ESA Climate Change Initiative "Plus" (CCI+)

Product User Guide (PUG) Version 5.0

-

For the RemoTeC GOSAT-2 SRON Full-Physics Products: XCO2 (CO2_GO2_SRFP) and XCH4 (CH4_GO2_SRFP)

Version 2.0.3

for the Essential Climate Variable (ECV)

Greenhouse Gases (GHG)

Written by:

GHG-CCI+ project team

Lead author:

Andrew Gerald Barr

(SRON Netherlands Institute for Space Research, Leiden, The Netherlands)

Co-authors:

Tobias Borsdorff, Jochen Landgraf

(SRON Netherlands Institute for Space Research, Leiden, The Netherlands)

Change log:

Version Nr.	Date	Status	Reason for change
Version 1	27. Oct. 2020	Draft	New document
Version 1.1	04. Jan. 2021	As submitted	- Update format - Update purpose of document
Version 1.1	04. Feb. 2021	As submitted	- Update after ESA reviews - Remote typos
Version 2.0	04. Nov. 2021	As submitted	- L2 data reprocessing: update filter criteria, selection of TCCON station, and bias correction
Version 3.0	27. Jan. 2022	As submitted	- Updated doc to version 3.0
Version 4.0	15. Aug. 2023	As submitted	- Update doc to version 4.0 - Quality filtering via random forest model prediction
Version 5.0	11. July 2024	As submitted	- Update doc to version 5.0 - New product version v2.0.3 - Merge of CO2 and CH4 Full physics products into one document - Timeseries extended to end of 2023

Table of Contents

1. Purpose of document

This document describes the Product User Guide (PUG) of the RemoTeC XCH_4 and XCO_2 GOSAT-2 SRON Full-Physics Products (CH4_GO2_SRFP & CO2_GO2_SRFP), which is a deliverable for the ESA GHG-CCI+ project led by University of Bremen, Germany.

Within the project, satellite-derived atmospheric Carbon Dioxide (CO_2) and Methane (CH_4) Essential Climate Variable (ECV) data products are generated and delivered to ESA for inclusion into the ESA-GHG-CCI+ database from which users can access these data products and the corresponding documentations.

The satellite-derived data products are:

- Column-averaged 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.

This document will be focused on the Level-2 products retrieved using the GOSAT-2 Full-Physics algorithm developed by SRON Netherlands Institute for Space Research, The Netherlands.

2. Greenhouse gases Observing SATellite-2 (GOSAT-2)

The Japanese Greenhouse gases Observing SATellite-2 (GOSAT2) was launched on 29th October 2018 and started operational observations form February 2019. GOSAT2 provides dedicated global measurements of total column CO_2 and CH_4 from its SWIR bands. It is equipped with two instruments, the Thermal And Near Infrared Sensor for carbon Observations - Fourier Transform Spectrometer-2 (TANSO-FTS2) as well as a dedicated Cloud and Aerosol Imager-2 (TANSO-CAI-2).

The TANSO-FTS2 instrument (Nakajima et al., 2017) has five spectral bands with a high spectral resolution 0.2 cm^{-1}. Three operate in the SWIR at 0.75-0.77, 1.56-1.69 and at the extended 1.92-2.33 µm range, providing sensitivity to the near-surface absorbers. The fourth and fifth channels operating in the thermal infrared between 5.5-8.4 and 8.4-14.3 µm providing mid-tropospheric sensitivity.

The measurement strategy of TANSO-FTS2 is optimized for the characterization of continental-scale sources and sinks. TANSO-FTS2 utilizes a pointing mirror to perform off-nadir measurements at the same location on each 6-day repeat cycle. The pointing mirror allows TANSO-FTS2 to observe up to ±35° across track and ±40° along-track. These measurements nominally consist of 5 across track points spaced ~160km apart with a ground footprint diameter of approximately 9.7 km and a 4 second exposure duration. The satellite has an intelligent pointing monitor camera which makes it possible to adjust the line of sight of the FTS to steer away from cloud contaminated areas. Whilst the majority of data is limited to measurements over land where the surface reflectance is high, TANSO-FTS2 also observes in sun-glint mode over the ocean.

3. RemoTeC retrieval algorithm

The Full Physics products are retrieved from GOSAT-2 TANSO-FTS spectra using the RemoTeC algorithm, that has been developed jointly by SRON and Karlsruhe Institute of Technology (KIT). The algorithm retrieves simultaneously XCH_4 and XCO_2. For the retrieval, we analyze four spectral regions, the 0.77 µm oxygen band, two CO_2 bands at 1.61 and 2.06 µm, as well as a CH_4 band at 1.64 µm. Within the retrieval procedure, the sub-columns of CO_2 and CH_4 in different altitude layers are being retrieved. To obtain the column averaged dry air mixing ratios XCO_2 and XCH_4 the sub-columns are summed up to get the total column, which is divided by the dry-air columns obtained from ECMWF model data in combination with a surface elevation data base.

The retrieved XCO_2 has been validated against ground based measurements from the Total Column Carbon Observing Network (TCCON). To further improve accuracy of XCO_2 product, a bias correction has been developed based on TCCON comparisons. We use the GGG2020 release of the TCCON data (Wunch et al., 2015, Laughner et al. 2021). More details on the technical aspects of the retrievals can be found in the ATBD GO2-SRFP document (Barr et al. 2023).

4. XCO_2 RemoTeC FP data product (Feb. 2019 – Dec 2023)

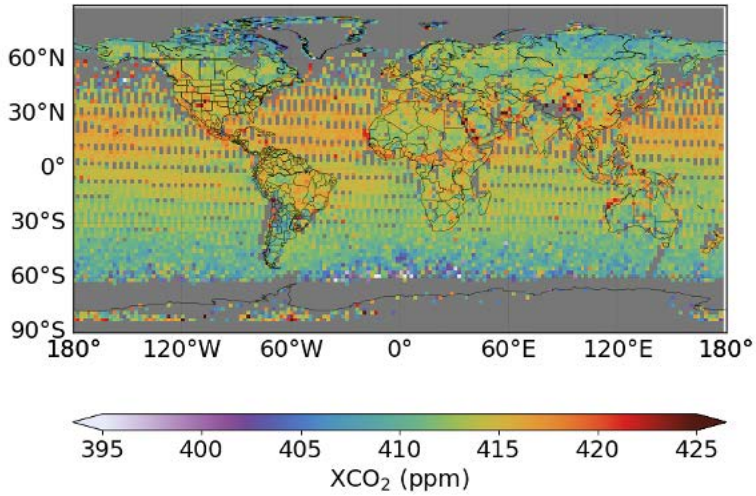
In this section, we show examples of the GOSAT-2 XCO_2 FP data product by showing global averaged maps (Sec. 4.1) and by giving a summary of the validation results relative to TCCON (Sec. 4.2).

4.1 Global maps

Figure 1 shows global average maps of the RemoTeC GOSAT-2 FP XCO_2 data product, with the bias-corrected XCO_2 data and the scaled random error (Barr et al. 2024). The GOSAT-2 XCO_2 product provides good global spatial coverage. As can be seen, in some regions the coverage is limited by cloud cover (the observations correspond to cloud free scenes), sun illumination conditions, etc. The error on XCO_2 varies from around 0.7 ppm to 5 ppm and is higher for high latitude regions and low albedo areas.

4.2 Validation with TCCON

This section summarizes the main validation results presented in the RemoTeC GOSAT-2 ESA GHG CCI+ End-to-End ECV Uncertainty Budget (E3UB) version 5.0 document (Barr et al. 2024). We used ground based TCCON GGG2020 (Laughner et al., 2021) data obtained from https://tccondata.org/ as reference data set. We co-located GOSAT-2 and TCCON measurements with a maximum time difference of 2.5h, a maximum distance of 300 km in both longitudinal and latitudinal directions. In cases of multiple TCCON measurements of the same site collocating with a GOSAT-2 sounding, we averaged the TCCON measurements.



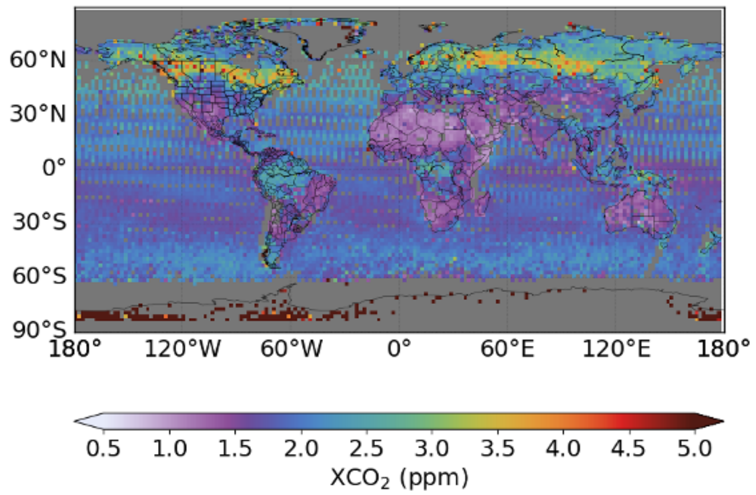


Figure 1: Global averaged XCO_2 (in ppm) and its error between February 2019 and Dec 2023 for the CO2_GO2_SRFP on a 2x2° grid, on top and bottom panels respectively.

The mean bias (global offset) amounts to -0.15 ppm. The standard deviation of the site biases (spatial accuracy or station-to-station variability) is 0.57 ppm for land and 0.48 ppm for ocean observations. TCCON observes these gases with a precision on mole fractions of ~0.15% and ~0.2% for CO_2 and CH_4, respectively (Toon et al., 2009). The single measurement precision of GOSAT-2 compared to TCCON amounts to 2.14 ppm. The validation results are summarised in Table 1.

Figure 2 shows the collocations of GOSAT-2 FP XCO_2 for land observations, and Figure 3 the same for observations over ocean (sunglint), with the TCCON sites. Detailed bias and scatter (i.e., single sounding precision measured by the standard deviation of the difference to TCCON after removing systematic effects) are described in the E3UB (Barr et al. 2024).

Table 1: Overview of the GOSAT-2 XCO_2 products vs TCCON co-located measurements. N is the number of collocated measurements, μ is the mean bias, σ is the single measurement precision, γ is the linear drift term, δ is the station-to-station bias, and R is the correlation coefficient.

	Full Physics					
Variable	N	μ (ppm)	σ (ppm)	γ (ppm yr^{-1})	δ (ppm)	R
GOSAT2 Land	17203	-0.15	2.14	0.48	0.57	0.88
GOSAT-2 Ocean	349	-0.35	2.49	-0.87	0.48	0.79

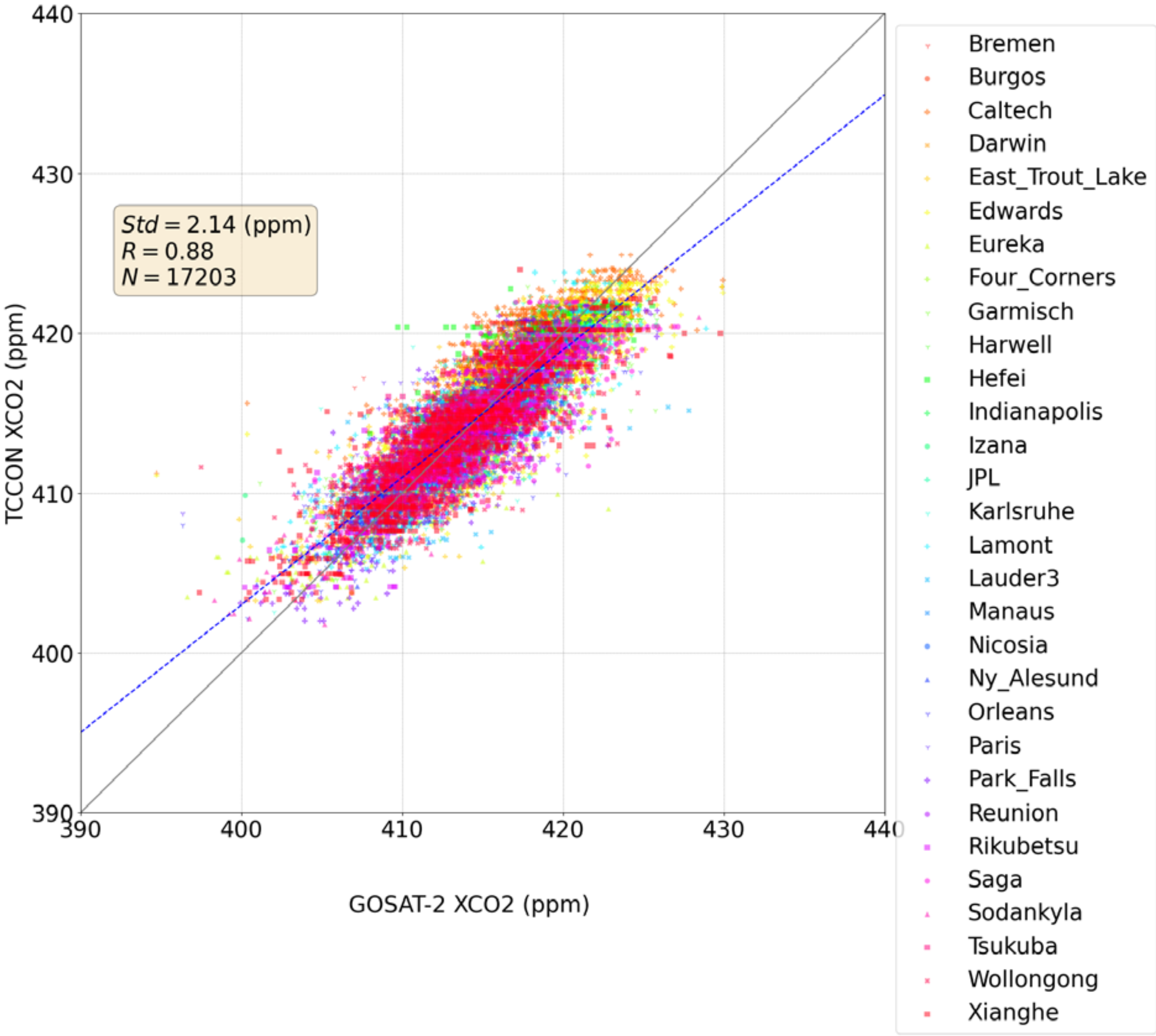


Figure 2: GOSAT-2 XCO_2 for soundings over land plotted against TCCON, for the RemoTeC Full Physics product. Data are compared only if they are fully colocated in space and time. The standard deviation of the population, Pearson's correlation coefficient and number of retrievals are given in the inset. The legend plots the different TCCON stations where markers are as follows. Stations that are along the coast and also sensitive to glint mode (ocean) measurements are indicated as circles. Those that have high latitudes in the northern and southern hemispheres are upward triangles and crosses, respectively. Stations in Asia, North America and Europe are indicated by squares, pluses and downward triangles respectively.

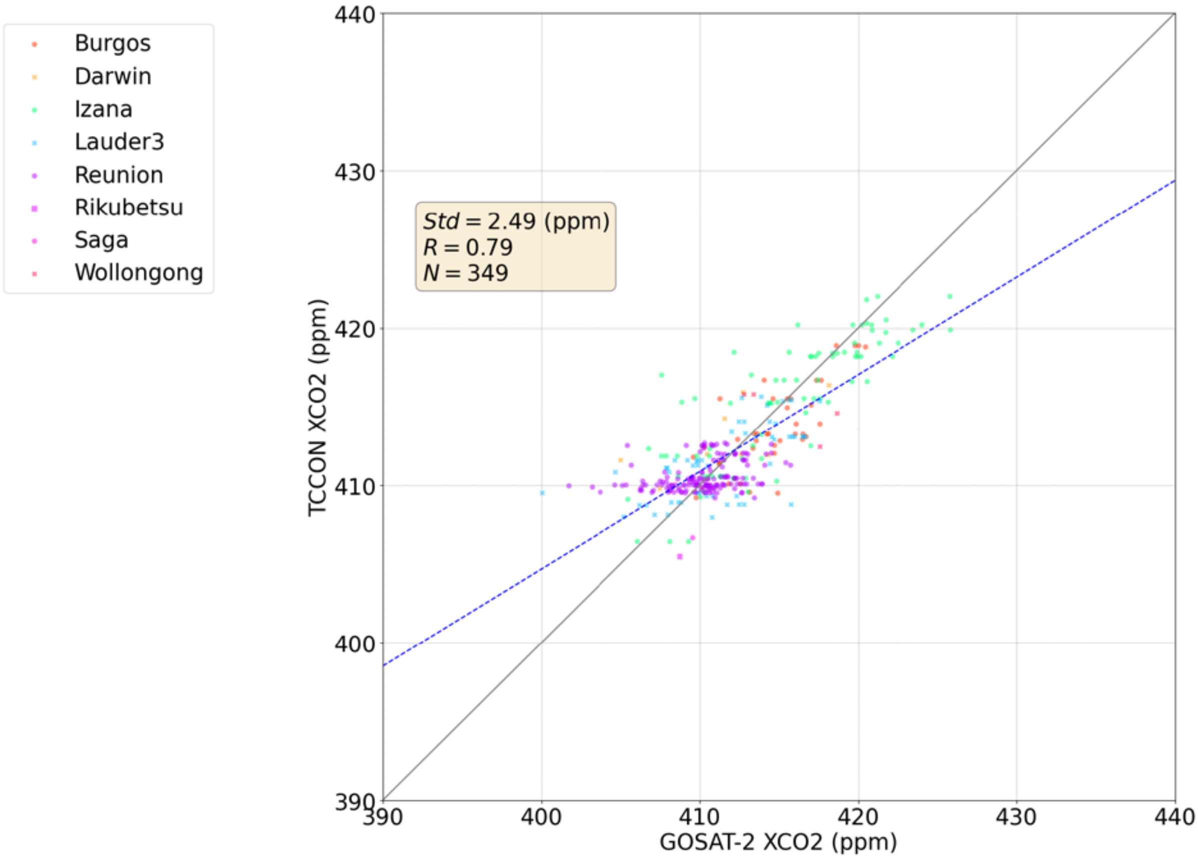


Figure 3 Same as Figure 2 but for soundings taken in glint mode.

4.3 Bias correction

From comparison with TCCON it was found, that the error in XCO_2 is highly correlated with the retrieved albedo α at window 2 (1600 nm). Based on this correlation, the following bias correction for land observations has been developed.

$$\mathrm{XCO2}_{\mathrm{corr}} = \mathrm{XCO2} * (a + b * \alpha)$$

with *a* = 0.98852, b = 0.04537. The bias correction parameters are obtained by the fitting of GOSAT-2 and TCCON differences.

For sun-glint observations, it is found that XCO_2 error is correlated with the O_2 ratio RO_2. It defines the ratio between retrieved and prior O_2 column. In this case, a similar correction function is applied,

$$\mathrm{XCO2}_{\mathrm{corr}} = \mathrm{XCO2} * (a + b * \mathrm{RO}_2)$$

with *a* = 1.4135, b = -0.4192.

5. XCH_4 RemoTeC FP data product (Feb. 2019 – Dec 2023)

In this section, we show examples of the GOSAT-2 XCH_4 FP data product by showing global averaged maps (Sec. 5.1) and by giving a summary of the validation results relative to TCCON (Sec. 5.2).

5.1 Global maps

Figure 4 shows global average maps of the RemoTeC GOSAT-2 FP XCH_4 data product, with the bias-corrected XCH_4 data and the scaled random error (Barr et al. 2024). The GOSAT-2 XCH_4 product provides good global spatial coverage. As can be seen, in some regions the coverage is limited by cloud cover (the observations correspond to cloud free scenes), sun illumination conditions, etc. The error on XCH_4 varies from around 10 ppb to 18 ppb and is higher for high latitude regions and low albedo areas.

5.2 Validation with TCCON

This section summarizes the main validation results presented in the RemoTeC GOSAT-2 ESA GHG CCI+ End-to-End ECV Uncertainty Budget (E3UB) version 5.0 document (Barr et al. 2024). We used ground based TCCON GGG2020 (Laughner et al., 2021) data obtained from https://tccondata.org/ as reference data set. We co-located GOSAT-2 and TCCON measurements with a maximum time difference of 2.5h, a maximum distance of 300 km in both longitudinal and latitudinal directions. In cases of multiple TCCON measurements of the same site collocating with a GOSAT-2 sounding, we averaged the TCCON measurements.

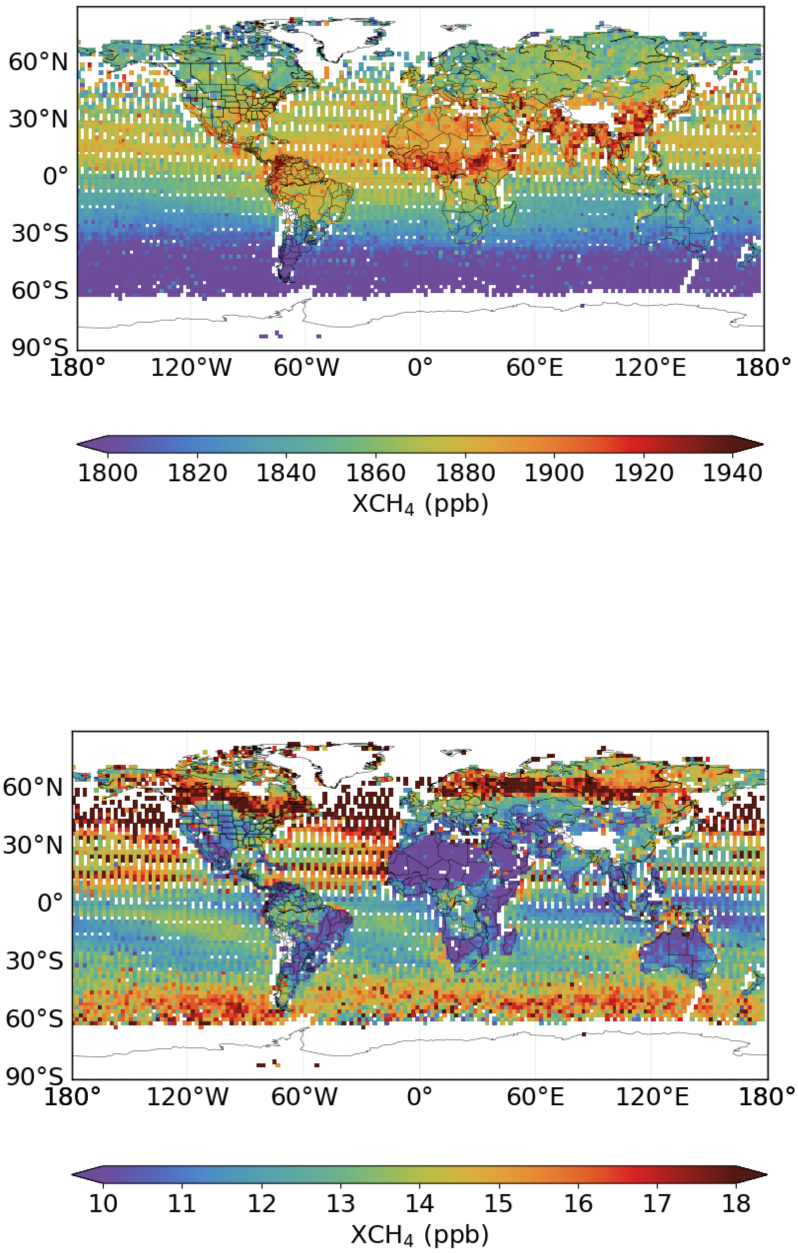


Figure 4: Global averaged XCH_4 (in ppb) and its error between February 2019 and Dec 2023 for the CH4_GO2_SRFP on a 2x2° grid, on top and bottom panels respectively.

The mean bias (global offset) amounts to 0.41 ppb. The standard deviation of the site biases (spatial accuracy or station-to-station variability) is 4.78 ppb for land and 11.57 ppb for ocean observations. TCCON observes these gases with a precision on mole fractions of ~0.15% and ~0.2% for CO_2 and CH_4, respectively (Toon et al., 2009)**.** The single measurement precision of GOSAT-2 compared to TCCON amounts to 15.2 ppb. The validation results are summarised in Table 2.

Figure 5 shows the collocations of GOSAT-2 FP XCH_4 for land observations, and Figure 6 the same for observations over ocean (sunglint), with the TCCON sites. Detailed bias and scatter (i.e., single sounding precision measured by the standard deviation of the difference to TCCON after removing systematic effects) are described in the E3UB (Barr et al. 2024).

Table 2: Overview of the GOSAT-2 XCH_4 products vs TCCON co-located measurements. N is the number of collocated measurements, μ is the mean bias, σ is the single measurement precision, γ is the linear drift term, δ is the station-to-station bias, and R is the correlation coefficient.

	Full Physics					
Variable	N	μ (ppb)	σ (ppb)	γ (ppb yr^{-1})	δ (ppb)	R
GOSAT2 Land	17319	0.41	15.2	0.77	4.78	0.89
GOSAT-2 Ocean	349	1.42	18.13	5.12	11.57	0.89

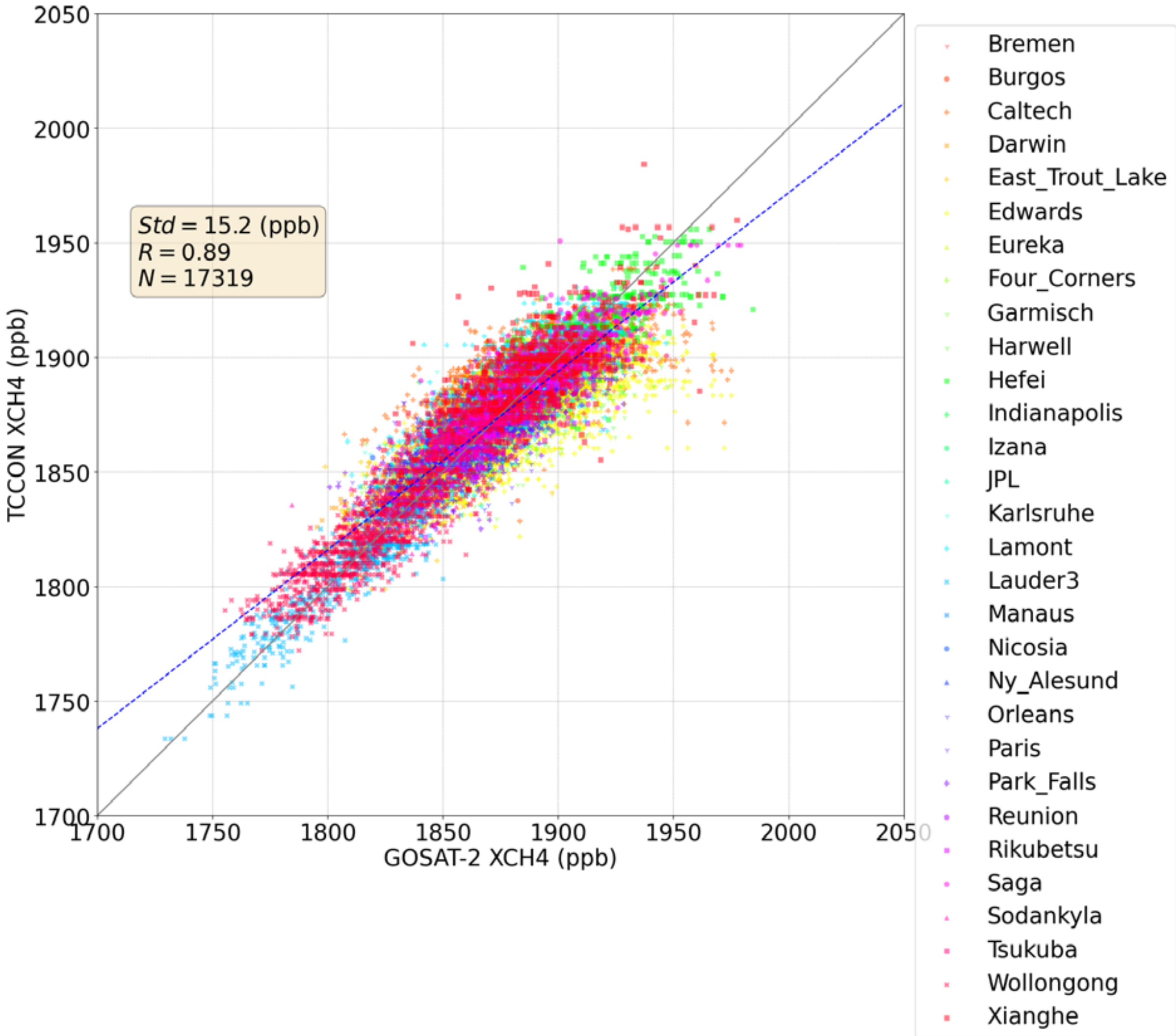


Figure 5: GOSAT-2 XCH_4 for soundings over land plotted against TCCON, for the RemoTeC Full Physics product. Data are compared only if they are fully colocated in space and time. The standard deviation of the population, Pearson's correlation coefficient and number of retrievals are given in the inset. The legend plots the different TCCON stations where markers are as follows. Stations that are along the coast and also sensitive to glint mode (ocean) measurements are indicated as circles. Those that have high latitudes in the northern and southern hemispheres are upward triangles and crosses, respectively. Stations in Asia, North America and Europe are indicated by squares, pluses and downward triangles respectively.

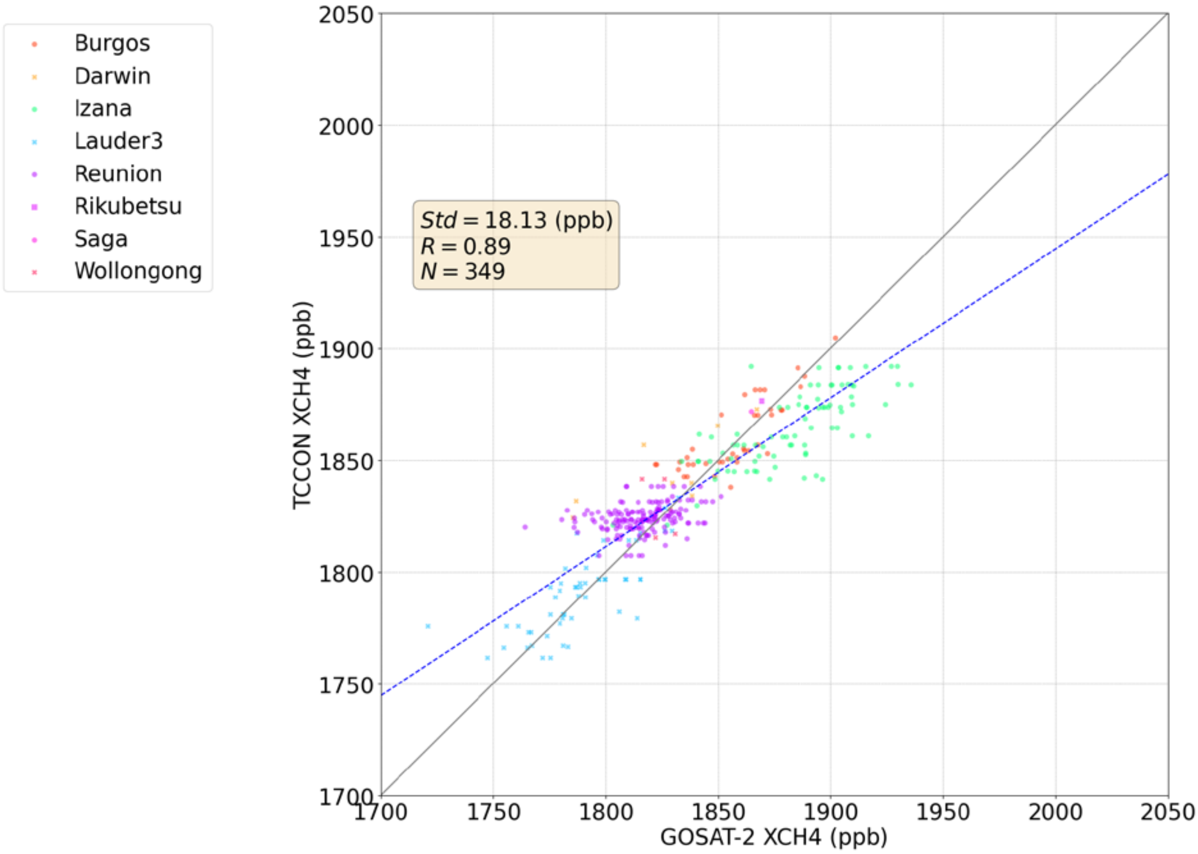


Figure 6 Same as Figure 5 but for soundings taken in glint mode.

5.3 Bias correction

From comparison with TCCON it was found, that the error in XCH_4 is highly correlated with the retrieved albedo α at window 2 (1600 nm). Based on this correlation, the following bias correction for land observations has been developed.

$$\mathrm{XCH4}_{\mathrm{corr}} = \mathrm{XCH4} * (a + b * \alpha)$$

with *a* = 0.98885, b = 0.03115. The bias correction parameters are obtained by the fitting of GOSAT-2 and TCCON differences.

For sun-glint observations, it is found that XCH_4 error is correlated with the O_2 ratio RO_2. It defines the ratio between retrieved and prior O_2 column. In this case, a similar correction function is applied,

$$\mathrm{XCH4}_{\mathrm{corr}} = \mathrm{XCH4} * (a + b * \mathrm{RO}_2)$$

with *a* = 1.4543, b = -0.4636.

6. Description of data format

6.1 Product Content and Format

The Full Physics data products are stored per day in a single NetCDF file. Retrieval results are provided for the individual GOSAT-2 spatial footprints, i.e. no averaging has been applied. The product file contains the key standard products, i.e. the retrieved column averaged dry air mixing ratios XCO_2 and XCH_4 with bias correction, averaging kernels and quality flags, as well as secondary products specific for the RemoTeC algorithm. Tables 3-6 list the key variables in the data product.

Table 3: Common dimensions for the Full Physics products.

Dimension	Type	Unlimited	Units	Description
sounding_dim	int	no		Number of sounding
polarization_dim	int	no		Number of polarization = 2
level_dim	int	no		Number of level = 13
layer_dim	int	no		Number of layer = 12
window_dim	int	no		Number of retrieval window = 4
char_l1bname	int	no		Number of character of L1B name = 44

Table 4: Common variables for the Full Physics products.

Name	Type	Dim.	Units	Description
solar_zenith_angle	float	n	degrees	Angle between line of sight to the sun and local vertical
sensor_zenith_angle	float	n	degrees	Angle between the line of sight to the sensor and the local vertical

time	float	n	seconds	Seconds since 1970-01-01 00:00:00
longitude	float	n	degrees _east	Center longitude
latitude	float	n	degrees _north	Center latitude
pressure_levels	float	n, 13	hPa	Pressure levels
pressure_weight	float	n, 12		Layer dependent weights needed to apply the averaging kernels
flag_landtype	int	n		0 = land, 1 = ocean
flag_sunglint	int	n		0 = no sunglint, 1 = sunglint
gain	char	n		Number of gain coefficient calculated from solar calibration mode data. [1P 1S 2P 2S 3P 3S]
exposure_id	int	n		Exposure identification number of the sounding
l1b_name	char	n		Name of the Level 1B file of the sounding
signal_to_noise_window	float	n, 4, 2		Signal to noise ratio per retrieval window and for both polarization directions
dry_airmass_layer	float	n, 12	m-2	Dry airmass per layer
altitude	float	n	m	Vertical distance above the surface
air_temperature	float	n, 13	K	The bulk temperature of the air at each level
surface_elevation_stdev	float	n	m	Standard deviation of the surface elevation within the sounding
x_wind	float	n, 13	m s-1	Eastward wind velocity
y_wind	float	n, 13	m s-1	Northward wind velocity
chi2	float	n		Chi-squared value of the sounding
optical_thickness_of_atmosphere_layer_due_to_ambient_aerosol	float	n, 4		Scattering optical thickness per retrieval window
h2o_column	float	n	m-2	Retrieved total water column

surface_albedo_758	float	n		The retrieved albedo at 758 nm
surface_albedo_1593	float	n		The retrieved albedo at 1593 nm
surface_albedo_1629	float	n		The retrieved albedo at 1629 nm
surface_albedo_2042	float	n		The retrieved albedo at 2042 nm
intensity_offset_o2a	float	n	W cm-2	The retrieved intensity offset in the O2A band
aerosol_size	float	n		Retrieved size parameter of the aerosol distribution
aerosol_central_height	float	n	m	Peak height of the aerosol Gaussian height distribution
aerosol_total_column	float	n	m-2	Retrieved total aerosol column

Table 5: Product specific (additional) variables for the CO2_GO2_SRFP product.

Name	Type	Dim.	Units	Description
raw_xco2	float	n	1e-6	Retrieved column dry-air mole fraction of atmospheric carbon dioxide (XCO2) in ppm before bias correction
raw_xco2_err	float	n	1e-6	1-sigma statistical uncertainty of the retrieved column-average dry-air mole fraction of atmospheric carbon dioxide
xco2	float	n	1e-6	Retrieved column dry-air mole fraction of atmospheric carbon dioxide (XCO2) in ppm
xco2_uncertainty	float	n	1e-6	1-sigma uncertainty of the retrieved column-average dry-air mole fraction of atmospheric carbon dioxide
xco2_averaging_kernel	float	n, 12		Normalized column averaging kernel

co2_profile_apriori	float	n, 12	1e-6	A priori dry-air mole fraction profile of atmospheric carbon dioxide
xco2_quality_flag	int	n		Quality assurance (QA) value for XCO2 retrieval. A value between 0 and 1 is given to each pixel, where 0 is the best. Pixels with QA value of 1 should never be used and are considered bad quality.

Table 6: Product specific (additional) variables for the CH4_GO2_SRFP product.

Name	Type	Dim.	Units	Description
raw_xch4	float	n	1e-9	Retrieved column dry-air mole fraction of atmospheric methane (XCH4) in ppb before bias correction
raw_xch4_err	float	n	1e-9	1-sigma statistical uncertainty of the retrieved column-average dry-air mole fraction of atmospheric methane
xch4	float	n	1e-9	Retrieved column dry-air mole fraction of atmospheric methane (XCH4) in ppb
xch4_uncertainty	float	n	1e-9	1-sigma uncertainty of the retrieved column-average dry-air mole fraction of atmospheric methane
xch4_averaging_kernel	float	n, 12		Normalized column averaging kernel
ch4_profile_apriori	float	n, 12	1e-9	A priori dry-air mole fraction profile of atmospheric methane
xch4_quality_flag	int	n		Quality assurance (QA) value for XCH4 retrieval. A value between 0 and 1 is given to each pixel, where 0 is the best. Pixels with QA value of 1 should never be used and are considered bad quality.

6.2 Quality Flags and Metadata

To use the data products, users are encouraged to check the corresponding quality flag. In the NetCDF files, the quality flag, namely xco2_quality_flag or xch4_quality_flag, has been generated. The quality flag takes the form of a quality assurance (QA) value, with a range of values from 0 to 1, where 0 is considered to be the best quality data. Increasing the QA value taken will give more data but increase the uncertainty, e.g. QA <= 0.4 will give better coverage but larger scatter in the bias with TCCON.

- 0.8 to 0: progressively better quality data (quality has been checked).
- 1: data should not be used (e.g. bad fit to data, residual cloud contamination)

TCCON validation results (precision) for each QA value can be seen in Figure 7 for both XCO_2 and XCH_4.

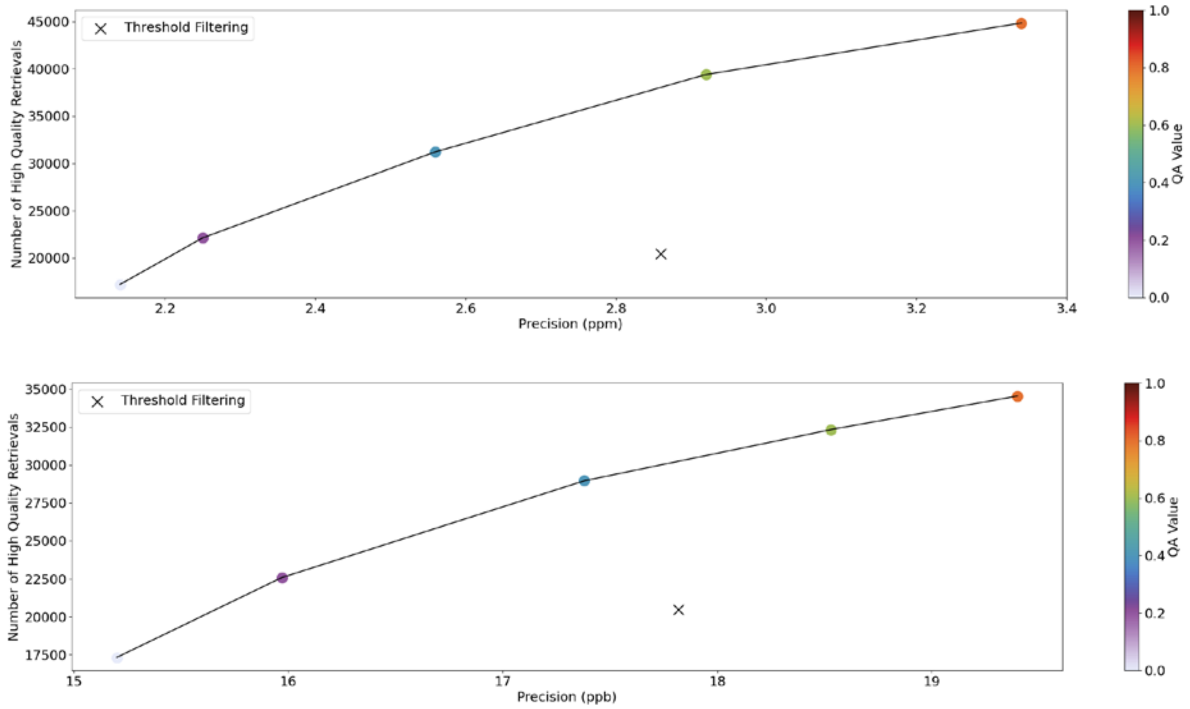


Figure 7: Plot of the number of retrievals flagged as good against the precision as a function of QA value for the XCO_2 (top) and XCH_4 (bottom) Full Physics products. The cross marks the parameter space for filtering the data product with threshold criteria (v2.0.0).

The different QA values are obtained by training random forest models with different strictness of labelling good and bad retrievals during training classification. We then filter the global dataset to make a prediction whether each pixel is good or bad, using the differntly trained models. Finally we take an average of the binary outputs to create a QA value between 0 and 1.

Quality Filtering

Quality filtering is conducted through the use of a trained random forest classifier model which predicts the quality of the retrievals based on a selection of retrieval parameters, such as the cirrus signal, intensity offset, slope of the continuum etc.

The random forest model is trained on GOSAT-2 colocations with TCCON, where retrievals are classified via the bias. Retrievals in the training set are flagged as good quality if the bias is within a certain range and those outwith this range are flagged as bad quality. The model then learns the relationship between the quality of the retrieval and the selection of retrieval parameters on which it is trained, and uses these to predict the quality of all future data.

Data from TCCON are also used in the validation of the final product. The random forest model is trained in a supervised way, where the model is given values for a selection of features as well as a classification label. Thus inclusion of the same data in both training and predicting can lead to artificial features, as the model has already seen the target label in training and therefore simply predicts this. In order to avoid this, we implement a temporal extrapolation by excluding one year from the training data, and using the resulting model to make the quality predictions for the excluded year. This results in one model per year of filtered data. Here we assume that the relationship between retrieval quality and features used in training is not temporally dependent. Figure 8 shows the an example of classification metrics from the filtering model for retrievals in the year 2021 for the XCO_2 product.

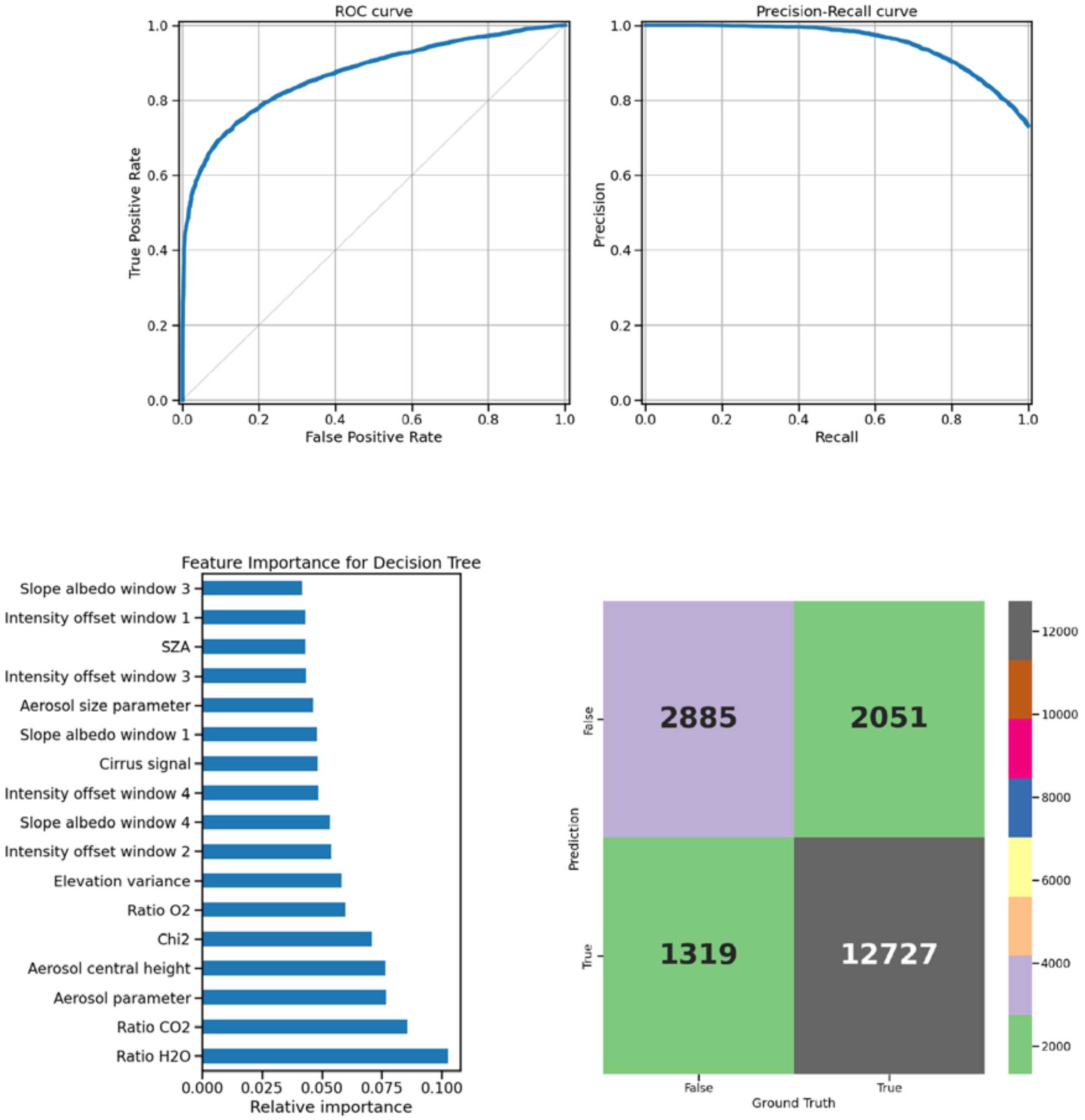


Figure 8: *Top*: The ROC curve (left) shows the true positive rate vs. false positive rate at different classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing both false positives and true positives. In the precision-recall (right) curve, a high area under the curve represents both high recall and high precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate. *Bottom left:* List of features used in the model in order of importance from top to bottom. *Bottom right*: Confusion matrix comparing number of correct predictions.

A consequence of training the random forest model on GOSAT-2 colocations with TCCON is that retrievals with surface albedo ≳ 0.4 are absent from the training sample, due to the lack of TCCON stations in high albedo areas. To circumvent this we define a set of high albedo data to include in the training set using a list of threshold criteria (following v2.0.0 of this product).

Furthermore the random forest method is limited only to land retrievals due to the low number of colocations over ocean. For ocean measurements we then also apply the threshold criteria for determining the quality flag. The criteria are as follows where any retrieval that does not satisfy <u>all</u> of these is assigned a 1 for the quality flag:

For land:

- Cost function (chi2) < 12.0
- Number of iterations < 31
- Signal to noise ratio (SNR) > 50
- Std. deviation of surface elevation within GOSAT-2 ground pixel < 100 m
- Solar zenith angle (SZA) < 75^0
- Aerosol optical thickness (AOT) window 1 < 1.0
- 3 < Aerosol size < 6
- 0 < Aerosol height < 10,000 m
- 0 < Blended Albedo < 1.4
- 0 < Cirrus signal < 2.0E-9
- 0.99 < CO_2 ratio < 1.018
- 0.96 < O_2 ratio < 1.04
- 0.95 < H_2O ratio < 1.08

For ocean (sun-glint):

- Cost function (chi2) < 12.0
- Number of iterations < 31
- Signal to noise ratio (SNR) > 50
- Std. deviation of surface elevation within GOSAT-2 ground pixel < 100 m
- Solar zenith angle (SZA) < 75^0
- 0 < Blended Albedo < 0.4
- 0 < Cirrus signal < 2.0E-9
- 0.99 < CO_2 ratio < 1.003
- 0.96 < O_2 ratio < 1.04
- 0.95 < H_2O ratio < 1.08

The yield and accuracy of the data of the final product is directly correlated to the strictness of the classification of the training dataset. The machine learning approach better captures the inter-dependencies of the retrieval parameters as it learns these during training, and thus less retrievals are filtered out when they should be kept in and vice versa.

6.3 Recommended data usage

It is strongly recommended to only use the bias-corrected data in, except if users explicitly correct for biases themselves (e.g. in an inverse modeling framework). The bias correction has been developed independently for land and sun-glint observations.

If the data are to be compared with other XCO_2 or XCH_4 data for which vertical profile information is available (e.g. inverse modeling, comparison to models, comparison to measured profiles), the column averaging kernels should be used. Here, it should be noted, that **the column averaging kernels are to be applied to layer sub-columns (m-2),** as these are the quantities directly retrieved by the RemoTeC algorithm.

For model comparisons, for example, the retrieved XCO_2 should be compared to $[VCO2]'_{model}/[VAIR]_{model}$, where $[VAIR]_{model}$ is the total dry air column provided by the model and $[VCO2]'_{model}$ is the model total CO_2 column after applying the column averaging kernel, viz.:

$$[VCO2]'_{\mathrm{model}} = [VCO2]_{\mathrm{prior}} + \mathbf{a}^T(\mathbf{x}_{\mathrm{model}} - \mathbf{x}_{\mathrm{prior}})$$

where $[VCO2]_{prior}$ is the prior CO_2 total column used in the retrieval, $\mathbf{x}_{model}$ is the vertical CO_2 profile from the model (as sub-columns) and $\mathbf{x}_{prior}$ is the prior vertical profile from the retrieval. For application of the column averaging kernel, the model vertical profile should be re-calculated on the vertical grid of the retrieval (preferred) or the averaging kernel has to be interpolated to the vertical grid of the model. The same applies to XCH_4.

6.4 Tools for Reading the Data

The data are stored in NetCDF format, which can be read with standard tools in the common programming languages (IDL, Matlab, Python, Fortran90, C++, etc).

7. References

/Barr et al. 2023/ A.G Bar, Borsdorff, T., Ladgraf, J.: ESA Climate Change Initiative "Plus" (CCI+) ATBD (ATBDv4) for the RemoTeC CO2 GOSAT-2 Data Product CO2_GO2_SRFP v2.0.2 for the Essential Climate Variable (ECV) Greenhouse Gases (GHG), ESA GHG CCI+, 2023.

/Barr et al. 2024/ A.G Bar, Borsdorff, T., Ladgraf, J.: ESA Climate Change Initiative "Plus" (CCI+) End-to-End ECV Uncertainty Budget Version 2.0.3 for the RemoTeC CO2 (CO2_GO2_SRFP) and CH4 (CH4_GO2_SRFP) GOSAT-2 Data Products for the Essential Climate Variable (ECV) Greenhouse Gases (GHG), ESA GHG CCI+, 2024.

/**Laughner et al., 2021**/: Laughner, J., Andrews, A., Roche, S., Kiel, M., and Toon, G. ginput v1.0.10: GGG2020 prior profile software (Version 1.0.10), https://doi.org/10.22002/D1.1944, 2021.

/Nakajima et al., 2017/ M. Nakajima, Hiroshi Suto, K. Yotsumoto, K. Shiomi, T. Hirabayashi, "Fourier transform spectrometer on GOSAT and GOSAT-2," Proc. SPIE 10563, International Conference on Space Optics — ICSO 2014, 105634O (17 November 2017); https://doi.org/10.1117/12.2304062.

/Toon et al., 2009/ Toon, G., Blavier, J.F., Washenfelder, R., Wunch, D., Keppel-Aleks, G., Wennberg, P., Connor, B., Sherlock, V., Griffith, D., Deutscher, N. and Notholt, J., 2009, April. Total column carbon observing network (TCCON). In Fourier Transform Spectroscopy (p. JMA3). Optical Society of America.

/Wunch et al., 2015/ Wunch, D., Toon, G.C., Sherlock, V., Deutscher, N.M., Liu, X., Feist, D.G., Wennberg, P.O., The Total Carbon Column Observing Network's GGG2014 Data Version. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA (available at: doi:10.14291/tccon.ggg2014.documentation.R0/1221662), 2015.

*** end of document