SRON NETHERLANDS INSTITUTE FOR SPACE RESEARCH

RemoTAP Algorithm Theoretical Baseline Document

Otto Hasekamp, Sha Lu, Guangliang Fu, Stephanie Rusli, Lianghai Wu, Jochen Landgraf

SRON Netherlands Institute for Space Research, Niels Bohrweg 4 2333 CA Leiden, the Netherlands

Document reference: SRON-ESG-RP-2021-010
Release data: 31st October 2021

Citation: Hasekamp, O.P., Lu, S., Fu, G., Rusli, S., Wu, L., Landgraf, J.: RemoTAP Algorithm Theoretical Baseline Document, SRON-ESG-RP-2021-010, SRON Netherlands Institute for Space Research, Leiden, 2021

1 Introduction to RemoTAP

To measure atmospheric CO_2 from space, several satellite missions in space observe Earth reflected radiances in the near-infrared (NIR) spectral range with the O_2 A absorption band, in the shortwave infrared at 1.6 μm (SWIR-1) with weak absorption lines of CO_2, CH_4, and H_2O, and around 2.0 μm (SWIR-2) with strong absorption lines of CO_2,and H_2O. Examples of such instruments are the GOSAT-1 and GOSAT-2 and OCO-2 spectrometers. Also the Copernicus CO_2 monitoring (CO2M) spectrometer includes these spectral bands as listed in Table 1.1. The reason to choose these spectral band is that aerosol information can be inferred from the O_2 A and the 2.0 μm band, whereas the the 1.6 μm band provides information on atmospheric XCO_2 and XCH_4 column mixing ratios. Buchwitz et al.[1] (and references therein) provide a good overview of currently available retrieval approaches.

The XCO_2 data products from NIR, SWIR-1 and SWIR-2 measurements suffers from significant aerosol induced errors depending on the atmospheric aerosol load. This means that using spectrometer measurement only the stringent XCO_2 accuracy requirements of the CO2M missions can only be met after significant a posteriori bias correction[2]. Therefore, a Multi-Angle Polarimeter (MAP) is added to the CO2M instrument suite to better correct for aerosols ([2] and[3]). The CO2M polarimeter concept is a band based polarimeter measuring at the UV-VIS-NIR spectral bands of 3MI[4] with a 40 viewing angle hyper-angular resolution.

For atmospheric passive remote sensing, polarimeters that measure both intensity and polarization and that observe a ground pixel under multiple viewing angles contain the richest set of information about aerosols in our atmosphere[5]. The reason is that the angular dependence of the scattering matrix elements related to linear polarization, depends strongly on the microphysical aerosol properties, like refractive index and particle size[6,7]. Furthermore, the polarization signal is mostly dominated for light that has been scattered only once, which means that the characteristics of the scattering matrix remain largely preserved in a top-of-atmosphere polarization measurement. The added value of polarization has been demonstrated by a number of studies on synthetic measurements[7–10], airborne measurements[11–17], and spaceborne measurements[18–20]. Algorithms for aerosol retrieval from MAP instruments can be divided in two main groups: LookUp-Table (LUT) based approaches[21–23] and full inversion approaches[12,13,18–20,24]. Generally speaking, LUT approaches are faster but less accurate than full inversion approaches because LUT approaches choose the best fitting aerosol model from a discrete lookup table. Therefore, LUT based algorithms are not suited to fully exploit the information contained in MAP measurements. Full inversion approaches are more accurate but slower because they require radiative transfer calculations as part of the retrieval procedure.

This section describes the theoretical baseline of the Remote Sensing of Trace Gas and Aerosol Properties (RemoTAP) algorithm for level-1 to level-2 processing of the (CO2M) mission. The RemoTAP algorithm as well as its software implementation is specifically designed for CO2M data processing exploiting both the CO_2 spectrometer (CO2I) measurements and the multi-angle polarimeter (MAP) observations in a flexible and synergistic manner. It finds it origin in the SRON aerosol retrieval algorithm[14,15,17,18,20,25–27]. The algorithm was recently extended to include spectroscopic measurements and the ability to retrieve trace gas columns[2]. This extension is partly based on the RemoTeC algorithm which was also developed at SRON and applied extensively for Green-House Gas (GHG) retrieval from GOSAT[28,29] as part of ESA's Climate Change Initiative (CCI)[1], OCO-2[30–32], and TROPOMI[33–35].

RemoTAP aims to server both the need of EUMETSAT for an operational algorithm as well as ESA's need for mission end-to-end performance analysis during the implementation phase B2,C,D and first data analysis during the commissioning phase E1. Figure 1.1 summarizes the high-level processing options of RemoTAP. Using the level-1B product of CO2I and the collocated level-1C MAP product, the algorithm allows for four different processing strategies.

1. **CO2I-only**: Here, the retrieval uses only measurements of the CO2I spectrometer instrument to infer XCO_2 and aerosol parameters to describe the atmospheric light path.
2. **MAP-only**: This processing line employs only MAP measurements to determine atmospheric aerosol properties. A direct link to the required XCO_2 performance is not given.

3. **sequential MAP→CO2I**: This processing starts with a aerosol retrieval using MAP measurements and subsequently uses the derived aerosol information as input to spectrometer-only retrieval. The latter infers XCO_2 and updates the aerosol product using the additional aerosol information of the spectrometer measurements
4. **simultaneous MAP+CO2I**: This option uses simultaneously MAP and CO2I spectrometer observations to infer both aerosol and XCO_2 information.

For ESA's objective of an end-to-end performance evaluation during the implementation phase B2,C,D and for a first data analysis during the commissioning phase E1, both the spectrometer-only and MAP-only processing option is essential. It supports to evaluate the performance of individual instruments. Moreover, to relate the MAP performance to the required XCO_2 performance, the sequential MAP-spectrometer processing line is required. The simultaneous MAP-spectrometer data processing is considered to be most appropriate to fully explore the CO2M mission concept. Because of the larger complexity of this algorithm, we consider this processing line as most appropriate for the operational data exploitation to be performed by EUMETSAT. Therefore, this document includes the description of the theoretical baseline of the RemoTAP algorithm for all four performance options of RemoTAP. The document specifies the required input, the provided output and the needed ancillary data. It provides a first performance analysis and specifies the product accuracy and the run-time performance of the RemoTAP software.

The section is structured as follows: Section 2 gives an introduction to the algorithm. Section 3 provides a detailed description of the RemoTAP algorithm, which includes a description of the model atmosphere, the forward model, the inversion approach, and the approach to obtain a suitable first guess state vector. We provide flow charts of the algorithm and its key components for the spectrometer-only, MAP-only and the sequential MAP→CO2I and simultaneous MAP→CO2I retrieval. The input and output fields are specified in detail in Sec. 48 and Sec. 56 describes the a priori data handling. The IDEF diagrams of Sec. 61 illustrates the data flow on the level of the RemoTAP main modules and so support the implementation of the software in an future CO2M processing framework. Section 62 yields an error analysis for the MAP+CO2I and CO2I-only retrievals and feasibility aspects are discussed in Sec. 73 and a validation strategy for the RemoTAP L2 data product is subject of Sec. 78. In a later phase, Sec. 86 will provide an analysis of instrument induced error on the RemoTAP data product. Finally, assumptions and limitations, also in perspective of potential algorithm improvements, are discussed in Sec. 87. A description of the radiative transfer solver LINTRAN V2.0 is given in the final section 90.

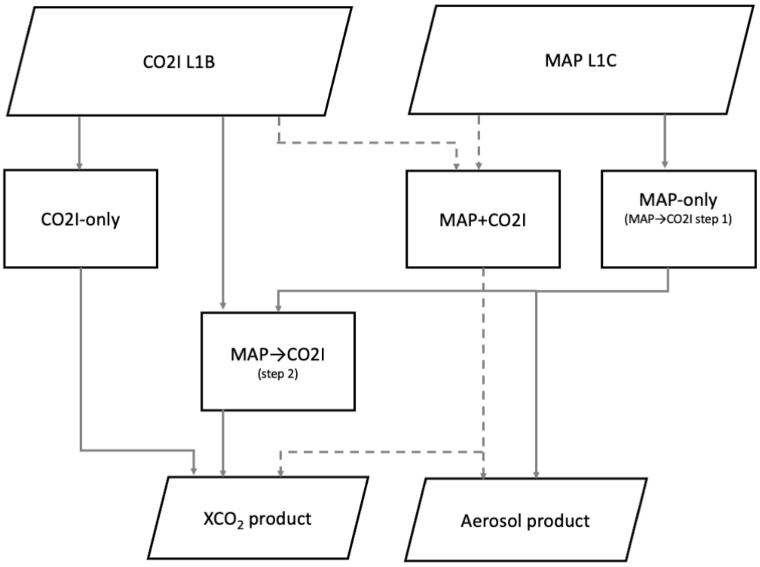

Figure 1.1: RemoTAP high-level processing options.

Table 1.1: Spectral windows <of the CO2M spectrometer.

band	**spectral window [nm]**	**target species**
NIR	747–773	O_2
SWIR-1	1590–1675	CO_2 and CH_4
SWIR-2	1993–2095	CO_2 and H_2O

2 RemoTAP processing logic

The RemoTAP algorithm is designed to use a mulit-core and multi-threading processing architecture such that single threads are utilized to process data from single ground pixels. Figure 2.2 summarizes the top level software structure from the perspective of an operating framework.

The software separates the initialization from the actual ground-pixel based retrieval, where data granules can be processed in a parallel manner. In this fashion, it optimizes the run time performance within an overall processing framework. We begin with the initialization of the processor. It provides input to the RemoTAP retrieval, which has to be allocated only once per processing call. It includes static data like the retrieval input settings, lookup tables for molecular absorption cross sections and aerosol optical properties. Also the instrument spectral response function (ISRF) as part of the instrument calibration key data and a solar reference spectrum is required. All these data are accessible for the remaining software via pointers to an allocated internal memory. The numerical effort is minor and any parallelization of the software is not considered. Next, we estimate solar line-by-line spectra from CO2I solar irradiance observations by a spectral deconvolution approach. This has to be performed for each swath position and the corresponding calculations can be parallelized.

Per data granule, ancillary input data are collocated to the CO2I L1B spatial grid. This module is not described in this ATBD. To a large extend, it is an algorithm independent operation and one may consider this as an general CO2M processing tool applicable also for processor components other than RemoTAP. In a loop over all spatial pixels of a data granule, the non-scattering retrieval is performed to derive, inter alia, the XCO_2 and XCH_4 proxy product followed by the full-physics retrieval. The latter has the option of the four different processing lines: (1) CO2I-only, (2) sequential MAP→CO2I, and (3) simultaneous MAP+CO2I retrievals. Additionally, RemoTAP includes the option of MAP-only data processing, which in the context of CO2M enables to evaluate the MAP instrument performance independent from the spectrometer. In all cases, the parallel computation over ground pixels is essential to satisfy the performance requirements.

Finally, the software architecture includes a dedicated module to perform an a posterior bias correction, although due to the novel combination of payload instrument the XCO_2 and XCH_4 biases should be reduced to a minimum.

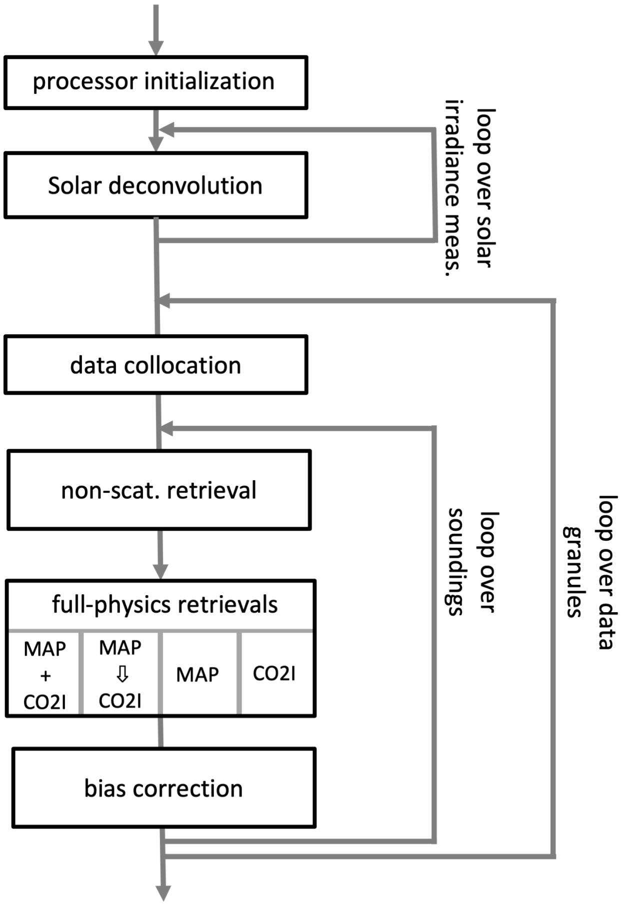

Figure 2.2: RemoTAP top level algorithm architecture.

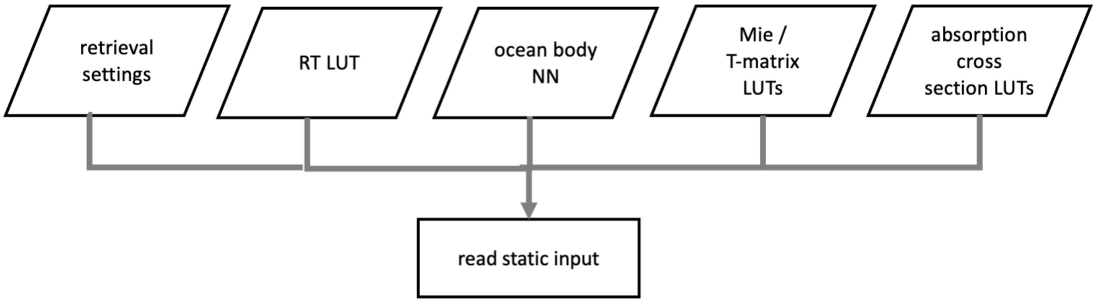

Figure 4.3: Schematic overview of the processor initialization using static input data.

3 Algorithm description

The previous section provided an outline of the RemoTAP algorithm, which is described in detail in this section.

4 Processor initializtion

The RemoTAP algorithm requires a series of static input data, which are the retrieval settings, the radiative transfer look up table RT-LUT for the 1st guess retrieval, the neural network (NN) for the ocean body reflection matrix, the Mie/T-Matrix LUTs, and absorption cross sections of telluric absorbing gases. The algorithm settings can be provided by a list of different settings or as a concatenated string, which is unraveled by the RemoTAP software. This initialization needs to happen only once per processor call and its data flow is illustrated in the flow chart of Fig. 4.3. All static input is stored in an allocated internal memory, and is available for the remaining processing via shared memory management. The static input variables are described in more detail in Sec. 49.

5 Deconvolution of the solar spectrum

For each across-track position of the CO2I observations, we infer a line-by-line solar spectrum from the CO2I solar irradiance measurements via a spectral deconvolution. The CO2I solar measurement $\mathbf{F}_{0,\text{meas}}$ can be simulated by

$$\mathbf{F}_{0,\text{meas}}\left(\lambda_i\right)=\mathbf{S}_{\text{sun}}\mathbf{F}_0\left(\lambda\right) \tag{5.1}$$

where $\mathbf{F}_{0,\text{meas}}$ is the measurement, $\mathbf{F}_0$ represents the deconvolved solar spectrum and and $\mathbf{S}_{\text{sun}}$ is the ISRF matrix of the solar measurement, containing in each row the sampled ISRF for each irradiance element of $\mathbf{F}_{0,\text{meas}}$.

First, we determine a spectral shift of the solar measurement to ensure that the Fraunhofer lines of the solar measurement are aligned with the model solar spectrum. For this purpose, we apply a line search algorithm, where the CO2I irradiance measurements are compared with the model solar spectrum $\mathbf{F}_0^*$ convolved with the ISRF,

$$\delta\lambda_s=\max_{\delta\lambda_s}\left\{\rho\left(\mathbf{F}_{0,\text{meas}}\left(\lambda+\delta\lambda_s\right),\mathbf{F}_0^*\left(\lambda\right)\right)\right\} \tag{5.2}$$

where $\rho\left(X,Y\right)$ is the Pearson correlation coefficient.

After correcting the spectral shift, we invert Eq. (5.1) by a deconvolution approach, where the ISRF matrix S_{sun} gives the linear forward model to be inverted. The inversion is a linear problem, which can be solved in one step. It is complicated by the fact that the problem is underdetermined. The measurement vector $\mathbf{F}_{0,\text{meas}}$ contains fewer values than the solar spectrum to be retrieved $\mathbf{F}_0$.

Van Deelen et al. (2007)[36] showed that the least squares minimum length solution, which minimizes the length of the solution vector as a side constraint to solve the underdetermined inversion problem, is of sufficient accuracy to simulate Earth radiance measurements of the GOME mission. Following this approach, we obtain

$$\mathbf{F}_0 = \mathbf{S}_{\text{sun}}^T(\mathbf{S}_{\text{sun}}\mathbf{S}_{\text{sun}}^T)^{-1}\mathbf{F}_{0,\text{meas}} \ , \tag{5.3}$$

which contains significant noise contributions. However, this noise is in the null-space of the ISRF. This means that after convolving the simulated line-by-line radiance spectra of the reflected sun light to the CO2I spectral resolution, the noise of the solar spectrum is reduced significantly and so will affect the simulation only little after the final convolution of the Earth radiance spectrum with its corresponding ISRF. To mitigate edge effects, we execute the deconvolution on a slightly extended spectral window ($\sim$ 3 nm extra on both sides of the window).

Assuming that the Earth radiance and the solar irradiance are affected similarly by instrument and calibration errors, the use of $\mathbf{F}_0$ instead of an a priori solar reference spectrum has a clear advantage to reduce spectral fit residuals. As the solar spectrum is already corrected for a spectral shift ($\delta\lambda_s$), any further spectral alignment fo the solar spectrum is not considered in the CO2M RemoTAP data processing. For further details and for an overview of the benefits of this method we refer to Deelen et al., 2007[36] and Wassmann et al., 2015[37].

6 Preprocessing

RemoTAP can be combined with SRON's preprocessing tool box MIPrep, which resamples input data to the CO2I spatial samples, the latter serving as a reference mesh for the CO2M data product. The toolbox includes a horizontal 2-dimensional interpolation of the meteo data, like temperature and humidity, and trace gas vertical profiles for $\mathrm{CO_2}$, $\mathrm{CH_4}$ and $\mathrm{H_2O}$. A schematic overview of the pre-processing step is depicted in Fig. 6.4. The output data can be provided as return values to a calling framework or can be written to an output file, which we consider as a relevant processor option during the development phase. The relevant inputs are a CLIM cloud mask, meteo data (ECMWF), and atmospheric composition data (CAMS). Optional, this processing step can be taken over by a framework tool supporting the multi-algorithm processor infrastructure. The current version of the RemoTAP ATBD does not describe the preprocessing of data in detail, which we consider a standard practice for atmospheric missions.

7 Full-physics Retrieval

8 Forward Model

The radiance and state of polarization of light at a given wavelength can be described by an intensity vector $\mathbf{I}$ which has the Stokes parameters I, Q, U, V as its components[38]:

$$\mathbf{I} = [I, Q, U, V]^T \ . \tag{8.4}$$

Here, T indicates the transposed vector, and the Stokes parameters are defined with respect to a certain reference plane. Here, we will use the local meridian plane as reference plane. The angular dependence of single scattering of polarized light can be described by means of the scattering phase matrix $\mathbf{P}$. We will restrict ourselves to scattering phase matrices of the form

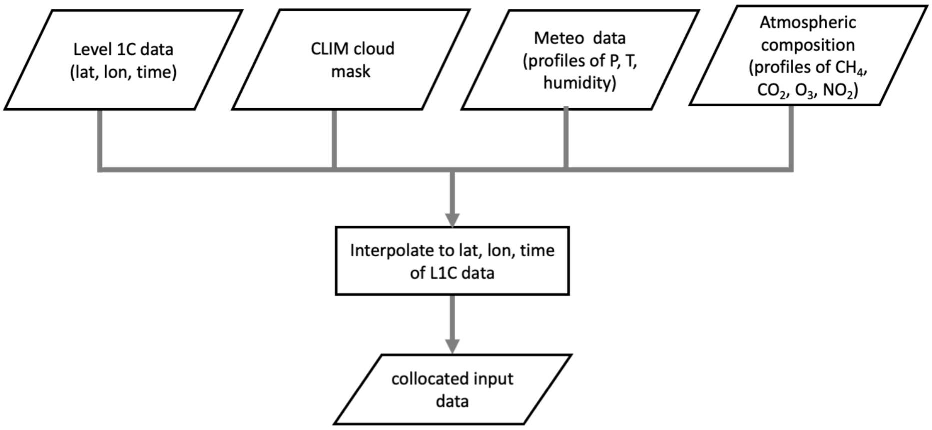

Figure 6.4: Schematic overview of the algorithm pre-processing step.

$$\mathbf{P}(\Theta) = \begin{pmatrix} p_1(\Theta) & p_5(\Theta) & 0 & 0 \\ p_5(\Theta) & p_2(\Theta) & 0 & 0 \\ 0 & 0 & p_3(\Theta) & p_6(\Theta) \\ 0 & 0 & -p_6(\Theta) & p_4(\Theta) \end{pmatrix} . \tag{8.5}$$

where $p_1, p_2, ..., p_6$ are certain functions of scattering angle Θ, i.e. the angle between the incoming and scattered beams, and the scattering plane is the plane of reference. This type of scattering matrix is valid for[39] (i) scattering by an assembly of randomly oriented particles each having a plane of symmetry, (ii) scattering by an assembly containing particles and their mirror particles in equal numbers and with random orientations, (iii) Rayleigh scattering with or without depolarization effects. In this work, we ignore Stokes parameter V, and the corresponding elements p_4 and p_6 of the scattering phase matrix, because it is very small and there are no indications it is useful for aerosol retrieval.

In general, the forward model should simulate the Stokes parameters I,Q, and U for different spectral bands/pixels and viewing angles. For a given spectral measurement in band/pixel i the forward model should simulate

$$\bar{E}(\lambda_i) = \int_{\lambda_{\text{start}}}^{\lambda_{\text{end}}} d\lambda \, s_{\text{earth,i}}(\lambda) \, E(\lambda) \; \approx \sum_{l=1}^{N_{\Delta\lambda}} \Delta\lambda_l s_{\text{earth,i}}(\lambda_l) \, E(\lambda_l) \; , \tag{8.6}$$

where E represents either I, Q, or U, s_{earth} is the Instrumental Spectral Response Function (ISRF), λ_i is the wavelength assigned to spectral measurement i, and $N_{\Delta\lambda}$ is a suitable number of wavelength bins to discretize the domain of integration $[\lambda_{\text{start}}, \lambda_{\text{end}}]$.

If the dependence of E on λ is weak within the spectral range of the ISRF (e.g. at continuum wavelengths or for spectrally smooth absorbing trace gases), we can approximate Eq. (8.6) by

$$\bar{E}(\lambda_i) = E(\lambda_i) \; , \tag{8.7}$$

In any case (Eq. 8.6 or Eq. 8.7) we need a Radiative Transfer (RT) model to compute $I(\lambda)$, $Q(\lambda)$, and $U(\lambda)$ for a given model atmosphere.

9 Model Atmosphere and Optical Properties

For the forward model calculations, the model atmosphere is discretized in a number N_{RT} of homogeneous vertical layers, provided as input as a fixed number of altitude levels z_{lev} above the surface in the atmosphere. In the standard setup we use 24 vertical layers (2 km spacing between 0-20 km, 4 km spacing between 20-36 km, 1 layer above 36 km). First, we determine the pressure at these altitude levels by linear interpolation of the input pressure profile p_{met} as a function of altitude z_{met} from a meteorological data source (ECMWF, or CAMS), to the forward model levels $z_{\mathrm{lev}} + z_{\mathrm{surf}}$, where z_{surf} is the surface height from a Digital Elevation Map (DEM).

Each layer in the model atmosphere is characterized by the scattering optical thickness, absorption optical thickness, and scattering phase function, determined by molecules and aerosols in the atmosphere. The molecular (Rayleigh) scattering optical thickness for layer k and at wavelength λ is given by

$$\tau_{\mathrm{sca,mol},k}(\lambda) = \sigma_{\mathrm{ray}}(\lambda)\; N_{\mathrm{air,k}} \tag{9.8}$$

where the sub-column of dry air in layer k is denoted with

$$N_{\mathrm{air,k}} = \frac{\left(p_{\mathrm{lev,k+1}} - p_{\mathrm{lev,k}}\right)\; N_A}{M_{\mathrm{air}}\; g_k\; \left(1 + \frac{r_{\mathrm{water,k}}}{\rho_{\mathrm{rel}}}\right)}\,. \tag{9.9}$$

Here, $p_{\mathrm{lev,k+1}}$ and $p_{\mathrm{lev,k}}$ denote the air pressure at the lower and upper layer boundaries, respectively, N_A is Avogado's number, M is the molecular mass of air, g_k is the gravity constant in altitude layer k, and $\rho_{\mathrm{rel}} = 1.60855$ is the mass of air relative to the mass of water[40]. Further, σ_{ray} is the Rayleigh scattering cross section given by[41]

$$\sigma_{\mathrm{ray}}\left(\lambda\right) = A_{\mathrm{ray}}\; \lambda^{-(4+X)} \tag{9.10}$$

$$X = B_{\mathrm{ray}}\; \lambda + \frac{C_{\mathrm{ray}}}{\lambda} - D_{\mathrm{ray}} \tag{9.11}$$

with $A_{\mathrm{ray}} = 4.02 \cdot 10^{-28}$ cm^2, $B_{\mathrm{ray}} = 0.389\,\mu$m^{-1}, $C_{\mathrm{ray}} = 0.04926\,\mu$m, and $D_{\mathrm{ray}} = 0.3228$. The relevant elements of the Rayleigh scattering phase matrix (Eq. 8.5) are given by (e.g.[6])

$$p_{1,\mathrm{ray}}\left(\Theta\right) = \frac{3}{4}\left(1+\cos^2\Theta\right)\frac{1-\delta}{1+\delta/2} \tag{9.12}$$

$$p_{2,\mathrm{ray}}\left(\Theta\right) = \frac{3}{4}\left(1+\cos^2\Theta\right) \tag{9.13}$$

$$p_{3,\mathrm{ray}}\left(\Theta\right) = \frac{3}{2}\cos\Theta \tag{9.14}$$

$$p_{5,\mathrm{ray}}\left(\Theta\right) = -\frac{3}{4}\sin^2\Theta \tag{9.15}$$

where Θ is the scattering angle. The depolarization ratio δ varies with wavelength and accounts for the anisotropy of the air molecule[41]. δ varies between 0.03178 at 300 nm and 0.0274 at 1000 nm.

The molecular absorption optical depth is first calculated per absorber (j) and per discretized atmospheric layer (k). To account for the temperature and pressure dependence of the molecular absorptions within a model layer, we divide each model layer k in $N_{\mathrm{sub},k}$ sublayers, such that the pressure thickness ΔP_{is} of the sublayer $<$ 10 hPa. The molecular absorption optical depth for layer k and absorber j is given by:

$$\tau_{\mathrm{abs,k,j}}\left(\lambda\right) = \sum_{is=1}^{N_{\mathrm{sub,k}}} \sigma_j\left(p_{is}, T_{is}, \lambda\right)\; N_{gas,k,j}\; \frac{\Delta p_{is}}{\Delta p_k}. \tag{9.16}$$

with the absorption cross-section $\sigma_j(p_{is}, T_{is}, \lambda)$ of molecule j at wavelength λ, pressure p_{is} and temperature T_{is} at the center of sublayer is of atmospheric layer k, Δp_{is} is the pressure thickness of sub-layer is and Δp_k of model layer k. Further, $N_{gas,k,j}$ is the sub-column of absorber j in layer k, given by

$$N_{gas,k,j} = r_{gas,k,j}\; N_{\mathrm{air,k}} \tag{9.17}$$

where $r_{gas,k,j}$ is the dry air mixing ratio of gas j in layer k. The total molecular absorption optical depth for model layer k is then given by

$$\tau_{\mathrm{abs,mol},k}(\lambda) = \sum_j \tau_{\mathrm{abs,k,j}}\,(\lambda)\,. \tag{9.18}$$

The model atmosphere also includes aerosols. The aerosol scattering optical thickness $\tau_{\mathrm{sca,aer,k}}$ and absorption optical thickness $\tau_{\mathrm{abs,aer,k}}$ are calculated for each layer k

$$\tau_{\mathrm{sca,aer,k}} = \sum_{i=1}^{N_{\mathrm{modes}}} \sigma_{\mathrm{sca,aer,i}}\, N_{\mathrm{aer},k,i} \tag{9.19}$$

$$\tau_{\mathrm{abs,aer,k}} = \sum_{i=1}^{N_{\mathrm{modes}}} \sigma_{\mathrm{abs,aer,i}}\, N_{\mathrm{aer},k,i} \tag{9.20}$$

$$\tag{9.21}$$

where N_{modes} is the number of size modes that describe the aerosol size distribution (see below), $N_{\mathrm{aer},k,i}$ is the aerosol number sub-column of mode i in layer k and $\sigma_{\mathrm{sca,aer,i}}$ and $\sigma_{\mathrm{abs,aer,i}}$ are the aerosol scattering and absorption cross-sections for mode i, respectively. Here, we assume that $\sigma_{\mathrm{sca,aer,i}}$ and $\sigma_{\mathrm{abs,aer,i}}$ are altitude independent.

They are obtained by:

$$\sigma_{\mathrm{sca},aer,i} = \sum_{j=1}^{N_{\mathrm{bins}}} \left(f_{\mathrm{sph,i}}\, K_{\mathrm{sph,sca,j}}\,(m_i) + (1 - f_{\mathrm{sph,i}})\; K_{\mathrm{el,sca,j}}\,(m_i)\right) \;\; r_j\; n_{\mathrm{aer,i}}\,(r_j)\;\; v\,(r_j) \tag{9.22}$$

$$K_{\mathrm{sph/el,sca,j}}\,(m_i) = \int\limits_{\Delta \ln r_j} \frac{C_{\mathrm{sph/el,sca}}\,(r, m_i)}{v\,(r)} d\ln r\;, \tag{9.23}$$

where $n_{\mathrm{aer,i}}$ is the aerosol number size distribution of mode i, v denotes particle volume and $C_{\mathrm{sph,sca}}$ and $C_{\mathrm{el,sca}}$ are the scattering cross sections of a spherical and ellipsoidal particle, respectively, of radius r and mode refractive index m_i, where for ellipsoidal particles the aspect ration distribution of Dubovik et al.[42] has been used. The relative contribution of spherical and ellipsoidal particles is determined by the parameter $f_{\mathrm{sph,i}}$. The kernels $K_{\mathrm{sph/el,scat,j}}$ are pre-calculated for N_{bins} size bins with representative radius r_j, $j = 1, \cdots, N_{\mathrm{bins}}$ according to Dubovik et al.[42] and are stored in a lookup table as a function of aerosol size parameter $x = 2\pi r/\lambda$, and the real and imaginary part of the complex refractive index m. The values for the actual aerosol real- and imaginary refractive index are obtained by linear interpolation from the tabulated values. Similar expressions hold for the absorption cross-section and the elements of the aerosol scattering phase matrix $\mathbf{P}_{\mathrm{aer}}$.

In addition to the aerosol modes, the model atmosphere also contains a cirrus cloud with cirrus scattering optical thickness $\tau_{\mathrm{sca,cir,k}}$ and absorption optical thickness $\tau_{\mathrm{abs,cir,k}}$ in layer k

$$\tau_{\mathrm{sca,cir,k}} = \sigma_{\mathrm{sca,cir}}\, N_{\mathrm{cir},k} \tag{9.24}$$

$$\tau_{\mathrm{abs,cir,k}} = \sigma_{\mathrm{abs,cir}}\, N_{\mathrm{cir},k} \tag{9.25}$$

$$\tag{9.26}$$

where $N_{\mathrm{cir},k}$ is the cirrus number sub-column in layer k and $\sigma_{\mathrm{sca,cir}}$ and $\sigma_{\mathrm{abs,cir}}$ are the cirrus scattering and absorption cross-sections. $\sigma_{\mathrm{sca,cir}}$ and $\sigma_{\mathrm{abs,cir}}$, as well as the elements of the cirrus scattering phase matrix $\mathbf{P}_{\mathrm{cir}}$, are obtained from a LUT, according to van Diedenhoven et al.[43], which stores the cirrus optical properties computed using the geometrics optics approximation for a cirrus ice crystal that are described as single hexagonal crystals with varying size ($r_{\mathrm{eff,cir}}$), aspect ratio (α_{cir}), and surface roughness (δ_{cir}). The optical properties ($\sigma_{\mathrm{sca,cir}}$, $\sigma_{\mathrm{abs,cir}}$, $\mathbf{P}_{\mathrm{cir}}$) for a given combination of $r_{\mathrm{eff,cir}}$, α_{cir}, and δ_{cir} are then obtained by linear interpolation from the tabulated values.

Finally, the total optical properties per layer in the model atmosphere are obtained by combining the contribution of molecules, aerosols, and cirrus:

$$\tau_{\mathrm{abs,k}} = \tau_{\mathrm{abs,mol,k}} + \tau_{\mathrm{abs,aer,k}}\tau_{\mathrm{abs,cir,k}} \tag{9.27}$$

$$\tau_{\mathrm{sca,k}} = \tau_{\mathrm{sca,mol,k}} + \tau_{\mathrm{sca,aer,k}} + \tau_{\mathrm{sca,cir,k}} \tag{9.28}$$

$$\mathbf{P}_k(\Theta) = \frac{\tau_{\mathrm{sca,mol,k}}\mathbf{P}_{\mathrm{mol}}(\Theta) + \tau_{\mathrm{sca,aer,k}}\mathbf{P}_{\mathrm{aer}} + \tau_{\mathrm{sca,cir,k}}\mathbf{P}_{\mathrm{cir}}\,(\Theta)}{\tau_{\mathrm{sca,k}}}\,. \tag{9.29}$$

For radiative transfer calculations, the scattering matrix $\mathbf{P}_k(\Theta)$ has to be transformed from the scattering plane to the local meridian plane

$$\mathbf{Z}_k(\lambda, \mu_{\mathrm{in}}, \mu_{\mathrm{out}}, \Delta\varphi) = \mathbf{L}_2\ \mathbf{P}_k(\Theta)\ \mathbf{L}_1 \tag{9.30}$$

with

$$\mathbf{L}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2i_2 & -\sin 2i_2 & 0 \\ 0 & \sin 2i_2 & \cos 2i_2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{9.31}$$

and

$$\mathbf{L}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2i_1 & -\sin 2i_1 & 0 \\ 0 & \sin 2i_1 & \cos 2i_1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{9.32}$$

The rotation angles i_1 and i_2 are given by

$$\cos i_1 = \frac{\mu_{\mathrm{out}}(1-\mu_{\mathrm{in}}^2)^{1/2} - \mu_{\mathrm{in}}(1-\mu_{\mathrm{out}}^2)^{1/2}\cos\Delta\varphi}{(1-\cos^2\Theta)^{1/2}} \tag{9.33}$$

$$\cos i_1 = \frac{\mu_{\mathrm{in}}(1-\mu_{\mathrm{out}}^2)^{1/2} - \mu_{\mathrm{out}}(1-\mu_{\mathrm{in}}^2)^{1/2}\cos\Delta\varphi}{(1-\cos^2\Theta)^{1/2}} \tag{9.34}$$

$$\tag{9.35}$$

where μ_{in} and μ_{out} are respectively the cosines of incoming and outgoing angles θ_{in} and θ_{out}, and $\Delta\varphi$ is the relative azimuth angle.

For multiple scattering calculations, the scattering phase matrix is expanded in a Fourier series

$$\begin{split}\mathbf{Z}_k(\mu_{\mathrm{in}}, \mu_{\mathrm{out}}, \Delta\varphi) = \frac{1}{2}\sum_{m=0}^{\infty}(2-\delta_{m0})\big[&\mathbf{B}^{+m}(\Delta\varphi)\mathbf{Z}_k^m(\mu_{\mathrm{in}}, \mu_{\mathrm{out}})(\mathbf{E}+\mathbf{\Lambda}) + \\ &\mathbf{B}^{-m}(\Delta\varphi)\mathbf{Z}_k^m(\mu_{\mathrm{in}}, \mu_{\mathrm{out}})(\mathbf{E}-\mathbf{\Lambda})\big],\end{split} \tag{9.36}$$

where

$$\mathbf{\Lambda} = \mathrm{diag}\,[1, 1, -1, -1]\,. \tag{9.37}$$

and

$$\mathbf{B}^{+m}(\Delta\varphi) = \text{diag}[\cos m\Delta\varphi, \cos m\Delta\varphi, \sin m\Delta\varphi, \sin m\Delta\varphi] \tag{9.38}$$
$$\mathbf{B}^{-m}(\Delta\varphi) = \text{diag}[-\sin m\Delta\varphi, -\sin m\Delta\varphi, \cos m\Delta\varphi, \cos m\Delta\varphi]. \tag{9.39}$$

The m-th Fourier coefficient of the phase matrix can be calculated by

$$\mathbf{Z}^m(\mu_{\text{in}}, \mu_{\text{out}}) = (-1)^m \sum_{l=m}^{L} \mathbf{P}^l_m(-\mu_{\text{out}})\, \mathbf{S}^l(z)\, \mathbf{P}^l_m(-\mu_{\text{in}}), \tag{9.40}$$

where L is a suitable truncation index[44] and $\mathbf{P}^l_m$ is the generalized spherical function matrix given by

$$\mathbf{P}^m_l(\mu) = \begin{pmatrix} P^l_{m0}(\mu) & 0 & 0 & 0 \\ 0 & P^l_{m+}(\mu) & P^l_{m-}(\mu) & 0 \\ 0 & P^l_{m-}(\mu) & P^l_{m+}(\mu) & 0 \\ 0 & 0 & 0 & P^l_{m0}(\mu) \end{pmatrix} \tag{9.41}$$

where

$$P^l_{m\pm} = \frac{1}{2}\left(P^l_{m,-2} \pm P^l_{m,2}\right) \tag{9.42}$$

and $P^l_{mn}(\mu)$ are the generalized spherical functions[45], which were introduced in atmospheric radiative transfer by[46]. $\mathbf{S}^l$ is the expansion coefficient matrix having the form

$$\mathbf{S}^l = \begin{pmatrix} \beta^l_1 & \beta^l_5 & 0 & 0 \\ \beta^l_5 & \beta^l_2 & 0 & 0 \\ 0 & 0 & \beta^l_3 & \beta^l_6 \\ 0 & 0 & -\beta^l_6 & \beta^l_4 \end{pmatrix}, \tag{9.43}$$

where the expansion coefficients follow from the scattering phase matrix $\mathbf{P}$ in Eq. (8.5) (see e.g.[47]):

$$\beta^l_1 = \frac{2l+1}{2} \int_{-1}^{1} P^l_{0,0}(\cos\theta) p_1(\theta)\, d(\cos\theta), \tag{9.44}$$
$$\beta^l_2 + \beta^l_3 = -\frac{2l+1}{2}\sqrt{\frac{(l-2)!}{(l+2)!}} \int_{-1}^{1} P^l_{2,2}(\cos\theta)(p_2(\theta)+p_3(\theta)) d(\cos\theta), \tag{9.45}$$
$$\beta^l_2 - \beta^l_3 = -\frac{2l+1}{2}\sqrt{\frac{(l-2)!}{(l+2)!}} \int_{-1}^{1} P^l_{2,-2}(\cos\theta)(p_2(\theta)+p_3(\theta)) d(\cos\theta), \tag{9.46}$$
$$\beta^l_4 = \frac{2l+1}{2} \int_{-1}^{1} P^l_{0,0}(\cos\theta) p_4(\theta)\, d(\cos\theta), \tag{9.47}$$
$$\beta^l_5 = \frac{2l+1}{2} \int_{-1}^{1} P^l_{0,2}(\cos\theta) p_5(\theta)\, d(\cos\theta), \tag{9.48}$$
$$\beta^l_6 = \frac{2l+1}{2} \int_{-1}^{1} P^l_{0,2}(\cos\theta) p_6(\theta)\, d(\cos\theta). \tag{9.49}$$

The vertically integrated absorption and scattering optical thickness τ_{abs} and τ_{sca} are the respective sums over all N_{RT} layers. Other derived variables are the total (extinction) optical thickness $\tau_k = \tau_{\text{sca,k}} + \tau_{\text{abs,k}}$ and $\tau = \tau_{\text{sca}} + \tau_{\text{abs}}$, and the single scattering albedo $\omega_k = \tau_{\text{sca,k}}/\tau_k$.

From the equations above, it is clear that in order to compute the optical properties of the model atmosphere, we need in the different layers of the model atmosphere the aerosol size distribution $n_{\mathrm{aer}}\left(r_i\right)$, aerosol number sub-column N_k, the fraction of spherical aerosol f_{sphere}, as well as the sub-column $N_{gas,k,j}$ of the j different trace gases. Additionally, we need the atmospheric pressure and temperature profile.

Concerning the aerosol parameters, the aerosol size distribution $n_{\mathrm{aer,i}}(r)$ for mode i is either describe by a log-normal function or power law. The log-normal size distribution is written as:

$$n_{\mathrm{aer}}(r) = \frac{1}{\sqrt{2\pi}\,\sigma_g\, r} \exp\left[-(\ln r - \ln r_g)^2/(2\sigma_g^2)\right], \tag{9.50}$$

where r is radius (or radius of a volume equivalent sphere), r_g is the median radius and σ_g is the standard deviation. Instead of r_g and σ_g we use the effective radius r_{eff} and effective variance v_{eff} because they are less dependent on the actual shape of the size distribution[6]. r_{eff} and v_{eff} are related to r_g and σ_g by:

$$\sigma_g^2 = \ln\left(1 + v_{eff}\right) \tag{9.51}$$

$$r_g = r_{eff}/(1 + v_{eff})^{5/2}. \tag{9.52}$$

The power-law function is given by

$$n_{\mathrm{aer}}\left(r\right) = \begin{cases} C & \text{for } r \le r_1 \\ C\left(\frac{r}{r_1}\right)^{-p} & \text{for } r_1 < r \le r_2 \\ 0 & \text{for } r > r_2 \end{cases} \tag{9.53}$$

where C is a scaling constant and the exponent p specifies the decrease of n_{aer} with particle size. The cut-offs are r_1 =0.1 µm, r_2 = 10 µm and the constant C is determined from normalization of the size distribution.

The number sub-column $N_{\mathrm{aer,k,i}}$ (see Eqs. 9.19 and 9.20) for layer k and mode i is parameterized through the total column of aerosol particles, $N_{\mathrm{aer,i}}$ for mode i, and a normalized Gaussian altitude distribution with center height z_{aer} and width w_0. Hence, for model layer k with a layer height z_k, and thickness Δz_k:

$$N_{\mathrm{aer,k,i}} = N_{\mathrm{aer,i}} B \exp\left[-\frac{\ln 2\left(z_k - z_{\mathrm{aer}}\right)^2}{w_0^2}\right] \Delta z_k, \tag{9.54}$$

where $N_{\mathrm{aer,i}}$ is the aerosol column number for mode i and B is the normalization constant of the Gaussian. A similar expression holds for the number sub-column $N_{\mathrm{cir,k}}$ of cirrus particles.

The spectrally dependent refractive index $m(\lambda)$ per mode is parameterized by

$$m(\lambda) = \sum_{k=1}^{n_\alpha} \alpha_k\, m^k(\lambda) \tag{9.55}$$

where $m^k(\lambda)$ are prescribed functions of wavelength, for which we use standard refractive index spectra for different aerosol components, i.e. Dust (DU)[48], Inorganic/Sulphate (INORG) and Black Carbon (BC)[49], and Organic Carbon (OC)[50].

The aerosol parameters to be included in the retrieval state vector are specified by the user in a retrieval initialization file. Here, we consider 4 different setups:

1. **Parametric 2-mode retrievals** where the size distribution is described by 2 log-normal modes (i.e. $N_{\mathrm{modes}} = 2$). For both modes the state vector includes r_{eff}, v_{eff}, N_{aer} and f_{sph}. For the mode with smallest particles, referred to as fine mode, the refractive index coefficients α_k are included in the state vector. These correspond to the standard refractive index spectra INORG, BC, OC. For the mode with largest particles, referred to as coarse mode, the α_k corresponding to INORG and DU. Further, one value for z_{aer} is included which is assumed to be the same for both modes. w_0 is fixed at 2000 m.

Table 9.2: Definition of the effective radius (r_{eff}) and the effective variance (v_{eff}) for the multi-mode (M-1 to M-10) retrievals.

	M-1	M-2	M-3	M-4	M-5	M-6	M-7	M8	M-9	M-10
r_{eff} (μm)	0.070	0.094	0.130	0.163	0.220	0.282	0.882	1.2	1.759	3.0
v_{eff}	0.130	0.130	0.130	0.130	0.130	0.130	0.284	1.0	1.718	1.718

2. **Parametric 3-mode retrievals** where the size distribution is described by 3 log-normal modes (i.e. $N_{\text{modes}} = 3$), with one fine mode and 2 coarse modes (soluble and insoluble). For the fine mode the state vector includes r_{eff}, v_{eff}, N_{aer} and f_{sph} and the refractive index coefficients α_k that correspond to the standard refractive index spectra INORG, BC, OC. The coarse insouble mode consists of non-spherical dust. for this mode the state vector includes r_{eff}, N_{aer}, and a coefficient for the imaginary part of the DU refractive index. The fixed parameters are $f_{\text{sph}} = 0$, $v_{eff} = 0.6$, $\alpha_k = 1$ for the DU real part refractive index. One value for z_{aer} is included which is assumed to be the same for modes 1 and 2. w_0 is fixed at 2000 m. The 3rd mode is a coarse soluble mode. For this mode the state vector includes r_{eff}, N_{aer}, and coefficient α_k of the INORG refractive index spectrum. The fixed parameters are $f_{\text{sph}} = 1$, $v_{eff} = 0.6$, and z_{aer}=0.5 km.
3. **Multi-mode retrievals** where the size distribution is described by a combination of $N_{obs} \leq 10$ log-normal modes. For each mode N_{aer} is included in the state vector while r_{eff}, v_{eff} are fixed to (a selection of) the values in Table 9.2. Here, mode 1-6 are considered as fine modes and 7-10 as coarse modes. The state vector includes one value for f_{sph} for all fine modes and one value for all coarse modes and for the fine mode refractive index (i.e. assumed the same for all fine modes) it includes the coefficients α_k that correspond to the standard refractive index spectra INORG, BC, OC, and for the coarse mode refractive index (i.e. assumed the same for all coarse modes) the coefficients corresponding to INORG and DU. Further, one value for z_{aer} is included which is assumed to be the same for all modes, i.e. all modes have the same vertical distribution. w_0 is fixed at 2000 m.
4. **Power-law retrievals** where the aerosol size distribution is described by a single power-law mode. Here, N_{aer} and power p are included in the state vector. The refractive index is fixed by $1.4 - 0.003i$ for all spectral bands. Further, z_{aer} is included in the state vector and w_0 is fixed at 2000 m.

Setup 1 and 2 are used for retrievals where MAP measurements are available, i.e. MAP-only or MAP+spectrometer retrievals, while setup 3 is being used when only spectrometer measurements are available. Note that this setup is very similar to the RemoTeC setup[51,52]. For all aerosol setups there is the option to add a mode cirrus particles with N_{cir}, $r_{\text{eff,cir}}$, α_{cir}, and δ_{cir}, and z_{cir} as state vector elements and w_0 fixed at 2000 m.

The sub-columns $N_{gas,k,j}$ of the j different trace gases in layer k of the model atmosphere are given by

$$N_{gas,k,j} = c_j \, N_{std,k,j}, \tag{9.56}$$

where $N_{std,k,j}$ is the sub-column for trace gas j corresponding to an *a priori* vertical profile and c_j is the corresponding scaling factor (see below). The O_2 vertical profile follows from the pressure profile. For MAP-only retrievals we fix the scaling factors c_j to 1.0, i.e. their atmospheric abundance is fixed to the a priori value through out the retrieval, while for MAP+spectrometer and spectrometer-only retrievals we include the scaling factors c_j corresponding to CO_2, CH_4, and H_2O in the retrieval state vector.

10 Land Surface Reflection Matrix

For retrievals over land, the surface reflection matrix is written as

$$\mathbf{R}_s(\lambda, \theta_{\text{in}}, \theta_{\text{out}}, \Delta\varphi) = r_{11}(\lambda, \theta_{\text{in}}, \theta_{\text{out}}, \Delta\varphi)\, \mathbf{D} + \mathbf{R}_{\text{pol}}, \tag{10.57}$$

where $\mathbf{D}$ is the null matrix except $\mathbf{D}_{11}$ = 1. For the Bi-Directional Reflection function (BDRF) $r_{11}(\lambda, \vartheta_{\rm in}, \vartheta_{\rm out}, \Delta\varphi)$ we have 2 options implemented. The first option is the Ross-Li BDRF model given by:

$$r_{11}(\lambda, \theta_{\rm in}, \theta_{\rm out}, \Delta\varphi) = A(\lambda)\ \ (1 + k_{\rm geo} f_{\rm geo}(\theta_{\rm in}, \theta_{\rm out}, \Delta\varphi) + k_{\rm vol} f_{\rm vol}(\theta_{\rm in}, \theta_{\rm out}, \Delta\varphi)) \tag{10.58}$$

where $f_{\rm geo}$ and $f_{\rm vol}$ are respectively the geometric (Li-Sparse) and volumetric (Ross-Thick) kernel[53] and references therein. They are given by

$$f_{\rm vol}(\theta_{\rm in}, \theta_{\rm out}, \Delta\varphi) = \frac{(\pi/2-\gamma)\cos\gamma + \sin\gamma}{\mu_{\rm in} + \mu_{\rm out}} - \frac{\pi}{4} \tag{10.59}$$

$$f_{\rm geo}(\theta_{\rm in}, \theta_{\rm out}, \Delta\varphi) = O(\theta'_{\rm in}, \theta'_{\rm out}, \Delta\varphi) - \sec\theta'_{\rm out} - \sec\theta'_{\rm in} + \frac{1}{2}(1-\cos\Theta')\ \sec\theta'_{\rm out}\ \sec\theta'_{\rm in} \tag{10.60}$$

$$O = \frac{1}{\pi}(t - \sin t\cos t)(\sec\theta'_{\rm out} + \theta'_{\rm in}) \tag{10.61}$$

$$\cos t = \frac{h}{b}\frac{\sqrt{D^2 + (\tan\theta'_{\rm in}\ \tan\theta'_{\rm out}\ \sin\Delta\phi)^2}}{\sec\theta'_{\rm out} + \sec\theta'_{\rm in}} \tag{10.62}$$

$$D = \sqrt{\tan^2\theta'_{\rm in} + \tan^2\theta'_{\rm out} - 2\ \tan\theta'_{\rm in}\ \tan\theta'_{\rm out}\ \cos\Delta\phi} \tag{10.63}$$

$$\cos\Theta' = -\cos\theta'_{\rm in}\ \cos\theta'_{\rm out} - \sin\theta'_{\rm in}\ \sin\theta'_{\rm out}\cos\Delta\phi \tag{10.64}$$

$$\theta'_{\rm out} = \tan^{-1}\left(\frac{b}{r}\ \tan|\theta'_{\rm out}|\right), \tag{10.65}$$

$$\theta'_{\rm in} = \tan^{-1}\left(\frac{b}{r}\ \tan|\theta'_{\rm in}|\right). \tag{10.66}$$

Numerically, the absolute value of the RHS of Eq. (10.62) can exceed 1. In this case we set $\cos t = 0$. The volumetric kernel represents the scattering within a dense vegetation canopy, and is based on a radiative transfer approximation of single scattering due to small, uniformly distributed and non-absorbing leaves. The angular behavior of this kernel is to have a minimum near the backscatter direction and bright limbs[54]. The geometric kernel represents surfaces with larger gaps between objects, and thus accounts for self shadowing. The angular behavior of this kernel is therefore to have a maximum at backscattering where there are no shadows. $f_{\rm geo}$ is based on the work of[53,55], but is used in the reciprocal form given in[56], for the case that the ratio of the height of the tree at the center of the crown to the vertical crown radius, $h/b = 2$, and the ratio of the vertical crown radius to the horizontal crown radius is (spherical, or compact crowns) $b/r = 1$.

The second option is the Rahman-Pinkty-Verstraete (RPV) model[13,57,58]:

$$r_{11}(\lambda, \theta_{\rm in}, \theta_{\rm out}, \Delta\varphi) = A(\lambda)\ \left(\frac{(\mu_{\rm in}\ \mu_{\rm out})^{k-1}}{(\mu_{\rm in} + \mu_{\rm out})^{1-k}}\ F(g,\Theta)[1 + R(G)]\right) \tag{10.67}$$

$\mu_{\rm in}$ and $\mu_{\rm out}$ are respectively the cosines of incoming and outgoing angles $\theta_{\rm in}$ and $\theta_{\rm out}$. g is the asymmetry parameter of the Henyey-Greenstein phase function $F(g,\Theta)$. Θ is the scattering angle. $1 + R(G)$ is an approximation of the hot spot effect[57], where

$$G = \sqrt{\tan^2\theta_0 + \tan^2\theta_v - 2\tan\theta_{\rm in}\tan|\theta_{\rm out}|\cos\Delta\varphi} \tag{10.68}$$

and $R(G) = \frac{1-A(\lambda)}{1+G}$.

In Eq. (10.57) $\mathbf{R}_{\rm pol}$ is given by[59]

$$\mathbf{R}_{\rm pol}(\theta_{\rm in}, \theta_{\rm out}, \phi_v - \phi_0) = B_{\rm pol}\ \left(\frac{\exp\left(-\tan(\frac{\pi-\Theta}{2})\right)\ \exp\left(-\nu\right)\ \mathbf{F}_p(m,\Theta)}{4(\mu_{\rm in} + \mu_{\rm out})}\right). \tag{10.69}$$

Here, $B_{\rm pol}$ is a scaling parameter (band-independent). $\mathbf{F}_p(m,\Theta)$ is the the Fresnel scattering matrix with refractive index $m = 1.5$. We use ν = 0.1[58].

The parameters included in the state vector are $f_{\rm geo}$ and $f_{\rm vol}$ when using the Ross-Li model and and k, g when using the RPV model. Further, we include $B_{\rm pol}$, assumed spectrally neutral, and for $A(\lambda)$ we retrieve one value for each MAP band. For spectrometer windows we describe the spectral dependence $A(\lambda)$ within a spectrometer window as a polynomial

$$A(\lambda) = \sum_{i=0}^{I} a_i \,(\lambda - \lambda_0)^i, \tag{10.70}$$

where λ_0 is a reference wavelength chosen as the first wavelength of the spectral window. We include the polynomial coefficients a_i in the state vector.

11 Ocean Reflection Matrix

For retrievals over ocean, we describe the ocean reflection matrix as

$$\mathbf{R}_s(\lambda, \theta_{\text{in}}, \theta_{\text{out}}, \Delta\varphi) = \mathbf{R}_{frn}(\theta_{\text{in}}, \theta_{\text{out}}, \Delta\varphi) + \mathbf{R}_{ul}(\lambda, \theta_{\text{in}}, \theta_{\text{out}}, \Delta\varphi) + A(\lambda)\,\mathbf{D} \tag{11.71}$$

where $\mathbf{R}_{frn}$ is the contribution of the ocean surface, which is described by Fresnel reflection on a rough ocean surface, depending on the wind speed- and direction to provide a Gaussian distribution of surface slopes[60]. $\mathbf{R}_{ul}$ is the ocean body (underlight) contribution. For the ocean body, we need a bio-optical model to compute optical properties of the ocean from bio-physical ocean parameters. We used the bio-optical model of Chowdhary et al.[61] for case-2 waters (open ocean) that has the chlorophyll-a concentration x_{chl} as the only bio-physical ocean parameter to compute the ocean optical properties (single scattering albedo, phase matrix). Using the hydrosol model of Chowdhary et al.[61], the ocean is described as a mixture of sea-water and a particulate component. The scattering and absorption coefficients of sea-water are taken from Smith and Baker[62], while the optical properties of the particulate components were calculated using detritus–plankton (D–P) mixtures. The particulates were assumed to be spherical, so the scattering phase matrix could be obtained using Mie calculations. The relative contribution of detritus and plankton are parameterized by X_{chl}. Here it should be noted that the underlight contribution is insensitive to the optical depth of ocean when the ocean optical thickness is larger than 10. In this study, we set the ocean optical depth to 20 and assumed a black ocean bottom surface. This ocean surface/body system was being solved by a vector radiative transfer model[63,64]. However, this model is computationally expensive because of the large ocean optical thickness. As an alternative, a Neural Network (NN) has been designed to simulate (Fourier coefficients of) the ocean body contribution to the reflection matrix just above the ocean surface, with as input the oceanic chlorophyll-a concentration. Finally, $A(\lambda)$ in Eq. (11.71) is a wavelength dependent Lambertian albedo term that accounts for oceanic foam but may also correct for errors in $\mathbf{R}_{ul}$. W_s, X_{chl}, and $A(\lambda)$ for the different spectral bands are included in the state vector.

12 Radiative transfer model

Having defined the relation between atmospheric/surface properties and optical properties for the different layers of the model atmosphere, the next step is to solve the RT equation for the given optical properties. The radiative transfer model comprises two contributions, single and multiple scattering contribution of the backscattered radiances I_{ss} and I_{ms}, respectively:

$$\mathbf{I} = \mathbf{I}_{\text{ss}} + \mathbf{I}_{\text{ms}} \tag{12.72}$$

Here, each component is simulated separately.

Before computing $\mathbf{I}_{\text{ss}}$ and $\mathbf{I}_{\text{ms}}$, we first apply the delta-M correction[65]:

$$\mathbf{Z}^*_{\text{ij,k}} = \mathbf{Z}_{\text{ij,k}}/(1 - f_k) \tag{12.73}$$

$$\mathbf{S}^{*l}_{\text{ij,k}} = \left(\mathbf{S}^{l}_{\text{ij,k}} - \delta_{\text{ij}} f_k \,(2l+1)\right)/(1 - f_k) \tag{12.74}$$

$$\tau^*_{\text{sca,k}} = \tau_{\text{sca,k}}\,(1 - f_k) \tag{12.75}$$

$$\tau^*_{\text{k}} = \tau^*_{\text{sca,k}} + \tau_{\text{abs,k}} \tag{12.76}$$

$$\omega^*_{\text{k}} = \tau^*_{\text{sca,k}}/\tau^*_{\text{k}} \tag{12.77}$$

where i and j denote the matrix element of $\mathbf{Z}$ or $\mathbf{S}$ in Eq. (9.30) and 9.43, respectively and

$$f_k = \mathbf{S}_{11,\mathrm{k}}^{N_{str}} / (2N_{str} + 1) \tag{12.78}$$

with N_{str} being the number of streams used in the multiple scattering calculations.

For a given wavelength, the computation of I_{ss} is straight forward:

$$\mathbf{I}_{\mathrm{ss}} = \frac{\tilde{\mu}}{4\pi\mu_v} \sum_{k=1}^{N_{\mathrm{RT}}} \left\{ \omega_k^* \left[1 - e^{-\tau_k^*/\tilde{\mu}} \right] e^{-\tau_1^{*k-1}/\tilde{\mu}} \mathbf{Z}_k^* \mathbf{F}_0 \right\} + \frac{\mu_0}{\pi}\, e^{\tau^*/\tilde{\mu}}\, \mathbf{R}_s \mathbf{F}_0 \tag{12.79}$$

where μ_0 is the cosine of the solar zenith angle, μ_v is the cosine of the viewing zenith angle, $\frac{1}{\tilde{\mu}} = \frac{1}{\mu_0} + \frac{1}{\mu_v}$. R_s is the surface reflection matrix of Eq. (10.57) and

$$\tau_{k_1}^{*k_2} = \sum_{k=k_1}^{k_2} \tau_k^* \; . \tag{12.80}$$

The incoming solar irradiance vector $\mathbf{F}_0 = [F_0, 0, 0]^T$, F_0 being the solar flux per unit area perpendicular to the direction of the solar beam.

The derivatives of the i-th element (Stokes index) of $\mathbf{I}_{\mathrm{ss}}$ with respect to $\tau_{\mathrm{sca,k}}^*$, $\tau_{\mathrm{abs,k}}$, $\omega^*{}_k$, the ij-th element $z_{k,ij}$ of $\mathbf{Z}_k^*$ and the elements $r_{s,ij}$ of $\mathbf{R}_s$ can be calculated in an analytical manner,

$$\begin{aligned}
\frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial \tau_k^*} &= \frac{1}{4\pi\mu_v} \left\{ \omega_k^* e^{-\tau_k^*/\tilde{\mu}} e^{-\tau_1^{*k-1}/\tilde{\mu}} \mathbf{Z}_k^* \mathbf{F}_0 - \sum_{k'=k+1}^{N_{\mathrm{RT}}} \omega_{k'}^* \left[1 - e^{-\tau_{k'}^*/\tilde{\mu}} \right] e^{-\tau_1^{*k'-1}/\tilde{\mu}} \mathbf{Z}_{k'}^* \mathbf{F}_0 \right\} \\
&\quad - \frac{\mu_0}{\pi\tilde{\mu}}\, e^{-\tau^*/\tilde{\mu}}\, \mathbf{R}_s \mathbf{F}_0 && (12.81) \\
\frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial z_{k,ij}} &= \frac{F_0 \tilde{\mu}}{4\pi\mu_v} \omega_k^* \left[1 - e^{-\tau_k^*/\tilde{\mu}} \right] e^{-\tau_1^{*k-1}/\tilde{\mu}} \delta_{j1} && (12.82) \\
\frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial \omega_k^*} &= \frac{\tilde{\mu}}{4\pi\mu_v} \left[1 - e^{-\tau_k^*/\tilde{\mu}} \right] e^{-\tau_1^{*k-1}/\tilde{\mu}} \mathbf{Z}_k^* \mathbf{F}_0 && (12.83) \\
\frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial r_{s,ij}} &= \frac{\mu_0 F_0}{\pi}\, e^{-\tau^*/\tilde{\mu}}\, \delta_{j1} && (12.84)
\end{aligned}$$

where δ_{j1} is the Kronecker delta, and with the deduced derivatives

$$\begin{aligned}
\frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial \tau_{\mathrm{sca,k}}^*} &= \frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial \tau_k^*} + \frac{1 - \omega_k^*}{\tau_k^*} \frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial \omega_k^*} && (12.85) \\
\frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial \tau_{\mathrm{abs,k}}^*} &= \frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial \tau_k^*} - \frac{\omega_k^*}{\tau_k^*} \frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial \omega_k^*} \; . && (12.86)
\end{aligned}$$

The computation of the multiply scattered radiation involves the solution of the plane-parallel radiative transfer equation. For this purpose, we use the LINTRAN V2.0 model[66], which provides the TOA radiance as well as its derivatives with respect to $\tau_{\mathrm{abs,k}}$, $\tau_{\mathrm{sca,k}}$ β_k^l(Eq. 9.43), for each layer k of the model atmosphere, as well as surface reflectance parameters, using the forward adjoint perturbation theory. LINTRAN is described in more detail in Appendix 91.

Based on derivatives of the radiance with respect to the optical parameters τ_{abs}, τ_{sca}, and P or its expansion coefficients β_k^l, the derivatives with respect to a physical parameter x (i.e. an element of the state vector) can be calculated in a straightforward manner by

$$\begin{aligned}
\frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial x} &= \sum_{k=1}^{N_{RT}} \left[\frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial \tau_{\mathrm{sca,k}}} \frac{\partial \tau_{\mathrm{sca,k}}}{\partial x} + \frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial \tau_{\mathrm{abs,k}}} \frac{\partial \tau_{\mathrm{abs,k}}}{\partial x} + \frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial P_k} \frac{\partial P_k}{\partial x} \right] && (12.87) \\
\frac{\partial \mathbf{I}_{\mathrm{ms}}}{\partial x} &= \sum_{k=1}^{N_{RT}} \left[\frac{\partial \mathbf{I}_{\mathrm{ms}}}{\partial \tau_{\mathrm{sca,k}}} \frac{\partial \tau_{\mathrm{sca,k}}}{\partial x} + \frac{\partial \mathbf{I}_{\mathrm{ms}}}{\partial \tau_{\mathrm{abs,k}}} \frac{\partial \tau_{\mathrm{abs,k}}}{\partial x} + \sum_{l=0}^{M} \frac{\partial \mathbf{I}_{\mathrm{ss}}}{\partial \beta_k^l} \frac{\partial \beta_k^l}{\partial x} \right] \; , && (12.88)
\end{aligned}$$

with

$$\frac{\partial \tau_{\mathrm{sca,k}}}{\partial x} = \frac{\partial \tau^*_{\mathrm{sca,k}}}{\partial x} \frac{\partial f}{\partial \tau^*_{\mathrm{sca,k}}} \tag{12.89}$$

$$\frac{\partial P_{\mathrm{k}}}{\partial x} = \frac{\partial P^*_{\mathrm{k}}}{\partial x} \frac{\partial f}{\partial P^*_{\mathrm{k}}} \tag{12.90}$$

13 Linear-k method for fine structure absorption bands

In order to avoid time consuming multiple scattering calculations on the line-by-line grid in fine-structure absorption bands, we aim at reducing the number of spectral calculations, following the linear-k approach of Hasekamp and Butz, 2008[67]. For this purpose, we consider the intensity $\mathbf{I}_{\mathrm{ms}}$ as a function of total absorption optical thickness τ_{abs} and its normalized vertical distribution $\mathbf{n}$

$$\mathbf{I}_{\mathrm{ms}}(\lambda) = \mathbf{I}_{\mathrm{ms}}(\tau_{\mathrm{abs}}(\lambda), \mathbf{n}(\lambda)) \ . \tag{13.91}$$

For a vertically homogenous atmosphere, analogously one can consider the dependence on the absorption coefficient k of the atmosphere and so explains the method name and its similarity with the k-distribution approach and spectral mapping methods.

We assume that the atmospheric scattering properties and surface reflection properties are constant over the spectral range under consideration. Element n_k of the vector $\mathbf{n}$ represents the relative contribution of the absorption optical thickness of altitude layer k of the model atmosphere to the total absorption optical thickness of the atmosphere, such that

$$\tau_{\mathrm{abs,k}}(\lambda) = n_k(\lambda)\, \tau_{\mathrm{abs}}(\lambda) \ . \tag{13.92}$$

For a vertically homogeneous atmosphere, $\mathbf{I}_{\mathrm{ms}}$ becomes a smooth function of absorption optical thickness. So only a limited number of calculations at preselected values of τ_{abs} are needed for interpolation (see e.g.[68]). Finally, the radiances can be mapped back into wavelength space. For a non-homogeneous atmosphere, one has to assume that the vertical distribution $\mathbf{n}(z)$ of τ_{abs} is independent of wavelength in the spectral interval under consideration, which is the underlying assumption of the correlated k method[69]. For the simulation of moderate- or high resolution spectra in the near- and shortwave infrared spectral ranges, this assumption causes errors of several percent in reflectance spectra for realistic inhomogeneous terrestrial atmospheres (see e.g.[70]).

Obviously, we need an efficient radiative transfer model that can account for the vertical distribution of absorption optical thickness at different wavelengths. For this purpose, we consider the radiance as a function of absorption optical thickness and perform calculations for a limited number interpolation points τ^l_{abs} and corresponding vertical distributions $\mathbf{n}^l$, with $l = 1, \cdots, L$. From the reference calculations $\mathbf{I}_{\mathrm{ms}}\left(\tau^l_{\mathrm{abs}}, n^l\right)$, we obtain the multiply scattered intensity vector $\mathbf{I}_{\mathrm{ms}}(\lambda)$ at any wavelength λ in the spectral range under consideration with absorption optical thickness $\tau_{\mathrm{abs}}(\lambda)$ and its vertical distribution $\mathbf{n}(\lambda)$ in two steps. First, we account for differences between the actual vertical distribution $\mathbf{n}(\lambda)$ and the vertical distributions $\mathbf{n}^l$ used in the reference calculations, by employing the linear approximation:

$$\mathbf{I}_{ms}(\tau^l_{\mathrm{abs}}, \mathbf{n}(\lambda)) \approx \mathbf{I}_{ms}(\tau^l_{\mathrm{abs}}, \mathbf{n}^l) + \frac{\partial \mathbf{I}_{ms}}{\partial \mathbf{n}} \left[\mathbf{n}(\lambda) - \mathbf{n}^l\right] \ , \tag{13.93}$$

where the derivatives with respect to $\mathbf{n}$ follow from the forward adjoint perturbation theory[71]. This yields the intensity vector of the multiply scattered radiation at the grid points τ^l_{abs}, corrected for the actual vertical distribution. To obtain the intensity I_{ms} at $\tau_{\mathrm{abs}}(\lambda)$, we fit a second order polynomial to the logarithm of the (absorption profile corrected) radiances at the grid points, using the grid point closest to $\tau_{\mathrm{abs}}(\lambda)$ and the two neighboring points. For Stokes parameters Q and U we perform the polynomial interpolation using stokes fractions $q = Q/I$ and $u = U/I$ but further the procedure is the same. Finally we correct for variation of scattering properties and surface albedo within the spectral window, applying a linear correction based on the derivatives of the spectrum with respect to scattering properties.

For the numerical implementation, we choose the grid points equidistant on a logarithmic scale, i.e.:

$$\tau_{\text{grid}}^{l} = \tau_{\text{min}} \cdot \left(\frac{\tau_{\text{max}}}{\tau_{\text{min}}}\right)^{\frac{l-1}{L-1}} . \tag{13.94}$$

where τ_{min} and τ_{max} are respectively the minimum and maximum absorption optical thickness in the spectral window under consideration. If $\tau_{\text{max}} > 15$ its value is set to 15, because for larger values of the absorption optical thickness the radiation field is dominated by single scattering (being calculated exactly) and hence interpolation errors are of minor importance. The rationale of choosing a logarithmic scale is to obtain more grid points at small values of absorption optical thickness, where multiple scattering effects are most important. We use $L = 10$ grid points for each of the three bands, but this number can be updated during the commissioning phase. For more information on the linear-k method we refer to[67].

14 Inversion Procedure

In the inversion procedure, we invert the linearized forward model of Eq. (19.112) for iteration step n to find the state vector $\mathbf{x}_{\text{n+1}}$ for iteration step $n+1$. Hereto, we minimize the following cost function[72]:

$$\mathbf{x}_{\text{n+1}} = \min_{\mathbf{x}} \left([\mathbf{K}\,\mathbf{x} - \mathbf{y}]^T \; \mathbf{S}_y^{-1} \; [\mathbf{K}\,\mathbf{x} - \mathbf{y}] \right) + \left([\mathbf{x} - \mathbf{x}_a]^T \; \gamma^2 \mathbf{H}^{-1} \; [\mathbf{x} - \mathbf{x}_a] \right), \tag{14.95}$$

which we transform to

$$\tilde{\mathbf{x}}_{\text{n+1}} = \min_{\tilde{\mathbf{x}}} \left([\tilde{\mathbf{K}}\,\tilde{\mathbf{x}} - \tilde{\mathbf{y}}]^T \; [\tilde{\mathbf{K}}\,\tilde{\mathbf{x}} - \tilde{\mathbf{y}}] \right) + \gamma^2 \;\; \left([\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_a]^T \; [\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_a] \right), \tag{14.96}$$

where $\tilde{\mathbf{K}} = \mathbf{S}_y^{-\frac{1}{2}} \mathbf{K} \mathbf{H}^{\frac{1}{2}}$, $\tilde{\mathbf{x}} = \mathbf{H}^{-\frac{1}{2}} \mathbf{x}$ and $\tilde{\mathbf{y}} = \mathbf{S}_y^{-\frac{1}{2}} (\mathbf{y} - \mathbf{F}(\mathbf{x}_{\mathbf{n}}))$. $\mathbf{x}_{\text{a}}$ is the a priori state vector, $\mathbf{S}_y$ is the measurement error covariance matrix, γ is a regularization parameters, and $\mathbf{H}$ is a regularization matrix that ensures that all state vector parameters range within the same order of magnitude and determine the relative weight of parameters in the side constraint[18]. Note that if $\gamma = 1$ and $\mathbf{H}$ is the prior error covariance matrix, Eq. (14.95) (and hence 14.96) reduce to the cost function of the optimal estimation method[73]. We use a diagonal matrix for $\mathbf{H}$ with diagonal elements $h_{ii} = w_i^2$.

The solution of Eq. (14.95) is given by:

$$\tilde{\mathbf{x}}_{n+1} = \tilde{\mathbf{x}}_n + \Lambda(\tilde{\mathbf{K}}^T \tilde{\mathbf{K}} + \gamma^2 \mathbf{I})^{-1} (\tilde{\mathbf{K}}^T \tilde{\mathbf{y}} - \gamma^2 (\tilde{\mathbf{x}}_n - \tilde{\mathbf{x}}_a)). \tag{14.97}$$

Λ is a filter/damping factor between 0 and 1, which limits the step size for each iteration of the state vector. In this way, we use a Gauss-Newton scheme with reduced step size to avoid diverging retrievals.

The error covariance matrix $\mathbf{S}_x$ of the retrieved state vector is given by

$$\mathbf{S}_x = \mathbf{S}_r + \mathbf{S}_e, \tag{14.98}$$

where $\mathbf{S}_r$ is the regularization error covariance matrix which describes the effect of the a priori error covariance matrix $\mathbf{S}_a$ on $\mathbf{x}$,

$$\mathbf{S}_r = (\mathbf{I} - \mathbf{A})\; \mathbf{S}_{\mathbf{a}}\; (\mathbf{I} - \mathbf{A})^{\mathbf{T}}, \tag{14.99}$$

and $\mathbf{S}_e$ is the retrieval error covariance matrix that describes the effect of measurement- and forward model errors on $\mathbf{x}$,

$$\mathbf{S}_e = \mathbf{D}\; \mathbf{S}_y\; \mathbf{D}^T, \tag{14.100}$$

where $\mathbf{D}$ is the contribution- or gain matrix

$$\mathbf{D} = \left(\mathbf{K}^T\; \mathbf{S}_y^{-1}\; \mathbf{K} + \gamma^2 \mathbf{H}^{-1}\right)^{-1}\; \mathbf{K}^T\; \mathbf{S}_y^{-1}, \tag{14.101}$$

and $\mathbf{A}$ is the averaging kernel

$$\mathbf{A} = \mathbf{D}\,\mathbf{K} \tag{14.102}$$

The standard deviation σ_τ on the AOD can be obtained from the retrieval error covariance matrix $\mathbf{S}_x$ via

$$\sigma_\tau = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} S_{i,j}\,\frac{\partial\tau}{\partial x_i}\,\frac{\partial\tau}{\partial x_j}} \tag{14.103}$$

where $S_{i,j}$ denotes element (i,j) of $\mathbf{S}_x$. A similar expression holds for the SSA.

For trace gas columns, the column averaging kernel for gas j is given by

$$a_{\mathrm{col,k,j}} = \mathbf{d}_{\mathrm{col,j}}\,\mathbf{k}_k\,\sum_{n=1}^{N_{\mathrm{RT}}} N_{\mathrm{std,n,j}} \tag{14.104}$$

where k indicates layer index, $\mathbf{k}_k = \frac{\partial\mathbf{F}}{\partial N_{gas,k,j}}$, and $\mathbf{d}_{\mathrm{col,j}}$ is the row of $\mathbf{D}$ that corresponds to the scaling factor for trace gas j. The final retrieval product for trace gases is the column averaged dry air mixing ratio

$$X_{\mathrm{gas,j}} = \frac{\sum_{k=1}^{N_{\mathrm{RT}}} N_{gas,k,j}}{\sum_{k=1}^{N_{\mathrm{RT}}} N_{\mathrm{air,k}}} \tag{14.105}$$

which is related to the true vertical profile of gas j through

$$X_{\mathrm{gas,j}} = \frac{\mathbf{a}_{\mathrm{col,j}}\,\mathbf{x}_{\mathrm{true}}}{\sum_{k=1}^{N_{\mathrm{RT}}} N_{\mathrm{air,k}}} \tag{14.106}$$

where $\mathbf{x}_{\mathrm{true}}$ is the vector containing the true sub-columns in the different vertical layers.

15 Choosing regularization parameter and step size

Within each iteration step, we compute $\tilde{\mathbf{x}}_{n+1}$ for different trial values for Λ (10 values between 0.1 and 1.0) and γ^2 (0.1, 1.3, 2.5, 3.8, 5.0) and compute the $\chi^2_{\mathrm{trial}} = \frac{1}{n_{\mathrm{meas}}}\sum_{i=1}^{n_{\mathrm{meas}}}\frac{(F_{\mathrm{appr}}(\mathbf{x}_{\mathrm{n}})-y_i)^2}{S_y(i,i)}$, where n_{meas} is the number of elements in $\mathbf{y}$ and $\mathbf{F}_{\mathrm{appr}}$. Here, $\mathbf{F}_{\mathrm{appr}}$ is given by:

$$\mathbf{F}_{\mathrm{appr}}(\mathbf{x}) = \mathbf{F}_{\mathrm{ms}}(\mathbf{x}_n) + \frac{\partial\mathbf{F}_{\mathrm{ms}}}{\partial\mathbf{x}}\,[\mathbf{x}-\mathbf{x}_{\mathrm{n}}] + \mathbf{F}_{\mathrm{ss}}(\mathbf{x}). \tag{15.107}$$

Here, $\mathbf{F}_{\mathrm{ms}}$ is the forward model only taking multiple scattering into account, which has already been computed for $\mathbf{x}_n$ in the previous iteration step, and $\mathbf{F}_{\mathrm{ss}}$ is the forward model only taking into account single scattering, which is calculated exactly for each trial Λ and γ^2 using Eq. (12.79). We select the combination of Λ and γ^2 that results in the smallest χ^2_{trial} to produce the state vector $\mathbf{x}_{\mathrm{n+1}}$ for the next iteration step.

16 Checking for Convergence

For each iteration step n, we compute χ^2 of the fit between forward model and measurement. Then we create an array χ^2_{order} of dimension n that contains all χ^2_i values for $i = 1, \cdots, n$, ordered from smallest to largest value. We consider the iteration to be converged if:

- $|\chi^2_n - \chi^2_{n-1}|/\chi^2_n < 0.01$ or $|\chi^2_n - \chi^2_{n-1}| < 0.05$.

- $|\chi^2_{\mathrm{order,1}} - \chi^2_{\mathrm{order,2}}|/\chi^2_{\mathrm{order,1}} < 0.01$ or $|\chi^2_{\mathrm{order,1}} - \chi^2_{\mathrm{order,2}}| < 0.05$.This criterion prevents the iteration to 'walk' away from the minimum.
- The difference between any 2 elements of $\chi^{\mathbf{2}}{}_{\mathrm{order}}$, $|\chi^2_{\mathrm{order,i}} - \chi^2_{\mathrm{order,i-1}}| < 0.01$.This criterion prevents the iteration to jump between 2 different χ^2 values.

The maximum number of iteration steps is set to 10. The iteration step that produced the minimum χ^2 is given as the output of the retrieval.

17 Look-up Table Retrieval to obtain 1st guess

The iterative retrieval starts with a 1st guess state vector $\mathbf{x}_1$. If $\mathbf{x}_1$ deviates too much from the truth the risk exists that the iteration diverges or ends up in a local minimum. We use a Look-Up Table (LUT) retrieval approach to obtain a reasonable 1st guess state vector. In the LUT-retrieval, we use a multi-mode aerosol description with 10 log-normal modes (see above). The state vector $\mathbf{x}_{\mathrm{lut}}$ for the LUT retrieval contains as aerosol parameters the mode aerosol column number $N_{\mathrm{aer,i}}, i = 1, \cdots, 10$. Here, each $N_{\mathrm{aer,i}}$ value corresponds to an AOD value at given wavelength (for different wavelengths the AOD is obviously different). Further $\mathbf{x}_{\mathrm{lut}}$ contains the same land surface or ocean parameters as the iterative retrieval. The forward model $\mathbf{F}_{\mathrm{lut}}$ for the LUT retrieval uses tabulated values of Stokes parameters $I_{\mathrm{ms,i}}, Q_{\mathrm{ms,i}}, U_{\mathrm{ms,i}}$ of multiple scattered light, calculated with the LINTRAN RT model for each aerosol mode i separately at discrete values of $N_{\mathrm{aer,i}}$, and the parameters describing the surface reflection matrix $\mathbf{R}_s$.

To obtain $\mathbf{F}_{\mathrm{lut}}(\mathbf{x}_{\mathrm{lut}})$ at a given wavelength λ with AOD$_\lambda$, we first (linearly) interpolate the tabulated values of the the Stokes parameters $I_{\mathrm{ms,node}}, Q_{\mathrm{ms,node}}, U_{\mathrm{ms,node}}$ at the node points of the LUT to the actual values of surface/ocean parameters in $\mathbf{x}_{\mathrm{lut}}$ and AOD$_\lambda$. Then we combine the Stokes parameters from the different size modes using the approach of[74]:

$$E_{\mathrm{ms}} = \sum_{i=1}^{N_{mode}} E_{\mathrm{ms,i}} \frac{\mathrm{AOD}_{\mathrm{i},\lambda}}{\mathrm{AOD}_\lambda} \tag{17.108}$$

where E denotes either I, Q, or U, and $\mathrm{AOD}_{\mathrm{i},\lambda}$ denotes the AOD of mode i at wavelength λ. The single scattering contribution to $\mathbf{F}_{\mathrm{lut}}$ is computed exactly using Eq. (12.79). Using the LUT based forward model $\mathbf{F}_{\mathrm{lut}}(\mathbf{x}_{\mathrm{lut}})$ we find $\mathbf{x}_{\mathrm{lut}}$ by using the same inversion procedure as described in section 14, with the difference that the approximate forward model of Eq. (15.107) is replaced by $\mathbf{F}_{\mathrm{lut}}$.

In the LUT based forward model, molecular absorption is neglected so only spectral bands are selected with very small contribution of molecular absorption. This means, the LUT retrieval is not suited for spectrometer-only retrievals. In MAP+spectrometer retrievals the LUT retrievals is used to obtain a 1st guess for the aerosol parameters. For spectr<ometer bands the surface parameter A from Eqs. 10.58 and 10.67 is obtained by

$$A_{1st,spec} = R_{cont}(\lambda_{\mathrm{cont}}) \frac{\pi}{u_0} / \frac{\partial R}{\partial A}(\lambda_{\mathrm{cont}}) \tag{17.109}$$

where $R_{cont}(\lambda_{\mathrm{cont}})$ is the largest value in the spectrometer window under consideration, which is assumed to represent the continuum reflectance value.

18 Nuisance

RemoTAP allows for fitting a spectral recalibration of the solar spectrum ($\delta\lambda_s$). To obtain the derivative of I in the line by line calculation with respect to $\delta\lambda_s$, we use the finite difference method by applying a numerical perturbation of the line-by-line wavelength grid of the solar spectrum.

Table 19.3: Spectral windows of the non-scattering retrievals and corresponding target species.

number	spectral window [nm]	target species
1	755–773	O2
2	1593–1621	CO2 (SWIR-1)
3	1629–1654	CH4 and H2O (SWIR-1)
4	1993–2095	CO2 and H2O (SWIR-2)

The shift in the Earth-shine radiance is retrieved throug a perturbing the ISRF s_{earth}. Instead of fitting the spectral shifts at each wavelength, we assume $\delta\lambda_s$ and $\delta\lambda_e$ to be spectral denpendent by a polynomial:

$$\delta\lambda_{s,e} = \sum_{0}^{N_{sh}} sh_i^{s,e}(\lambda - \lambda_0)^i \tag{18.110}$$

In this way, we fit the polynomial coefficients $sh_i^{s,e}$ for spectral shifts.

We also include intensity offset in the state vector, which is defined as constant over each spectrometer band by I_{off}.

19 Non-scattering Retrievals and the Proxy Product

Before we discuss the non-scattering retrieval, we reflect upon general aspects that are common to all the inversions in this document. Any retrieval algorithm aims at inferring an atmospheric state vector $\mathbf{x}$ from a measurement vector $\mathbf{y}$. The state vector is linked to the measurement vector through a forward model $\mathbf{F}(\mathbf{x},\mathbf{b})$ that depends on the state vector $\mathbf{x}$ and the vector $\mathbf{b}$ containing ancillary parameters that are not retrieved,

$$\mathbf{y} = \mathbf{F}(\mathbf{x},\mathbf{b}) + \mathbf{e}_y \tag{19.111}$$

where $\mathbf{e}_y$ represents the measurement error vector. In our case the measurement vector may consist of multi-spectral, multi-angle measurements of intensity and state of polarization performed by the MAP instrument and/or multi-spectral single-viewing-angle measurements in trace-gas absorption bands performed by the CO2I spectrometer. In the following, we will omit the dependence of $\mathbf{F}$ on $\mathbf{b}$.

For the retrieval procedure it is needed that the non-linear forward model is linearized so that the retrieval problem can be solved iteratively. For iteration step n the forward model is approximated by

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{x}_{\mathrm{n}}) + \mathbf{K}\left[\mathbf{x} - \mathbf{x}_{\mathrm{n}}\right] \tag{19.112}$$

where $\mathbf{x}_{\mathrm{n}}$ is the state vector for the current iteration step, and $\mathbf{K}$ is the Jacobian matrix with elements

$$K_{ij} = \frac{\partial F_i}{\partial x_j}(\mathbf{x}_{\mathrm{n}}). \tag{19.113}$$

Below, we will describe the algorithm for non-scattering retrieval in more detail and will address the more complex full-physics retrieval in the next section.

Retrievals of trace gas columns from spectrometer measurements under the assumption of an atmosphere without scattering are useful for cloud filtering[75] and for the proxy approach[76]. The non-scattering retrieval is performed for each spectral band of the spectrometer separately, where for each spectral band the relevant absorbers is taken into account: O_2 for the NIR band, CO_2, CH_4, and H_2O for the SWIR-1 band, and CO_2, CH_4, and H_2O for the SWIR-2 band.

20 Forward model

The non-scattering forward model calculates the backscattered top-of-the-atmosphere radiance I as a function of wavelength λ on a line-by-line basis, so that the convolution with the instrument spectral response function according to Eq. (8.6) can be performed.

For the non-scattering forward model, by definition, scattering is neglected and Lambert-Beer's law is applicable:

$$I(\lambda) \,=\, F_0\,(\lambda)\; A\,(\lambda)\; \frac{\mu_0}{\pi}\, e^{-\tau_{\rm abs}(\lambda)/\tilde{\mu}} \tag{20.114}$$

with the air mass factor

$$\frac{1}{\tilde{\mu}} = \frac{1}{\mu_0} + \frac{1}{\mu_v}\;, \tag{20.115}$$

where μ_0 is the (absolute value) cosine of the solar zenith angle, μ_v is the (absolute value) cosine of the viewing zenith angle. F_0 is the incoming solar irradiance and $\tau_{\rm abs}$ is the absorption optical thickness due to molecular absorption. The spectral dependence of the Lambertian surface albedo A is parametrized through an n-th order spectral polynomial, according to Eq. (10.70). The absorption optical depth per layer is computed according to Eq. (9.16). The total vertically integrated molecular absorption optical thickness $\tau_{\rm abs}$ is

$$\tau_{\rm abs} = \sum_j \sum_{k=1}^{K_{\rm atm}} \tau_{\rm abs,k,j} \tag{20.116}$$

The line-by-line forward model also provides the derivatives of I with respect to $\tau_{\rm abs}$, the surface albedo and the solar irradiance, namely

$$\frac{\partial I}{\partial \tau_{\rm abs}} = -\frac{I}{\tilde{\mu}} \tag{20.117}$$

$$\frac{\partial I}{\partial A} = \frac{I}{A} \tag{20.118}$$

$$\frac{\partial I}{\partial F_0} = \frac{I}{F_0} \tag{20.119}$$

These derivatives are progressed to derivatives with respect to the parameters to be retrieved via the chain rules:

$$\frac{\partial I}{\partial N_{\rm gas,k,j}} = -\sigma_j(p_k, T_k, \lambda)\,\frac{I}{\tilde{\mu}} \tag{20.120}$$

$$\frac{\partial I}{\partial a_i} = (\lambda - \lambda_0)^i\,\frac{I}{A(\lambda)} \tag{20.121}$$

where a_i is the i-th albedo coefficient in Eq. (10.70).

The spectral shifts are fitted in the measurements of both solar irradiance and Earth-shine radiandce for non-scattering retrieval.

21 Inversion

The state vector parameters of the non-scattering retrievals in each band, are restricted to the profile scaling parameters c_j of Eq. (9.56), the coefficients corresponding to a 1st order polynomial to describe the spectral dependence of the surface albedo, and a wavelength shift between the earth radiance and solar irradiance.

Similar to the regular retrieval, the non-scattering inversion is performed iteratively. However, instead of minimizing the regularized cost function of Eq. (14.95), we minimize the least-squares cost function for each iteration step

$$\mathbf{x}_{\mathrm{n}+1} = \min_{\mathbf{x}} \left([\mathbf{K}\,\mathbf{x} - \mathbf{y}]^T\, \mathbf{S}_y^{-1}\, [\mathbf{K}\,\mathbf{x} - \mathbf{y}] \right), \tag{21.122}$$

which is solved by

$$\mathbf{x}_{\mathrm{n}+1} = \left(\mathbf{K}^T\, \mathbf{S}_y^{-1}\, \mathbf{K}\right)^{-1}\, \mathbf{K}^T\, \mathbf{S}_y^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x_n})). \tag{21.123}$$

Since the non-scattering forward model is sufficiently linear, it is not needed to use a step-size reduction as in the regular retrieval.

22 Proxy retrieval

The proxy methane product uses the non-scattering output to construct a best estimate of the column-average dry-air mole fraction of methane ($\text{XCH}_4^{\mathrm{proxy}}$) according to

$$\text{XCH}_4^{\mathrm{proxy}} = \frac{C_{\mathrm{CH_4}}}{C_{\mathrm{CO_2}}} \text{XCO}_2^{\mathrm{apr}}, \tag{22.124}$$

where total columns $C_{\mathrm{CH_4}}$ and $C_{\mathrm{CO_2}}$ are retrieved under the non-scattering assumption from the sub-windows indicated in Table 19.3. $\text{XCO}_2^{\mathrm{apr}}$ is the total dry air mixing ratio of carbon dioxide coming from the CAMS forecast.

Similarly, one can use the proxy to retrieve the column-average dry-air mole fraction of carbon dioxide ($\text{XCO}_2^{\mathrm{proxy}}$) according to

$$\text{XCO}_2^{\mathrm{proxy}} = \frac{C_{\mathrm{CO_2}}}{C_{\mathrm{CH_4}}} \text{XCH}_4^{\mathrm{apr}}, \tag{22.125}$$

As outlined before, the first key assumption of the proxy approach is that scattering effects cancel in the $\frac{C_{\mathrm{CH_4}}}{C_{\mathrm{CO_2}}}$ ratio, which is only true if the respective CH_4 and CO_2 columns are inferred from spectral absorption bands of similar strength spectrally close to each other. Therefore, the proxy approach must use the SWIR-1 band. The other key assumption is that the a priori estimate of $\text{XCO}_2^{\mathrm{apr}}$ or $\text{XCH}_4^{\mathrm{apr}}$is sufficiently close to the true CO_2 or CH_4 total column mixing ratio such that the respective error contribution to $\text{XCH}_4^{\mathrm{proxy}}$ or $\text{XCO}_2^{\mathrm{proxy}}$ does not jeopardize the accuracy.

The precision of the proxy product is defined as the instrument noise propagation on the proxy product, so it is derived from the noise propagations of the non-scattering retrieval results. As the spectral windows of the non-scattering retrievals of CH_4 and CO_2 do not overlap, and instrument noise is assumed to be uncorrelated among spectral pixels, there is no correlated noise on the two non-scattering retrieval results. The precision on the proxy XCH_4 product is therefore defined as:

$$\varepsilon_{\text{XCH}_4}^{\mathrm{proxy}} = \text{XCH}_4^{\mathrm{proxy}} \sqrt{\left(\frac{\epsilon_{\mathrm{CH_4}}^{\mathrm{ns}}}{CH_4^{\mathrm{ns}}}\right)^2 + \left(\frac{\epsilon_{\mathrm{CO_2}}^{\mathrm{ns}}}{CO_2^{\mathrm{ns}}}\right)^2}, \tag{22.126}$$

where $\epsilon_{\mathrm{CH_4}}^{\mathrm{ns}}$ and $\epsilon_{\mathrm{CO_2}}^{\mathrm{ns}}$ denote the precision on the CH_4 column CH_4^{ns} and retrieved CO_2 column CO_2^{ns}, respectively, from the non-scattering retrieval approach. The same expression holds for $\varepsilon_{\text{XCO}_2}^{\mathrm{proxy}}$. This precision term only contains instrument noise propagation. Estimated errors on XCO_2 or XCH_4 forecast from CAMS, and biases because of imperfect cancellation of scattering effects are not included in this quantity.

23 Detailed Algorithm Setup

This section provides more detail on the different processing chains of RemoTAP.

24 Detailed Algorithm Setup MAP-only retrievals

25 State Vector Definition

For MAP-only retrievals, RemoTAP has the option the option to choose the multi-mode aerosol description, ithe parametric 2-mode, or the parametric 3-mode aerosol description. The corresponding state vector elements are given in Tables 25.4, 25.5, and 25.6, respectively. Currently, the parametric 3-mode setup is selected as the baseline. For all setups there is the option to add a cirrus mode to the state vector. The corresponding elements are given in Table 25.7.

Table 25.4: State vector elements, prior values, and corresponding values w_i that define the diagonal regularization matrix $\mathbf{H}$ by $h_{ii} = w_i^2$, for the multi-mode aerosol description. Note that for the commonly used 5-mode setup we use (from Table 9.2) modes 2, 4, 6 (fine modes),and 7, 9 (coarse modes). If $w_i = 0$, indicated by *, the corresponding parameters is fixed. σ_{ext} denotes the aerosol extinction cross-section at 550 nm.

state vector element	prior value	w_i	min - max
Aerosol parameters			
N$_i, i = 1, \cdots, N_{\text{mode}}$	0.0001/σ_{ext}	0.5/$N_{\text{mode}}/\sigma_{\text{ext}}$	0.001 - n/a
f^f_{sphere}	0.95	1.0	0.0 - 1.0
α^f_{inorg}	1.0	0.1	such that $1.3 < m_r < 1.69$
α^f_{bc}	0.015	0.1	such that $1.3 < m_r < 1.69$
α^f_{oc}	0.10	0.1	such that $1.3 < m_r < 1.69$
f^c_{sphere}	0.10	0.1	0.0 - 1.0
α^c_{inorg}	1.02	0.1	such that $1.3 < m_r < 1.69$
α^c_{du}	0.95	1.0	such that $1.3 < m_r < 1.69$
z_{aer}	2 km	4 km	-2 km - 10 km
Option 1: Ross-Li land surface parameters			
$A(\lambda_i), i = 1, \cdots, N_{\text{band,map}}$	0.05	0.50	0.0 - 1.0
k_{geo}	0.1	0.1	0.0 - 0.35
k_{vol}	0.5	0.5	0.0 - 1.5
Option 2: RPV land surface parameters			
$A(\lambda_i), i = 1, \cdots, N_{\text{band,map}}$	0.05	0.50	0.0 - 1.0
g	-0.09	0.5	-0.5 - 0.5
k	0.80	0.5	0.05 - 0.95

Land surface parameters (Maignan)			
B	1.0	1.0	0.2 - 10.0
ν	0.1	0.0*	n/a
Ocean parameters			
windspeed	7 m/s	30 m/s	0.1 m/s - 100 m/s
x_{chl}	2 mg/m^3	20 mg/m^3	0 mg/m^3 - 30 mg/m^3
$A(\lambda_i), i = 1, \cdots, N_{\mathrm{band,map}}$	0.0	0.01	-0.10 - 0.10

Table 25.5: State vector elements, prior values, and corresponding values w_i that define the diagonal regularization matrix $\mathbf{H}$ by $h_{ii} = w_i^2$, for the parametric 2-mode aerosol description. If $w_i = 0$, indicated by *, the corresponding parameters is fixed.

state vector element	prior value	w_i	min - max
Aerosol parameters			
r_{eff}^f	0.20 µm	0.10 µm	0.02 µm - 0.30 µm
v_{eff}^f	0.20	0.05	0.01 - 0.80
N^f	0.20/σ_{ext}	2.0/σ_{ext}	0.001 - n/a
f_{sphere}^f	0.95	1.0	0.0 - 1.0
$\alpha_{\mathrm{inorg}}^f$	1.0	0.1	such that $1.3 < m_r < 1.69$
α_{bc}^f	0.015	0.1	such that $1.3 < m_r < 1.69$
α_{oc}^f	0.10	0.1	such that $1.3 < m_r < 1.69$
r_{eff}^c	1.5 µm	1.0 µm	0.2 µm - 4.0 µm
v_{eff}^c	0.20	0.05	0.01 - 0.80
N^c	0.20/σ_{ext}	2.0/σ_{ext}	0.001 - n/a
f_{sphere}^c	0.10	1.0	0.0 - 1.0
$\alpha_{\mathrm{inorg}}^c$	1.02	0.05	such that $1.3 < m_r < 1.69$
α_{du}^c	0.95	1.0	such that $1.3 < m_r < 1.69$
z_{aer}	2 km	4 km	-2 km - 10 km
Option 1: Ross-Li land surface parameters			

$A(\lambda_i), i = 1, \cdots, N_{\text{band,map}}$	0.05	0.50	0.0 - 1.0
k_{geo}	0.1	0.1	0.0 - 0.35
k_{vol}	0.5	0.5	0.0 - 1.5
Option 2: RPV land surface parameters			
$A(\lambda_i), i = 1, \cdots, N_{\text{band,map}}$	0.05	0.50	0.0 - 1.0
g	-0.09	0.5	-0.5 - 0.5
k	0.80	0.5	0.05 - 0.95
Land surface parameters (Maignan)			
B	1.0	1.0	0.2 - 10.0
ν	0.1	0.0^*	n/a
Ocean parameters			
windspeed	7 m/s	30 m/s	0.1 m/s - 100 m/s
x_{chl}	2 mg/m^3	20 mg/m^3	0 mg/m^3 - 30 mg/m^3
$A(\lambda_i), i = 1, \cdots, N_{\text{band,map}}$	0.0	0.01	-0.10 - 0.10

Table 25.6: State vector elements, prior values, and corresponding values w_i that define the diagonal regularization matrix $\mathbf{H}$ by $h_{ii} = w_i^2$, for the parametric 3-mode aerosol description. The superscript ^{ci} and ^{cs} denote 'coarse insoluble' and 'coarse soluble', respectively. If $w_i = 0$, indicated by *, the corresponding parameters is fixed.

state vector element	prior value	w_i	min - max
Aerosol parameters			
r_{eff}^f	0.15 µm	0.20 µm	0.02 µm - 0.30 µm
v_{eff}^f	0.20	0.05	0.01 - 0.80
N^f	0.0001/σ_{ext}	0.5/σ_{ext}	0.001 - n/a
f_{sphere}^f	0.95	0.1	0.0 - 1.0
α_{inorg}^f	1.0	0.1	such that $1.3 < m_r < 1.69$
α_{bc}^f	0.015	0.1	such that $1.3 < m_r < 1.69$
α_{oc}^f	0.10	0.1	such that $1.3 < m_r < 1.69$
r_{eff}^{ci}	1.0 µm	1.0 µm	0.7 µm - 5.0 µm

v_{eff}^{ci}	0.60	0*	n/a
N^{ci}	0.0001/σ_{ext}	0.5/σ_{ext}	0.001 - n/a
f_{sphere}^{ci}	0.0	0*	n/a
$\alpha_{\text{du,im}}^{ci}$	0.95	0.10	0-1
$z_{\text{aer}}^{f,ci}$	2 km	4 km	-2 km - 10 km
r_{eff}^{cs}	2.5 µm	1.0 µm	0.7 µm - 5.0 µm
v_{eff}^{cs}	0.60	0*	n/a
N^{cs}	0.0001/σ_{ext}	0.5/σ_{ext}	0.001 - n/a
f_{sphere}^{cs}	1.0	0*	n/a
$\alpha_{\text{inorg,im}}$	0.9	0.10	such that $1.3 < m_r < 1.69$
z_{aer}^{cs}	0.5 km	0*	n/a
Option 1: Ross-Li land surface parameters			
$A(\lambda_i), i = 1, \cdots, N_{\text{band,map}}$	0.05	0.50	0.0 - 1.0
k_{geo}	0.1	0.1	0.0 - 0.35
k_{vol}	0.5	0.5	0.0 - 1.5
Option 2: RPV land surface parameters			
$A(\lambda_i), i = 1, \cdots, N_{\text{band,map}}$	0.05	0.50	0.0 - 1.0
g	-0.09	0.5	-0.5 - 0.5
k	0.80	0.5	0.05 - 0.95
Land surface parameters (Maignan)			
B	1.0	1.0	0.2 - 10.0
ν	0.1	0.0*	n/a
Ocean parameters			
windspeed	7 m/s	30 m/s	0.1 m/s - 100 m/s
x_{chl}	2 mg/m^3	20 mg/m^3	0 mg/m^3 - 30 mg/m^3
$A(\lambda_i), i = 1, \cdots, N_{\text{band,map}}$	0.0	0.01	-0.10 - 0.10

Table 25.7: State vector elements, prior values, and corresponding values w_i that define the diagonal regularization matrix $\mathbf{H}$ by $h_{ii} = w_i^2$, for the cirrus mode. If $w_i = 0$, indicated by *, the corresponding parameters is fixed.

state vector element	prior value	w_i	min - max
$r_{\mathrm{eff,cir}}$	25.0 µm	10.0 µm	5.0 µm - 50.0 µm
α_{cir}	0.65	1.0	0.18 - 5.5
δ_{cir}	0.63	0.2	0.0 - 0.7
N_{cir}	0.04	0.05	0.0 - n/a
z_{cir}	tropopause height	4 km	5 km - 18 km

26 Measurement vector definition

The measurement vector for MAP-only retrievals contains the reflectance $R_I = I/F_{0,meas}$, $F_{0,\text{meas}}$ being the model solar flux perpendicular to the solar beam, and the relative Stokes parameters $q = Q/I$, $u = U/I$ at multiple wavelengths (to be specified in the initialization file), and for the different viewing angles provided by the MAP. The number of wavelengths can be different for R_I on the one hand and q and u on the other hand. As an alternative for using q and u the code also works with Degree of Linear Polarization (DoLP=$\sqrt{q^2 + u^2}$) in the measurement vector. This would ignore potential information present in the Angle of Linear Polarization but may have the advantage of having smaller measurement uncertainties. The best choice is to be made based on evaluation of retrievals on real measurements.

27 Algorithm flow

The MAP-only processing chain requires the level 1C data and the pre-processor data as input. We assume the module is called by a processing framework as illustrated in Fig. 2.2, which is optimized for the processing of data granules. A schematic overview of the MAP-only retrieval is given in Fig. 27.5 where the software is thread-safe and so appropriate for parallelization. Numerical tests were performed using the openMP and openMPI application. For each L1C ground pixel, the observation is checked for clouds and only cloud-free pixels are being further processed. Subsequently, the first guess retrieval and iterative retrieval are being performed.Finally, the results are returned to the calling framework and all dynamic memory is deallocated. Section 53 gives a detailed overview of the input and output of the MAP-only processor.

The iterative retrieval procedure is outlined in more detail in Fig. 27.6. The iterative retrieval starts with the 1st guess state vector that comes from the LUT retrieval. In the iteration scheme, the forward model and inversion procedures are called for each iteration step till convergence has been reached. The forward model and inversion are shown in more detail in Figs. 27.7 and 27.8.

28 Cloud filtering

For MAP-only retrievals, we perform an *a priori* and an *a posteriori* cloud filtering. The *a priori* cloud filtering is based on a cloud classification from an external instrument. CO2M mission the external

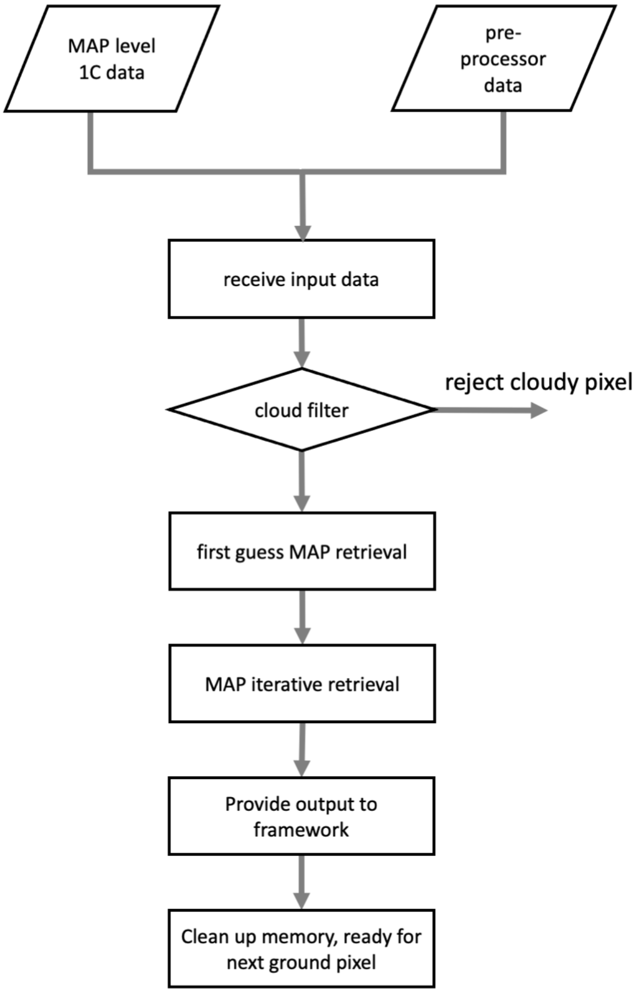

Figure 27.5: Schematic overview of the general retrieval procedure of MAP-only retrievals. The pre-processed data is output from the pre-processor of Fig. 6.4. The iterative retrieval is further detailed in Fig. 27.6.

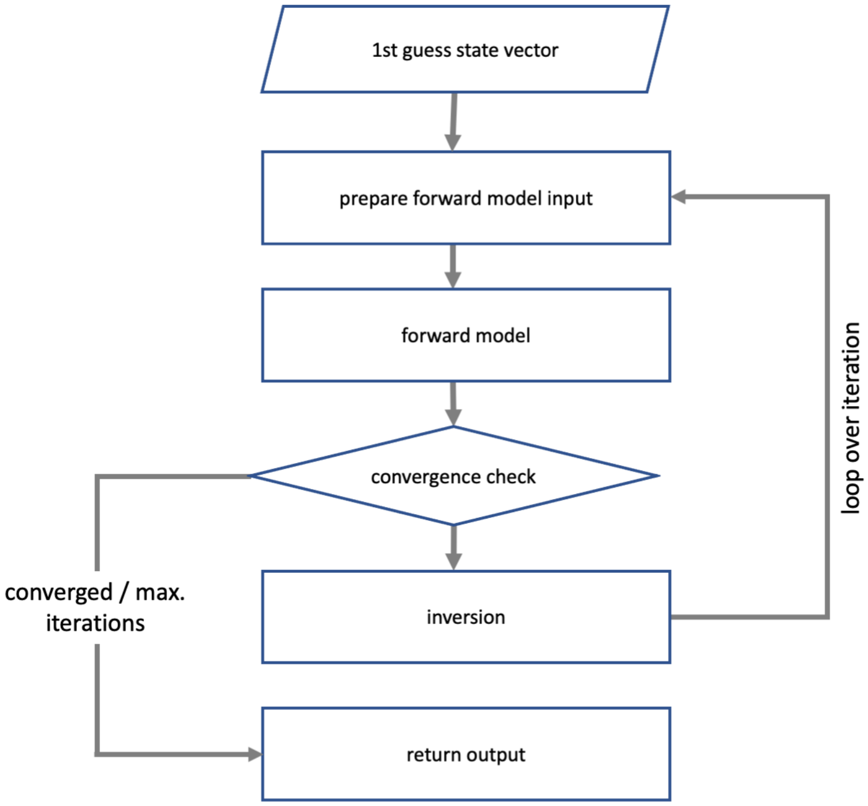

Figure 27.6: Flow chart of the iterative retrieval procedure. The forward model is further detailed in Fig. 27.7 and the inversion procedure in Fig. 27.8.

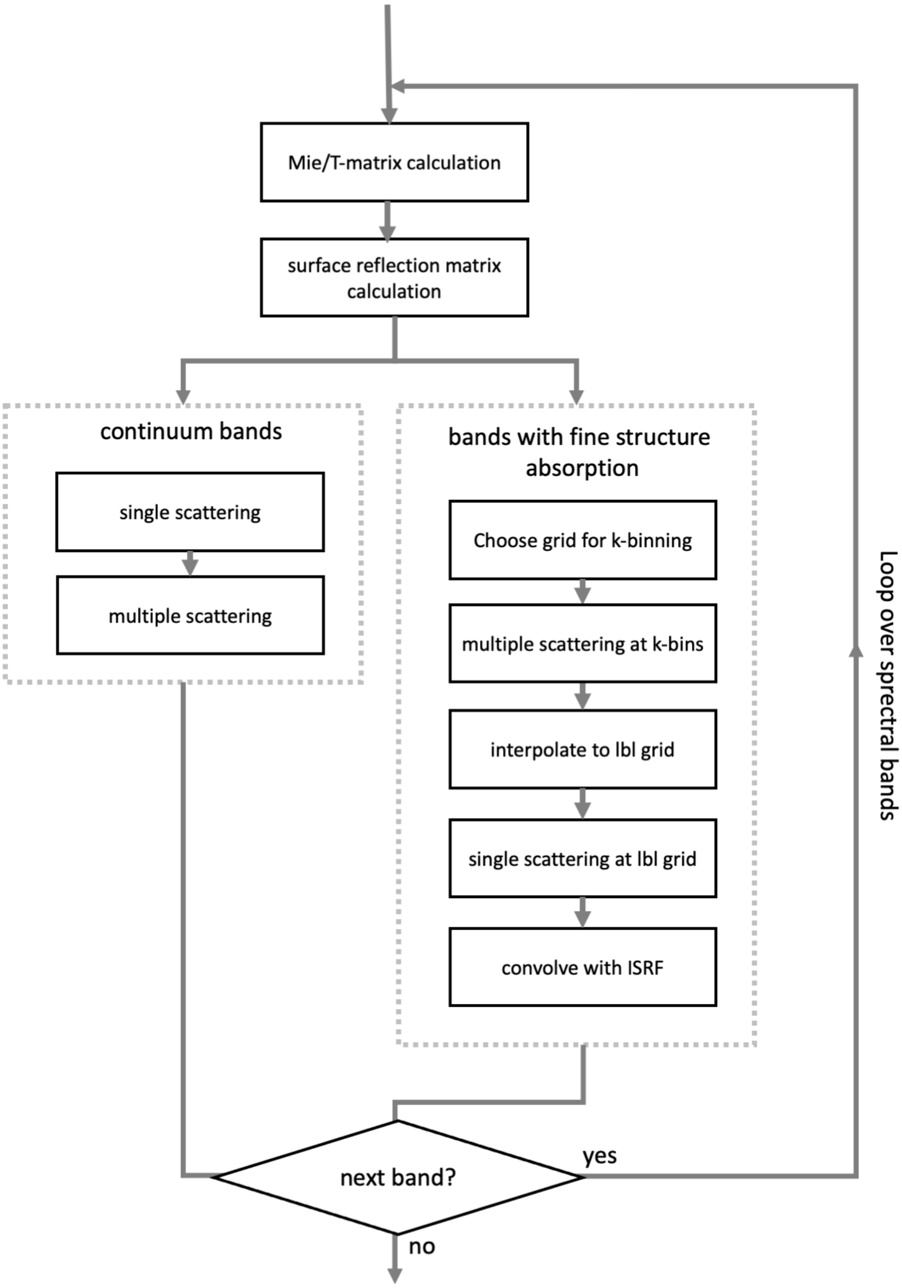

Figure 27.7: Flow chart of the forward model used in the iterative retrieval.

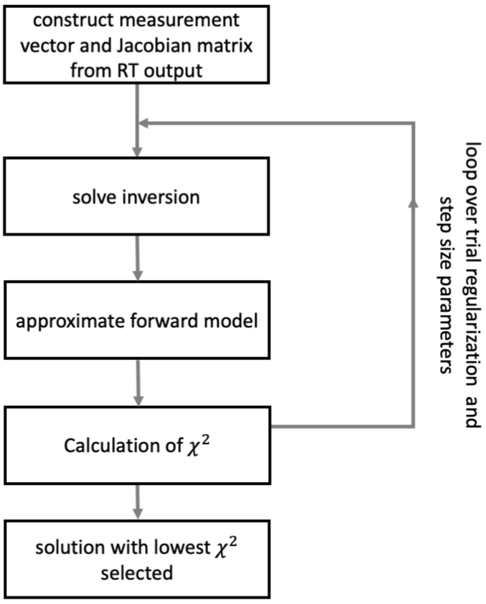

Figure 27.8: Flow chart of the inversion procedure used in the iterative retrieval.

cloud information comes from the Cloud Imager (CLIM). In a pre-processing step, the cloud information should be aggregated to the spatial grid of the L1C MAP data (see below). We expect that for each spatial grid cell of the L1C grid the following information:

- The fraction of the L1C grid cell that is confidently clear $f_{\mathrm{conf,clear}}$.
- The fraction of the L1C grid cell that is probably clear $f_{\mathrm{prob,clear}}$.
- The fraction of the L1C grid cell that is confidently cloudy $f_{\mathrm{conf,cloud}}$.
- The fraction of the L1C grid cell that is probably cloudy $f_{\mathrm{prob,cloud}}$.

We only start the retrieval for a L1C grid cell if $f_{\mathrm{conf,clear}} + f_{\mathrm{prob,clear}} > T_{\mathrm{cl}}$, where T_{cl} is a threshold to be determined based on validation of the aerosol products. As a start, we will use $T_{\mathrm{cl}} = 0.90$.

Our *a posteriori* cloud filtering is based on the χ^2 difference between forward model and measurement. Using POLDER-3 data[25], we found that this is an effective way to screen out clouds. MAP measurements have a unique capability to distinguish between scattering by (liquid) cloud droplets and aerosol particles and therefore, if a L1C grid cell is cloud contaminated, the forward model will not be able to reproduce the measurement, resulting in a large χ^2. The final threshold of the χ^2 has to be determined based on real measurements. For POLDER-3, we found χ^2=2 was appropriate.

29 Detailed Algorithm Setup Non-Scattering Retrieval

30 State Vector Definition

For the non-scattering retrievals, we fit the different trace gas columns (in the form of profile scaling factors) separately for each spectral band (NIR, SWIR-1, SWIR-2). Further, only surface albedo and its spectral dependence is fitted. The corresponding state vector elements are given in Table 30.8.

Table 30.8: State vector elements, prior values, and corresponding values w_i that define the diagonal regularization matrix $\mathbf{H}$ by $h_{ii} = w_i^2$, for the elements related to trace gas columns and surface reflection. If $w_i = 0$, indicated by *, the corresponding parameters is fixed.

state vector element	prior value	w_i	min - max
Trace gas parameters			
c_{co2} (SWIR-1, SWIR-2)	1.0	1.0	0.0 - 2.0
c_{ch4} (SWIR-1)	1.0	1.0	0.0 - 2.0
c_{h2o} (SWIR-1, SWIR-2)	1.0	1.0	0.0 - 2.0
c_{o2} (NIR)	1.0	1.0	0.0 - 2.0
Land surface parameters (retrieved)			
a_0(NIR, SWIR-1, SWIR-2)	0.1	0.50	0.0 - 1.0
a_1(NIR, SWIR-1, SWIR-2)	0.0	1e-3	-1e-3 - 1e-3
a_2(NIR, SWIR-1, SWIR-2), (optional)	0.0	1e-5	-1e-5 - 1e-5

Land surface parameters (fixed)			
k_{geo}	0.0	0.0*	
k_{vol}	0.0	0.0*	
B	1.0	0.0*	
ν	0.1	0.0*	
Ocean parameters			
windspeed	7 m/s	30 m/s	0.1 m/s - 100 m/s
a_0^{ocean}(NIR, SWIR-1, SWIR-2)	0.0	0.50	-0.1 - 0.1
a_1^{ocean}(NIR, SWIR-1, SWIR-2)	0.0	1e-3	-1e-3 - 1e-3
a_2^{ocean}(NIR, SWIR-1, SWIR-2)	0.0	1e-5	-1e-5 - 1e-5
Nuisance			
$sh_0^{s,e}$(NIR, SWIR-1, SWIR-2)	0.0	0.1	-0.1 - 0.1
$sh_1^{s,e}$(NIR, SWIR-1, SWIR-2)	0.0	0.1	-1e-4 - 1e-4
$sh_2^{s,e}$(NIR, SWIR-1, SWIR-2), optional	0.0	0.1	-1e-6 - 1e-6
I_{off}(NIR, SWIR-1, SWIR-2), optional	0.0	1e-14 (photons/cm2/nm/sr)	-1e-9 - 1e-9 (photons/cm2/nm/sr)

31 Measurement vector definition

The measurement vector for CO2I-only retrievals contains the Earth radiance I together with the line-by-line irradiance spectrum derived in Sec. 5, both as function of wavelengths for the 3 spectrometer bands NIR, SWIR-1, and SWIR-2. Alternatively, the reflectance $R_I = I/F_0$ can be provided as input.

32 Algorithm flow, input, output

The flow of the non-scattering retrieval is given in Fig. 32.9. The non-scattering will be performed for both clear and cloudy scenes so no cloud screening is performed here.

33 Detailed Algorithm Setup CO2I-only Retrieval

34 State Vector Definition

For CO2I-only retrievals, we use the single mode power-law description for aerosols, like in RemoTeC[51].The corresponding state vector elements are given in Table 34.9 and 34.10.

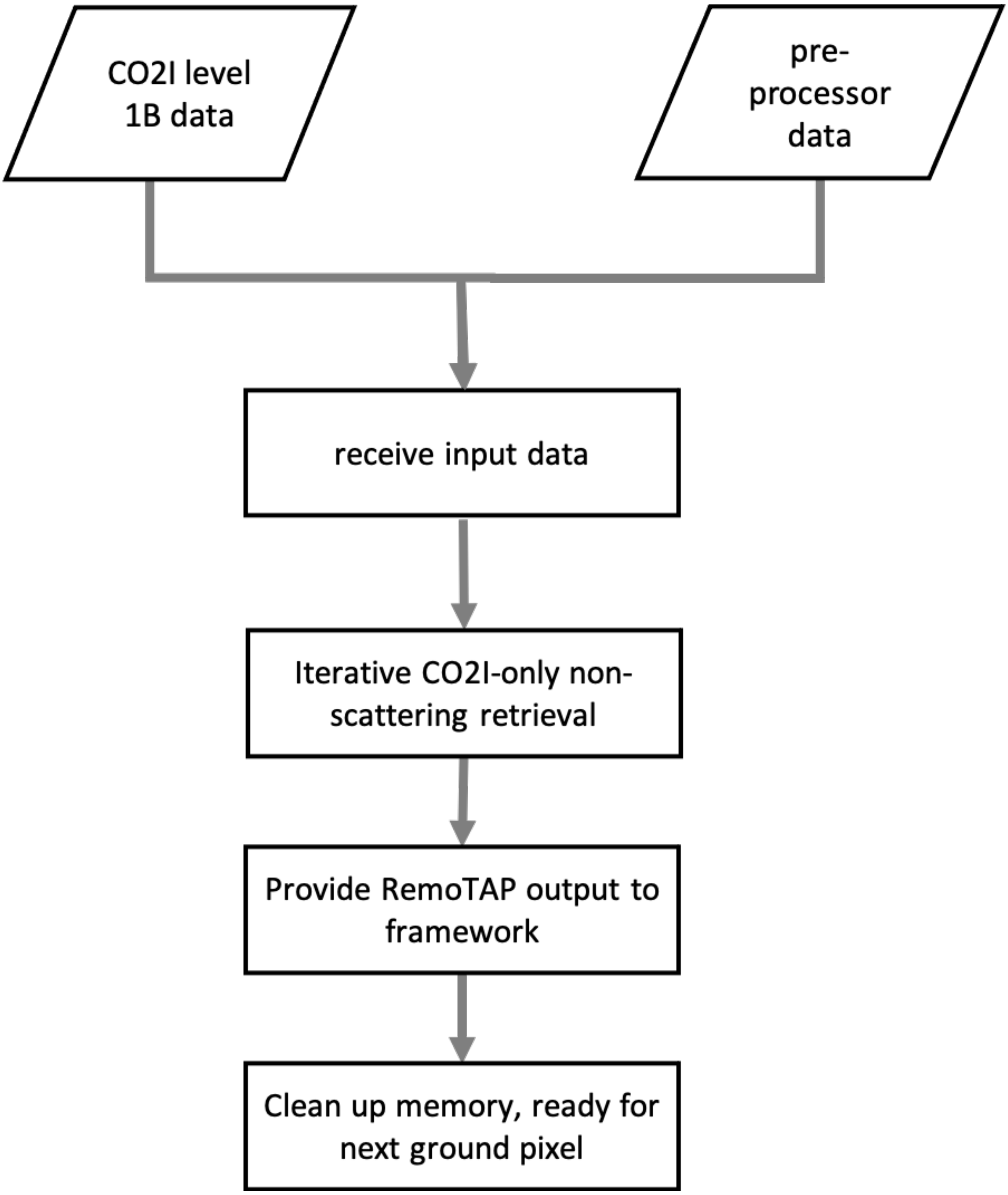

Figure 32.9: Schematic overview of the general retrieval procedure of non-scattering retrievals.

Table 34.9: State vector elements, prior values, and corresponding values w_i that define the diagonal regularization matrix $\mathbf{H}$ by $h_{ii} = w_i^2$, for the power-law aerosol description (see Eq. 9.53). If $w_i = 0$, indicated by *, the corresponding parameters is fixed. σ_{ext} denotes the aerosol extinction cross-section at 550 nm.

state vector element	prior value	w_i	min - max
Retrieved aerosol parameters			
p	4.0	0.10	0.02- 6.0
N	0.10/σ_{ext}	0.50/σ_{ext}	0.0- n/a
z_{aer}	2 km	1 km	-2 km - 10 km
Fixed aerosol parameters			
f_{sphere}	1.0	0.0*	
m_r	1.4	0.0*	(same for all wavelengths)
m_i	0.003	0.0*	(same for all wavelengths)
w_0	2 km	0.0*	

Table 34.10: State vector elements, prior values, and corresponding values w_i that define the diagonal regularization matrix $\mathbf{H}$ by $h_{ii} = w_i^2$, for the elements related to trace gas columns and surface reflection. If $w_i = 0$, indicated by *, the corresponding parameters is fixed.

state vector element	prior value	w_i	min - max
Trace gas parameters			
c_{co2}	1.0	1.0	0.0 - 2.0
c_{ch4}	1.0	1.0	0.0 - 2.0
c_{h2o}	1.0	1.0	0.0 - 2.0
Land surface parameters (retrieved)			
a_0(NIR, SWIR-1, SWIR-2)	0.1	0.50	0.0 - 1.0
a_1(NIR, SWIR-1, SWIR-2)	0.0	1e-3	-1e-3 - 1.e-3
a_2(NIR, SWIR-1, SWIR-2), (optional)	0.0	1e-5	-1e-5 - 1.e-5
Land surface parameters (fixed)			
k_{geo}	0.0	0.0*	

k_{vol}	0.0	0.0*	
B	1.0	0.0*	
ν	0.1	0.0*	
Ocean parameters			
windspeed	7 m/s	30 m/s	0.1 m/s - 100 m/s
a_0^{ocean}(NIR, SWIR-1, SWIR-2)	0.0	0.50	-0.1 - 0.1
a_1^{ocean}(NIR, SWIR-1, SWIR-2)	0.0	1e-3	-1e-3 - 1.e-3
a_2^{ocean}(NIR, SWIR-1, SWIR-2),	0.0	1e-5	-1e-5 - 1.e-5
Nuisance			
$sh_0^{s,e}$(NIR, SWIR-1, SWIR-2)	0.0	0.1	-0.1 - 0.1
$sh_1^{s,e}$(NIR, SWIR-1, SWIR-2)	0.0	0.1	-1e-4 - 1e-4
$sh_2^{s,e}$(NIR, SWIR-1, SWIR-2) (optional)	0.0	0.1	-1e-6 - 1e-6
I_{off}(NIR, SWIR-1, SWIR-2), (optional)	0.0	1e-14 (photons/cm2/nm/sr)	-1e-9 - 1e-9 (photons/cm2/nm/sr)

35 Measurement vector definition

The measurement vector for CO2I-only retrievals contains the Earth radiance I together with the line-by-line irradiance spectrum derived in Sec. 5, both as function of wavelengths for the 3 spectrometer bands NIR, SWIR-1, and SWIR-2. Alternatively, the reflectance $R_I = I/F_0$ can be provided as input.

36 Algorithm flow, input, output

The flow of the retrieval is slightly modified compared to MAP-only retrievals, see Fig. 36.10, as it does not include the LUT retrieval but receives the results of the non-scattering retrievals from Sec. 19 as input for the cloud clearing of observations. A fixed 1st guess for the aerosol properties is chosen of $p = 2$, AOD=0.10, and $z_{\text{aer}} = 2.0\ km$. The forward model of the iterative retrieval scheme and inversion procedure are the same as for the MAP-only retrieval, see Figs. 27.7 and 27.8. Note that for CO2I-only retrievals the continuum band approach is not applicable as all spectrometer bands include absorption bands with fine spectral features.

37 Cloud filtering

For spectrometer retrievals, we will use the result from non-scattering retrievals in the different spectral bands (NIR, SWIR-1, SWIR-2) to filter for clouds, following[75]. Here, we retrieve independent estimates of the CO2 and H2O column abundances using observations taken at 1.61 µm (SWIR-1 weak CO2

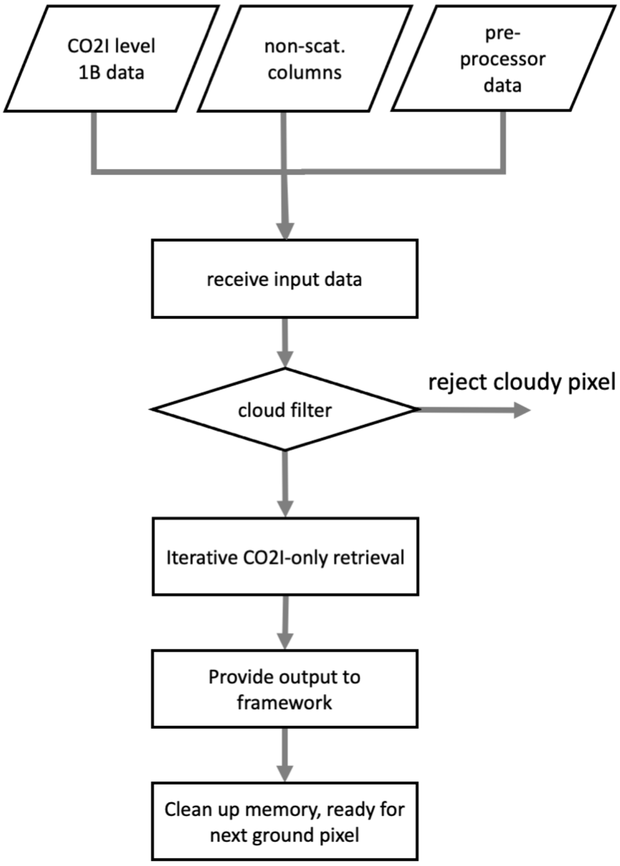

Figure 36.10: Schematic overview of the general retrieval procedure of CO2I-only retrievals. The iterative retrieval is further detailed in Fig. 27.6.

band) and 2.06 µm (SWIR-2 strong CO2 band), while neglecting atmospheric scattering (see above). The CO2 and H2O column abundances retrieved in these two spectral regions differ significantly in the presence of cloud and scattering aerosols. Ratios of the retrieved CO2 (R_{CO2}) and H2O (R_{H2O}) column abundances are computed as

$$R_{\mathrm{gas}} = \frac{C^W_{\mathrm{gas}}}{C^S_{\mathrm{gas}}}, \tag{37.127}$$

where C^W_{gas} and C^S_{gas} represents the vertical column density of the retrieved gas (CO2 or H2O) in the weak and strong absorption bands, respectively. Clouds and aerosols modify the optical path length in the two bands differently, producing column abundance ratios significantly different from unity[75]. There are two fundamental reasons why the ratio deviates from unity in the presence of scattering. First, for most terrestrial surfaces, albedos in the 1.6 µm band are most often higher than at 2.0 µm. This yields a variable fractional contribution of scattered light to the measured radiances. Second, the 1.6 and 2.06 µm band strengths are very different, resulting in different sensitivities to atmospheric scattering. If no scattering is assumed in the retrieval, a deviation from unity in the ratio thus indicates a substantial variation in the photon path-length (PPL) distribution between the two bands, while, in the absence of scattering, this ratio approaches unity.

Using OCO-2 data[30], we found that suitable values are $0.990 < R_{\mathrm{co2}} < 1.035$ and $0.950 < R_{\mathrm{h2o}} < 1.060$. Further, we found the useful filter $0.885 < \frac{C^{\mathrm{ret}}_{O_2}}{C^{\mathrm{met}}_{O_2}} < 1.020$, where $C^{\mathrm{ret}}_{O_2}$ is the retrieved O_2 column from the NIR band under non-scattering assumption and $C^{\mathrm{met}}_{O_2}$ is the O_2 column as computed from the surface pressure from the meteorological input data.

38 Detailed Algorithm Setup Simultaneous Retrieval MAP+CO2I

39 State Vector Definition

For the aerosol part of the state vector, we have the option to choose the multi-mode aerosol description, the parametric 2-mode, or the parametric 3-mode aerosol description. The corresponding state vector elements are given in Tables 25.4, 25.5, and 25.6 in section 24, respectively. The cirrus related state vector elements are given in 25.7. Currently, the parametric 3-mode setup is selected as the baseline. For the part of the state vector related to the spectrometer bands, the state vector elements are given in Table 34.10.

40 Measurement vector definition

The measurement vectors for the combined MAP+CO2I retrieval contains:

- From CO2I the reflectance, $R_I = I/F_0$ or the Earth shine radiance together with the line-by-line irradiance spectrum derived in Sec. 5, both as function of wavelengths for the 3 spectrometer bands NIR, SWIR-1, and SWIR-2.
- From MAP the reflectance $R_I = I/F_{0,\mathrm{meas}}$ and the relative Stokes parameters $q = Q/I$, $u = U/I$ at multiple wavelengths (to be specified in the initialization file), and for the different viewing angles provided by the MAP. The number of wavelengths can be different for R_I on the one hand and q and u on the other hand. As an alternative for using q and u the code also works with Degree of Linear Polarization (DoLP=$\sqrt{q^2+u^2}$) in the measurement vector. This would ignore potential information present in the Angle of Linear Polarization but may have the advantage of having smaller measurement uncertainties. The best choice is to be made based on evaluation of retrievals on real measurements.

41 Algorithm flow, input, output

The flow of the algorithm is very similar to that of the MAP-only retrieval, except that the combined retrieval approach requires the results of the non-scattering retrieval (non. scat columns) as input for data filtering, see Fig. 41.11. Also the forward model of the iterative retrieval scheme and inversion procedure (per step) are the same as for the MAP-only and CO2I-only retrieval.

42 Cloud filtering

The cloud filtering of the simultaneous MAP+CO2I retrieval approach is not yet defined. As baseline we consider the cloud screening scheme of the CO2I-only approach, where appropriate threshold values should be defined during the CO2M commissioning phase.

43 Detailed Algorithm Setup Sequential Two-Step Retrieval MAP→CO2I

The sequential two-step retrieval first performs a MAP-only retrieval, followed by a CO2I-only retrieval, which uses the retrieved aerosol properties of the first step (MAP-only) as input. In the 2nd step (CO2I-only) certain aerosol parameters are still being fitted, together with the trace gas columns.

44 State Vector Definition

For the MAP-only step of the two-step retrieval, we use the 2-mode state vector definition of Table 25.5. The retrieved aerosol parameters serve as input for the CO2I-only step, which has the same aerosol parameters in the state vector as the first step.The corresponding state vector elements for this second step are given in Table 44.11. As follows from this table, only 2 aerosol parameters are being fitted in the 2nd step (N^c and r^c_{eff}) while all other aerosol parameters are being fixed to the values of the MAP-only retrieval.

Table 44.11: State vector elements, prior values, and corresponding values w_i that define the diagonal regularization matrix $\mathbf{H}$ by $h_{ii} = w_i^2$, for the CO2I-only step in the two-step retrieval. If $w_i = 0$, indicated by *, the corresponding parameters is fixed. σ_{ext} denotes the aerosol extinction cross-section at 550 nm.

state vector element	prior value	w_i	min - max
Retrieved aerosol parameters			
N^c	MAP-only	2.0/σ_{ext}	0.001 - n/a
r^c_{eff}	MAP-only	1.0 µm	0.2 µm - 4.0 µm
Fixed aerosol parameters			
r^f_{eff}	MAP-only	0.0*	0.02 µm - 0.30 µm
v^f_{eff}	MAP-only	0.0*	0.01 - 0.80
N^f	MAP-only	0.0*	0.001 - n/a

f^f_{sphere}	MAP-only	0.0*	0.0 - 1.0
$\alpha^f_{\mathrm{inorg}}$	MAP-only	0.0*	such that $1.3 < m_r < 1.69$
α^f_{bc}	MAP-only	0.0*	such that $1.3 < m_r < 1.69$
α^f_{oc}	MAP-only	0.0*	such that $1.3 < m_r < 1.69$
v^c_{eff}	MAP-only	0.0*	0.01 - 0.80
f^c_{sphere}	MAP-only	0.0*	0.0 - 1.0
$\alpha^c_{\mathrm{inorg}}$	MAP-only	0.0*	such that $1.3 < m_r < 1.69$
α^c_{du}	MAP-only	0.0*	such that $1.3 < m_r < 1.69$
z_{aer}	MAP-only	0.0*	-2 km - 10 km
Land surface parameters (retrieved)			
a_0(NIR, SWIR-1, SWIR-2)	0.1	0.50	0.0 - 1.0
a_1(NIR, SWIR-1, SWIR-2)	0.0	1e-3	-1e-3 - 1.e-3
a_2(NIR, SWIR-1, SWIR-2) (optional)	0.0	1e-5	-1e-5 - 1.e-5
Land surface parameters (fixed)			
k_{geo}	MAP-only	0.0*	
k_{vol}	MAP-only	0.0*	
B	MAP-only	0.0*	
ν	0.1	0.0*	
Ocean parameters			
windspeed	7 m/s	30 m/s	0.1 m/s - 100 m/s
a_0^{ocean}(NIR, SWIR-1, SWIR-2)	0.0	0.50	-0.1 - 0.1
a_1^{ocean}(NIR, SWIR-1, SWIR-2)	0.0	1e-3	-1e-3 - 1e-3
a_2^{ocean}(NIR, SWIR-1, SWIR-2)	0.0	1e-5	-1e-5 - 1e-5
Trace gas parameters			
c$_{\mathrm{co2}}$	1.0	1.0	0.0 - 2.0
c$_{\mathrm{ch4}}$	1.0	1.0	0.0 - 2.0
c$_{\mathrm{h2o}}$	1.0	1.0	0.0 - 2.0

Nuisance			
$sh_0^{s,e}$(NIR, SWIR-1, SWIR-2)	0.0	0.1	-0.1 - 0.1
$sh_1^{s,e}$(NIR, SWIR-1, SWIR-2)	0.0	0.1	-1e-4 - 1e-4
$sh_2^{s,e}$(NIR, SWIR-1, SWIR-2) (optional)	0.0	0.1	-1e-6 - 1e-6
I_{off}(NIR, SWIR-1, SWIR-2) (optional)	0.0	1e-14 (photons/cm2/nm/sr)	-1e-9 - 1e-9 (photons/cm2/nm/sr)

45 Measurement vector definition

The measurement vectors for the 1st and 2nd step are identical to the one of the MAP-only retrieval described in Sec. 24 and the CO2I-only retrieval described in Sec. 33.

46 Algorithm flow, input, output

The flow of the retrieval is slightly modified compared to MAP-only and CO2I-only retrievals as it performs those retrievals in 2 separate steps, see Fig. 46.12. Also the forward model of the iterative retrieval scheme and inversion procedure (per step) are the same as for the MAP-only and CO2I-only retrieval.

47 Cloud filtering

The cloud filtering of the sequential two-step retrieval approach is not yet defined. As baseline we consider the cloud screening scheme of the MAP-only retrieval, combined with the CO2I-only approach, where appropriate threshold values should be defined during the CO2M commissioning phase.

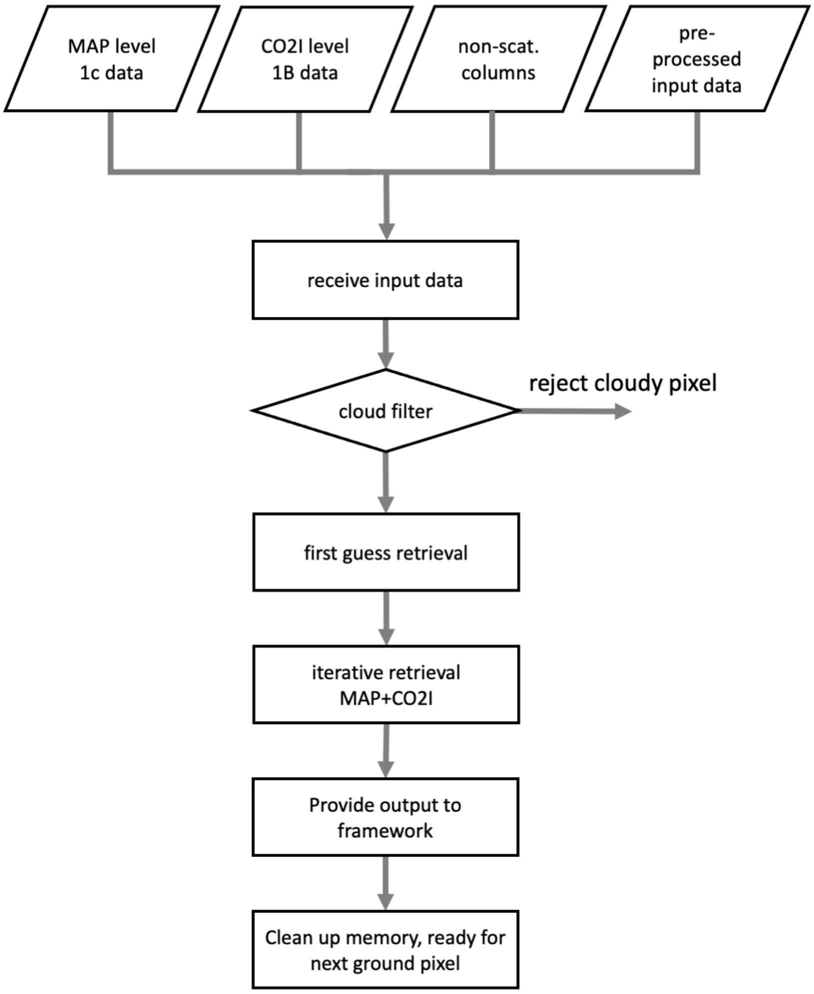

Figure 41.11: Schematic overview of the general retrieval procedure of the simultaneous CO2I+MAP retrievals. The iterative retrieval is further detailed in Fig. 27.6.

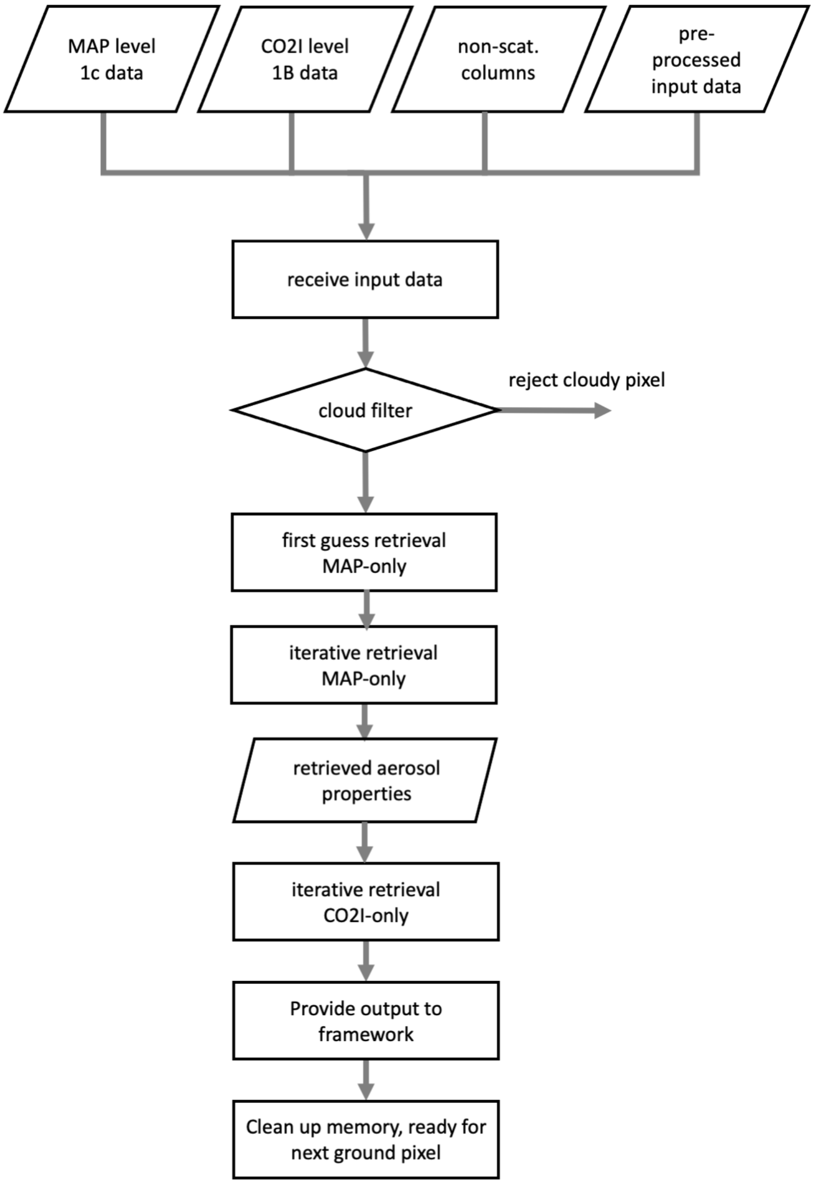

Figure 46.12: Schematic overview of the general retrieval procedure of the sequential MAP→CO2I retrievals. The iterative retrieval is further detailed in Fig. 27.6.

48 Input/Output

49 Static input for processor initialization

This section describes the static input to the processor initialization. It comprises

- the Mie/T matrix lookup tables (49.12)
- Radiative tranfer lookup tables for first guess retrievals (49.13)
- absorption cross section lookup tables (49.14)

Table 49.12: The Mie/T-matrix lookup table: $N_{\mathrm{rri}} = 22$ denotes the number of LUT node points for the real refractive index, $N_{\mathrm{iri}} = 16$ the number of node points for the imaginary refractive index, $N_{\mathrm{size}} = 41$ the number of node points for the aerosol size parameter ($\frac{2\,\pi\,r}{\lambda}$, $N_{\mathrm{ang}} = 181$ the number of node points for the scattering angle.

parameter	array size	units	comments
Mie/T-Matrix LUT			
grid RRI	N_{rri}	-	
grid IRI	N_{iri}	-	
grid size parameter	N_{size}	-	
Sphere extinction cross section	$N_{\mathrm{rri}} \times N_{\mathrm{iri}} \times N_{\mathrm{size}}$	cm^2	
Sphere absorption cross section	$N_{\mathrm{rri}} \times N_{\mathrm{iri}} \times N_{\mathrm{size}}$	cm^2	
Spheroid extinction cross section	$N_{\mathrm{rri}} \times N_{\mathrm{iri}} \times N_{\mathrm{size}}$	cm^2	
Spheroid absorption cross section	$N_{\mathrm{rri}} \times N_{\mathrm{iri}} \times N_{\mathrm{size}}$	cm^2	
Sphere scattering matrix (element 11)	$N_{\mathrm{rri}} \times N_{\mathrm{iri}} \times N_{\mathrm{size}} \times N_{\mathrm{ang}}$	-	
Sphere scattering matrix (element 22)	$N_{\mathrm{rri}} \times N_{\mathrm{iri}} \times N_{\mathrm{size}} \times N_{\mathrm{ang}}$	-	
Sphere scattering matrix (element 33)	$N_{\mathrm{rri}} \times N_{\mathrm{iri}} \times N_{\mathrm{size}} \times N_{\mathrm{ang}}$	-	
Sphere scattering matrix (element 44)	$N_{\mathrm{rri}} \times N_{\mathrm{iri}} \times N_{\mathrm{size}} \times N_{\mathrm{ang}}$	-	
Sphere scattering matrix (element 21)	$N_{\mathrm{rri}} \times N_{\mathrm{iri}} \times N_{\mathrm{size}} \times N_{\mathrm{ang}}$	-	

Sphere scattering matrix (element 43)	$N_{\mathrm{rri}} \times N_{\mathrm{iri}} \times N_{\mathrm{size}} \times N_{\mathrm{ang}}$	-	

Table 49.13: Radiative Transfer Lookup tables for land and ocean scenes

parameter	dimension	units	comments
RT LUT (1st guess retrieval) land			
wavelength grid	$N_{\mathrm{wave,lut}}$	nm	
Rayleigh cross sections	$N_{\mathrm{wave,lut}}$	cm^2	
mode effective radius	$N_{\mathrm{modes,lut}}$	µm	
mode effective variance	$N_{\mathrm{modes,lut}}$		
mode fraction spheres	$N_{\mathrm{modes,lut}}$		
mode RRI	$N_{\mathrm{modes,lut}} \times N_{\mathrm{wave,lut}}$		
mode IRI	$N_{\mathrm{modes,lut}} \times N_{\mathrm{wave,lut}}$		
viewing zenith angle grid	$N_{\mathrm{vza,lut}}$	degree	
solar zenith angle grid	$N_{\mathrm{sza,lut}}$	degree	
surface pressure grid	$N_{\mathrm{psurf,lut}}$	hPa	
AOD grid	$N_{\mathrm{aod,lut}}$	-	
BDRF A grid	$N_{\mathrm{bdrf,wave,lut}}$	-	
BDRF geo par1 grid	$N_{\mathrm{bdrf,geo1,lut}}$	-	Ross-Li kernel coefficient of RPV parameter
BDRF geo par2 grid	$N_{\mathrm{bdrf,geo2,lut}}$	-	Ross-Li kernel coefficient of RPV parameter
Stokes parameter I reflectance	$N_{\mathrm{modes,lut}} \times N_{\mathrm{wave,lut}} \times N_{\mathrm{vza,lut}} \times N_{\mathrm{sza,lut}} \times N_{\mathrm{fou,lut}} \times N_{\mathrm{psurf,lut}} \times N_{\mathrm{aod,lut}} \times N_{\mathrm{bdrf,wave,lut}} \times N_{\mathrm{bdrf,geo1,lut}} \times N_{\mathrm{bdrf,geo2,lut}}$	-	$N_{\mathrm{fou,lut}}$ Fourier coefficients
Stokes parameter Q reflectance	$N_{\mathrm{modes,lut}} \times N_{\mathrm{wave,lut}} \times N_{\mathrm{vza,lut}} \times N_{\mathrm{sza,lut}} \times N_{\mathrm{fou,lut}} \times N_{\mathrm{psurf,lut}} \times N_{\mathrm{aod,lut}} \times N_{\mathrm{bdrf,wave,lut}}$	-	$N_{\mathrm{fou,lut}}$ Fourier coefficients

Stokes parameter U reflectance	$N_{\mathrm{modes,lut}} \times N_{\mathrm{wave,lut}} \times N_{\mathrm{vza,lut}} \times N_{\mathrm{sza,lut}} \times N_{\mathrm{fou,lut}} \times N_{\mathrm{psurf,lut}} \times N_{\mathrm{aod,lut}} \times N_{\mathrm{bdrf,wave,lut}}$	-	$N_{\mathrm{fou,lut}}$ Fourier coefficients
RT LUT (1st guess retrieval) ocean			
wavelength grid	$N_{\mathrm{wave,lut}}$	nm	
Rayleigh cross sections	$N_{\mathrm{wave,lut}}$	cm^2	
mode effective radius	$N_{\mathrm{modes,lut}}$	μm	
mode effective variance	$N_{\mathrm{modes,lut}}$		
mode fraction spheres	$N_{\mathrm{modes,lut}}$		
mode RRI	$N_{\mathrm{modes,lut}} \times N_{\mathrm{wave,lut}}$		
mode IRI	$N_{\mathrm{modes,lut}} \times N_{\mathrm{wave,lut}}$		
viewing zenith angle grid	$N_{\mathrm{vza,lut}}$	degree	
solar zenith angle grid	$N_{\mathrm{sza,lut}}$	degree	
surface pressure grid	$N_{\mathrm{psurf,lut}}$	hPa	
AOD grid	$N_{\mathrm{aod,lut}}$	-	
wind speed grid	$N_{\mathrm{ws,lut}}$	m/s	
x_{chl} grid	$N_{\mathrm{chl,lut}}$		Ross-Li kernel coefficient of RPV parameter
Stokes parameter I reflectance	$N_{\mathrm{modes,lut}} \times N_{\mathrm{wave,lut}} \times N_{\mathrm{vza,lut}} \times N_{\mathrm{sza,lut}} \times N_{\mathrm{fou,lut}} \times N_{\mathrm{psurf,lut}} \times N_{\mathrm{aod,lut}} \times N_{\mathrm{ws,lut}} \times N_{\mathrm{chl,lut}}$	-	$N_{\mathrm{fou,lut}}$ Fourier coefficients
Stokes parameter Q reflectance	$N_{\mathrm{modes,lut}} \times N_{\mathrm{wave,lut}} \times N_{\mathrm{vza,lut}} \times N_{\mathrm{sza,lut}} \times N_{\mathrm{fou,lut}} \times N_{\mathrm{psurf,lut}} \times N_{\mathrm{aod,lut}} \times N_{\mathrm{ws,lut}} \times N_{\mathrm{chl,lut}}$	-	$N_{\mathrm{fou,lut}}$ Fourier coefficients
Stokes parameter U reflectance	$N_{\mathrm{modes,lut}} \times N_{\mathrm{wave,lut}} \times N_{\mathrm{vza,lut}} \times N_{\mathrm{sza,lut}} \times N_{\mathrm{fou,lut}} \times N_{\mathrm{psurf,lut}} \times N_{\mathrm{aod,lut}} \times N_{\mathrm{ws,lut}} \times N_{\mathrm{chl,lut}}$	-	$N_{\mathrm{fou,lut}}$ Fourier coefficients

Table 49.14: Trace gas absorption cross-section lookup table: $N_{\mathrm{p,xs}}$ denotes the number of pressure node points, $N_{\mathrm{T,xs}}$ the number of temperature node points, $N_{\lambda,gas,win}$, with gas is O_3, NO_2, O_4, O_2 or H_2O, CO_2, and CH_4,denotes the number of wavelengths at which cross sections are stored in a certain spectral window, $N_{\mathrm{win,gas}}$ the number of spectral windows for a certain gas.

grid pressure	$N_{\mathrm{p,xs}}$	Pa	
grid temperature	$N_{\mathrm{T,xs}}$	K	
wavenumber grid O_2	$N_{\lambda,\mathrm{wino2}} \times N_{\mathrm{wino2}}$	cm^{-1}	
wavenumber grid H_2O	$N_{\lambda,\mathrm{winh2o}} \times N_{\mathrm{winh2o}}$	cm^{-1}	
wavenumber grid O_3	$N_{\lambda,\mathrm{wino3}} \times N_{\mathrm{wino3}}$	cm^{-1}	
wavenumber grid NO_2	$N_{\lambda,\mathrm{winno2}} \times N_{\mathrm{winno2}}$	cm^{-1}	
wavenumber grid O_4	$N_{\lambda,\mathrm{wino4}} \times N_{\mathrm{wino4}}$	cm^{-1}	
cross-sections O_2	$N_{\mathrm{p,xs}} \times N_{\mathrm{T,xs}} \times N_{\lambda,\mathrm{wino2}} \times N_{\mathrm{wino2}}$	cm^2	
cross-sections H_2O	$N_{\mathrm{p,xs}} \times N_{\mathrm{T,xs}} \times N_{\lambda,\mathrm{winh2o}} \times N_{\mathrm{winh2o}}$	cm^2	
cross-sections CO_2	$N_{\mathrm{p,xs}} \times N_{\mathrm{T,xs}} \times N_{\lambda,\mathrm{winco2}} \times N_{\mathrm{winco2}}$	cm^2	
cross-sections CH_4	$N_{\mathrm{p,xs}} \times N_{\mathrm{T,xs}} \times N_{\lambda,winch4} \times N_{\mathrm{winch4}}$	cm^2	
cross-sections O_3	$N_{\mathrm{p,xs}} \times N_{\mathrm{T,xs}} \times N_{\lambda,wino3} \times N_{\mathrm{wino3}}$	cm^2	
cross-sections NO_2	$N_{\mathrm{p,xs}} \times N_{\mathrm{T,xs}} \times N_{\lambda,winno2} \times N_{\mathrm{winno2}}$	cm^2	
cross-sections O_4	$N_{\mathrm{p,xs}} \times N_{\mathrm{T,xs}} \times N_{\lambda,wino4} \times N_{\mathrm{wino4}}$	cm^2	

Table 49.15: ISRF and solar reference spectrum. $N_{\lambda,\mathrm{map}}$ is the number of MAP bands and $N_{\lambda,\mathrm{fine,isrf-map}}$ is the number of spectral samplings of the ISRF per MAP band. $N_{\lambda,\mathrm{co2i}}$ and $N_{\lambda,\mathrm{fine,isrf-xo2i}}$ are the corresponding parameters for the spectrometer ISRF, and $N_{\lambda,\mathrm{fine,\ sun}}$ is the number of spectral samples of the solar reference spectrum.

ISRF-MAP	$N_{\lambda,\mathrm{map}} \times N_{\lambda,\mathrm{fine,isrf-map}}$		see Eq. 8.6
ISRF-CO2I	$N_{\lambda,\mathrm{co2i}} \times N_{\lambda,\mathrm{fine,isrf-co2i}}$		
ISRF-CO2I-SUN	$N_{\lambda,\mathrm{co2i}} \times N_{\lambda,\mathrm{fine,isrf-co2i}}$		
solar_irradiance_model	$N_{\lambda,\mathrm{fine,\ sun}}$		

50 Static input for scan line intitialization

This section describes the static input to the initialization of the procesing of each data granule. It comprises ISRF and solar irradiance related input.

Table 50.16: Input for scanline initialization.

parameter	array size	units	comments
ISRF related			
ISRF-CO2I-SUN	$N_{\lambda,\mathrm{co2i}} \times N_{\lambda,\mathrm{fine,isrf-co2i}}$		see Eq. 8.6
solar_irradiance_model	$N_{\lambda,\mathrm{fine},sun}$		

51 Dynamic input pre-processor

52 Input pre-processor

This section describes the input to the pre-processor tool, which basically allocates the input to the CO2I (/MAP) ground pixels.

Table 52.17: Dynamic input for the pre-processor (ECMWF-based). $N_{\mathrm{alt,clim}}$ denotes the number of along-track CLIM pixels, $N_{\mathrm{act,clim}}$ denotes the number of across-track CLIM pixels, $N_{\mathrm{alt,co2i}}$ denotes the number of along-track CO2I pixels, $N_{\mathrm{act,co2i}}$ denotes the number of across-track CO2I pixels.$N_{\mathrm{lat,met}}$ and $N_{\mathrm{lon,met}}$ denote the the number of latitudes and longtitudes of the meteo data, $N_{\mathrm{lev,met}}$ denotes the number of vertical levels of the meteo data.

parameter	**array size**	**Source**	**units**
Surface pressure meteo	$N_{\mathrm{lat,met}} \times N_{\mathrm{lon,met}}$	CAMS/ ECMWF	hPa
Surface elevation meteo	$N_{\mathrm{lat,met}} \times N_{\mathrm{lon,met}}$	CAMS/ ECMWF	m
Hybrid coefficients A, B	$N_{\mathrm{lev,met}}$	CAMS/ ECMWF	-
Temperature profile	$N_{\mathrm{lat,met}} \times N_{\mathrm{lon,met}} \times N_{\mathrm{lev,met}}$	CAMS/ ECMWF	K
Geopotenital height	$N_{\mathrm{lat,met}} \times N_{\mathrm{lon,met}} \times N_{\mathrm{lev,met}}$	CAMS/ ECMWF	m
Specific humidity	$N_{\mathrm{lat,met}} \times N_{\mathrm{lon,met}} \times N_{\mathrm{lev,met}}$	CAMS/ ECMWF	kg /kg

O_3 profile	$N_{\mathrm{lat,met}} \times N_{\mathrm{lon,met}} \times N_{\mathrm{lev,met}}$	CAMS/ ECMWF	TBD
CO_2 profile	$N_{\mathrm{lat,met}} \times N_{\mathrm{lon,met}} \times N_{\mathrm{lev,met}}$	CAMS/ ECMWF	TBD
CH_4 profile	$N_{\mathrm{lat,met}} \times N_{\mathrm{lon,met}} \times N_{\mathrm{lev,met}}$	CAMS/ ECMWF	TBD
NO_2 profile	$N_{\mathrm{lat,met}} \times N_{\mathrm{lon,met}} \times N_{\mathrm{lev,met}}$	CAMS/ ECMWF	TBD
Cloud fraction	$N_{\mathrm{alt,clim}} \times N_{\mathrm{act,clim}}$	CLIM	-
latitude CLIM	$N_{\mathrm{alt,clim}} \times N_{\mathrm{act,clim}}$	CLIM L2	degrees
longitude CLIM	$N_{\mathrm{alt,clim}} \times N_{\mathrm{act,clim}}$	CLIM L2	degrees
latitude CO2I	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,co2i}}$	CO2I L1B	degrees
longitude CO2I	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,co2i}}$	CO2I L1B	degrees
latitude MAP	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,co2i}}$	CO2I L1B	degrees
longitude MAP	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,co2i}}$	CO2I L1B	degrees
altitude	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,co2i}}$	CO2I L1B	m
altitude variation	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,co2i}}$	CO2I L1B	m
land-water mask	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,co2i}}$	CO2I L1B -	

53 MAP L1C data

Table 53.18: Dynamic input of MAP L1C data. $N_{\mathrm{alt,map}}$ denotes the number of along-track MAP pixels, $N_{\mathrm{act,map}}$ is the number of across-track MAP pixels, N_{vp} is the number of viewing angles of the MAP, $N_{\lambda,\mathrm{pol}}$ indicates the number of spectral samples for polarization of the MAP, $N_{\lambda,\mathrm{rad}}$ is the number of spectral samples for radiance of the MAP.

parameter	**array size**	**units**	**comment**
MAP L1C file			
latitude MAP	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}}$	degree	
longitude MAP	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}}$	degree	
date and time of measurement	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}}$	Julian day	mean of all aggregated measurements

sensor_azimuth	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}} \times N_{\mathrm{vp}}$	degree	clockwise from north
sensor_zenith	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}} \times N_{\mathrm{vp}}$	degree	
solar_azimuth	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}} \times N_{\mathrm{vp}}$	degree	clockwise from north
solar_zenith	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}} \times N_{\mathrm{vp}}$	degree	
I_polsample	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}} \times N_{\mathrm{vp}} \times N_{\lambda,\mathrm{pol}}$	Wm^{-2}sr^{-1} μm^{-1}	intensity at spectral sampling and resolution of polarization measurements.
I_polsample_noise	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}} \times N_{\mathrm{vp}} \times N_{\lambda,\mathrm{pol}}$	Wm^{-2}sr^{-1} μm^{-1}	
Q	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}} \times N_{\mathrm{vp}} \times N_{\lambda,\mathrm{pol}}$	Wm^{-2}sr^{-1} μm^{-1}	
Q_noise	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}} \times N_{\mathrm{vp}} \times N_{\lambda,\mathrm{pol}}$	Wm^{-2}sr^{-1} μm^{-1}	
U	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}} \times N_{\mathrm{vp}} \times N_{\lambda,\mathrm{pol}}$	Wm^{-2}sr^{-1} μm^{-1}	
U_noise	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}} \times N_{\mathrm{vp}} \times N_{\lambda,\mathrm{pol}}$	Wm^{-2}sr^{-1} μm^{-1}	
I	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}} \times N_{\mathrm{vp}} \times N_{\lambda,\mathrm{rad}}$	Wm^{-2}sr^{-1} μm^{-1}	at native sampling/resolution
I_noise	$N_{\mathrm{alt,map}} \times N_{\mathrm{act,map}} \times N_{\mathrm{vp}} \times N_{\lambda,\mathrm{rad}}$	Wm^{-2}sr^{-1} μm^{-1}	
intensity_F0	$N_{\mathrm{vp}} \times N_{\lambda,\mathrm{rad}}$	Wm^{-2}	model solar spectrum at sampling/resolution of intensity wavelength grid
polarization_F0	$N_{\mathrm{vp}} \times N_{\lambda,\mathrm{pol}}$	Wm^{-2}	model solar spectrum at sampling/resolution of polarization wavelength grid
sun_earth_distance		AU	
polarization_wavelengths	$N_{\lambda,\mathrm{pol}}$	nm	
polarization_bandpasses	$N_{\lambda,\mathrm{pol}}$	nm	FWHM
intensity_wavelengths	$N_{\lambda,\mathrm{rad}}$	nm	
intensity_bandpasses	$N_{\lambda,\mathrm{rad}}$	nm	FWHM

54 Input from CO2I L1B data

Table 54.19: Dynamic input for the aerosol L2 processor (CO2I L1B file). $N_{\mathrm{alt,spec}}$ denotes the number of along-track CO2I pixels, $N_{\mathrm{act,spec}}$ denotes the number of across-track CO2I pixels. N_{win} denotes the number of spectral windows of the CO2I (e.g. NIR, SWIR-1, SWIR-2), $N_{\lambda,\mathrm{spec},win}$ denotes the number of wavelength pixels per spectral window.

parameter	**array size**	**units**	**comment**
CO2I L1B file			
radiance spectrum	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,spec}} \times N_{\mathrm{win}} \times N_{\lambda,\mathrm{spec},win}$	Wm^{-2}sr^{-1} μm^{-1}	
radiance spectrum noise	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,spec}} \times N_{\mathrm{win}} \times N_{\lambda,\mathrm{spec,win}}$	Wm^{-2}sr^{-1} μm^{-1}	
radiance wavelengths	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,spec}} \times N_{\mathrm{win}} \times N_{\lambda,\mathrm{spec},win}$	nm	
latitude center	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,spec}}$	degree	
longitude center	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,sec}}$	degree	
latitude corner	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,spec}} \times 4$	degree	
longitude corner	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,spec}} \times 4$	degree	
date and time of measurement	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,spec}}$	Julian day	
sensor_azimuth	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,spec}}$	degree	clockwise from north
sensor_zenith	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,spec}}$	degree	
solar_azimuth	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,spec}}$	degree	clockwise from north
solar_zenith	$N_{\mathrm{alt,spec}} \times N_{\mathrm{act,spec}}$	degree	

55 RemoTAP output

Table 55.20: MAP related output of RemoTAP. $N_{\mathrm{wave,optic}}$ denotes the number of wavelengths at which optical properties are computed, $N_{\mathrm{wave,retr}}$ denotes the number of wavelength bands used in the retrieval, N_{modes} the number of aerosol size modes used in the retrieval, N_{pix} the number of spatial pixels for which a retrieval has been performed.

parameter	dimension	units	range	comments
Static output				

wavelengths optical properties	$N_{\mathrm{wave,optic}}$	nm	350-2200	wavelengths at which optical properties (AOD, SSA) are given.
wavelengths retrieval	$N_{\mathrm{wave,retr}}$	nm	410-865	wavelengths used in the retrieval.
number of modes	1	-	1-20	N_{modes}
effective radius flag	N_{modes}	-	0 or 1	1 if fitted, 0 if not fitted
effective variance flag	N_{modes}	-	0 or 1	1 if fitted, 0 if not fitted
real refractive index flag	N_{modes}	-	0 or 1	1 if fitted, 0 if not fitted
imaginary refractive index flag	N_{modes}	-	0 or 1	1 if fitted, 0 if not fitted
fraction of spheres flag	N_{modes}	-	0 or 1	1 if fitted, 0 if not fitted
Aerosol Layer Height flag	N_{modes}	-	0 or 1	1 if fitted, 0 if not fitted
Cirrus flag	N_{modes}	-	0 or 1	1 if cirrus, 0 if not cirrus
Surface brdf model	1	-	n/a	Rahman-Pinkty-Verstraete or Ross-Li
Surface bpdf model	1	-	n/a	Maignan B or Maignan B and ν
Dynamic output				
AOD	$N_{\mathrm{pix}} \times N_{\mathrm{wave,optic}}$	-	0-10	
AOD uncertainty	$N_{\mathrm{pix}} \times N_{\mathrm{wave,optic}}$	-	0-10	1-σ
SSA	$N_{\mathrm{pix}} \times N_{\mathrm{wave,optic}}$	-	0-1	
SSA uncertainty	$N_{\mathrm{pix}} \times N_{\mathrm{wave,optic}}$	-	0-1	1-σ
effective radius	$N_{\mathrm{pix}} \times N_{\mathrm{mode}}$	µm	0.01-10.00	-
effective radius uncertainty	$N_{\mathrm{pix}} \times N_{\mathrm{mode}}$	µm	0.01-10.00	1-σ

effective variance	$N_{\rm pix} \times N_{\rm mode}$		0.01-1.0	-
effective variance uncertainty	$N_{\rm pix} \times N_{\rm mode}$		0.01-1.0	1-σ
real refractive index	$N_{\rm pix} \times N_{\rm mode} \times N_{\rm wave,optic}$		1.30-1.70	-
real refractive index uncertainty	$N_{\rm pix} \times N_{\rm mode} \times N_{\rm wave,optic}$		0.0-0.20	-
fraction spheres	$N_{\rm pix} \times N_{\rm mode}$		0-1	-
fraction spheres uncertainty	$N_{\rm pix} \times N_{\rm mode}$		0-1	1-σ
Aerosol Layer Height (ALH)	$N_{\rm pix} \times N_{\rm mode}$	km	0-20	-
ALH uncertainty	$N_{\rm pix} \times N_{\rm mode}$	km	0-20	1-σ
mode AOD	$N_{\rm pix} \times N_{\rm mode} \times N_{\rm wave,optic}$	-	0-10	
mode AOD uncertainty	$N_{\rm pix} \times N_{\rm mode} \times N_{\rm wave,optic}$	-	0-10	1-σ
mode column number	$N_{\rm pix} \times N_{\rm mode}$	m^{-2}	0-10^{16}	-
mode column number uncertainty	$N_{\rm pix} \times N_{\rm mode}$	m^{-2}	0-10^{17}	1-σ
brdf scaling parameter	$N_{\rm pix} \times N_{\rm wave,retr}$	-	0-1	-
brdf geometric parameters	$N_{\rm pix} \times 2$	-	0-1.5	-
surface polarization parameters	$N_{\rm pix} \times 2$	-	0-15	B and ν
chi squared of fit	$N_{\rm pix}$	-	0-1000	
quality flag	$N_{\rm pix}$	-	0-100	0 for good retrievals. Other TBD
cloud fraction	$N_{\rm pix}$	-	0-1	
solar zenith angle	$N_{\rm pix} \times N_{\rm vp}$	degree	0-80	
viewing zenith angle	$N_{\rm pix} \times N_{\rm vp}$	degree	0-70	
relative azimuth angle	$N_{\rm pix} \times N_{\rm vp}$	degree	0-70	

scattering angle	$N_{\mathrm{pix}} \times N_{\mathrm{vp}}$	degree	0-180	
Aerosol degrees of freedom for signal	N_{pix}	-	$1-3$	
Processing quality flag	PQF	-	-	

Table 55.21: CO2I related output of RemoTAP. $N_{\mathrm{wave,co2i}}$ denotes the number of CO2I bands which are used in the retrieval, and N_{pix} the number of spatial pixels for which a retrieval has been performed.

parameter	dimension	units	range	comments
Static output				
CO2I bands	$N_{\mathrm{band,co2i}} \times 2$	nm	747-2090	start and end wavelengths of CO2I bands.
Dynamic output				
Pressure grid	$N_{\mathrm{pix}}\times\ N_{\mathrm{lay,retr}}$	hPa	Pressure at layer boundaries	
altitude grid	$N_{\mathrm{pix}}\times\ N_{\mathrm{lay,retr}}$	m	$0 - 1.0 \cdot 10^5$	
Dry air column	N_{pix}	mol / m^2	$3 \cdot 10^5 - 4 \cdot 10^5$	
A-priori CH_4 profile	$N_{\mathrm{pix}}\times\ N_{\mathrm{lay,retr}}$	mol / m^2	$0 - 0.1$	
A-priori CO_2 profile	$N_{\mathrm{pix}}\times\ N_{\mathrm{lay,retr}}$	mol / m^2	$0 - 20$	
A-priori O_2 column	N_{pix}	mol / m^2	$7 \cdot 10^4 - 8 \cdot 10^4$	
A-priori H_2O column	N_{pix}	mol / m^2	$0 - 5 \cdot 10^3$	
Proxy XCH_4	N_{pix}	ppb	$1750 - 1850$	
Proxy XCH_4 precision	N_{pix}	ppb	$0 - 40$	1-σ, only includes retrieval noise
Physics XCH_4	N_{pix}	ppb	$1750 - 1850$	
Physics XCH_4 precision	N_{pix}	ppb	$0 - 40$	1-σ, only includes retrieval noise
Physics CH_4 column averaging kernel	$N_{\mathrm{pix}}\times\ N_{\mathrm{lay,retr}}$	1	$0 - 3$	
Physics XCO_2	N_{pix}	ppm	$300 - 600$	
Physics XCO_2 precision	N_{pix}	ppm	$0 - 10$	1-σ, only includes retrieval noise
Physics CO_2 column averaging kernel	$N_{\mathrm{pix}}\times\ N_{\mathrm{lay,retr}}$	1	$0 - 3$	
H_2O column	N_{pix}	mol / m^2	$0 - 1 \cdot 10^4$	
CO_2 column	N_{pix}	mol / m^2	$0 - 300$	

CH_4 column	N_{pix}	mol / m^2	$0-1$	
Temperature offset	N_{pix}	K	$\pm 0-5$	
Number of iterations	N_{pix}	-	$0-30$	
surface brdf scaling parameter	$N_{\mathrm{pix}} \times N_{\mathrm{band,co2i}}$	1	$0-1$	
a1 surface brdf scaling parameter linear dependence	$N_{\mathrm{pix}} \times N_{\mathrm{band,co2i}}$	1	$\pm 0-1 \cdot 10^{-3}$	
a2 surface brdf scaling parameter quadratic dependence	$N_{\mathrm{pix}} \times N_{\mathrm{band,co2i}}$	1	$\pm 0-1 \cdot 10^{-5}$	
spectral shift	$N_{\mathrm{pix}} \times N_{\mathrm{band,co2i}}$	nm	$\pm 0-0.1$	
spsh1 spectral shift linear dependence	$N_{\mathrm{pix}} \times N_{\mathrm{band,co2i}}$	nm	$\pm 0-0.1$	
spsh2 spectral shift quadratic dependence	$N_{\mathrm{pix}} \times N_{\mathrm{band,co2i}}$	nm	$\pm 0-0.1$	
spectral shift sun	$N_{\mathrm{pix}} \times N_{\mathrm{band,co2i}}$	nm	$\pm 0-0.1$	
spshs1 spectral shift sun linear dependence	$N_{\mathrm{pix}} \times N_{\mathrm{band,co2i}}$	nm	$\pm 0-0.1$	
spshs2 spectral shift sun quadratic dependence	$N_{\mathrm{pix}} \times N_{\mathrm{band,co2i}}$	nm	$\pm 0-0.1$	
Intensity offset (reflectance)	$N_{\mathrm{pix}} \times N_{\mathrm{band,co2i}}$	1	$\pm 0-0.001$	
NIR χ^2 of the spectral fit	N_{pix}	-	tbd	
SWIR-1 χ^2 of the spectral fit	N_{pix}	-	tbd	
SWIR-2 χ^2 of the spectral fit	N_{pix}	-	tbd	
Total reduced χ^2 of the spectral fit	N_{pix}	-	tbd	
Total degrees of freedom for signal	N_{pix}	-	tbd	
CH_4 degrees of freedom for signal	N_{pix}	-	tbd	

CO_2 degrees of freedom for signal	N_{pix}	-	tbd	
H_2O degrees of freedom for signal	N_{pix}	-	tbd	
Processing quality flag	PQF	-	-	
Aerosol size parameter	N_{pix}	-	$1-5$	Only for spectrometer-only retrievals
Aerosol particle column	N_{pix}	mol / m^2	$0-10^{-6}$	Only for spectrometer-only retrievals
Aerosol layer height	N_{pix}	m	$0-15000$	Only for spectrometer-only retrievals
Aerosol degrees of freedom for signal	N_{pix}	-	$1-3$	Only for spectrometer-only retrievals

Table 55.22: Non-scattering retrieval output of RemoTAP related to CO2I. $N_{\mathrm{band,ns}}$ denotes the total number of CO2I bands used in the non-scattering retrieval, $N_{\mathrm{band,ns,o2}}$, $N_{\mathrm{band,ns,h2o}}$, $N_{\mathrm{band,ns,ch4}}$, $N_{\mathrm{band,ns,co2}}$ the numbers bands used specifically in the retrieval of O_2, H_2O, CH_4 and CO_2, respectively, and N_{pix} the number of spatial pixels for which a retrieval has been performed.

parameter	dimension	units	range	comments
Static output				
non-scattering bands	$N_{\mathrm{band,ns}} \times 2$	nm	747-2090	start and end wavelengths of bands for non-scattering retrieval.
non-scattering bands for O_2	$N_{\mathrm{band,ns,o2}} \times 2$	nm	747-2090	start and end wavelengths of bands for non-scattering retrieval for O_2.
non-scattering bands for H_2O	$N_{\mathrm{band,ns,h2o}} \times 2$	nm	747-2090	start and end wavelengths of bands for non-scattering retrieval for H_2O.
non-scattering bands for CH_4	$N_{\mathrm{band,ns,ch4}} \times 2$	nm	747-2090	start and end wavelengths of bands for non-scattering retrieval for CH_4.
non-scattering bands for CO_2	$N_{\mathrm{band,ns,co2}} \times 2$	nm	747-2090	start and end wavelengths of bands for non-scattering retrieval for CO_2.
Dynamic output				
Non-scattering O_2 column	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns,o2}}$	mol / m^2	$0 - 1 \cdot 10^5$	
Non-scattering H_2O column	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns,h2o}}$	mol / m^2	$0 - 1 \cdot 10^4$	
Non-scattering CH_4 column	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns,ch4}}$	mol / m^2	$0 - 1$	
Non-scattering CO_2 column	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns,co2}}$	mol / m^2	$0 - 300$	
Non-scattering CH_4 column averaging kernel	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns,ch4}} \times N_{\mathrm{lay,retr}}$	1	$0 - 3$	

Non-scattering CO_2 column averaging kernel	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns,co2}} \times N_{\mathrm{lay,retr}}$	1	$0-3$	
Proxy XCO_2	N_{pix}	ppm	$300-600$	
Proxy XCO_2 precision	N_{pix}	ppm	$0-10$	1-σ, only includes retrieval noise
surface brdf scaling parameter	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns}}$	1	$0-1$	
a1 surface brdf scaling parameter linear dependence	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns}}$	1	$\pm 0-1\cdot 10^{-3}$	
a2 surface brdf scaling parameter quadratic dependence	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns}}$	1	$\pm 0-1\cdot 10^{-5}$	
spectral shift	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns}}$	nm	$\pm 0-0.1$	
spsh1 spectral shift linear dependence	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns}}$	nm	$\pm 0-0.1$	
spsh2 spectral shift quadratic dependence	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns}}$	nm	$\pm 0-0.1$	
spectral shift sun	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns}}$	nm	$\pm 0-0.1$	
spshs1 spectral shift sun linear dependence	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns}}$	nm	$\pm 0-0.1$	
spshs2 spectral shift sun quadratic dependence	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns}}$	nm	$\pm 0-0.1$	
Intensity offset (reflectance)	$N_{\mathrm{pix}} \times N_{\mathrm{band,ns}}$	1	$\pm 0-0.001$	

56 A Posteriori data handling and bias correction

57 Output quality assessment and assurance

To assess and assure the quality of the output variables, we propose to use a set of processing quality flags (PQFs), which are an important element of the dynamic algorithm output. The flags indicate the status, errors, data filtering and warnings to evaluate the retrieval quality. A first proposal of the PQFs is summarized in Tab. 57.23, which we expect to optimize and adapted during future developments of the software.

58 Bias correction

The CO2M mission as well as the RemoTAP software is designed to minimize biases on the XCO$_2$ and XCH$_4$ data product, which is inferred from a combination of the CO2I spectrometer and MAP measurements. For CO2M payload instruments compliant with the mission requirements the RemoTAP XCO$_2$ and XCH$_4$ data accuracy and precision is also compliant with the mission objective without any a posteriori correction. However, based on experience with earlier missions, we consider to implement an a posteriori correction for systematic errors in RemoTAP. Here, biases are a result of instrument and algorithm specific features and an overarching bias correction valid for multiple algorithms is less likely to work. The CO2M mission payload, in turn, is unique both on the level of the spectrometer CO2I and in its combination with a multiangle polarimeter. We expect that the data accuracy will outperform those of current $\mathrm{CO_2}$ missions in space. Thus, it is fair to assume that in case a bias correction is required, it should be simple and of little complexity. This drives our current preliminary design of a potential bias correction, which is based on experience with previous data sets.

59 Single pixel bias correction (input / output definition and preparation)

Biases in the XCO$_2$ and XCH$_4$ data product will be identified during the validation with

- independent on-ground and aircraft measurements,
- spatial dependence of the XCO$_2$ and XCH$_4$ data product over regions with no significant emission sources.

Both approaches require an analysis of sufficient large data sets covering CO2M observations over a period that is sufficient to capture seasonal bias dependencies. Obviously, for real time data processing, a basic assumption is that any XCO$_2$ and XCH$_4$ bias correction has no significant time dependence over periods less than a year.

Figures 59.13 and 59.14 shows the XCO$_2$ bias of GOSAT and OCO-2 data with respect to XCO$_2$ ground based measurements of the TCCON network as discussed by Guerlet et al. 2013[29] and Wu et al. 2018[30]. To derive a bias correction, correlations with selected parameters of the state vector are depicted. For GOSAT retrievals, correlations are largest with the inverse of the aerosol size parameter α_s and the product of the inferred scattering layer height and aerosol scattering optical depth. However, for OCO-2 data, the correlation with the χ^2 of the fitting residuals is most striking. Based on this findings, both products are corrected linearly, where the linear bias function is determined from the validation data.

Moreover, Wu et al., 2018[30] evaluated the across track dependence of the retrieved XCO$_2$ amount, assuming that under certain conditions a constant XCO$_2$ mixing ratio should be observed on spatial scales of the OCO-2 swath. Thus, any measured variability can be used to correct for a swath dependent XCO_2 bias. With an imaging spectrometer, this basic idea can be exploit even further. Analyzing Sentinel-5 Precursor XCH_4 data, Lorente et al, 2021,[35] evaluated XCH_4 retrievals of areas, where only minor sources of methane are expected. The variability of the surface brightness within these areas can so be used to determine a bias correction as a function of the retrieved surface albedo in a certain spectral band. Combining selected areas of different brightness, the entire albedo range could be covered and so Lorente et al., 2021, derived an albedo dependent bias correction relative to a reference albedo. Both, the swath dependent bias correction of Wu et al, 2018[30] and the small area approach of Lorente et al, 2021[35] do not depend on external validation measurement, but only require a sufficient amount of data for statistically significant data analyses.

For CO2M, we should not decide in advance on the specific method to derive a bias correction. Moreover, the analysis to derive a bias correction in this way is not embedded in the retrieval software but must be performed by 'external' tools. Therefore, we propose to implement a general module, which has the capability for a multi-polynomial correction of single observations as a function of selected parameter of the state vector and the measurement. The proposed processing flow of the bias correction is depicted in Fig. 59.17.

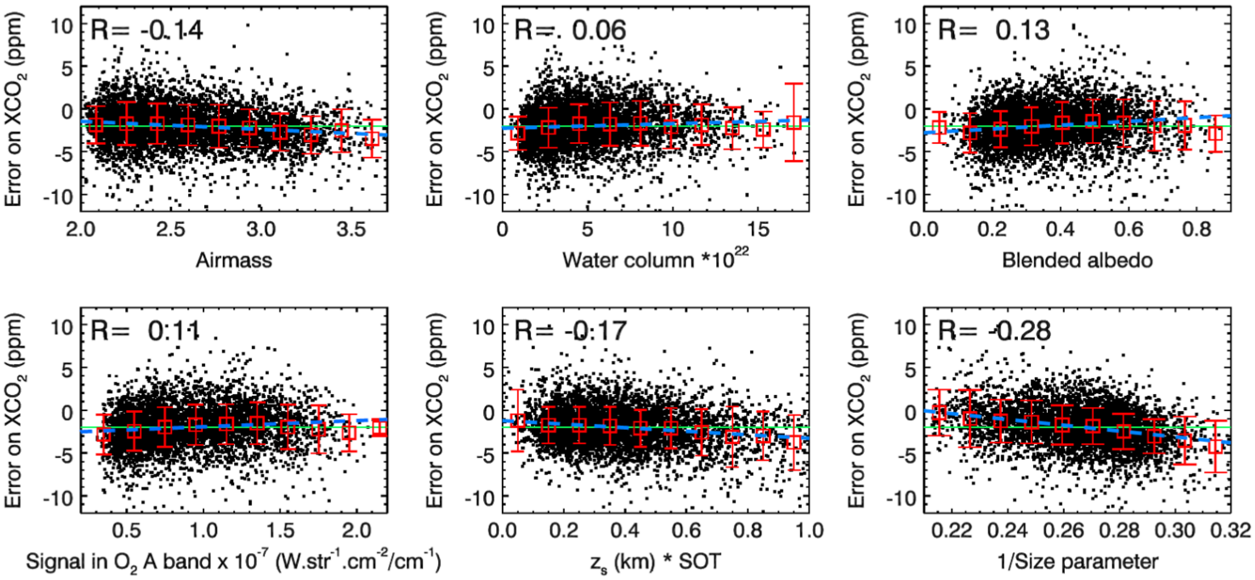

Figure 59.13: Error on XCO_2, defined as the difference between colocated GOSAT and TCCON retrievals, as a function of six parameters: air mass, water column, blended albedo, signal in O2 A-band, scattering optical thickness (SOT), and the reciprocal value of the aerosol size parameter $1/\alpha_s$. The green solid line represents the mean error, and the blue dashed line is a linear regression fit to the data. Mean values within 10 bins are shown in red squares along with 1-σ standard deviation of the mean in each bin. Correlation R with each variable is given. (Guerlet et al., 2013, Fig. 11[29])

60 Bias correction with multi-pixel interdependence

Currently, we do not foresee a bias correction that requires simultaneously data from different CO2M observations.

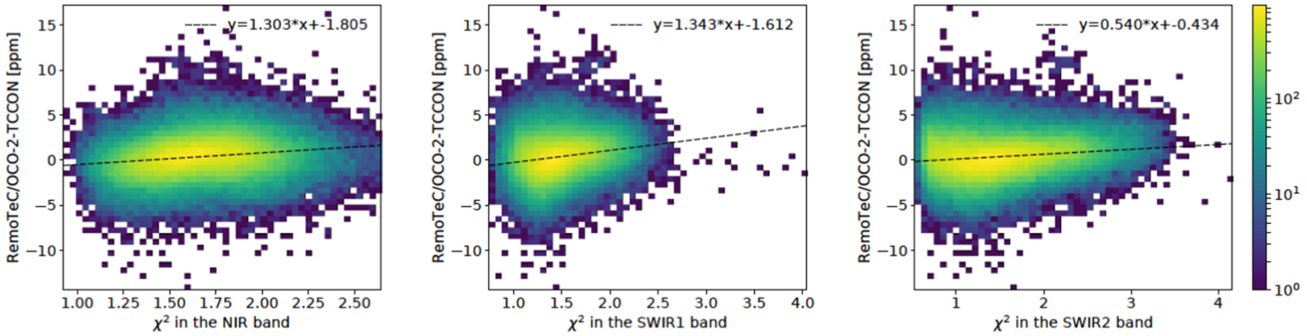

Figure 59.14: Error on OCO-2 XCO_2 retrievals as a function of the goodness of fit in the NIR, SWIR-1 and SWIR-2 bands. Different colors represent the frequency of point occurrence. The dashed line is a linear regression fit to the data (Wu et al. 2018, Fig 4)[30]

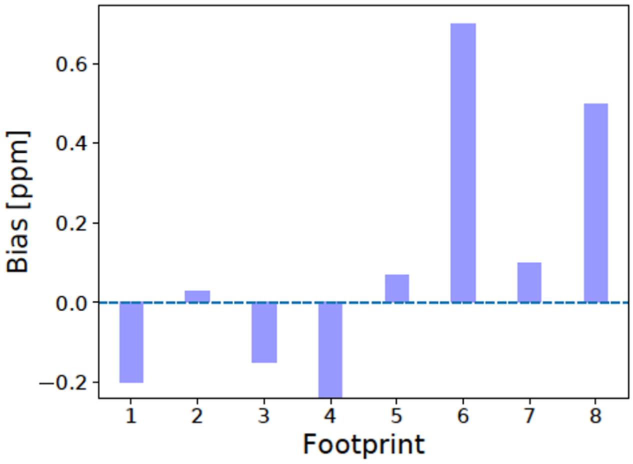

Figure 59.15: Estimated swath-dependent OCO-2 XCO_2 biases using Target mode observations.(Wu et al. 2018, Fig 3[30])

Table 57.23: Processing Quality Flags (PQFs) of RemoTeC. Four types of flags are provided: status (S), error (E), filter (F) and warning (W)

Short name	**Type**	**Description**
successfull_retrieval	S	No failures, output contains value. Warnings still possible.
low_signal_filter	F	Scene too dark to perform a retrieval.
sza_range_filter	F	Solar zenith angle out of range.
vza_range_error	F	Viewing zenith angle out of range.
cloud_filter_band	F	Scene flagged as cloudy by internal non-scat. retrievals
cloud_filter_clim	F	Scene flagged as cloudy from CLIM
cloud_filter_map	F	Scene flagged as cloudy from MAP
altitude_consistency_filter	F	Too large difference between model altitude and scene altitude value.
altitude_roughness_filter	F	Too large standard deviation of altitude in DEM.
clim_scene_heterogeneity_filter	F	Scene flagged as heterogenous.
clim_cirrus_filter	F	Scene flagged as cirrus-contaminated from dedicated cirrus CLIM input data.
ocean_filter	F	The ground pixel is over ocean (and ocean glint retrievals are not switched on or no glint geometry).
input_spectrum_missing	E	Too few non-flagged spectral pixels in both radiance and irradiance to perform retrieval.
filter_retrieval_failure	E	A retrieval used for filtering failed.
numerical_error	E	General fatal numerical error occurred during inversion.
lut_range_error	E	Case out of range of lookup table and no extrapolation is desired.
convergence_error	E	Inversion did not converge.
max_optical_thickness_error	E	Maximum aerosol optical thickness exceeded during iterations.
aerosol_boundary_error	E	Boundary hit of aerosol parameters at last iteration.
boundary_hit_error	E	Fatal boundary hit during iterations.
wavelength_calibration_warning	W	Offset from wavelength fit is larger than threshold from configuration.

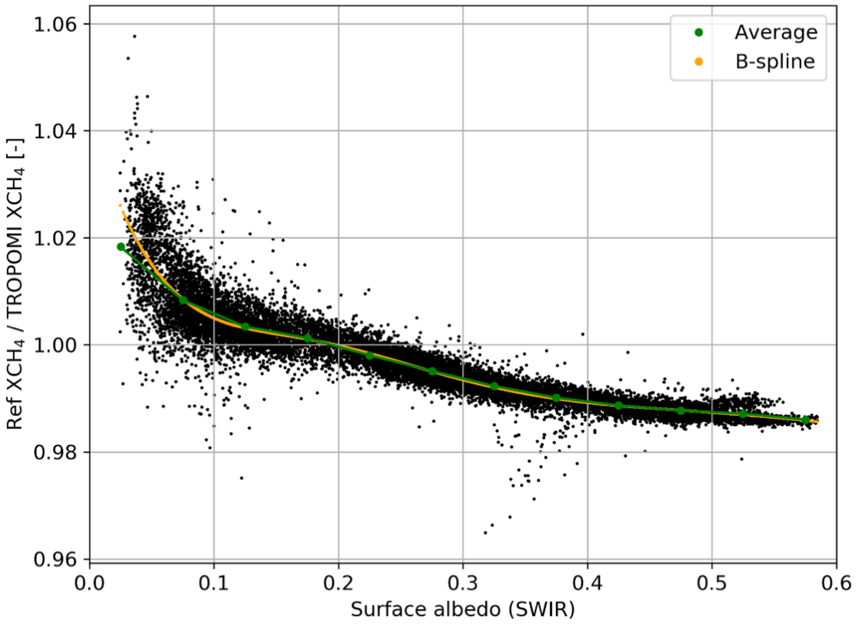

Figure 59.16: Ratio of reference XCH_4 to TROPOMI XCH_4 as a function of the retrieved surface albedo. Green dots show the average ratio in 0.05 albedo bins, and the orange line shows the B-spline fit used to derive the bias correction. Data are averaged from 1 January 2018 until 31 December (Lorente et al. 2021 Fig. 4.,[35]

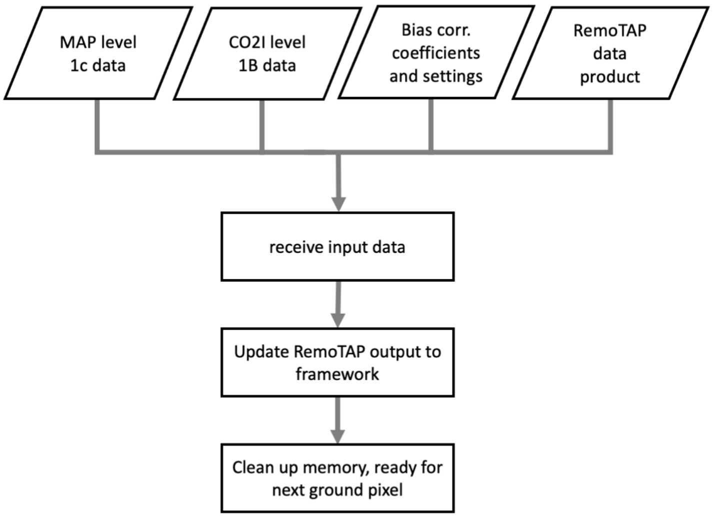

Figure 59.17: Flow diagram of the RemoTAP bias correction.

61 Data-Flows / Processing logic

The proposed RemoTAP data flow is visualized in the integrated definition data flow (IDEF) diagram in Fig. 61.18. It includes the main processing elements to be called from the higher level processing layer, called the processor frame work in the following. The orange lines in the diagram shows processing loops, which requires parallelization at frame work level to optimize the numerical performance. The module interfaces are specified by different groups of input data, which are specified in more detail in Sec. 48 Moreover, the different module can exchange data via shared memory.

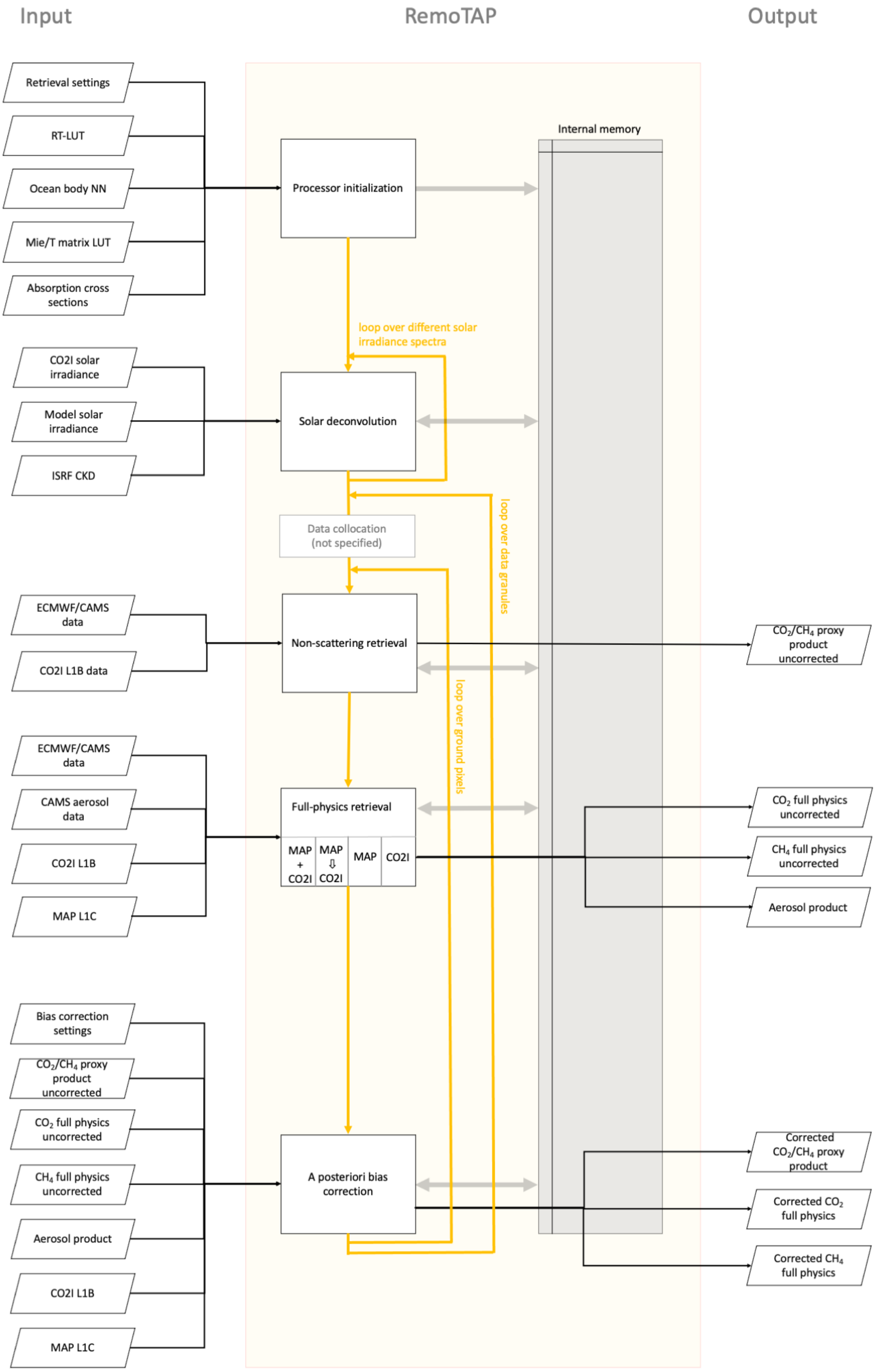

Figure 61.18: IDEF of the RemoTAP module.

62 RemoTAP performance

We investigated the performance of MAP+CO2I simultaneous retrievals and CO2I-only retrievals based on synthetic measurements created for the MAP-band and CO2I instruments over simulated CO2M orbits.

63 Instrument Description

For MAP-band, we use measurements of radiance and the degree of linear polarization (DoLP) at 6 wavelengths of 410, 440, 490, 550, 669.9, 863.4 nm and at 41 viewing angles (see Tab. 63.24) according to the CO2M requirements[3]. Here, we use a simplified MAP error model that assumes the total error on reflectance and DoLP can be described as a random error with a magnitude corresponding to the instruments accuracy requirements, which are assumed to be: 3% on reflectance I/F_0 and 0.003 on DoLP. For the 3-band CO2I spectrometer, which consider radiance spectra in a NIR band at 765 nm and two shortwave infrared bands at 1.6 μm (SWIR1) and 2.0 μm (SWIR2), with the spectral properties given in Tab. 63.25. The spectrometer noise model employs the formula

$$SNR = a_{\text{noise}} I / \sqrt{(a_{\text{noise}} I + b_{\text{noise}})}, \tag{63.128}$$

where a_{noise} and b_{noise} are spectrometer specific constants given for each spectral window[77].

Table 63.24: Characteristics of the MAP-band

Features	Setup
Number of VZAs	41
Viewing angles [degree]	$0,\ \pm3,\ \pm6, ..., \pm54,\ \pm57,\ \pm60$
Wavelengths [nm]	410, 440, 490, 550, 669.9, 863.4
Number of radiance measurements	246
Number of DoLP measurements	246
Total number of measurements	492
Radiance uncertainty	3%
DoLP uncertainty	0.003

64 Test Data Description

65 Orbit simulation

We simulated four days of synthetic measurements, each day consisting of 15 orbits of which the location (latitude, longitude) and geometry (solar zenith angle (SZA), viewing zenith angles (VZA), relative azimuth angles (AZI)) are obtained from the SRON orbit simulator[78] for CO2M orbits, as shown in Figure 65.19. The simulation uses MODIS cloud data of the year 2006 to simulate a realistic cloud mask and CALIOP cirrus data of 2007 for a cirrus mask. Retrievals are only being performed for cloud-free

Table 63.25: Characteristics of the CO2I spectrometer

Band ID	spectral range	spectral resolution	internal lbl sampling	a_{noise}	b_{noise}	telluric absorption
units	nm	nm	nm	photons^{-1} cm^2 s nm sr	1	-
NIR	747-773	0.04	0.003	2.0×10^{-8}	19600	O_2
SWIR1	1590-1675	0.10	0.01	1.32×10^{-7}	202500	H_2O, CO_2, CH_4
SWIR2	1900-2095	0.116	0.01	1.54×10^{-7}	202500	H_2O, CO_2

and cirrus-free pixels over land. To keep numerical efforts within reasonable limits, we performed the test on 1% of the pixels, which are distributed evenly over the globe.

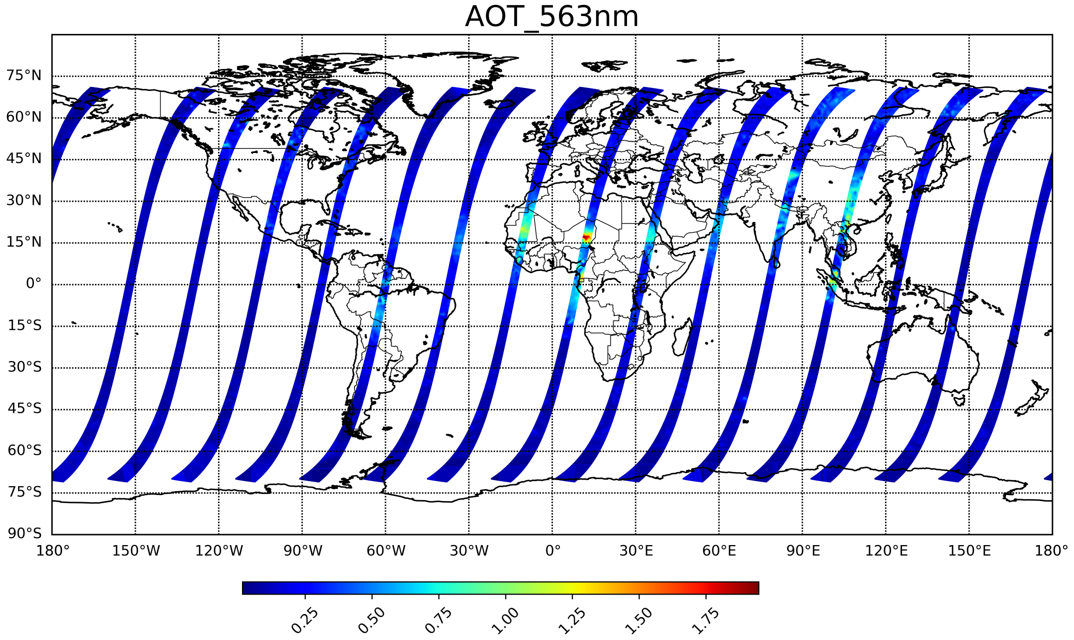

Figure 65.19: Illustration of CO2M orbits showing the aerosol optical thickness at 563 nm.

66 Ensemble Description

To generate the model atmosphere for each ground pixel different data sources are combined.

Aerosol data: We take the microphysical aerosol properties of our synthetic ensemble from the ECHAM-HAM model. ECHAM-HAM provides mass-mixing ratio in different vertical layers of the atmosphere of different aerosol species (Sulfate, Organic Carbon, Black Carbon, Dust, Sea Salt) in seven different size modes: Nucleation Soluble (NS), Aitken Soluble (KS), Accumulation Soluble (AS), Coarse Soluble (CS), Aitken Insoluble (KI), Accumulation Insoluble (AI), Coarse Insoluble (CI). Using the airmass in each model layer, we compute total mass in each layer per species per mode, which is translated into total volume in each layer per species per mode using the specific density per species. Also, for each layer the sub-column number of aerosol particles and the volume of aerosol water is provided. We sum up the different layers to obtain per mode the total column volume per species (including water), as well as the column number per mode. From the total volume V (all species together) and column number per mode N, we compute the mode radius under the assumption of a log-normal mode description (9.50):

$$r_g = \left(\frac{3V}{4\pi N} e^{-\frac{9 ln^2 \sigma_g}{2}} \right)^{\frac{1}{3}} \tag{66.129}$$

with $\sigma_g = 1.59$ for modes NS, KS, AS, KI, AI, and $\sigma_g = 2.0$ for modes CS and CI. The refractive index for each mode is obtained using a volume weighted mean of the refractive index of each species. We take the altitude of the layer with maximum sub-column number as the aerosol layer height z_{aer} for that mode and further assume a normal Gaussian altitude distribution with the center layer height z_{aer}, and width $w_0 = 2.0$ km for each mode. ECHAM-HAM does not provide information on the shape of particles. Here, we assume that dust is purely non-spherical and hence take the fraction of non-spherical particles as the volume fraction of dust.

The aerosol optical depth (AOD) of the ECHAM-HAM ensemble is severely underestimated compared to POLDER-3 satellite retrievals. Therefore, we scaled the column number of all fine and coarse modes such that they agree with the fine- and coarse mode AOD as retrieved from POLDER-3[79]. Figure 66.20(a)-(d) illustrates the total AOD and single scattering albedo (SSA) at 563 nm, real refractive index (RRI) at 550 nm for the AS mode and the CS mode.

Cirrus and cloud data: The cirrus optical depth (COD), the cirrus layer height and a cirrus mask are adapted from CALIOP data. Aspect ratio and roughness are provided by B. van Diedenhoven (personal communication, 2020), winch are inferred from POLDER observations. For the implementation in the RemoTAP, we use the gridded monthly averages calculated from both datasets. Cloud information of effective radius, optical depth, and cloud mask are taken from MODIS data.

Land and ocean data: The isotropic reflection of the land BDRF is derived from GOME-2 (335-772 nm) and MODIS (645-2130 nm) data. The directional parameters (k_{geo} and k_{vol}) of the Ross-Li model in Equation (10.58) are also from GOME-2 observation. The land surface BPDF scaling factor(x_{bpdf}^{scale}) is from POLDER-3 retrievals. For synthetic measurements over ocean, we use the chlorophyll-a concentration from MODIS and the wind speed from NCEP meteorological data. Figure 66.20(e)-(f) illustrates the total albedo at 1640nm and BPDF scaling parameter.

Trace gas data: We use the column averaged volume mixing ratio XCO_2 from the 3-hourly CAMS v18r1 data and the volume mixing ratio XCH_4 from the CAMS v17r1 data. Both data sets are averaged to a daily basis. The vertical profile is adapted from AFGL data base. Figure 66.20 (g) and (h) depicts the derived volume mixing ratio of XCO_2 and XCH_4.

Wavelenth shifts: The wavelength shift in the measurement of solar irradiance is assumed to be constant over each spectrometer band with a value equal to 0.1 of its measurement resolution. A linear dependence on the spectrum is assumed for the wavelength shift in the measurement of Earth-shine radiance, where the 0th-order term is equal to 0.05 of the measurement resolution for each band, and the 1st-order coefficient is equal to 1e-7 for all spectrometer bands.

67 Errors and Uncertainty Characterization

68 Performance analysis for the simultaneous MAP+CO2I retrievals

The state vector of MAP+CO2I retrieval case is given in section 38. We evaluate the performance of the 70th percentile best $\chi 2^2$ retrievals, and the results are shown in Fig. 68.21. We can observe that for AOD retrieval, MAP+CO2I case results in an Root-Mean-Square-Error (RMSE) of 0.0209 and small bias in terms of magnitude valued as -0.0044. XCO_2 retrieval has an RMSE of 0.6585 ppm and a

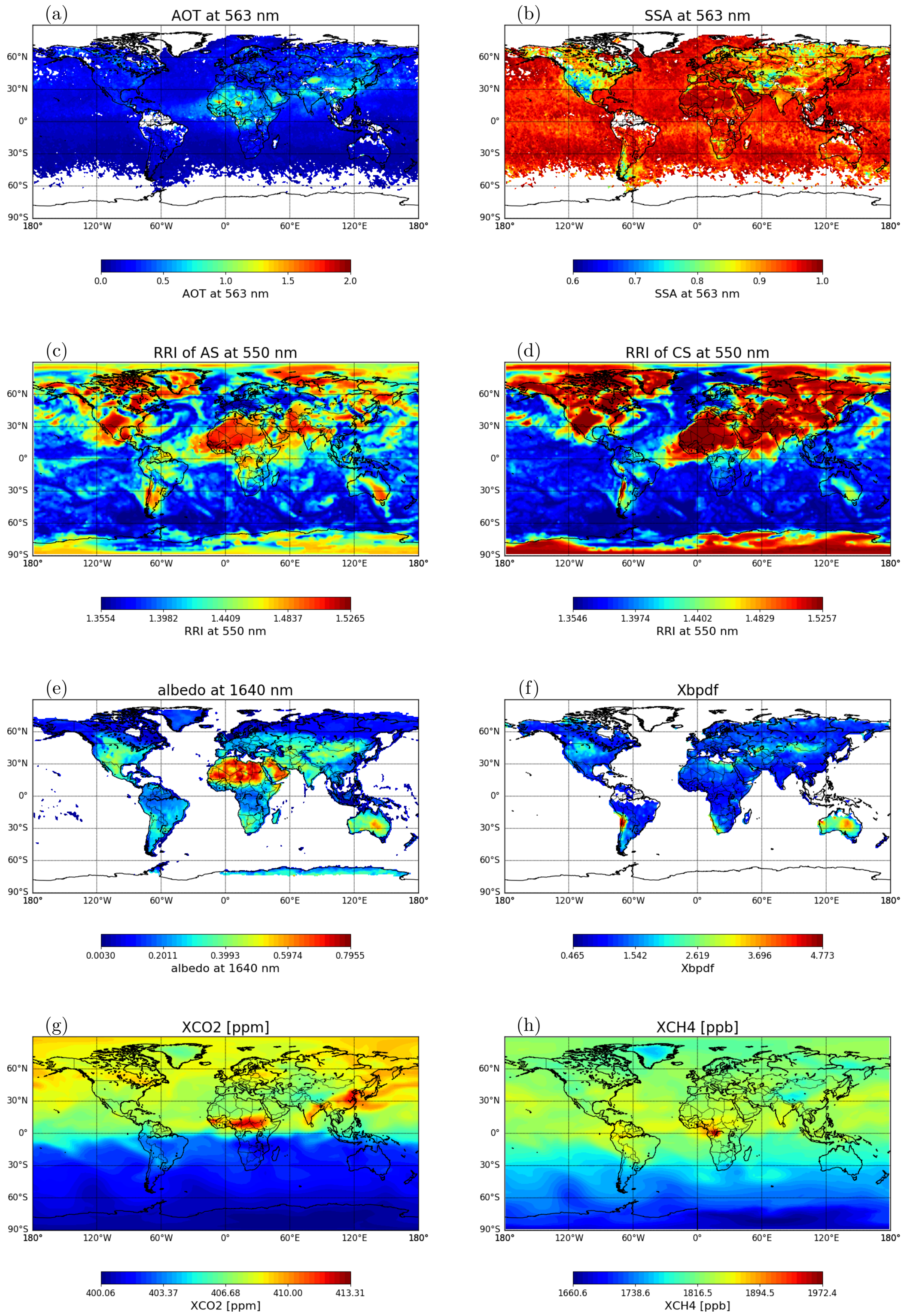

Figure 66.20: Illustration of the ensemble data on May 15 for aerosol properties (a)-(d), for surface properties (e)-(f) and for trace gases (g) and (h).

bias of 0.1096 ppm. The numbers are compliant with the mission requirements. XCH_4 retrieval has an RMSE of 3.8192 ppb and a bias of 1.2853 ppb, which on a relative scale agree very well with the XCO_2 performance estimates.

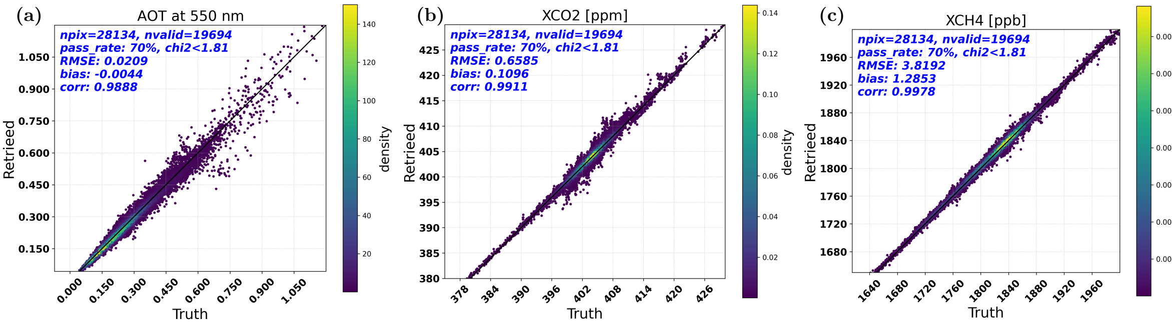

Figure 68.21: Performance of the 70th percentile best MAP plus CO2I retrievals for (a) AOD at 550 nm, (b) XCO_2 and (c) XCH4.

Figure 68.22 shows the XCO_2 error ΔXCO_2 for the 70th percentile best $\chi2^2$ MAP+CO2I retrievals as a function of SZA, the BDRF isotropic parameter at 2000 nm, the BPDF scaling factor and AOD at 550 nm. Excluding a few obvious outliers, the general trend is as follows. With respect to SZA, a slightly larger scatter of ΔXCO_2 can be observed at SZA $\approx 18°$, and the scatter starts to decrease until SZA $\approx 60°$. With respect to surface properties, the magnitude of ΔXCO_2 decreases obviously for brighter scenes or larger BPDF scaling factor. As AOD gets larger, the scatter of ΔXCO_2 increases mildly.

In Fig. 68.23 we also show the XCH_4 error ΔXCH_4 in dependence of the parameters mentioned above. There is no clear relationship between ΔXCH_4 and SZA, except for a few outliers at SZA $\approx 18°$. Analogous to Fig. 68.22 , the scatter in ΔXCH_4 reduces with scene brightness and larger BPDF scaling factor. The bias of ΔXCH_4 decreases with larger AOD value.

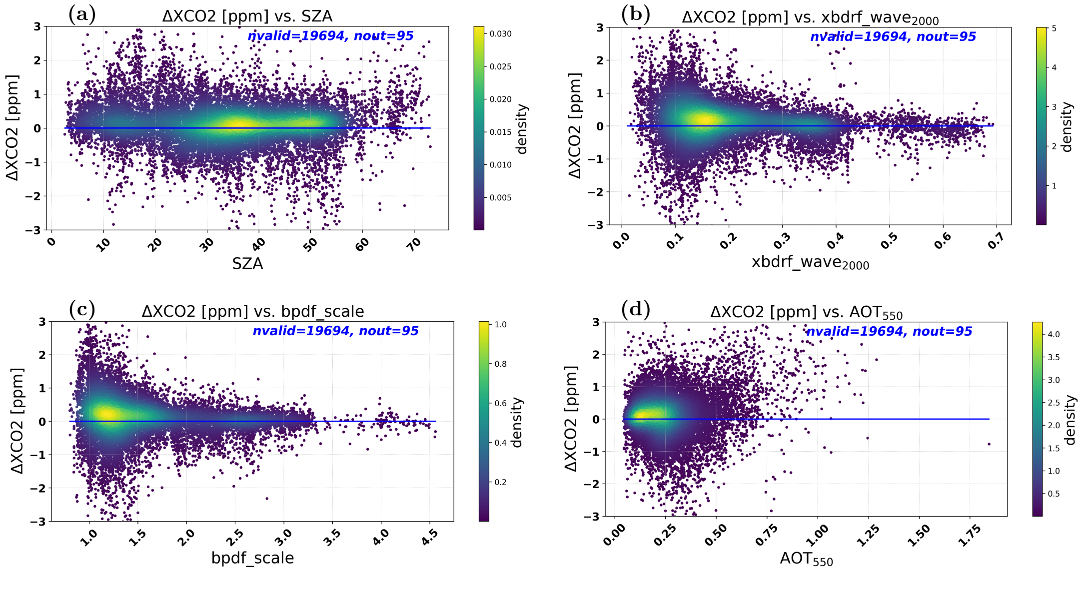

Figure 68.22: Residual of XCO_2 for the 70th percentile best $\chi2^2$ MAP+CO2I retrievals as a function of (a) SZA, (b) isotropic surface reflection at 2000 nm, (c) BPDF scaling factor and (d) AOT at 550 nm.

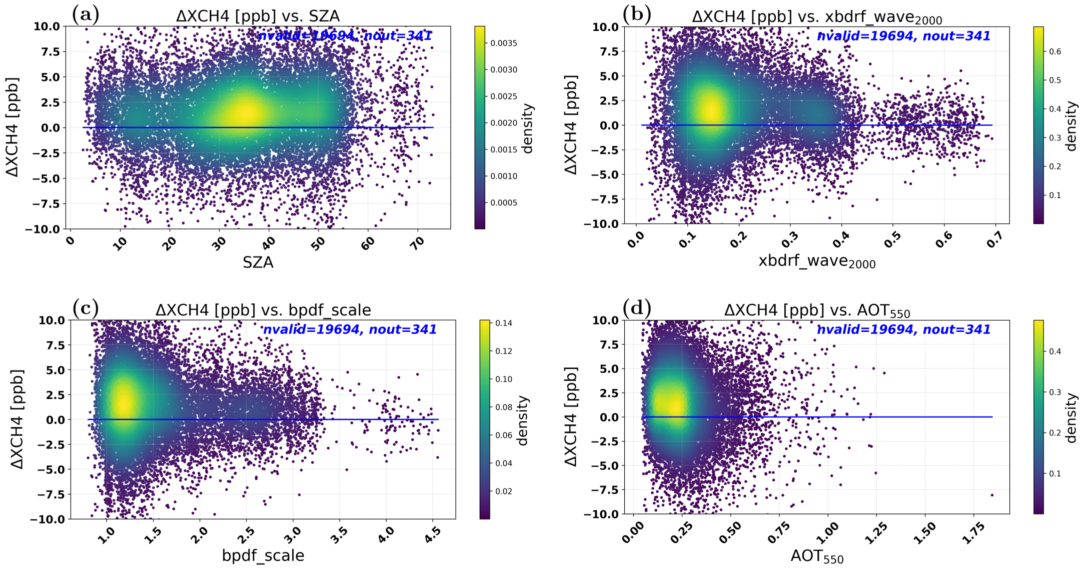

Figure 68.23: Same as Fig. 68.22 but for XCH$_4$.

69 Performance analysis on CO2I-only retrievals

The retrieved state variables for the CO2I-only case are given in section 33. Here the atmospheric aerosol is described by a 1-mode power-law size distribution, and surface reflection is assumed to be Lambertian in the retrieval. The performance is shown by Fig. 69.24. We can observe that by lacking of information CO2I-only case is not able to well retrieve AOD, which results in a large RMSE of 0.1840 and bias of -0.1250. As a result of the poor AOD retrieval, XCO$_2$ retrieval has an RMSE of 1.5587 ppm and a bias of 0.2545 ppm. The XCH$_4$ retrieval has an RMSE of 6.4262 ppb and a bias of -1.5283 ppb, which is non-compliant with the CO2M mission requirements.

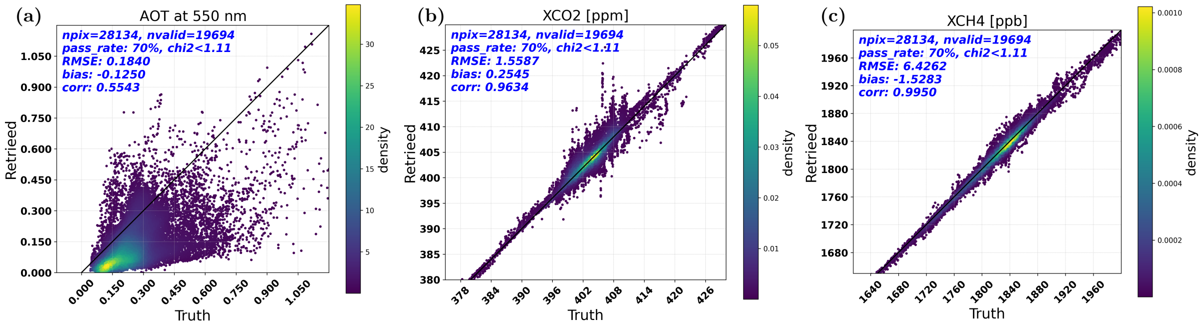

Figure 69.24: Performance of the 70th percentile best $\chi 2^2$ CO2I-only retrievals for (a) AOT at 550 nm, (b) XCO$_2$ and (c) XCH$_4$.

70 Influence of cirrus on MAP+CO2I retrievals

To investigate the effect of cirrus clouds, we first test the performance of MAP+CO2I retrieval where we only fit aerosol properties (not cirrus) for the cirrus-contaminated pixels (with cirrus_mask$=1$), with the results shown by Fig. 70.25. If we condsider the 70th percentile best $\chi 2^2$ retrievals with best chi2 (upper panel of Fig. 70.25), the RMSEs of AOT at 550 nm, XCO$_2$ and XCH$_4$ are 0.1410, 6.4187 ppm and 29.9628 ppb, respectively, increasing by a factor of around 10 compared to cirrus-free case (Fig. 68.21). However, these retrievals include a lot of cases with poor fit between forward model and measurement as is apparent from the 70 percentile chi2 value of 111. If we now apply a filter on the goodness of fit of chi2<2 (lower panel of Fig. 70.25), many of the bad retrievals are filtered out (only 4% of the cirrus pixels remain) and the RMSE and bias are reduced to 2.09 and -1.68 ppm for XCO_2 and 9.22 and 6.66 ppb for XCH_4. This shows chi2 filtering is an effective way of screening for cirrus and with this filtering cirrus contamination would only affect a small number of pixels. On the other hand, when we would be able to perform retrievals for cirrus contaminated pixels, we would increase the number of valid retrievals and therewith the effective coverage of the XCO_2 and XCH_4 products

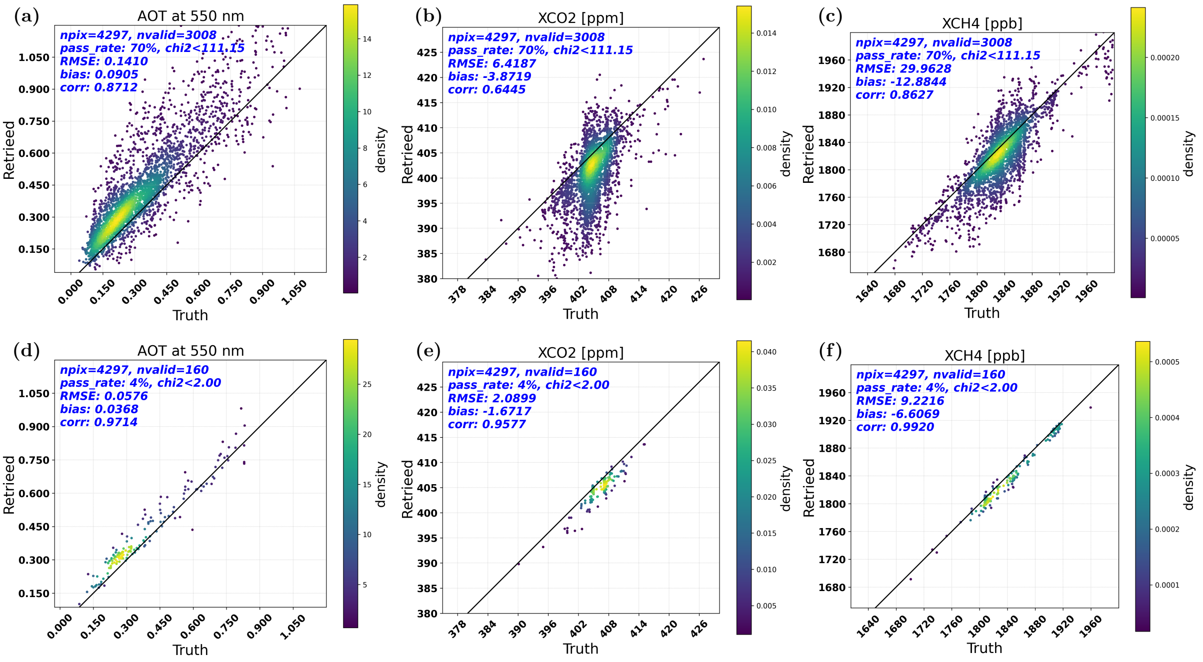

Figure 70.25: Performance of the 70th percentile best $\chi 2^2$ MAP+CO2I retrievals at cirrus-contaminated pixels by not fitting cirrus for (a) AOT at 550 nm, (b) XCO_2 and (c) XCH_4.

We therefore conduct the retrieval by fitting cirrus on all pixels including both cirrus-contaminated and cirrus-free ones, where the parameters for cirrus in the state vector can be seen in section 38. The retrievals on cirrus-contaminated pixels are shown in the upper panel of Fig. 70.26, and on cirrus-free pixels in the lower panel. We can see that COT at 550 nm is well retrieved with an RMSE of 0.0139 and a bias of -0.0046. The cirrus-induced error in terms of RMSE is significantly reduced to 0.0357, 1.1813 ppm, 6.1200 ppb for AOT at 550 nm, XCO_2 and XCH_4 retrievals. Fitting cirrus does not influence the accuracy of the retrievals at cirrus-free pixels, since we can observe similar values on the error terms between the case of fitting cirrus (lower panel of Fig. 70.26) and the case of not fitting (Fig. 68.21). For the two cases, the RMSEs are 0.0212 compared to 0.0209 for AOT at 550nm, 0.6588 ppm to 0.6585 ppm for XCO_2 and 3.8503 ppb to 3.8192 ppb for XCH_4.

71 Influence of cirrus on CO2I-only retrievals

Similar tests are also performed for CO2I-only retrieval. Fig. 71.27 shows the results of not fitting cirrus for cirrus-contaminated pixels. The RMSE of AOT at 550 nm is 0.2402, which is similar to the cirrus-free case (Fig. 69.24). The RMSEs of XCO_2 and XCH_4 are 3.6883 ppm and 16.2801 ppb, respectively, increasing by a factor of 2.5 compared to cirrus-free case (Fig. 69.24). This means cirrus contamination can be more easily compensated by fitting aerosol properties for the CO2I measurements than for the MAP measurements.

The performance can be somewhat improved by fitting cirrus in addition to aerosol properties, as illustrated in Fig. 71.28. If we fit cirrus at cirrus-contaminated pixels (upper panel of Fig. 71.28), the

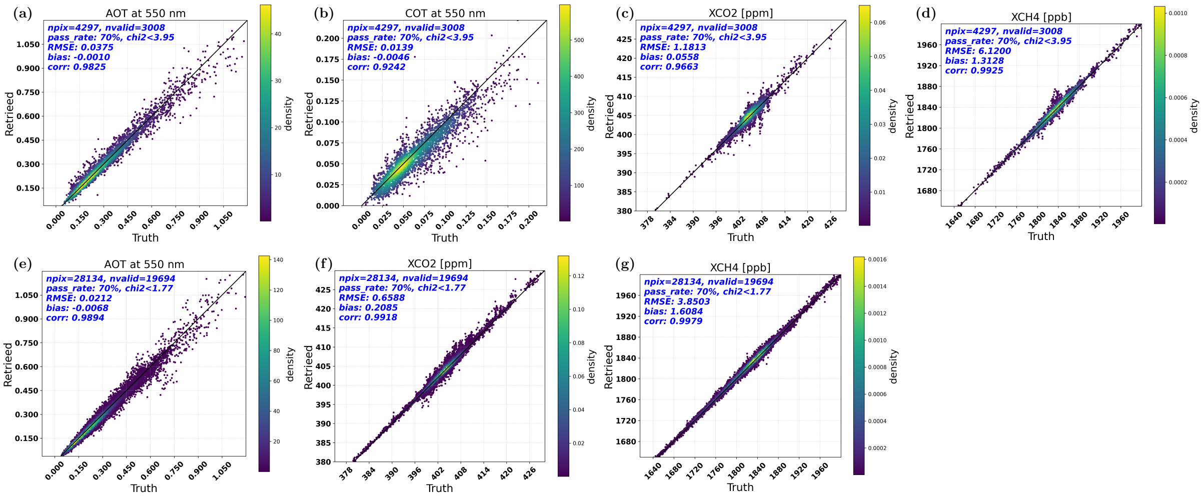

Figure 70.26: Performance of the 70th percentile best $\chi 2^2$ MAP+CO2I retrievals by fitting cirrus at cirrus-contaminated pixels for (a) AOT at 550 nm, (b) COT at 550 nm (c) XCO$_2$ and (d) XCH$_4$, and at cirrus-free pixels for (e) AOT at 550 nm, (f) XCO$_2$ and (g) XCH$_4$.

retrievals slightly improves with the RMSEs of 0.2781, 0.0422, 3.0529 ppm and 14.2012 ppb for AOT at 550 nm, COT at 550 nm, XCO$_2$ and XCH$_4$, respectively. With respect to the cirrus-free pixels, the retrievals of fitting cirrus (lower panel of Fig. 71.28) are also slightly better than that of not fitting (Fig. 69.24).

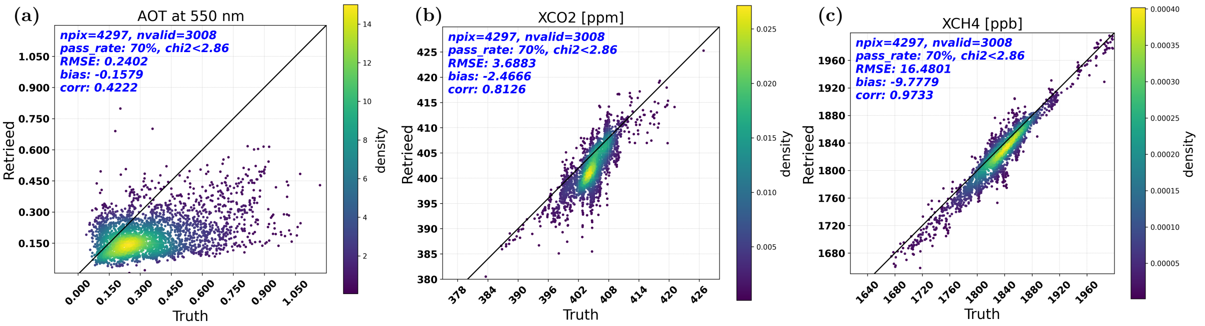

Figure 71.27: Performance of the 70th percentile best $\chi 2^2$ CO2I-only retrievals at cirrus-contaminted pixels by not fitting cirrus for (a) AOT at 550 nm, (b) XCO$_2$ and (c) XCH$_4$.

72 Initial results of non-scattering retrieval and proxy retrieval

Fig. 72.29 shows the ratio of C_{CH_4}/C_{CO_2}, XCO$_2$ and XCH$_4$ for the proxy retrievals, with the upper panel for the results at cirrus-free pixels and the lower panel for cirrus-contaminated pixels. We use the Mean Relative Error (MRE), which is the mean of absolute error scaled by the truth, to describe the accuracy of the ratio XCH$_4$/XCO$_2$. The retrievals for cirrus-free pixels (the upper panel of Fig. 72.29), have an MRE of the ratio C_{CH_4}/C_{CO_2} of 0.18%, leading to an RMSE and bias of XCO$_2$ as 0.9280 ppm and 0.2677 ppm, and an RMSE and bias of XCH$_4$ 4.1996 ppb and -1.2097 ppb, under assumption of perfect prior XCO$_2$ (for XCH$_4$ proxy) or prior XCH$_4$ (for XCO$_2$ proxy). For cirrus-contaminated pixels, the MRE of the ratio C_{CH_4}/C_{CO_2} is 0.2%, the RMSE and bias of XCO$_2$ are 1.0572 ppm and 0.4415 ppm, and the RMSE and bias of XCH$_4$ are 4.7787 ppb and -1.9952 ppb, which are similar to the cirrus-free pixels and indicate small influence of cirrus on proxy retrieval. The results indicate that even with perfect prior information, the proxy retrievals have non-negligible error. In this case, the proxy retrievals are significantly better

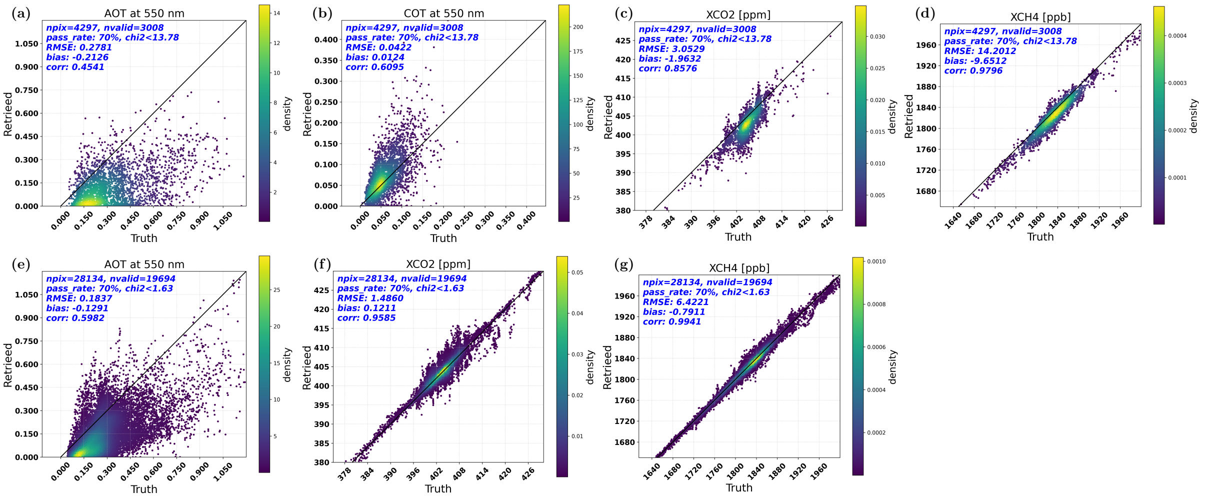

Figure 71.28: Performance of the 70th percentile best $\chi 2^2$ CO2I-only retrievals by fitting cirrus at cirrus-contaminted pixels for (a) AOT at 550 nm, (b) COT at 550 nm (c) XCO$_2$ and (d) XCH$_4$, and at cirrus-free pixels for (e) AOT at 550 nm, (f) XCO$_2$ and (g) XCH$_4$.

than the CO2I only full physics retrievals, especially when cirrus is present. Compared to CO2I+MAP retrievals, the proxy retrievals have slightly larger errors for the cirrus free pixels and a slightly smaller error for the cirrus contaminated pixels. It should be noted however that there will be a significant error in the prior XCH$_4$ and XCO$_2$ needed for the proxy method. Taken this into account, the error in the proxy will probably be (much) larger than in the CO2I+MAP retrievals.

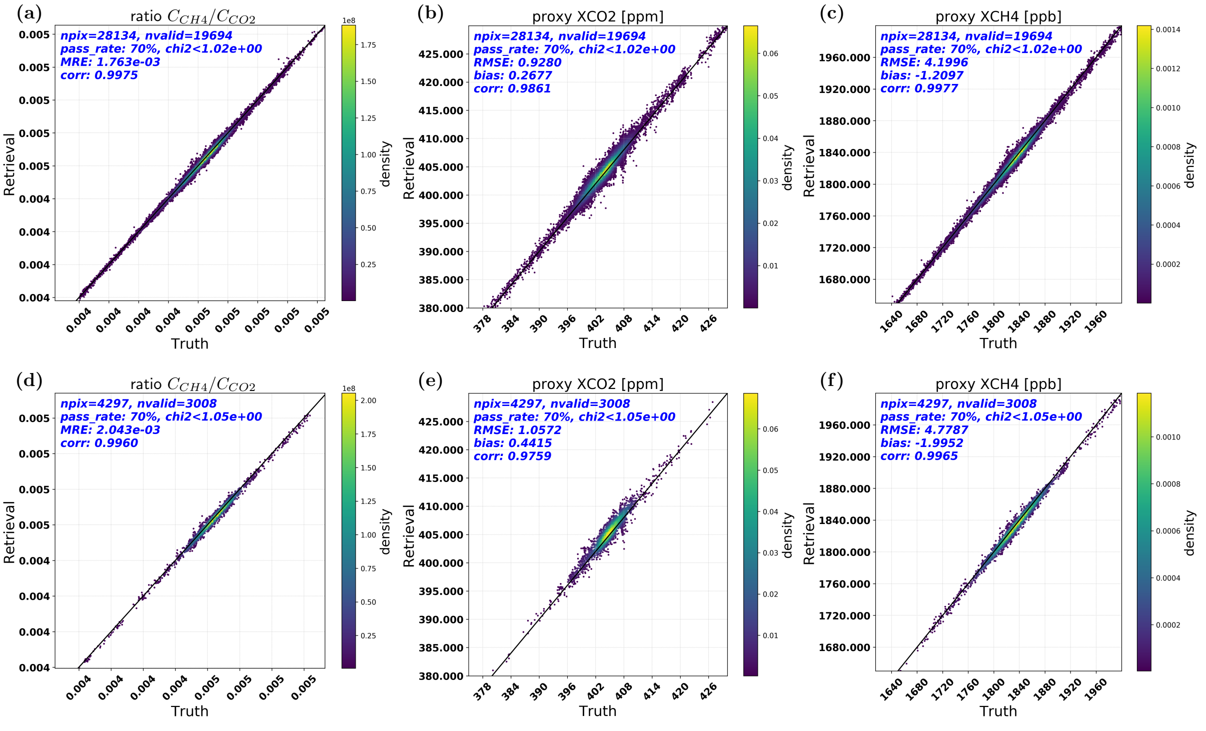

Figure 72.29: Performance of the 70th percentile best $\chi 2^2$ proxy retrievals out of SWIR-1 band at cirrus-free pixels for (a) ratio of C_{CH_4}/C_{CO_2}, (b) XCO$_2$ and (c) XCH$_4$, and at cirrus-contaminated pixels for (d) ratio of C_{CH_4}/C_{CO_2}, (e) XCO$_2$ and (f) XCH$_4$.

73 Feasibility

74 Computational Effort

The RemoTAP algorithm and software have strong heritage in application to POLDER-3/PARASOL satellite measurements[18,20,25,27,79,80], airborne measurements of the Research Scanning Polarimeter[14,15,26], SPEX airborne[16,17], airMSPI[17]. It has proven to run stable on large data sets, both for real and synthetic measurements, at the SRON internal computer clusters as well as on the Spider supercomputer of the NL Surfsara computing facility.

Table 74.26 summarizes the computational time of simultaneous MAP+CO2I, CO2I-only and MAP-only case over land averaged over their 70th percentile best chi2 retrievals. The retrievals were performed with the ifort v20 compiler. The CO2I-only case without fitting cirrus needs 6.26 s per pixel per thread on average for the retrieval, with 1.21 s per iteration. MAP+CO2I case without fitting cirrus needs 24.18 s per pixel per thread on average, with 3.22 s per iteration. For the cases of fitting cirrus, CO2I-only retrieval needs 12.48 s per pixel per thread on average, with 1.88 s per iteration. The total retrieval time increases with a large increasement in the number of iteration and a small increasement in the processing time of cirrus. The MAP+CO2I case, however, has a small increasement in the processing time of cirrus by 0.13 s, resulting a total processing time of 26.33 per pixel per thread on average with the number of iterations remained the same with cirrus-free case. MAP-only retrievals were only performed for cirrus-free case, which needs 7.33 s per pixel per thread on average, with 1.49 s per iteration.

Table 74.26: Computational time of MAP+spectrometer and spectrometer-only retrievals for the orbit case.

retrieval case	fitting cirrus	time/iteration /pixel/thread [s]	total time /pixel/thread [s]	# of iterations /pixel
CO2I-only	no	1.21	6.26	5.17
MAP+CO2I	no	3.22	24.18	7.7
CO2I-only	yes	1.88	12.48	7.5
MAP+CO2I	yes	3.35	26.33	7.7
MAP-only	no	1.49	7.33	4.9

For the present version of the code, the computational speed for gfortran is similar to that for ifort. This was tested on the SRON computational resources. This means the gfortran computational speed has increased by a factor 2-3 compared to the previous code version. We expect that this improvement also applies when running the code on the EUMETSAT TCE. For the previous code version, the computational speed using gfortran was very similar between the SRON environment and the EUMETSAT TCE.

75 Programming and Procedural Considerations

RemoTAP is written in FORTRAN-90 and tested with two compliers, the gfortran and the ifort. For the tests, we used versions gfortran-9 and the intel-v20 ifort. The proposed software implementation assumes an overarching framework interface, calling the different modules. The software design foresees that memory is shared between the different modules. For optimal implementation in a framework, a trade-off analysis has to be made between file-based and interface-based data exchange. Overall, the software is thread-safe and parallelized processing is possible, tested with openMP and openMPI.

76 Quality Assessment and Diagnostics

In case problems arise when executing RemoTAP, the software returns different PQFs (Tab. 57.23) and status flags (Tab. 77.28) indicating warnings and/or serious software failures. The provided information can be supported by basic break point outputs. Obviously, this cannot replace a detailed debugging but may help to trace errors in the developing and implementation phase of RemoTAP.

77 Exception Handling

Any exception handling of the RemoTAP software should depend on the corresponding requirements of the calling framework. To facilitate proper error handling, we propose that each individual routine returns a status flag to the framework. The return status flag of zero indicates a successful processing whereas any value other than zero indicates an error. If the operational processor is properly installed, only the success flag will be returned as status flag in any of the routines.

The return status flag should not be confused with processing quality flags (PQFs) as defined in Sec. 57, which are meant to indicate that a specific calculation may not provide proper output (e.g. because of non-convergence or data filtering). An unsuccessful calculation for such reasons does not imply that the algorithm or the framework is wrongly implemented. Moreover, the processing quality flag can also indicate a warning. In this case, proper output is given, but it is advised to handle the data with care. The return status, on the other hand, are essential for proper handling of errors occurring in the RemoTAP Fortran routines. They must be consistent with return status flags used inside the framework. The symbols of the possible return status flags should be common to all algorithms and should be defined by the framework. Our suggestion of return flags is listed in Tab. 77.28.

Table 76.27: Breakpoint output (BPO) of RemoTAP.

Parameter	**Symbol**	**Dimension**	**Physical Unit**
wavelength I MAP	$\text{BPO} - \lambda_{\text{pol,MAP}}$	$N_{\lambda,\text{rad}}$	nm
measured I MAP	$\text{BPO} - I_{\text{meas,MAP}}$	$N_{\text{vp}} \times N_{\lambda,\text{rad}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
noise radiance	$\text{BPO} - I_{\text{noise,MAP}}$	$N_{\text{vp}} \times N_{\lambda,\text{rad}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
simulated I MAP	$\text{BPO} - I_{\text{sim,MAP}}$	$N_{\text{vp}} \times N_{\lambda,\text{rad}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
wavelength Q/U MAP	$\text{BPO} - \lambda_{\text{pol,MAP}}$	$N_{\lambda,\text{pol}}$	nm
measured Q MAP	$\text{BPO} - Q_{\text{meas,MAP}}$	$N_{\text{vp}} \times N_{\lambda,\text{pol}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
noise Q MAP	$\text{BPO} - Q_{\text{noise,MAP}}$	$N_{\text{vp}} \times N_{\lambda,\text{pol}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
simulated Q MAP	$\text{BPO} - Q_{\text{sim,MAP}}$	$N_{\text{vp}} \times N_{\lambda,\text{pol}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
measured U MAP	$\text{BPO} - U_{\text{meas,MAP}}$	$N_{\text{vp}} \times N_{\lambda,\text{pol}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
noise U MAP	$\text{BPO} - U_{\text{noise,MAP}}$	$N_{\text{vp}} \times N_{\lambda,\text{pol}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
simulated U MAP	$\text{BPO} - U_{\text{sim,MAP}}$	$N_{\text{vp}} \times N_{\lambda,\text{pol}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
wavelengths NIR CO2I	$\text{BPO} - \lambda_{\text{NIR}}$	$\text{IR} - N_{\text{NIR}}$	nm
measured spectrum NIR CO2I	$\text{BPO} - I_{\text{meas,NIR}}$	$\text{IR} - N_{\text{NIR}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
noise spectrum NIR CO2I	$\text{BPO} - I_{\text{noise,NIR}}$	$\text{IR} - N_{\text{NIR}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
modeled spectrum NIR CO2I	$\text{BPO} - I_{\text{mod,NIR}}$	$\text{IR} - N_{\text{NIR}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
wavelengths SW1 CO2I	$\text{BPO} - \lambda_{\text{SW1}}$	$\text{IR} - N_{\text{SW1}}$	nm
measured spectrum SW1 CO2I	$\text{BPO} - I_{\text{meas,SW1}}$	$\text{IR} - N_{\text{SW1}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
noise spectrum SW1 CO2I	$\text{BPO} - I_{\text{noise,SW1}}$	$\text{IR} - N_{\text{SW1}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
modeled spectrum SW1 CO2I	$\text{BPO} - I_{\text{mod,SW1}}$	$\text{IR} - N_{\text{SW1}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
wavelengths SW2 CO2I	$\text{BPO} - \lambda_{\text{SW2}}$	$\text{IR} - N_{\text{SW2}}$	nm
measured spectrum SW2 CO2I	$\text{BPO} - I_{\text{meas,SW2}}$	$\text{IR} - N_{\text{SW2}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
noise spectrum SW2 CO2I	$\text{BPO} - I_{\text{noise,SW2}}$	$\text{IR} - N_{\text{SW2}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$
modeled spectrum SW2 CO2I	$\text{BPO} - I_{\text{mod,SW2}}$	$\text{IR} - N_{\text{SW2}}$	$\text{phot}/\left(\text{m}^2\,\text{s}\,\text{sr}\,\text{nm}\right)$

Table 77.28: Proposed RemoTAP return status flags and their meaning.

Flag symbol	**Meaning**
STATUS_S_SUCCESS	No error.
STATUS_E_ASSERTION_ERROR	Assertion failed.
STATUS_E_GENERIC_EXCEPTION	Catch all error.
STATUS_E_CONFIGURATION_ERROR	Error in the configuration of the processor.
STATUS_E_DIMENSIONS_ERROR	The number of dimensions or the dimension sizes are not as expected.
STATUS_E_MEMORY_ALLOCATION_ERROR	Memory allocation failed.
STATUS_E_CONSISTENCY_ERROR	Provided input is not internally consistent.
STATUS_E_MEMORY_DEALLOCATION_ERROR	Error while releasing memory.

78 Validation

This section discusses the validation needs for the RemoTAP XCO_2, XCH_4 and aerosol data product.

79 Ground based

80 The Total Carbon Column Observatory Network (TCCON)

In 2004 the TCCON network was founded in preparation for the validation of the OCO mission. Since then the network has become the standard for validating satellite based column measurements of CO_2 and CH_4[40,81]. TCCON is a network of inter-calibrated ground-based Fourier transform spectrometers (Bruker 125HR) that measure the absorption in the NIR/SWIR of direct sunlight by trace gas species such as CO_2, CH_4, CO, HDO, etc. Due to the viewing geometry, TCCON measurements are much less influenced by atmospheric scattering by cirrus and aerosols than satellite observations of backscattered/reflected sunlight. TCCON XCO_2 and XCH_4 measurements have been calibrated and validated against the WMO-standard of in-situ measurements using dedicated aircraft campaigns of XCH_4 profiles and their resulting accuracy have been estimated to 0.4% (2σ value)[40]. Table 80.29 shows an overview of stations, which provided data for the year 2020 according to https://tccondata.org. Although the TCCON data base is the most accurate set of reference data, an important limitation of the TCCON network is the limited coverage with respect to surface albedo and its sparse spatial coverage. Also, for most stations regular delivery of data takes at least a year and so specific effort is required to allow for a timely validation of CO2M data. Moreover, the TCCON data policy is on conflict with that of the Copernicus programme.

Table 80.29: Overview of TCCON stations with data publicly available from the year 2020 (https://tccondata.org, March 2021)

station	**latitude** [degree]	**longitude** [degree]	**height** [m]	**period**	**comment**
Burgos	18.5325	120.6496	35	since Mar. 2017	
CalTech	34.136	-118.127	230	since Sept 2012	
Darwin	-12.4245	130.892	30	since Aug. 2005	
Edwards	34.95	-117.83	699	since July 2013	
East Trout Lake	54.34	104.99	501	since Oct. 2016	
Eureka	80.05	-86.42	610	since July 2010	too northerly
Lzaña	28.3	-16.5	2370	since May 2007	too high
Karlsruhe	49.1	8.438	111	since Apr 2010	
Lauder	-45.038	169.684	370	since June 2004	
Paris	48.486	2.356	60	since Sept 2014	
Park Falls	45.945	-90.273	442	since May 2004	
Reunion Island	-20.9014	55.4847	87	since Oct. 2011	
Saga	33.241	130.288	7	since June 2011	
Sodankylä	67.368	26.633	179	since Feb. 2009	
Wollongong	-34.406	150.879	30	since June 2008	

81 The COCCON network

COCCON is an important validation data source that is complementary to the TCCON network. The COCCON network is based on portable EM27 Bruker spectrometer[82]. Since 2016, the network of inter-calibrated spectrometer is running, which at the meantime comprises 40 devices operating around the world. Although less accurate, measurement of this network provides a very valuable extension to the TCCON network regarding its global coverage due to the stable performance of the spectrometer.

82 Mobile EM27 observations

A main objective of CO2M is to determine anthropogenic $\mathrm{CO_2}$ and $\mathrm{CH_4}$ emissions of localized point sources. The accuracy of these estimates reply on accurate measurements of the XCO$_2$ anomalies, which is the enhancement of the XCO$_2$ signal with respect to a situation without the local source. To validate the accuracy of these type of CO2M observations, it is essential to measure the XCO$_2$

enhancement in downward direction of an emission plume with sufficient spatial samples. Locating portable EM27 in downwind direction may provide an adequate validation concept. However, the optimal location of the spectrometers depends on the local wind field and so requires very flexible logistics. Mobile EM27 may provide an appropriate alternative[83], as they ensure even more flexible use of the devices. Finally, mobile EM27 may provide an important validation source for CO2M ocean sun glint observations as demonstrated by Klappenbach et al., 2015[84] and Knapp et al., 2021[85].

83 Aeronet

The MAP observations and the inferred aerosol product can be validated with ground-based measurements of the Aerosol Robotic Network (AERONET). For more than 25 years, the network has provided long-term, continuous, and readily accessible public domain database of aerosol optical, microphysical and radiative properties for aerosol research and characterization, validation of satellite retrievals, and synergism with other databases. The network imposes standardization of instruments, calibration, processing and distribution. Although globally distributed, currently no collocated TCCON and AeroNET measurements are performed on a regular basis. Therefore, an equipment of several TCCON sites with an AeroNET instrument would be a very valuable extension of the CO2M validation data set.

84 Aircraft measurements

In Europe, aircraft measurements of XCO_2 and XCH_4 total column mixing ratio and aerosol properties are preformed sporadically, although they can provide an important element to validate CO2M measurements over cities. This was demonstrated by several aircraft campaigns (e.g.[17,86]) and therefore we advise to consider these observations as an essential element of regular CO2M validation campaigns.

85 In Situ Measurements

In situ measurements of the full vertical XCO_2 and XCH_4 profile will be extremely useful as they allow the column averaging kernel to be applied and hence the retrieved XCO_2 and XCH_4 from CO2M can be compared with truly the same quantities. TCCON only delivers the column integrated values and hence do not allow the averaging kernel to be applied, which leads to a remaining uncertainty in the comparison. Vertical profiles can be obtained by aircrafts as is done for calibrating the TCCON network, but they do not sample the full total column. In this respect a very interesting new development is provided by the AirCore initiative[87]. AirCore is an innovative atmospheric sampling system that consists of a long tubing, usually in the shape of a coil, that can sample the surrounding atmosphere and preserve a profile of the trace gas of interest. The narrow diameter and long length are designed to minimize the diffusive mixing occurring inside the tubing between sampling and analysis. AirCore, invented and patented by Pieter Tans (the head of the Carbon Cycle Group at NOAA/ESRL), can in principle provide measurement precisions equal or even better than the in-situ flask sampling for CO_2 and CH_4 (and also CO). A clear advantage of AirCore over aircraft measurements is the lower cost and the fact that it can in principle sample the profile across the entire column. The most obvious application of AirCore would be as an alternative to the aircraft spirals that are undertaken for the absolute calibration of TCCON to the in-situ standard. Not only does it lower the cost it also allows for measurements higher up in the atmosphere where now climatology has to be used to extrapolate the aircraft measurements and which poses the largest error source to this absolute TCCON calibration particularly for CH_4[40]. Another possible application of AirCore would be to use it in areas where there are no TCCON validation sites.

86 Verification of instrument performance

This section will be added later.

87 Assumption and Limitations

88 Performance Assumptions

The RemoTAP software is based on the following assumptions:

- Level 1 input data are all collocated and no adjustment of measurements is required because of spatial misregistration, nor between measurements of different payload instruments neither between spectral band measurements of CO2I.
- The RemoTAP software requires an overarching framework, which calls the different models in sequential order as depicted in Fig. 2.2.
- Each module requires access to the internal memory to share variables. Data exchange between framework and the RemoTAP tools is not yet defined and a thorough trade-off between file-based and interface exchange is required for further interface implementation.
- RemoTAP assumes a sequential approach for cloud clearing, where in first instance, CLIM data are used for pre-selecting data. Subsequently, the cloud clearing is based on non-scattering retrievals of $\mathrm{CO_2}$ and $\mathrm{CH_4}$ total column densities from weak and strong absorption bands. Moreover, MAP observations can be used to identify cloudy pixels.

89 Potential Improvements

Cloud filtering based on MAP observations requires further development including an assessment of its numerical efficiency. This should also include the aspect of cirrus detection, which represents a major challenge for the $\mathrm{CO_2}$ retrieval. Here both CLIM observations in the 1.38 μm band and MAP observations may be used in a synergistic manner to enhance the accuracy of the CO2M data product, Lastly, to enhance the maturity of the algorithm, we advice to perform dedicated retrieval studies using OCO-2 and Polder data. With the launch of Sentinel-5, a unique test environment is given for the synergistic exploitation of MAP and spectrometer data. Applying RemoTAP to Sentinel-5 and 3MI data, will enhance the maturity and confidence in the CO2M RemoTAP software and so is strongly recommended.

90 Numerical Tools

91 LINTRAN V2.0

To calculate the multiple scattering radiance field at the spectral grid points of the k-binning approach, we employ the scalar version of the LINTRAN radiative transfer model. The model is already described in the literature[66] and its latest version 2.0 is already used for the operational processing of the S5P methane data product. Therefore, we restrict our discussion to the basic features of this model. LINTRAN solves the scalar transport equation in an atmosphere that is illuminated by a radiation source $\mathbf{S}$. In its forward formulation, the transport equation is written as

$$\hat{\mathbf{L}} \circ \mathbf{I} = \mathbf{S}_0, \tag{91.130}$$

where I denotes the monochromatic intensity field. The forward transport operator $\hat{\mathbf{L}}$ comprises an operator to describe extinction $\hat{\mathbf{L}}_e$ and a scattering operator $\hat{\mathbf{L}}_s$ to account for all scattering events including surface reflection at the bottom boundary of the atmosphere[88,89].

$$\hat{\mathbf{L}} = \hat{\mathbf{L}}_e - \hat{\mathbf{L}}_s \tag{91.131}$$

with

$$\hat{\mathbf{L}}_e = \left[\mu \frac{\delta}{\delta z} + \beta_e\left(z\right)\right] \tag{91.132}$$

and

$$\hat{\mathbf{L}}_s = \int_{4\pi} d\tilde{\mathbf{\Omega}} \left\{ \frac{\beta_s\left(z\right)}{4\pi} P\left(z, \tilde{\mathbf{\Omega}}, \mathbf{\Omega}\right) + \rho\left(\tilde{\mathbf{\Omega}}, \mathbf{\Omega}\right) \delta\left(z\right) \Theta\left(\mu\right) \left|\mu\right| \Theta\left(-\tilde{\mu}\right) \left|\tilde{\mu}\right| \right\}. \tag{91.133}$$

Here z denotes the vertical coordinate running from the surface to the top of the atmosphere, $\mathbf{\Omega}$ is the solid angle determined by the relative azimuth angle φ and the cosine of the zenith angle μ ($\mu < 0$ for downward directions, $\mu > 0$ for upward directions) and β_e and β_s represent the extinction and scattering coefficients, respectively. P denotes the phase function and ρ represents the bidirectional reflection function (BDRF) that characterizes the surface reflection. δ and Θ represent the Dirac delta function and the Heaviside step function, respectively.

The source term on the right-hand side of Eq. (91.130) is the unidirectional solar source $\mathbf{S}_0$, incident at the top of the atmosphere,

$$\mathbf{S}_0 = u_0 \delta\left(z - z_{\mathrm{TOA}}\right) \delta\left(\mathbf{\Omega} - \mathbf{\Omega}_0\right) \mathbf{F}_0, \tag{91.134}$$

Here F_0 denotes a monochromatic solar flux of a certain wavelength incident on the top of the atmosphere (TOA) at z_{TOA}, along solid angle $\mathbf{\Omega}_0 = (-u_0, \varphi_0)$ (with $u_0 > 0$ by definition).

Solving the transport equation produces the entire intensity field, although in the context of satellite remote sensing one is generally interested in simulating a certain observation. Any observable or radiative effect that is a linear function of the intensity field $\mathbf{I}$ can be described by a suitable response function R[90,91] through an inner product:

$$\mathcal{F} = \langle \mathbf{R}, \mathbf{I} \rangle. \tag{91.135}$$

Here, the inner product of two arbitrary functions a and b is defined as

$$\langle \mathbf{a}, \mathbf{b} \rangle = \int dz \int d\mathbf{\Omega} \quad a\left(z, \mathbf{\Omega}\right) b\left(z, \mathbf{\Omega}\right). \tag{91.136}$$

Considering a downward viewing instrument measuring intensity at the top of the atmosphere in viewing direction $\mathbf{\Omega}_v$, the response function is given by

$$R\left(z, \mathbf{\Omega}\right) = \frac{1}{F_0} \delta\left(\mathbf{\Omega} - \mathbf{\Omega}_v\right) \delta\left(z - z_{\mathrm{TOA}}\right) \ . \tag{91.137}$$

The discussion above gives the theoretical framework for the forward-adjoint approach, which is used to linearize the multiple scattering simulations with respect to the optical properties of the model atmosphere. To solve the radiative transfer equation numerically, we separate first and second order scattering from higher scattering orders in the solution of Eq. (91.130). The lower order of scattering radiance field can be solved analytically, whereas the simulation of higher order of scattering requires a numerical approach.

Finite orders of scattering are calculated analytically using the separation between $\hat{L}_e$ and $\hat{L}_s$ in Eq. (91.131) as shown in Schepers et al.(2014)[64]. Here, the order of scattering is defined as follows. The source of Eq. (91.130) has scattering order zero. Solving the radiative transfer equation without $\hat{L}_s$ results in I with the same order of scattering as S, thus

$$\hat{\mathbf{L}}_e \circ \mathbf{I}_n = \mathbf{S}_n \tag{91.138}$$

Here, the subscript n is the order of scattering. Solving the full radiative transfer equation results in I with at least the same order of scattering as S, or

$$\left(\hat{\mathbf{L}}_e - \hat{\mathbf{L}}_s\right) \circ \mathbf{I}_{n+} = \mathbf{S}_n \tag{91.139}$$

where for $n = 0$, $\mathbf{I}_{0+} = \mathbf{I}$ and it will be exactly Eq. (91.130). Combining Eqs. (91.138) and (91.139) gives.

$$\hat{\mathbf{L}}_e \circ \mathbf{I}_{(n+1)+} = \hat{\mathbf{L}}_s \circ \mathbf{I}_{n+} \tag{91.140}$$

If this is true for any n, we can conclude.

$$\hat{\mathbf{L}}_e \circ \mathbf{I}_{n+1} = \hat{\mathbf{L}}_s \circ \mathbf{I}_n \tag{91.141}$$

And substituting the left hand side with Eq.(91.138) for $n+1$ gives

$$\hat{\mathbf{L}}_s \circ \mathbf{I}_n = \mathbf{S}_{n+1} \tag{91.142}$$

Using Eqs.(91.138), (91.142) and (91.135), any finite order of scattering can be solved analytically.

$$\mathcal{F}_n = \langle \mathbf{R}, \left(\hat{\mathbf{L}}_e^{-1}\hat{\mathbf{L}}_s\right)^n \hat{\mathbf{L}}_e^{-1}\mathbf{S}_0\rangle \tag{91.143}$$

Here, $\hat{\mathbf{L}}_e^{-1}$ is the transmission operator and can be defined analytically. Derivatives of $\mathcal{F}_n$ with respect to any optical properties affecting $\hat{\mathbf{L}}_e$ or $\hat{\mathbf{L}}_s$ can be obtained by straight forward differentiation.

Higher order of scattering are calculated by solving Eq.(91.139) numerically. An essential step towards the solution is the stream approximation, where we approximate the zenith angle dependence of the radiance field by available streams along Gaussian quadrature point. The accuracy of this approximation is governed by the order of the Gaussian quadrature. Finally, the integro-differential equation can be converted to a block-diagonal matrix equation assuming a linear radiance field within the model layer. This assumption means a rough estimate for optically thick media, and errors introduced by this approximation can be mitigated by chopping model layers in a number of optically thin sub-layers. Here, the degree of layer splitting depends on the optical properties of the layer and can be controlled by external setting parameters. The multi-scattering observable becomes.

$$\mathcal{F}_{n+} = \langle \mathbf{R}, \mathbf{I}_{n+}\rangle \tag{91.144}$$

Similarly to how orders of scattering can be included in the source $\mathbf{S}_n$, orders of scattering can be included in the response function as well. The most efficient way to calculate high order of scattering is to split the orders as evenly as possible between R and S. As single and double scattering were solved analytically, the high order scattering includes third and higher orders. This is done by using $\mathbf{S}_2$ and $\mathbf{R}_1$.

To linearize $\mathcal{F}$ with respect to atmospheric and surface parameters, we come back to the operator formulation of atmospheric radiative transfer and employ the forward-adjoint perturbation theory (e.g.[90,91]). We will drop any order of scattering subscripts to simplify the equations. First, we introduce the adjoint intensity field $\mathbf{I}^\dagger$ defined by the adjoint transport equation taking the response function as the adjoint source, namely

$$\hat{\mathbf{L}}^\dagger \circ \mathbf{I}^\dagger = \mathbf{R} \tag{91.145}$$

Here $\hat{\mathbf{L}}^\dagger$ represents the adjoint radiative transfer operator[88,89,91]. Within the forward-adjoint perturbation theory, the adjoint intensity field represents the importance of sources within the atmosphere for a given observation $\mathcal{F}$[92,93]. This follows from the definition of the adjoint operator $\hat{\mathbf{L}}^\dagger$ with

$$\mathcal{F} = \langle \mathbf{R}, \mathbf{I} \rangle = \langle \hat{\mathbf{L}}^\dagger \circ \mathbf{I}^\dagger, \mathbf{I} \rangle = \langle \mathbf{I}^\dagger, \hat{\mathbf{L}} \circ \mathbf{I} \rangle = \langle \mathbf{I}^\dagger, \mathbf{S} \rangle \tag{91.146}$$

The third equal sign uses the definition of an adjoint operator and the other equal signs use Eqs. (91.135), (91.145) and (91.130).

For the forward and adjoint intensity fields and the associated source and response functions, the derivative of the observable with respect to an atmospheric parameter x reads[88,94]

$$\frac{\partial \mathcal{F}}{\partial x} = -\langle \mathbf{I}^\dagger, \hat{\mathbf{L}}' \mathbf{I} \rangle + \langle \mathbf{I}^\dagger, \mathbf{S}' \rangle + \langle \mathbf{R}', \mathbf{I} \rangle, \tag{91.147}$$

where $\hat{\mathbf{L}}'$ denotes the derivative of the transport operator with respect to x, i.e.

$$\hat{\mathbf{L}}' = \lim_{\Delta x \to 0} \frac{\Delta \hat{\mathbf{L}}}{\Delta x} \tag{91.148}$$

and $\mathbf{S}'$ and $\mathbf{R}'$ denote derivatives of the source and response functions, respectively.

Analytical expressions for $\hat{\mathbf{L}}'$ concerning the derivative with respect to absorption and scattering coefficients, phase function and BDRF parameters have previously been derived (e.g.[63,89,95]). Summarizing the use of the foward-adjoint perturbation theory for the linearization of our problem is justified by the fact that the inner product in Eq. (91.147) can be calculated in a straight forward manner, which provides the derivatives of $\mathcal{F}$ with respect to the optical parameters $\tau_{\mathrm{abs,k}}$, $\tau_{\mathrm{sca,k}}$ α_k^l, and A using only two radiative transfer solutions, the forward and the adjoint intensity field $\mathbf{I}$ and $\mathbf{I}^\dagger$. This represents a significant numerical advantage with respect to a numerical scheme based on a finite difference of the forward model.

List of TBD & TBC

TBC-ID	Section	Title	Description
TBC-1	section **??**	configuration parameters for v0	The configuration parameter list is to be completed for v0. The number of possible "BatchType" values is to be completed.

Table 91.1: List of TBCs/Assumptions

TBD-ID	Section	Title	Description
TBD-1	section **??**	~~configuration parameters~~	~~Dimensions of some configuration parameters are TBD.~~

Table 91.2: List of TBDs/Assumptions

References

[1] M. Buchwitz, M. Reuter, O. Schneising, W. Hewson, R. G. Detmers, H. Boesch, O. P. Hasekamp, I. Aben, H. Bovensmann, J. P. Burrows, A. Butz, F. Chevallier, B. Dils, C. Frankenberg, J. Heymann, G. Lichtenberg, M. De Mazière, J. Notholt, R. Parker, T. Warneke, C. Zehner, D. W. T. Griffith, N. M. Deutscher, A. Kuze, H. Suto, and D. Wunch. Global satellite observations of column-averaged carbon dioxide and methane: The GHG-CCI XCO 2 and XCH 4 CRDP3 data set. *Remote Sensing of Environment*, 203:276–295, December 2017.

[2] S. P. Rusli, O. Hasekamp, J. aan de Brugh, G. Fu, Y. Meijer, and J. Landgraf. Anthropogenic co$_2$ monitoring satellite mission: the need for multi-angle polarimetric observations. *Atmospheric Measurement Techniques Discussions*, 2020:1–31, 2020.

[3] Y. Meijer, H. Boesch, A. Bombelli, M. Buchwitzh, P. Ciais, D. Crisp, R. Englen, K. Holmlund, S. Houweling, G. Janssens-Maenhout, J. Marshall, M. Nakajima, B. Pinty, M. Scholze, J.-L. Bezy, M. R. Drinkwater, T. Fehr, V. Fernandez, A. Loescher, H. Nett, and B. Sierk. Copernicus co2 monitoring mission requirements document. techreport 3.0, European Space Agency, October 2020.

[4] B. Fougnie, T. Marbach, A. Lacan, R. Lang, P. Schlüssel, G. Poli, R. Munro, and A. B. Couto. The multi-viewing multi-channel multi-polarisation imager - Overview of the 3MI polarimetric mission for aerosol and cloud characterization. *J. Quant. Spectrosc. Radiat. Transfer*, 219:23–32, November 2018.

[5] O. Dubovik, Z. Li, M. I. Mishchenko, D. Tanré, Y. Karol, B. Bojkov, B. Cairns, D. J. Diner, W. R. Espinosa, P. Goloub, X. Gu, O. Hasekamp, J. Hong, W. Hou, K. D. Knobelspiesse, J. Landgraf, L. Li, P. Litvinov, Y. Liu, A. Lopatin, T. Marbach, H. Maring, V. Martins, Y. Meijer, G. Milinevsky, S. Mukai, F. Parol, Y. Qiao, L. Remer, J. Rietjens, I. Sano, P. Stammes, S. Stamnes, X. Sun, P. Tabary, L. D. Travis, F. Waquet, F. Xu, C. Yan, and D. Yin. Polarimetric remote sensing of atmospheric aerosols: instruments, methodologies, results, and perspectives. *J. Quant. Spectrosc. Radiat. Transfer*, 224:474–511, 2019.

[6] J. E. Hansen and L. D. Travis. Light scattering in planetary atmospheres. *Space Sci. Rev.*, 16:527–610, 1974.

[7] M. I. Mishchenko and L. D. Travis. Satellite retrieval of aerosol properties over the ocean using polarization as well as intensity of reflected sunlight. *J. Geophys. Res.*, 102:16,989–17,013, 1997.

[8] O. P. Hasekamp and J. Landgraf. Retrieval of aerosol properties over land surfaces: capabilities of multiple-viewing-angle intensity and polarization measurements. *Appl. Opt.*, 46:3332–3344, June 2007.

[9] O. P. Hasekamp. Capability of multi-viewing-angle photo-polarimetric measurements for the simultaneous retrieval of aerosol and cloud properties. *Atmospheric Measurement Techniques*, 3(4):839–851, 2010.

[10] K. Knobelspiesse, B. Cairns, M. Mishchenko, J. Chowdhary, K. Tsigaridis, B. van Diedenhoven, W. Martin, M. Ottaviani, and M. Alexandrov. Analysis of fine-mode aerosol retrieval capabilities by different passive remote sensing instrument designs. *Opt. Express*, 20:21457–21484, 2012.

[11] J. Chowdhary, B. Cairns, M. Mishchenko, P. Hobbs, G. Cota, J. Redemann, K. Rutledge, B. Holben, and Ed. Russel. Retrieval of aerosol scattering and absorption properties from photo-polarimetric observations over the ocean during the clams experiment. *J. Atmos. Sci.*, 62(4):1093–1117, 2005.

[12] F. Waquet, B. Cairns, K. Knobelspiesse, J. Chowdhary, L. D. Travis, B. Schmid, and M. I. Mishchenko. Polarimetric remote sensing of aerosols over land. *Journal of Geophysical Research (Atmospheres)*, 114(D13):1206–+, January 2009.

[13] F. Xu, G. van Harten, D. J. Diner, O. V. Kalashnikova, F. C. Seidel, C. J. Bruegge, and O. Dubovik. Coupled retrieval of aerosol properties and land surface reflection using the Airborne Multiangle SpectroPolarimetric Imager. *Journal of Geophysical Research (Atmospheres)*, 122:7004–7026, July 2017.

[14] L. Wu, O. Hasekamp, B. van Diedenhoven, and B. Cairns. Aerosol retrieval from multiangle, multispectral photopolarimetric measurements: Importance of spectral range and angular resolution. *Atmos. Meas. Tech.*, 8:2625–2638, 2015.

[15] L. Wu, O. Hasekamp, B. van Diedenhoven, B. Cairns, J. E. Yorks, and J. Chowdhary. Passive remote sensing of aerosol layer height using near-uv multi-angle polarization measurements. *Geophys. Res. Lett.*, 43(16):8783–8790, 2016.

[16] Cheng Fan, Guangliang Fu, Antonio Di Noia, Martijn Smit, Jeroen H.H. Rietjens, Richard A. Ferrare, Sharon Burton, Zhengqiang Li, and Otto P. Hasekamp. Use of a neural network-based ocean body radiative transfer model for aerosol retrievals from multi-angle polarimetric measurements. *Remote Sensing*, 11(23), 2019.

[17] G. Fu, O. Hasekamp, J. Rietjens, M. Smit, A. Di Noia, B. Cairns, A. Wasilewski, D. Diner, F. Seidel, F. Xu, K. Knobelspiesse, M. Gao, A. da Silva, S. Burton, C. Hostetler, J. Hair, and R. Ferrare. Aerosol retrievals from different polarimeters during the acepol campaign using a common retrieval algorithm. *Atmospheric Measurement Techniques*, 13(2):553–573, 2020.

[18] Otto P. Hasekamp, Pavel Litvinov, and André Butz. Aerosol properties over the ocean from PARASOL multiangle photopolarimetric measurements. *J. Geophys. Res.*, 116(D14):D14204+, July 2011.

[19] O. Dubovik, M. Herman, A. Holdak, T. Lapyonok, D. Tanré, J. L. Deuzé, F. Ducos, A. Sinyuk, and A. Lopatin. Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations. *Atmospheric Measurement Techniques*, 4(5):975–1018, 2011.

[20] G. Fu and O. Hasekamp. Retrieval of aerosol microphysical and optical properties over land using a multimode approach. *Atmospheric Measurement Techniques*, 11(12):6627–6650, 2018.

[21] J. L. Deuzé, P. Goloub, M. Herman, A. Marchand, G. Perry, S. Susana, and D. Tanré. Estimate of the aerosol properties over the ocean with POLDER. *J. Geophys. Res.*, 105(14):15329–15346, 2000.

[22] J. L. Deuzé, F. M. Bréon, C. Devaux, P. Goloub, M. Herman, B. Lafrance, F. Maignan, A. Marchand, F. Nadal, G. Perry, and D. Tanré. Remote sensing of aerosols over land surfaces from POLDER-ADEOS-1 polarized measurements. *J. Geophys. Res.*, 106(15):4913–4926, March 2001.

[23] M. Herman, J. L. Deuzé, C. Devaux, P. Goloub, F. M. Bréon, and D. Tanré. Remote sensing of aerosols over land surfaces, including polarisation measurements: Application to POLDER Measurements. *J. Geophys. Res.*, 102:17,039–17,049, 1997.

[24] S. Stamnes, C. Hostetler, R. Ferrare, S. Burton, X. Liu, J. Hair, Y. Hu, A. Wasilewski, W. Martin, B. Van Diedenhoven, J. Chowdhary, I. Cetinić, L. K. Berg, K. Stamnes, and B. Cairns. Simultaneous polarimeter retrievals of microphysical aerosol and ocean color parameters from the "MAPP" algorithm with comparison to high-spectral-resolution lidar aerosol and ocean products. *Appl. Opt.*, 57, April 2018.

[25] F. A. Stap, O. P. Hasekamp, and T. Röckmann. Sensitivity of parasol multi-angle photopolarimetric aerosol retrievals to cloud contamination. *Atmospheric Measurement Techniques*, 8(3):1287–1301, 2015.

[26] A. Di Noia, O. P. Hasekamp, L. Wu, B. van Diedenhoven, B. Cairns, and J. E. Yorks. Combined neural network/phillips-tikhonov approach to aerosol retrievals over land from the nasa research scanning polarimeter. *Atmos. Meas. Tech.*, 10:4235–4252, 2017.

[27] Otto P. Hasekamp, Edward Gryspeerdt, and Johannes Quaas. Analysis of polarimetric satellite measurements suggests stronger cooling due to aerosol-cloud interactions. *Nature Communications*, 10(1):1–7, November 2019.

[28] A. Butz, Guerlet S., Hasekamp O., Schepers D. andGalli A., Aben I., Frankenberg C., Hartmann J. M., Tran H., Kuze A., Keppel-Aleks G., Toon G., Wunch D., Wennberg P., Deutscher N., Griffith D., Macatangay R., Messerschmidt J., Notholt J., and Warneke T. Toward accurate CO_2 and CH_4 observations from GOSAT. *Geophys. Res. Let.*, 38:doi:10.1029/2011GL047888, 2011.

[29] S. Guerlet, A. Butz, D. Schepers, S. Basu, O. P. Hasekamp, A. Kuze, T. Yokota, J.-F. Blavier, N. M. Deutscher, D. W. T. Griffith, F. Hase, E. Kyro, I. Morino, V. Sherlock, R. Sussmann, A. Galli, and I. Aben. Impact of aerosol and thin cirrus on retrieving and validating xco2 from gosat shortwave infrared measurements. *Journal of Geophysical Research*, 118(10):4887–4905, 2013.

[30] L. Wu, O. Hasekamp, H. Hu, J. Landgraf, A. Butz, J. aan de Brugh, I. Aben, D. F. Pollard, D. W. T. Griffith, D. G. Feist, D. Koshelev, F. Hase, G. C. Toon, H. Ohyama, I. Morino, J. Notholt, K. Shiomi, L. Iraci, M. Schneider, M. de Mazière, R. Sussmann, R. Kivi, T. Warneke, T.-Y. Goo, and Y. Té. Carbon dioxide retrieval from oco-2 satellite observations using the remotec algorithm and validation with tccon measurements. *Atmospheric Measurement Techniques*, 11(5):3111–3130, 2018.

[31] L. Wu, O. Hasekamp, H. Hu, J. aan de Brugh, J. Landgraf, A. Butz, and I. Aben. Full-physics carbon dioxide retrievals from the orbiting carbon observatory-2 (oco-2) satellite by only using the $2.06\mu m$ band. *Atmospheric Measurement Techniques*, 12(11):6049–6058, 2019.

[32] L. Wu, J. aan de Brugh, Y. Meijer, B. Sierk, O. Hasekamp, A. Butz, and J. Landgraf. Xco2 observations using satellite measurements with moderate spectral resolution: investigation using gosat and oco-2 measurements. *Atmospheric Measurement Techniques*, 13(2):713–729, 2020.

[33] Haili Hu, Jochen Landgraf, Rob Detmers, Tobias Borsdorff, Joost Aan de Brugh, Ilse Aben, André Butz, and Otto Hasekamp. Toward global mapping of methane with tropomi: First results and intersatellite comparison to gosat. *Geophysical Research Letters*, 45(8):3682–3689, 2018.

[34] O. Hasekamp, A. Lorente, H. Hu, A. Butz, J. aan de Brugh, and J. Landgraf. Algorithm theoretical baseline document for sentinel-5 precursor methane retrieval. Report, Netherlands Insitute for Space Research, SRON, Utrecht, The Netherlands, 2019.

[35] A. Lorente, T. Borsdorff, A. Butz, O. Hasekamp, J. aan de Brugh, A. Schneider, L. Wu, F. Hase, R. Kivi, D. Wunch, K. Pollard, D.F. Shiomi, V.A. Deutscher, N.M. Velazco, C.M. Roehl, P.O. Wennberg, T. Warneke, and Landgraf J. Methane retrieved from TROPOMI: Improvement of the data product and validation of the first 2 years of measurements. *Atmos. Meas. Tech.*, 14:665–684, 2021.

[36] R. van Deelen, O. P. Hasekamp, and J. Landgraf. Accurate modeling of spectral fine-structure in earth radiance spectra measured with the global ozone monitoring experiment. *Appl. Optics*, 46:243–252, 2007.

[37] A. Wassmann, T. Borsdorff, J. M. J. aan de Brugh, O. P. Hasekamp, I. Aben, and J. Landgraf. The direct fitting approach for total ozone column retrievals: a sensitivity study on GOME-2/MetOp-A measurements. *Atmospheric Measurement Techniques*, 8:4429–4451, October 2015.

[38] S. Chandrasekhar. *Radiative transfer*. Dover Publications, Inc., New York, 1960.

[39] H. C. van de Hulst. *Light scattering by small particles*. J. Wiley and sons, New York, 1957.

[40] D. Wunch, G. C. Toon, P. O. Wennberg, S. C. Wofsy, B. B. Stephens, M. L. Fischer, O. Uchino, J. B. Abshire, P. Bernath, S. C. Biraud, J.-F. L. Blavier, C. Boone, K. P. Bowman, E. V. Browell, T. Campos, B. J. Connor, B. C. Daube, N. M. Deutscher, M. Diao, J. W. Elkins, C. Gerbig, E. Gottlieb, D. W. T. Griffith, D. F. Hurst, R. Jiménez, G. Keppel-Aleks, E. A. Kort, R. Macatangay, T. Machida, H. Matsueda, F. Moore, I. Morino, S. Park, J. Robinson, C. M. Roehl, Y. Sawa, V. Sherlock, C. Sweeney, T. Tanaka, and M. A. Zondlo. Calibration of the Total Carbon Column Observing Network using aircraft profile data. *Atmospheric Measurement Techniques*, pages 1351–1362, 2010.

[41] A. Bucholtz. Rayleigh-scattering calculations for the terrestrial atmosphere. *Appl. Optics*, 35:2765–2773, 1995.

[42] O. Dubovik, A. Sinyuk, T. Lapyonok, B. N. Holben, M. Mishchenko, P. Yang, T. F. Eck, H. Volten, O. MuñOz, B. Veihelmann, W. J. van der Zande, J.-F. Leon, M. Sorokin, and I. Slutsker. Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust. *J. Geophys. Res.*, 111:D11208, 2006.

[43] B. van Diedenhoven, A. S. Ackerman, B. Cairns, and A. M. Fridlind. A flexible parameterization for shortwave optical properties of ice crystals. *J. Atmos. Sci.*, 71(5):1763–1782, 2014.

[44] E. A. Ustinov. Methods of spherical harmonics: Application to the transfer of polarized radiation in a vertically non-uniform planetary atmosphere. mathematical apparatus. *Cosmic Res*, 26:473, 1988.

[45] I.M. Gel'fand, R.A. Minlos, and Z. Shapiro. *Representations of the rotation and Lorentz groups and their applications*. Pergamon Press, Oxford, 1963.

[46] I. Kuščer and M. Ribarič. Matrix formalism in the theory of diffusion of light. *Opt. Acta*, 6:42–51, 1959.

[47] W. A. de Rooij and C. C. A. H. van der Stap. Expansion of Mie scattering matrices in generalized spherical functions. *Astron. Astrophys.*, 131:237–248, 1984.

[48] O. Torres, R. Decae, P. Veefkind, and G. de Leeuw. Omi aerosol retrieval algorithm. *ATBD-OMI-03*, pages 47–69, 2001.

[49] G.A. d'Almeida, P. Koepke, and E.P. Shettle. *Atmospheric Aerosols: Global Climatology and Radiative Characteristics*. A. Deepak Publishing, Hampton, VA, ISA, 1991.

[50] Thomas W. Kirchstetter, T. Novakov, and Peter V. Hobbs. Evidence that the spectral dependence of light absorption by aerosols is affected by organic carbon. *Journal of Geophysical Research: Atmospheres*, 109(D21), 2004.

[51] A. Butz, O. P. Hasekamp, C. Frankenberg, and I. Aben. Retrievals of atmospheric CO_2 from simulated space-borne measurements of backscattered near-infrared sunlight: accounting for aerosol effects. *Appl. Optics*, 48:3322, 2009.

[52] A. Butz, O. P. Hasekamp, C. Frankenberg, J. Vidot, and I. Aben. CH_4 retrievals from space-based solar backscatter measurements: Performance evaluation against simulated aerosol and cirrus loaded scenes. *J. Geophys. Res.*, 115:D24302, 2010.

[53] W. Wanner, X. Li, and A. H. Strahler. On the derivation of kernels for kernel-driven models of bidirectional reflectance. *Journal of Geophysical Research: Atmospheres*, 100(D10):21077–21089, 1995.

[54] Kirk D. Knobelspiesse, Brian Cairns, Beat Schmid, Miguel O. Román, and Crystal B. Schaaf. Surface brdf estimation from an aircraft compared to modis and ground estimates at the southern great plains site. *Journal of Geophysical Research: Atmospheres*, 113(D20), 2008.

[55] X. Li and A. H. Strahler. Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: effect of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2):276–292, 1992.

[56] Wolfgang Lucht, Andrew Hyman, Alan Strahler, Mike Barnsley, Paul Hobson, and J.-P Muller. A comparison of satellite-derived spectral albedos to ground-based broadband albedo measurements modeled to satellite spatial scale for a semidesert landscape. *Remote Sensing of Environment*, 74:85–98, 10 2000.

[57] Hafizur Rahman, Bernard Pinty, and Michel M. Verstraete. Coupled surface-atmosphere reflectance (CSAR) model: 2. Semiempirical surface model usable with NOAA advanced very high resolution radiometer data. *Journal of Geophysical Research: Atmospheres*, 98(D11):20791–20801, 1993.

[58] Pavel Litvinov, Otto Hasekamp, and Brian Cairns. Models for surface reflection of radiance and polarized radiance: Comparison with airborne multi-angle photopolarimetric measurements and implications for modeling top-of-atmosphere measurements. *Remote Sensing of Environment*, 115(2):781–792, 2011.

[59] Fabienne Maignan, François-Marie Bréon, Emilie Fédèle, and Marc Bouvier. Polarized reflectances of natural surfaces: Spaceborne measurements and analytical modeling. *Remote Sensing of Environment*, 113(12):2642–2650, 2009.

[60] C. Cox and W. Munk. Statistics of the sea surface derived from sun glitter. *J. Mar. Res.*, 13:198–227, 1954.

[61] J. Chowdhary, B. Cairns, F. Waquet, K. Knobelspiesse, M. Ottaviani, J. Redemann, L. Travis, and M. Mishchenko. Sensitivity of multiangle, multispectral polarimetric remote sensing over open oceans to water-leaving radiance: Analyses of rsp data acquired during the milagro campaign. *Remote Sens. Environ.*, 118:284–308, 2012.

[62] R.C. Smith and K.S. Baker. Optical properties of the clearest natural waters. *Appl. Opt.*, 20:177–184, 1981.

[63] O. Hasekamp and J. Landgraf. A linearized vector radiative transfer model for atmospheric trace gas retrieval. *J. Quant. Spectrosc. Radiat. Transfer*, 75:221–238, 2002.

[64] D. Schepers, J. M. J. aan de Brugh, Ph. Hahne, A. Butz, O. P. Hasekamp, and J. Landgraf. LINTRAN v2.0: A linearised vector radiative transfer model for efficient simulation of satellite-born nadir-viewing reflection measurements of cloudy atmospheres. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 149:347–359, December 2014.

[65] W. Wiscombe. The delta-m method: rapid yet accurate radiative flux calculations for strongly asymmetric phase functions. *J. Atmos. Sci.*, 34:1408–1422, 1977.

[66] J.M.J. aan de Brugh. The lintran radiative transfer solver. Report, Netherlands Insitute for Space Research, SRON, Utrecht, The Netherlands, 2018.

[67] O. P. Hasekamp and A. Butz. Efficient calculation of intensity and polarization spectra in vertically inhomogeneous scattering and absorbing atmospheres. *J. Geophys. Res.*, 113:D20309, 2008.

[68] B. van Diedenhoven, O. P. Hasekamp, and J. Landgraf. Efficient vector radiative transfer calculations in vertically inhomogeneous cloudy atmospheres. *Appl. Optics*, 45:5993–6006, 2006.

[69] Q. Fu and K. N. Liou. On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres. *Journal of Atmospheric Sciences*, 49:2139–2156, November 1992.

[70] M. Duan, Q. Min, and J. Li. A fast radiative transfer model for simulating high-resolution absorption bands. *J. Geophys. Res.*, 110(D15), 2005. D15201.

[71] O. P. Hasekamp and J. Landgraf. Linearization of vector radiative transfer with respect to aerosol properties and its use in satellite remote sensing. *J. Geophys. Res.*, 110:D04203, 2005.

[72] A. Tikhonov. On the solution of incorrectly stated problems and a method of regularization. *Dokl. Akad. Nauk SSSR*, 151:501–504, 1963.

[73] C.D. Rodgers. *Inverse methods for Atmospheric Sounding: Theory and Practice*. World Sc.,River Edge, N. J., 2000.

[74] M.Wang and H.R. Gordon. Radiance reflected from the ocean-atmosphere system: synthesis from individual components of the aerosol size distribution. *Appl. Opt.*, 33:7088–7095, 1994.

[75] T. E. Taylor, C. W. O'Dell, C. Frankenberg, P. T. Partain, H. Q. Cronk, A. Savtchenko, R. R. Nelson, E. J. Rosenthal, A. Y. Chang, B. Fisher, G. B. Osterman, R. H. Pollock, D. Crisp, A. Eldering, and M. R. Gunson. Orbiting carbon observatory-2 (oco-2) cloud screening algorithms: validation against collocated modis and caliop data. *Atmospheric Measurement Techniques*, 9(3):973–989, 2016.

[76] C. Frankenberg, J. F. Meirink, P. Bergamaschi, A. P. H. Goede, M. Heimann, S. Körner, U. Platt, M. van Weele, and T. Wagner. Satellite chartography of atmospheric methane from sciamachy on board envisat: Analysis of the years 2003 and 2004. *Journal of Geophysical Research: Atmospheres*, 111(D7), 2006.

[77] J. Landgraf, J.M.J. aan de Brugh, L. Wu, O. Hasekamp, S. Rusli, H. van Heck, A. Butz, K. Graph, H. Boesch, L. Vogel, M. Buchwitz, C. Chen, D. Fuertes, Y. Karol, and O. Dubovik. Study on spectral sizing for co2 observations: Final report. Report SRON-CSS-TN-2020-002, SRON Netherlands Insitute for Space Research, Utrecht, The Netherlands, 2020.

[78] J.M.J. aan de Brugh. Orbit simulator. Report SRON-CSS-TN-2016-01, SRON Netherlands Insitute for Space Research, Utrecht, The Netherlands, 2016.

[79] C. Lacagnina, O. P. Hasekamp, and O. Torres. Direct radiative effect of aerosols based on PARASOL and OMI satellite observations. *Journal of Geophysical Research (Atmospheres)*, 122:2366–2388, February 2017.

[80] C. Lacagnina, O. P. Hasekamp, H. Bian, G. Curci, G. Myhre, T. van Noije, M. Schulz, R. B. Skeie, T. Takemura, and K. Zhang. Aerosol single-scattering albedo over the global oceans: Comparing PARASOL retrievals with AERONET, OMI, and AeroCom models estimates. *Journal of Geophysical Research (Atmospheres)*, 120:9814–9836, September 2016.

[81] Wunch D., Toon G.C., J.-F. L. Blavier, R. A Washenfelder, J Notholt, B. J Connor, D. W. T Griffith, V Sherlock, and P. O Wennberg. The total carbon column observing network. *Philos. T. R. Soc. A.*, 369(1943):2087–2112, May 2011.

[82] M. Frey, M. K. Sha, F. Hase, M. Kiel, T. Blumenstock, R. Harig, N. M. Surawicz, G.and Deutscher, J. E. Shiomi, K.and Franklin, H. Bösch, J. Chen, M. Grutter, H. Ohyama, Y. Sun, A. Butz, G. Mengistu Tsidu, D. Ene, D. Wunch, Z. Cao, O. Garcia, M. Ramonet, F. Vogel, , and J. Orphal. "building the collaborative carbon column observing network (coccon): long-term stability and ensemble performance of the em27/sun fourier transform spectrometer". *Atmos. Meas. Tech.*, 13:1513–1530, 2019.

[83] A. Butz, A. S. Dinger, N. Bobrowski, J. Kostinek, L. Fieber, C. Fischerkeller, G. B. Giuffrida, F. Hase, F. Klappenbach, J. Kuhn, P. Lübcke, L. Tirpitz, and Q. Tu. Remote sensing of volcanic co$_2$, hf, hcl, so$_2$, and bro in the downwind plume of mt. etna. *Atmospheric Measurement Techniques*, 10(1):1–14, 2017.

[84] F. Klappenbach, M. Bertleff, J. Kostinek, F. Hase, T Blumenstock, A. Agusti-Panareda, Razinger M., and A. Butz. "$\mathrm{x}co_2$ and $\mathrm{x}ch_4$ total column measurement during polarstern cruise ps83 (ant-xxix/10), north-south gradient". *PANGAEA*, 8:5023–5038, 2015.

[85] M. Knapp, R. Kleinschek, F. Hase, A. Agustí-Panareda, A. Inness, J. Barré, J. Landgraf, T. Borsdorff, S. Kinne, and A. Butz. Shipborne measurements of $\mathrm{x}co_2$, $\mathrm{x}ch_4$, and xco above the pacific ocean and comparison to cams atmospheric analyses and s5p/tropomi. *Earth System Science Data*, 13(1):199–211, 2021.

[86] T. Krings, B. Neininger, K. Gerilowski, S. Krautwurst, M. Buchwitz, J. P. Burrows, C. Lindemann, T. Ruhtz, D. Schüttemeyer, and H. Bovensmann. Airborne remote sensing and in situ measurements of atmospheric CO_2 to quantify point source emissions. *Atmospheric Measurement Techniques*, 11(2):721–739, 2018.

[87] A. Karion, C. Sweenney, P. Tans, and T. Newberger. AirCore: An Innovative Atmospheric Sampling System. *J. Atm. Ocean Tech.*, 27:1839–1849, 2010.

[88] E. A. Ustinov. Adjoint sensitivity analysis of radiative transfer equation: Temperature and gas mixing ratio weighting functions for remote sensing of scattering atmospheres in thermal IR. *J. Quant. Spectrosc. Radiat. Transfer*, 68:195–211, 2001.

[89] J. Landgraf, O. Hasekamp, and T. Trautmann. Linearization of radiative transfer with respect to surface properties. *J. Quant. Spectrosc. Radiat. Transfer*, 72:327–339, 2002.

[90] G.I. Marchuk. Equation for the value of information from weather satellites and formulation of inverse problems. *Cosmic Res*, 2:394–409, 1964.

[91] M.A. Box, S.A.W. Gerstl, and C. Simmer. Application of the adjoint formulation to the calculation of atmospheric radiative effects. *Beitr. Phys. Atmos.*, 61:303–311, 1988.

[92] G. I. Bell and S. Glasstone. *Nuclear Reactor Theory*. Van Nostrand Reinhold Company, New York, 1970.

[93] J. Lewins. *Importance, the adjoint function*. Pergamon Press, Oxford, England, 1965.

[94] H. Walter, J. Landgraf, and O. Hasekamp. Linearization of a pseudo-spherical radiative transfer model. *J. Quant. Spectrosc. Radiat. Transfer*, 85:251–283, 2004.

[95] M.A. Box, S.A.W. Gerstl, and C. Simmer. Computation of atmospheric radiative effects via perturbation theory. *Beitr. Phys. Atmos.*, 62:193–199, 1989.