	ESA Climate Change Initiative “Plus” (CCI+)	Page 1
	Algorithm Theoretical Basis Document Version 1 (ATBDv1) - FOCAL OCO-2	
	for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Version 1 – Final 23. Aug. 2019

ESA Climate Change Initiative “Plus” (CCI+)

Algorithm Theoretical Basis Document Version 1 (ATBDv1)


Retrieval of XCO₂ from the OCO-2 satellite using the Fast Atmospheric Trace Gas Retrieval (FOCAL)

for the Essential Climate Variable (ECV)
Greenhouse Gases (GHG)

Written by:

GHG-CCI group at IUP

Lead author: M. Reuter, IUP, Univ. Bremen, Germany

	ESA Climate Change Initiative “Plus” (CCI+) Algorithm Theoretical Basis Document Version 1 (ATBDv1) - FOCAL OCO-2 for the Essential Climate Variable (ECV) Greenhouse Gases (GHG)	Page 2
		Version 1 – Final
		23. Aug. 2019

Change log:

Version Nr.	Date	Status	Reason for change
Version 1 – Final 1	23. Aug. 2019	Final Version	New document.

Algorithm Theoretical Basis Document Version 1 (ATBDv1)

–

Retrieval of XCO₂ from the OCO-2 satellite using the Fast Atmospheric Trace Gas Retrieval (FOCAL)

ESA Climate Change Initiative "Plus" (CCI+)
for the Essential Climate Variable (ECV)

Greenhouse Gases (GHG)

Prepared by:

M. Reuter, M. Buchwitz, O. Schneising, H. Bovensmann, J. P. Burrows

Institute of Environmental Physics (IUP)
University of Bremen, FB1
PO Box 33 04 40
D-28334 Bremen
Germany

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	4
---------------------	---	---	----------

Contents

1	Introduction	6
2	Algorithm Overview	8
2.1	Physical Basis	8
2.2	Performance	10
2.3	Computational Efficiency	11
2.4	Input Data	12
2.5	Output Data	12
3	Radiative Transfer	14
3.1	Radiance transmission	15
3.2	Irradiance transmission	16
3.3	Solar radiation	17
3.4	I_C	17
3.5	I_{SD}	17
3.6	I_{CD}	18
3.7	I_{SI}	18
3.8	I_{CI}	19
3.9	I_{SIF}	19
3.10	Approximations	20
3.11	Pseudo-spherical geometry	21
4	Retrieval	23
4.1	Measurement vector y	23
4.2	Measurement error covariance matrix \mathbf{S}_ϵ	25
4.3	Forward model \vec{F}	25
4.4	State vector \vec{x}	26
4.5	A priori error covariance matrix \mathbf{S}_a	26
4.6	Jacobian matrix \mathbf{K}	28
4.7	Parameter vector \vec{b}	30
4.8	A posteriori error covariance matrix $\hat{\mathbf{S}}$	30
4.9	Convergence	30
5	Preprocessing	31
5.1	Filtering	31
5.2	Noise Model	34
5.3	Zero level offset correction	38

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	5
---------------------	---	---	----------

6	Postprocessing	41
6.1	Filtering	41
6.2	Bias correction	42
7	Error Characterization	48
7.1	Retrieval setups	48
7.2	Scenarios	49
7.3	Results	50
8	Validation	59
8.1	Model comparison	59
8.2	Comparison with NASA's operational OCO-2 L2 product . . .	59
8.3	Validation with TCCON	61
9	Version History	65
9.1	v08	65
9.2	v06	65
9.3	v01	65
	References	66

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	6
---------------------	---	---	----------

1 Introduction

Satellite retrievals of the atmospheric dry-air column-average mole fraction of CO₂ (XCO₂) based on hyperspectral measurements in appropriate near (NIR) and short wave infrared (SWIR) O₂ and CO₂ absorption bands can help to answer pressing questions about the carbon cycle (e.g., Reuter et al., 2017a). However, the precision and even more the accuracy requirements for applications like surface flux inversion or emission monitoring are demanding (e.g., Miller et al., 2007; Chevallier et al., 2007; Bovensmann et al., 2010). As an example, large scale biases of a few tenths of a ppm can already hamper an inversion with mass-conserving global inversion models (Miller et al., 2007; Chevallier et al., 2007).

The Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY, Burrows et al., 1995; Bovensmann et al., 1999) became operational in 2002 and its radiance measurements allowed to start the time series of NIR/SWIR XCO₂ retrievals. With an overlap of about three years, the Greenhouse Gases Observing Satellite (GOSAT, Kuze et al., 2009) allowed complementation and continuation of this time series in 2009.

The Orbiting Carbon Observatory-2 (OCO-2) was launched in 2014 also aiming at continuing and improving XCO₂ observations from space. As part of the A-train satellite constellation, OCO-2 flies in a sun-synchronous orbit crossing the equator at 13:36 local time. It measures one polarization direction of the solar backscattered radiance in three independent wavelength bands: the O₂-A band at around 760 nm (band1) with a spectral resolution of about 0.042 nm and a spectral sampling of about 0.015 nm, the weak CO₂ band at around 1610 nm (band2) with a spectral resolution of about 0.080 nm and a spectral sampling of about 0.031 nm, and the strong CO₂ band at around 2060 nm (band3) with a spectral resolution of about 0.103 nm and a spectral sampling of about 0.040 nm. OCO-2 is operated in a near-push-broom fashion and has eight footprints across track and an integration time of 0.333 s. The instrument's spatial resolution at ground is 1.29 km across track and 2.25 km along track. See Crisp et al. (2004) for more information on the OCO-2 instrument.

Multiple scattering of light at aerosols and clouds can be a significant error source for XCO₂ retrievals. Therefore, so called full physics retrieval algorithms were developed aiming to minimize scattering related errors by explicitly fitting scattering related properties such as cloud water/ice content, aerosol optical thickness, cloud height, etc. However, the computational costs for multiple scattering radiative transfer (RT) calculations can be immense. Processing

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	7
---------------------	---	---	----------

all data of the Orbiting Carbon Observatory-2 (OCO-2) can require up to thousands of CPU cores and the next generation of CO₂ monitoring satellites will produce at least an order of magnitude more data. For this reason, the Fast atmOspheric traCe gAs retrieval FOCAL has been developed reducing the computational costs by orders of magnitude by approximating multiple scattering effects with an analytic solution of the RT problem of an isotropic scattering layer.

This algorithm theoretical basis document (ATBD) describes FOCAL in detail as used for the retrieval of XCO₂ from OCO-2. In large parts, this document is compiled from text and figures of the publications of Reuter et al. (2017c,b). Reuter et al. (2017c) described the physical and mathematical basis of FOCAL's radiative transfer (RT) and assessed the quality of a proposed FOCAL based OCO-2 XCO₂ retrieval algorithm by confronting it with accurate multiple scattering vector RT simulations covering, among others, some typical cloud and aerosol scattering scenarios. This initial FOCAL OCO-2 XCO₂ algorithm with the version number v01 has only been used for theoretical studies based on simulated measurements.

Reuter et al. (2017b) adapted this algorithm and confronted FOCAL for the first time with actually measured OCO-2 data and protocolled the steps undertaken to transform the input data (most importantly, the OCO-2 radiances) into a validated XCO₂ data product. This includes preprocessing, adaptation of the noise model, zero level offset correction, post-filtering, bias correction, comparison with the CAMS (Copernicus Atmosphere Monitoring Service) greenhouse gas flux inversion model, comparison with NASA's operational OCO-2 XCO₂ product, and validation with ground based Total Carbon Column Observing Network (TCCON) data. Their FOCAL OCO-2 XCO₂ algorithm has the version number v06 and is the bases for further developments also described in this ATBD.

The FOCAL OCO-2 XCO₂ algorithm (in the following for the sake of simplicity referred to as FOCAL) is being continuously developed further and the most recent version is v08. Several sections of this document correspond to a specific version number. This is mentioned at the beginning or within the heading of the particular sections. A version history itemizing the main changes from version to version can be found in Section 9.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	8
---------------------	---	---	----------

2 Algorithm Overview

2.1 Physical Basis

The FOCAL OCO-2 XCO₂ algorithm described in this ATBD fits the OCO-2 measured radiance simultaneously in four fit windows: SIF (~ 758.26 – 759.24 nm), O₂ (~ 757.65 – 772.56 nm), wCO₂ (~ 1595.0 – 1620.6 nm), and sCO₂ (~ 2047.3 – 2080.9 nm). This is achieved by iteratively optimizing the state vector including the following geophysical parameters: five layered CO₂ and H₂O concentration profiles, the pressure (i.e., height), scattering optical thickness at 760 nm, and the Ångström exponent of a scattering layer, solar induced chlorophyll fluorescence (SIF), and polynomial coefficients describing the spectral albedo in each fit window. The fit is performed using the optimal estimation formalism (Rodgers, 2000) and Levenberg-Marquardt minimization of the cost function.

The RT model of FOCAL approximates multiple scattering effects at an optically thin isotropic scattering layer. It splits up the top of atmosphere (TOA) radiance into parts originating from direct reflection at the scattering layer or the surface and parts originating from multiple scattering of the diffuse radiant flux between scattering layer and surface. FOCAL's relatively simple approximation of the RT problem allows unphysical inputs such as negative scattering optical thicknesses or albedos. This can be an advantage when analyzing measurements including noise and assuming Gaussian a priori error statistics. FOCAL accounts for polarization only implicitly by the retrieval of a variable scattering optical thickness.

The PPDF (photon path-length distribution function) method (e.g., Bril et al., 2007, 2012) gains its computational efficiency by applying the theorem of equivalence to replace computationally expensive multiple scattering RT computations with a set of fast transmission computations. This is conceptually similar to FOCAL which uses an effective transmission function for the diffuse flux. However, different from the PPDF method, FOCAL accounts for multiple scattering by solving the geometric series of successive (flux) scattering events.

In principle, the PPDF method can simulate arbitrary scattering phase functions (SPFs). This is not possible for FOCAL which can only simulate an isotropic scattering layer. However, splitting the radiance into direct and diffuse parts can be interpreted as a SPF with a sharp forward peak and which is isotropic otherwise. This still represents typical Mie SPFs not very well but much better than an entirely isotropic SPF.

Strictly, the theorem of equivalence only applies for spectral regions with constant scattering and reflection properties (Bennartz and Preusker, 2006)

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	9
---------------------	---	---	----------

making the PPDF shape, e.g., depending on surface albedo. This can make it complicated to transfer scattering information from one fit window into another. Reflection and scattering properties of FOCAL are allowed to vary within the fit windows and can be used to transfer information between fit windows, e.g., via the Ångström exponent.

Despite FOCAL is in principle able to account for scattering at an optically thin scattering layer, pre- and post-filtering as well as bias correction is still needed. The strict pre-filtering bases on sounding quality, NASA's spike EOF analyses, OMI UV aerosol index, and MODIS Aqua cloud coverage. Due to the wider swath, MODIS cloud masking has the potential advantage to better account for 3D-effects caused by neighboring cloud contamination. However, as Aqua is lagging OCO-2 by 15 minutes, we chose a cloud filtering radius of 10 km, to prevent potential cloud movement from introducing cloud contamination. As a result this filter has a throughput of only about 11% and dominates the total pre-filtering throughput of about 4%. This makes this filter the main reason for FOCAL v06 having about three times less data points than NASA's operational product and an OCO-2 based cloud filtering as also done by NASA (Taylor et al., 2016) is a potential solution for future FOCAL versions.

In order to consider not only instrumental noise but also (pseudo) noise of the forward model, we set up a noise model that depends on the instrument noise and one free fit parameter which we determined from the residuals of a set of relatively unconstrained retrievals. The noise model suggests that forward model errors (plus potential pseudo noise of the instrument) have a magnitude of 0.5‰ – 3.2‰ of the continuum radiance. This means that in dark scenes the mismatch of simulated and measured radiance is still dominated by the noise of the instrument but in bright scenes (e.g., above deserts) the forward model error dominates.

Apparent or effective zero level offsets can have various reasons such as residual calibration errors or unconsidered spectroscopic effects. For the SIF, and both CO₂ fit windows, we found linear relationships between the retrieved zero level offsets and the continuum radiances with slopes between 0.8% and 1.8%. As FOCAL v06 usually does not retrieve the zero level offset (ZLO) per sounding, we correct the measured radiance with the derived linear relationships before the retrieval.

Post-filtering checks for convergence, for fit window residuals being smaller than the thresholds derived from the noise model analyses, and for potential outliers. Non converging soundings are often found near the South Atlantic Anomaly (SAA), the Saharan desert, and the Arabian peninsula. Soundings with too large residuals are often found above the tropical oceans. The filter

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	10
---------------------	---	---	-----------

for potential outliers is most active in high latitudes and its dominant input parameter (above land) is the retrieved Ångström exponent. The total post-filtering throughput is about 42%. The average residual to continuum signal ratio (RSR) is 2.2‰, 3.0‰, 2.8‰, and 3.4‰ for the SIF, the O₂, the wCO₂ and the sCO₂ fit window, respectively.

A bias correction is applied to the post-filtered results which primarily bases on the assumption that XCO₂ has (on average) little variations on small scales so that correlations to more variable parameters can be used to quantify biases. As a consequence, the bias correction does not require any ground truth data except for a globally constant offset of -1.67 ppm. We found a distinct OCO-2 footprint dependent bias in the range between -0.97 ppm and 1.22 ppm but the most prominent global bias pattern results from the land/sea bias of 1.80 ppm.

2.2 Performance

We assessed the performance of FOCAL v01 by means of simulated radiance measurements. As accurate XCO₂ retrievals will probably always require a rigorous cloud and aerosol screening, our simulation-based error characterization concentrated on scenarios with scattering optical thicknesses in the range of about 0.03 and 0.70.

The quality of the spectral fits in the O₂ fit window is usually 2.5 to 4 times better than expected from instrumental noise (without added pseudo noise). In the CO₂ fit windows, the quality of the spectral fits is usually at least 7 times better than expected from instrumental noise and even smaller fit residuals are obtained in the SIF fit window.

Systematic errors of XCO₂ range from -2.5 ppm to 3.0 ppm and are usually smaller than ±0.3 ppm (for the tested scenarios). The stochastic uncertainty of XCO₂ is typically about 1.0 ppm. Systematic errors of XH₂O range from -243 ppm to 0 ppm and are usually smaller than ±6 ppm. The stochastic uncertainty of XH₂O is typically about 9 ppm. Note, 1000 ppm ≅ 6.44 kg/m² for the analyzed H₂O profiles. The degree of freedom for the retrieved five-layered CO₂ and H₂O profiles is typically 2.2. As SIF is retrieved from Fraunhofer lines in a spectral region with negligible gaseous absorption features, it can be retrieved without significant interferences with the retrieved scattering properties. The systematic SIF errors are always below 0.02 mW/m²/sr/nm, i.e., it can be expected that instrumental or forward model effects causing an in-filling (a reduction of the line depths) of the used Fraunhofer lines can dominate the systematic errors when analyzing actually measured data without additional corrections. The stochastic uncertainty of SIF is usually below 0.3 mW/m²/sr/nm.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	11
---------------------	---	---	-----------

Without understating the importance of analyzing synthetic measurements, the actual retrieval performance can only be assessed by analyzing measured data including, e.g., pre- and post-filtering, and all kinds of instrumental effects. Therefore, we compared FOCAL v06 XCO₂ results of actually measured (not simulated) OCO-2 data with co-located values of the CAMS v15r4 model. Both data sets show similar large scale spatial patterns and the systematic temporal and regional differences have a standard deviation of 1.0 ppm. The standard deviation of the single sounding mismatches amounts to 1.1 ppm which agrees reasonably well with the average reported uncertainty of 1.2 ppm.

We also compared FOCAL's v06 XCO₂ with the operational NASA OCO-2 product. Large scale patterns of both data sets are similar and the most prominent difference is that the NASA product has about three times more soundings. The primary reason for this is the inherently poor throughput (11%) of the MODIS based cloud screening of FOCAL's preprocessor. The NASA product has on average 0.7 ppm larger values than FOCAL v06. The standard deviation of the difference between both products is 1.1 ppm.

Finally, we validated one year (2015) of FOCAL v06 XCO₂ data with and without bias correction as well as NASA's operational OCO-2 XCO₂ product with and without bias correction with co-located ground based TCCON observations. The algorithms show similarities in the site-to-site patterns of bias and scatter. The standard deviations of the site biases are 0.82 ppm and 0.67 ppm for the NASA product and FOCAL, respectively (0.69 ppm and 0.78 ppm without bias correction). These algorithm-to-algorithm differences are barely significant because TCCON's per site accuracy is about 0.4 ppm (1 σ) (Wunch et al., 2011). The average scatter relative to TCCON is 1.31 ppm and 1.34 ppm for NASA and FOCAL, respectively (1.62 ppm and 1.60 ppm without bias correction).

Additionally, we performed an initial validation of one year (2015) of FOCAL v06 XH₂O data with co-located ground based TCCON observations and found site-to-site biases with a standard deviation of 206 ppm and an average single sounding precision of 293 ppm. However, due to the much larger natural variability of XH₂O compared to XCO₂, future studies are needed to quantify or minimize the influence of representation errors.

2.3 Computational Efficiency

The computational performance of FOCAL is similar to an absorption only retrieval and currently determined by the convolution of the simulated spectra with the instrumental line shape function (ILS).

A typical SCIATRAN (programmed with FORTRAN) vector RT simulation for

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	12
---------------------	---	---	-----------

an OCO-2-like instrument including an optically thin aerosol and cloud layer takes about 32000 s using one core of an Intel Core i7-3770 CPU running at 3.4 GHz (released in 2012). This compares to 0.06 s for FOCAL v01 (programmed in IDL) if only the spectrum and 0.11 s if also the Jacobian is computed. The convolution of spectrum and Jacobian adds 0.22 s and is, therefore, currently the main driver of the total computation time of 0.33 s needed for the forward model of the retrieval.

Processing an entire year of OCO-2 data with FOCAL v06 takes about two weeks on a small cluster with 8 Intel Xeon E5-2687W CPUs with eight cores running at 3.1 GHz (released in 2012). This means, FOCAL is fast enough to process data from current and future satellites similar to CarbonSat (Bovensmann et al., 2010; Buchwitz et al., 2013) providing at least an order of magnitude more data with a reasonable amount of CPU cores - especially, when taking the to be expected CPU developments until launch date into account. Additionally, FOCAL's computations are simple enough for an adaptation to GPU architecture which have been achieved recently and which further substantially improved the computational efficiency.

2.4 Input Data

OCO-2 v8 L1b data (Eldering et al., 2015) are the main input for the FOCAL OCO-2 L2 retrieval. One year has a volume of about 6TB. MODIS Aqua MYD35 collection 6 L2 cloud mask data (Ackerman et al., 2010) are the basis for the cloud filtering. One year has a volume of about 1TB. OMI (Ozone Monitoring Instrument aboard Aura) L3 global daily gridded 1°x 1°UV aerosol index data (Stammes, 2002) are used for the detection of potentially aerosol contaminated scenes. One year has a volume of about 250MB. The ECMWF ERA5 dataset (<http://www.ecmwf.int>) used by FOCAL has a data volume of about 19TB per year. Gaseous absorption cross sections are calculated from NASA's (National Aeronautics and Space Administration) tabulated absorption cross section database ABSCO v4.0 (H₂O) and v5.0 (O₂ and CO₂) (Thompson et al., 2012). We use a high resolution solar irradiance spectrum which we generated by fitting the solar irradiance spectrum of Kurucz (1995) with the high resolution solar transmittance spectrum used by O'Dell et al. (2012). The pre-processed L1 database has a data volume of about 350GB per year.

2.5 Output Data

Only those measurements which fulfill all quality criteria are stored in daily result files in Network Common Data Format (NetCDF). These files contain all the

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	13
---------------------	---	---	-----------

Table 1: List of output parameters contained in daily FOCAL result files in NetCDF file format. Dimensions are defined as number of pixels per orbit (n) and number of profile layers (m=5). More details can be found in the product specification document (PSDv3, Buchwitz et al., 2014).

Parameter	Type	Dimension	Unit	Description
solar_zenith_angle	Float	n	Degrees	Solar zenith angle (0°=zenith)
sensor_zenith_angle	Float	n	Degrees	Viewing zenith angle (0°=nadir)
time	Double	n	Seconds	Seconds since 01.01.1970 00:00 UTC
longitude	Float	n	Degrees	Longitude of pixel centre
latitude	Float	n	Degrees	Latitude of pixel centre
pressure_levels	Float	n×(m+1)	hPa	Retrieval pressure levels
pressure_weight	Float	n×m	-	Pressure weights
sif_760nm	Float	n	mW/m ² /sr/nm	Solar-induced chlorophyll fluorescence at 760nm
xh2o	Float	n	ppm	Retrieved XH ₂ O
xh2o_uncertainty	Float	n	ppm	Uncertainty in retrieved XH ₂ O
xh2o_averaging_kernel	Float	n×m	-	Normalized column averaging kernel for XH ₂ O
h2o_profile_apriori	Float	n×m	ppm	A priori H ₂ O profile
xh2o_quality_flag	Float	n	-	Quality flag for XH ₂ O retrieval (0=good)
xco2	Float	n	ppm	Retrieved XCO ₂
xco2_uncertainty	Float	n	ppm	Uncertainty in retrieved XCO ₂
xco2_averaging_kernel	Float	n×m	-	Normalized column averaging kernel for XCO ₂
co2_profile_apriori	Float	n×m	ppm	A priori CO ₂ profile
xco2_quality_flag	Float	n	-	Quality flag for XCO ₂ retrieval (0=good)

information required for, e.g., surface flux inverse modeling such as retrieved XCO₂ values for individual ground pixels, their errors, corresponding averaging kernels, used a priori profiles, etc. In order to provide this and allow comparisons with other remote sensing or in-situ measurements as well as models, the result files contain the parameters found in Tab. 1. A detailed description of the file format and the primary parameters as well as a manual on how to correctly use them can be found in the product specification document (PSDv3, Buchwitz et al., 2014). The final L2 database has a data volume of about 1.5GB per year.

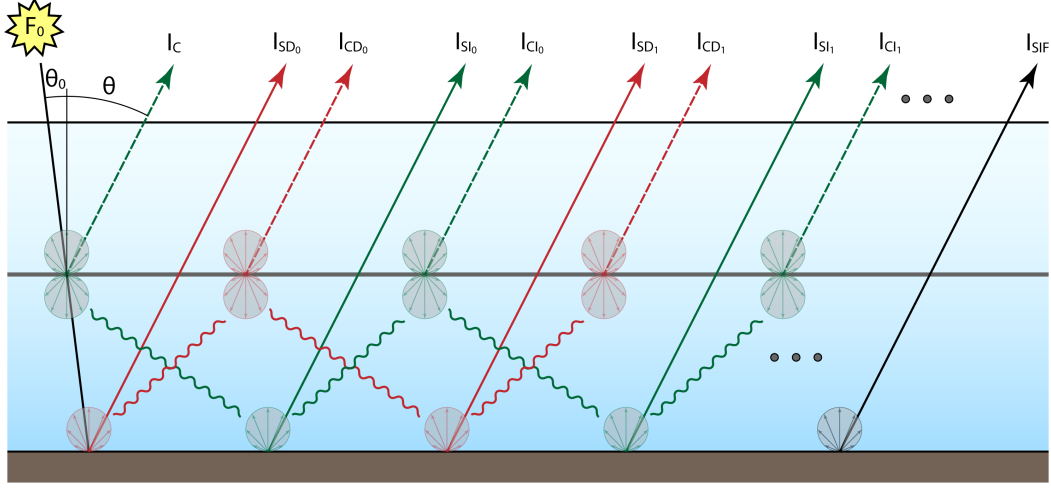


Figure 1: Basic radiative transfer setup with an absorbing atmosphere, a surface with Lambertian reflectance, and an optically thin semi-transparent layer which can partly transmit, absorb, or scatter light in an isotropic way. F_0 is the solar incoming flux, θ_0 and θ are the solar and satellite zenith angles, and I is the radiance reaching the satellite instrument split into components as discussed in the main text. Red represents radiation originating from direct illumination of the surface. Green represents radiation originating from direct illumination of the scattering layer. Arrows represent radiance components reaching the satellite instrument originating from the surface (solid) or from the scattering layer (dashed). Waved lines represent diffuse radiant fluxes.

3 Radiative Transfer

Let, for now, the model atmosphere consist of a plane parallel, vertically heterogeneous, absorbing atmosphere, a surface with Lambertian reflectance, and an optically thin scattering layer of infinitesimal geometrical thickness (Fig. 1). Light hitting the scattering layer may either be transmitted without interaction, absorbed, or isotropically scattered into the upper or lower hemisphere (or half-space). In the following, we derive an equation for the satellite measured radiance I for a plane parallel geometry; in Sec. 3.11, we adapt our results for a pseudo spherical geometry.

We separate the radiance reaching the satellite instrument in the components I_C , I_{SD} , I_{CD} , I_{SI} , I_{CI} , and I_{SIF} :

$$I = I_C + I_{SD} + I_{CD} + I_{SI} + I_{CI} + I_{SIF} \quad (1)$$

I_C is the radiance directly scattered from the scattering layer to the satellite. I_{SD} represents the radiance originating from the surface due to direct illumination

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	15
---------------------	---	---	-----------

of the surface and includes components due to multiple scattering of the Lambertian surface flux (I_{SD_i}). I_{CD} represents the radiance originating from the scattering layer due to direct illumination of the surface including components due to multiple scattering (I_{CD_i}). I_{SI} represents the radiance originating from the surface due to diffuse illumination of the surface including components due to multiple scattering (I_{SI_i}). I_{CI} represents the radiance originating from the scattering layer due to diffuse illumination of the surface including components due to multiple scattering (I_{CI_i}). I_{SIF} is the radiance originating from solar induced chlorophyll fluorescence at 760 nm (SIF) transmitted through the scattering layer but ignoring multiple scattering because of the weak signal.

If not otherwise noted, in the following, F stands for flux, I for intensity (radiance), T for transmittance, τ for vertical optical thickness, and g for gaseous absorption. A superscript s stands for the scattering layer in general. A subscript e, a, and s stand for extinction, absorption, and scattering of the scattering layer, respectively. As an example, the term T_I^g represents a transmittance of intensity through a gaseous absorber.

3.1 Radiance transmission

The transmittance T_I^g along a slant light path through a plane parallel atmospheric layer with gaseous absorption can be computed with Beer-Lambert's law

$$\begin{aligned} T_I^g(\tau_g, \mu) &= e^{-\mu \int K(z) dz} \\ &= e^{-\mu \tau_g} \end{aligned} \quad (2)$$

with K being the absorption coefficient, z the height above the surface, τ_g the total vertical optical thickness, and $\mu = 1/\cos \theta$ the light path extension for the zenith angle θ .

Considering light scattering and absorption within the scattering layer, the fraction of light transmitted through the scattering layer becomes

$$T_I^s(\tau_e, \mu) = e^{-\tau_e \mu} = 1 - S_I(\tau_s, \tau_e, \mu) - A_I(\tau_s, \tau_e, \mu); \quad (3)$$

with $\tau_e = \tau_a + \tau_s$ being the extinction optical thickness, i.e., the sum of absorption (not to be confused with gaseous absorption) and scattering optical thickness. S_I and A_I are the fraction of scattered and absorbed radiance within

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	16
---------------------	---	---	-----------

the scattering layer:

$$S_I(\tau_s, \tau_e, \mu) = \frac{\tau_s}{\tau_e} [1 - T_I^s(\tau_e, \mu)] \quad (4)$$

$$A_I(\tau_s, \tau_e, \mu) = \frac{\tau_a}{\tau_e} [1 - T_I^s(\tau_e, \mu)] \quad (5)$$

3.2 Irradiance transmission

The transmittance of the radiant flux originating from a Lambertian source through a plane parallel atmospheric layer can be computed by integrating over the hemisphere (see, e.g., the textbook of Roedel and Wagner (2011)):

$$T_F^g(\tau_g) = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\frac{\pi}{2}} e^{-\frac{\tau_g}{\cos \theta}} \cos \theta \sin \theta d\theta d\varphi. \quad (6)$$

Integration over the azimuth angle φ and substituting $\mu = 1/\cos \theta$ gives

$$T_F^g(\tau_g) = 2 \int_1^\infty \frac{e^{-\tau_g \mu}}{\mu^3} d\mu, \quad (7)$$

which is basically the definition of the third exponential integral E_3

$$T_F^g(\tau_g) = 2 E_3(\tau_g). \quad (8)$$

Analogously, the flux transmitted through the atmosphere below the scattering layer (with gaseous optical thickness τ_\downarrow) plus the scattering layer becomes

$$T_F^{gs}(\tau_\downarrow + \tau_e) = 2 E_3(\tau_\downarrow + \tau_e). \quad (9)$$

So that the relative additional extinction due to the scattering layer becomes

$$E_F(\tau_e, \tau_\downarrow) = 1 - \frac{E_3(\tau_\downarrow + \tau_e)}{E_3(\tau_\downarrow)}. \quad (10)$$

This can be separated into a fraction of scattered and absorbed flux within the scattering layer:

$$S_F(\tau_s, \tau_e, \tau_\downarrow) = \frac{\tau_s}{\tau_e} E_F(\tau_e, \tau_\downarrow) \quad (11)$$

$$A_F(\tau_s, \tau_e, \tau_\downarrow) = \frac{\tau_a}{\tau_e} E_F(\tau_e, \tau_\downarrow) \quad (12)$$

Note that Eq. 7 could also be interpreted as theorem of equivalence in the form used by Bennartz and Preusker (2006) but accounting only for photon path extensions and a PPDF specific for an isotropic scattering layer.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	17
---------------------	---	---	-----------

3.3 Solar radiation

The solar incoming flux shall be F_0 . As only Lambertian surfaces are considered in our model, the radiance components I_C , I_{SD} , I_{CD} , I_{SI} , and I_{CI} become proportional to

$$I_0 = \frac{F_0}{\pi \mu_0} T_I^g(\tau_{\uparrow}, \mu_0 + \mu). \quad (13)$$

Here τ_{\uparrow} is the gaseous optical thickness above the scattering layer and μ_0 or μ the light path extension for the solar or satellite zenith angle θ_0 or θ . $T_I^g(\tau_{\uparrow}, \mu_0 + \mu)$ corresponds to the transmission along the slant light path from the sun to the scattering layer and from the scattering layer to the satellite.

3.4 I_C

I_C is the radiance directly scattered from the scattering layer to the satellite

$$I_C = I_0 S_I(\tau_s, \tau_e, \mu_0) b, \quad (14)$$

where b corresponds to the fraction of radiation scattered into the hemisphere in backward direction, i.e., the upper or lower hemisphere for light coming from the sun or the surface. Analogously, f is the fraction of radiation scattered into the hemisphere in forward direction and

$$1 = f + b. \quad (15)$$

3.5 I_{SD}

I_{SD} represents the radiance originating from the surface due to direct illumination of the surface and includes components due to multiple scattering of the Lambertian surface flux (I_{SDi}). This means, solar radiation transmits directly through the scattering layer ($T_I^s(\tau_e, \mu_0)$) and the atmosphere below ($T_I^g(\tau_{\downarrow}, \mu_0)$) and illuminates the surface with an Lambertian albedo α . This produces an upward flux which is in parts transmitted, absorbed, and scattered into the upper hemisphere, or back scattered into the lower hemisphere when reaching the scattering layer. The back scattered part contributes to the illumination of the surface and so on. The radiance component I_{SDi} corresponds to the directly transmitted radiance from the surface through the lower atmosphere ($T_I^g(\tau_{\downarrow}, \mu)$), the scattering layer ($T_I^s(\tau_e, \mu)$), and the upper atmosphere after i -times of diffuse reflection between surface and scattering layer ($\alpha S_F(\tau_s, \tau_e, \tau_{\downarrow}) b [T_F^g(\tau_{\downarrow})]^2$).

Summing up all individual radiance components I_{SD_i} results in the following geometric series:

$$I_{SD} = I_0 \alpha T_I^s(\tau_e, \mu_0) T_I^s(\tau_e, \mu) T_I^g(\tau_\downarrow, \mu_0) T_I^g(\tau_\downarrow, \mu) \sum_{i=0}^{\infty} \left(\alpha S_F(\tau_s, \tau_e, \tau_\downarrow) b [T_F^g(\tau_\downarrow)]^2 \right)^i \quad (16)$$

$$= I_0 \alpha T_I^s(\tau_e, \mu_0) T_I^s(\tau_e, \mu) T_I^g(\tau_\downarrow, \mu_0) T_I^g(\tau_\downarrow, \mu) \frac{1}{1 - \alpha S_F(\tau_s, \tau_e, \tau_\downarrow) b [T_F^g(\tau_\downarrow)]^2} \quad (17)$$

3.6 I_{CD}

I_{CD} represents the radiance originating from the scattering layer due to direct illumination of the surface and includes components due to multiple scattering of the Lambertian surface flux (I_{CD_i}). As for I_{SD} , solar radiation transmits directly through the scattering layer ($T_I^s(\tau_e, \mu_0)$) and the atmosphere below ($T_I^g(\tau_\downarrow, \mu_0)$) and illuminates the surface with an Lambertian albedo α . This produces an upward flux which is in parts transmitted, absorbed, and scattered into the upper hemisphere, or back scattered into the lower hemisphere when reaching the scattering layer. The back scattered part contributes to the illumination of the surface and so on. The radiance component I_{CD_i} originates from the scattering layer due to the diffuse surface flux transmitting the lower atmosphere ($T_F^g(\tau_\downarrow)$) and getting scattered into the upper hemisphere ($f S_F(\tau_s, \tau_e, \tau_\downarrow)$) after i -times of diffuse reflection between surface and scattering layer ($\alpha S_F(\tau_s, \tau_e, \tau_\downarrow) b [T_F^g(\tau_\downarrow)]^2$). Summing up all individual radiance components I_{CD_i} results in the following geometric series:

$$I_{CD} = I_0 \alpha T_I^s(\tau_e, \mu_0) S_F(\tau_s, \tau_e, \tau_\downarrow) f T_I^g(\tau_\downarrow, \mu_0) T_F^g(\tau_\downarrow) \frac{1}{1 - \alpha S_F(\tau_s, \tau_e, \tau_\downarrow) b [T_F^g(\tau_\downarrow)]^2} \quad (18)$$

3.7 I_{SI}

I_{SI} represents the radiance originating from the surface due to diffuse illumination of the surface by the scattering layer and includes components due to multiple scattering of the isotropic downward flux of the scattering layer (I_{SI_i}). Here we follow that part of the solar radiation which is diffusely scattered downward by the scattering layer ($f S_I(\tau_s, \tau_e, \mu_0)$) and transmitted to the surface ($T_F^g(\tau_\downarrow)$). The illuminated surface produces an upward flux which is in parts transmitted,

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	19
---------------------	---	---	-----------

absorbed, and scattered into the upper hemisphere, or back scattered into the lower hemisphere when reaching the scattering layer. The back scattered part contributes to the diffuse illumination of the surface and so on. The radiance component I_{SI_i} corresponds to the directly transmitted radiance from the surface through the lower atmosphere ($T_l^g(\tau_{\downarrow}, \mu)$), the scattering layer ($T_l^s(\tau_e, \mu)$), and the upper atmosphere after i-times of diffuse reflection between surface and scattering layer ($\alpha S_F(\tau_s, \tau_e, \tau_{\downarrow}) b [T_F^g(\tau_{\downarrow})]^2$). Summing up all individual radiance components I_{SI_i} results in the following geometric series:

$$I_{SI} = I_0 \alpha S_l(\tau_s, \tau_e, \mu_0) f T_l^s(\tau_e, \mu) T_F^g(\tau_{\downarrow}) T_l^g(\tau_{\downarrow}, \mu) \frac{1}{1 - \alpha S_F(\tau_s, \tau_e, \tau_{\downarrow}) b [T_F^g(\tau_{\downarrow})]^2} \quad (19)$$

3.8 I_{CI}

I_{CI} represents the radiance originating from the scattering layer due to diffuse illumination of the scattering layer and includes components due to multiple scattering of the isotropic downward flux of the scattering layer (I_{CI_i}). Again we follow that part of the solar radiation which is diffusely scattered downward by the scattering layer ($f S_l(\tau_s, \tau_e, \mu_0)$) and transmitted to the surface ($T_F^g(\tau_{\downarrow})$). The illuminated surface produces an upward flux which is in parts transmitted, absorbed, and scattered into the upper hemisphere, or back scattered into the lower hemisphere when reaching the scattering layer. The back scattered part contributes to the diffuse illumination of the surface and so on. The radiance component I_{CI_i} originates from the scattering layer due to the diffuse surface flux transmitting the lower atmosphere ($T_F^g(\tau_{\downarrow})$) and getting scattered into the upper hemisphere ($f S_F(\tau_s, \tau_e, \tau_{\downarrow})$) after i-times of diffuse reflection between surface and scattering layer ($\alpha S_F(\tau_s, \tau_e, \tau_{\downarrow}) b [T_F^g(\tau_{\downarrow})]^2$). Summing up all individual radiance components I_{CI_i} results in the following geometric series:

$$I_{CI} = I_0 \alpha S_l(\tau_s, \tau_e, \mu_0) S_F(\tau_s, \tau_e, \tau_{\downarrow}) f^2 [T_F^g(\tau_{\downarrow})]^2 \frac{1}{1 - \alpha S_F(\tau_s, \tau_e, \tau_{\downarrow}) b [T_F^g(\tau_{\downarrow})]^2} \quad (20)$$

3.9 I_{SIF}

I_{SIF} is the radiance originating from the isotropic solar induced chlorophyll fluorescence flux F_{SIF}^0 at the surface transmitted through the atmosphere

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	20
---------------------	---	---	-----------

($T_l^g(\tau_\downarrow + \tau_\uparrow, \mu)$) and the scattering layer ($T_l^s(\tau_e, \mu)$) but ignoring multiple scattering because of the weak signal.

$$I_{SIF} = \frac{F_{SIF}^0}{\pi} T_l^s(\tau_e, \mu) T_l^g(\tau_\downarrow + \tau_\uparrow, \mu) \quad (21)$$

3.10 Approximations

By means of the following approximations, we are reducing the complexity of the final result which further enhances the computational efficiency. Note that this also considerably reduces the complexity of the (analytic) partial derivatives needed to compute the Jacobian (used by the retrieval).

Due to the high accuracy requirements for the retrieval of greenhouse gases, we are primarily interested in scenarios where scattering at aerosols and clouds is minimal, even if the retrieval algorithm is, in principle, capable of reducing scattering related errors.

Additionally, we are primarily interested in accurate greenhouse gas concentrations; inaccuracies in the retrieved scattering properties are less important. For these reasons, we make an approximation for small extinction optical thicknesses.

Further, we assume that the spectral signal produced by absorption within the scattering layer cannot easily be disentangled from an albedo and scattering signal. For some cases, it is even identical; e.g., when the single scattering albedo ($\omega = \tau_s/\tau_e$) becomes zero, the absorption and the albedo signal become identical. Therefore, we are not aiming to explicitly retrieve the absorption within the scattering layer and approximate that $\tau_a = 0$ (i.e., $\tau_e = \tau_s$). As a result, the retrieved albedo and the amount of scattered radiation may be slightly off, which does not pose a problem as long as the retrieved greenhouse gas concentrations are not affected.

Additionally, we assume that the light is scattered in same parts into the upper and lower hemisphere at the scattering layer ($f = b = 1/2$), which is reasonable especially for an optically thin scattering layer.

First order Taylor series approximation of Eq. 4 and Eq. 3 gives

$$S_l(\tau_s, \mu) \approx \mu \tau_s \text{ and} \quad (22)$$

$$T_l^s(\tau_s, \mu) \approx 1 - S_l(\tau_s, \mu). \quad (23)$$

The amount of diffuse scattered radiant flux (Eq. 11) simplifies to

$$S_F(\tau_s, \tau_\downarrow) \approx \frac{E_2(\tau_\downarrow)}{E_3(\tau_\downarrow)} \tau_s. \quad (24)$$

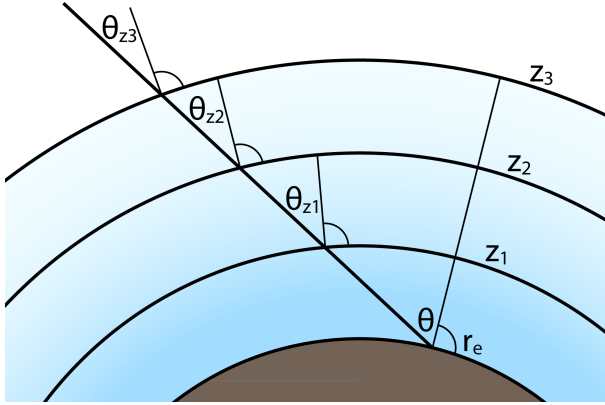


Figure 2: Spherical geometry of the Earth's atmosphere with the Earth's radius r_e , the (solar or satellite) zenith angle θ at the surface and at the heights $z_{1,2,3}$.

Here E_2 is the second exponential integral and $E_2(\tau_\downarrow)/E_3(\tau_\downarrow)$ a number always between 1 and 2.

Substituting Eq. 22–24 into Eq. 17–21 and subsequently first order Taylor series approximation of Eq. 1 at $\tau_s = 0$ yields:

$$\begin{aligned}
 I \approx & \frac{F_0}{\pi \mu_0} T_l^g(\tau_\uparrow, \mu_0 + \mu) \left[\frac{1}{2} \mu_0 \tau_s + \right. \\
 & \alpha \left(T_l^g(\tau_\downarrow, \mu_0 + \mu) [1 - (\mu_0 + \mu) \tau_s + 2 \alpha E_2(\tau_\downarrow) E_3(\tau_\downarrow) \tau_s] + \right. \\
 & \left. T_l^g(\tau_\downarrow, \mu_0) E_2(\tau_\downarrow) \tau_s + T_l^g(\tau_\downarrow, \mu) E_3(\tau_\downarrow) \mu_0 \tau_s \right] + \\
 & \left. \frac{F_{SIF}^0}{\pi} T_l^g(\tau_\downarrow + \tau_\uparrow, \mu) [1 - \mu \tau_s] \right]. \quad (25)
 \end{aligned}$$

3.11 Pseudo-spherical geometry

Due to the spherical geometry of the Earth's atmosphere (Fig. 2), the (solar and satellite) zenith angle changes with height z .

$$\theta(z) = \arcsin\left(\frac{r_e}{r_e + z} \sin \theta\right), \quad (26)$$

with r_e being the Earth's radius and θ the (solar or satellite) zenith angle at the surface.

Correspondingly, also the light path extensions μ and μ_0 become height dependent. In the following, θ , θ_0 , μ , and μ_0 shall refer to values defined at the surface. $\theta(z)$, $\theta_0(z)$, $\mu(z)$, and $\mu_0(z)$ shall refer to height z (Eq. 26) and θ^s ,

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	22
---------------------	---	---	-----------

θ_0^s , μ^s , and μ_0^s shall refer to the scattering layer. This has implications for Eq. 2 which now becomes

$$T_l^g(K(z), \mu(z)) = e^{-\int K(z) \mu(z) dz}. \quad (27)$$

Additionally, μ in Eq. 3, 4, 5, 22, and 24 has to be replaced with the corresponding value at the scattering layer μ^s .

In order to keep the integral in Eq. 6 simple, we do not account for the spherical geometry for the transmission of the diffuse fluxes contributing to multiple scattering. For this reason, we consider this approach a pseudo-spherical approximation.

4 Retrieval

Within this section, we describe how FOCAL v01 retrieves XCO₂ (plus XH₂O and SIF) from an OCO-2 like satellite instrument. However, in principle, the retrieval may be applied to various passive hyperspectral satellite instruments operating in the NIR or SWIR and may be used to gain information on various gaseous species with suitable absorption bands.

The aim of the retrieval is to find the most probable atmospheric state (especially the CO₂ concentration) given an OCO-2 measurement and some a priori knowledge. According to Rodgers (2000) and as done by, e.g., Reuter et al. (2009), this can be achieved by minimizing the cost function

$$\chi^2 = \frac{1}{m+n} [(\vec{y} - \vec{F}(\vec{x}, \vec{b}))^T \mathbf{S}_\epsilon^{-1} (\vec{y} - \vec{F}(\vec{x}, \vec{b})) + (\vec{x} - \vec{x}_a)^T \mathbf{S}_a^{-1} (\vec{x} - \vec{x}_a)]. \quad (28)$$

iteratively with the Gauss-Newton method until convergence is reached.

$$\vec{x}_{i+1} = \vec{x}_i + \hat{\mathbf{S}}_i [\mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} (\vec{y} - \vec{F}(\vec{x}_i, \vec{b})) - \mathbf{S}_a^{-1} (\vec{x}_i - \vec{x}_a)] \quad (29)$$

$$\hat{\mathbf{S}}_i = (\mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} \mathbf{K}_i + \mathbf{S}_a^{-1})^{-1} \quad (30)$$

All quantities used in these equations are explained and discussed in the following.

4.1 Measurement vector y

The measurement vector contains that data measured by the instrument from which we want to gain knowledge about the atmosphere (e.g., the CO₂ concentration). Each of OCO-2's bands consists of 1016 spectral pixels which we group into four fit windows: SIF (~758.26–759.24 nm), O₂ (~757.65–772.56 nm), wCO₂ (~1595.0–1620.6 nm), and sCO₂ (~2047.3–2080.9 nm). The center wavelengths of the individual spectral pixels have been obtained from an example OCO-2 L1b file (oco2_L1bScGL_04243a_150419_B7000r_150608142047.h5, <https://daac.gsfc.nasa.gov>). The separate SIF fit window ensures that the SIF information solely comes from free Fraunhofer lines rather than from O₂ absorption features which makes it much easier to avoid misinterpretations with scattering properties (Frankenberg et al., 2011). The measurement vector is of dimension $m \times 1$ ($m \approx 2600$) and an example is illustrated in Fig. 3 (top). Note that within this section, the measurement vector consists of simulated observations for which the true atmospheric state is known.

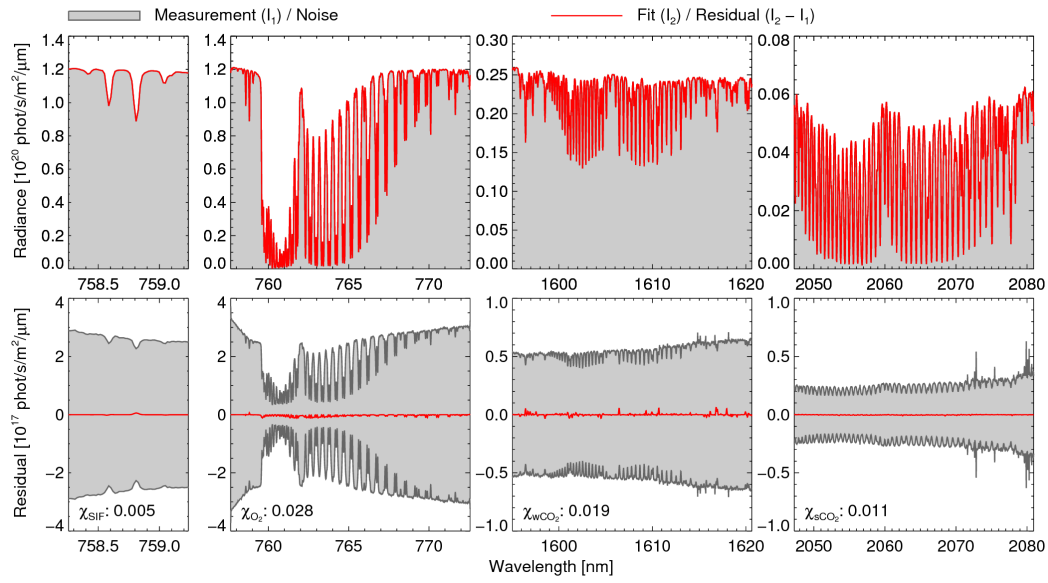


Figure 3: SCIATRAN simulated OCO-2 measurement fitted with FOCAL. Geophysical *baseline* scenario and *0-Scat* retrieval setup, $\theta_0 = 40^\circ$, parallel polarization. See Sec. 7 for definitions of geophysical scenarios and retrieval setups. **Top:** Simulated and fitted radiance measurement in gray and red, respectively. **Bottom:** Simulated measurement noise and fit residual $\vec{\Delta y} = \vec{I}_2 - \vec{I}_1$ (fit minus measurement) in gray and red, respectively. An estimate of the goodness of fit (relative to the noise) in fit window j is computed by $\chi_j = (\frac{1}{m_j} \vec{\Delta y}_j^T \mathbf{S}_{\epsilon_j}^{-1} \vec{\Delta y}_j)^{1/2}$.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	25
---------------------	---	---	-----------

4.2 Measurement error covariance matrix \mathbf{S}_ϵ

Strictly speaking, the measurement error covariance matrix does not only quantify the measurement errors and their correlations; it, additionally, accounts for the forward model error. However, for now, we assume the measurement error to dominate and that no cross correlations exist, i.e., \mathbf{S}_ϵ becomes diagonal. We use the noise parameterization as provided by the same OCO-2 L1b example file mentioned above to compute the diagonal elements of \mathbf{S}_ϵ . The measurement error covariance matrix is of dimension $m \times m$ and an example is illustrated in Fig. 3 (bottom).

4.3 Forward model \vec{F}

The forward model is a vector function of dimension $m \times 1$ that simulates the measurement vector, i.e., OCO-2 measurements. Its inputs are the state and parameter vector defining the geophysical and instrumental state. Primarily, the forward model consists of the RT model described in Sec. 3. The RT computations require a discretization of the atmosphere which we split into 20 homogeneous layers, each containing the same number of dry-air particles (i.e., molecules).

Additionally to the RT calculations, the forward model simulates the instrument by convolving the RT simulations performed on a fixed high resolution wavelength grid with the ILS obtained from the same OCO-2 L1b example file mentioned above. Furthermore, the forward model has the ability to simulate zero level offsets (i.e., additive radiance offsets), shift and squeeze the wavelength axes of the fit windows according to Eq. 31, and squeeze the ILS according to Eq. 33.

$$\lambda' = \lambda + \lambda_{sh} + \lambda_n \lambda_{sq} \quad (31)$$

$$\lambda_n = 2 - 4 \frac{\lambda_1 - \lambda}{\lambda_1 - \lambda_0} \quad (32)$$

Here λ' is the modified wavelength, λ the nominal wavelength, λ_{sh} the wavelength shift parameter, λ_n the normalized nominal wavelength, λ_{sq} the wavelength squeeze parameter, and $\lambda_{0,1}$ the minimum or maximum of λ , respectively. The normalization of λ is done in a way that the average absolute value of λ_n is approximately one.

$$\lambda'_{ILS} = \lambda_{ILS} ILS_{sq} \quad (33)$$

Here λ'_{ILS} is the modified ILS wavelength computed from the nominal ILS λ_{ILS} wavelength and the squeeze parameter ILS_{sq} .

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	26
---------------------	---	---	-----------

4.4 State vector \vec{x}

The state vector consists of all quantities which we retrieve from the measurement and is of dimension $n \times 1$ with $n = 36$. The dry-air mole fractions of water vapor (H_2O) and CO_2 are retrieved from both CO_2 fit windows within five layers splitting the atmosphere into parts containing the same number of dry-air particles. This means, each CO_2 and H_2O layer spans over four atmospheric layers used for the discretized RT calculations. The CO_2 and H_2O concentrations are homogeneous within each of the five layers. XCO_2 and XH_2O are not part of the state vector but are calculated during the post processing from the layer concentrations.

SIF at 760 nm is derived from the SIF fit window by scaling the SIF reference spectrum F_{SIF}^0 . The scattering parameters pressure (i.e., height) of the scattering layer p_s (in units of the surface pressure p_0), scattering optical thickness at 760 nm τ_s , and Ångström exponent \AA are derived from all fit windows simultaneously.

Within the SIF fit window, FOCAL additionally fits a first order polynomial of the spectral albedo $\alpha P_{0,1}$ and shift and squeeze of the wavelength axis $\lambda_{\text{sh,sq}}$. Within the other fit windows, FOCAL additionally fits a second order polynomial of the spectral albedo $\alpha P_{0,1,2}$, shift and squeeze of the wavelength axis, and a squeeze of the instrumental line shape function ILS_{sq} .

We estimate the first guess zeroth order albedo polynomial coefficients αP_0 from the continuum reflectivities $R_0 = \pi \mu_0 I / F_0$ using up to nine spectral pixels at the fit windows' lower wavelength length ends. The first guess profiles of H_2O and CO_2 are obtained from ECMWF (European Centre for Medium-Range Weather Forecasts) analysis fields and SECM2016, respectively. SECM2016 corresponds to the simple empirical carbon model of Reuter et al. (2012) but trained with version CT2016 of the CarbonTracker model (Peters et al., 2007). All other first guess state vector elements are scene independent and the a priori state vector \vec{x}_a equals the first guess state vector \vec{x}_0 .

Tab. 2 summarizes the state vector composition including the used fit windows, a priori \vec{x}_a and first guess \vec{x}_0 values, a priori uncertainties $\sigma \vec{x}_a$, and typical values of a posteriori uncertainties $\sigma \hat{\vec{x}}$ and the degrees of freedom for signal d_s .

4.5 A priori error covariance matrix \mathbf{S}_a

The a priori error covariance matrix defines the uncertainties of the a priori state vector elements and their correlations. Its dimensionality is $n \times n$. Except for the CO_2 and H_2O profile layers, we assume \mathbf{S}_a to be diagonal. As described by Reuter et al. (2012), we compute the CO_2 layer-to-layer covariances by

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	27
---------------------	---	---	-----------

Table 2: State vector composition of the baseline (FOCALv01), i.e., the *3-Scat* retrieval setup (see Sec. 7 for definition of retrieval setups). From left to right, the columns represent the name of the state vector element, its sensitivity within the four fit windows, a priori \vec{x}_a and first guess \vec{x}_0 value, the a priori uncertainty $\sigma\vec{x}_a$, the a posteriori uncertainty $\sigma\hat{\vec{x}}$, and the degrees of freedom d_s . A posteriori uncertainty and degrees of freedom represent results of the geophysical *Rayleigh* scenario, $\theta_0 = 40^\circ$, and perpendicular polarization.

State vector element	Fit window sensitivity				\vec{x}_a, \vec{x}_0	$\sigma\vec{x}_a$	$\sigma\hat{\vec{x}}$	d_s
	SIF	O ₂	wCO ₂	sCO ₂				
αP_0^{SIF}	•				R_0^{SIF}	0.1	0.0016	1.00
αP_1^{SIF}	•				0.0	0.01	0.0008	0.99
$\alpha P_0^{\text{O}_2}$		•			$R_0^{\text{O}_2}$	0.1	0.0000	1.00
$\alpha P_1^{\text{O}_2}$		•			0.0	0.01	0.0000	1.00
$\alpha P_2^{\text{O}_2}$		•			0.0	0.01	0.0000	1.00
$\alpha P_0^{\text{wCO}_2}$			•		$R_0^{\text{wCO}_2}$	0.1	0.0001	1.00
$\alpha P_1^{\text{wCO}_2}$			•		0.0	0.01	0.0000	1.00
$\alpha P_2^{\text{wCO}_2}$			•		0.0	0.01	0.0000	1.00
$\alpha P_0^{\text{sCO}_2}$				•	$R_0^{\text{sCO}_2}$	0.1	0.0000	1.00
$\alpha P_1^{\text{sCO}_2}$				•	0.0	0.01	0.0000	1.00
$\alpha P_2^{\text{sCO}_2}$				•	0.0	0.01	0.0000	1.00
$\lambda_{\text{sh}}^{\text{SIF}}$ [nm]	•				0.0	0.01	0.0001	1.00
$\lambda_{\text{sq}}^{\text{SIF}}$ [nm]	•				0.0	0.01	0.0002	0.94
$\lambda_{\text{sh}}^{\text{O}_2}$ [nm]		•			0.0	0.01	0.0000	1.00
$\lambda_{\text{sq}}^{\text{O}_2}$ [nm]		•			0.0	0.01	0.0000	1.00
$\text{ILS}_{\text{sq}}^{\text{O}_2}$		•			1.0	0.01	0.0000	1.00
$\lambda_{\text{sh}}^{\text{wCO}_2}$ [nm]			•		0.0	0.01	0.0000	1.00
$\lambda_{\text{sq}}^{\text{wCO}_2}$ [nm]			•		0.0	0.01	0.0000	1.00
$\text{ILS}_{\text{sq}}^{\text{wCO}_2}$			•		1.0	0.01	0.0000	1.00
$\lambda_{\text{sh}}^{\text{sCO}_2}$ [nm]				•	0.0	0.01	0.0000	1.00
$\lambda_{\text{sq}}^{\text{sCO}_2}$ [nm]				•	0.0	0.01	0.0000	1.00
$\text{ILS}_{\text{sq}}^{\text{sCO}_2}$				•	1.0	0.01	0.0000	1.00
SIF [mW/m ² /sr/nm]	•				0.0	10.0	0.29	1.00
p_s [p_0]	•	•	•	•	0.2	1.0	0.002	1.00
τ_s	•	•	•	•	0.01	0.1	0.0001	1.00
\hat{A}	•	•	•	•	4.0	2.0	0.29	0.98
H ₂ O L ₀ [ppm]			•	•	ECMWF	2179.9	497.5	0.84
H ₂ O L ₁ [ppm]			•	•	ECMWF	2186.9	849.8	0.55
H ₂ O L ₂ [ppm]			•	•	ECMWF	1066.0	415.2	0.59
H ₂ O L ₃ [ppm]			•	•	ECMWF	205.4	151.4	0.21
H ₂ O L ₄ [ppm]			•	•	ECMWF	2.67	2.59	0.00
CO ₂ L ₀ [ppm]			•	•	SECM2016	21.8	6.5	0.65
CO ₂ L ₁ [ppm]			•	•	SECM2016	14.1	4.3	0.29
CO ₂ L ₂ [ppm]			•	•	SECM2016	12.7	6.1	0.27
CO ₂ L ₃ [ppm]			•	•	SECM2016	12.0	5.9	0.42
CO ₂ L ₄ [ppm]			•	•	SECM2016	16.8	9.9	0.60
XH ₂ O [ppm]					ECMWF	898.2	9.2	2.2
XCO ₂ [ppm]					SECM2016	10.0	1.0	2.2

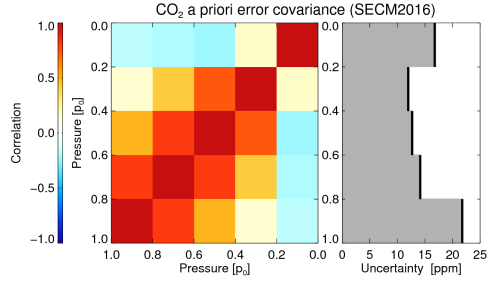


Figure 4: CO₂ a priori error covariance computed from randomly chosen SECM2016 profiles and corresponding CT2016 profiles. The CO₂ layer variances have been up-scaled so that the a priori XCO₂ uncertainty becomes 10 ppm (1 ppm without scaling). **Left:** Layer-to-layer correlation matrix of the a priori uncertainty. **Right:** 1 σ a priori uncertainty.

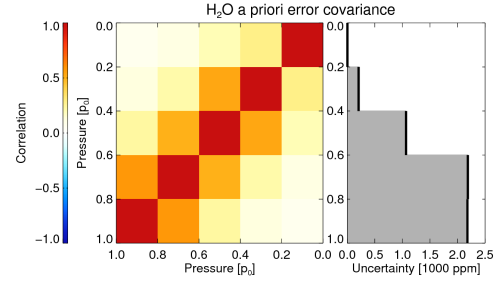


Figure 5: As Fig. 4 but for H₂O and estimated from day-to-day variations of ECMWF analysis profiles (without variance scaling as done for CO₂).

comparing randomly chosen SECM2016 profiles with corresponding CT2016 model profiles. The CO₂ layer variances have been up-scaled so that the a priori XCO₂ uncertainty becomes 10 ppm (1 ppm $\hat{=}$ 2.5‰ without scaling). This ensures retrievals to be dominated by the measurement but not the a priori. We estimated the H₂O layer-to-layer covariances by analyzing H₂O day-to-day variations of ECMWF analysis profiles. CO₂ and H₂O a priori error covariances are shown in Fig. 4 and Fig. 5. All other (diagonal) elements of \mathbf{S}_a are listed in row $\sigma\bar{x}_a$ of Tab. 2.

4.6 Jacobian matrix K

The Jacobian matrix includes the first order derivatives of the forward model with respect to all state vector elements and has a dimensionality of $m \times n$. A measurement can only include information on those state vector elements which have sufficiently linearly independent derivatives. Fig. 6 illustrates the content of a typical example of a Jacobian matrix. Note that the sensitivity to SIF has artificially been set to zero in the O₂ fit window in order to ensure, that the SIF information solely comes from the SIF fit window and misinterpretations with scattering parameters are avoided (Frankenberg et al., 2011).

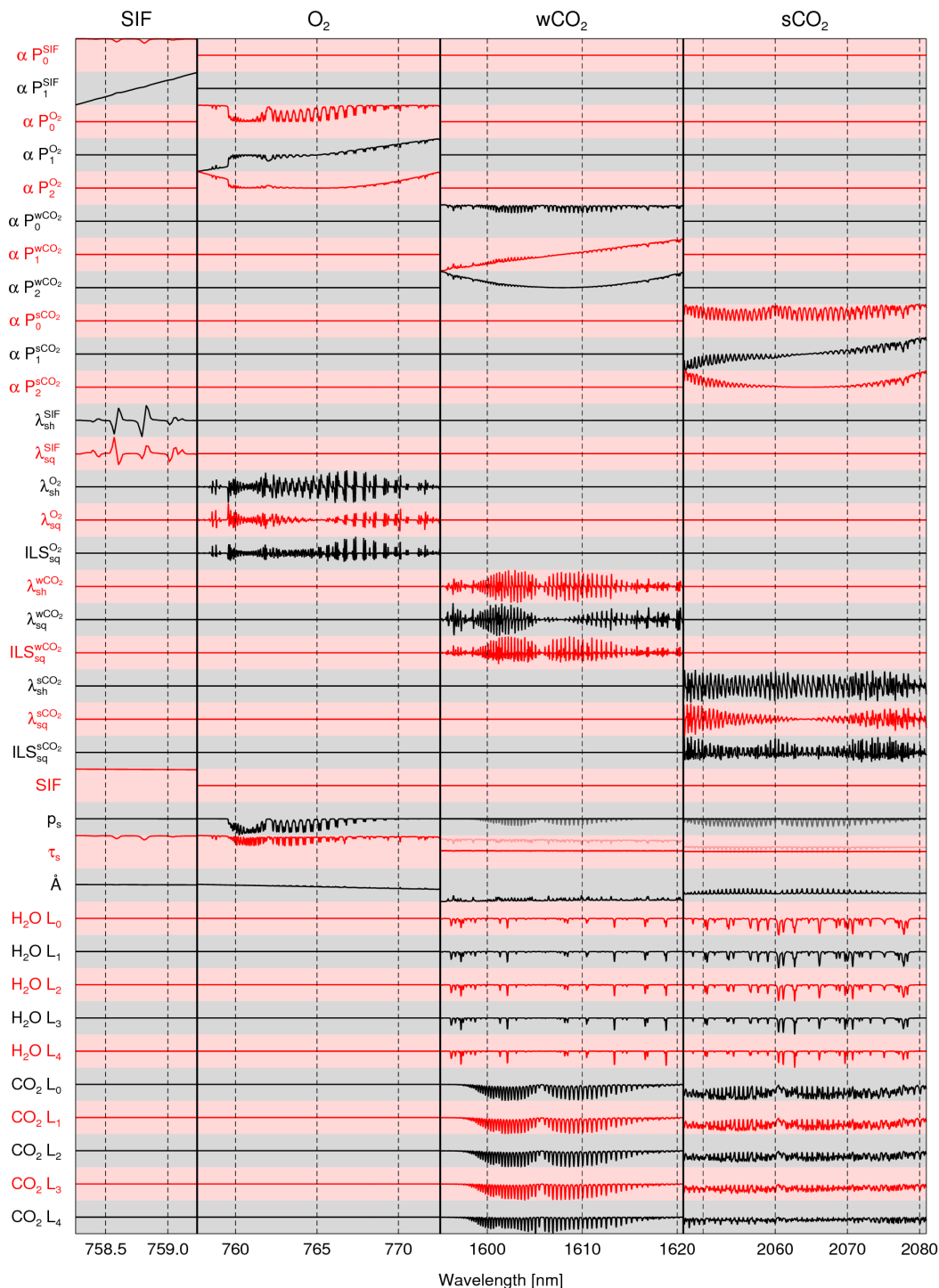


Figure 6: Jacobian matrix computed with FOCAL v01 for the geophysical *Rayleigh* scenario and the *3-Scat* retrieval setup. Within the CO₂ fit windows, an additional line in light colors shows the partial derivatives according to τ_s and p_s scaled by a factor of 10 and 20, respectively.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	30
---------------------	---	---	-----------

4.7 Parameter vector \vec{b}

The state vector includes only a small subset of geophysical and instrumental properties that influence a simulated radiance measurement. All these additional properties are assumed to be known and form the parameter vector \vec{b} .

The observation geometry (particularly, the solar and satellite zenith angles θ_0 and θ), Earth/Sun distance, Doppler shifts, ILS, measurement wavelength grid, etc. are used as provided or calculated from data in the satellite L1b orbit files. Atmospheric temperature, pressure, and dry-air sub-column profiles are obtained from ECMWF analysis data. Gaseous absorption cross sections are calculated from NASA's (National Aeronautics and Space Administration) tabulated absorption cross section database ABSCO v4.0 (H₂O) and v5.0 (O₂ and CO₂) (Thompson et al., 2012).

We use a high resolution solar irradiance spectrum (F_0) which we generated by fitting the solar irradiance spectrum of Kurucz (1995) with the high resolution solar transmittance spectrum used by O'Dell et al. (2012), a forth order polynomial, and a Gaussian ILS. The used solar induced chlorophyll fluorescence irradiance spectrum (F_{SIF}^0) has been obtained from the publication of Rascher et al. (2009) and scaled to 1.0 mW/m²/sr/nm at 760 nm. In order to account for OCO-2 measuring one polarization direction only, we divided the solar and the chlorophyll fluorescence irradiance spectrum by a factor of two.

All FOCAL v01 RT simulations are performed at a high resolution wavelength grid (not to be confused with the measurement wavelength grid) with a sampling distance of 0.001 nm for the SIF and the O₂ fit window and 0.005 nm for both CO₂ fit windows.

4.8 A posteriori error covariance matrix $\hat{\mathbf{S}}$

Once convergence is achieved, the a posteriori error covariance matrix includes the a posteriori uncertainties of the retrieved state vector elements and their correlations. It has a dimensionality of $n \times n$.

4.9 Convergence

We define that convergence is achieved when the state vector increment is small compared to the a posteriori error. Specifically, we stop iterating once:

$$\frac{1}{n} [(\vec{x}_i - \vec{x}_{i-1})^T \hat{\mathbf{S}}^{-1} (\vec{x}_i - \vec{x}_{i-1})] < 0.2. \quad (34)$$

Additionally, we test if χ^2 is smaller than 2. The maximum number of allowed iterations is 15.

5 Preprocessing

The mathematical bases of FOCAL's retrieval method is described in detail in Sec. 3 and 4. However, in order to analyze actually measured data instead of simulations, pre-filtering of the OCO-2 L1b calibrated radiances, adjustments of the noise model, and accounting for potential zero level offsets is required. This section refers to FOCAL v06.

5.1 Filtering

During preprocessing, we collect all datasets that are needed to run the retrievals and pre-filter soundings with potentially degraded quality or potential cloud or aerosol contamination. Due to the demanding precision and accuracy requirements for XCO₂ retrievals (e.g., Miller et al., 2007; Chevallier et al., 2007; Bovensmann et al., 2010) and the large amount of OCO-2 data, we prioritize quality over quantity in the course of pre-filtering.

The primary input data used in this section are global OCO-2 L1b calibrated radiances (i.e., a single linear component of the polarization of the incoming light) version 7r (Eldering et al., 2015; Crisp et al., 2017) of the year 2015 in glint (GL), nadir (ND), target (TG), and transition (XS) mode which have been obtained from <https://daac.gsfc.nasa.gov>. Each of OCO-2's three bands consists of 1016 spectral pixels which we group into the four fit windows illustrated in Fig. 7 showing a typical OCO-2 measurement fitted with FOCAL.

Each L1b orbit file includes information on spectral pixels with potentially reduced quality, e.g., due to radiometric problems. Based on the last nadir orbit in 2015 (oco2_L1bScND_07974a_151231_B7200r_160121043229.h5), we generated a dead or bad pixel mask which we use for the O₂ and both CO₂ fit windows. For the SIF fit window, we ignore the dead or bad pixel mask because it is located in a spectral region generally flagged as potentially bad.

We reject all soundings flagged to have potentially reduced quality (quality flag≠0) or failing a data integrity test (e.g., unreasonable sounding ID or time). We filter out potentially "tricky" scenes with solar or satellite zenith angles greater than 70°, latitudes beyond ±70°, or extreme surface roughnesses (standard deviation of the surface elevation) greater than 1000 m. In Fig. 8, this filter is referred to as LAT/SUZ/SAZ/ σ ALT.

We use the spike EOF analysis provided with the OCO-2 L1b data (Eldering et al., 2015) and accept only soundings with less than 60 spectral pixels with potentially poor quality (referred to as bad colors) in the O₂ and no bad colors in the CO₂ bands. This primarily filters out soundings above South America

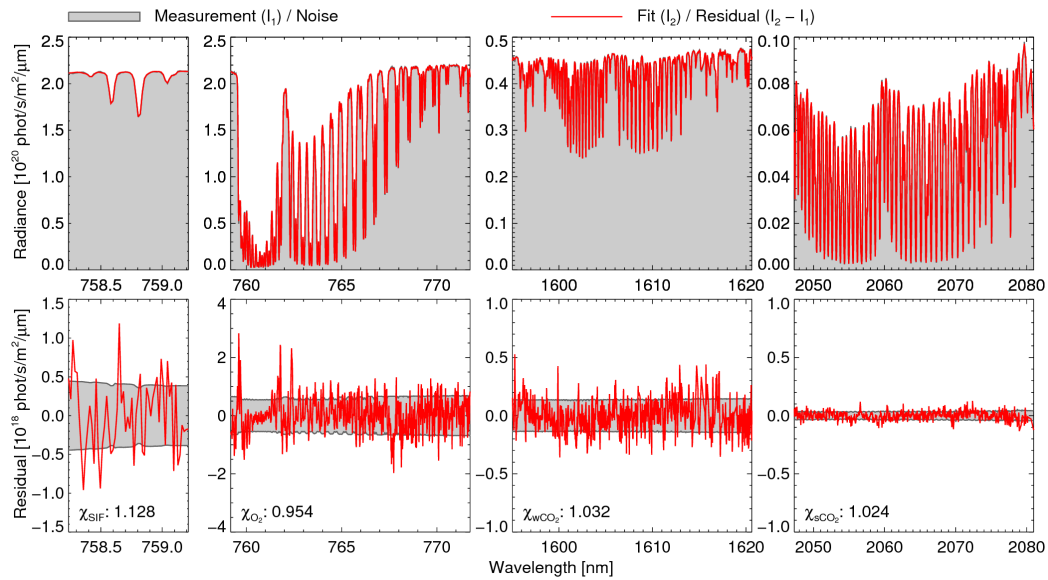


Figure 7: OCO-2 measurement of June 6, 2015, 12:01 UTC near Hamburg, Germany (sounding ID: 2015060512011938) fitted with FOCAL v06. **Top:** Simulated and fitted radiance measurement in gray and red. **Bottom:** Adapted measurement noise (see Sec. 5.2) and fit residual (fit minus measurement) in gray and red; the root-mean-square deviation (RMSD) is in units of the residual. The RMSD expected from noise is given in brackets.

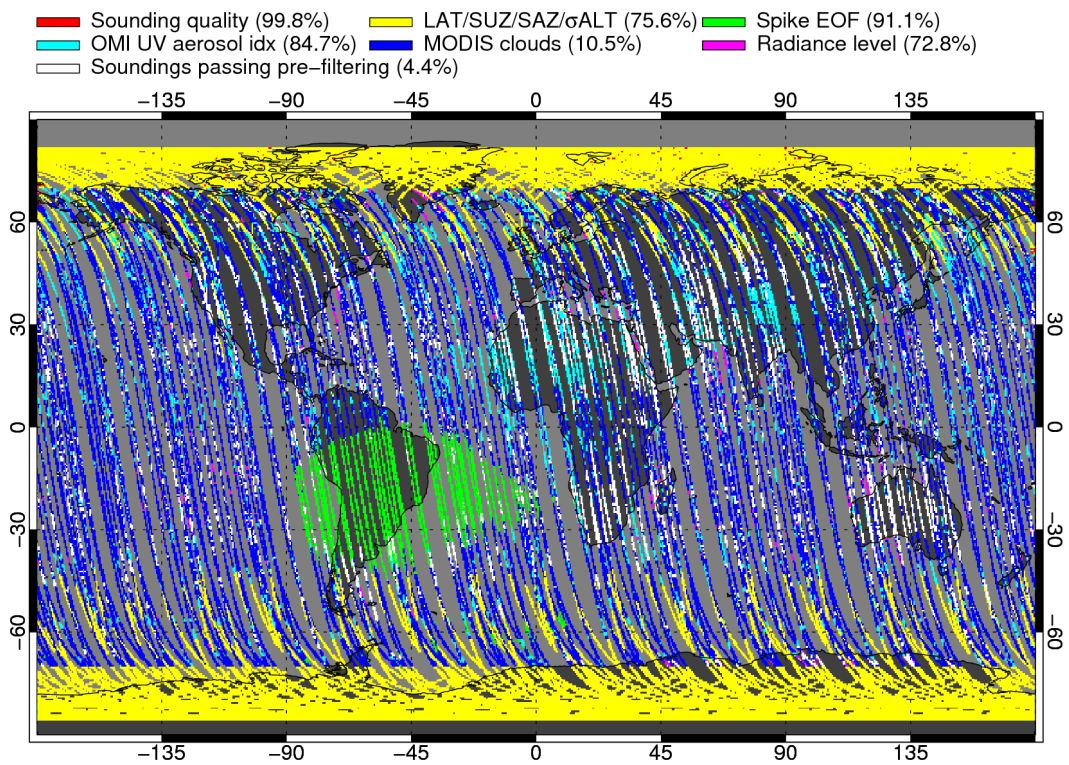


Figure 8: Pre-filtering statistics of the 24 days data subset used for the noise model analysis (Sec. 5.2). The filters are applied in the order: Sounding quality, LAT/SUZ/SAZ/ σ ALT, Spike EOF, OMI UV aerosol idx, MODIS clouds, and Radiance level (see main text for a description). The colors represent filter activity and soundings passing all filters are shown in white. Numbers in brackets represent filter throughputs.

and the South Atlantic because of contamination by cosmic rays within the SAA caused by the shape of the inner Van Allen radiation belt (Fig. 8).

Potentially aerosol contaminated scenes are filtered using OMI (Ozone Monitoring Instrument aboard Aura lagging OCO-2 by 30 minutes) L3 global daily gridded $1^\circ \times 1^\circ$ UV aerosol index data (OMAERUVd v003 obtained from <https://daac.gsfc.nasa.gov>, Stammes, 2002) with a filtering threshold of one. As described by Stammes (2002), the UV aerosol index is obtained by comparing the measured reflectance ratio at two wavelengths (342.5 nm and 388.0 nm) to the calculated reflectance ratio using a Rayleigh atmosphere with an assumed surface albedo. The UV index is relatively insensitive to scattering aerosol layers or clouds, because it is mainly determined by the reduction of Rayleigh multiple scattering due to aerosol absorption. As illustrated in Fig. 8, this filter most prominently impacts regions contaminated with desert dust aerosols.

Potentially cloud contaminated scenes are filtered using MODIS Aqua (moderate-resolution imaging spectroradiometer aboard Aqua) L2 cloud mask data with about $1 \text{ km} \times 1 \text{ km}$ resolution (collection 6, MYD35, obtained from <https://ladsweb.modaps.eosdis.nasa.gov>, Ackerman et al., 2010). All MODIS ground pixels which are not flagged as clear or probably clear are considered as potentially cloudy. Aqua is lagging OCO-2 by 15 minutes and in order to account for the parallax effect and potential cloud movements, we use only OCO-2 data with at least 10 km distance to the nearest MODIS cloud. Even though 10 km is not overly conservative, this filter has a throughput of only about 11% and dominates the total pre-filtering throughput of about 4% (Fig. 8).

Additionally, we filter out very dark or bright scenes, i.e., extreme detector fillings. Specifically, we ensure that the continuum radiance in each band is between 5% and 95% of the maximum band radiance as specified by Eldering et al. (2015).

Meteorological profiles come from ECMWF operational analysis data (<http://apps.ecmwf.int>) and have a resolution of six hours, $0.75^\circ \times 0.75^\circ$, and 137 height layers. As part of the preprocessor, these profiles are corrected for the actual surface height of the OCO-2 soundings and split into 20 layers containing the same number of dry-air particles.

5.2 Noise Model

The measurement error covariance matrix (Reuter et al., 2017c) has to account not only for the measurement noise but for the total error including also the forward model error. The measurement noise of the instrument is well known from laboratory measurements and in-flight estimates. In theoretical studies,

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	35
---------------------	---	---	-----------

as in Sec. 7, it is often assumed for convenience, that the measurement noise dominates and that other error components can be neglected, i.e., the noise model is approximated by the measurement noise.

Especially, when analyzing measured data, unknown inaccuracies of the forward model can violate this assumption and lead to larger fit residuals and unrealistic results (and error estimates) because the optimal estimation retrieval puts too much trust in the measurement. This may happen, e.g., due to imperfect knowledge of the ILS, unconsidered spectroscopic effects such as Raman scattering, inaccuracies of the spectroscopic data bases, approximations of the radiative transfer model, or imperfect meteorology.

Ideally, one would reduce the fit residuals to the instrument's noise level by improving the forward model, but this is often not possible. A potential solution is to fit parts of the residuum by empirical orthogonal functions (EOF) computed from a representative set of measurements as done by Boesch et al. (2015). Another approach is to adjust the noise model so that it accounts for measurement noise plus forward model error (e.g., O'Dell et al., 2012; Yoshida et al., 2013; Heymann et al., 2015) and a variant of this approach is also used by us.

Most forward model errors can be interpreted to result from inaccuracies of the computed (effective) atmospheric transmittance. However, the largest scene-to-scene variability of the simulated radiance is due to changes of, e.g., albedo and solar zenith angle. Therefore, it is reasonable to assume forward model errors to be approximately proportional to the continuum signal I_{cont} which we obtain from up to nine spectral pixels at the fit windows' lower wavelength length ends.

We model the root mean square RSR by

$$RSR = \sqrt{NSR^2 + \delta F^2}, \quad (35)$$

where NSR represents the root mean square of the spectral 1σ radiance noise (as reported in the OCO-2 L1b data) to continuum signal ratio and δF the relative forward model error.

In order to estimate the free parameter δF , we analyzed a representative set of pre-filtered soundings (Fig. 9) with a modified FOCAL setup for which we (quadratically) added 2% of the continuum radiance to the measurement noise. This overestimation of the expected total error effects that the retrieval usually converges towards values being not very far away from the a priori, i.e., values being more or less realistic. Additionally, we switched off the SIF retrieval (which is basically identical to a zero level offset in the SIF fit window) and switched on the retrieval of zero level offsets in all four fit windows.

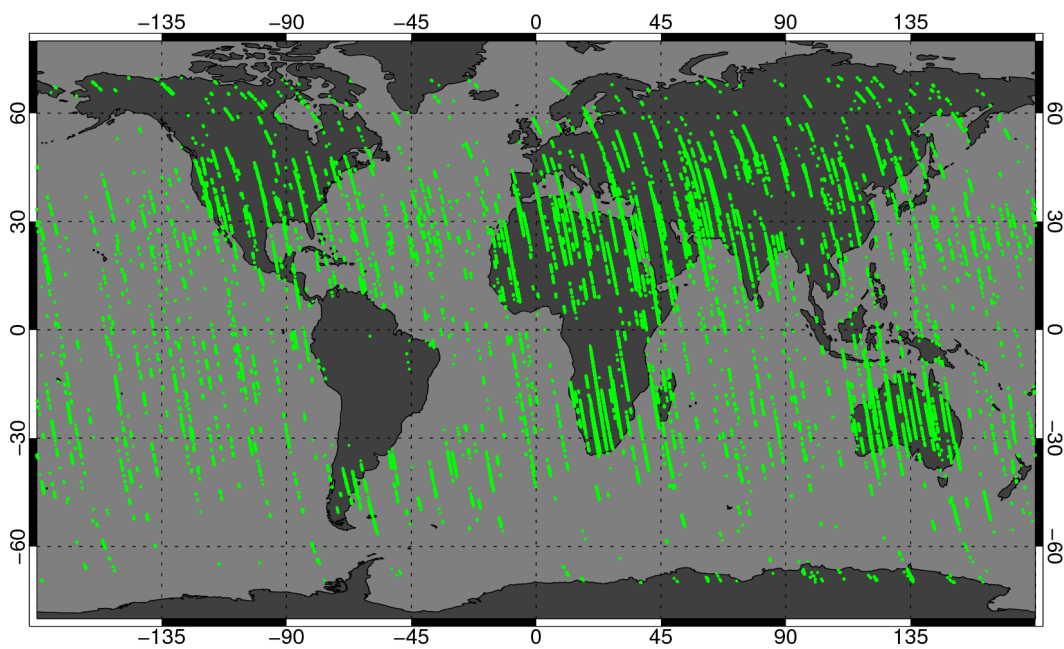


Figure 9: Sampling of all pre-filtered soundings analyzed in order to determine the noise model. The data set consists of 10% of all pre-filtered OCO-2 soundings (randomly selected) of 24 days in 2015 (13.01., 15.01., 14.02., 16.02., 10.03., 20.03., 03.04., 19.04., 08.05., 23.05., 08.06., 24.06., 15.07., 16.07., 15.08., 16.08., 15.09., 16.09., 15.10., 16.10., 15.11., 17.11., 12.12., 18.12.). This results in a manageable but still representative data set with respect to nadir/glint observation geometry, season, and spatial distribution.

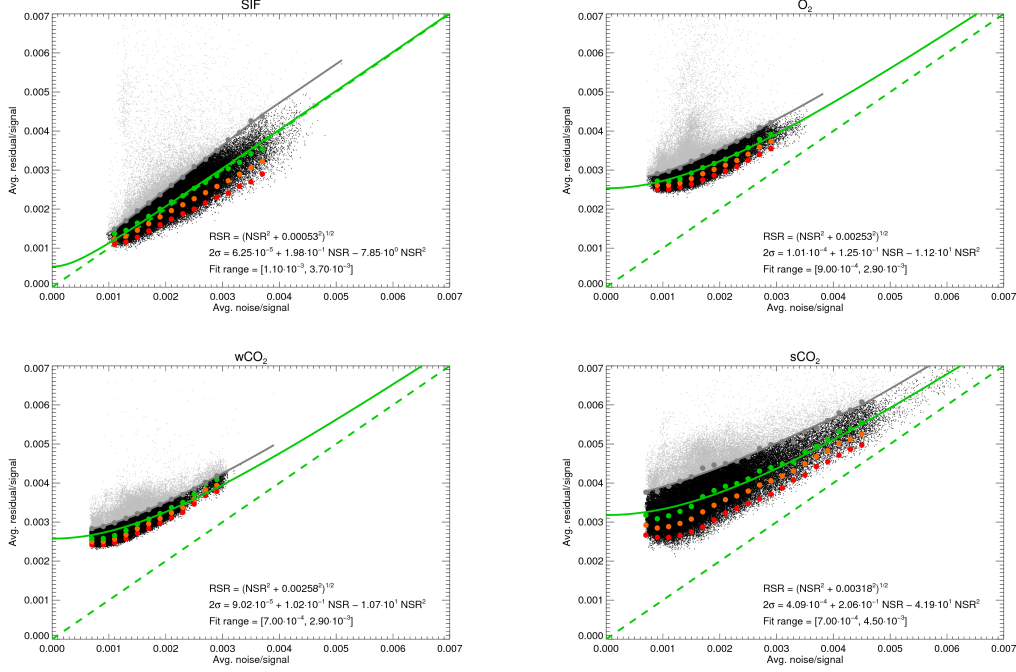


Figure 10: Root mean square noise to signal ratio NSR versus root mean square residual to signal ratio RSR for all four fit windows. **red points:** 2.28th percentile within bins with more than 500 samples (35 bins in total). **orange points:** 15.9th percentile. **green points:** expectation value estimated from the 2.28th and 15.9th percentile. **solid green line:** RSR as computed from the RSR model (Eq. 35). **gray points:** RSR model plus 2σ estimated from the 2.28th and 15.9th percentile. **gray line:** outlier threshold. **gray dots:** potential outliers. **dashed green line:** one-to-one line.

If the instrument noise would dominate the total error, RSR and NSR would (statistically) lie on a 1:1 line. After the removal of outliers (Fig. 10, gray dots), this is basically the case for the SIF fit window with forward model errors estimated to be about 0.5‰ of the (continuum) signal (Fig. 10). The forward model error within the other fit windows is estimated to be between 2.5‰ and 3.2‰ (Fig. 10). This means, the total error in dark scenes (large NSR) is still dominated by the instrumental noise but in bright scenes (small NSR), the forward model error dominates.

Outliers are removed as follows: The data set is grouped in 35 NSR bins. Only bins with more than 500 samples are further considered. Within each bin, RSR should follow a χ^2 -distribution with as many degrees of freedom as spectral pixels of the fit window. The number of spectral pixels is always large enough to approximate the χ^2 -distribution with a Gaussian distribution.

Outliers represent poor fits, e.g., due to complicated atmospheric conditions which cannot be well described by the forward model. As they usually enhance the RSR, we have to approach the expectation value of RSR from the lowermost values. The 2.28th and 15.9th percentile (Fig. 10, red and orange points) of the Gaussian distribution are two and one standard deviations (2σ) smaller than the expectation value. We used this to estimate the expectation value (Fig. 10, green points) from which we determined the free fit parameter δF of Eq. 35 (numerical values are shown in Fig. 10). Note that adding 4% instead of 2% of the continuum radiance to the measurement noise gave similar results (not shown here).

Soundings with a RSR being more than 2σ larger than expected from Eq. 35 are considered outliers. For this purpose, we fitted the second order polynomial

$$2\sigma = a_0 + a_1 NSR + a_2 NSR^2 \quad (36)$$

and use it as threshold for the maximal allowed deviation from the RSR model (Fig. 10, gray lines).

We define the noise model which modifies the reported OCO-2 L1b radiance noise N analog to Eq. 35:

$$N' = \sqrt{N^2 + I_{\text{cont}}^2 \delta F^2}. \quad (37)$$

5.3 Zero level offset correction

We define as ZLO an additive fit window-wide radiance offset. An apparent or effective ZLO can have various reasons such as residual calibration errors or unconsidered spectroscopic effects. Many of these effects can be expected to result in ZLOs being approximately proportional to the fit window's continuum radiance. In order to study potential ZLOs, we used the same modified FOCAL setup as in the last section but with the just defined noise model. The simultaneous retrieval of ZLOs reduce the uncertainty reduction for XCO₂ and renders the SIF retrieval impossible. Therefore, we aimed at a ZLO correction rather than a ZLO retrieval per sounding. We analyzed the same 24 days of OCO-2 data as in the last section but filtered for potential contamination with chlorophyll fluorescence because in the SIF fit window it is not possible to disentangle ZLO and SIF (Fig. 11). For this purpose, we used monthly L3 MODIS Aqua chlorophyll-a data (obtained from https://modis.gsfc.nasa.gov/data/dataproduct/chlor_a.php, Hu et al., 2012) over ocean and normalized difference vegetation index (NDVI) data over land (obtained from <https://modis.gsfc.nasa.gov/data/dataproduct/mod13.php>).

Fig. 12 shows that we find a reasonably linear relationship (with correlations around 0.9) between the retrieved ZLO and the continuum radiance within the

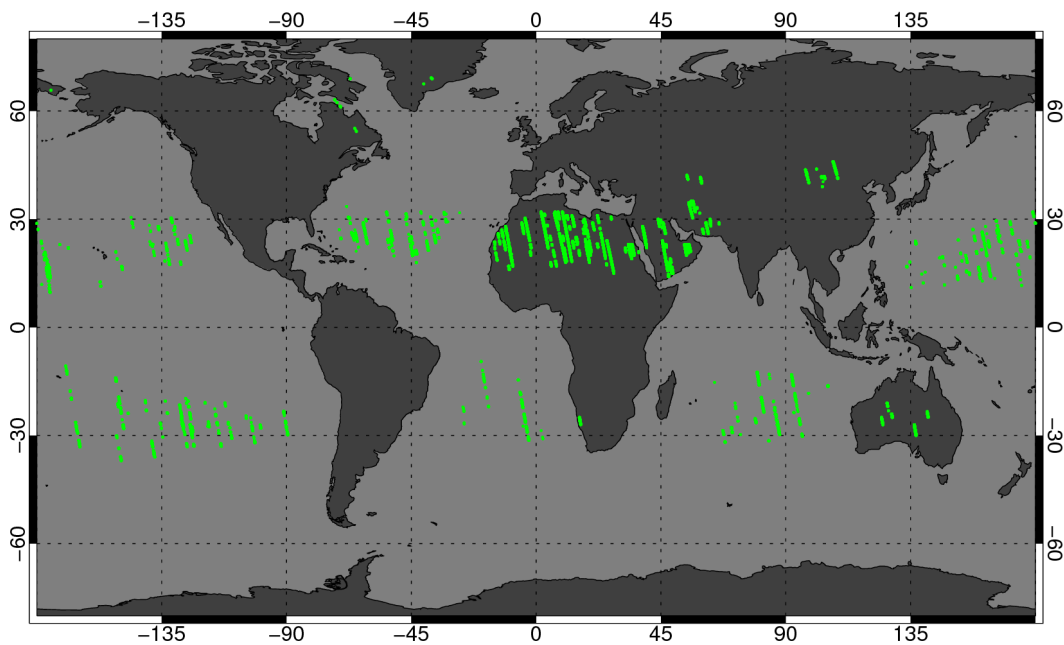


Figure 11: Sampling of all pre-filtered soundings analyzed in order to determine the ZLO correction. The data set consists of all pre-filtered OCO-2 soundings of 24 days in 2015 (13.01., 15.01., 14.02., 16.02., 10.03., 20.03., 03.04., 19.04., 08.05., 23.05., 08.06., 24.06., 15.07., 16.07., 15.08., 16.08., 15.09., 16.09., 15.10., 16.10., 15.11., 17.11., 12.12., 18.12.) additionally filtered for potential contamination with chlorophyll fluorescence (see main text).

SIF and both CO₂ fit windows hinting at ZLOs in the range of 0.8%-1.8% of the continuum radiance. In the following, we use the fitted linear relationship as ZLO correction for these three fit windows. In the O₂ fit window, the correlation between ZLO and continuum radiance is poor and the linear fit suggests a small negative slope. Therefore, we decided to not apply a ZLO correction for this fit window.

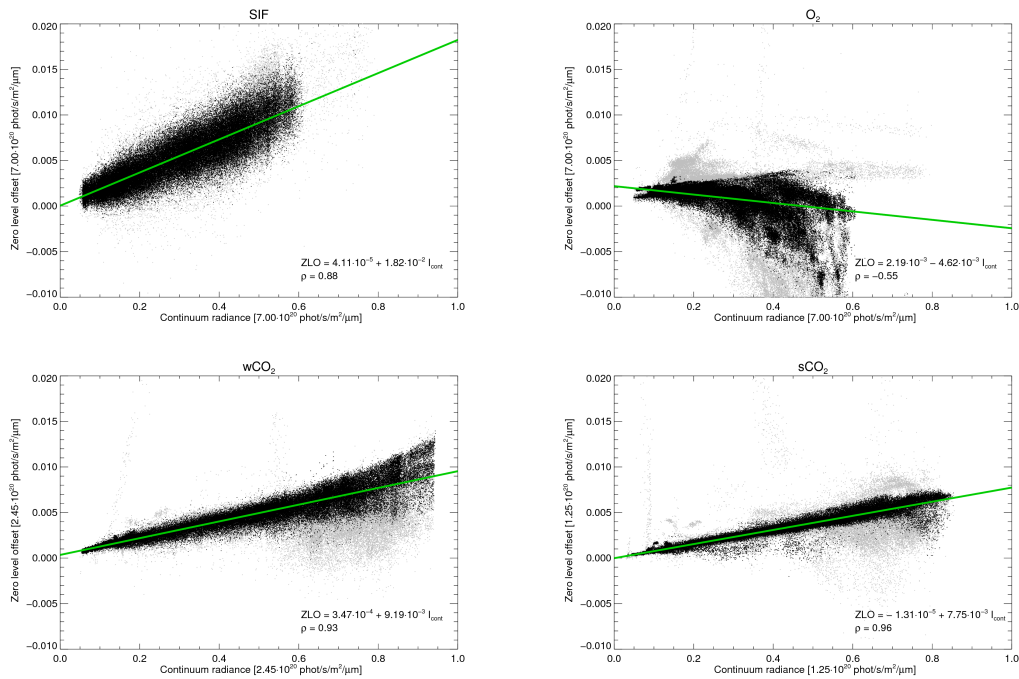


Figure 12: Retrieved zero level offset (ZLO) versus continuum radiance (I_{cont}) for all four fit windows. **gray dots:** potential outliers, (i.e., no convergence, $\chi^2 > 2$, or RSR exceeding threshold (see Fig. 10)). **green line:** linear fit.

6 Postprocessing

This section describes all postprocessing steps performed by FOCAL v06.

6.1 Filtering

First of all, we check for convergence, i.e., the state vector increment has to be small compared to the a posteriori uncertainty, the maximum number of iterations must not exceed 15, and χ^2 must not exceed 2 (for more details, see Reuter et al., 2017c). Convergence is achieved in about 74% of all pre-filtered OCO-2 soundings. Many non-converging soundings can be found near the SAA, the Saharan desert, and the Arabian peninsula (Fig. 13).

In the next step, we check for each fit window if the RSR is smaller than the threshold for potential outliers defined in Sec. 5.2. The throughput of this filter, which is most active above the tropical oceans (Fig. 13), is about 68% .

Additionally, we filter for potential outliers by parameters that have a unexpectedly large influence on the retrieved local XCO₂ variability. For the example data shown in Fig. 13, this filter is most active in high latitudes and has a throughput of about 84%.

This filter bases on the idea that XCO₂ outliers increase the local retrieved XCO₂ variability and are likely correlated with extreme values of some of the candidate parameters: XCO₂ uncertainty σ_{XCO_2} , lowermost layer of the CO₂ column averaging kernel, XH₂O, XH₂O uncertainty $\sigma_{\text{XH}_2\text{O}}$, XH₂O difference to the a priori, continuum radiance in the O₂ ($I_{\text{cont}}^{\text{O}_2}$), wCO₂ ($I_{\text{cont}}^{\text{wCO}_2}$), and sCO₂ ($I_{\text{cont}}^{\text{sCO}_2}$) fit window, gradient between first and second CO₂ layer ∇CO_2 , albedo difference between the O₂ and sCO₂ fit window, and all non CO₂ and H₂O state vector elements (see Reuter et al., 2017c).

For a representative two months data set (April and August 2015), we estimated the local retrieved XCO₂ variability $\text{VAR}(\Delta \text{XCO}_2)$ as follows: For each sounding, we computed the difference ΔXCO_2 between XCO₂ and its 5°×5°daily median and subsequently, we computed the variance of all ΔXCO_2 values falling in grid boxes with more than 100 samples. Now we searched for an upper or lower threshold for that candidate parameter which reduces $\text{VAR}(\Delta \text{XCO}_2)$ most when removing 1‰ of all data points. We repeated this until 15% of all data points were removed. In order to reduce the complexity of the postprocessing filter procedure, we now identified the 10 most promising parameters separately for land and ocean and repeated the whole exercise to find filter thresholds for these 10 parameters.

Fig. 14 shows (especially for land) that the decrease in variability somewhat

reduces after the removal of the first 5%-10%. A potential interpretation is that in this range indeed primarily outliers are removed. After the removal of approximately 15% the decrease in variability is relatively constant over a larger range before it drops to zero when the last data points are removed. As the curves do not show a distinct kink, the choice to remove 15% of all data points is a bit arbitrary but seemed to be a good compromise.

Above land (Fig. 14, left), the potential outliers filter reduces the variance of ΔXCO_2 from 2.05 ppm^2 to 1.68 ppm^2 . The Ångström exponent \AA is the dominant parameter, contributing 38% to the variance reduction. All parameter thresholds found for the potential outliers filter above land are listed in Tab. 3 (left).

Above sea (Fig. 14, right), this filter reduces the variance of ΔXCO_2 from 1.22 ppm^2 to 1.09 ppm^2 . In glint geometry, scattering is less important and the dominant parameter is the wavelength squeeze in the wCO_2 fit window $\lambda_{sq}^{wCO_2}$, contributing 32% to the variance reduction. All parameter thresholds found for the potential outliers filter above land are listed in Tab. 3 (right).

The combined throughput of all three post-filters (convergence, residual, and potential outliers) is about 42%.

6.2 Bias correction

The basic assumption of the bias correction scheme is that on average XCO_2 has little variations on small scales so that correlations to more variable parameters can be used to quantify biases. As a consequence, the bias correction does not require any ground truth data except for the quantification of a globally constant offset.

The final bias model consists of four components: a footprint bias B_f , a land/sea bias B_{ls} , a linear bias model B_{lin} , and a globally constant bias B_g :

$$B = B_f + B_{ls} + B_{lin} + B_g . \quad (38)$$

These four components are successively derived from analyses of the same two months data set (April and August 2015) used to determine the thresholds of the potential outliers filter (Sec. 6.1).

The swath of OCO-2 consists of eight neighboring footprints across track. In order to determine the mean footprint anomaly, we used only soundings belonging to complete sets of eight neighboring soundings which all passed the post-filtering and which were entirely over land or sea. Fig. 15 (right) shows the sampling of the roughly 180000 soundings where this is the case. For each of these sets of eight soundings, we compute the footprint anomaly and

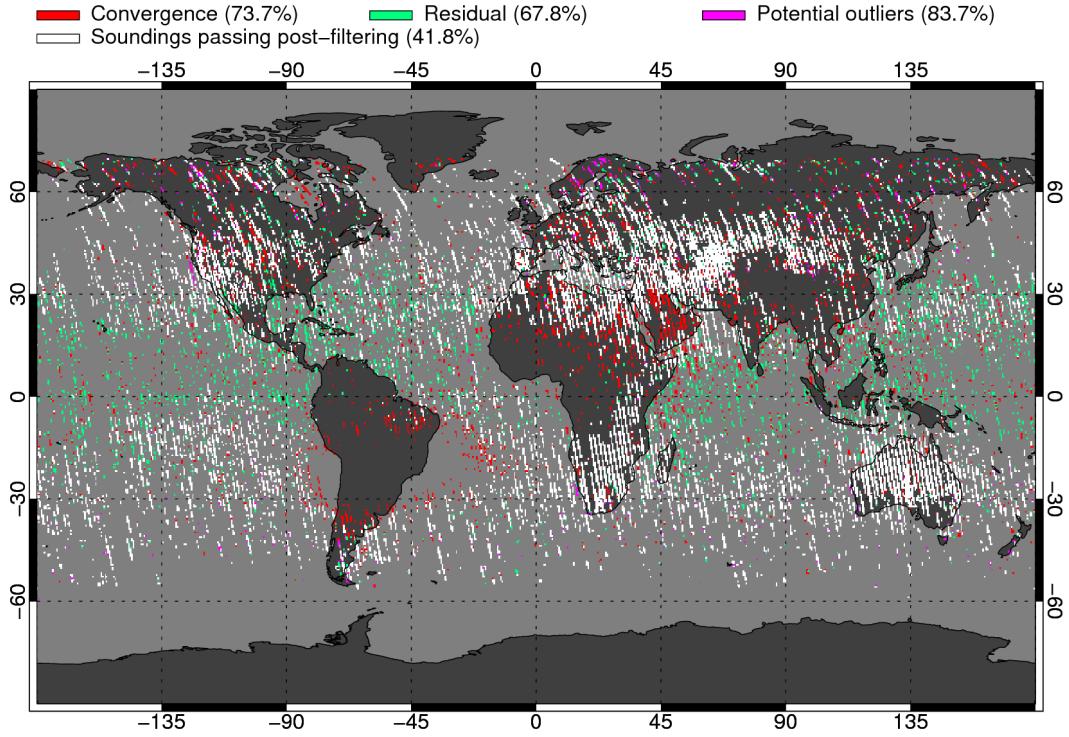


Figure 13: Post-filtering statistics for April and August 2015. The filters are applied in the order: convergence, residual, and potential outliers (see main text for a description). The colors represent filter activity and soundings passing all filters are shown in white. Numbers in brackets represent filter throughputs.

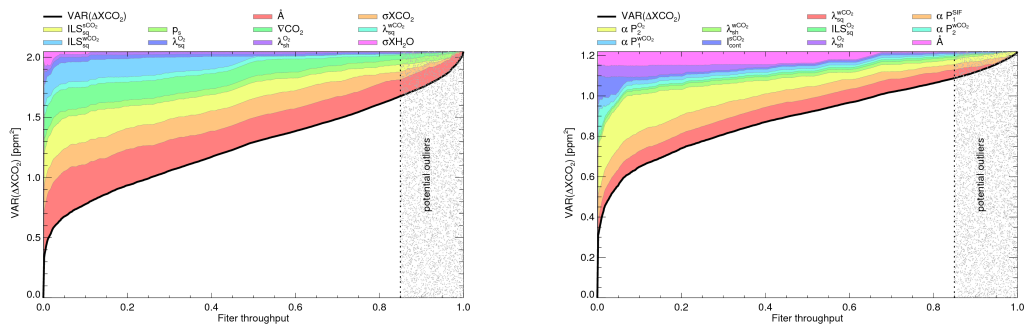


Figure 14: Variance versus filter throughput for the 10 most promising parameters identified for the potential outliers filter. The colors represent the prorated variance reduction of the individual parameters. See Reuter et al. (2017c) and the main text for a description of the individual parameters. **left:** Land. **right:** Sea.

Table 3: Thresholds and prorated variance reduction of the 10 parameters of the potential outliers filter for soundings above land (**top**) and sea (**bottom**). In total, the variance of ΔXCO_2 is reduced from 2.05 ppm² to 1.68 ppm² above land and reduced from 1.22 ppm² to 1.09 ppm² above sea. See Reuter et al. (2017c) and the main text for a description of the individual parameters.

	Parameter	Lower threshold	Upper threshold	Variance reduction [%]
Land	\AA	1.6669	-	38
	σXCO_2 [ppm]	-	1.2963	17
	$ILS_{sq}^{sCO_2}$	-	1.0022	16
	p_s [p ₀]	-	$2.2603 \cdot 10^{-1}$	11
		$1.6435 \cdot 10^{-1}$		
	∇CO_2 [ppm]	5.2509	5.9995	11
	$\lambda_{sq}^{wCO_2}$ [nm]	-	$3.9367 \cdot 10^{-5}$	2
		$5.2186 \cdot 10^{-4}$		
	$ILS_{sq}^{wCO_2}$	-	1.0041	1
	$\lambda_{sq}^{O_2}$ [nm]	-	-	2
			$2.5907 \cdot 10^{-5}$	
	$\lambda_{sh}^{O_2}$ [nm]	-	$9.2043 \cdot 10^{-4}$	1
Sea		$6.6146 \cdot 10^{-4}$		
	σXH_2O [ppm]	-	15.705	1
	$\lambda_{sq}^{wCO_2}$ [nm]	-	-	32
		$3.1372 \cdot 10^{-4}$	$8.2869 \cdot 10^{-5}$	
	αP_1^{SIF}	-	$3.4846 \cdot 10^{-3}$	20
		$2.3184 \cdot 10^{-3}$		
	$\alpha P_2^{O_2}$	-	$1.7900 \cdot 10^{-4}$	19
	$\lambda_{sh}^{wCO_2}$ [nm]	-	$2.1023 \cdot 10^{-3}$	9
	$ILS_{sq}^{O_2}$	-	1.0175	4
	$\alpha P_2^{wCO_2}$	-	-	5
			$2.1247 \cdot 10^{-4}$	
	$\alpha P_1^{wCO_2}$	$3.2736 \cdot 10^{-4}$	-	4
	$I_{cont}^{sCO_2}$ [$1.25 \cdot 10^{20}$]	$5.6468 \cdot 10^{-2}$	-	4
	Ph/s/m ² /μm]			
	$\lambda_{sh}^{O_2}$ [nm]	-	-	2
		$3.4860 \cdot 10^{-4}$		
	\AA	1.9014	-	2

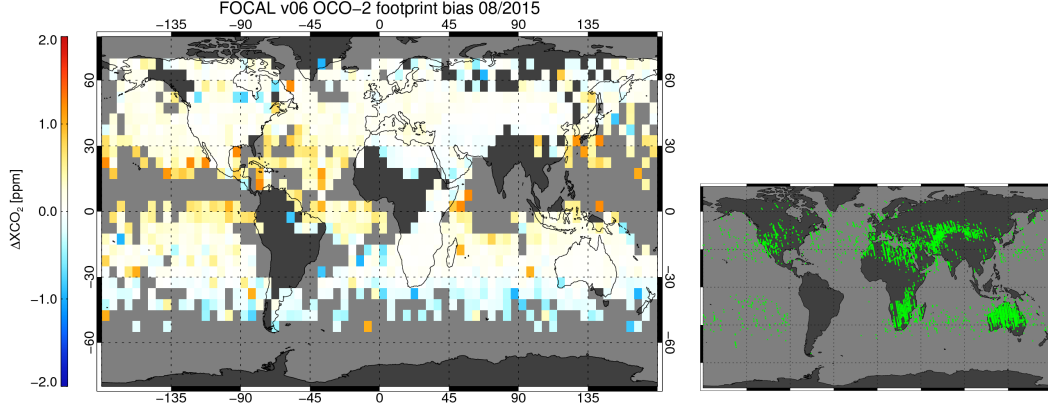


Figure 15: FOCAL v06 OCO-2 footprint bias pattern (Eq. 39) at the example of August 2015 (**left**) and sampling of soundings used to determine the footprint bias (**right**).

subsequently the average footprint anomaly of all sets which we then use as bias function. Fig. 15 (left) shows the footprint bias pattern at the example of all soundings passing the post-filtering in August 2015 and the corresponding bias function depending on the footprint f is:

$$B_f(f) = \begin{cases} -0.974 \text{ ppm} & \text{if } f = 1, \\ -0.336 \text{ ppm} & \text{if } f = 2. \\ -0.234 \text{ ppm} & \text{if } f = 3. \\ -0.315 \text{ ppm} & \text{if } f = 4. \\ -0.856 \text{ ppm} & \text{if } f = 5. \\ 1.013 \text{ ppm} & \text{if } f = 6. \\ 0.484 \text{ ppm} & \text{if } f = 7. \\ 1.219 \text{ ppm} & \text{if } f = 8. \end{cases} \quad (39)$$

In order to determine the land/sea bias, we corrected all post-filtered results for the footprint bias and analyzed all coastline overpasses with a maximum duration of 120 s (≈ 800 km along track), at least 100 soundings, and a land fraction between 30% and 70%. For each of these coastline overpass (162 with about 102000 soundings, Fig. 16, right) we computed the land/sea anomaly and hence the average land/sea anomaly (± 0.8986 ppm). Fig. 16 (left) shows the land/sea bias pattern corresponding to the bias function

$$B_{ls}(l) = 0.8986 \text{ ppm } (2l - 1) \quad (40)$$

with l being the land/sea fraction.

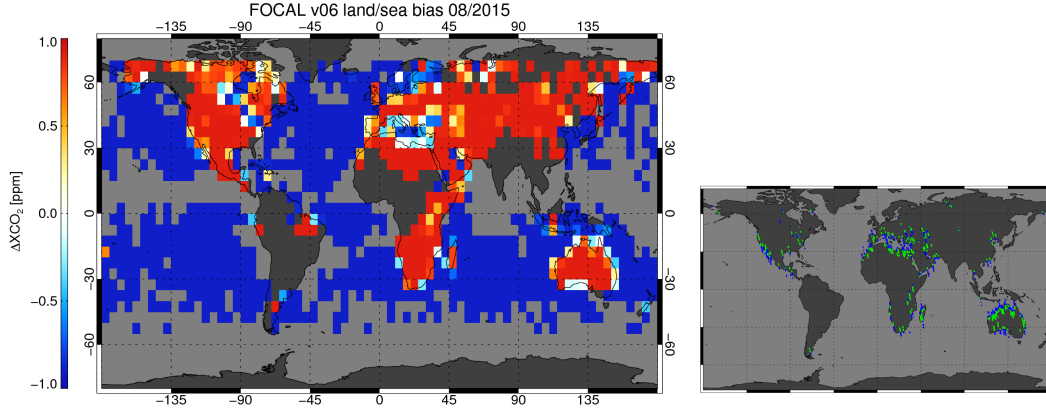


Figure 16: FOCAL v06 land/sea bias pattern (Eq. 40) at the example of August 2015 (**left**) and sampling of soundings used to determine the land/sea bias (**right**).

Additionally to the footprint and land/sea biases, we found a small correlation ($\rho = 0.16$) between small scale anomalies of XCO_2 and the retrieved ILS squeeze in the wCO_2 fit window $ILS_{sq}^{wCO_2}$. The small scale anomalies have been computed from 60 s chunks (≈ 400 km) of post-filtered and footprint and land/sea bias corrected OCO-2 orbit data with more than 10 soundings. In total about 5000 chunks with one million soundings (Fig. 17, right) have been analyzed to compute the linear bias model (Fig. 17, left):

$$B_{lin}(ILS_{sq}^{wCO_2}) = 107.936 \text{ ppm} \cdot ILS_{sq}^{wCO_2} - 107.862 \text{ ppm} . \quad (41)$$

Finally, we correct for a global offset in respect to the optimized CO_2 concentration fields of the CAMS (Copernicus Atmosphere Monitoring Service) greenhouse gas flux inversion model (Chevallier et al., 2010) obtained from <http://apps.ecmwf.int>. CAMS is the CO_2 atmospheric inversion product of the European Union programme Copernicus that develops information services based on satellite earth observation and other data (<http://www.copernicus.eu/>). The product is released twice per year and each time covers the full period from 1979 until the year before the release. It results from an analysis of CO_2 surface air sample measurements over the corresponding period and consists of optimized CO_2 surface fluxes over the globe and of associated 3D CO_2 concentrations. Version 15r4 used here analyzed 37 years of surface measurements (1979-2015). Its spatial resolution is of 3.75° in longitude and 1.875° in latitude, with 39 hybrid layers in the vertical. A full description of v15r4 is given by Chevallier (2017), that shows, among other validation results, that its root-mean-square fit to TCCON measurements is usually close to 1 ppm. The global offset of FOCAL v06 relative to the CAMS model amounts:

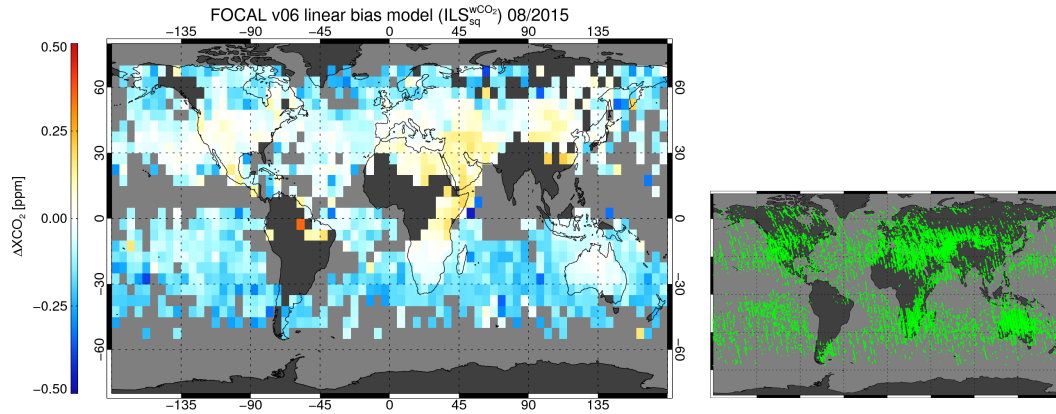


Figure 17: FOCAL v06 bias pattern of the linear bias model (Eq. 41) at the example of August 2015 (**left**) and sampling of soundings used to determine the linear bias model (**right**).

$$B_g = -1.673 \text{ ppm} . \quad (42)$$

Fig. 18 (left) shows an example of the total bias pattern consisting of all four components (footprint bias, land/sea bias, linear bias model, global bias). As can be seen, the large scale pattern of the total bias is dominated by the land/sea bias followed by the footprint bias and the linear bias model plays only a minor role. For comparison, Fig. 18 (right) shows the total bias pattern of NASA's operational OCO-2 L2 product v7.3.05b (Boesch et al., 2015; Eldering et al., 2017) obtained from <https://daac.gsfc.nasa.gov> and in the following referred to as NASA v7.3.05b. The overall variability is similar (0.82 ppm and 0.71 ppm for FOCAL v06 and NASA v7.3.05b, respectively) and the NASA product also has a distinct land/sea bias but with opposite sign, i.e., with largest values over sea (note the reversed color bar in Fig. 18, right).

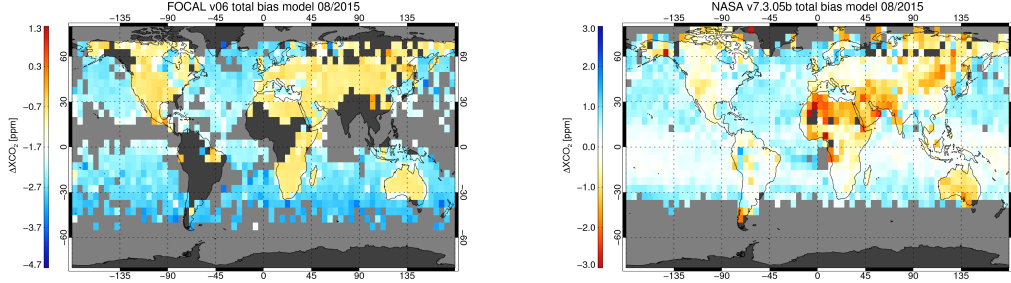


Figure 18: FOCAL v06 total bias pattern (Eq. 38) at the example of August 2015 (**left**) and NASA OCO-2 v7.3.05b total bias pattern (**right**).

7 Error Characterization

In order to assess FOCAL's theoretical capabilities (primarily in retrieving XCO_2 , XH_2O , and SIF), we confront it with radiance measurements simulated with the accurate RT code SCIATRAN (Rozanov et al., 2014). The performed analyses can be understood also as test of the suitability of the approximations made in FOCAL's RT and of the retrieval setup. Hereby, we primarily concentrate on scattering related errors and analyze the systematic and stochastic, i.e., the a posteriori errors of several different retrieval setups and geophysical scenarios.

We are not aiming to comprehensively cover the majority of potential geophysical scenarios, because the final quality depends on the full retrieval scheme including, e.g., potential instrument and forward model errors and different post-filtering capabilities. The aim of the inversion experiments shown in this section is rather to show that the FOCAL v01 setup is a promising candidate to serve as starting point for the development of a full retrieval scheme and its application to actually measured OCO-2 data. This is presented in Sec. 8 quantifying the final quality of the retrieval by comparing retrievals of actually measured data with independent ground truth measurements.

7.1 Retrieval setups

The baseline retrieval setup (FOCAL v01) is described in Sec. 4. As this setup accounts for scattering with three scattering related state vector elements (pressure, i.e., height of the scattering layer p_s , scattering optical thickness τ_s , and Ångström exponent \tilde{A}), it is referred to as *3-Scat* setup in the following. All other tested retrieval setups are descendants of this setup. The *4-Scat* setup has an extended state vector, additionally fitting the fraction of radiation scattered into the hemisphere in forward direction (f in Eq. 15). The *0-Scat*

setup equals an absorption only retrieval; this means, the state vector does not include any scattering related parameters and the fit is limited to the CO₂ fit windows. The *3-Scat-O₂* setup equals the baseline setup except for scattering parameter derivatives which have artificially been set to zero in the CO₂ bands in order to ensure that the scattering information solely comes from the O₂ band. Accordingly, the *3-Scat-CO₂* setup ensures that the scattering information solely comes from the CO₂ bands. The scenarios *3-Scat-synth* and *0-Scat-synth* use a synthetic a priori error covariance matrix for the CO₂ profile as proposed by Reuter et al. (2012) but with a correlation length of $1.0 p_0$ instead of $0.3 p_0$. The scenarios *3-Scat-stiff* and *0-Scat-stiff* use a similar synthetic a priori error correlation matrix but computed with a correlation length of $100 p_0$. This “stiffens” the a priori error covariance matrix so that the departure from the a priori profile becomes basically proportional to the uncertainty profile. For these scenarios, the a priori error covariance matrix of the H₂O profile has been stiffened in the same way.

7.2 Scenarios

The geophysical *baseline* scenario has a spectrally flat albedo of 0.2, 0.2, 0.1, and 0.05 in the SIF, O₂, wCO₂, and sCO₂ fit window (values which have also been used by, e.g., Bovensmann et al., 2010). It does not include chlorophyll fluorescence, scattering by aerosols, clouds, or Rayleigh. Its temperature, pressure, and water vapor ($X_{H_2O} = 3031 \text{ ppm} \cong 19.52 \text{ kg/m}^2$) profiles are taken from an ECMWF analysis of August 28, 2015, 12:00 UTC, 9°E, 53°N. Its CO₂ profile is calculated with SECM2016 and corresponds to an XCO₂ value of about 395 ppm. Note that ECMWF and SECM2016 are also used to compute the first guess and a priori H₂O and CO₂ profiles (Tab. 2). All other scenarios are descendants of the *baseline* scenario.

Each scenario is analyzed for three solar zenith angles (20°, 40°, and 60°) and for two directions of polarization (parallel and perpendicular to the SPP). The satellite zenith angle is set to 0°(nadir).

The *SIF* scenario adds $1 \text{ mW/m}^2/\text{sr/nm}$ chlorophyll fluorescence at 760 nm to the simulated measurement of the *baseline* scenario. The *XCO₂+6 ppm* scenario has an increased CO₂ concentration of 15 ppm, 10 ppm, and 5 ppm in the three lowermost layers, so that the column-average concentration is enhanced by 6 ppm.

All scattering related scenarios are more complex for the retrieval because of FOCAL’s scattering approximations. The *Rayleigh* scenario adds Rayleigh scattering to the baseline scenario; the Rayleigh optical thickness at 760 nm

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	50
---------------------	---	---	-----------

for this scenario is about 0.026. *Rayleigh+Aerosol BG* additionally includes a (primarily) stratospheric background aerosol with an AOT (aerosol optical thickness at 760 nm) of 0.019 (0.003 at 1600 nm and 0.001 at 2050 nm). *Rayleigh+Aerosol cont* adds a continental aerosol to the boundary layer so that the total AOT becomes 0.158 (0.060 at 1600 nm and 0.037 at 2050 nm). *Rayleigh+Aerosol urban* adds a strong contamination with urban aerosol to the boundary layer and the total AOT becomes 0.702 (0.245 at 1600 nm and 0.151 at 2050 nm).

The scenarios *Rayleigh+Dark surface*, *Rayleigh+Bright surface*, and *Rayleigh+Ocean glint* distinguish from the *Rayleigh* scenario only by their surface reflection properties. *Rayleigh+Dark surface* and *Rayleigh+Bright surface* correspond to the *Rayleigh* scenario but with an albedo multiplied with 0.7 and 1.4, respectively. The *Rayleigh+Ocean glint* scenario deviates from the assumption of a Lambertian surface bidirectional reflectance distribution function (BRDF); it includes an ocean surface at a wind speed of 5 m/s, 37° to the solar principal plane (SPP). Additionally, the satellite zenith angle of this scenario is set to 0.75 times the solar zenith angle so that the satellite looks near the glint spot of specular reflectance.

Two cloud scenarios (*Rayleigh+Aerosol BG+Water cloud* and *Rayleigh+Aerosol BG+Ice cloud*) add a sub-visible water or ice cloud to the *Rayleigh+Aerosol BG* scenario. The water cloud has a height of 3 km, droplets with an effective radius of 12 µm, and a COT (cloud optical thickness at 500 nm) of 0.039. The ice cloud is made of fractal particles with an effective radius of 50 µm, has a height of 8 km, and a COT of 0.033.

7.3 Results

Primarily, we are interested in XCO₂ retrieval results of high quality; the correct retrieval of other state vector elements is less important as long as the XCO₂ quality is not affected. Fig. 19 summarizes the systematic errors and stochastic uncertainties of the retrieved XCO₂ for all retrieval setups and geophysical scenarios.

The *baseline* scenario is mainly to ensure consistency of the RT used to simulate the measurements (SCIATRAN) and the RT of the retrieval (FOCAL). Additionally, the *baseline* scenario allows estimates of the retrieval's noise error. With SCIATRAN, it is not simply possible to simulate FOCAL's scattering approximations, that is why this scenario excludes scattering. The systematic errors of the *baseline* scenario are always very small (0.03 ppm at maximum), which confirms the RT consistency in the absorption only case and ensures that,

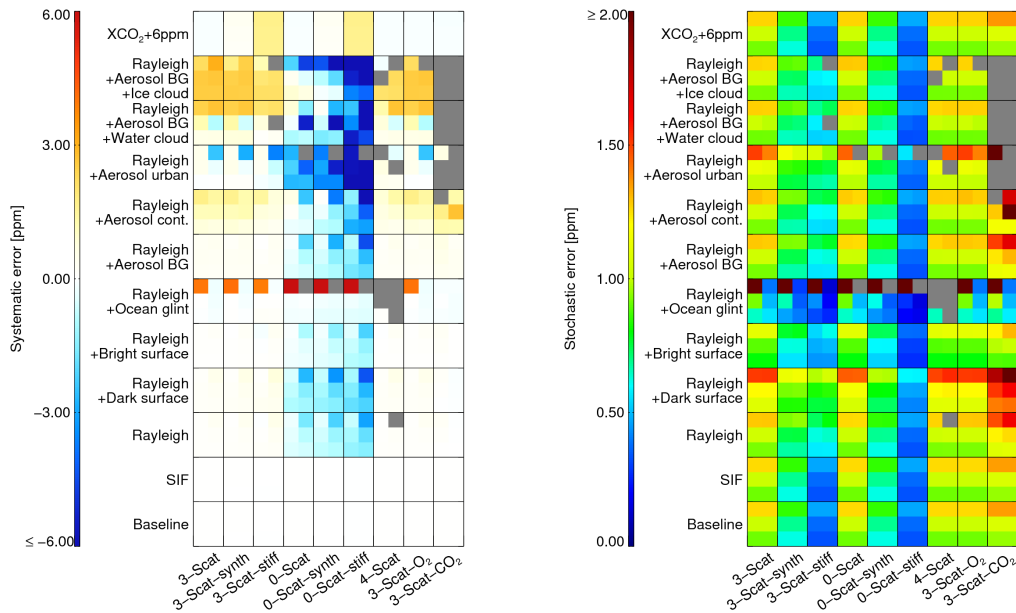


Figure 19: Error characteristics of nine retrieval setups and twelve geophysical scenarios. Each box includes six sub-boxes representing polarization parallel (left) and perpendicular (right) to the SPP as well as three solar zenith angles (20°, 40°, and 60°, from bottom to top). Gray boxes represent not converging retrievals. **Left:** Systematic error (retrieved minus true XCO₂). **Right:** Stochastic uncertainty as reported by the optimal estimation retrieval.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	52
---------------------	---	---	-----------

e.g., the number of particles is basically identical in the SCIATRAN and the FOCAL “world”.

The systematic errors of the *SIF* scenario are not larger than for the *baseline* scenario, because i) SIF is solely determined from the SIF fit window and ii) there is no SIF flux emitted in the CO₂ fit windows.

A more complex case for FOCAL is the *Rayleigh* scenario, because Rayleigh scattering takes place in the entire atmospheric column with a peanut-shaped SPF. This means, it cannot be expected that FOCAL is able to perfectly fit the simulated measurement. Fig. 20 (top) shows a spectral fit in all fit windows but with a state vector not including any scattering parameter, so that the geophysical results (e.g., XCO₂) become identical with those of the *0-Scat* setup.

Not surprisingly, the residual in the O₂ fit window becomes large compared to the simulated measurement noise ($\chi_{O_2} = 6.825$). The residuals in the CO₂ fit windows are already small compared to the instrumental noise even without fitting scattering parameters ($\chi_{wCO_2} = 0.026$, $\chi_{sCO_2} = 0.049$). This is only partly explained by Rayleigh scattering having an Ångström exponent of four and, therefore, a much smaller scattering optical thickness at longer wavelengths. It also indicates that disentangling scattering parameters and CO₂ concentration from measurements in the CO₂ fit windows may be difficult. In other words, most of the scattering information must be imprinted in the residual of the O₂ fit window. This is also why the results of the *3-Scat-O₂* are similar to the *3-Scat* setup and why the *3-Scat-CO₂* retrievals are often not converging (Fig. 19).

Allowing the *3-Scat* retrieval setup to fit the scattering parameters p_s , τ_s , and \tilde{A} , reduces the O₂ residual to become typically four times smaller than expected from instrumental noise ($\chi_{O_2} = 0.250$, Fig. 20, middle). Simultaneously, the XCO₂ error reduces from -0.43 ppm to 0.10 ppm (-0.89 ppm and 0.16 ppm for perpendicular polarization).

All other scattering related scenarios are even more “complicated” for FOCAL because different particles contribute to scattering. For example, cloud particles have different properties like height or Ångström exponent as aerosol particles, but FOCAL can only retrieve one effective height and one effective Ångström exponent. Additionally, the SPFs of aerosols and clouds are less isotropic. Therefore, the residuals (Fig. 20, bottom) and more importantly, the systematic errors typically increase for these scenarios (Fig. 19, left).

Fig. 21 shows the retrieved scattering parameters for the *3-Scat* setup and a set of scattering related plus the baseline scenario. As the baseline scenario does not include any scattering, the retrieved p_s and \tilde{A} are close to their

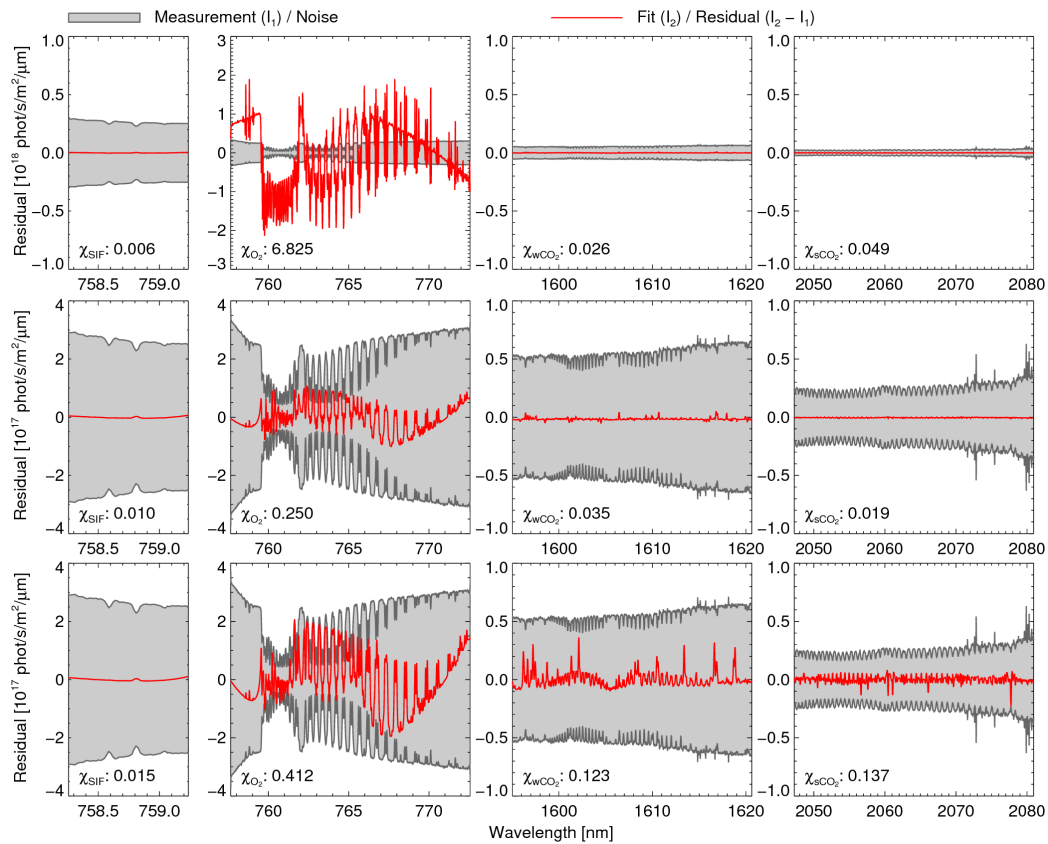


Figure 20: As Fig.3 (bottom) but for the *Rayleigh* scenario and the *0-Scat* setup (**top**), the *Rayleigh* scenario and the *3-Scat* setup (**middle**), and the *Rayleigh+Aerosol BG+Water cloud* scenario and the *3-Scat* setup (**bottom**).

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	54
---------------------	---	---	-----------

a priori values and have a large a posteriori uncertainty. Consistent with the expectations, the retrieved effective Ångström exponent is close to four (about 3.8) for the Rayleigh scenario and reduces to 2.8–3.6 for the aerosol and 2.1–2.6 for cloud scenarios. This means the scattering optical thickness at longer wavelengths increases relative to the shorter wavelengths. Rayleigh scattered light is unpolarized in forward and backward scattering direction but polarized perpendicular to the incident beam for scattering angles of 90°. For this reason, the retrieved τ_s is always larger for the polarization direction perpendicular to the SPP. As expected, this effect is more/less pronounced for larger/smaller solar zenith angles (not shown). In contrast to the *3-Scat* and *4-Scat* setups, the *0-Scat* retrievals cannot fit τ_s which results in a larger polarization dependency of the resulting systematic errors (Fig. 19, left).

As shown in Fig. 21, the highest scattering optical thicknesses at 760 nm are obtained for the urban aerosol and the cloud scenarios. However, the quantitative interpretation of the retrieved values of τ_s and p_s is difficult because they are effective values representing all kinds of scattering in the atmospheric column. Additionally, τ_s and p_s may not be perfectly independent because light path modifications are expected to become larger when enhancing the height of the scattering layer. It can be observed that the retrieved values of τ_s are generally smaller than the scattering optical thicknesses computed by SCIATRAN (Sec. 7.2). This is expected because of the different SPFs assumed by SCIATRAN and FOCAL. Especially for Mie scattering of cloud and aerosol particles, the SCIATRAN simulations use SPFs with a distinct forward peak contributing to the total scattering optical thickness. FOCAL, however, interprets scattering in forward direction as transmission (not contributing to τ_s). This means, τ_s is best comparable for the *Rayleigh* scenario with a SPF without forward peak.

The scenarios *Rayleigh*, *Rayleigh+Dark surface*, and *Rayleigh+Bright surface* differ by their surface albedo. However, the retrieved scattering parameters show little differences because in FOCAL these parameters represent (within the limits of the made assumptions) approximations of real geophysical quantities.

Applying FOCAL to the *Rayleigh+Ocean glint* scenario with a highly non-Lambertian surface BRDF results in systematic XCO₂ and XH₂O errors usually comparable to the *Rayleigh* scenario (Fig. 19 and 26, left) except for solar zenith angles of 60° and polarization parallel to the SPP. In near-glint geometry, specular reflectance dominates the radiation field but with increasing solar zenith angle the reflected radiation becomes more and more polarized. As a result the direct photon path often dominates (if not observing parallel polarization at large solar zenith angles) and an imperfect parameterization of scattering becomes less

important. The domination of the direct photon path also results in a larger total radiance and, correspondingly, smaller stochastic errors in perpendicular polarization (Fig. 19 and 26, right). The larger systematic XCO_2 errors of about 4 ppm at 60° and parallel polarization are a result of the poor surface reflectivity in this observation geometry and associated with large stochastic errors of about 8 ppm and little error reduction (analog for XH_2O). This means, applied to real measurements, such retrievals would most certainly be filtered during post processing. Note that due to the non-Lambertian surface, the retrieved albedo may have values larger than one.

Fig. 19 (right) shows that the shape of the CO_2 a priori error covariance matrix can considerably influence the stochastic XCO_2 a posteriori uncertainty, even though the a priori XCO_2 uncertainty has not been changed. Stiffening the covariance matrix by enhancing the layer-to-layer correlations as done for the *synth* and *stiff* setups (Fig. 22 and Fig. 23), reduces the stochastic XCO_2 uncertainty from typically about 1 ppm to 0.7 ppm (*synth*) and 0.4 ppm to 0.6 ppm (*stiff*), which does not necessarily mean that results actually improve.

Except for the $XCO_2 + 6 \text{ ppm}$ scenario, the systematic errors of the *3-Scat*, *3-Scat-synth*, and *3-Scat-stiff* setups are very similar. This is not the case for the *0-Scat*, *0-Scat-synth*, and *0-Scat-stiff* setups for which the systematic errors increase with stiffness of the CO_2 a priori error covariance matrix. Apparently, the (loose) profile retrieval of the *0-Scat* scenario happens to somewhat compensate light path related errors. In the case of the *3-Scat* setups, the scattering parameters are doing this job. Fig. 24 shows, that the largest deviations of the retrieved profiles from the true profile (a priori) indeed occur for the *0-Scat* setup.

The degree of freedom for the CO_2 profile is about 2.2 for the *3-Scat* setup and reduces to 1.8 for the *3-Scat-synth* and 1.0 for the *3-Scat-stiff* setup. The degree of freedom for the H_2O profile reduces from 2.2 for the *3-Scat* setup to 1.0 for the *3-Scat-stiff* setup. Additionally, the column averaging kernels (AKs) change and show larger deviations from unity; specifically, as illustrated in Fig. 25, the XCO_2 AK increases to about 1.2 in the boundary layer and reduces to 0.6 in the stratosphere. As a result, the systematic error (in this particular case, the smoothing error) increases for the *stiff* setups to about 1.6 ppm (Fig. 19, left).

As illustrated in Fig. 26, scattering related systematic XH_2O errors are usually negative and larger for the *0-Scat* setups. Stiffening the H_2O a priori error covariance matrix has little influence on the systematic or stochastic error which is usually about 10 ppm. SIF is almost not influenced by the mostly low scattering optical thicknesses of the tested scenarios and the stochastic

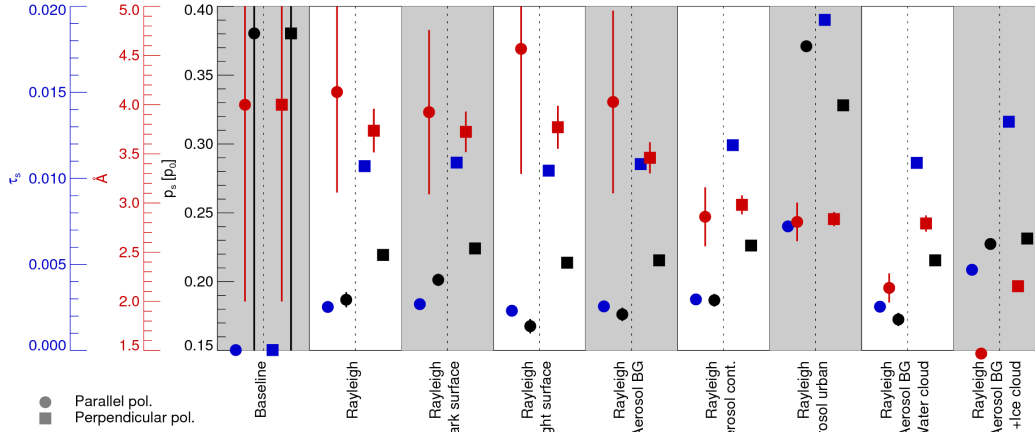


Figure 21: Retrieved scattering optical thickness, Ångström exponent, and height of the scattering layer for the 3-Scat setup and the scenarios ($\theta_0 = 40^\circ$): *baseline*, *Rayleigh*, *Rayleigh+Dark surface*, *Rayleigh+Bright surface*, *Rayleigh+Aerosol BG*, *Rayleigh+Aerosol cont.*, *Rayleigh+Aerosol urban*, *Rayleigh+Aerosol BG+Water cloud*, and *Rayleigh+Aerosol BG+Ice cloud*.

a posteriori error is usually between 0.2 and 0.3 mW/m²/sr/nm (Fig. 27).

All tested retrieval setups do not have the ability to change the number of dry-air particles in the atmospheric column, e.g., by fitting the surface pressure, or a shift of the temperature profile. As a result, relative errors of the number of dry-air particles computed from the meteorological profiles directly translate into relative errors of the retrieved XCO₂ and XH₂O. For example, a 1 hPa error of the surface pressure will result in a XCO₂ error of about 0.4 ppm.

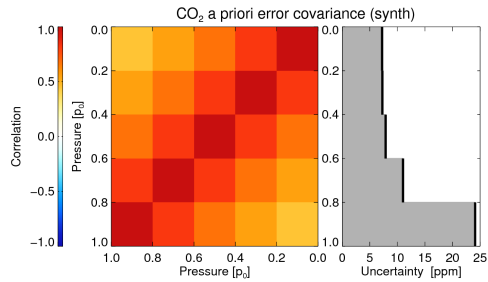


Figure 22: As Fig. 4 but for a synthetic a priori error covariance matrix as proposed by Reuter et al. (2012) with a correlation length of $1.0 p_0$.

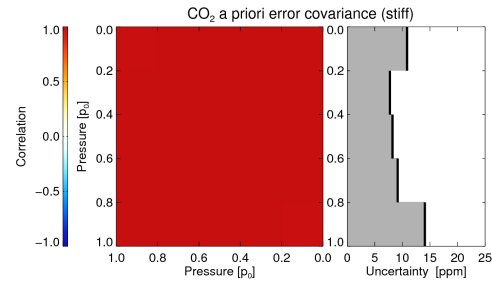


Figure 23: Same as Fig. 4 but for a synthetic a priori error correlation matrix as proposed by Reuter et al. (2012) with a correlation length of $100 p_0$.

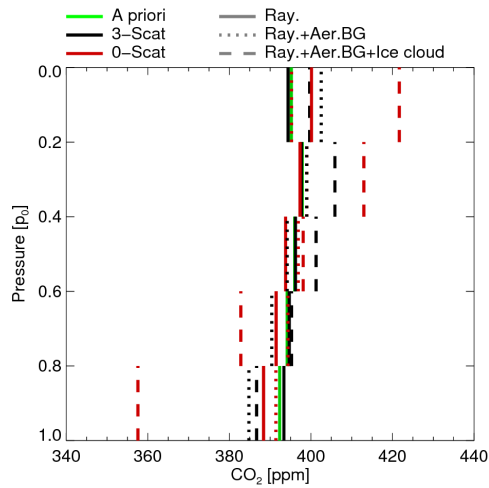


Figure 24: Retrieved and a priori CO_2 profiles for the 0-Scat and 3-Scat retrieval setups and the geophysical scenarios *Rayleigh*, *Rayleigh+Aerosol BG*, and *Rayleigh+Aerosol BG+Ice cloud* ($\theta_0 = 40^\circ$, perpendicular polarization).

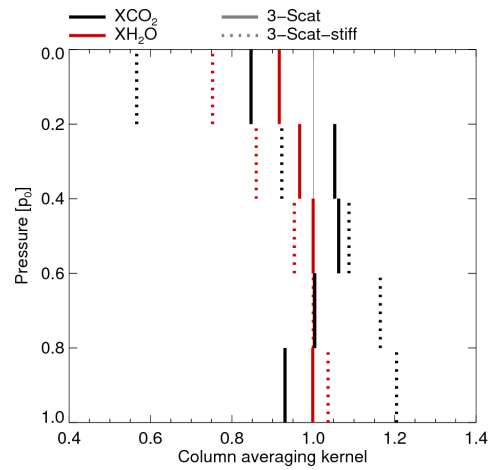


Figure 25: XCO_2 and XH_2O column averaging kernels for the 3-Scat and 3-Scat-stiff retrieval setups and the geophysical *Rayleigh* scenario ($\theta_0 = 40^\circ$, perpendicular polarization).

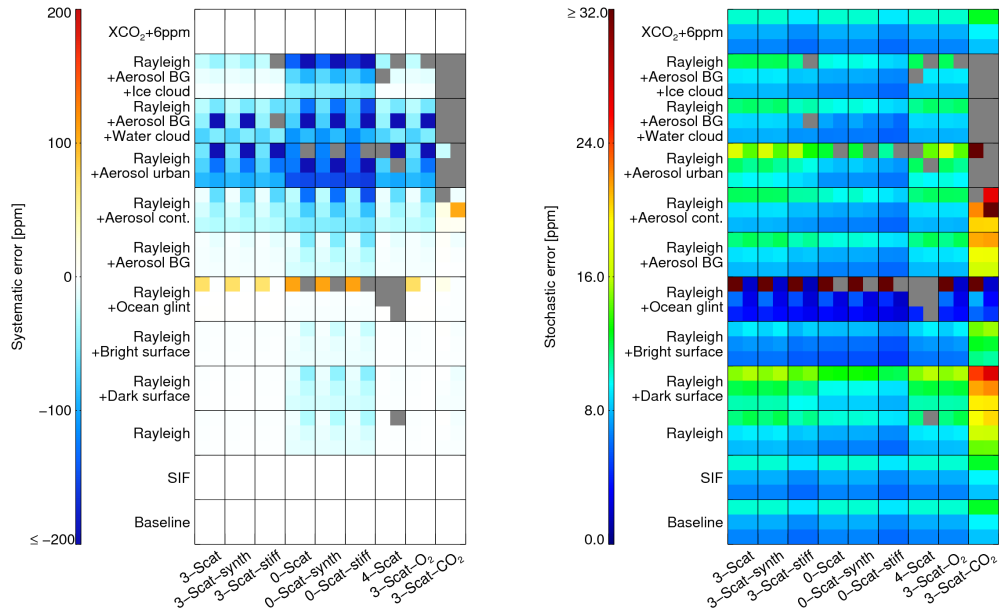


Figure 26: As Fig. 19 but for XH_2O .

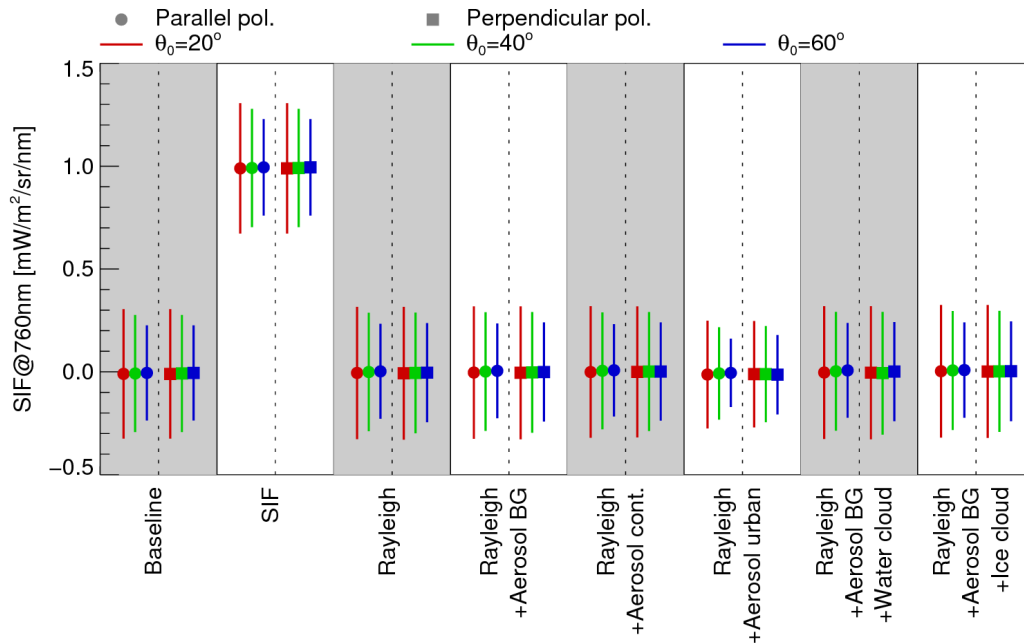


Figure 27: Retrieved solar induced chlorophyll fluorescence for the 3-Scat retrieval setup and the geophysical *baseline*, *SIF*, and all scattering related scenarios. The error bars represent the 1σ a posteriori uncertainty.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	59
---------------------	---	---	-----------

8 Validation

8.1 Model comparison

In this section we compare two months (April and August 2015) of post-filtered and bias corrected FOCAL v06 XCO₂ results with corresponding values of the CAMS v15r4 model accounting for FOCAL's column averaging kernels (e.g., Rodgers, 2000). Fig. 28 shows 5°×5° monthly gridded values for both months, FOCAL, and CAMS. The main spatial and temporal patterns are similar for FOCAL and CAMS with largest and smallest values in the northern hemisphere in April and August, respectively. Differences become larger at smaller scales, e.g., FOCAL sees larger values in natural and anthropogenic source regions of Sub-Saharan Africa and East Asia in April but also above the Sahara in August. However, it shall be noted that often only few data points are in the corresponding grid boxes.

In grid boxes with more than 100 soundings, the standard error of the mean becomes negligible (≈ 0.1 ppm). Therefore, the difference between FOCAL and CAMS in such grid boxes can be interpreted as systematic temporal and regional mismatch or bias. The standard deviation of this systematic mismatch (including also representation errors) amounts to 1.0 ppm. The standard deviation of the single sounding mismatch after subtracting the systematic mismatch amounts to 1.1 ppm which agrees reasonably well with the average reported uncertainty of 1.2 ppm.

8.2 Comparison with NASA's operational OCO-2 L2 product

In this section we compare the same two months (April and August 2015) of post-filtered and bias corrected FOCAL v06 XCO₂ results with NASA's operational OCO-2 L2 product. Comparing Fig. 29 with Fig. 28 (top) shows similar large scale temporal and spatial patterns and also the relative enhancement in the anthropogenic source regions of East Asia in April are similar. The most obvious difference is that the NASA product has about three times more soundings. The primary reason for this is the inherently poor throughput (11%) of the MODIS based cloud screening of the preprocessor (see discussion in Sec. 5).

Analyzing only the same soundings in both data sets and considering the column averaging kernels, the NASA product has on average 0.7 ppm larger values than FOCAL which is (due to the used color table) most noticeable in the northern hemisphere of Fig. 29 (right). The standard deviation of the difference is 1.1 ppm. As done in the last section, we separate the systematic mismatch from the stochastic mismatch by analyzing grid boxes with more than

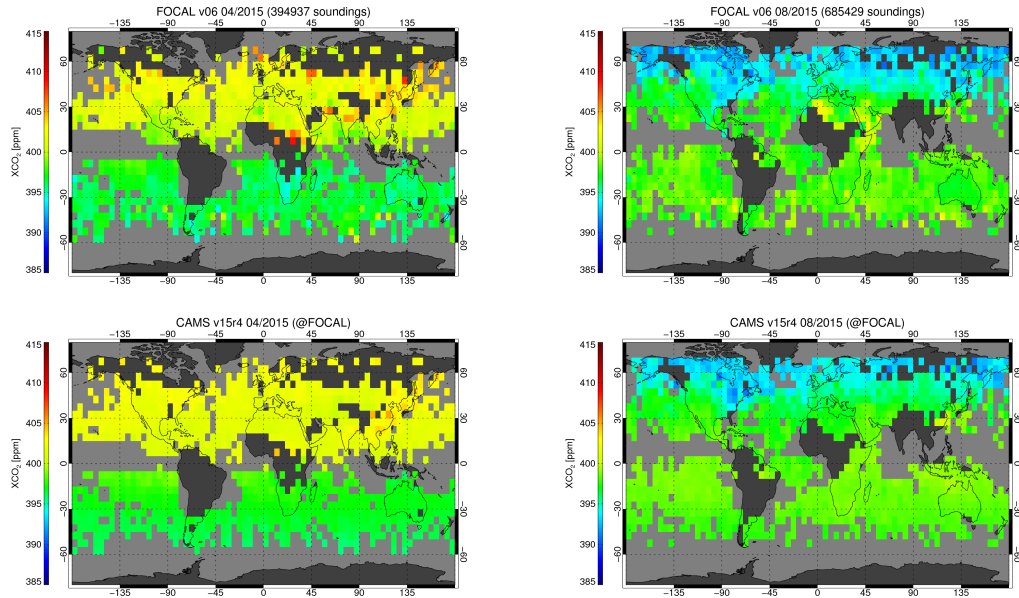


Figure 28: Monthly mean XCO₂ at 5°x5°. **Top:** FOCAL v06. **Bottom:** CAMS v15r4 sampled as FOCAL. **Left:** April 2015. **Right:** August 2015.

100 co-locations. The standard deviation of the stochastic and the systematic mismatch amounts 0.91 ppm and 0.83 ppm, respectively. It is no surprise, that the stochastic mismatch is smaller than expected from the combined reported uncertainties because both data products base on the same L1b input data including the same noise spectra.

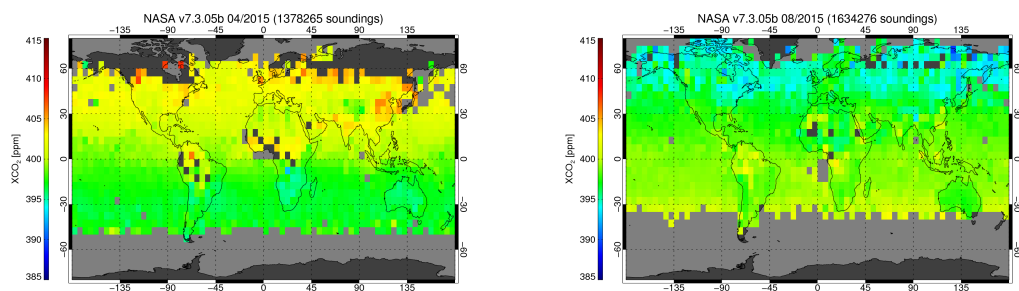


Figure 29: NASA v7.3.05b monthly mean XCO₂ at 5°x5°. **Left:** April 2015. **Right:** August 2015.

8.3 Validation with TCCON

In this section we show validation results for one year (2015) of FOCAL v06 and NASA's operational OCO-2 L2 data and analyze the influence of the bias corrections. We used ground based TCCON (Wunch et al., 2011) GGG2014 data obtained from <http://tcccon.ornl.gov> as reference data set and a similar validation protocol as Reuter et al. (2013) (2013, 2011). We considered the column averaging kernels of all data products and co-located OCO-2 and TCCON measurements with a maximum time difference of 2h, a maximum distance of 500 km, and a maximum surface elevation difference of 250 m. In cases with multiple TCCON measurements of the same site co-locating with an OCO-2 sounding, we averaged the TCCON measurements. In total we found about 179000 and 378000 co-locations for FOCAL and the NASA product, respectively.

Fig. 30 shows the co-locations of all 19 sites with more than 250 co-locations. Per site statistics (bias and scatter, i.e., single sounding precision measured by the standard deviation of the difference to TCCON) are shown from north to south in Fig. 31. Note that global offsets have been removed for both figures (-0.29 ppm, 0.35 ppm, -0.62 ppm, and -0.94 ppm for NASA, NASA not bias corrected, FOCAL, and FOCAL not bias corrected).

Both algorithms show a somewhat similar bias site-to-site pattern regardless whether the bias correction is applied or not. The largest differences of the bias corrected satellite products can be found in Sodankylä and Tsukuba with larger than 1 ppm biases of the NASA product and FOCAL. The standard deviations of the site biases are 0.82 ppm and 0.67 ppm for the NASA product and FOCAL (0.69 ppm and 0.78 ppm if no bias correction is applied). These algorithm-to-algorithm differences are barely significant because TCCON's per site accuracy is about 0.4 ppm (1σ) (Wunch et al., 2011).

The analyzed algorithms also show a similar site-to-site pattern for the scatter with lowest values for the southern hemispheric sites probably due to smaller natural variability and, consequently, smaller representation errors. Both algorithms have a similar average scatter relative to TCCON before bias correction (1.62 ppm and 1.60 ppm for NASA and FOCAL) and after bias correction (1.31 ppm and 1.34 ppm for NASA and FOCAL). This means, both bias corrections primarily reduce the scatter rather than the site biases. However, according to Fig. 18, the influence of the bias correction on the spatial bias pattern may be larger elsewhere.

FOCAL's retrieved XH_2O has also been initially compared with TCCON in the same manner. However, due to the much larger natural variability of

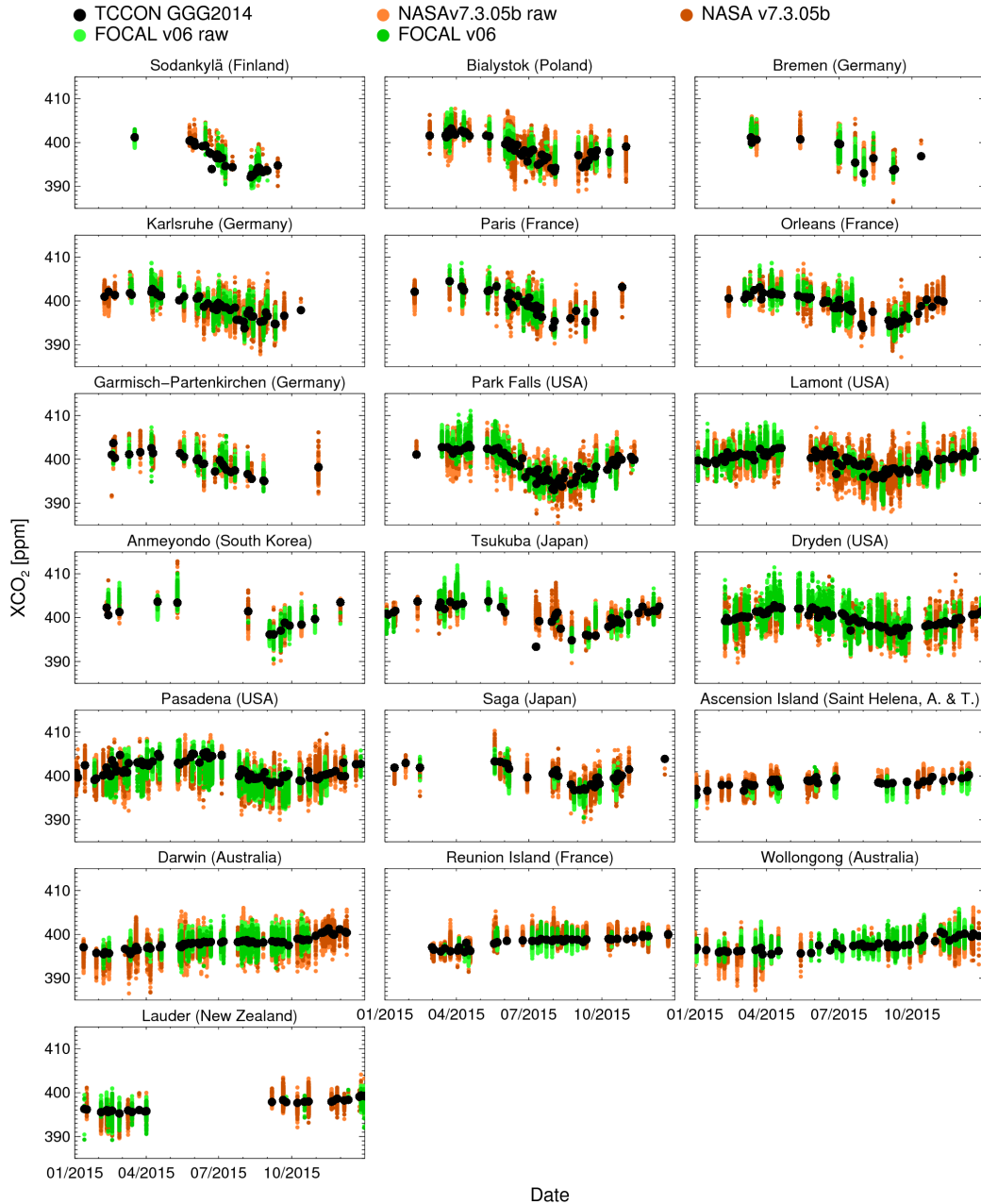


Figure 30: Validation of FOCAL v06 and NASA's operational OCO-2 L2 product (both with and without bias correction) with TCCON data from sites with more than 250 co-locations. The sites are ordered from north (top/left) to south (bottom/right): Sodankylä (Kivi et al., 2014), Bialystok (Deutscher et al., 2014), Bremen (Notholt et al., 2014), Karlsruhe (Hase et al., 2014), Paris (Te et al., 2014), Orleans (Warneke et al., 2014), Garmisch-Partenkirchen (Sussmann and Rettinger, 2014), Park Falls (Wennberg et al., 2014a), Lamont (Wennberg et al., 2014b), Anmeyondo (Goo et al., 2014), Tsukuba (Morino et al., 2014), Dryden (Iraci et al., 2014), Pasadena (Wennberg et al., 2014c), Saga (Shiomi et al., 2014), Ascension Island (Feist et al., 2014), Darwin (Griffith et al., 2014a), Reunion Island (De Maziere et al., 2014), Wollongong (Griffith et al., 2014b), and Lauder (Sherlock et al., 2014).

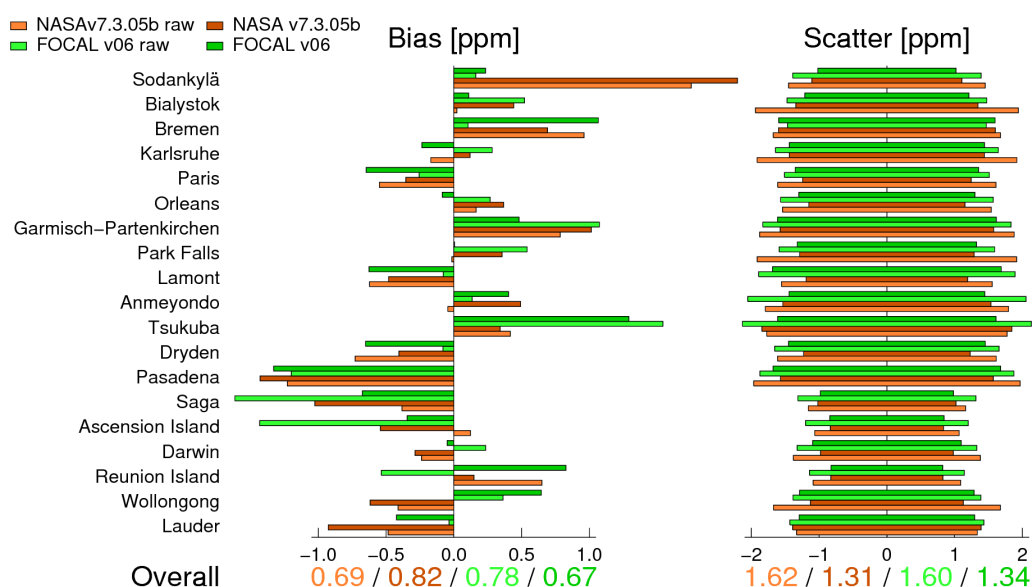


Figure 31: Validation statistics bias and scatter per TCCON site with more than 250 co-locations for FOCAL v06 and NASA's operational OCO-2 L2 product (both with and without bias correction). The summarizing values ("overall") represent the standard deviation of the site biases and the average scatter relative to TCCON. The sites are ordered from north (top) to south (bottom).

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	64
---------------------	---	---	-----------

water vapor (typically spanning a range from 500 ppm to 7000 ppm), we used stricter co-location criteria (1h maximum time difference and 150 km maximum distance) reducing the number of co-locations roughly by a factor of five. The global offset amounts to -150 ppm, the standard deviation of the site biases is 206 ppm, and the average single sounding precision is 293 ppm. It should be mentioned that, in contrast to XCO₂, the agreement significantly reduces when relaxing the co-location criteria. Conversely, a significant part of the observed deviations could still be due to representation errors which are expected to reduce for even stricter co-location criteria. This, however, would also further reduce the number of co-locations.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	65
---------------------	---	---	-----------

9 Version History

9.1 v08

Generation of a global 4-years data set.

Changes over v06:

- Improved cross section data bases with finer temperature, pressure, and wavelength grid in the wCO₂ (0.0026nm) and sCO₂ (0.0044nm) band.
- Quadratic wavelength and linear pressure interpolation of the cross section data base.
- Usage of HITRAN2016 as H₂O spectroscopy.
- Allowing negative values of p_s for improved convergence behavior.
- Widened limits for improved convergence behavior.
- Improved smoothing and noise error diagnostics.
- Usage of ECMWF ERA5 meteorological data.
- Bug fixes.

9.2 v06

First application and validation of the FOCAL OCO-2 XCO₂ algorithm to a larger global dataset of actually measured OCO-2 data as described by Reuter et al. (2017b).

Changes over v01:

- Development of a preprocessor including filtering, adaptation of the noise model, and zero level offset correction.
- Development of a postprocessor including filtering and bias correction.
- Implementation of the Levenberg-Marquardt minimizer.
- Bug fixes.

9.3 v01

The initial version of the FOCAL OCO-2 XCO₂ algorithm as described by Reuter et al. (2017c). This version has been used to analyzed simulated OCO-2 measurements.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	66
---------------------	---	---	-----------

References

- Ackerman, S., Frey, R., Strabala, K., Liu, Y., Gumley, L., Baum, B., and Menzel, P.: Discriminating clear-sky from cloud with MODIS - Algorithm Theoretical Basis Document (MOD35), Version 6.1, Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin - Madison, 2010.
- Bennartz, R. and Preusker, R.: Representation of the photon pathlength distribution in a cloudy atmosphere using finite elements, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 98, 202–219, 2006.
- Boesch, H., Brown, L., Castano, R., Christi, M., Connor, B., Crisp, D., Eldering, A., Fisher, B., Frankenberg, C., Gunson, M., Granat, R., McDuffie, J., Miller, C., Natraj, V., O'Brien, D., O'Dell, C., Osterman, G., Oyafuso, F., Payne, V., Polonski, I., Smyth, M., Spurr, R., Thompson, D., and Toon, G.: Orbiting Carbon Observatory-2 (OCO-2) - Level 2 Full Physics Retrieval - Algorithm Theoretical Basis, Version 2.0 Rev 2, National Aeronautics and Space Administration, Jet Propulsion Laboratory, California Institute of Technology, URL https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/OC02_L2_ATBD.V6.pdf, 2015.
- Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A.: SCIAMACHY – Mission Objectives and Measurement Modes, *Journal of the Atmospheric Sciences*, 56, 127–150, URL [http://dx.doi.org/10.1175/1520-0469\(1999\)056<0127:SMOAMM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1999)056<0127:SMOAMM>2.0.CO;2), 1999.
- Bovensmann, H., Buchwitz, M., Burrows, J. P., Reuter, M., Krings, T., Gerilowski, K., Schneising, O., Heymann, J., Tretner, A., and Erzinger, J.: A remote sensing technique for global monitoring of power plant CO₂ emissions from space and related applications, *Atmospheric Measurement Techniques*, 3, 781–811, doi:10.5194/amt-3-781-2010, URL <http://www.atmos-meas-tech.net/3/781/2010/>, 2010.
- Bril, A., Oshchepkov, S., and Yokota, T.: Application of a probability density function-based atmospheric light-scattering correction to carbon dioxide retrievals from GOSAT over-sea observations, *Remote Sensing of the Environment*, 117, 301–306, 2012.
- Bril, A., Oshchepkov, S., Yokota, T., and Inoue, G.: Parameterization of aerosol and cirrus cloud effects on reflected sunlight spectra measured from space: application of the equivalence theorem, *Applied Optics*, 46, 2460–2470, 2007.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	67
---------------------	---	---	-----------

Buchwitz, M., Reuter, M., Bovensmann, H., Pillai, D., Heymann, J., Schneising, O., Rozanov, V., Krings, T., Burrows, J. P., Boesch, H., Gerbig, C., Meijer, Y., and Löscher, A.: Carbon Monitoring Satellite (Carbon-Sat): assessment of atmospheric CO₂ and CH₄ retrieval errors by error parameterization, *Atmospheric Measurement Techniques*, 6, 3477–3500, doi:10.5194/amt-6-3477-2013, URL <http://www.atmos-meas-tech.net/6/3477/2013/>, 2013.

Buchwitz, M., Detmers, R., and GHG-CCI project team with contributions from NASA/ACOS/OCO-2 team members: ESA Climate Change Initiative (CCI) Product Specification Document (PSD) for the Essential Climate Variable (ECV) Greenhouse Gases (GHG) - Description of Common Parameters for core (ECA) products, URL http://www.esa-ghg-cci.org/index.php?q=webfm_send/160, 2014.

Burrows, J. P., Hölzle, E., Goede, A. P. H., Visser, H., and Fricke, W.: SCIAMACHY – Scanning Imaging Absorption Spectrometer for Atmospheric Cartography, *Acta Astronautica*, 35, 445–451, 1995.

Chevallier, F.: Validation report for the inverted CO₂ fluxes, v16r1. CAMS deliverable CAMS73_2015S2_D73.1.4.2-1979-2016-v1_201707CO2, URL https://atmosphere.copernicus.eu/sites/default/files/FileRepository/Resources/Validation-reports/Fluxes/CAMS73_2015SC2_D73.1.4.2-1979-2016-v1_201707_final.pdf, 2017.

Chevallier, F., Bréon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO₂ sources and sinks: Theoretical study in a variational data assimilation framework, *Journal of Geophysical Research*, 112, D09307, doi:10.1029/2006JD007375, URL <http://dx.doi.org/10.1029/2006JD007375>, 2007.

Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morguí, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO₂ surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, *Journal of Geophysical Research*, 115, doi:10.1029/2010jd013887, URL <http://dx.doi.org/10.1029/2010JD013887>, 2010.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	68
---------------------	---	---	-----------

Crisp, D., Atlas, R. M., Bréon, F.-M., Brown, L. R., Burrows, J. P., Ciais, P., Connor, B. J., Doney, S. C., Fung, I. Y., Jacob, D. J., Miller, C. E., O'Brien, D., Pawson, S., Randerson, J. T., Rayner, P., Salawitch, R. S., Sander, S. P., Sen, B., Stephens, G. L., Tans, P. P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., Kuang, Z., Chudasama, B., Sprague, G., Weiss, P., Pollock, R., Kenyon, D., and Schroll, S.: The Orbiting Carbon Observatory (OCO) mission, *Advances in Space Research*, 34, 700–709, 2004.

Crisp, D., Pollock, H. R., Rosenberg, R., Chapsky, L., Lee, R. A. M., Oyafuso, F. A., Frankenberg, C., O'Dell, C. W., Bruegge, C. J., Doran, G. B., Eldering, A., Fisher, B. M., Fu, D., Gunson, M. R., Mandrake, L., Osterman, G. B., Schwandner, F. M., Sun, K., Taylor, T. E., Wennberg, P. O., and Wunch, D.: The on-orbit performance of the Orbiting Carbon Observatory-2 (OCO-2) instrument and its radiometrically calibrated products, *Atmospheric Measurement Techniques*, 10, 59–81, doi:10.5194/amt-10-59-2017, URL <https://www.atmos-meas-tech.net/10/59/2017/>, 2017.

De Maziere, M., Sha, M. K., Desmet, F., Hermans, C., Scolas, F., Kumps, N., Metzger, J.-M., Duflot, V., and Cammas, J.-P.: TCCON data from Reunion Island (La Reunion), France, Release GGG2014R0., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.reunion01.R0/1149288, 2014.

Deutscher, N., Notholt, J., Messerschmidt, J., Weinzierl, C., Warneke, T., Petri, C., Grupe, P., and Katrynski, K.: TCCON data from Bialystok, Poland, Release GGG2014R1., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.bialystok01.R1/1183984, 2014.

Eldering, A., Pollock, R., Lee, R., Rosenberg, R., Oyafuso, F., Crisp, D., Chapsky, L., and Granat, R.: Orbiting Carbon Observatory-2 (OCO-2) - LEVEL 1B - Algorithm Theoretical Basis, Version 1.2 Rev 1, National Aeronautics and Space Administration, Jet Propulsion Laboratory, California Institute of Technology, URL http://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/OC02_L1B_ATBD.V7.pdf, 2015.

Eldering, A., O'Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., Avis, C., Braverman, A., Castano, R., Chang, A., et al.: The Orbiting Carbon Observatory-2: first 18 months of science data products, *Atmospheric Measurement Techniques*, 10, 549, 2017.

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

GGG2014R0., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.ascension01.R0/1149285, 2014.

Frankenberg, C., Butz, A., and Toon, G. C.: Disentangling chlorophyll fluorescence from atmospheric scattering effects in O2 A-band spectra of reflected sun-light, *Geophysical Research Letters*, 38, n/a–n/a, doi:10.1029/2010GL045896, URL <http://dx.doi.org/10.1029/2010GL045896>, 103801, 2011.

Goo, T.-Y., Oh, Y.-S., and Velazco, V. A.: TCCON data from Anmeyondo, South Korea, Release GGG2014R0., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.anmeyondo01.R0/1149284, 2014.

Griffith, D. W. T., Deutscher, N., Velazco, V. A., Wennberg, P. O., Yavin, Y., Aleks, G. K., Washenfelder, R., Toon, G. C., Blavier, J.-F., Murphy, C., Jones, N., Kettlewell, G., Connor, B., Macatangay, R., Roehl, C., Ryzek, M., Glowacki, J., Culgan, T., and Bryant, G.: TCCON data from Darwin, Australia, Release GGG2014R0., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.darwin01.R0/1149290, 2014a.

Griffith, D. W. T., Velazco, V. A., Deutscher, N., Murphy, C., Jones, N., Wilson, S., Macatangay, R., Kettlewell, G., Buchholz, R. R., and Riggensbach, M.: TCCON data from Wollongong, Australia, Release GGG2014R0., TCCON data archive, hosted by CDIAC, doi: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 GGG2014R1., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.karlsruhe01.R1/1182416, 2014.

Heymann, J., Reuter, M., Hilker, M., Buchwitz, M., Schneising, O., Bovensmann, H., Burrows, J. P., Kuze, A., Suto, H., Deutscher, N. M., Dubey, M. K., Griffith, D. W. T., Hase, F., Kawakami, S., Kivi, R., Morino, I., Petri, C., Roehl, C., Schneider, M., Sherlock, V., Sussmann, R., Velazco, V. A., Warneke, T., and Wunch, D.: Consistent satellite XCO₂ retrievals from SCIAMACHY and GOSAT using the BESD algorithm, *Atmospheric Measurement Techniques*, 8, 2961–2980, doi:10.5194/amt-8-2961-2015, URL <http://www.atmos-meas-tech.net/8/2961/2015/>, 2015.

Hu, C., Lee, Z., and Franz, B.: Chlorophyll-a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference, *Journal of Geophysical Research: Oceans*, 117, 2012.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	70
---------------------	---	---	-----------

Iraci, L., Podolske, J., Hillyard, P., Roehl, C., Wennberg, P. O., Blavier, J.-F., Landeros, J., Allen, N., Wunch, D., Zavaleta, J., Quigley, E., Osterman, G., Albertson, R., Dunwoody, K., and Boyden, H.: TCCON data from Armstrong Flight Research Center, Edwards, CA, USA, Release GGG2014R1., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.edwards01.R1/1255068, 2014.

Kivi, R., Heikkinen, P., and Kyro, E.: TCCON data from Sodankyla, Finland, Release GGG2014R0., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.sodankyla01.R0/1149280, 2014.

Kurucz, H. L.: The solar spectrum: atlases and line identifications, workshop on laboratory and astronomical high resolution spectra. astronomical society of the pacific conference series, Proceedings of ASP Conference No.81 Held in Brussels, Belgium 29 August-2 September 1994, pp. 17–31, URL www.scopus.com, 1995.

Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, Applied Optics, 48, 6716, doi:10.1364/AO.48.006716, URL <http://dx.doi.org/10.1364/AO.48.006716>, 2009.

Miller, C. E., Crisp, D., DeCola, P. L., Olsen, S. C., Randerson, J. T., Michalak, A. M., Alkhaled, A., Rayner, P., Jacob, D. J., Suntharalingam, P., Jones, D. B. A., Denning, A. S., Nicholls, M. E., Doney, S. C., Pawson, S., Bösch, H., Connor, B. J., Fung, I. Y., O'Brien, D., Salawitch, R. J., Sander, S. P., Sen, B., Tans, P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., and Law, R. M.: Precision requirements for space-based X_{CO_2} data, Journal of Geophysical Research, 112, D10314, doi:10.1029/2006JD007659, 2007.

Morino, I., Matsuzaki, T., and Shishime, A.: TCCON data from Tsukuba, Ibaraki, Japan, 125HR, Release GGG2014R1., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.tsukuba02.R1/1241486, 2014.

Notholt, J., Petri, C., Warneke, T., Deutscher, N., Buschmann, M., Weinzierl, C., Macatangay, R., and Grupe, P.: TCCON data from Bremen, Germany, Release GGG2014R0., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.bremen01.R0/1149275, 2014.

O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller,

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	71
---------------------	---	---	-----------

C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: The ACOS CO₂ retrieval algorithm - Part 1: Description and validation against synthetic observations, *Atmospheric Measurement Techniques*, 5, 99–121, doi:10.5194/amt-5-99-2012, URL <http://www.atmos-meas-tech.net/5/99/2012/>, 2012.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *Proceedings of the National Academy of Sciences of the United States of America*, 104, 18 925–18 930, doi:10.1073/pnas.0708986104, URL <http://www.pnas.org/content/104/48/18925.abstract>, updates documented at <http://carbontracker.noaa.gov>, 2007.

Rascher, U., Agati, G., Alonso, L., Cecchi, G., Champagne, S., Colombo, R., Damm, A., Daumard, F., de Miguel, E., Fernandez, G., Franch, B., Franke, J., Gerbig, C., Gioli, B., Gómez, J. A., Goulas, Y., Guanter, L., Gutiérrez-de-la Cámara, O., Hamdi, K., Hostert, P., Jiménez, M., Kosvanova, M., Lognoli, D., Meroni, M., Miglietta, F., Moersch, A., Moreno, J., Moya, I., Neininger, B., Okujeni, A., Ounis, A., Palombi, L., Raimondi, V., Schickling, A., Sobrino, J. A., Stellmes, M., Toci, G., Toscano, P., Udelhoven, T., van der Linden, S., and Zaldei, A.: CEFLES2: the remote sensing component to quantify photosynthetic efficiency from the leaf to the region by measuring sun-induced fluorescence in the oxygen absorption bands, *Biogeosciences*, 6, 1181–1198, doi:10.5194/bg-6-1181-2009, URL <http://www.biogeosciences.net/6/1181/2009/>, 2009.

Reuter, M., Thomas, W., Albert, P., Lockhoff, M., Weber, R., Karlsson, K.-G., and Fischer, J.: The CM-SAF and FUB cloud detection schemes for SEVIRI: Validation with synoptic data and initial comparison with MODIS and CALIPSO, *Journal of Applied Meteorology and Climatology*, 48, 301–316, doi:10.1175/2008JAMC1982.1, URL <http://dx.doi.org/10.1175/2008JAMC1982.1>, 2009.

Reuter, M., Bovensmann, H., Buchwitz, M., Burrows, J. P., Connor, B. J., Deutscher, N. M., Griffith, D. W. T., Heymann, J., Keppel-Aleks, G., Messerschmidt, J., Notholt, J., Petri, C., Robinson, J., Schneising, O., Sherlock, V., Velazco, V., Warneke, T., Wennberg, P. O., and Wunch, D.: Retrieval of atmospheric CO₂ with enhanced accuracy and precision

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	72
---------------------	---	---	-----------

- from SCIAMACHY: Validation with FTS measurements and comparison with model results., J. Geophys. Res., 116, doi:10.1029/2010JD015047, URL <http://dx.doi.org/10.1029/2010JD015047>, 2011.
- Reuter, M., Buchwitz, M., Schneising, O., Hase, F., Heymann, J., Guerlet, S., Cogan, A. J., Bovensmann, H., and Burrows, J. P.: A simple empirical model estimating atmospheric CO₂ background concentrations, Atmospheric Measurement Techniques, 5, 1349–1357, doi:10.5194/amt-5-1349-2012, URL <http://dx.doi.org/10.5194/amt-5-1349-2012>, 2012.
- Reuter, M., Bösch, H., Bovensmann, H., Bril, A., Buchwitz, M., Butz, A., Burrows, J. P., O'Dell, C. W., Guerlet, S., Hasekamp, O., Heymann, J., Kikuchi, N., Oshchepkov, S., Parker, R., Pfeifer, S., Schneising, O., Yokota, T., and Yoshida, Y.: A joint effort to deliver satellite retrieved atmospheric CO₂ concentrations for surface flux inversions: the ensemble median algorithm EMMA, Atmospheric Chemistry and Physics, 13, 1771–1780, doi: 10.5194/acp-13-1771-2013, URL <http://www.atmos-chem-phys.net/13/1771/2013/>, 2013.
- Reuter, M., Buchwitz, M., Hilker, M., Heymann, J., Bovensmann, H., Burrows, J. P., Houweling, S., Liu, Y. Y., Nassar, R., Chevallier, F., Ciais, P., Marshall, J., and Reichstein, M.: How Much CO₂ Is Taken Up by the European Terrestrial Biosphere?, Bulletin of the American Meteorological Society, 98, 665–671, doi:10.1175/BAMS-D-15-00310.1, 2017a.
- Reuter, M., Buchwitz, M., Schneising, O., Noël, S., Bovensmann, H., and Burrows, J. P.: A fast atmospheric trace gas retrieval for hyperspectral instruments approximating multiple scattering - Part 2: application to XCO₂ retrievals from OCO-2, Remote Sensing, 9, doi:10.3390/rs9111102, URL <http://www.mdpi.com/2072-4292/9/11/1102>, 2017b.
- Reuter, M., Buchwitz, M., Schneising, O., Noël, S., Rozanov, V., Bovensmann, H., and Burrows, J. P.: A fast atmospheric trace gas retrieval for hyperspectral instruments approximating multiple scattering - Part 1: radiative transfer and a potential OCO-2 XCO₂ retrieval setup, Remote Sensing, 9, doi:10.3390/rs9111159, URL <http://www.mdpi.com/2072-4292/9/11/1159>, 2017c.
- Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific Publishing, Singapore, 2000.
- Roedel, W. and Wagner, T.: Physik unserer Umwelt: Die Atmosphäre, Springer, 2011.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	73
---------------------	---	---	-----------

- Rozanov, V., Rozanov, A., Kokhanovsky, A., and Burrows, J.: Radiative transfer through terrestrial atmosphere and ocean: Software package SCIATRAN, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 133, 13 – 71, doi:<https://doi.org/10.1016/j.jqsrt.2013.07.004>, URL <http://www.sciencedirect.com/science/article/pii/S0022407313002872>, 2014.
- Sherlock, V., Connor, B., Robinson, J., Shiona, H., Smale, D., and Pollard, D.: TCCON data from Lauder, New Zealand, 125HR, Release GGG2014R0., TCCON data archive, hosted by CDIAC, doi: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 GGG2014R0., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.saga01.R0/1149283, 2014.
- Stammes, P.: OMI Algorithm Theoretical Basis Document, Volume III, Clouds, Aerosols, and Surface UV Irradiance (ATBD-OMI-03), Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands, URL <http://eospso.gsfc.nasa.gov/sites/default/files/atbd/ATBD-OMI-03.pdf>, 2002.
- Sussmann, R. and Rettinger, M.: TCCON data from Garmisch, Germany, Release GGG2014R0., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.garmisch01.R0/1149299, 2014.
- Taylor, T. E., O'Dell, C. W., Partain, P. T., Cronk, H. Q., Nelson, R. R., Rothstein, E. J., Chang, A. Y., Osterman, G. B., Pollock, R. H., and Gunson, M. R.: Orbiting Carbon Observatory-2 (OCO-2) cloud screening algorithms: validation against collocated MODIS and CALIOP data, *Atmospheric Measurement Techniques*, 9, 973, 2016.
- Te, Y., Jeseck, P., and Janssen, C.: TCCON data from Paris, France, Release GGG2014R0., TCCON data archive, hosted by CDIAC, doi:10.14291/tccon.ggg2014.paris01.R0/1149279, 2014.
- Thompson, D. R., Chris Benner, D., Brown, L. R., Crisp, D., Malathy Devi, V., Jiang, Y., Natraj, V., Oyafuso, F., Sung, K., Wunch, D., and et al.: Atmospheric validation of high accuracy CO₂ absorption coefficients for the OCO-2 mission, *J Quant Spectrosc Radiat Transfer*, 113, 2265–2276, doi:10.1016/j.jqsrt.2012.05.021, URL <http://dx.doi.org/10.1016/j.jqsrt.2012.05.021>, 2012.

ESA CCI+ ECV GHG	ATBD FOCAL OCO-2 Version 1 August 2019	Institute of Env. Physics, University of Bremen	74
---------------------	---	---	-----------

Warneke, T., Messerschmidt, J., Notholt, J., Weinzierl, C., Deutscher, N., Petri, C., Grupe, P., Vuillemin, C., Truong, F., Schmidt, M., Ramonet, M., and Parmentier, E.: TCCON data from Orleans, France, Release GGG2014R0., TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.orleans01.R0/1149276, 2014.

Wennberg, P. O., Roehl, C., Wunch, D., Toon, G. C., Blavier, J.-F., Washenfelder, R., Keppel-Aleks, G., Allen, N., and Ayers, J.: TCCON data from Park Falls, Wisconsin, USA, Release GGG2014R0., TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.parkfalls01.R0/1149161, 2014a.

Wennberg, P. O., Wunch, D., Roehl, C., Blavier, J.-F., Toon, G. C., Allen, N., Dowell, P., Teske, K., Martin, C., and Martin, J.: TCCON data from Lamont, Oklahoma, USA, Release GGG2014R1., TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.lamont01.R1/1255070, 2014b.

Wennberg, P. O., Wunch, D., Yavin, Y., Toon, G. C., Blavier, J.-F., Allen, N., and Keppel-Aleks, G.: TCCON data from Jet Propulsion Laboratory, Pasadena, California, USA, Release GGG2014R0., TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.jpl01.R0/1149163, 2014c.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network (TCCON), Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences, 369, 2087–2112, doi:10.1098/rsta.2010.0240, URL <http://dx.doi.org/10.1098/rsta.2010.0240>, 2011.

Yoshida, Y., Kikuchi, N., Morino, I., Uchino, O., Oshchepkov, S., Bril, A., Saeki, T., Schutgens, N., Toon, G. C., Wunch, D., and et al.: Improvement of the retrieval algorithm for GOSAT SWIR XCO₂ and XCH₄ and their validation using TCCON data, Atmospheric Measurement Techniques, 6, 1533–1547, doi:10.5194/amt-6-1533-2013, URL <http://dx.doi.org/10.5194/amt-6-1533-2013>, 2013.