ESA Climate Change Initiative “Plus” (CCI+)

Product User Guide (PUG) Version 4.0 - for the RemoTeC XCH_4 GOSAT-2 SRON Full-Physics Product (CH4_GO2_SRFP) Version 2.0.2

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 - Remove 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

Table of Contents

1. Purpose of document

This document describes the Product User Guide (PUG) of the RemoTeC XCH_4 GOSAT-2 SRON Full-Physics Product (CH4_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 XCH_4 Level-2 product 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 CH4_GO2_SRFP product is 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 XCH_4 has been validated against ground based TCCON measurements. To further improve accuracy of XCH_4 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. 2023a).

4. XCH_4 RemoTeC FP data product (Feb. 2019 – Dec 2021)

In this section, we show examples of the GOSAT-2 XCH_4 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 and Figure 2 show global average maps of the RemoTeC GOSAT-2 FP XCH_4 data product. Figure 1 shows the bias-corrected XCH_4 data and Figure 2 is the scaled random error (Barr et al. 2023b). The GOSAT-2 FP 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.

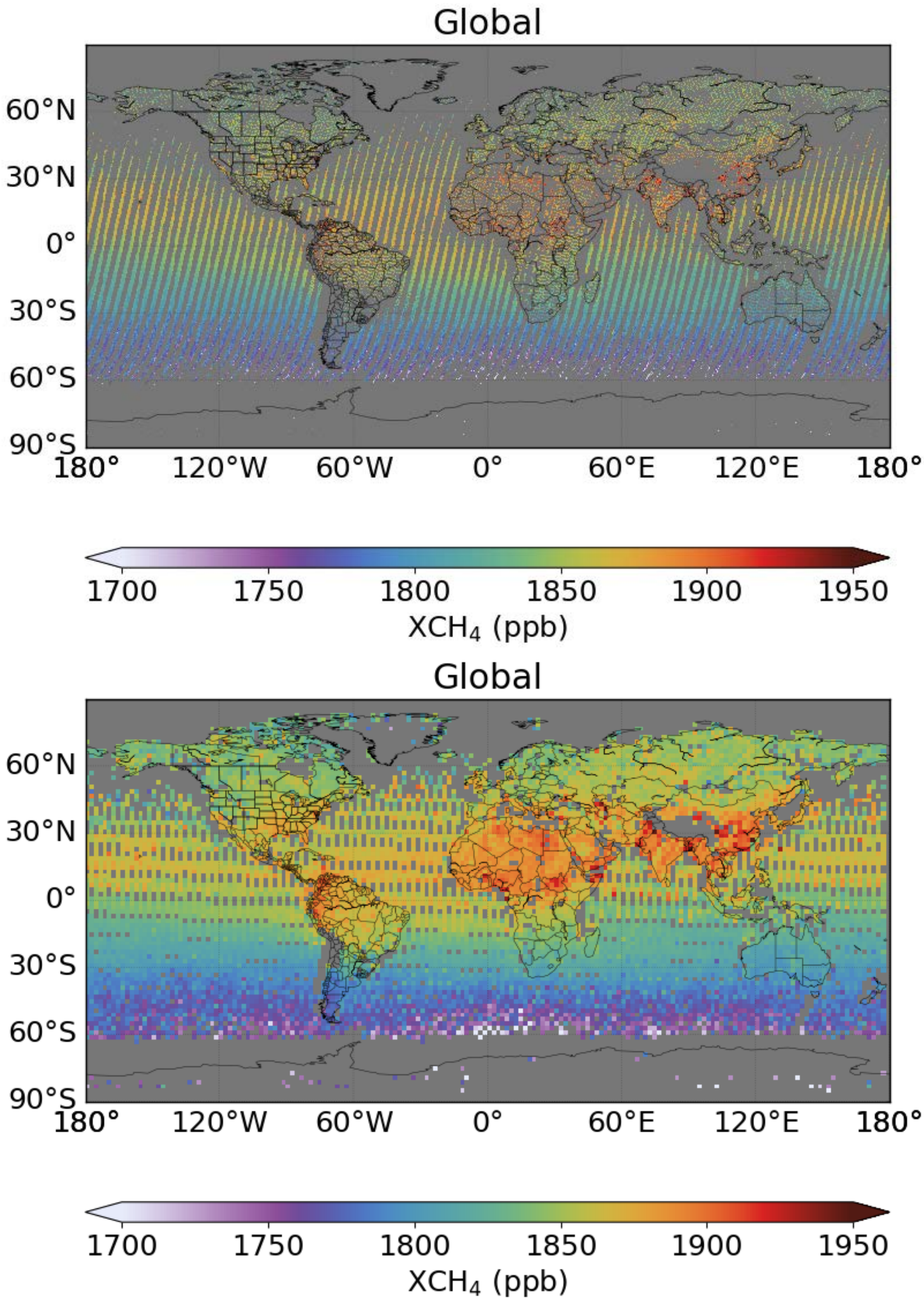

Figure 1: Global averaged XCH_4 between February 2019 and December 2021 for the CH4_GO2_SRFP product on a 0.5x0.5° latitude/longitude grid (top) and 2x2° grid (bottom).

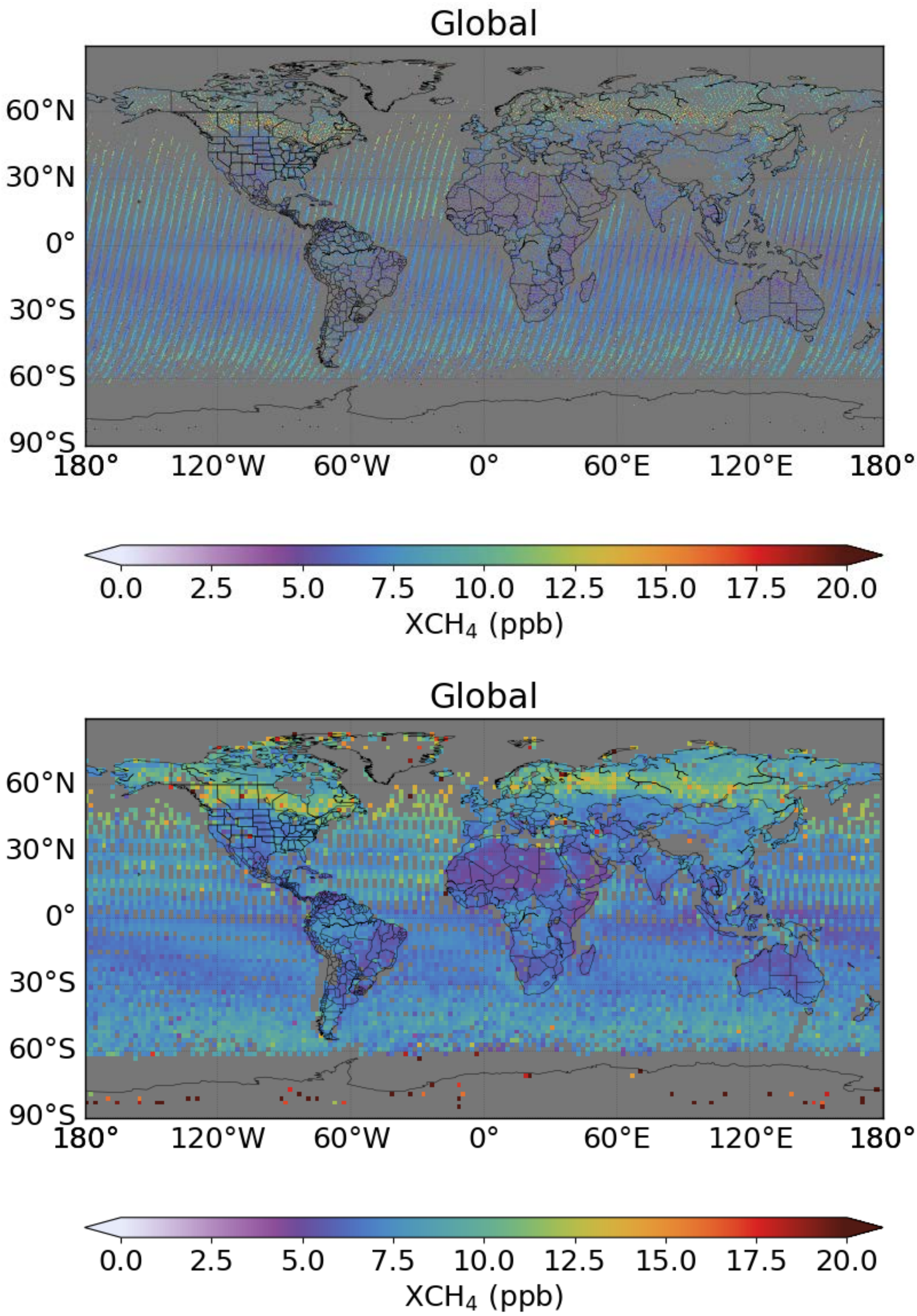

Figure 2: Same as Figure 1 but for the corresponding error.

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) document Version 4.0 (Barr et al. 2023b). 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.

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 4.3 ppb. 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 1.

Figure 3 shows the collocations of GOSAT-2 FP XCH_4 for land observations, and Figure 4 the same for observations over ocean (sunglint), with the TCCON sites. Statistics per site for each mode are shown in Figure 5 and Figure 6, respectively. 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. 2023b).

	Full Physics					
Variable	N	μ (ppb)	σ (ppb)	$\bar{\mu} \pm \sigma_{\bar{\mu}}$ (ppb)	$\bar{\sigma} \pm \sigma_{\bar{\sigma}}$ (ppb)	R
GOSAT2 Land	8399	-0.41	15.19	-0.40 ± 4.30	14.68 ± 2.29	0.85
GOSAT-2 Ocean	109	-0.17	15.01	-2.38 ± 7.73	12.01 ± 2.76	0.92

Table 1: Overview of the GOSAT-2 XCH_4 products vs TCCON co-located measurements. The mean bias μ and single measurement precision σ are calculated by taking the mean and standard deviation of the differences of all GOSAT-2 and TCCON pairs. The mean of the site means $\bar{\mu}$ and the spatial accuracy $\sigma_{\bar{\mu}}$ are calculated by taking the mean and standard deviation of the site means. The mean standard deviation $\bar{\sigma}$ and and standard deviation of the standard deviations $\sigma_{\bar{\sigma}}$ are calculated by taking the mean and the standard deviation of the site standard deviations.

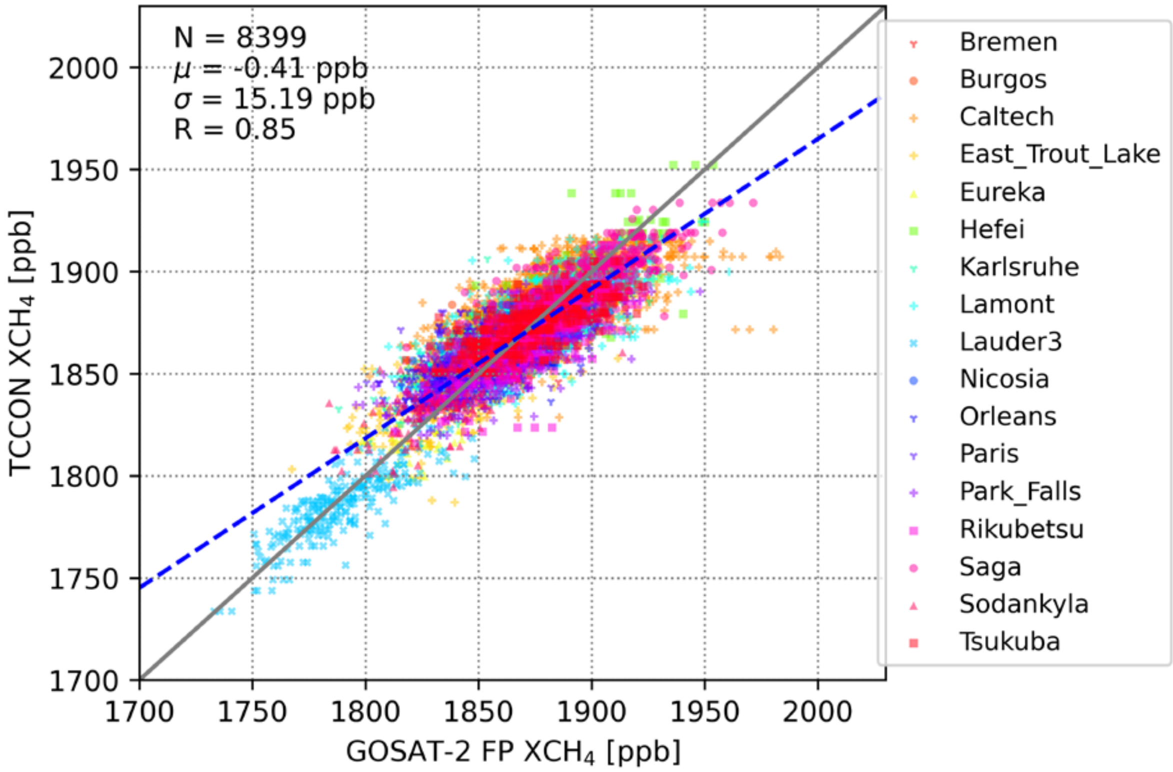

Figure 3: Validation of land soundings of FP XCH_4 with collocated TCCON measurements at all TCCON sites for the period Feb. 2019 - Dec 2021. Numbers in the figures: μ = bias, i.e., average of the difference; σ = single measurement precision, i.e., standard deviation of the difference; N = number of co-locations; R = Pearson correlation coefficient.

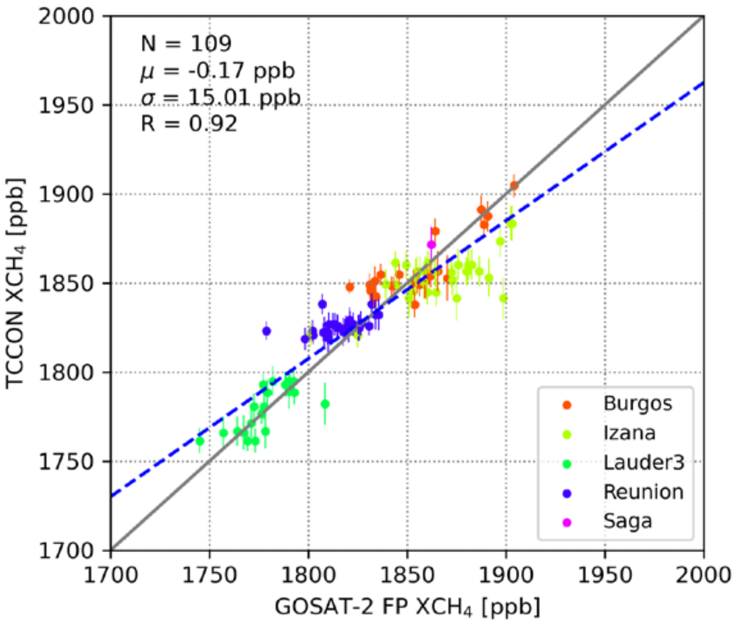

Figure 4: Validation of ocean soundings of XCH_4 with collocated TCCON measurements at all TCCON sites for the period Feb. 2019 - Dec 2021. Numbers in the figures: μ = bias, i.e., average of the difference; σ = single measurement precision, i.e., standard deviation of the difference; N = number of co-locations; R = Pearson correlation coefficient.

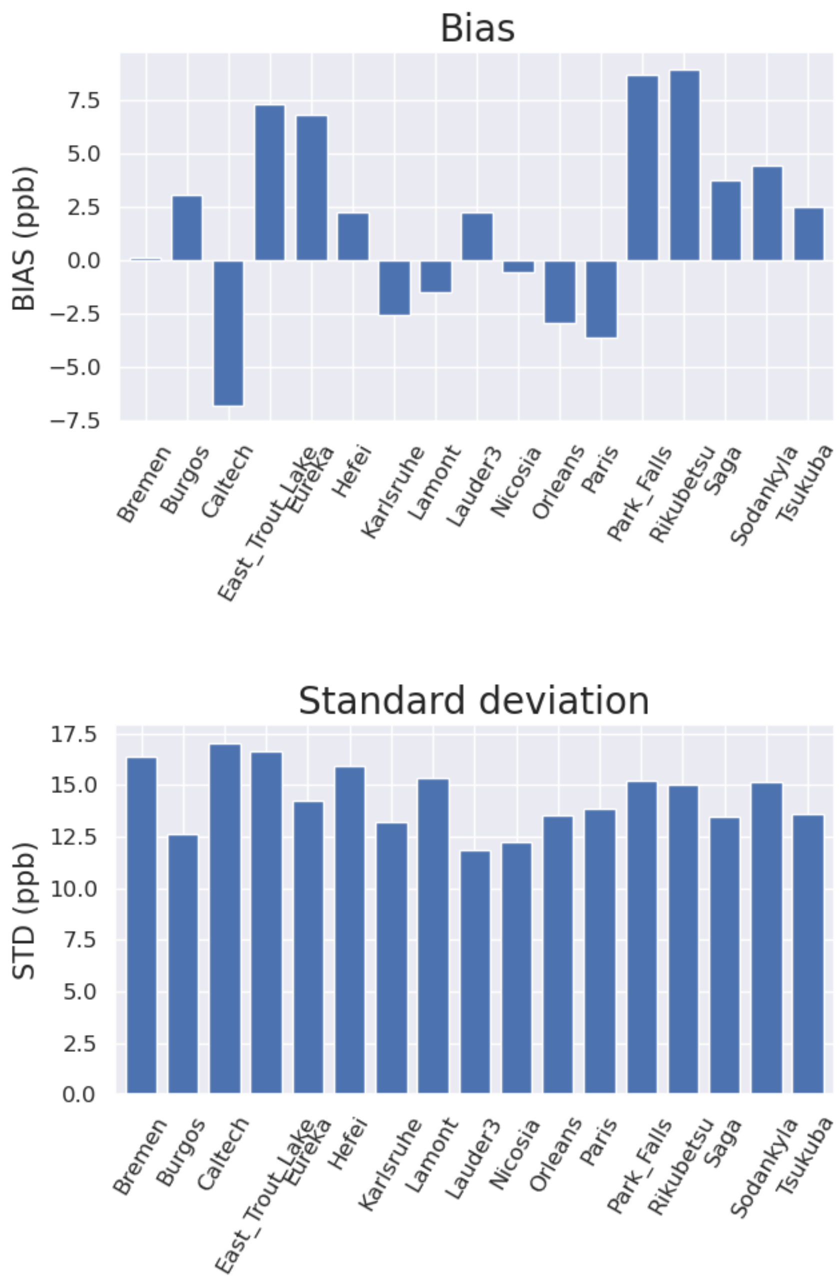

Figure 5: Validation statistics bias (top) and scatter (bottom) per TCCON site for land observations (bias corrected). The summarizing values represent the standard deviation of the site biases and the average scatter relative to TCCON.

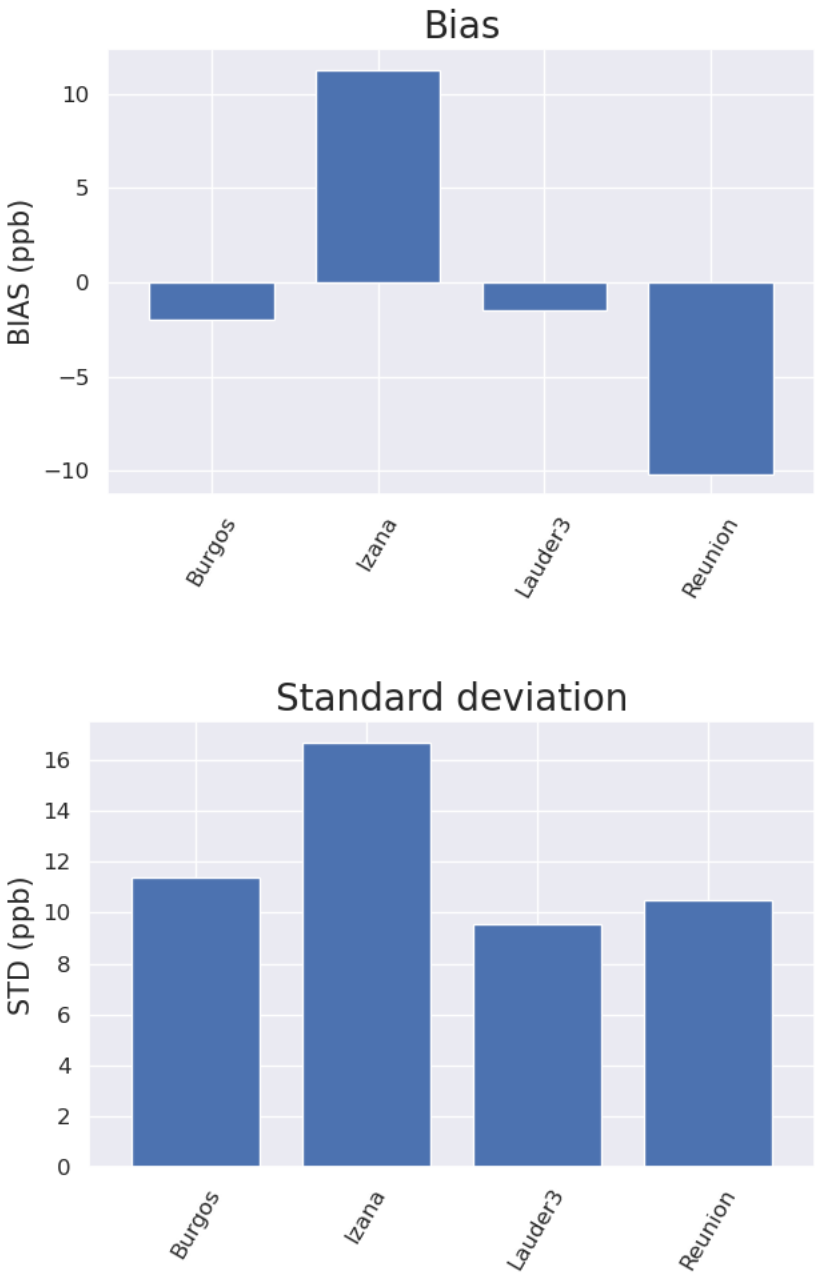

Figure 6: Same as Figure 5 but for sun-glint observations.

4.3 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.

$$\mathrm{XCH4_{corr}} = \mathrm{XCH4} * (a + b * \alpha)$$

with a = 0.99034 and b = 0.04061. 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 $\mathrm{RO_2}$. It defines the ratio between retrieved and prior O_2 column. In this case, a similar correction function is applied,

$$\mathrm{XCH4_{corr}} = \mathrm{XCH4} * (a + b * \mathrm{RO_2})$$

with a = 1.46506 and b = -0.47338.

5. Description of data format

5.1 Product Content and Format

The CH4_GOS_SRFP data are stored per day in a single NetCDF file. Retrieval results are provided for the individual GOSAT 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 2-4 list the key variables in the data product.

Table 2: Common dimensions for the CH4_GO2_SRFP product.

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 3: Common variables for the CH4_GO2_SRFP product.

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
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 flag for XCH4 retrieval, 0 = good, 1 = bad

Table 4: Product specific (additional) variables for the CH4_GO2_SRFP product.

Name	Type	Dim.	Units	Description
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 altitude above the surface
air_temperature	float	n, 13	K	The bulk temperature of the air at each level
surface_altitude_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
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
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

5.2 Quality Flags and Metadata

To use data of GOSAT-2 FP XCH_4, users are encouraged to check the corresponding quality flag. In the NetCDF files, the quality flag, namely xch4_quality_flag, has been generated. It can have two values,

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

<u>Filtering</u>

In v2.0.2 of the XCH_4 full physics product, 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 of XCH_4. 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.

XCH_4 from TCCON is also used in the validation of the final product (sections 4.2 and 4.3). 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 7 shows the classification metrics for the random forest model for 2021.

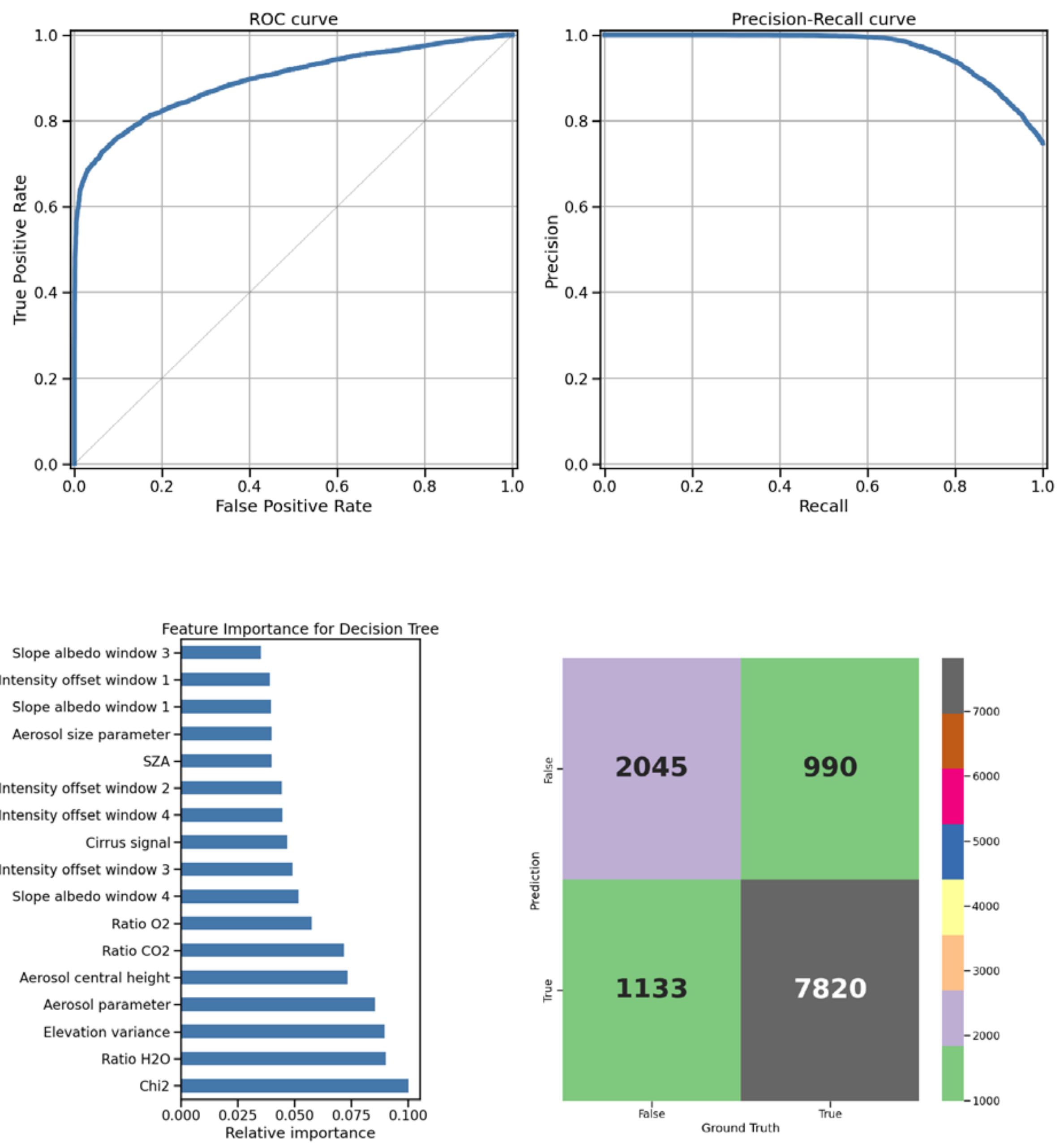

Figure 7: *Top*: The ROC curve (left) shows the true positive rate vs. false positive rate at different classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing both false positives and true positives. In the precision-recall (right) curve, a 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 all 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. For the data product described here, filtering using the random forest classifier results in an increase of 13 % and 2 % for both data yield and precision respectively in comparison to v2.0.0. 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.

Finally, we find an excellent correlation (R=0.96) and low bias (1.5 ppb) between v2.0.0 and v2.0.2 for land retrievals, illustrated in Figure 8, meaning that, while TCCON stations are limited in global coverage, training on TCCON data does not produce obvious biases when extrapolating the filtering to global scales.

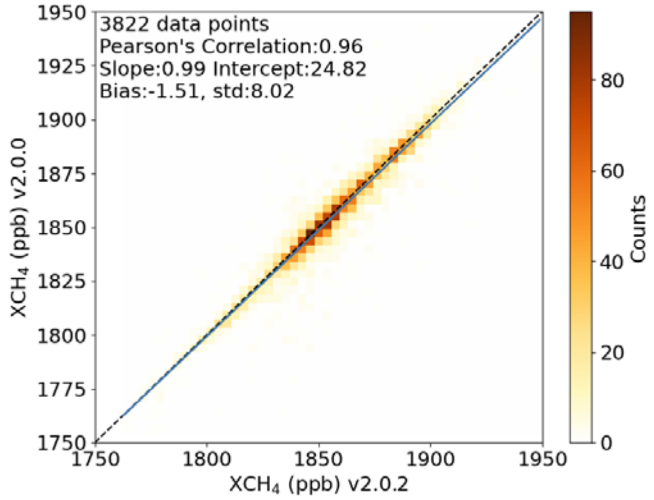

Figure 8: Correlation plot of XCH_4 from v2.0.2 against v2.0.0. Data are sampled on a 1x1° grid across all latitude/longitudes for land measurements over the time period Feb 2019-Dec 2021.

5.3 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.

Important! Please only use data with xch4_quality_flag = 0 (for land and sun-glint observation)

If the data are to be compared with other 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, the retrieved XCH_4 should be compared to $[VCH4]'_{model}/[VAIR]_{model}$, where $[VAIR]_{model}$ is the total dry air column provided by the model and $[VCH4]'_{model}$ is the model total CH_4 column after applying the column averaging kernel, viz.:

$$[VCH4]'_{\mathrm{model}} = [VCH4]_{\mathrm{prior}} + \mathbf{a}^T(\mathbf{x}_{\mathrm{model}} - \mathbf{x}_{\mathrm{prior}})$$

where $[VCH4]_{prior}$ is the prior CH_4 total column used in the retrieval, $\mathbf{x}_{model}$ is the vertical CH_4 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.

5.4 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).

6. References

/Barr et al. 2023a/ A.G Barr and the GHG-CCI group at SRON: ESA Climate Change Initiative "Plus" (CCI+) ATBD (ATBDv4) for the RemoTeC CH4 GOSAT-2 Data Product CH4_GO2_SRFP v2.0.2 for the Essential Climate Variable (ECV) Greenhouse Gases (GHG), ESA GHG CCI+, 2023.

/Barr et al. 2023b/ A.G Barr and the GHG-CCI+ group at SRON: ESA Climate Change Initiative "Plus" (CCI+) End-to-End ECV Uncertainty Budget Version 2.0.2 for the RemoTeC CH4 GOSAT-2 Data Product CH4_GO2_SRFP for the Essential Climate Variable (ECV) Greenhouse Gases (GHG), ESA GHG CCI+, 2023.

/**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