ESA Climate Change Initiative "Plus" (CCI+)

Algorithm Theoretical Basis Document (ATBD)

-

TROPOMI WFM-DOAS XCH$_4$

for the Essential Climate Variable (ECV)

Greenhouse Gases (GHG)

Written by:

GHG-CCI group at IUP

Lead author: O. Schneising, IUP, Univ. Bremen, Germany

Change log:

Version Nr.	Date	Status	Reason for change
Version 1	20 Aug 2019	Final	New document

Algorithm Theoretical Basis Document (ATBD)

TROPOMI WFM-DOAS (TROPOMI/WFMD) XCH_4

Prepared by:

Oliver Schneising

Valid for:

TROPOMI/WFMD

Product	Version
Methane column-averaged dry air mole fraction (XCH_4)	v1.2

Contents

1 Purpose of document

This document provides the theoretical basis of the WFM-DOAS algorithm designed to retrieve column-averaged dry air mole fractions of methane, XCH$_4$, from the shortwave-infrared (SWIR) nadir spectra of the TROPOMI instrument onboard the Sentinel-5 Precursor (Sentinel-5P) satellite (Product ID: CH4_S5P_WFMD).

2 Version history

This section summarises the version history of TROPOMI/WFMD XCH$_4$ in detail (Schneising et al., 2019).

v1.2

The data product is based on TROPOMI Level 1b V01.00.00 files comprising spectra from the nominal operational mode, which started end of April 2018, and reprocessed spectra from the previous six-month commissioning phase. All settings are as described in the following.

3 Algorithm overview

The Weighting Function Modified Differential Optical Absorption Spectroscopy (WFM-DOAS) algorithm (Buchwitz et al., 2006, 2007; Schneising et al., 2011, 2019) is a linear least-squares method based on scaling (or shifting) pre-selected atmospheric vertical profiles. The vertical columns of the desired gases are determined from the measured sun-normalised radiance by fitting a linearised radiative transfer model to it.

3.1 Forward model

The forward model is derived from the radiative transfer model SCIATRAN (Rozanov et al., 2002, 2014) in pseudo-spherical atmosphere mode. To enable a fast retrieval, a look-up table scheme for the radiances and their derivatives has been implemented containing 17280 reference spectra for varying solar zenith angle, altitude, albedo, water vapour, and temperature. The reference spectra are computed with high spectral resolution in line-by-line mode and subsequently convolved to TROPOMI spectral resolution of the SWIR bands using an instrument specific fixed spectral response function extracted from the TROPOMI ISRF Calibration Key Data v1.0.0 for nadir at 2338 nm.

Auxiliary input data:

- US Standard atmospheric profiles with methane scaled to 1850 ppb.

- SCIATRAN aerosol model using the background scenario described in Schneising et al. (2008, 2009).

- HITRAN 2016 spectroscopic parameters (Gordon et al., 2017).

- Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) and Global Land Cover Characterization (GLCC) of the United States Geological Survey (USGS) (United States Geological Survey, 2018a,b)

3.2 Inversion procedure

The linearised radiative transfer model (appropriately chosen from the look-up table according to the relevant parameters) plus a low order polynomial is linear least squares fitted to the logarithm of the measured sun-normalised radiance. The trace gas vertical profiles (CH$_4$, CO, H$_2$O) are scaled for the fit (i.e., the profile shape is not varied). Additional fit parameters are the shift of a pre-selected temperature profile, a scaling factor for the pressure profile, and parameters for a second order polynomial.

Due to the potential non-linear dependencies of the radiances with respect to water vapour and temperature within their natural variability, the algorithm treats both parameters iteratively. The algorithm starts with look-up table elements representing U.S. Standard Atmosphere water vapour amount and temperature. If the retrieved parameter pair after the fit is closer to another look-up table element, the process is repeated with the corresponding reference spectrum. Usually convergence is achieved after one iteration step.

The spectral fitting windows in TROPOMI band 7 were optimised to retrieve CH$_4$ and CO simultaneously as accurately as possible (determined by an error analysis based on simulated measurements). They are shown in Figure 3.1 together with the absorption features of the relevant trace gases. Note that CO is a much weaker absorber compared to CH$_4$ and H$_2$O. The apparent albedo is retrieved in the pre-processing by comparison of the measured continuum radiance with pre-calculated values from a look-up table. Cloud information is obtained from strong H$_2$O absorption lines in band 8 (see Figure 3.1) by comparing the measured radiances to reference radiances for cloud-free conditions. As the absorption in these lines is strong, the measured radiance is small in the clear sky case. In the presence of clouds, most of the atmospheric H$_2$O is shielded and the measured backscattered radiance coherently increases (Heymann et al.,

ALGORITHM THEORETICAL BASIS DOCUMENT
TROPOMI WFM-DOAS XCH$_4$
ESA CLIMATE CHANGE INITIATIVE "PLUS" (CCI+)

INSTITUTE OF
ENVIRONMENTAL PHYSICS,
UNIVERSITY OF BREMEN

5
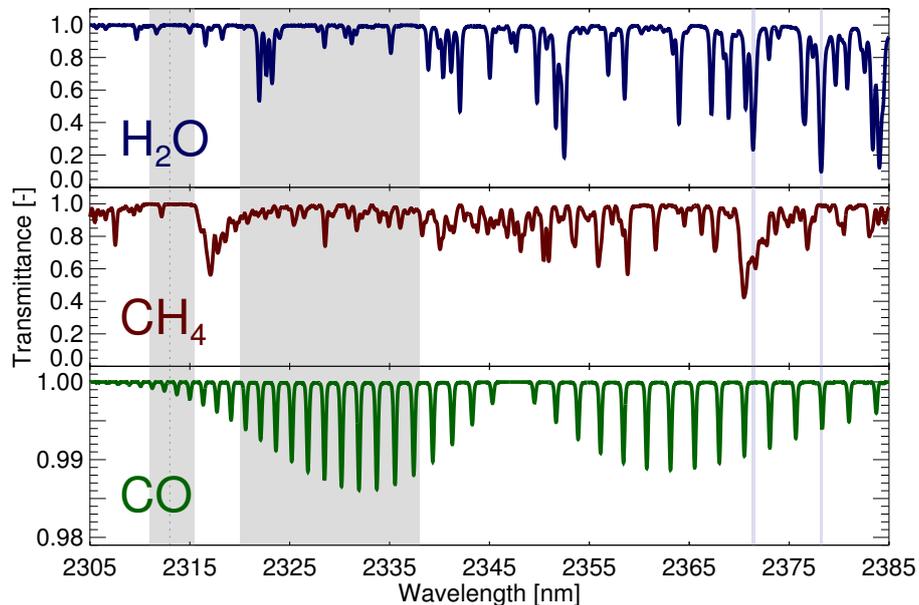
Figure 3.1: Fitting windows (grey) and trace gas transmittances for the SWIR bands of TROPOMI for U.S. Standard atmosphere concentrations. The strong H$_2$O absorption lines between 2370 and 2380 nm used to obtain cloud information are shown in light blue . The apparent albedo is retrieved in the continuum at 2313 nm (dashed line).

2012). The corresponding ratio r_{cld} of measured to reference radiance for the selected strong absorption lines is thus an indicator of cloud contamination.

Apparent albedo retrieval: 2313 nm

Spectral fitting windows: 2311–2315.5 nm, 2320–2338 nm

Fit parameters:

- Scaling factor for CH$_4$ column

- Scaling factor for CO column

- Scaling factor for H$_2$O column

- Shift parameter for temperature profile

- Scaling factor for pressure profile

- Parameters for low order polynomial

In order to convert the retrieved vertical columns into column-averaged dry air mole fractions (denoted XCH$_4$ and XCO), the columns are divided by the dry air column obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) analysis. Thereby, the ECMWF dry columns are corrected for the actual surface elevation of the individual TROPOMI measurements (based on the deviation from the mean altitude of the coarser model grid) inheriting the high spatial resolution of the satellite data.

4 Detailed mathematical algorithm description

This description of the algorithm is based on Schneising (2008) and Schneising et al. (2019).

4.1 The standard DOAS equation

Given a reflecting surface illuminated by a light source under zenith angle θ', the ratio of reflected radiant flux density (integrated radiance L over the half-space) to incident radiant flux density (irradiance) $E = E_0 \cos \theta'$ defines the albedo of the surface

$$\rho = \frac{\int_{\Omega=2\pi} L \cos \theta d\Omega}{E} \quad , \tag{4.1}$$

where Ω is the solid angle and θ the angle between reflection direction and the surface normal (Figure 4.1). Hence, albedo is the fraction of incident radiation that is reflected by a surface, with $\rho \in [0, 1]$ being a dimensionless quantity. In case of a Lambertian surface, the reflected radiance L is isotropic and therefore independent of the viewing geometry which simplifies the relation to

$$\rho = \frac{L \int_{\Omega=2\pi} \cos \theta d\Omega}{E} = \frac{\pi L}{E} \quad . \tag{4.2}$$

The albedo is characteristic for a particular material and usually depends on wavelength for natural surface types.

Considering a surface of albedo ρ and a light path corresponding to a top-of-atmosphere measurement of reflected solar radiation (satellite geometry, see Figure 4.1) in a plane-parallel atmosphere, Beer-Lambert's law

$$L(s) = L_0 e^{-\tau(s)} \quad , \quad \tau(s) = \int_{s_1}^{s_2} \kappa(s) ds \tag{4.3}$$

describing the exponential attenuation of incoming radiation L_0 due to extinction with τ being the slant optical density of the atmosphere corresponding to the
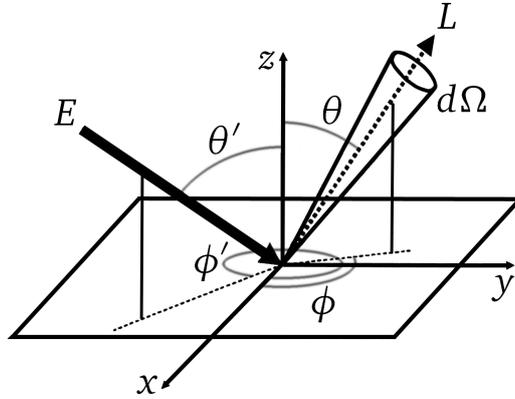
Figure 4.1: Reflection on a surface.

given finite light path can be rewritten as

$$L = \rho \frac{E_0 \cos \theta'}{\pi} e^{-\tau} \quad .$$

$$(4.4)$$

Assuming a spectral window with specific gaseous absorption, the slant optical density can be split into the sum

$$\tau(\lambda) = \tau_g(\lambda) + \tau_r(\lambda)$$

$$(4.5)$$

with slant optical density of the narrow band absorption τ_g of the considered species and residual optical density τ_r of the other processes. Because of the smooth broad band dependencies on wavelength of the Rayleigh scattering coefficient $\sigma_R \propto \lambda^{-4}$ and the aerosol extinction coefficient $\kappa_A \propto \lambda^{-a}$, τ_r can be approximated by a low-order polynomial as function of wavelength if the spectral window is small enough. Hence, the logarithm of the sun-normalised radiance $I(\lambda)$ can be written as

$$\ln I(\lambda) = \ln \left(\frac{\pi L(\lambda)}{E_0(\lambda)} \right) = -\tau_g(\lambda) + \ln \left(\rho(\lambda) \cos \theta' \right) + \sum_{j=0}^{N} \tilde{a}_j \lambda^j$$

$$= -\tau_g(\lambda) + \sum_{j=0}^{N} a_j \lambda^j$$

$$(4.6)$$

using (4.4), (4.5), and the assumption that the albedo term also shows only broad band spectral features and thus can be approximated by a low-order polynomial, too.

Assuming that the absorption cross-section k is independent of height and hence, does not depend on temperature and pressure, the slant optical density respecting the gaseous absorption is given by

$$\tau_g(\lambda) = \int_{s_1}^{s_2} k(\lambda)n(s)ds = k(\lambda)\int_{s_1}^{s_2} n(s)ds = k(\lambda)S \qquad (4.7)$$

with trace gas number density n and slant column S which can be converted in the desired vertical column

$$V = \int_0^{z_0} n(z)dz \qquad (4.8)$$

via division by the air mass factor $A = S/V$ describing the absorption enhancement due to the slant light path. If scattering is neglected the air mass factor is only dependent on the viewing geometry,

$$A = \frac{1}{\cos\theta'} + \frac{1}{\cos\theta} \quad , \qquad (4.9)$$

but generally depends explicitly on wavelength and is not exactly known because the actual light path is unknown due to multiple scattering.

With this notation (4.6) becomes the standard DOAS equation:

$$\ln I(\lambda) = -VA_\lambda k(\lambda) + \sum_{j=0}^{N} a_j\lambda^j \quad . \qquad (4.10)$$

The wavelength dependent air mass factor A_λ is usually approximated by a constant value taking into account scattering properties in the respective wavelength range.

However, the strong absorption lines of the considered gases in the near-infrared are highly pressure and temperature dependent, so that (4.7) is not satisfied and the standard DOAS algorithm is not applicable. This drawback can be tackled by a weighting function modified (WFM) DOAS approach using linearisation points.

4.2 Linearisation and weighting functions

A linear relation between $\ln I(\lambda)$ and the vertical column cannot be derived without any assumption because V is not an independent parameter in the radiative transfer equation. This can be seen by the fact that the sun-normalised radiance depends on the number density vertical profile $n(z)$ of the absorber

rather than on its vertical column amount because the absorption coefficient is given by

$$\alpha_\lambda(z) = n(z)k_\lambda(z) \ . \tag{4.11}$$

Actually, the same vertical column can be achieved by different profiles and thus can occur concurrently with different sun-normalised radiances.

However, $\ln I(\lambda)$ can be expanded by a functional Taylor series with respect to the absorption coefficient around $\bar{\alpha}_\lambda(z)$, thus

$$\ln I(\lambda, \alpha_\lambda(z)) = \ln I(\lambda, \bar{\alpha}_\lambda(z)) + \int_0^{z_0} \left. \frac{\delta \ln I(\lambda, \alpha_\lambda(z))}{\delta \alpha_\lambda(z)} \right|_{\bar{\alpha}_\lambda(z)} \delta\alpha_\lambda(z)dz \tag{4.12}$$

in first order approximation neglecting linearisation error.

Assuming that the variation of the absorption coefficient is only due to variation of the number density profile

$$\delta\alpha_\lambda(z) = k_\lambda(z)\delta n(z) \quad , \quad \delta n(z) = n(z) - \bar{n}(z) \tag{4.13}$$

equation (4.12) becomes

$$\ln I(\lambda, n(z)) = \ln I(\lambda, \bar{n}(z)) + \int_0^{z_0} w_\lambda(z)\delta n(z)dz \tag{4.14}$$

with the height dependent weighting function w_λ being defined by the functional derivative

$$w_\lambda(z) = \left. \frac{\delta \ln I(\lambda, n(z))}{\delta n(z)} \right|_{\bar{n}(z)} \ . \tag{4.15}$$

In order to introduce the vertical column V as a variable, one assumes that the variation in the number density can be achieved by a scaling of the original profile:

$$n(z) = C\bar{n}(z) \tag{4.16}$$

As a consequence the same scaling relation is true for the vertical columns ($V = C\bar{V}$), hence

$$n(z) = \frac{V}{\bar{V}}\bar{n}(z) \tag{4.17}$$

and (4.14) can be rewritten as

$$\ln I(\lambda, n(z)) = \ln I(\lambda, \bar{n}(z)) + \int_0^{z_0} w_\lambda(z)\left(\frac{V}{\bar{V}} - 1\right)\bar{n}(z)dz \ . \tag{4.18}$$

With the columnar weighting function W which is defined by

$$W(\lambda) = \frac{\int_0^{z_0} w_\lambda(z)\bar{n}(z)dz}{\bar{V}} \tag{4.19}$$

the logarithm of the sun-normalised radiance can then be considered locally as a linear function of the absorber's vertical column rather than a functional of its number density profile:

$$\ln I(\lambda, V) = \ln I(\lambda, \bar{V}) + W(\lambda)(V - \bar{V}) \ . \tag{4.20}$$

So far we have only considered a single absorber and neglected other atmospheric processes whose variation can also cause a variation in the sun-normalised radiance. This can be taken into account by introducing additional weighting functions for each considered parameter, respectively. Although, the Rayleigh scattering coefficient, aerosol extinction coefficient, and surface albedo term only show broad band spectral dependence, their weighting functions can contain spectral features of the absorbers due to scattering which enhances the light path. Assuming that the spectral dependence of the weighting functions corresponding to scattering parameters and albedo can be approximated to first order by a polynomial, the linearisation of the logarithm of the sun-normalised radiance is given by

$$\ln I(\lambda, \mathbf{V}) = \ln I(\lambda, \bar{\mathbf{V}}) + \sum_{j=1}^{J} W_j(\lambda)(V_j - \bar{V}_j) + \sum_{i=0}^{N} b_i \lambda^i \tag{4.21}$$

where the components of vector $\mathbf{V}$, denoted V_j, are the vertical columns of all trace gases which have absorption lines in the selected spectral region.

4.3 Inversion and WFM-DOAS equation

We have seen in the previous section that the logarithm of the sun-normalised radiance can be assumed locally as a linear function of the vertical column under the scaling assumption of the vertical profiles of the absorbing gases if the linearisation point $\bar{\mathbf{V}}$ is close enough to $\mathbf{V}$.

In case of the TROPOMI retrieval one wants to determine the vertical column amounts from the measured sun-normalised radiance. As a forward model the discretisation of the deduced relation is used, since TROPOMI provides radiances at discrete wavelength values. Hence, the modelled radiation is given by

$$\ln I_{\lambda_m}^{mod}(\mathbf{V}, \mathbf{b}) = \ln I_{\lambda_m}^{mod}(\bar{\mathbf{V}}) + \sum_{j=1}^{J} \frac{\partial \ln I_{\lambda_m}^{mod}}{\partial V_j}\bigg|_{\bar{V}_j} (V_j - \bar{V}_j) + P_{\lambda_m}(\mathbf{b}) \tag{4.22}$$

with the center wavelength λ_m of detector pixel number m and vector of polynomial coefficients $\mathbf{b}$ of polynomial P. A derivative with respect to a vertical column refers thereby to the change of the top-of-atmosphere radiance caused by a scaling of a pre-selected absorber concentration vertical profile. There are

m equations of this type, one for each detector pixel in the fitting window. The objective is to find the optimal state, so that the linear model best fits the observed radiance.

Since the number of spectral points m is greater than the number of parameters to retrieve n, the problem is overconstrained and a linear least-squares approach is suitable for the retrieval of the desired vertical columns. The problem can be rewritten as

$$\mathbf{y} = \mathbf{Ax} + \mathbf{e} \tag{4.23}$$

with (log-)radiance difference $\mathbf{y} \in \mathbb{R}^m$ of measurement and linearised model due to a deviation $\mathbf{x} \in \mathbb{R}^n$ of the state vector from the multidimensional linearisation point, weighting function (Jacobian) matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ (with derivatives at the linearisation point and polynomial basis functions as columns), as well as sum of forward model error and (normally distributed log-transformed) instrument noise $\mathbf{e} \in \mathbb{R}^m$.

The covariance matrix associated with measurement noise is given by $\mathbf{C_y} = \text{diag}(\sigma_1^2, \ldots, \sigma_m^2) \in \mathbb{R}^{m \times m}$. To give larger weight to spectral points with smaller error variances and to obtain error estimates of the retrieval parameters via error propagation from the uncorrelated measurement errors σ_i, a weighted least squares approach is applied with matrix of weights defined by $\mathbf{W} = \mathbf{C_y}^{-1}$. With the posterior probability $p(\mathbf{x}|\mathbf{y})$ of $\mathbf{x}$ given $\mathbf{y}$, the most probable inference of the inversion $\hat{\mathbf{x}} = \arg\max_{\mathbf{x} \in \mathbb{R}^n} p(\mathbf{x}|\mathbf{y})$ is obtained by minimising

$$f(\mathbf{x}) = \left\| \mathbf{W}^{\frac{1}{2}}(\mathbf{y} - \mathbf{Ax}) \right\|_2^2 = (\mathbf{y} - \mathbf{Ax})^{\mathrm{T}} \mathbf{W}(\mathbf{y} - \mathbf{Ax}) \tag{4.24}$$

with respect to $\mathbf{x}$, where T is the matrix transpose. Hence,

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = 2(\mathbf{A}^{\mathrm{T}} \mathbf{WAx} - \mathbf{A}^{\mathrm{T}} \mathbf{Wy}) \overset{!}{=} 0 \tag{4.25}$$

provides the solution $\hat{\mathbf{x}} = \mathbf{C_x} \mathbf{A}^{\mathrm{T}} \mathbf{Wy}$ of the inverse problem, where $\mathbf{C_x} = (\mathbf{A}^{\mathrm{T}} \mathbf{WA})^{-1}$ is the covariance matrix of solution $\hat{\mathbf{x}}$. The errors of the retrieval parameters are estimated by

$$\hat{\sigma}_j = \sqrt{(\mathbf{C_x})_{jj}} \tag{4.26}$$

where $(\mathbf{C_x})_{jj}$ is the j-th diagonal element of the covariance matrix.

4.4 Look-up table approach

To avoid time-consuming on-line radiative transfer simulations, a fast look-up table scheme has been implemented. The pre-computed spectral radiances and their derivatives (e.g., with respect to trace gas concentration and temperature

or pressure profile changes) depend on solar zenith angle, surface elevation (pressure), surface albedo, water vapour, and temperature amount to consider possible non-linearities caused by the high variability of atmospheric water vapour and temperature. The reference spectra are computed with the radiative transfer model SCIATRAN in line-by-line mode for assumed (e.g., climatological) "mean" columns $\bar{\mathbf{V}}$ depending on surface elevation. Multiple scattering is fully taken into account.

For a specific satellite measurement (characterised by the parameters solar zenith angle, surface elevation, and surface albedo) the appropriate modelled radiance and weighting functions which are used in the fitting procedure are deduced by multidimensional interpolation of the neighbouring reference spectra in parameter space. The required surface albedo is retrieved beforehand by comparing the measured sun-normalised radiance at a selected wavelength, in a transparent region of the fitting window where no significant gaseous or particulate absorptions occur, to pre-calculated radiances for different surface albedos.

The reference spectra are calculated for water vapour amounts corresponding to the U.S. Standard Atmosphere scaled by 0.5, 1, 1.5, 2, 3, and 4, as well as for temperatures profiles shifted by -15, 0, and 15 K. The algorithm starts with an assumed scaling factor of 1 for H$_2$O and a shift of 0 K for the temperature, i.e., with U.S. Standard Atmosphere values. If the retrieved parameter pair after the fit is closer to another look-up table element, the process is repeated with the corresponding reference spectrum. Usually convergence is achieved after one iteration step.

5 Post-processing

This section is largely adopted from Schneising et al. (2019).

5.1 Column-averaged dry air mole fractions

The methane vertical columns are retrieved using small spectral fitting windows in the shortwave-infrared located in TROPOMI band 7. As methane is long-lived and therefore well mixed in the atmosphere, the vertical columns are dominated by topography and surface pressure. Thus, the rather small source and sink signals are not identifiable in the vertical columns. However, by normalising with the dry air column the small differences between the well correlated columns can be observed in the resulting column-averaged mole fraction.

To this end, dry air column obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) analysis are used. Thereby, the ECMWF dry

ALGORITHM THEORETICAL BASIS DOCUMENT
TROPOMI WFM-DOAS XCH$_4$
ESA CLIMATE CHANGE INITIATIVE "PLUS" (CCI+)

INSTITUTE OF
ENVIRONMENTAL PHYSICS,
UNIVERSITY OF BREMEN

13

columns are corrected for the actual surface elevation of the individual TROPOMI measurements (based on the deviation from the mean altitude of the coarser model grid) inheriting the high spatial resolution of the satellite data.

An analysis based on simulated measurements has indicated that this approach is superior to a normalisation by simultaneously retrieved oxygen (O$_2$ A-band) from TROPOMI band 6 for off-nadir conditions and/or in the presence of strong scatterers in the atmosphere (aerosol, clouds) as a consequence of the spectral distance in combination with the albedo differences of natural surface types between NIR band 6 and SWIR band 7. For these reasons, O$_2$ is a barely sufficient proxy for the lightpath in the 2.3 μm spectral range in a scattering atmosphere. In addition to the better accuracy of the ECMWF-based mole fraction computation, this approach is also faster, because the oxygen fit and the interband coregistration mapping can be omitted. The out-of-spectral-band straylight issue of the TROPOMI band 6 (Kleipool et al., 2018) would potentially further hamper the O$_2$-proxy approach.

5.2 Quality filtering

To enable a fast processing speed to handle the huge amount of TROPOMI data, the look-up table is limited to rather simple physical conditions (e.g., cloud-free scenes). Thus, a quality screening algorithm excluding measurements not sufficiently characterized by the forward model had to be implemented. First of all, challenging conditions with solar zenith angles larger than 75°, which are increasingly prone to scattering and saturation related issues due to the weakening signal and lengthening of the light path, are cut off. To be independent of other data sets and their ongoing availability, it was aimed at filtering based on parameters directly included in the retrieval output. This was achieved by using a machine learning approach based on a random forest classifier, which is a meta estimator growing many independent decision trees on different subsamples of the data set and uses averaging to improve the predictive accuracy and prevent overfitting. Thereby, each tree of the ensemble is grown in the following way (Breiman, 2001):

1. Randomly draw N samples from the training set of size N with replacement (bootstrap sample). For large N, a fraction of about 63.2% unique samples is expected, the remainder being duplicates.

2. From the F input variables, $f \ll F$ are randomly chosen out of F and the best split according to minimisation of Gini impurity on these f is used to split the node (Breiman, 1996a). The value of f is held constant during the forest growing.

3. There is no pruning of the decision trees, i.e., each tree is grown to the largest possible extent.

To classify a new previously unseen measurement after growing the forest with the training data, each decision tree gives a classification according to the input features of the measurement and the forest chooses the majority vote over all trees in the forest. The combination of the tree results each based on different bootstrap replicates of the learning set is called *bootstrap aggregating* or *bagging* (Breiman, 1996b). The forest error rate depends on the *correlation* between the trees in the forest and the *strength* of the individual trees. The forest error rate decreases with decreasing correlation and increasing strength of the trees. Reducing f reduces both the correlation and the strength, while increasing f increases both. Hence, there is an optimal range of f minimising the forest error rate.

We use a forest size of 200 trees and the well-recognised standard choice of $f = \sqrt{F}$. The training data set comprises 16 randomly chosen days and for each day, 5 million measurements are randomly selected. Thus, the training subset consists of 80 million measurements, which are classified based on cloud information from the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard Suomi NPP (Hutchison and Cracknell, 2005), which flies in loose formation configuration with Sentinel-5 Precursor (S5P trails behind by 3.5 minutes). This classification is augmented by additionally flagging distinct XCH$_4$ deviations relative to a climatology consisting of MACC-2 flux inversion system (Bergamaschi et al., 2013) averages on a $6° \times 4°$ grid for the years 2003-2005 adjusted by an accumulated increase until the time of the measurement based on globally averaged marine NOAA surface data (Dlugokencky, 2018), identifying scenes obviously not well characterized by the forward model, in particular conspicuously decreased methane abundances in the presence of clouds due to shielding of the underlying atmosphere or in the case of very low surface reflectances.

To train the forest, a set F of 25 feature variables is selected by feature ranking with recursive feature elimination and cross-validated selection of the best features. As widely used, 20% of the training data are randomly drawn and retained as test data. The corresponding predictive accuracy as function of the selected features confirms that the random forest does not overfit, because the accuracy has its global maximum when using all 25 features (Schneising et al., 2019). The selected variables in order of importance are: 1.) H$_2$O column difference to ECMWF, 2.) cloud parameter r_{cld}, 3.) simplified surface type (water, coastal, land, desert, ice), 4.) linear polynomial coefficient p_1, 5.) pressure difference to ECMWF, , 6.) altitude, 7.) latitude, 8.) CO fit error, 9.) temperature, 10.) root mean square of fit residual, 11.) temperature difference to ECMWF, 12.) H$_2$O fit

error, 13.) pressure fit error, 14.) H$_2$O column, 15.) longitude, 16.) solar zenith angle, 17.) pressure, 18.) quadratic polynomial coefficient p_2, 19.) radiance ratio strong H$_2$O absorption to continuum, 20.) dry air column from ECMWF, 21.) retrieved apparent albedo, 22.) continuum radiance, 23.) relative azimuth angle, 24.) across-track dimension index, and 25.) strong H$_2$O absorption radiance. The predictive accuracy when using all 25 features amounts to 0.983, which means that 98.3% of all scenes are correctly classified.

A more detailed analysis of the predictive power of the random forest can be obtained from the confusion matrix of the test data set (Schneising et al., 2019). The percentage of all good measurements that are incorrectly excluded (false negative rate of class 0) amounts to about 13% (9% for land scenes). For these cases the filter is too strict, but the quality of the data passing the filter is not compromised. The percentage of all the measurements predicted to be good that are incorrectly classified and should actually be excluded (false discovery rate of class 0) amounts to about 11%. For these cases the filter appears not stringent enough. However, as the training classification is quite strict, that does not necessarily mean that all these measurements are actually of low quality. The rate can rather be interpreted as an upper bound of potentially remaining challenging retrievals on the verge of sufficient characterisation by the forward model, e.g., observations near cloud edges. The effective diagnostic performance of the quality filter will emerge from the validation.

Adding additional parameters to F does not significantly improve the predictive accuracy further. It is important to note that the resulting classification is independent of the absolute abundances of the primary retrieval parameters CH$_4$ and CO. The performance of the classification algorithm is demonstrated in Schneising et al. (2019) confirming that cloudy scenes are reliably excluded in general and that the quality filter is usually stricter than the VIIRS classification, in particular over the weakly reflecting ocean. Measurements classified as cloudy by VIIRS but still passing the quality filter are rare and not associated with conspicuous methane abundances.

5.3 Shallow learning calibration

The implemented machine learning-based quality filter described in the previous subsection removes observations not sufficiently characterized by the forward model. Although this procedure typically excludes scenes exhibiting large systematic errors, smaller systematic errors may remain in the residual data set. In particular, there seems to be a systematic albedo dependence of unknown origin of retrieved methane abundances with an underestimation over dark surfaces. As a consequence of the fairly stringent quality requirements for methane, a random

ALGORITHM THEORETICAL BASIS DOCUMENT
TROPOMI WFM-DOAS XCH$_4$
ESA CLIMATE CHANGE INITIATIVE "PLUS" (CCI+)

INSTITUTE OF
ENVIRONMENTAL PHYSICS,
UNIVERSITY OF BREMEN

16

forest regressor algorithm was implemented to reduce the remaining systematic methane errors after the retrieval by calibrating against an assumed standard defined below, which is deemed insensitive to surface reflectance variations.

Like the classification algorithm described in the previous subsection, the random forest regressor (Criminisi and Shotton, 2013) grows an ensemble of decision trees, training each tree on a different data sample applying the bootstrap aggregating technique. From the f randomly chosen parameters the optimal split maximising the variance reduction in the child nodes is used to split the nodes. To focus on the most prominent features (shallow learning of systematic errors caused by surface albedo variations), the tree growing is limited to 500 leaf nodes. Again, a forest size of 200 trees and $f = \sqrt{F}$ is used, where F consists of 5 feature variables, which are in order of importance: retrieved apparent albedo, solar zenith angle, cloud parameter r_{cld}, strong H$_2$O absorption radiance, and across-track dimension index.

To compute the correction for a new measurement after growing the forest with the training data, each decision tree provides a regression according to the input features of the measurement and the random forest uses the average over all tree regressions as final calibration value for this observation. In other words, the random forest regressor uses averaging in the bagging procedure to combine the individual tree results (in contrast to voting used in the classification case).

The calibration data set consists of the XCH$_4$ climatology introduced in the previous subsection evaluated for selected regions (Arctic, Western United States, Central Europe, Japan, Sahara, South Atlantic, and Southeast Australia; defined in Schneising et al. (2019)) spanning a wide range of albedos and solar zenith angles. For individual regions, the climatology is roughly corrected for potential systematic overall biases by adding up a single region-specific correction value based on a comparison to nearby sites of the Total Carbon Column Observing Network (TCCON) (Wunch et al., 2011) for the year 2017. In any case, the seasonal and intra-regional spatial variations are solely determined by the climatology.

6 Error characterisation

From the validations with ground-based Fourier Transform Spectroscopy (FTS) measurements at the 21 Total Carbon Column Observing Network (TCCON) sites listed in Table 6.1, realistic error estimates of the satellite data are provided and summarised in Table 6.2 (Schneising et al., 2019).

Station	Latitude [°]	Longitude [°]	Altitude [km]	Reference
Eureka	80.05	−86.42	0.61	Strong et al. (2019)
Ny-Ålesund	78.92	11.92	0.02	Notholt et al. (2017)
Sodankylä	67.37	26.63	0.19	Kivi et al. (2014)
East Trout Lake	54.35	−104.99	0.50	Wunch et al. (2018)
Białystok	53.23	23.03	0.19	Deutscher et al. (2015)
Karlsruhe	49.10	8.44	0.11	Hase et al. (2015)
Orléans	47.97	2.11	0.13	Warneke et al. (2014)
Garmisch	47.48	11.06	0.75	Sussmann and Rettinger (2018)
Park Falls	45.94	−90.27	0.44	Wennberg et al. (2017)
Lamont	36.60	−97.49	0.32	Wennberg et al. (2016b)
Tsukuba	36.05	140.12	0.03	Morino et al. (2018a)
Edwards	34.96	−117.88	0.70	Iraci et al. (2016)
JPL	34.20	−118.18	0.39	Wennberg et al. (2016a)
Caltech	34.14	−118.13	0.24	Wennberg et al. (2015)
Saga	33.24	130.29	0.01	Shiomi et al. (2014)
Burgos	18.53	120.65	0.04	Morino et al. (2018b)
Ascension Island	−7.92	−14.33	0.03	Feist et al. (2014)
Darwin	−12.46	130.93	0.04	Griffith et al. (2014a)
Réunion	−20.90	55.49	0.09	De Mazière et al. (2017)
Wollongong	−34.41	150.88	0.03	Griffith et al. (2014b)
Lauder	−45.04	169.68	0.37	Sherlock et al. (2014)

Table 6.1: TCCON sites used in the validation ordered according to latitude from north to south.

	XCH$_4$ [ppb]	XCO [ppb]
Random Error	14.0	5.1
Systematic Error	4.3	2.0

Table 6.2: Error characterisation of the WFM-DOAS data products (valid for TROPOMI/WFMD v1.2 XCH$_4$ and XCO).

References

Bergamaschi, P., Houweling, S., Segers, A., Krol, M., Frankenberg, C., Scheepmaker, R. A., Dlugokencky, E., Wofsy, S. C., Kort, E. A., Sweeney, C., Schuck, T., Brenninkmeijer, C., Chen, H., Beck, V., and Gerbig, C.: Atmospheric CH$_4$ in the first decade of the 21st century: Inverse modeling analysis using SCIAMACHY satellite retrievals and NOAA surface measurements, J. Geophys. Res., 118, 7350–7369, https://doi.org/10.1002/jgrd.50480, 2013.

Breiman, L.: Technical Note: Some Properties of Splitting Criteria, Machine Learning, 24, 41–47, https://doi.org/10.1023/A:1018094028462, 1996a.

Breiman, L.: Bagging Predictors, Machine Learning, 24, 123–140, https://doi.org/10.1023/A:1018054314350, 1996b.

Breiman, L.: Random forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Buchwitz, M., de Beek, R., Noël, S., Burrows, J. P., Bovensmann, H., Schneising, O., Khlystova, I., Bruns, M., Bremer, H., Bergamaschi, P., Körner, S., and Heimann, M.: Atmospheric carbon gases retrieved from SCIAMACHY by WFM-DOAS: version 0.5 CO and CH$_4$ and impact of calibration improvements on CO$_2$ retrieval, Atmos. Chem. Phys., 6, 2727–2751, https://doi.org/10.5194/acp-6-2727-2006, 2006.

Buchwitz, M., Khlystova, I., Bovensmann, H., and Burrows, J. P.: Three years of global carbon monoxide from SCIAMACHY: comparison with MOPITT and first results related to the detection of enhanced CO over cities, Atmos. Chem. Phys., 7, 2399–2411, https://doi.org/10.5194/acp-7-2399-2007, 2007.

Criminisi, A. and Shotton, J.: Decision Forests for Computer Vision and Medical Image Analysis. Advances in Computer Vision and Pattern Recognition, Springer, London, 2013.

De Mazière, M., Sha, M. K., Desmet, F., Hermans, C., Scolas, F., Kumps, N., Metzger, J.-M., Duflot, V., and Cammas, J.-P.: TCCON data from Réunion Island (La Réunion), France, Release GGG2014.R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.reunion01.r1, 2017.

Deutscher, N. M., Notholt, J., Messerschmidt, J., Weinzierl, C., Warneke, T., Petri, C., and Grupe, P.: TCCON data from Białystok, Poland, Release GGG2014.R1.

TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.bialystok01.r1/1183984, 2015.

Dlugokencky, E. J.: Annual Increase in Globally-Averaged Atmospheric Methane, NOAA ESRL, (last access: 22 October 2018), URL `www.esrl.noaa.gov/gmd/ccgg/trends_ch4/`, 2018.

Feist, D. G., Arnold, S. G., John, N., and Geibel, M. C.: TCCON data from Ascension Island, Saint Helena, Ascension and Tristan da Cunha, Release GGG2014.R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.ascension01.r0/1149285, 2014.

Gordon, I. E., Rothman, L. S., Hill, C., Kochanov, R. V., Tan, Y., Bernath, P. F., Birk, M., Boudon, V., Campargue, A., Chance, K. V., Drouin, B. J., Flaud, J.-M., Gamache, R. R., Hodges, J. T., Jacquemart, D., Perevalov, V. I., Perrin, A., Shine, K. P, Smith, M.-A., Tennyson, J., Toon, G. C., Tran, H., Tyuterev, V. G., Barbe, A., Császár, A. G., Devi, V. M., Furtenbacher, T., Harrison, J. J., Hartmann, J.-M., Jolly, A., Johnson, T., Karman, T., Kleiner, I., Kyuberis, A., Loos, J., Lyulin, O. M., Massie, S. T., Mikhailenko, S. N., Moazzen-Ahmadi, N., Müller, H. S. P., Naumenko, O. V., Nikitin, A. V., Polyansky, O. L., Rey, M., Rotger, M., Sharpe, S., Sung, K., Starikova, E., Tashkun, S. A., Vander Auwera, J., Wagner, G., Wilzewski, J., Wcisło, P., Yu, S., and Zak, E. J.: The HITRAN2016 molecular spectroscopic database, J. Quant. Spectrosc. Radiat. Transfer, 203, 3–69, https://doi.org/10.1016/j.jqsrt.2017.06.038, 2017.

Griffith, D. W., Deutscher, N. M., Velazco, V. A., Wennberg, P. O., Yavin, Y., Keppel-Aleks, G., Washenfelder, R. A., Toon, G. C., Blavier, J.-F., Paton-Walsh, C., Jones, N. B., Kettlewell, G. C., Connor, B. J., Macatangay, R. C., Roehl, C., Ryczek, M., Glowacki, J., Culgan, T., and Bryant, G. W.: TCCON data from Darwin, Australia, Release GGG2014.R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.darwin01.r0/1149290, 2014a.

Griffith, D. W., Velazco, V. A., Deutscher, N. M., Paton-Walsh, C., Jones, N. B., Wilson, S. R., Macatangay, R. C., Kettlewell, G. C., Buchholz, R. R., and Riggenbach, M. O.: TCCON data from Wollongong, Australia, Release GGG2014.R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.wollongong01.r0/1149291, 2014b.

Hase, F., Blumenstock, T., Dohe, S., Groß, J., and Kiel, M.: TCCON data from Karlsruhe, Germany, Release GGG2014.R1. TCCON data archive, hosted by

ALGORITHM THEORETICAL BASIS DOCUMENT
TROPOMI WFM-DOAS XCH$_4$
ESA CLIMATE CHANGE INITIATIVE "PLUS" (CCI+)

INSTITUTE OF
ENVIRONMENTAL PHYSICS,
UNIVERSITY OF BREMEN

20

CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.karlsruhe01.r1/1182416, 2015.

Heymann, J., Bovensmann, H., Buchwitz, M., Burrows, J. P., Deutscher, N. M., Notholt, J., Rettinger, M., Reuter, M., Schneising, O., Sussmann, R., and Warneke, T.: SCIAMACHY WFM-DOAS XCO$_2$: reduction of scattering related errors, Atmos. Meas. Tech., 5, 2375–2390, https://doi.org/10.5194/amt-5-2375-2012, 2012.

Hutchison, K. D. and Cracknell, A. P.: Visible Infrared Imager Radiometer Suite: A New Operational Cloud Imager, CRC Press of Taylor and Francis, London, 2005.

Iraci, L. T., Podolske, J. R., Hillyard, P. W., Roehl, C., Wennberg, P. O., Blavier, J.-F., Landeros, J., Allen, N., Wunch, D., Zavaleta, J., Quigley, E., Osterman, G. B., Albertson, R., Dunwoody, K., and Boyden, H.: TCCON data from Edwards, California, USA, Release GGG2014.R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.edwards01.r1/1255068, 2016.

Kivi, R., Heikkinen, P., and Kyrö, E.: TCCON data from Sodankylä, Finland, Release GGG2014.R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.sodankyla01.r0/1149280, 2014.

Kleipool, Q., Ludewig, A., Babić, L., Bartstra, R., Braak, R., Dierssen, W., Dewitte, P.-J., Kenter, P., Landzaat, R., Leloux, J., Loots, E., Meijering, P., van der Plas, E., Rozemeijer, N., Schepers, D., Schiavini, D., Smeets, J., Vacanti, G., Vonk, F., and Veefkind, P.: Pre-launch calibration results of the TROPOMI payload on-board the Sentinel-5 Precursor satellite, Atmos. Meas. Tech., 11, 6439–6479, https://doi.org/10.5194/amt-11-6439-2018, 2018.

Morino, I., Matsuzaki, T., and Horikawa, M.: TCCON data from Tsukuba, Ibaraki, Japan, 125HR, Release GGG2014.R2. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.tsukuba02.r2, 2018a.

Morino, I., Velazco, V. A., Hori, A., Uchino, O., and Griffith, D. W. T.: TCCON data from Burgos, Philippines, Release GGG2014.R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.burgos01.r0, 2018b.

ALGORITHM THEORETICAL BASIS DOCUMENT
TROPOMI WFM-DOAS XCH$_4$
ESA CLIMATE CHANGE INITIATIVE "PLUS" (CCI+)

INSTITUTE OF
ENVIRONMENTAL PHYSICS,
UNIVERSITY OF BREMEN

21

Notholt, J., Warneke, T., Petri, C., Deutscher, N. M., Weinzierl, C., Palm, M., and Buschmann, M.: TCCON data from Ny-Ålesund, Spitzbergen, Norway, Release GGG2014.R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.nyalesund01.r0/1149278, 2017.

Rozanov, V. V., Buchwitz, M., Eichmann, K.-U., de Beek, R., and Burrows, J. P.: SCIATRAN - a new radiative transfer model for geophysical applications in the 240–2400 nm spectral region: The pseudo-spherical version, Adv. Space Res., 29, 1831–1835, 2002.

Rozanov, V. V., Rozanov, A. V., Kokhanovsky, A. A., and Burrows, J. P.: Radiative transfer through terrestrial atmosphere and ocean: Software package SCIATRAN, J. Quant. Spectrosc. Radiat. Transfer, 133, 13–71, https://doi.org/10.1016/j.jqsrt.2013.07.004, 2014.

Schneising, O.: Analysis and interpretation of satellite measurements in the near-infrared spectral region: Atmospheric carbon dioxide and methane, Ph.D. thesis, University of Bremen, 2008.

Schneising, O., Buchwitz, M., Burrows, J. P., Bovensmann, H., Reuter, M., Notholt, J., Macatangay, R., and Warneke, T.: Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite - Part 1: Carbon dioxide, Atmos. Chem. Phys., 8, 3827–3853, https://doi.org/10.5194/acp-8-3827-2008, 2008.

Schneising, O., Buchwitz, M., Burrows, J. P., Bovensmann, H., Bergamaschi, P., and Peters, W.: Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite - Part 2: Methane, Atmos. Chem. Phys., 9, 443–465, https://doi.org/10.5194/acp-9-443-2009, 2009.

Schneising, O., Buchwitz, M., Reuter, M., Heymann, J., Bovensmann, H., and Burrows, J. P.: Long-term analysis of carbon dioxide and methane column-averaged mole fractions retrieved from SCIAMACHY, Atmos. Chem. Phys., 11, 2863–2880, https://doi.org/10.5194/acp-11-2863-2011, 2011.

Schneising, O., Buchwitz, M., Reuter, M., Bovensmann, H., Burrows, J. P., Borsdorff, T., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Hermans, C., Iraci, L. T., Kivi, R., Landgraf, J., Morino, I., Notholt, J., Petri, C., Pollard, D. F., Roche, S., Shiomi, K., Strong, K., Sussmann, R., Velazco, V. A., Warneke, T., and Wunch, D.: A scientific algorithm to simultaneously retrieve carbon monoxide and methane from TROPOMI onboard Sentinel-5

ALGORITHM THEORETICAL BASIS DOCUMENT
TROPOMI WFM-DOAS XCH$_4$
ESA CLIMATE CHANGE INITIATIVE "PLUS" (CCI+)

INSTITUTE OF
ENVIRONMENTAL PHYSICS,
UNIVERSITY OF BREMEN

22

Precursor, Atmos. Meas. Tech. Discuss., 2019, 1–44, in review, https://doi.org/10.5194/amt-2019-243, 2019.

Sherlock, V., Connor, B., Robinson, J., Shiona, H., Smale, D., and Pollard, D. F.: TCCON data from Lauder, New Zealand, 125HR, Release GGG2014.R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.lauder02.r0/1149298, 2014.

Shiomi, K., Kawakami, S., Ohyama, H., Arai, K., Okumura, H., Taura, C., Fukamachi, T., and Sakashita, M.: TCCON data from Saga, Japan, Release GGG2014.R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.saga01.r0/1149283, 2014.

Strong, K., Roche, S., Franklin, J. E., Mendonca, J., Lutsch, E., Weaver, D., Fogal, P. F., Drummond, J. R., Batchelor, R., and Lindenmaier, R.: TCCON data from Eureka, Canada, Release GGG2014.R3. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.eureka01.r3, 2019.

Sussmann, R. and Rettinger, M.: TCCON data from Garmisch, Germany, Release GGG2014.R2. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.garmisch01.r2, 2018.

United States Geological Survey: Land Cover Products – Global Land Cover Characterization (GLCC), https://doi.org/10.5066/F7GB230D, 2018a.

United States Geological Survey: Digital Elevation - Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010), https://doi.org/10.5066/F7J38R2N, 2018b.

Warneke, T., Messerschmidt, J., Notholt, J., Weinzierl, C., Deutscher, N. M., Petri, C., and Grupe, P.: TCCON data from Orléans, France, Release GGG2014.R0. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.orleans01.r0/1149276, 2014.

Wennberg, P. O., Wunch, D., Roehl, C. M., Blavier, J.-F., Toon, G. C., and Allen, N. T.: TCCON data from California Institute of Technology, Pasadena, California, USA, Release GGG2014.R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.pasadena01.r1/1182415, 2015.

ALGORITHM THEORETICAL BASIS DOCUMENT
TROPOMI WFM-DOAS XCH$_4$
ESA CLIMATE CHANGE INITIATIVE "PLUS" (CCI+)

INSTITUTE OF
ENVIRONMENTAL PHYSICS,
UNIVERSITY OF BREMEN

23

Wennberg, P. O., Roehl, C. M., Blavier, J.-F., Wunch, D., Landeros, J., and Allen, N. T.: TCCON data from Jet Propulsion Laboratory, Pasadena, California, USA, Release GGG2014.R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.jpl02.r1/1330096, 2016a.

Wennberg, P. O., Wunch, D., Roehl, C. M., Blavier, J.-F., Toon, G. C., and Allen, N. T.: TCCON data from Lamont, Oklahoma, USA, Release GGG2014.R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.lamont01.r1/1255070, 2016b.

Wennberg, P. O., Roehl, C. M., Wunch, D., Toon, G. C., Blavier, J.-F., Washenfelder, R., Keppel-Aleks, G., Allen, N. T., and Ayers, J.: TCCON data from Park Falls, Wisconsin, USA, Release GGG2014.R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.parkfalls01.r1, 2017.

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, Phil. Trans. R. Soc. A, 369, 2087–2112, https://doi.org/10.1098/rsta.2010.0240, 2011.

Wunch, D., Mendonca, J., Colebatch, O., Allen, N. T., Blavier, J.-F., Roche, S., Hedelius, J., Neufeld, G., Springett, S., Worthy, D., Kessler, R., and Strong, K.: TCCON data from East Trout Lake, Canada, Release GGG2014.R1. TCCON data archive, hosted by CaltechDATA, California Institute of Technology, https://doi.org/10.14291/tccon.ggg2014.easttroutlake01.r1, 2018.