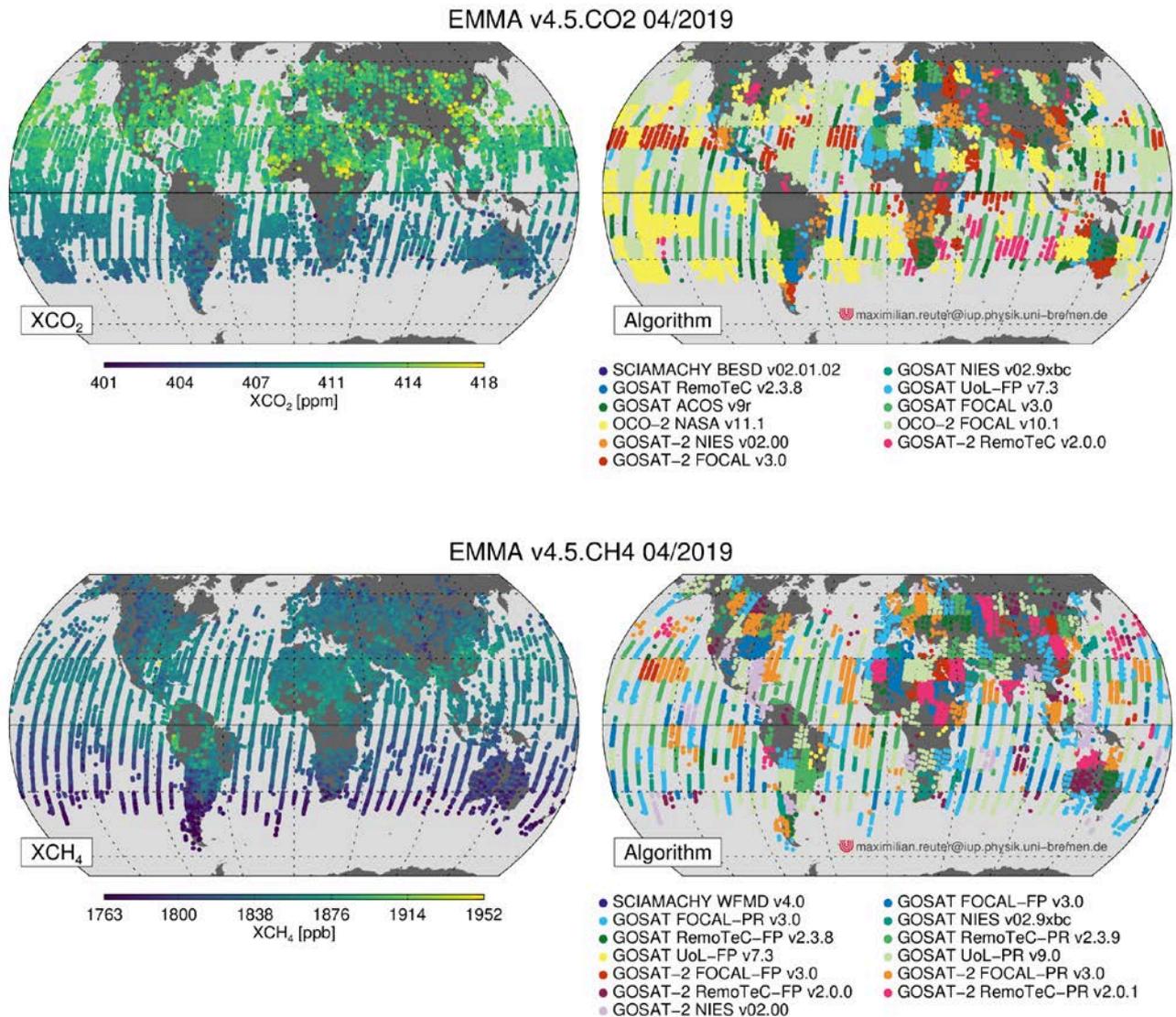


Figure 11: EMMA L2 XCO₂ and XCH₄ (left) and corresponding selected algorithm (right) for EMMA CO₂ for the example of April 2019 (top) and EMMA CH₄ for the example of April 2019 (bottom).

4.2 Obs4MIPs products

The XCO₂_OBS4MIPs and XCH₄_OBS4MIPs products consist of monthly, 5°x5° gridded level 3 XCO₂ or XCH₄ data computed from the corresponding EMMA databases. Additionally, the output files include gridded information about the number of averaged soundings, column averaging kernels, a priori profiles, standard deviation of XCO₂ or XCH₄, and an estimate of the total uncertainty accounting for measurement noise plus potential spatial and/or temporal biases.

For the example of August 2015, Figure 12 shows the Obs4MIPs XCO₂ or Obs4MIPs XCH₄ values as well as the corresponding total uncertainty.



The format of the output data (Obs4MIPs NetCDF) and the file name convention is described in the separate Product User Guide and Specification (PUGS) main document (D3). A list and description of the included EMMA specific parameters can be found in the EMMA related PUGS annex (D4).

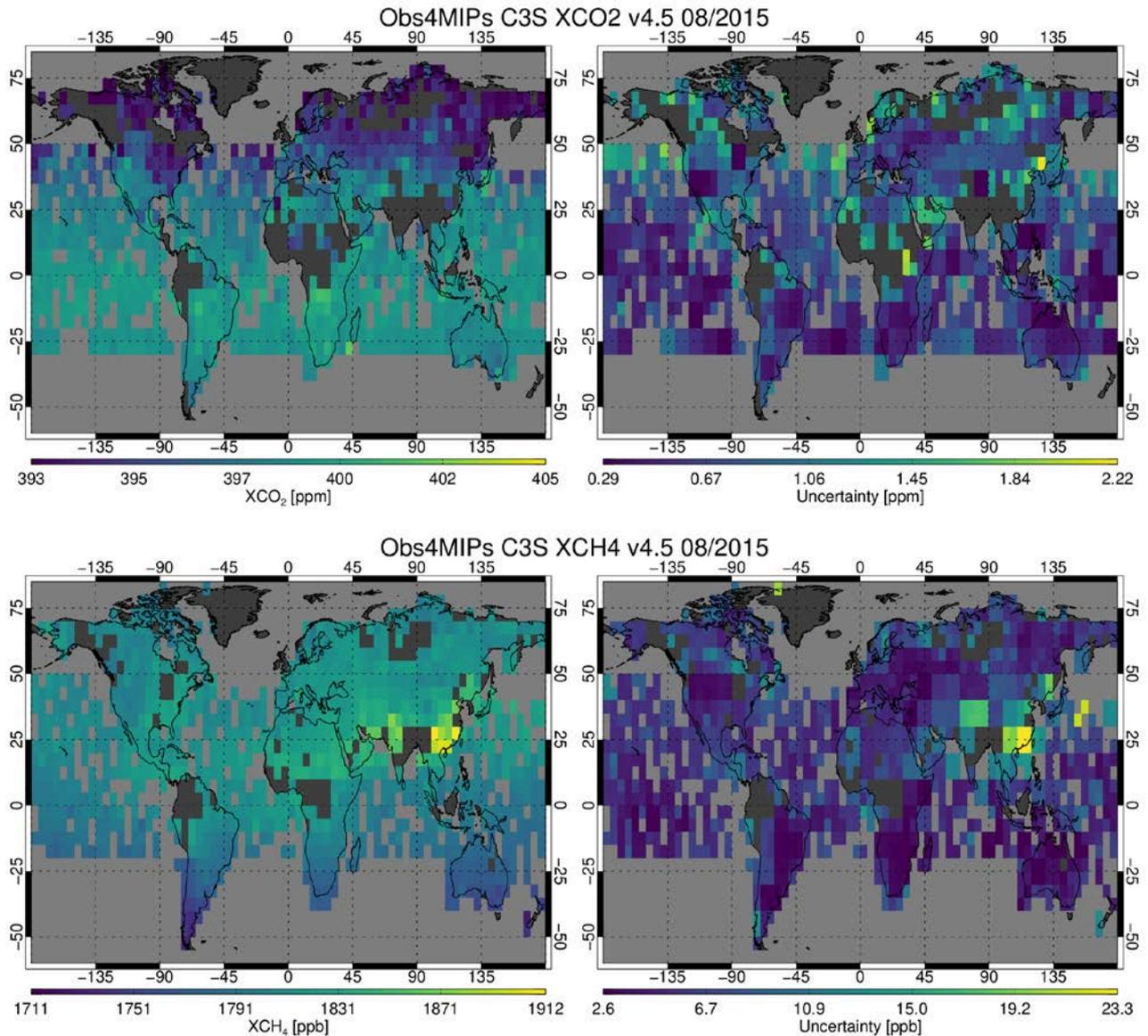


Figure 12: Top: XCO₂_OBS4MIPS for August 2015 (left) and its uncertainty computed from the retrieval noise and EMMA’s inter-algorithm spread (right). Bottom: Same for XCH₄_OBS4MIPS.



5 References

- Araujo and New, 2007:** Araujo, M. B. and New, M.: Ensemble forecasting of species distributions, *Trends Ecol. Evol.*, 22, 42–47, doi:10.1016/j.tree.2006.09.010, 2007.
- Boesch and Anand, 2017:** H. Boesch and J. Anand, Algorithm Theoretical Basis Document (ATBD) – ANNEX A for products CO₂_GOS_OCFP, CH₄_GOS_OCFP & CH₄_GOS_OCP, Copernicus Climate Change Service (C3S) project on satellite-derived Essential Climate Variable (ECV) Greenhouse Gases (CO₂ and CH₄) data products (project C3S_312a_Lot6), Version 1 (21/08/2017), 2017.
- Bojinski et al., 2014:** Bojinski, S., M. Verstraete, T. C. Peterson, C. Richter, A. Simmons and M. Zemp, The concept of essential climate variables in support of climate research, applications and policy, *Bulletin of the American Meteorological Society (BAMS)*, 1431-1443, September 2014.
http://www.wmo.int/pages/prog/gcos/documents/bams_ECV_article.pdf
- Bovensmann et al., 1999:** Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A.: SCIAMACHY –Mission Objectives and Measurement Modes, *J. Atmos. Sci.*, 56, 127–150, 1999.
- Buchwitz et al., 2023c:** Buchwitz, M., et al., Algorithm Theoretical Basis Document (ATBD) – Main document for Greenhouse Gas (GHG: CO₂ & CH₄) data set CDR 7 (2003-2022), project C3S2_312a_Lot2_DLR – Atmosphere, v7.0, 2023.
- Burrows et al., 1995:** Burrows, J. P., Hölzle, E., Goede, A. P. H., Visser, H., and Fricke, W.: SCIAMACHY – Scanning Imaging Absorption Spectrometer for Atmospheric Chartography, *Acta Astronaut.*, 35, 445–451, 1995.
- Butz et al., 2011:** Butz, A., Guerlet, S., Hasekamp, O., Schepers, D., Galli, A., Aben, I., Frankenberg, C., Hartmann, J.-M., Tran, H., Kuze, A., Keppel-Aleks, G., Toon, G., Wunch, D., Wennberg, P., Deutscher, N., Griffith, D., Macatangay, R., Messerschmidt, J., Notholt, J., and Warneke, T.: Toward accurate CO₂ and CH₄ observations from GOSAT, *Geophys. Res. Lett.*, 38, L14812, <https://doi.org/10.1029/2011GL047888>, 2011
- Chevallier et al., 2007:** Chevallier, F., Bréon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO₂ sources and sinks: Theoretical study in a variational data assimilation framework, *J. Geophys. Res.*, 112, D09307, doi:10.1029/2006JD007375, 2007.
- Cogan et al., 2012:** Cogan, A. J., Boesch, H., Parker, R. J., Feng, L., Palmer, P. I., Blavier, J.-F. L., Deutscher, N. M., Macatangay, R., Notholt, J., Roehl, C., Warneke, T., and Wunch, D.: Atmospheric carbon dioxide retrieved from the Greenhouse gases Observing SATellite (GOSAT): Comparison with ground-based TCCON observations and GEOS-Chem model calculations, *J. Geophys. Res. Atmos.*, 117, D21301, <https://doi.org/10.1029/2012JD018087>, 2012
- Crisp et al., 2017:** Crisp, D.; Pollock, H.R.; Rosenberg, R.; Chapsky, L.; Lee, R.A.M.; Oyafuso, F.A.; Frankenberg, C.; O'Dell, C.W.; Bruegge, C.J.; Doran, G.B.; et al. The on-orbit performance of the Orbiting Carbon Observatory-2 (OCO-2) instrument and its radiometrically calibrated products. *Atmos. Meas. Tech.* 2017, 10, 59–81.



Detmers, 2017a: R. Detmers, Algorithm Theoretical Basis Document (ATBD) – ANNEX B for products CO2_GOS_SRFP & CH4_GOS_SRFP, Copernicus Climate Change Service (C3S) project on satellite-derived Essential Climate Variable (ECV) Greenhouse Gases (CO₂ and CH₄) data products (project C3S_312a_Lot6), Version 1 (21/08/2017), 2017

Detmers, 2017b: R. Detmers, Algorithm Theoretical Basis Document (ATBD) – ANNEX C for product CH4_GOS_SRFP, Copernicus Climate Change Service (C3S) project on satellite-derived Essential Climate Variable (ECV) Greenhouse Gases (CO₂ and CH₄) data products (project C3S_312a_Lot6), Version 1 (21/08/2017), 2017

Gier et al., 2020: Gier, B. K., Buchwitz, M., Reuter, M., Cox, P. M., Friedlingstein, P., and Eyring, V.: Spatially resolved evaluation of Earth system models with satellite column-averaged CO₂, *Biogeosciences*, 17, 6115–6144, <https://doi.org/10.5194/bg-17-6115-2020>, 2020.

Hagedorn et al., 2005: Hagedorn, R., Doblas-Reyes, F., and Palmer, T.: The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept, *Tellus A*, 57, 219–233, doi:10.1111/j.1600-0870.2005.00103.x, 2005.

Houweling et al., 2004: Houweling, S., Breon, F.-M., Aben, I., Rönodenbeck, C., Gloor, M., Heimann, M., and Ciais, P.: Inverse modeling of CO₂ sources and sinks using satellite data: a synthetic inter-comparison of measurement techniques and their performance as a function of space and time, *Atmos. Chem. Phys.*, 4, 523–538, doi:10.5194/acp-4-523-2004, 2004.

Kiel et al., 2019: Kiel, M., O’Dell, C. W., Fisher, B., Eldering, A., Nassar, R., MacDonald, C. G., and Wennberg, P. O.: How bias correction goes wrong: measurement of XCO₂ affected by erroneous surface pressure estimates, *Atmos. Meas. Tech.*, 12, 2241–2259, <https://doi.org/10.5194/amt-12-2241-2019>, 2019

Kharin and Zwiers, 2002: Kharin, V. and Zwiers, F.: Climate predictions with multimodel ensembles, *J. Climate*, 15, 793–799, doi:10.1175/1520-0442(2002)015<0793:CPWME>2.0.CO;2, 2002.

Krisna et al., 2021: Trismono Candra Krisna, Ilse Aben, Lianghai Wu, Otto Hasekamp, Jochen Landgraf: ESA Climate Change Initiative “Plus” (CCI+) Algorithm Theoretical Basis Document (ATBD) Version 1.3 – For the RemoTeC XCO₂ and XCH₄ GOSAT-2 SRON Full Physics Products (CO₂_GO₂_SRFP and CH₄_GO₂_SRFP) Version 2.0.0 for the Essential Climate Variable (ECV) Greenhouse Gases (GHG), https://www.iup.uni-bremen.de/carbon_ghg/docs/GHG-CClplus/CRDP7/ATBDv3_GHG-CCI_CO2_CH4_GO2_SRFP_v2p0p0.pdf, 2021.

Kuze et al., 2009: Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T. (2009), Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, *Appl. Opt.*, 48, 6716–6733, 2009.

Kuze et al., 2016: Kuze, A., Suto, H., Shiomi, K., Kawakami, S., Tanaka, M., Ueda, Y., Deguchi, A., Yoshida, J., Yamamoto, Y., Kataoka, F., Taylor, T. E., and Buijs, H. L.: Update on GOSAT TANSO-FTS performance, operations, and data products after more than 6 years in space, *Atmos. Meas. Tech.*, 9, 2445–2461, doi:10.5194/amt-9-2445-2016, 2016.



Miller et al., 2007: Miller, C. E., Crisp, D., DeCola, P. L., Olsen, S. C., Randerson, J. T., Michalak, A. M., Alkhaled, A., Rayner, P., Jacob, D. J., Suntharalingam, P., Jones, D. B. A., Denning, A. S., Nicholls, M. E., Doney, S. C., Pawson, S., Bösch, H., Connor, B. J., Fung, I. Y., O'Brien, D., Salawitch, R. J., Sander, S. P., Sen, B., Tans, P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., and Law, R. M.: Precision requirements for space-based XCO₂ data, *J. Geophys. Res.*, 112, D10314, doi:10.1029/2006JD007659, 2007.

National Research Council, 2004: National Research Council, *Climate Data Records from Environmental Satellites: Interim Report 2004*, 150 pp., ISBN: 978-0-309-09168-8, DOI: <https://doi.org/10.17226/10944>, 2004.

Noël et al., 2022: S. Noël, M. Reuter, M. Buchwitz, J. Borchardt, M. Hilker, O. Schneising, H. Bovensmann, J.P. Burrows, A. Di Noia, R.J. Parker, H. Suto, Y. Yoshida, M. Buschmann, N.M. Deutscher, D.G. Feist, D.W.T. Griffith, F. Hase, R. Kivi, C. Liu, I. Morino, J. Notholt, Y.-S. Oh, H. Ohyama, C. Petri, D.F. Pollard, M. Rettinger, C. Roehl, C. Rousogonous, M.K. Sha, K. Shiomi, K. Strong, R. Sussmann, Y. Té, V.A. Velazco, M. Vrekoussis, and T. Warneke: Retrieval of greenhouse gases from GOSAT and GOSAT-2 using the FOCAL algorithm, *Atmos. Meas. Tech.*, 15, 3401-3437, <https://doi.org/10.5194/amt-15-3401-2022>, 2022.

O'Dell et al., 2012: O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: The ACOS CO₂ retrieval algorithm – Part 1: Description and validation against synthetic observations, *Atmos. Meas. Tech.*, 5, 99–121, doi:10.5194/amt-5-99-2012, 2012.

Parkinson et al., 2006: Parkinson, C., A. Ward and M. King, *Earth Science Reference Handbook*, NASA, Washington DC, 2006.

Rayner and O'Brien, 2001: Rayner, P. J. and O'Brien, D. M.: The utility of remotely sensed CO₂ concentration data in surface inversions, *Geophys. Res. Lett.*, 28, 175–178, 2001.

Reuter et al., 2010: M. Reuter, M. Buchwitz, O. Schneising, J. Heymann, H. Bovensmann, J. P. Burrows: A method for improved SCIAMACHY CO₂ retrieval in the presence of optically thin clouds. *Atmospheric Measurement Techniques*, 3, 209-232, 2010.

Reuter et al., 2011: 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, D. Wunch: Retrieval of atmospheric CO₂ with enhanced accuracy and precision from SCIAMACHY: Validation with FTS measurements and comparison with model results. *Journal of Geophysical Research - Atmospheres*, 116, D04301, doi: 10.1029/2010JD015047, 2011.

Reuter et al., 2013: M. Reuter, H. Bösch, H. Bovensmann, A. Bril, M. Buchwitz, A. Butz, J. P. Burrows, C. W. O'Dell, S. Guerlet, O. Hasekamp, J. Heymann, N. Kikuchi, S. Oshchepkov, R. Parker, S. Pfeifer, O. Schneising, T. Yokota, and Y. Yoshida: A joint effort to deliver satellite retrieved atmospheric CO₂ concentrations for surface flux inversions: the ensemble median algorithm EMMA. *Atmospheric Chemistry and Physics*, doi:10.5194/acp-13-1771-2013, 13, 1771-1780, 2013.



Reuter et al., 2016: M. Reuter, H. Bovensmann, M. Buchwitz, J. P. Burrows, J. Heymann, O. Schneising: Algorithm Theoretical Basis Document Version 5 (ATBDv5) - The Bremen Optimal Estimation DOAS (BESD) algorithm for the retrieval of XCO₂ for the Essential Climate Variable (ECV) Greenhouse Gases (GHG), 2016.

Reuter et al., 2017a: M.Reuter, M.Buchwitz, O.Schneising, S.Noël, V.Rozanov, H.Bovensmann and J.P.Burrows: A Fast Atmospheric Trace Gas Retrieval for Hyperspectral Instruments Approximating Multiple Scattering - Part 1: Radiative Transfer and a Potential OCO-2 XCO₂ Retrieval Setup, *Remote Sensing*, 9(11), 1159; doi:10.3390/rs9111159, 2017.

Reuter et al., 2017b: M.Reuter, M.Buchwitz, O.Schneising, S.Noël, H.Bovensmann and J.P.Burrows: A Fast Atmospheric Trace Gas Retrieval for Hyperspectral Instruments Approximating Multiple Scattering - Part 2: Application to XCO₂ Retrievals from OCO-2, *Remote Sensing*, 9(11), 1102; doi:10.3390/rs9111102, 2017.

Reuter et al., 2020: M. Reuter, M. Buchwitz, O. Schneising, S. Noël, H. Bovensmann, J.P. Burrows, H. Boesch, A. Di Noia, J. Anand, R.J. Parker, P. Somkuti, L. Wu, O.P. Hasekamp, I. Aben, A. Kuze, H. Suto, K. Shiomi, Y. Yoshida, I. Morino, D. Crisp, C.W. O'Dell, J. Notholt, C. Petri, T. Warneke, V.A. Velazco, N.M. Deutscher, D.W.T. Griffith, R. Kivi, D.F. Pollard, F. Hase, R. Sussmann, Y.V. Té, K. Strong, S. Roche, M.K. Sha, M. De Mazière, D.G. Feist, L.T. Iraci, C.M. Roehl, C. Retscher, and D. Schepers: Ensemble-based satellite-derived carbon dioxide and methane column-averaged dry-air mole fraction data sets (2003-2018) for carbon and climate applications, *Atmos. Meas. Tech.*, <https://www.atmos-meas-tech.net/13/789/2020>, 2020.

Reuter et al., 2021: M. Reuter, M. Hilker, S. Noël, M. Buchwitz, O. Schneising, H. Bovensmann, and J. P. Burrows: ESA Climate Change Initiative “Plus” (CCI+) Algorithm Theoretical Basis Document Version 3 (ATBDv3) - Retrieval of XCO₂ from the OCO-2 satellite using the Fast Atmospheric Trace Gas Retrieval (FOCAL) for the Essential Climate Variable (ECV) Greenhouse Gases (GHG), http://www.iup.uni-bremen.de/carbon_ghg/docs/GHG-CCIplus/CRDP7/ATBDv3_GHG-CCI_CO2_OC2_FOCA_v10.pdf, 2021

Rötter et al., 2011: Rötter, R. P., Carter, T. R., Olesen, J. E., and Porter, J. R.: Crop climate models need an overhaul, *Nat. Clim. Change*, 1, 175–177, 2011.

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

Schneising et al. (2018): O. Schneising and the ESA CCI GHG project team: ESA Greenhouse Gases Climate Change Initiative (GHG_cci): Column-averaged CH₄ from SCIAMACHY generated with the WFMD algorithm (CH₄_SCI_WFMD), version 4.0. Centre for Environmental Data Analysis, date of citation. <https://catalogue.ceda.ac.uk/uuid/aa09603e91b44f3cb1573c9dd415e8a8>, 2018

Stephens et al., 2007: Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., Ciais, P., Ramonet, M., Bousquet, P., Nakazawa, T., Aoki, S., Machida, T., Inoue, G., Vinnichenko, N., Lloyd, J., Jordan, A., Heimann, M., Shibistova, O., Langenfelds, R. L., Steele, L. P., Francey, R. J., and Denning, A. S.: Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO₂, *Science*, 316, 1732–1735, doi:10.1126/science.1137004, 2007



Suto et al., 2021: Suto, H., Kataoka, F., Kikuchi, N., Knuteson, R. O., Butz, A., Haun, M., Buijs, H., Shiomi, K., Imai, H., and Kuze, A.: Thermal and nearinfrared sensor for carbon observation Fourier transform spectrometer-2 (TANSO-FTS-2) on the Greenhouse gases Observing SATellite-2 (GOSAT-2) during its first year in orbit, *Atmos. Meas. Tech.*, 14, 2013–2039, <https://doi.org/10.5194/amt-14-2013-2021>, 2021

Taylor et al., 2022: Taylor, T. E., O'Dell, C. W., Crisp, D., Kuze, A., Lindqvist, H., Wennberg, P. O., Chatterjee, A., Gunson, M., Eldering, A., Fisher, B., Kiel, M., Nelson, R. R., Merrelli, A., Osterman, G., Chevallier, F., Palmer, P. I., Feng, L., Deutscher, N. M., Dubey, M. K., Feist, D. G., García, O. E., Griffith, D. W. T., Hase, F., Iraci, L. T., Kivi, R., Liu, C., De Mazière, M., Morino, I., Notholt, J., Oh, Y.-S., Ohyama, H., Pollard, D. F., Rettinger, M., Schneider, M., Roehl, C. M., Sha, M. K., Shiomi, K., Strong, K., Sussmann, R., Té, Y., Velazco, V. A., Vrekoussis, M., Warneke, T., and Wunch, D.: An 11-year record of XCO₂ estimates derived from GOSAT measurements using the NASA ACOS version 9 retrieval algorithm, *Earth Syst. Sci. Data*, 14, 325–360, <https://doi.org/10.5194/essd-14-325-2022>, 2022

Tebaldi, C. and Knutti, 2007: Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, *Philos. Trans. R. Soc. A*, 365, 2053–2075, doi:10.1098/rsta.2007.2076, 2007.

Vautard et al., 2009: Vautard, R., Schaap, M., Bergstrom, R., Bessagnet, B., Brandt, J., Builtjes, P. J. H., Christensen, J. H., Cuvelier, C., Foltescu, V., Graff, A., Kerschbaumer, A., Krol, M., Roberts, P., Rouil, L., Stern, R., Tarrason, L., Thunis, P., Vignati, E., and Wind, P.: Skill and uncertainty of a regional air quality model ensemble, *Atmos. Environ.*, 43, 4822–4832, doi:10.1016/j.atmosenv.2008.09.083, 2009.

Werschek, 2015: Werschek, M., EUMETSAT Satellite Application Facility on Climate Monitoring, C3S Climate Data Store workshop, Reading, UK, 3-6 March 2015.

<http://www.ecmwf.int/sites/default/files/elibrary/2015/13546-existing-solutions-eumetsat-satellite-application-facility-climate-monitoring.pdf>

Wunch et al., 2011: Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network (TCCON), *Phil. Trans. R. Soc. A*, 369, 2087–2112, doi:10.1098/rsta.2010.0240, 2011.

Yokota et al., 2004: Yokota, T., Oguma, H., Morino, I., and Inoue, G.: A nadir looking SWIR sensor to monitor CO₂ column density for Japanese GOSAT project, *Proceedings of the twenty-fourth international symposium on space technology and science*, Miyazaki: Japan Society for Aeronautical and Space Sciences and ISTS, 887–889, 2004.

Yoshida et al., 2013: Yoshida, Y., Kikuchi, N., Morino, I., Uchino, O., Oshchepkov, S., Bril, A., Saeki, T., Schutgens, N., Toon, G. C., Wunch, D., Roehl, C. M., Wennberg, P. O., Griffith, D. W. T., Deutscher, N. M., Warneke, T., Notholt, J., Robinson, J., Sherlock, V., Connor, B., Rettinger, M., Sussmann, R., Ahonen, P., Heikkinen, P., Kyrö, E., Mendonca, J., Strong, K., Hase, F., Dohe, S., and Yokota, T.: Improvement of the retrieval algorithm for GOSAT SWIR XCO₂ and XCH₄ and their validation using TCCON data, *Atmos. Meas. Tech.*, 6, 1533–1547, <https://doi.org/10.5194/amt-6-1533-2013>, 2013



Yoshida and Oshio, 2020: Y. Yoshida and H. Oshio: GOSAT-2 TANSO-FTS-2 SWIR L2 Retrieval Algorithm Theoretical Basis Document, National Institute for Environmental Studies, GOSAT-2 Project https://prdct.gosat-2.nies.go.jp/documents/pdf/ATBD_FTS-2_L2_SWL2_en_00.pdf, 2020

Copernicus Climate Change Service

ECMWF - Shinfield Park, Reading RG2 9AX, UK

Contact: <https://support.ecmwf.int/>

climate.copernicus.eu

copernicus.eu

ecmwf.int