	<p>S5L2PP Methane ATBD</p>	<p>Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 1/91</p>
--	--------------------------------	--

Sentinel 5 L2 Prototype Processors



Algorithm Theoretical Baseline Document: Methane Retrieval

J. Landgraf, A. Butz, O. Hasekamp, H. Hu, J. aan de Brugh

Prepared by : Jochen Landgraf Andre Butz Joost aan de Brugh	
Checked by : Jochen Landgraf Joost aan de Brugh	
Approved by : Jochen Landgraf	

Citation: Landgraf, J., Butz, A., Hasekamp, O., Hu, H., aan de Brugh, J., Sentinel 5 L2 Prototype Processors, Algorithm Theoretical Baseline Document: Methane Retrieval, SRON-ESA-S5L2PP-ATBD-001-v3.1-20190517-CH4, SRON Netherlands Institute for Space Research, Utrecht, The Netherlands, 2019

The copyright in this document is vested in SRON and University of Heidelberg. This document may only be reproduced in whole or in part, or stored in a retrieval system, or transmitted in any form, or by any means electronic, mechanical, photocopying or otherwise, either with the prior permission of SRON and University of Heidelberg or in accordance with the terms of ESA Contract No. 4000118463/16/NL/AI.



 	<p style="text-align: center;">S5L2PP Methane ATBD</p>	<p>Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 2/91</p>
--	--	--

Change Log

Issue	Date	Item	Comments
1.0	2017-02-15	All	Document creation
2.0	2017-10-02	Sec. 6.11	Section 'Detailed algorithm setup' added
2.0	2017-10-02	Sec. 7	Preliminary error analysis added based on S5P work
2.0	2017-10-02	Sec. 8	Section on product validation added
2.0	2017-10-02	Sec. 9	Section includes detailed input/output information and a first estimate on required computational resources.
2.1	2018-03-08	All	Applied corrections from PDR RIDs. Started error analysis for S5 (Sect. 7). General revision of whole document.
2.2	2018-11-14	Sec. 7	Extended error analysis.
2.2	2018-11-14	Sec. 9 and Appendix B	Added breakpoint output, revised processing quality flags. Added algorithm settings appendix.
2.2	2018-12-15	Sec. 9	Clarified remarks on coregistration.
3.0	2018-04-25	All	CDR RIDs are taken into account. In particular, the definition of spectral bands is corrected and I/O tables are updated

Table of Contents


Change Log	2
List of Tables	4
List of Figures	5
1 Introduction	8
1.1 Purpose and Objective	8
1.2 Document Overview	8
2 Applicable and Reference Documents	9
2.1 Applicable Documents	9
2.2 Reference Documents	9
3 Terms, Definitions and Abbreviated Terms	15
3.1 Terms and Definitions	15
3.2 Acronyms and Abbreviations	15
4 Introduction to Methane Retrieval algorithm	16
4.1 The S5 spectral range	16
4.2 Heritage	17
4.3 Requirements	17
5 Instrument Overview	20
6 Detailed Algorithm Description	21
6.1 Input	21
6.1.1 Absorption cross sections	21
6.1.2 METimage	23
6.2 Initial Data filtering A	23
6.3 Non-scattering retrievals	24
6.3.1 The model atmosphere	24
6.3.2 Non-scattering forward model	25
6.3.3 Non-scattering inversion	26
6.4 Proxy cloud filter B	29
6.5 Proxy retrieval	29
6.6 Cloud filter C of the physics-based retrieval	30
6.7 Fluorescence retrieval	31
6.8 Physics-based retrieval	33
6.8.1 Optical properties of the model atmosphere	33
6.8.2 Physics-based forward model	35
6.8.3 Physics-based inversion	39
6.9 Deconvolution solar spectrum	41
6.10 A posteriori data filtering and bias correction	42
6.11 Detailed Algorithm Setup	42
6.11.1 Serial Initialization	44
6.11.2 Parallel initialization	45
6.11.3 Retrievals	45
7 Error Analysis	50
7.1 Sensitivity to instrument errors	50
7.1.1 Generic ensembles	50

 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001
		Version : 3.1 Page
		Date : 17 May 2019 4/91

7.1.2	Noise analysis	51
7.1.3	ISRF distortion	52
7.1.4	Additive and multiplicative radiometric offset	53
7.1.5	Spectral alignment	53
7.1.6	Spectrometer stray light	58
7.2	Sensitivity to atmospheric input	63
7.2.1	Profiles of atmospheric trace gases	63
7.2.2	Temperature offset	63
7.2.3	Aerosol induced errors	65
8	Validation	69
8.1	Ground based	69
8.1.1	The Total Carbon Column Observatory Network (TCCON).....	69
8.1.2	In Situ Measurements	71
8.2	Satellite Intercomparison	71
9	Feasibility	72
9.1	Estimated Computational Effort	72
9.2	Inputs	72
9.2.1	Static Inputs	73
9.2.2	Dynamic Inputs	77
9.3	Outputs	80
9.3.1	Static Output	80
9.3.2	Dynamic Output	80
9.3.3	Breakpoint Output	82
	Appendices	83
A	Appendix A: LINTRAN V2.0	83
B	Algorithm settings	86
B.1	Filter settings	86
B.2	Fluorescence retrieval settings	86
B.3	Atmospheric gridding.....	87
B.4	Trace gas retrieval settings	87
C	Usage of L1b-flagging at framework level	89

List of Tables

1	Spectral ranges of the NIR-2a, NIR-2, SWIR-1 and SWIR-3 band [AD2, AD3, RD34].....	21
2	The S5 L1B and L2 Met data product used for cloud filtering, for more detail see Tab. 9.....	23
3	Spectral windows of the non-scattering retrievals and corresponding target species.	24
4	State vector elements and their a priori and first guess values for the non-scattering retrieval.....	27
5	Spectral fit windows of the physics-based retrieval.....	33
6	State vector elements and their a priori and first guess values for the physics-based three band retrieval. The state vector of a two band retrieval is defined by a subset, omitting the CO ₂ total column and all entries related to the SWIR-1 band. The symbol \mathbf{L}_1 and \mathbf{L}_o indicates first and zero order Tikhonov regularization, \emptyset indicates no side-constraint for these parameters.	39
7	Static input of the serial algorithm initialization.	44
8	Output data of the static initialization.....	45
9	Dynamic input to the CH ₄ retrieval algorithm. Metimage input as specified in Tab. 2.	46
10	Dynamic output data.	49

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001
		Version : 3.1 Page
		Date : 17 May 2019 5/91

11	Parameters a , b and N of the S5 instrument noise model in Eq. 105 at the reference wavelength λ_{ref} .	50
12	Reference scenario on which the generic ensembles are based.....	50
13	Interdependence of surface albedo in the three bands used for the generic ensemble for increasing scene brightness.	51
14	Mean XCH ₄ bias, standard deviation and data yield for the stray light analysis using the ensemble of Fig. 21. Corresponding numbers for a purely clear sky ensemble are given in parenthesis.....	60
15	Overview of TCCON stations (information from 2017).....	69
16	Static input data of RemoTeC-S5	73
17	Dynamic input of RemoTeC-S5.....	77
18	Output data of the static initialization	80
19	Dynamic output of RemoTeC. Processing Quality Flags are specified separately in Tab. 20.	80
20	Processing Quality Flags (PQFs) of RemoTeC. Four types of flags are provided: status (S), error (E), filter (F) and warning (W)	82
21	Breakpoint output of RemoTeC.	83
22	Filter settings used in RemoTeC.....	86
23	Fluorescence retrieval settings used in RemoTeC.	86
24	Atmospheric settings used in RemoTeC.	87
25	Window unspecific retrieval settings used in RemoTeC.....	87
26	Window settings used in RemoTeC.	87
27	Additional settings for physics-based retrieval used in RemoTeC.	88
28	Usage of L1b flags by algorithm.....	91



List of Figures

1	Simulated spectral transmission in the NIR spectral range for a light path from the TOA to the surface and finally to a nadir viewing satellite. Here we neglect any kind of atmospheric scattering and assume a solar zenith angle of 60 degree and an instrument viewing zenith angle of 0 degree. The spectrum is convolved with a Gaussian instrument spectral response with a resolution of 0.4 nm.	17
2	Same as Fig. 1 but for the SWIR-1 spectral window: total transmission (upper panel), contribution by CO ₂ (second panel), contribution by CH ₄ (third panel), and the water vapor contribution (fourth panel). The spectrum is convolved with a Gaussian instrument spectral response with a resolution of 0.25 nm. Note the different scales of the y-axes.	18
3	Same as Fig. 1 but for the SWIR-3 spectral window: total transmission (upper panel), contribution by CH ₄ (second panel), contribution by H ₂ O (third panel), and the CO contribution (fourth panel). Note the different scales of the y-axes.	19
4	Overall design of the RemoTeC algorithm including the CH ₄ proxy and physics-based retrieval branches. The input data include algorithm settings, lookup tables (LUTs) of absorption and scattering optical properties, the instrument spectral response function (ISRF) as part of the calibration key data set (CKD), atmospheric input generated from a digital elevation model (DEM) and Chemistry Transport Modeling (CTM), NIR-2, SWIR-1, and SWIR-3 S5 level 1B data (S5-L1B), and METimage L1B and cloud L2 products. The green boxes indicate the input and output layers interfacing the CH ₄ algorithm to the general S5 processing infrastructure. The blue boxes are the core parts of the RemoTeC CH ₄ retrieval algorithm.	22
5	Non-scattering retrieval results of CH ₄ and H ₂ O in the weak and strong absorption bands as percentage of the truth with a varying cloud height. Left panel: Thin cloud ($\tau = 2$). Right panel: Thick cloud ($\tau = 10$).	30
6	Particle size distribution $n_{\text{aer}}(r)$ as a function of particle radius r . The retrieval method relies on a power law (red solid) size distribution. Also shown are more realistic multi-modal lognormal size distributions for a fine mode (black dashed) and a coarse mode (black dotted) dominated aerosol type.	34

7	Relative error of the difference between a spectrum calculated using the linear- k method with respect to line-by-line calculations for the NIR-2 band (upper panel), SWIR-1 band (middle panel) and SWIR-3 band (lower panel). The spectra have been convolved with a Gaussian spectral response function with a Full Width at Half Maximum (FWHM) in Table 1. For the calculations a boundary layer aerosol was used with an optically thickness of 0.3 at 765 nm. Furthermore, we used a solar zenith angle (SZA) of 50° and a viewing zenith angle of 0°	38
8	RemoTeC algorithm architecture for initialization and pixel processing optimized for parallel computing. Here, two parallelization options are foreseen: first, the initialization of the algorithm with respect to the detector spatial channel index of the measurements and second the actual retrievals with respect to the pixels to be processed within a data granule.	43
9	Serial initialization, which is executed once per processor call. The solid green boxes indicate static input, the dashed green boxes indicate output to the remaining RemoTeC software provided by a pointer to the internal memory.	44
10	Parallel initialization: Based on input from the internal memory, the ISRF from the calibration key data is adapted to each spectral pixel. Moreover, the spectral shift between solar reference spectrum and observation is determined. This is needed as input for the subsequent de-convolution of the solar spectrum.	45
11	Software architecture of the S5 methane L2 algorithm. The shaded boxes indicated two specific calls of the retrieval model (non-scattering retrieval and physics-based retrieval), which is explained in more detail in Fig. 12.	47
12	Software architecture of the non-scattering and physics-based retrieval.	48
13	Vertical profiles of CH ₄ , CO ₂ , H ₂ O and CO used for the generic ensembles.	51
14	XCH ₄ precision of the physics-based and proxy retrieval product for the reference scenario but for varying surface albedo and solar zenith angle. Surface albedos in NIR-2, SWIR-1 and SWIR-3 are interdependent according to Tab. 13.	52
15	Averaging kernel as a function of solar zenith angle for a surface albedo of $(A_{\text{NIR}}, A_{\text{SWIR-1}}, A_{\text{SWIR-3}}) = (0.15, 0.1, 0.05)$	53
16	XCH ₄ retrieval bias due to distortions of the FWHM of the ISRF: (upper left) physics-based retrieval ISRF distortion in the NIR-2 band, (upper right) same but for the ISRF distortion in the SWIR-1 band, (lower left) same but for the ISRF distortion in the SWIR-3 band, (lower right) same but for the proxy product. Simulations are performed for the reference scenario in Tab. 12.	54
17	Same as Fig. 16 but for a varying additive radiometric offset.	55
18	Same as Fig. 16 but for a varying multiplicative radiometric offset.	56
19	Same as Fig. 16 but as a function of the spectral stretch parameter S	57
20	TROPOMI stray light kernel measured during the instrument on-ground calibration campaign (left). Stray light kernel used in this study (right).	59
21	(Left panel) MODIS cloud fraction for the stray light data granule, (right panel) surface elevation.	60
22	Retrieval bias of the physics-based and proxy XCH ₄ product for stray light in the individual bands.	61
23	Bias distribution function due to stray light in the different bands of the physics-based retrieval and the proxy retrieval.	62
24	XCH ₄ retrieval bias of the physics-based and proxy products with varying added H ₂ O in the boundary layer.	63
25	Same as Fig. 24 but with varying added CH ₄ in the boundary layer.	64
26	Same as Fig. 24 but with varying added CO ₂ in the boundary layer.	64
27	Retrieval bias of the physics-based and proxy XCH ₄ products with varying shifts of the atmospheric temperature.	65
28	Retrieval bias of the physics-based and proxy XCH ₄ product with varying surface albedo and solar zenith angle.	65
29	Retrieval bias of the physics-based and proxy XCH ₄ product with varying boundary layer aerosol optical depth at 765 nm.	66
30	Retrieval bias of the physics-based and proxy XCH ₄ products for the global ensemble. No filters apart from the convergence check are applied.	67
31	Bias distribution function of Fig. 30 for the physics-based retrieval and the proxy retrieval	68

 	<p style="text-align: center;">S5L2PP Methane ATBD</p>	<p>Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 7/91</p>
--	--	--

32 Range of the central 80% of the land albedo values in an area of about $500 \times 500 \text{ km}^2$ around each TCCON station, except Ascension Island, for which no land data is available. A dot indicates the median value. Sea values are excluded, but they may contaminate coastal land. 70

 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 8/91
--	------------------------	---


1 Introduction

1.1 Purpose and Objective

This document describes the algorithm for the retrieval of the column average dry air mixing ratio of methane, X_{CH_4} , from Sentinel-5 (S5) measurements in the near infrared (NIR) and shortwave infrared (SWIR) spectral range. The algorithm name is RemoTeC-S5 and it is one of the deliverables of the ESA project 'Sentinel-5 level 2 prototype processor' [AD5]. The purpose of the document is to describe the theoretical baseline of the algorithm, the input and ancillary data that are needed, and the output that will be generated. In addition, information about expected calculation times, and the expected accuracy are provided.

1.2 Document Overview

The document is structured as follows: References are provided in Chapter 2 and Chapter 3 contains a list of abbreviations used in this document. Chapter 4 provides a short introduction to satellite remote sensing of atmospheric CH_4 abundance and the heritage of the presented algorithm. Moreover, we recall the level-2 requirement for the CH_4 product, which represents the underlying criterion for the performance analysis of the presented algorithm. The theoretical concept of the CH_4 retrieval algorithm RemoTeC is summarized in Chapter 6, comprising a description of the lightpath proxy and the physics-based retrieval scheme. The parameters to be retrieved, ancillary data and a priori knowledge are discussed including the final data product of the algorithm. Chapter 7 considers the performance of the retrieval algorithm based on a set of generic measurement ensembles and a global geo-physical ensemble of simulated measurements for four days representing the four seasons. Here, we investigate the CH_4 retrieval noise and CH_4 retrieval biases due to forward model errors, erroneous atmospheric input data and instrument artifacts. The validation approach of the CH_4 data product is broached in Chapter 8. Based on the retrievals of simulated measurements, we evaluate the algorithm performance in the context of the S5 level-1 and 2 requirements. The numerical feasibility is the subject of Chapter 9, which comprises an estimate of the numerical effort, a high level data product description and the spatial data selection criteria of the measurements to be processed. Additional material is provided in the appendices.

	<p style="text-align: center;">S5L2PP Methane ATBD</p>	<p>Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 9/91</p>
---	--	--

2 Applicable and Reference Documents

2.1 Applicable Documents

- [AD1] S5L2PP Project Management Plan.
source: S&T; **ref:** ST-ESA-S5L2PP-PMP-001; **issue:** 2.4; **date:** 2019-05-17.
- [AD2] GMES Sentinels 4 and 5 mission requirements document.
source: ESA; **ref:** EOP-SMA/1507/JL-dr; **date:** 2011.
- [AD3] GMES Sentinels 4 and 5 mission requirements traceability document.
source: ESA; **ref:** EOP-SM/2413/BV-bv; **date:** 2017.
- [AD4] Sentinel-5/UVNS L1BPP Algorithm Theoretical Baseline Document.
source: AIRBUS; **ref:** GS5.RP.ASG.UVNS.00044; **issue:** 8.0; **date:** 03-10-2017.
- [AD5] Sentinel-5 Level-2 Prototype Processor Development - Statement of Work.
source: ESA; **ref:** S5-SOW-ESA-GR-0130; **issue:** 1.7; **date:** 2018-06-28.
- [AD6] Sentinel-5 Level-2 Prototype Processor Development Requirements Specification.
source: ESA; **ref:** S5-RS-ESA-GR-0131; **issue:** 1.7; **date:** 2018-06-29.
- [AD7] Requirements for the Geophysical Validation of Sentinel-5 Precursor Products, draft version.
source: ESA; **ref:** S5P-RS-ESA-SY-164; **date:** 2014.

2.2 Reference Documents


- [RD1] S5L2PP Terms, Definitions and Abbreviations.
source: S&T; **ref:** ST-ESA-S5L2PP-LST-001; **issue:** 2.1; **date:** 2018-08-28.
- [RD2] S. Solomon, D. Qin, M. Manning *et al.* (editors) *IPCC, Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007).
- [RD3] H. Bovensmann, J. P. Burrows, M. Buchwitz *et al.*; SCIAMACHY: Mission Objectives and Measurement Modes. *Journal of Atmospheric Sciences*; **56** (1999), 127.
- [RD4] T. Yokota, H. Oguma, I. Morino *et al.*; Test measurements by a BBM of the nadir-looking SWIR FTS aboard GOSAT to monitor CO₂ column density from space. *In Passive Optical Remote Sensing of the Atmosphere and Clouds IV* (edited by S. C. Tsay, T. Yokota and M.-H. Ahn); volume 5652 of *Proceedings of the SPIE*; (pp. 182–188) (2004); 10.1117/12.578497.
- [RD5] A. Kuze, H. Suto, M. Nakajima *et al.*; Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring. *Appl. Opt.*; **48** (2009), 6716; 10.1364/AO.48.006716.
- [RD6] P. Veefkind; TROPOMI on the ESA Sentinel-5 Precursor: a GMES mission for Global Observations of the Atmospheric Composition for Climate and Air Quality Applications. *Remote Sens. Environ.*; **120** (2012), 70.
- [RD7] J. F. Meirink, H. J. Eskes and A. P. H. Goede; Sensitivity analysis of methane emissions derived from SCIAMACHY observations through inverse modelling. *Atmos. Chem. Phys.*; **6** (2006), 1275.
- [RD8] P. Bergamaschi, C. Frankenberg, J. F. Meirink *et al.*; Satellite cartography of atmospheric methane from SCIAMACHY on board ENVISAT: 2. Evaluation based on inverse model simulations. *J. Geophys. Res.*; **112** (2007), D02304; 10.1029/2006JD007268.
- [RD9] A. Butz, O. P. Hasekamp, C. Frankenberg *et al.*; CH₄ retrievals from space-based solar backscatter measurements: Performance evaluation against simulated aerosol and cirrus loaded scenes. *J. Geophys. Res.*; **115** (2010), D24302; 10.1029/2010JD014514.

	<p style="text-align: center;">S5L2PP Methane ATBD</p>	<p>Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 10/91</p>
---	--	---



- [RD10] I. Aben, O. Hasekamp and W. Hartmann; Uncertainties in the space-based measurements of CO₂ columns due to scattering in the Earth's atmosphere. *J. Quant. Spectrosc. Radiat. Transfer*; **104** (2007), 450; 10.1016/j.jqsrt.2006.09.013.
- [RD11] A.M.S. Gloudemans, H. Schrijver, O.P. Hasekamp *et al.*; Error analysis for CO and CH₄ total column retrieval from SCIAMACHY 2.3 μ m spectra. *Atmos. Chem. Phys.*; **8** (2008), 3999.
- [RD12] C. Frankenberg, U. Platt and T. Wagner; Retrieval of CO from SCIAMACHY onboard ENVISAT: detection of strongly polluted areas and seasonal patterns in global CO abundances. *Atmos. Chem. Phys.*; **4** (2005), 8425.
- [RD13] C. Frankenberg, P. Bergamaschi, A. Butz *et al.*; Tropical methane emissions: A revised view from SCIAMACHY onboard ENVISAT. *Geophys. Res. Lett.*; **35** (2008), 15811; 10.1029/2008GL034300.
- [RD14] O. Schneising, M. Buchwitz, M. Reuter *et al.*; Long-term analysis of carbon dioxide and methane column-averaged mole fractions retrieved from SCIAMACHY. *Atmos. Chem. Phys.*; **11** (2011), 2863; 10.5194/acp-11-2863-2011.
- [RD15] R. Parker, H. Bösch, A. Cogan *et al.*; Methane observations from the Greenhouse Gases Observing SATellite: Comparison to ground-based TCCON data and model calculations. *Geophys. Res. Lett.*; **38** (2011), L15807; 10.1029/2011GL047871.
- [RD16] D. Schepers, S. Guerlet, A. Butz *et al.*; Methane retrievals from Greenhouse Gases Observing Satellite (GOSAT) shortwave infrared measurements: Performance comparison of proxy and physics retrieval algorithms. *J. Geophys. Res.*; **117** (2012), D10307; 10.1029/2012JD017549.
- [RD17] H. Bösch, G. C. Toon, B. Sen *et al.*; Space-based near-infrared CO₂ measurements: Testing the Orbiting Carbon Observatory retrieval algorithm and validation concept using SCIAMACHY observations over Park Falls, Wisconsin. *J. Geophys. Res.*; **111** (2006), D23302; 10.1029/2006JD007080.
- [RD18] B. J. Connor, H. Bösch, G. Toon *et al.*; Orbiting Carbon Observatory: Inverse method and prospective error analysis. *J. Geophys. Res.*; **113** (2008), D05305; 10.1029/2006JD008336.
- [RD19] S. Oshchepkov, A. Bril and T. Yokota; PPDF-based method to account for atmospheric light scattering in observations of carbon dioxide from space. *J. Geophys. Res.*; **113** (2008), D23210; 10.1029/2008JD010061.
- [RD20] A. Butz, O. P. Hasekamp, C. Frankenberg *et al.*; Retrievals of atmospheric CO₂ from simulated space-borne measurements of backscattered near-infrared sunlight: accounting for aerosol effects. *Appl. Opt.*; **48** (2009), 3322; 10.1364/AO.48.003322.
- [RD21] A. Bril, S. Oshchepkov and T. Yokota; Retrieval of atmospheric methane from high spectral resolution satellite measurements: a correction for cirrus cloud effects. *Appl. Opt.*; **48** (2009), 2139; 10.1364/AO.48.002139.
- [RD22] M. Reuter, M. Buchwitz, O. Schneising *et al.*; A method for improved SCIAMACHY CO₂ retrieval in the presence of optically thin clouds. *Atmos. Meas. Tech.*; **3** (2010), 209.
- [RD23] A. Butz, S. Guerlet, O. Hasekamp *et al.*; Toward accurate CO₂ and CH₄ observations from GOSAT. *Geophys. Res. Lett.*; **38** (2011), doi:10.1029/2011GL047888.
- [RD24] H. Bösch, D. Baker, B. Connor *et al.*; Global Characterization of CO₂ Column Retrievals from Shortwave-Infrared Satellite Observations of the Orbiting Carbon Observatory-2 Mission. *Remote Sens. Environ.*; **3** (2011), 270; 10.3390/rs3020270.
- [RD25] Y. Yoshida, Y. Ota, N. Eguchi *et al.*; Retrieval algorithm for CO₂ and CH₄ column abundances from short-wavelength infrared spectral observations by the Greenhouse gases observing satellite. *Atmos. Meas. Tech.*; **4** (2011), 717; 10.5194/amt-4-717-2011.
- [RD26] C. W. O'Dell, B. Connor, H. Bösch *et al.*; The ACOS CO₂ retrieval algorithm - Part 1: Description and validation against synthetic observations. *Atmos. Meas. Tech.*; **5** (2012), 99; 10.5194/amt-5-99-2012.
- [RD27] Sentinel-5/UVNS L1BPP Algorithm Theoretical Basis Document (ATBD).
source: AIRBUS; **ref:** GS5.RP.ASG.UVNS.00044; **issue:** 7.0; **date:** 2017-05-08.
- [RD28] Consolidation of S5-SWIR requirements, Final report.
source: Netherlands Institute for Space Research, SRON; **ref:** RfQ 3-13741/12/NL/CT/If; **date:** 2014.

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 11/91
---	------------------------	--



- [RD29] O. P. Hasekamp and A. Butz; Efficient calculation of intensity and polarization spectra in vertically inhomogeneous scattering and absorbing atmospheres. *J. Geophys. Res.*; **113** (2008), D20309; 10.1029/2008JD010379.
- [RD30] S. Guerlet, A. Butz, D. Schepers *et al.*; Impact of aerosol and thin cirrus on retrieving and validating XCO₂ from GOSAT shortwave infrared measurements. *J. Geophys. Res.*; **118** (2013), 4887; 10.1002/jgrd.50332.
- [RD31] H. Hu, O. Hasekamp, A. Butz *et al.*; The operational methane retrieval algorithm for TROPOMI. *Atmospheric Measurement Techniques*; **9** (2016), 5423; 10.5194/amt-9-5423-2016.
- [RD32] Algorithm Theoretical Baseline Document for Sentinel-5 Precursor methane retrieval.
source: SRON; **ref:** SRON-S5P-LEV2-RP-001; **date:** 2016.
- [RD33] S. Guerlet, S. Basu, A. Butz *et al.*; Reduced carbon uptake during the 2010 Northern Hemisphere summer from GOSAT. *Geophys. Res. Lett.*; **40** (2013), 2378; 10.1002/grl.50402.
- [RD34] Sentinel-5/UVNS L1BPP Algorithm Theoretical Baseline Document.
source: AIRBUS; **ref:** GS5.RP.ASG.UVNS.00044; **issue:** 8.0; **date:** 03-10-2017.
- [RD35] Sentinel 5 L2 Prototype Processors: Algorithm Theoretical Baseline Document S5-METImage Product.
source: STFC Rutherford Appleton Laboratory; **ref:** RAL-ESA-S5L2PP-ATBD-002; **issue:** 3.0; **date:** 15-12-2018.
- [RD36] J.J. Danielson and D.B. Gesch; Global multi-resolution terrain elevation data 2010 (GMTED2010). *U.S. Geological Survey Open-File Report*; **2011-1073** (2011), 26.
- [RD37] A. Jenouvrier, L. Daumont, L. Régalia-Jarlot *et al.*; Fourier transform measurements of water vapor line parameters in the 4200 6600 cm⁻¹ region. *J. Quant. Spectrosc. Radiat. Transfer*; **105** (2007), 326; 10.1016/j.jqsrt.2006.11.007.
- [RD38] L.S. Rothman, I.E. Gordon, A. Barbe *et al.*; The HITRAN 2008 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transfer*; **110** (2009), 533.
- [RD39] R. A. Scheepmaker, C. Frankenberg, A. Galli *et al.*; Improved water vapour spectroscopy in the 4174–4300 cm⁻¹ region and its impact on SCIAMACHY HDO/H₂O measurements. *Atmos. Meas. Tech.*; **6** (2013) (4), 879; 10.5194/amt-6-879-2013. URL <http://www.atmos-meas-tech.net/6/879/2013/>.
- [RD40] H. Tran, C. Boulet and J.-M. Hartmann; Line mixing and collision-induced absorption by oxygen in the A band: Laboratory measurements, model, and tools for atmospheric spectra computations. *J. Geophys. Res.*; **111** (2006), D15210; 10.1029/2005JD006869.
- [RD41] C. Frankenberg, D. Wunch, G. Toon *et al.*; Water vapor isotopologue retrievals from high-resolution GOSAT shortwave infrared spectra. *Atmospheric Measurement Techniques*; **6** (2013), 263; 10.5194/amt-6-263-2013.
- [RD42] D. Wunch, G. C. Toon, P. O. Wennberg *et al.*; Calibration of the Total Carbon Column Observing Network using aircraft profile data. *Atmospheric Measurement Techniques*; (2010), 1351; 10.5194/amt-3-1351-2010.
- [RD43] C.D. Rodgers; *Inverse Methods for Atmospheres: Theory and Practice*; volume 2 (World Scientific, 2000).
- [RD44] T. E. Taylor, C. W. O'Dell, C. Frankenberg *et al.*; Orbiting Carbon Observatory-2 (OCO-2) cloud screening algorithms; validation against collocated MODIS and CALIOP data. *Atmospheric Measurement Techniques Discussions*; **8** (2015), 12663; 10.5194/amtd-8-12663-2015.
- [RD45] B. van Diedenhoven, O. P. Hasekamp and I. Aben; Surface pressure retrieval from SCIAMACHY measurements in the O₂ A Band: validation of the measurements and sensitivity on aerosols. *Atmos. Chem. Phys.*; **5** (2005), 2109.
- [RD46] Algorithm Theoretical Baseline Document for Sentinel-5 Precursor: Carbon Monoxide Total Column Retrieval.
source: SRON; **ref:** SRON-S5P-LEV2-RP-002; **date:** 2016.
- [RD47] L. Guanter, I. Aben, P. Tol *et al.*; Potential of the TROPospheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor for the monitoring of terrestrial chlorophyll fluorescence. *Atmospheric Measurement Techniques*; **8** (2015), 1337; 10.5194/amt-8-1337-2015.

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 12/91
---	------------------------	--


- [RD48] Consolidation of Requirements for the Products Derived from the Shortwave-Infrared Channels of the GMES Sentinel-5 UVNS Instrument.
source: Karlsruhe Institute of Technology (KIT); **ref:** RfQ 3-13741/12/NL/CT/If; **date:** 2013.
- [RD49] A. Bucholtz; Rayleigh-Scattering calculations for the terrestrial atmosphere. *Appl. Opt.*; **35** (1995), 2765.
- [RD50] J. E. Hansen and L. D. Travis; Light scattering in planetary atmospheres. *Space Science Reviews*; **16** (1974), 527; 10.1007/BF00168069.
- [RD51] M. I. Mishchenko, I. V. Geogdzhayev, B. Cairns *et al.*; Aerosol retrievals over the ocean by use of channels 1 and 2 AVHRR data: sensitivity analysis and preliminary results. *Appl. Opt.*; **38** (1999), 7325; 10.1364/AO.38.007325.
- [RD52] P. Stier, J. Feichter, S. Kinne *et al.*; The aerosol-climate model ECHAM5-HAM. *Atmos. Chem. Phys.*; **5** (2005), 1125.
- [RD53] O. Dubovik, A. Sinyuk, T. Lapyonok *et al.*; Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust. *J. Geophys. Res.*; **111** (2006), D11208; 10.1029/2005JD006619.
- [RD54] J. F. de Haan, P. B. Bosma and J. W. Hovenier; The adding method for multiple scattering calculations of polarized light. *Astron. and Astrophys.*; **183** (1987), 371.
- [RD55] The Lintran radiative transfer solve.
source: Netherlands Institute for Space Research, SRON; **ref:** SRON-SWD-TN-2018-001; **date:** 2018.
- [RD56] B. van Diedenoven, O. P. Hasekamp and J. Landgraf; Efficient vector radiative transfer calculations in vertically inhomogeneous cloudy atmospheres. *Appl. Opt.*; **45** (2006), 5993; 10.1364/AO.45.005993.
- [RD57] Q. Fu and K. N. Liou; On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres. *Journal of Atmospheric Sciences*; **49** (1992), 2139.
- [RD58] M. Duan, Q. Min and J. Li; A fast radiative transfer model for simulating high-resolution absorption bands. *J. Geophys. Res.*; **110** (2005) (D15); 10.1029/2004JD005590. D15201; URL <http://dx.doi.org/10.1029/2004JD005590>.
- [RD59] 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** (2005), D04203; 10.1029/2004JD005260.
- [RD60] C. Frankenberg, C. O'Dell, L. Guanter *et al.*; Remote sensing of near-infrared chlorophyll fluorescence from space in scattering atmospheres: implications for its retrieval and interferences with atmospheric CO₂ retrievals. *Atmos. Meas. Tech.*; **5** (2012), 2081; 10.5194/amt-5-2081-2012.
- [RD61] C. D. Rodgers and B. J. Connor; Intercomparison of remote sounding instruments. *Journal of Geophysical Research (Atmospheres)*; **108** (2003), 4116; 10.1029/2002JD002299.
- [RD62] 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. Opt.*; **46** (2007), 243; 10.1364/AO.46.000243.
- [RD63] A. Wassmann, T. Borsdorff, J. M. J. aan de Brugh *et al.*; The direct fitting approach for total ozone column retrievals: a sensitivity study on GOME-2/MetOp-A measurements. *Atmospheric Measurement Techniques*; **8** (2015), 4429; 10.5194/amt-8-4429-2015.
- [RD64] Algorithm Theoretical Baseline Document for the RemoTeC XCH₄ Full Physics Product v.2.3.8 of the Essential Climate Variable (ECV) Greenhouse Gases (GHG).
source: SRON; **ref:** version 5.0; **date:** 2016.
- [RD65] Algorithm Theoretical Baseline Document Version 2 (ATBDv2)- The Bremen Optimal Estimation DOAS (BESD) algorithm for the retrieval of XCO₂ for the Essential Climate Variable (ECV) Greenhouse Gases (GHG).
source: IUP, University Bremen; **ref:** ; **date:** 2013.
- [RD66] TROPOMI CO and CH₄ data processing on a GRID infrastructure.
source: SRON Netherlands Institute for Space Research; **ref:** SRON-EARTH-TN-2017-01.

 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 13/91
--	------------------------	--

- [RD67] T. G. Farr, P. A. Rosen, E. Caro *et al.*; The Shuttle Radar Topography Mission. *Reviews of Geophysics*; **45** (2007), RG2004; 10.1029/2005RG000183.
- [RD68] A. Butz, A. Galli, O.P. Hasekamp *et al.*; ROPOMI aboard Sentinel-5 Precursor: Prospective performance of CH₄ retrievals for aerosol and cirrus loaded atmospheres. *Remote Sens. Environ.*; **120** (2012), 267.
- [RD69] L. A. Remer, Y. J. Kaufman, D. Tanré *et al.*; The MODIS Aerosol Algorithm, Products, and Validation. *Journal of Atmospheric Sciences*; **62** (2005), 947; 10.1175/JAS3385.1.
- [RD70] D. M. Winker, W. H. Hunt and M. J. McGill; Initial performance assessment of CALIOP. *Geophys. Res. Lett.*; **34** (2007), L19803; 10.1029/2007GL030135.
- [RD71] H. Schrijver, A. M. S. Gloudemans, C. Frankenberg *et al.*; Water vapour total columns from SCIAMACHY spectra in the 2.36 μm window. *Atmos. Meas. Tech.*; **2** (2009), 561.
- [RD72] D. Wunch, G. C. Toon, J.-F. L. Blavier *et al.*; The Total Carbon Column Observing Network. *Philos. T. R. Soc. A.*; **369** (2011) (1943), 2087; 10.1098/rsta.2010.0240.
- [RD73] M. Reuter, H. Bösch, H. Bovensmann *et al.*; A joint effort to deliver satellite retrieved atmospheric CO₂ concentrations for surface flux inversions: the ensemble median algorithm EMMA. *Atmos. Chem. Phys.*; **13** (2013), 1771; 10.5194/acp-13-1771-2013.
- [RD74] C. Petri, T. Warneke, N. Jones *et al.*; Remote sensing of CO₂ and CH₄ using solar absorption spectrometry with a low resolution spectrometer. *Atmos. Meas. Tech.*; **5** (2012), 1627; 10.5194/amt-5-1627-2012.
- [RD75] M. Gisi, F. Hase, S. Dohe *et al.*; XCO₂-measurements with a tabletop FTS using solar absorption spectroscopy. *Atmos. Meas. Tech.*; **5** (2012), 2969; 10.5194/amt-5-2969-2012.
- [RD76] A. Karion, C. Sweeney, P. Tans *et al.*; AirCore: An Innovative Atmospheric Sampling System. *J. Atm. Ocean Tech.*; **27** (2010), 1839; doi:10.1175/2010JTECHA1448.1.
- [RD77] Algorithm Theoretical Baseline Document for the RemoTeC XCH₄ Full Proxy Product v.2.3.8 of the Essential Climate Variable (ECV) Greenhouse Gases (GHG).
source: SRON; **ref:** version 5.0; **date:** 2016.
- [RD78] SURFsara GRID documentation.
url: URL <http://doc.grid.surfsara.nl/en/latest/>.
- [RD79] Sentinel 5 L2 Prototype Processors, Co-registration processing description.
source: TriOpSys; **ref:** TOS-S5L2PP-TN-1501; **issue:** 2.0; **date:** 15-12-2018.
- [RD80] 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** (2001), 195.
- [RD81] J. Landgraf, O.P. Hasekamp and T. Trautmann; Linearization of radiative transfer with respect to surface properties. *J. Quant. Spectrosc. Radiat. Transfer*; **72** (2002), 327.
- [RD82] G.I. Marchuk; Equation for the Value of Information from Weather Satellites and Formulation of Inverse Problems. *Cosmic Res.*; **2** (1964), 394.
- [RD83] 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** (1988), 303.
- [RD84] D. Schepers, J.M.J. van de Brugh, Ph. Hahne *et al.*; LINTRAN v2.0: A linearised vector radiative transfer model for efficient simulation of satellite-borne nadir-viewing reflection measurements of cloudy atmospheres. *Journal of Quantitative Spectroscopy and Radiative Transfer*; **149** (2014), 347 ;
<https://doi.org/10.1016/j.jqsrt.2014.08.019>. URL
<http://www.sciencedirect.com/science/article/pii/S002240731400363X>.
- [RD85] G. I. Bell and S. Glasstone; *Nuclear Reactor Theory* (Van Nostrand Reinhold Company, New York, 1970).
- [RD86] J. Lewins; *Importance, the adjoint function* (Pergamon Press, Oxford, England, 1965).

 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 14/91
--	------------------------	--

- [RD87] H. H. Walter, J. Landgraf and O. P. Hasekamp; Linearization of a pseudo-spherical vector radiative transfer model. *J. Quant. Spectrosc. Radiat. Transfer*; **85** (2004), 251.
- [RD88] M.A. Box, S.A.W. Gerstl and C. Simmer; Computation of atmospheric radiative effects via perturbation theory. *Beitr. Phys. Atmos.*; **62** (1989), 193.
- [RD89] O.P. Hasekamp and J. Landgraf; A linearized vector radiative transfer model for atmospheric trace gas retrieval. *J. Quant. Spectrosc. Radiat. Transfer*; **75** (2002).

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 15/91
---	------------------------	--

3 Terms, Definitions and Abbreviated Terms



Terms, definitions and abbreviated terms that are used in development program for the Sentinel 5 data processors are described in [RD1]. Terms, definitions and abbreviated terms that are specific for this document can be found below.

3.1 Terms and Definitions

There are no document specific terms and definitions.

3.2 Acronyms and Abbreviations

ATBD	Algorithm Theoretical Baseline Document
ECMWF	European Centre for Medium-Range Weather Forecasts
ENVISAT	Environmental Monitoring Satellite
GMTED2010	Global Multi-resolution Terrain Elevation Data 2010
GOSAT	Greenhouse gas Observing Sattellite
METimage	Advanced multispectral imaging radiometer for meteorological applications
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
NPP	NPOESS Preparatory Project
NIR	Near Infrared
OCO	Orbiting Carbon Observatory
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Chartography
SWIR	Shortwave Infrared
S5P	Sentinel-5 Precursor
S5	Sentinel-5
TCCON	Total Carbon Column Observing Network
TROPOMI	Tropospheric Monitoring Instrument
UVN	Ultraviolet, Visible, Near-Infrared
VIIRS	Visible Infrared Imaging Radiometer Suite
XCH ₄	Column averaged dry air mixing ratio of methane

 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 16/91
--	------------------------	--

4 Introduction to Methane Retrieval algorithm

Methane (CH_4) is, after carbon dioxide (CO_2), the most important contributor to the anthropogenically enhanced greenhouse effect [RD2]. Monitoring CH_4 abundances in the earth's atmosphere is the dedicated goal of several current and future satellite missions. Such space borne observations aim at providing CH_4 column concentrations with high sensitivity at the earth's surface, with good spatiotemporal coverage, and with sufficient accuracy to facilitate inverse modeling of sources and sinks. The Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) on board ENVISAT [RD3], which was operational 2002-2012, the Greenhouse Gases Observing Satellite (GOSAT) [RD4], [RD5], launched 2009, and the Sentinel 5 Precursor (S5P) [RD6], launched 2017, missions have the capability to achieve these goals. Their observation strategy relies on measuring spectra of sunlight backscattered by the earth's surface and atmosphere in the shortwave infrared (SWIR) spectral range. Absorption features of CH_4 allow for retrieval of its atmospheric concentration with high sensitivity to the ground and the lower atmosphere where the major CH_4 sources are located. The benefit of such measurements for estimating source/sink strengths, however, strongly depends on the precision and accuracy of the data product. When correlated on the regional or seasonal scale, systematic biases of a few tenths of a percent can jeopardize the usefulness of satellite-measured CH_4 concentrations for source/sink estimates [RD7, RD8, RD9]. Scattering by aerosols and cirrus clouds is the major challenge for retrievals of methane from space-borne observations of backscattered sunlight in the SWIR spectral range. While contamination by optically thick clouds can be filtered out reliably, optically thin scatterers are much harder to detect yet still modify the light path of the observed backscattered sunlight and thus, can lead to underestimation or overestimation of the true methane column if not appropriately accounted for. This light path effect strongly depends on the amount, the microphysical properties, and the height distribution of the scatterers as well as on the reflectance of the underlying ground surface [RD10, RD11]. Therefore, retrieval strategies rely on inferring CH_4 either simultaneously with scattering properties of the atmosphere or with a light path proxy. The latter 'proxy' approach has been successfully implemented for methane retrieval from SCIAMACHY measurements around $1.6 \mu\text{m}$, by using the CO_2 column, also retrieved from SCIAMACHY in the same spectral range, as a lightpath proxy [RD12]. The 'proxy' approach relies on the assumptions that scattering effects cancel in the ratio of the CH_4 column and the CO_2 column and that a prior estimate of the CO_2 column is sufficiently accurate to reliably re-calculate the methane column from the ratio. Further applications of the proxy approach for methane retrieval from SCIAMACHY are described by Frankenberg et al., 2008 [RD13] and Schneising et al., 2011 [RD14]. For GOSAT, the proxy approach has been successfully applied by Parker et al., 2011 [RD15] and Schepers et al., 2012 [RD16].

Alternatively, scattering induced lightpath modification can be taken into account by simultaneously inferring the atmospheric CH_4 concentration and physical scattering properties of the atmosphere. Such 'physics-based' methods have been developed for space-based CO_2 and CH_4 measurements from SCIAMACHY, GOSAT, and the Orbiting Carbon Observatory (OCO) (e.g. [RD17, RD18, RD19, RD20, RD21, RD22, RD23, RD24, RD25, RD26]). The physics based methods make use of the oxygen A band around 760 nm and absorption bands of the target absorber in the SWIR spectral range. The advantage of physics based methods for methane retrieval compared to a proxy method is that they do not depend on accurate prior information on the CO_2 column. On the other hand, the physics based algorithms are more complex and may be limited by the information content of the measurement with respect to aerosol properties and/or forward model errors in the description of aerosols. A detailed comparison between the two methods for GOSAT is provided by Schepers et al., 2012 [RD16].

4.1 The S5 spectral range

The spectral range measured by the S5 spectrometer (see [RD27]) allows for both a light-path-proxy and the physics based method and this ATBD describes both approaches. The spectral ranges to be used are shown in Figures 1–3. The proxy approach relies on measurements in the $1.6 \mu\text{m}$ range only and uses a non-scattering radiative transfer model. It results in a fast and numerically robust algorithm. The physics based method may operate on two different combinations of spectral bands. Here the combination of the NIR, the SWIR-1 ($1.6 \mu\text{m}$), and the SWIR-3 ($2.3 \mu\text{m}$) range is preferable with respect to the retrieval precision but puts stringent requirements on the spatial collocations of the three bands ([RD28]). The combination of the NIR and the $2.3 \mu\text{m}$ range reflects the window selection of the operational S5P algorithm. In both cases, the physics based algorithm retrieves 3 aerosol parameters (amount, size, height) simultaneously with the methane column (and other parameters such as surface albedo) in order to account for light path modification by aerosols. The information on aerosol parameters comes from the parts of the spectrum with

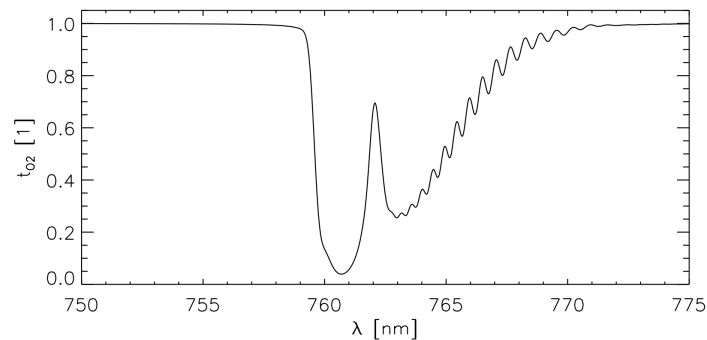


Figure 1: Simulated spectral transmission in the NIR spectral range for a light path from the TOA to the surface and finally to a nadir viewing satellite. Here we neglect any kind of atmospheric scattering and assume a solar zenith angle of 60 degree and an instrument viewing zenith angle of 0 degree. The spectrum is convolved with a Gaussian instrument spectral response with a resolution of 0.4 nm.

strong absorption lines (O_2 in the NIR band, CH_4 and H_2O in the 2.3 μm band) of which the depth and shape is modified by aerosol scattering. The basis of the algorithm is to fit a forward model, which is capable to handle light scattering by molecules and particles in the atmosphere. An important aspect of an operational physics based algorithm is computational speed and to serve this overall need, we developed a highly efficient radiative transfer model using a k -binning approach [RD29] that avoids time consuming line-by-line calculations.

4.2 Heritage

The algorithm for retrieval of methane columns from the S5 measurements is based on earlier developments of a CO_2 and CH_4 retrieval algorithm from GOSAT and S5P, called RemoTeC [RD20, RD9, RD23, RD16, RD30, RD31, RD32]. It makes use of measurements at the O_2 A band around 765 nm, at the CH_4 and CO_2 bands at 1.6 μm and around the strong carbon dioxide band at 2.0 μm in case of GOSAT measurements and at the O_2 A band, and at the 2.3 μm band in case of S5P. For CO_2 and CH_4 retrieval from GOSAT the algorithm has been thoroughly tested for simulations where the actual scattering properties are unknown [RD20, RD9], and has been successfully applied to real GOSAT measurements [RD23, RD16, RD30, RD33]. For the operational S5P methane column retrieval algorithm, we built on the RemoTeC algorithm for GOSAT and improved its numerical implementation and robustness. The S5P algorithm has been tested extensively for an ensemble of scenarios with realistic combinations of aerosol properties, cirrus properties, surface albedo, and solar zenith angle, similarly to what was done for GOSAT prior to the GOSAT launch [RD31, RD32].

4.3 Requirements

To improve our present knowledge on CH_4 on a global scale, the following requirements are formulated as a baseline for the Sentinel-5 prototype algorithm development [AD6]:

S5-L2-PRO-270 The precision in the total CH_4 column density shall be smaller than 10 ppbv or 0.5% (goal) to 18 ppbv or 1% (threshold). Depending on the scenario the least stringent of the absolute and the relative requirement applies.

S5-L2-PRO-280 The regional bias in the total CH_4 column density shall be smaller than 10 ppbv or 0.5% (goal) to 18 ppbv or 1% (threshold). Depending on the scenario the least stringent of the absolute and the relative requirements applies.

S5-L2-PRO-290 The stability in the CH_4 total column density shall be smaller than 4 ppbv or 0.25% per decade. Depending on the scenario the least stringent of the absolute and the relative requirements applies.

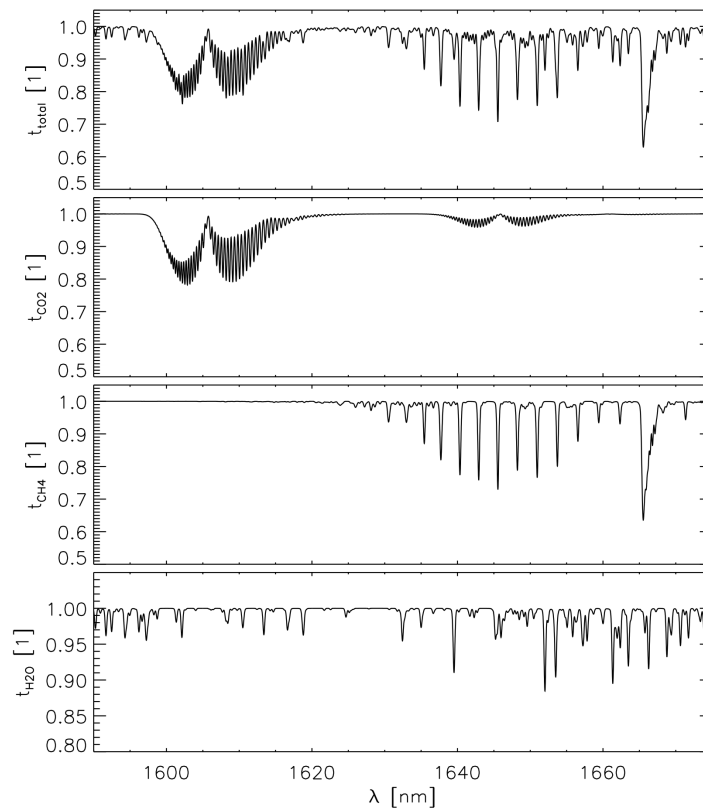


Figure 2: Same as Fig. 1 but for the SWIR-1 spectral window: total transmission (upper panel), contribution by CO_2 (second panel), contribution by CH_4 (third panel), and the water vapor contribution (fourth panel). The spectrum is convolved with a Gaussian instrument spectral response with a resolution of 0.25 nm. Note the different scales of the y-axes.

Here, requirement S5-L2-PRO-290 concerns the stability of the CH_4 product, which in turn is determined by the stability of the instrument in space rather than the algorithm. Therefore, we do not consider the requirement as a criterion to investigate the algorithm performance. The threshold requirements concerning precision and bias are very similar to the corresponding requirements of the S5P mission [AD7]. Based on the RemoTeC S5P heritage and the error analysis in Sect. 7, we conclude that the RemoTeC algorithm is compliant with the requirement S5-L2-PRO-270 and S5-L2-PRO-280 for solar zenith angles not exceeding 70° and observation zenith angle not exceeding 60° .

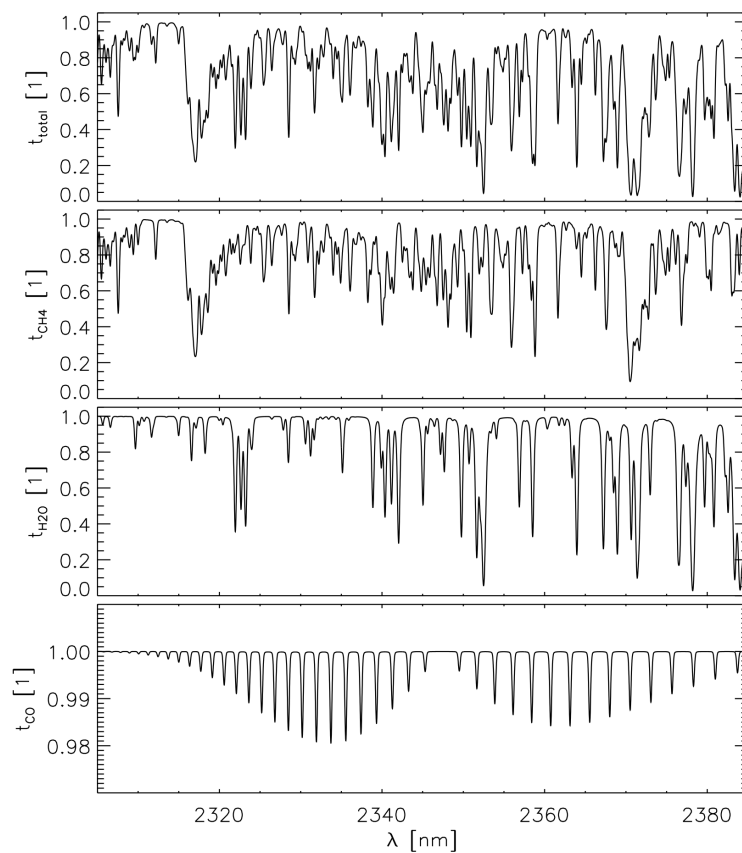



Figure 3: Same as Fig. 1 but for the SWIR-3 spectral window: total transmission (upper panel), contribution by CH₄ (second panel), contribution by H₂O (third panel), and the CO contribution (fourth panel). Note the different scales of the y-axes.

 	<p>S5L2PP Methane ATBD</p>	<p>Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 20/91</p>
--	--------------------------------	---

5 Instrument Overview

A description of the Sentinel 5 instrument and performance can be found in [RD34].



 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Date : 17 May 2019	Page 21/91
--	------------------------	---	---------------

Table 1: Spectral ranges of the NIR-2a, NIR-2, SWIR-1 and SWIR-3 band [AD2, AD3, RD34].

band	spectral range (nm)	spectral resolution (nm)	spectral sampling (nm)	sampling ratio
NIR-2a (band 3b)	745–755	0.4	0.1333	3
NIR-2 (band 3c)	755–773	0.4	0.1333	3
SWIR-1 (band 4)	1590–1675	0.25	0.1	3
SWIR-3 (band 5)	2305–2385	0.25	0.1	3

6 Detailed Algorithm Description

The RemoTeC algorithm comprises two different methods, the physics-based and the proxy approach, to infer methane total columns from SWIR measurements. The methods are complementary with respect to accuracy and data coverage. The overall architecture of the S5 RemoTeC algorithm is depicted in Fig. 4, which is used in the following subsections to explain the theoretical baseline. Details on the algorithm setup are provided later in Section 6.11.

After collecting and preparing the input data, an initial data filtering screens soundings to be discarded due to high and optically thick clouds or inferior quality and signal-to-noise ratio of the S5-L1B measurements. For data passing the initial filter, non-scattering retrievals of O₂, CH₄, CO₂, and H₂O vapor from specific spectral bands are used for both refined cloud filtering and input to the proxy retrieval module. Rigorous cloud filtering is needed as a pre-requisite for the physics-based retrieval and both the proxy and the physics-based retrieval products are provided as the S5 level-2 product. The input and output layer (green boxes in Fig. 4) interfaces the CH₄ retrieval algorithm to general S5 processing infrastructure.

The key differences between the CH₄ algorithm for S5 and its precursor S5P are the following. First, the S5 CH₄ algorithm includes the proxy retrieval, which S5P could not perform because they lack the SWIR-1 band. Second, auxiliary retrievals performed for backup cloud filtering as well as the retrieval of solar induced fluorescence are included in the CH₄ algorithm where for S5P, these results were imported from other S5P algorithms.

In the following, we describe the theoretical baseline of the different algorithm elements in more detail and conclude the section with more details on the algorithm architecture.

6.1 Input

All input data are shown in the green boxes in the top of Fig. 4. We distinguish between static and dynamic input, where the static input (e.g. retrieval input settings, cross section lookup tables and lookup tables for aerosol optical properties, calibration key data) is required to initialize the methane processor and needs to be provided only once when the processing starts. The dynamic input consists e.g. of the NIR-2, SWIR-1 and SWIR-3 level 1b radiance data, which are spatially resampled to the SWIR-1 ground pixel (spectral ranges of these bands ¹ are summarized in Tab. 1). The spatial resampling is not part of the methane algorithm but is assumed to be performed in advance of the methane processing. Moreover, the input comprises S5 METimage radiance and cloud data [RD35] and the atmospheric input generated from (static) DEM (GMTED2010, [RD36]) and dynamic CTM (ECMWF, CAMS) data. All dynamic input is assumed to be co-located for each S5 ground pixel by the processor frame accordingly.

6.1.1 Absorption cross sections

The absorption cross-sections of CH₄, CO₂, CO, H₂O, and O₂ are pre-calculated from the latest spectroscopic databases [RD37, RD38] assuming Voigt line shapes and stored in a lookup-table as a function of pressure, temperature, and wavenumber. For water vapor, the updated spectroscopic line list of Scheepmaker et al., 2013, [RD39] is employed. The absorption cross sections of O₂ and CO₂ (NIR-2 and SWIR-1) are calculated according to Tran et al., 2006 [RD40] taking into account line mixing and collision induced absorption. As H₂O concentrations are

¹ The band nomenclature is adapted from [AD2, AD3], where the SWIR-2 band covers the 2.0 μm range, which was discarded later in the project.

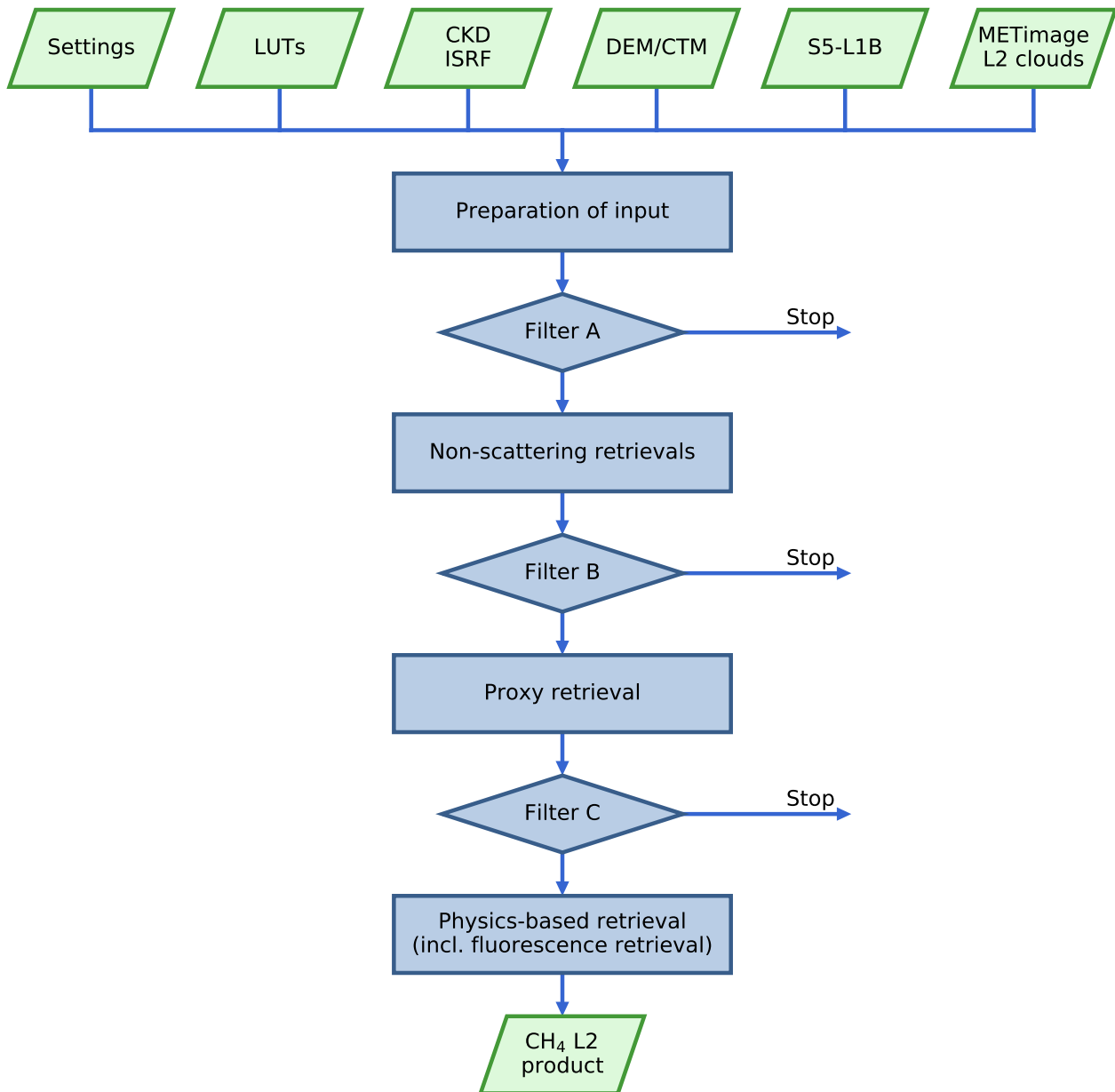


Figure 4: Overall design of the RemoTeC algorithm including the CH₄ proxy and physics-based retrieval branches. The input data include algorithm settings, lookup tables (LUTs) of absorption and scattering optical properties, the instrument spectral response function (ISRF) as part of the calibration key data set (CKD), atmospheric input generated from a digital elevation model (DEM) and Chemistry Transport Modeling (CTM), NIR-2, SWIR-1, and SWIR-3 S5 level 1B data (S5-L1B), and METimage L1B and cloud L2 products. The green boxes indicate the input and output layers interfacing the CH₄ algorithm to the general S5 processing infrastructure. The blue boxes are the core parts of the RemoTeC CH₄ retrieval algorithm.



 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Date : 17 May 2019	Page 23/91
--	--------------------------------------	---	---------------

Table 2: The S5 L1B and L2 Met data product used for cloud filtering, for more detail see Tab. 9.

METimage parameter	symbol	averaging approach	data product
FOV averaged radiance (band 4, 9)	$I_{b4,b9}^{imag}$	FOV	S5 L2 HET
Standard deviation of $I_{b4,b9}^{imag}$	$\sigma_{b4,b9}^{imag}$	FOV	S5 L2 HET
Number of METimage pixels confidently cloudy / probably cloudy / probably clear / confidently clear	$N_{c-cld}^{imag}, N_{p-cld}^{imag}, N_{c-clr}^{imag}, N_{p-clr}^{imag}$	FOV	S5 L2 CLA
METimage mean ice cloud optical thickness (at 550 nm)	τ_{ice}	SRF	S5 L2 CLA
METimage ice cloud fraction	f_{ice}	SRF	S5 L2 CLA
METimage ice cloud top pressure	p_{ice}	SRF	S5 L2 CLA
METimage ice cloud effective radius	r_{ice}	SRF	S5 L2 CLA
METimage liquid cloud fraction	f_{liq}	SRF	S5 L2 CLA
METimage liquid cloud top pressure	p_{liq}	SRF	S5 L2 CLA
METimage liquid cloud effective radius	r_{liq}	SRF	S5 L2 CLA
METimage liquid cloud optical depth (at 550 nm)	τ_{liq}	SRF	S5 L2 CLA
METimage coverage of valid data	N_{valid}	SRF	S5 L2 CLA

relatively high, self-broadening effects are taken into account by using an effective pressure as outlined in [RD41]. This is done the same as for GOSAT, OCO-2 and S5P.

6.1.2 METimage

Where S5P used the Visible Infrared Imaging Radiometer Suite (VIIRS), S5 uses METimage data for cloud filtering. The METimage data is explained in more detail in [RD35]. The METimage observations are either weighted averages using the instrument spatial response function (SRF) or averaged over four different fields of views (FOV) with a quadrilateral projection on ground. Here, the smallest field of view indicates the inner field of view (IFOV) and the remaining FOV averages refer to the outer field of view (OFOV). The METimage input is summarized in Tab. 2 and allows us to derive the effective fraction of 'confidently' clear and cloudy observations,

$$f_{cld}^{imag} = \frac{N_{c-cld}^{imag}}{N_{c-cld}^{imag} + N_{p-cld}^{imag} + N_{p-clr}^{imag} + N_{c-clr}^{imag}} \quad (1)$$

$$f_{clr}^{imag} = \frac{N_{c-clr}^{imag}}{N_{c-cld}^{imag} + N_{p-cld}^{imag} + N_{p-clr}^{imag} + N_{c-clr}^{imag}} \quad (2)$$

Moreover, we define two parameters to characterize the scene heterogeneity using FOV averages of the METimage radiances at 752 nm (b4) and 1375 nm (b9),

$$\eta_{b4}^{imag} = \frac{\sigma_{b4}^{imag}}{I_{b4}^{imag}} \quad (3)$$

and a corresponding expression for η_{b9}^{imag} .

6.2 Initial Data filtering A

Overall, we assume that only radiance and irradiance data are provided, which passes at the calling framework level data quality control. Appendix C gives an overview on the use of L1b flagging information for data control prior to the RemoTeC retrieval. Data with any warning are processed in the current baseline setting but later during the S5



 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001
		Version : 3.1 Date : 17 May 2019

Table 3: Spectral windows of the non-scattering retrievals and corresponding target species.

number	spectral window [nm]	target species
1	755–773	O ₂
2	1593–1621	CO ₂
3	1629–1654	CH ₄ /H ₂ O
4	2363–2373	CH ₄
5	2367–2377	H ₂ O

commission phase one may consider to use warning-flags of the L1B data product to potentially filter the the input data. Moreover, as shown in Fig. 4, the algorithm includes three separate filtering steps, indicated by A, B and C. The initial filtering step (A) will discard heavily cloud-contaminated scenes and soundings with low quality.

For the initial screening A of the S5 observations, we propose the following filter:

- A1** Measurements with too low signal can easily be filtered out considering the spectral maximum of a S5 radiance measurement in the SWIR-3 spectral range.

$$\max[I_{\text{SWIR-3}}] > T_{A1} \quad (4)$$

Alternatively, the spectral maximum of the SWIR-3 signal-to-noise values can be used.

- A2** For low sun, forward model errors increases substantially. Therefore, we propose to consider only measurements with sufficiently small solar zenith angle (SZA).

$$\text{SZA} < T_{A2} \quad (5)$$

- A3** Scenes, which are detected by METimage as confidently cloudy, will not be processed.

$$f_{\text{clid}}^{\text{imag}} < T_{A3} \quad (6)$$

- A4** Alternatively, one may filter on the occurrence of high, optically thick clouds, using the METimage cloud parameter.

$$\frac{\tau_{\text{liq},f_{\text{liq}}}}{p_{\text{liq}}} < T_{A4} \quad (7)$$

The threshold values T_{A1} – T_{A4} will be determined first during the commissioning phase using real data.

6.3 Non-scattering retrievals


The proxy retrieval and the cloud screening of S5 data requires non-scattering retrievals of O₂, CO₂, CH₄, and H₂O total columns from the five sub-windows as defined in Tab. 3. Per band an independent retrieval is preformed, providing in total six trace columns, two for CH₄ and H₂O and one for O₂ and CO₂, respectively.

6.3.1 The model atmosphere

At start-up of any retrieval, pressure, temperature, and the sub-columns of the gases are taken from CAMS forecast fields and are mapped by linear interpolation onto the K_{atm} -layer (72 in the baseline) vertical grid. Here, first the surface pressure p_{surf} is obtained by interpolating or extrapolating the input pressure profile as function of height to the actual mean surface height of the ground pixel (input from DEM). Next, the pressure values at the layer boundaries are calculated, with the pressure $p_{\text{lev},k}$ at the lower boundary of layer k (counting from top to bottom) is given by:

$$p_{\text{lev},k} = p_{\text{min}} + \Delta p \cdot k \quad (8)$$

$$\Delta p = (p_{\text{surf}} - p_{\text{min}}) / K_{\text{atm}} \quad (9)$$

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 25/91
---	------------------------	--

where p_{\min} is the pressure value of the upper boundary of the input atmosphere. The atmospheric profiles of temperature and of the gaseous constituents are constructed on this pressure grid via interpolation. For example, the methane sub-column $c_{\text{CH}_4,k}$ for the layer bounded by pressure levels $p_{\text{lev},k-1}$ and $p_{\text{lev},k}$ is given by:

$$c_{\text{CH}_4,k} = r_{\text{CH}_4,k} c_{\text{air},k} \quad (10)$$

where $r_{\text{CH}_4,k}$ is the methane dry air mixing ratio linearly interpolated from the input pressure grid to the pressure at the 'middle' of layer k defined by $(p_{\text{lev},k} + p_{\text{lev},k-1})/2$. The sub-column of dry air in layer k is denoted with $c_{\text{air},k}$.

$$c_{\text{air},k} = \frac{(p_{\text{lev},k+1} - p_{\text{lev},k}) N_A}{M g_k \left(1 + \frac{r_{\text{H}_2\text{O},k}}{\rho_{\text{rel}}}\right)}. \quad (11)$$

Here N_A is Avogadro's number, M is the molecular mass of air, g_k is the gravity constant in altitude layer k , and $\rho_{\text{rel}} = 1.60855$ is the mass of air relative to the mass of water [RD42]. The sub-columns of CO_2 , CO and H_2O are computed in the same manner. The O_2 sub-columns are obtained by multiplying the air sub-columns with the constant O_2 mixing ratio, which is 0.2095.

6.3.2 Non-scattering forward model

The line-by-line non-scattering forward model calculates the backscattered top-of-the-atmosphere radiance I as a function of wavelength λ . Here, the spectral sampling is fine enough to resolve all relevant atmospheric absorption lines. This means that the line-by-line spectral sampling must be at least 0.1 cm^{-1} ($\sim 5.8 \text{ pm}$) in the NIR-2 band and 0.02 cm^{-1} in both the SWIR bands ($\sim 5.3 \text{ pm}$ and 11 pm in SWIR-1 and SWIR-3 respectively). Further, the model calculates the derivatives of the radiance with respect to the retrieval parameters, which is necessary for the inversion.

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

$$I(\lambda) = F_0(\lambda) A(\lambda) \frac{\mu_0}{\pi} e^{-\tau_{\text{abs}}(\lambda)/\tilde{\mu}} \quad (12)$$

with the air mass factor

$$\frac{1}{\tilde{\mu}} = \frac{1}{\mu_0} + \frac{1}{\mu_v}, \quad (13)$$

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 τ_{abs} is the absorption optical thickness due to molecular absorption. The spectral dependence of the Lambertian surface albedo A is parameterized through an n -th order spectral polynomial,

$$A(\lambda) = \sum_{i=0}^n a_i (\lambda - \lambda_0)^i \quad (14)$$

with an arbitrary reference wavelength λ_0 . This λ_0 is chosen inside the window with the sole purpose of making parameter a_0 a representative value for the albedo in that window. The absorption optical depth is first calculated per absorber (j) per discretized atmospheric layer (k) with


$$\tau_{\text{abs},k,j}(\lambda) = \sigma_j(p_k, T_k, \lambda) c_{k,j} \quad (15)$$

with the absorption cross-section $\sigma_j(p_k, T_k, \lambda)$ of molecule j at wavelength λ , pressure p_k and temperature T_k at the center of atmospheric layer k . And $c_{k,j}$ is the sub-column of absorber j in layer k .

The cross-section for pressure p_k , temperature T_k and wavelength λ is obtained by linear interpolation from the tabulated values. The total vertically integrated molecular absorption optical thickness τ_{abs} is

$$\tau_{\text{abs}} = \sum_j \sum_{k=1}^{K_{\text{atm}}} \tau_{\text{abs},k,j} \quad (16)$$

The line-by-line solar spectrum (F_0) can be obtained in two ways. First one can adopt F_0 from an a priori model spectrum and second F_0 can be deduced from S5 solar irradiance measurements. For the latter, the corresponding

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 26/91
---	------------------------	--

algorithm is described in Sect. 6.9. The decision on which approach will be used for the operational data processing will be taken during the commissioning phase of the S5 instrument.

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

$$\frac{\partial I}{\partial \tau_{\text{abs}}} = -\frac{I}{\tilde{\mu}} \quad (17)$$

$$\frac{\partial I}{\partial A} = \frac{I}{A} \quad (18)$$

$$\frac{\partial I}{\partial F_0} = \frac{I}{F_0} \quad (19)$$

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

$$\frac{\partial I}{\partial c_{k,j}} = -\sigma_j(p_k, T_k, \lambda) \frac{I}{\tilde{\mu}} \quad (20)$$

$$\frac{\partial I}{\partial a_i} = (\lambda - \lambda_0)^i \frac{I}{A(\lambda)} \quad (21)$$

where a_i is the i -th albedo coefficient in Eq. (14) and $c_{k,j}$ are the subcolumns of the absorbers from Eq. (15). RemoTeC allows for fitting a spectral recalibration of the solar spectrum ($\delta\lambda_s$) using the derivative that comes from a straight forward chain rule from Eq. (19).

$$\frac{\partial I}{\partial \delta\lambda_s} = \frac{I}{F_0} \frac{\partial F_0}{\partial \lambda} \quad (22)$$

Given the line-by-line radiance $I(\lambda)$ and its derivatives, sensor spectral resolution is taken into account by convolution with the ISRF s_{earth} and so the forward model reads

$$\mathcal{F}(\lambda_i) = \int d\lambda s_{\text{earth}}(\lambda_i, \lambda) I(\lambda), \quad (23)$$

where λ_i is the wavelength assigned to spectral measurement i . When accounting for the line-by-line spectral sampling of the radiance spectrum, Eq. (23) rewrites

$$\mathcal{F} = \mathbf{S}_{\text{earth}} \mathbf{I} \quad (24)$$

where \mathbf{I} is the line-by-line radiance vector, \mathcal{F} is forward model vector to be used by the inversion containing the modeled radiances for all spectral sensor pixels under consideration, and matrix \mathbf{S} compromises the ISRF for all relevant spectral measurements. (Note the different units of the ISRF s_{earth} with inverse wavelength units, and the corresponding matrix \mathbf{S} , which is unitless, because it includes the differential $d\lambda$.) An analogue convolution is carried out for the derivatives in equations (20), (21) and (22) to calculate the corresponding derivatives of \mathcal{F} with respect to absorber sub-columns, surface albedo and the spectral shift of the solar spectrum.

It is also possible to calculate derivatives with respect to instrument calibration parameters, which do not affect \mathbf{I} , but $\mathbf{S}_{\text{earth}}$. This way RemoTeC allows fitting a spectral shift in the measurement of the Earth-shine radiance. The derivative with respect to this spectral shift is

$$\frac{\partial \mathcal{F}}{\partial \delta\lambda_e} = \frac{\partial \mathbf{S}_{\text{earth}}}{\partial \lambda_i} \mathbf{I} \quad (25)$$

6.3.3 Non-scattering inversion

For each of the retrievals listed in Tab. 3, the non-scattering inversion is performed. To discuss the inversion, we first define the state vector \mathbf{x} , the measurement vector \mathbf{y} , which includes the S5-L1B radiances at all spectral points inside the relevant window, and the ancillary vector \mathbf{b} containing parameters that are not retrieved.

The non-scattering state vector \mathbf{x} covers the elements listed in table 4. It constrains, in particular, the (sub-)columns of the relevant molecular absorbers. The ancillary parameter vector \mathbf{b} comprises all non-retrieved parameters required for evaluating the forward model. The most important ones are:



 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Date : 17 May 2019	Page 27/91

Table 4: State vector elements and their a priori and first guess values for the non-scattering retrieval.

Number of elements	State vector element	A priori and first guess value
K_{ret} or $2K_{\text{ret}}$	$K_{\text{ret}} = 12$ sub-columns for per target trace gas (see Tab. 3)	CTM/DEM
0, 1 or 2	Total columns of other trace gases that absorb in the window	CTM/DEM
3	Albedo polynomial parameters	Lambertian equivalent value (LER) for a_0 , 0 for a_1 and a_2
1	Spectral shift $\delta\lambda_e$ of earth radiances	zero
1	Spectral shift $\delta\lambda_s$ of solar irradiances	zero

- Temperature vertical profile at K_{atm} layers.
- Pressure vertical profile at K_{atm} layers.
- Molecular absorption cross sections.
- Solar irradiance spectrum
- Instrument spectral response function (ISRF).
- Observation geometry (μ_0 and μ_v).

The forward model as explained in Sect. 6.3.2 defines the relationship:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \mathbf{b}) + \mathbf{e}_y. \quad (26)$$

Given the state vector \mathbf{x} , the measurements \mathbf{y} , and the forward model \mathcal{F} , the inversion procedure aims at estimating the state vector by inverting equation (26). To this end, the forward model is linearized around a state vector \mathbf{x}_n according to

$$\mathbf{y} = \mathcal{F}(\mathbf{x}_n, \mathbf{b}) + \mathbf{K}(\mathbf{x} - \mathbf{x}_n) + \mathbf{e}_y. \quad (27)$$



The forward model calculates the Jacobian for the sub-columns in all discretized layers, which are $K_{\text{atm}} = 72$ layers in the baseline, but the state vector only contains $K_{\text{ret}} = 12$ sub-columns for the target absorbers and only one total column for the non-target absorbers. This reduction in vertical resolution is performed by (sub-)profile scaling, namely

$$K_{jk'} = \sum_{k \in k'} K_{jk} \frac{c_{jk}^{\text{ref}}}{C_{jk'}^{\text{ref}}} \quad (28)$$

$$c_{jk} = \mathbf{x}_{jk'} \frac{c_{jk}^{\text{ref}}}{C_{jk'}^{\text{ref}}} \quad (29)$$

where k' refers to a retrieval layer and k is an atmospheric layer in the forward model. c_{jk}^{ref} is the reference (a-priori) sub-column of absorber j in atmospheric layer k and $C_{jk'}^{\text{ref}}$ is the total sub-column of absorber j within retrieval layer k' , thus $C_{jk'}^{\text{ref}} = \sum_{k \in k'} c_{jk}^{\text{ref}}$. These equations reflect that the relative profiles of any absorber within one retrieval layer are kept constant.

The state vector is found by inverting Eq. (27), where the inversion method is based on a Tikhonov regularization scheme. Regularization is necessary because the inverse problem is ill-posed. This means that the measurements

 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 28/91
--	------------------------	--

contain insufficient information to retrieve all state vector elements independently, especially the K_{ret} sub-columns of a target absorber. Without Tikhonov regularization, these sub-columns will be overwhelmed by noise contributions. The regularization scheme constrains the state vector to overcome this, while keeping enough freedom to retrieve valuable information from the measurement. To find the right balance between avoiding noise and keeping valuable information, the total columns are left unconstrained, but the relative profiles are constrained to some extent to allow limited deviations from the a-priori relative profiles. The regularization strength can be fine-tuned during the commissioning phase of S5.

With Tikhonov regularization, the following minimization problem is solved,

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \left(\|\mathbf{S}_y^{-1/2} (\mathcal{F}(\mathbf{x}) - \mathbf{y})\|^2 + \|\mathbf{W}(\mathbf{x} - \mathbf{x}_a)\|^2 \right) \quad (30)$$

with the measurement error covariance \mathbf{S}_y and the regularization matrix \mathbf{W} . The regularization matrix \mathbf{W} maps the state vector to a set of values that will be constrained. In this case, we constrain deviations in the relative profile of the target absorber. This is done by setting \mathbf{W} such that it produces first derivatives of the relative abundance with respect to the reference profile, namely

$$\frac{\mathbf{x}_{jk'}}{C_{jk'}^{\text{ref}}} \quad (31)$$

To this end, we use the discretized first derivative matrix \mathbf{L}_1 , which produces $K_{\text{ret}} - 1$ first derivatives out of a K_{ret} -element vector.

$$\mathbf{L}_1 = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & \dots \\ 0 & 0 & 1 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (32)$$

Applying this matrix to the relative profile deviations defined in Eq. (31) and applying an arbitrary scaling factor γ_j for absorber j to control the regularization strength, results in the following regularization matrix.

$$\mathbf{W} = \begin{bmatrix} \frac{\gamma_j}{C_{j1}^{\text{ref}}} & -\frac{\gamma_j}{C_{j2}^{\text{ref}}} & 0 & 0 & \dots & 0 & 0 & \dots \\ 0 & \frac{\gamma_j}{C_{j2}^{\text{ref}}} & -\frac{\gamma_j}{C_{j3}^{\text{ref}}} & 0 & \dots & 0 & 0 & \dots \\ 0 & 0 & \frac{\gamma_j}{C_{j3}^{\text{ref}}} & -\frac{\gamma_j}{C_{j4}^{\text{ref}}} & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{bmatrix} \quad (33)$$

This matrix has the dimension $K_{\text{ret}} - 1 \times \dim(\mathbf{x})$. The zeros on the right indicate that other state parameters than the target absorber profile are not regularized. For non-scattering number 3 (Tab. 3), there are two target absorbers and therefore matrix \mathbf{W} has $2(K_{\text{ret}} - 1)$ rows, ($K_{\text{ret}} - 1$ first derivatives for each of the two target absorbers), which can be scaled by two independent scaling factors γ_j .

The inverse problem is solved iteratively, in which after each iteration, the forward model is re-linearized around the solution of the previous iteration. So for each iteration step, the minimization solution in Eq. (30) is

$$\mathbf{x}_{n+1} = \mathbf{G}(\mathbf{y} - \mathcal{F}(\mathbf{x}_n)) + \mathbf{A}\mathbf{x}_n + (\mathbf{I} - \mathbf{A})\mathbf{x}_a \quad (34)$$

with the gain matrix


$$\mathbf{G} = (\mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{W}^T \mathbf{W})^{-1} \mathbf{K}^T \mathbf{S}_y^{-1} \quad (35)$$

and averaging kernel matrix

$$\mathbf{A} = \mathbf{G}\mathbf{K} \quad (36)$$

To improve the convergence of this non-linear inversion, we use a Gauss-Newton scheme (see e.g. [RD43]) with reduced step-size by introducing a filter factor that limits the update per iteration step of the state vector. The updated state vector is then given by:

$$\mathbf{x}_{n+1} = \frac{1}{1 + \Lambda} \mathbf{x}_{\text{des}} + \frac{\Lambda}{1 + \Lambda} \mathbf{x}_n \quad (37)$$

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 29/91
---	------------------------	--

where \mathbf{x}_{des} is the desired \mathbf{x}_{n+1} without filter factor from Eq. 34. If Λ is large the update of the state vector is small. If $\Lambda = 0$, the equation for the updated state vector is equivalent to the pure Tikhonov equation. The iteration is started with a large Λ , typically on the order of 10. It is then reduced or increased in the following iteration steps according to an empirically found scheme similar to Levenberg-Marquardt strategies [RD43]. We accept solution \mathbf{x}_{n+1} and decrease Λ by a factor of 2.0 if the least squares norm ($\|\mathbf{S}_y^{-1/2}(\mathcal{F}(\mathbf{x}) - \mathbf{y})\|^2$) of iteration $n + 1$ is smaller than 1.1 times the least squares norm of iteration n . Otherwise, we discard the solution of iteration $n + 1$, increase Λ by a factor of 2.5 and solve again for \mathbf{x}_{n+1} . If Λ is smaller than a threshold value of 0.05, it is set to zero and the iteration is continued without a reduction in step size, assuming that the current state vector is sufficiently close to the true solution to finally converge.

Convergence is triggered when the state vector changes by an amount smaller than a preset threshold and the step-reducing parameter Λ has reached zero. The iterative process is terminated if state parameter elements reach unphysical values, if Λ grows too large by rejected steps or if convergence is not achieved within a given number of iterations. Results of non-successful inversions are discarded.

6.4 Proxy cloud filter B

The proxy CH_4 retrieval requires only a weak cloud filter (data filter B in Fig. 4). Therefore, we filter out all soundings for which the non-scattering retrieval delivers O_2 total column densities that largely deviate from the a priori estimate (based on CTM and DEM input data), i.e.

B1 We only process spectra if

$$\left| \frac{C_{\text{O}_2}}{C_{\text{O}_2}^{\text{apr}}} - 1 \right| < T_{\text{B1}} \quad (38)$$

where T_{B1} is an empirical threshold on the order of 0.1. $C_{\text{O}_2}^{\text{apr}}$ is the CAMS a priori O_2 total column density derived from the corresponding CAMS surface pressure, which is corrected for the pixel surface elevation using the pixel height from a digital elevation map. The value for the threshold T_{B1} will be determined first during the commissioning phase and further optimized during the first year of operation (for re-processing purposes).

6.5 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 ($\text{XCH}_4^{\text{proxy}}$) according to

$$\text{XCH}_4^{\text{proxy}} = \frac{C_{\text{CH}_4}}{C_{\text{CO}_2}} \times \text{XCO}_2^{\text{apr}}, \quad (39)$$

where C_{CH_4} and C_{CO_2} are retrieved under the non-scattering assumption from the SWIR-1 sub-windows indicated in Table 3. $\text{XCO}_2^{\text{apr}}$ is the total dry air mixing ratio of carbon dioxide coming from the CAMS forecast. As outlined before, the first key assumption of the proxy approach is that scattering effects cancel in the $\frac{C_{\text{CH}_4}}{C_{\text{CO}_2}}$ 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 $\text{XCO}_2^{\text{apr}}$ is sufficiently close to the true XCO_2 total column mixing ratio such that the respective error contribution to $\text{XCH}_4^{\text{proxy}}$ does not jeopardize the accuracy goal.

The precision of the proxy XCH_4 product is defined as the instrument noise propagation on the proxy XCH_4 , 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 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:

$$\mathcal{E}_{\text{XCH}_4}^{\text{proxy}} = \text{XCH}_4^{\text{proxy}} \sqrt{\left(\frac{\epsilon_{\text{CH}_4}^{\text{ns}}}{\text{CH}_4^{\text{ns}}}\right)^2 + \left(\frac{\epsilon_{\text{CO}_2}^{\text{ns}}}{\text{CO}_2^{\text{ns}}}\right)^2} \quad (40)$$

This precision term only contains instrument noise propagation. Estimated errors on XCO_2 forecast from CAMS, and biases because of imperfect cancellation of scattering effects are not included in this quantity.

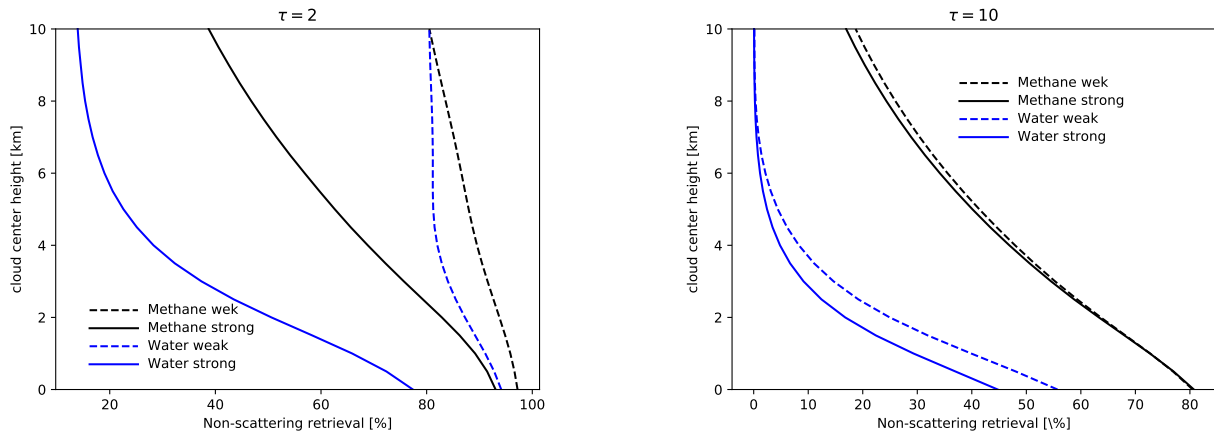


Figure 5: Non-scattering retrieval results of CH₄ and H₂O in the weak and strong absorption bands as percentage of the truth with a varying cloud height. Left panel: Thin cloud ($\tau = 2$). Right panel: Thick cloud ($\tau = 10$).


6.6 Cloud filter C of the physics-based retrieval

The full-physics methane retrievals can only be performed for cloud free ground pixels and thus requires a strict data filtering (data filter C in Fig. 4). The algorithm baseline is based on the METImage cloud mask defined in Eqs. (1) and (2) for the inner and outer FOVs. For GOSAT, where cloud information is provided by the Cloud and Aerosol Imager (CAI), we found that an appropriate criterion is that 99% of an area equal 4 times that of a ground pixel should be confidently clear. Additionally, we use the METImage ice cloud and band 9 radiance product to screen scenes with high cirrus contamination and with scene heterogeneity.

In case METImage data are not available, we base the data filtering on the S5 non-scattering retrieval of O₂, CH₄ and H₂O total column as discussed in Sec. 6.3. Here the difference between CH₄ and H₂O columns from weak and strong absorption bands indicate the presence of clouds in the observed scene. A similar cloud screening method was developed by the ACOS/OCO-2 team for cloud screening for GOSAT and OCO-2 CO₂ retrievals [RD44] and is used as backup cloud filter for operational CH₄ retrievals from S5P measurements. Moreover, we employ the proxy cloud filter, which is based on the O₂ column from S5 NIR-2 measurements as defined in Eq. (38) with a threshold value defined specifically for the physics-based retrieval. The most important limitation of this cloud filter is that for certain values of the surface albedo light path effects in the O₂ A band cancel out while clouds are present [RD45]. Furthermore, data filtering requires also to account for atmospheric scenes adjacent to the ground scene under consideration. Using S5 cloud observations only, this means that at least the entire adjacent ground pixels, in total nine pixels, should be considered. On the contrary, the METImage cloud filtering introduces much more flexibility in defining the adjacent scenes. However, due to the design of RemoTeC-S5 as a single pixel processor such data filtering is only possible a posteriori and so not considered as part of this operational processing unit. Due to the limitations in cloud flagging from S5 itself, it is strongly recommended to use S5 L2 Met data for cloud and cirrus filtering and to consider cloud filtering from S5 measurements as a backup approach.

The weak and strong band cloud filter is based on the fact that the bias of a non-scattering retrieval caused by ignoring scattering effects depends on the strength of the absorption lines. By comparing the retrievals in the weak and strong bands, a filter can be designed that does not depend on accurate a-priori knowledge of CH₄ or H₂O. This principle is also explained in [RD46], where the similar principle is done for S5P. It relies on the fact that with a scattering atmosphere, a combination of different light paths are observed. In strong absorption bands, the contribution of the longer light paths is very small, so the shorter light path is dominant. This leads to a retrieval of less CH₄ or H₂O than for the weak absorption band, where also long light path contributions are relevant. An important consequence is that this principle works best for optically thin clouds. Optically thick clouds are more efficiently filtered by the non-scattering O₂ column. The two band cloud filter has been tested with a synthetic case with a thin ($\tau = 2$) and a thick ($\tau = 10$) cloud. Figure 5 shows that the two-band cloud filter detects the thin cloud very well.

To summarize, the following checks for cloudiness and/or inhomogeneity are performed within the baseline approach using METImage:

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 31/91
---	------------------------	--

C1 Fraction of 'confidently clear' METimage pixels

$$f_{\text{clear}}^{\text{Met}} > T_{C1} \quad (41)$$

Here, threshold T_{C1} contains values for the four scales of the FOV.

C2 Average METimage cirrus reflectance within the inner FOV

$$I_{b9}^{\text{Met}} < T_{C2} \quad (42)$$

C3 The difference in METimage average cirrus reflectance between the SWIR IFOV and OFOV

$$\Delta I_{b9}^{\text{Met}} < T_{C3} \quad (43)$$

Here, threshold T_{C3} contains values for the three scales of the outer FOV.

C4 Scene heterogeneity

$$\eta_{b4}^{\text{Met}} < T_{C4} \quad (44)$$

T_{C4} contains values for the four scales of the FOV.

As a fallback in case S5 L2 Met data are not available, we consider the following checks:

- The difference between the water column $C_{\text{H}_2\text{O,weak}}$ retrieved from the weak water band and the water column $C_{\text{H}_2\text{O,strong}}$ retrieved from the strong water band should be

$$\frac{|C_{\text{H}_2\text{O,weak}} - C_{\text{H}_2\text{O,strong}}|}{C_{\text{H}_2\text{O,weak}}} < T_{C5} . \quad (45)$$

- The difference between $C_{\text{CH}_4,\text{weak}}$ and $C_{\text{CH}_4,\text{strong}}$

$$\frac{|C_{\text{CH}_4,\text{weak}} - C_{\text{CH}_4,\text{strong}}|}{C_{\text{CH}_4,\text{weak}}} < T_{C6} . \quad (46)$$

- The difference between the non-scattering S5 and a-priori O_2 column


$$\left| \frac{C_{\text{O}_2}}{C_{\text{O}_2}^{\text{ap}}} - 1 \right| < T_{C7} . \quad (47)$$

where $C_{\text{O}_2}^{\text{ap}}$ is derived from the corresponding CAMS surface pressure and corrected for the pixel surface elevation.

Analogous to the previous data filter A and B, all threshold values will be determined in the commissioning phase and further optimized during the first year of operation.

6.7 Fluorescence retrieval

To derive a priori information on the solar induced fluorescence (SIF) signal for the physics-based methane retrieval, we infer fluorescent emission of near infrared radiation by photosynthesizing plants from Sentinel-5 measurements in the spectral window 745 – 755 nm. This spectral range is between the absorption bands of water vapor and oxygen does not include relevant telluric absorption but contains solar Fraunhofer lines, whose relative depths in the reflected Earth radiance depend on the degree of fluorescent emission by vegetation. [RD47] analyzed the SIF retrieval performance in this window and found a relatively high SIF sensitivity to the measurement noise and so concluded that a spectral fit window 735-758 would be best compromise between SIF precision and window width. This spectral range is not covered by Sentinel 5 with the adjacent NIR-1 band at 685-710 nm. However, keeping in mind that for our algorithm the SIF product is only needed as prior information to the CH_4 retrieval, which is further adjusted using the SIF sensitivity within the spectral range of the O_2 A band, we conclude that the spectral coverage of the NIR-2a band is sufficient for our purpose.

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 32/91
---	------------------------	--

For the fluorescence retrieval, we simulate Earth radiances $I(\lambda)$ with a simplified radiative transfer model assuming reflection by a Lambertian surface combined with fluorescent emission neglecting any scattering or absorption in the atmosphere.

$$I(\lambda) = F_0(\lambda) A(\lambda) \frac{\mu_0}{\pi} + F_s(\lambda) \quad (48)$$

with linear spectral dependent Lambertian albedo

$$A(\lambda) = a_0 + a_1 (\lambda - \lambda_0) , \quad (49)$$

around a reference wavelength λ_0 analogous to Eq. (14). Analogously, we approximate the spectral dependence of fluorescent emission F_s by a 3rd order polynomial

$$F_s(\lambda) = \sum_{j=0}^3 f_j (\lambda - \lambda_0)^j . \quad (50)$$

The retrieval is made numerically efficient by avoiding any ISRF convolutions after algorithm initialization. Here, ISRF convolution on the polynomial functions is ignored, because these polynomials have no spectral features beyond the sensor resolution. The solar irradiance spectrum is convolved with the ISRF at model initialization. To keep the ability of fitting a spectral shift, the solar spectrum is not only convolved on the baseline S5 measurement wavelengths λ_i , but also at shifted wavelengths.

$$F_0^*(\lambda_i, \delta\lambda_e) = \int d\lambda s_{\text{earth}}(\lambda_i, \lambda) F_0(\lambda + \delta\lambda_e) , \quad (51)$$

so that the convolved solar spectrum with any realistic spectral shift can be accurately obtained by interpolation of Eq. (51).

Finally, our forward model reads

$$\mathcal{F}_i(a_0, a_1, f_0, f_1, f_2, f_3, \delta\lambda_e) = \frac{\mu_0}{\pi} F_0^*(\lambda_i, \delta\lambda_e) (a_0 + a_1 (\lambda_i + \delta\lambda_e - \lambda_0)) + \sum_{j=0}^3 f_j (\lambda_i + \delta\lambda_e - \lambda_0)^j \quad (52)$$

where we have substituted $\lambda = \lambda_i + \delta\lambda_e$ in Eq 49 and 50 to account for the spectral shift $\delta\lambda_e$.

The Jacobian of \mathcal{F} with respect to the surface albedo and fluorescence parameter can be derived in a straight forward manner,

$$\frac{\partial \mathcal{F}_i}{\partial a_j} = \frac{\mu_0}{\pi} F_0^*(\lambda_i + \delta\lambda_e - \lambda_0)^j \quad (53)$$

$$\frac{\partial \mathcal{F}_i}{\partial f_j} = (\lambda_i + \delta\lambda_e - \lambda_0)^j \quad (54)$$

For the derivatives with respect to $\delta\lambda_e$, the derivative of F_0^* is obtained during interpolation of the pre-calculated values of $F_0^*(\lambda_i, \delta\lambda_e)$ and the change of λ with $\delta\lambda_e$ gives straight forward derivatives.

$$\frac{\partial \mathcal{F}_i}{\partial \delta\lambda_e} = \frac{\mu_0}{\pi} \frac{\partial F_0^*}{\partial \delta\lambda_e} (a_0 + a_1 (\lambda_i + \delta\lambda_e - \lambda_0)) + \frac{\mu_0}{\pi} F_0^* a_1 + \sum_{j=0}^3 j f_j (\lambda_i + \delta\lambda_e - \lambda_0)^{j-1} \quad (55)$$

The different sensitivity of the forward model with respect to the surface albedo, solar fluorescence parameters and $\delta\lambda_e$ allows us to infer these quantities from the Sentinel-5 measurements and so the state vector comprises

$$\mathbf{x} = (a_0, a_1, f_0, f_1, f_2, f_3, \delta\lambda_e) \quad (56)$$

This inversion does not require any regularization, so the minimization problem is

$$\hat{\mathbf{x}} = \min_x \|\mathbf{S}_y^{-1/2} [\mathbf{y} - \mathcal{F}(\mathbf{x})]\|^2 \quad (57)$$



 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Date : 17 May 2019	Page 33/91
--	------------------------	---	---------------

Table 5: Spectral fit windows of the physics-based retrieval.

band	spectral fit window (nm)
NIR-2 (band 3c)	755–773
SWIR-1 (band 4)	1629–1675
SWIR-3 (band 5)	2305–2385

where S_y denotes the error covariance of the measurement vector y . We apply a Gauss Newton iteration scheme and so per iteration n , the solution is given by

$$\hat{x} = x_{n-1} + (\mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K})^{-1} \mathbf{K}^T \mathbf{S}_y^{-1} (y - \mathcal{F}(x_{n-1})) \quad (58)$$

with the Jacobian $\mathbf{K}_{ij} = \partial \mathcal{F}_i / \partial x_j(x_{n-1})$. The iteration is terminated using a convergence criterion based on the fit residuals. Because the spectral shift $\delta \lambda_e$ is the only reason for non-linearity, convergence is not an issue for this retrieval.

To make the SIF product more easily interpretable, the retrieved polynomial parameters (f_0, f_1, f_2, f_3) are translated to fluorescent emissions at four different wavelengths in the window. For the full-physics retrieval, the fluorescent emissivity at 755 nm and its spectral derivative are taken over as a-priori values, because the retrieval window of the full-physics retrieval is on the long-wave side of the fluorescence retrieval window.

6.8 Physics-based retrieval

The physics-based retrieval approach infers simultaneously gas concentrations and scattering properties of the atmosphere in order to model adequately the light path through the earth's atmosphere. Therefore, RemoTeC aims at retrieving the CH₄ vertical profile (with slightly more than 1 degree of freedom) and 3 scattering parameters characterizing the particle amount, size and height using multiple spectral bands. The algorithm is fully flexible concerning the selected spectral measurement bands and thus allows both a two-band retrieval combining the NIR-2 and SWIR-3 band, which is the baseline of the Sentinel-5 Precursor processing, and a combination of the NIR-2, SWIR-1 and SWIR-3 band. The three-band retrieval with the spectral fit windows as indicated in Tab. 5 is a new feature of Sentinel-5 CH₄ retrievals and its performance is already studied by Butz et al., 2013 [RD48] on simulated measurements showing a clear advantage of this spectral combination for tropical dark scenes. This section summarizes the theoretical baseline of the RemoTeC physics-based approach starting with a description of the optical properties of the model atmosphere. Hereafter, we summarize the radiative transfer simulations and the RemoTeC inversion approach.

6.8.1 Optical properties of the model atmosphere

Starting with the calculation of the absorption optical depth, we adapt the calculation of the molecular absorption properties $\tau_{\text{abs,mol}}$ from section 6.3.2 (there shortly described as τ_{abs}). In order to save computational cost for the physics-based radiative transfer calculations, we perform the simulations on a reduced vertical with $K_{\text{RT}} = 36$ layers, which is deduced from the atmospheric layering with $K_{\text{atm}} = 72$ by merging each two adjacent layers into one aggregated layer. The calculation of molecular scattering properties starts out from the reduced vertical RT grid. The Rayleigh scattering optical thickness for layer k and wavelength λ is given by

$$\tau_{\text{sca,mol},k}(\lambda) = \sigma_{\text{ray}}(\lambda) c_{\text{air},k} \quad (59)$$

where σ_{ray} is the Rayleigh scattering cross section given by [RD49]

$$\sigma_{\text{ray}}(\lambda) = A \lambda^{-(4+X)} \quad (60)$$

$$X = B\lambda + \frac{C}{\lambda} - D \quad (61)$$

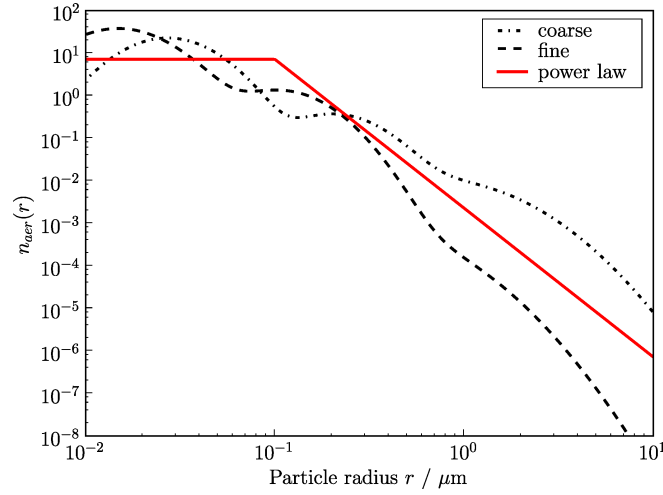


Figure 6: Particle size distribution $n_{\text{aer}}(r)$ as a function of particle radius r . The retrieval method relies on a power law (red solid) size distribution. Also shown are more realistic multi-modal lognormal size distributions for a fine mode (black dashed) and a coarse mode (black dotted) dominated aerosol type.

with $A = 4.02 \cdot 10^{-28} \text{ cm}^2$, $B = 0.389 \mu\text{m}^{-1}$, $C = 0.04926 \mu\text{m}$, and $D = 0.3228$. The Rayleigh scattering phase function is given by (e.g. [RD50])

$$P_{\text{mol}}(\Theta) = \frac{3}{4} (1 + \cos^2 \Theta) \frac{1 - \delta}{1 + \delta/2} \quad (62)$$

The depolarization ratio δ varies with wavelength and accounts for the anisotropy of the air molecule [RD49].

The model atmosphere also includes aerosols. For the reduced vertical RT grid ($K_{RT} = 36$), the RemoTeC algorithm describes aerosols by the following parameters (following [RD20, RD9]):

1. Number of particles in each layer of the model atmosphere. This is parameterized through the total column of aerosol particles, N_{aer} , 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 , the aerosol particle number is given by

$$h(z_k) = N_{\text{aer}} B \exp \left[-\frac{\ln 2 (z_k - z_{\text{aer}})^2}{w_0^2} \right] \quad (63)$$

where B is the normalization constant of the Gaussian.

2. A power law size distribution of $n_{\text{aer}}(r)$ with r aerosol radius, characterized by the power law exponent and an upper and lower cut-off (e.g. [RD51]):

$$n_{\text{aer}}(r) = \begin{cases} A & \text{for } r \leq r_1 \\ A \left(\frac{r}{r_1}\right)^{-\alpha} & \text{for } r_1 < r \leq r_2 \\ 0 & \text{for } r > r_2 \end{cases} \quad (64)$$



The cut-offs are $r_1 = 0.1 \mu\text{m}$, $r_2 = 10 \mu\text{m}$ and the constant A is determined from normalization of the size distribution. Figure 6 illustrates $n_{\text{aer}}(r)$ and compares it to a more realistic multimodal lognormal size distribution [RD52]. Through its parameter α , the particle size distribution controls the spectral dependence of aerosol optical properties among the considered retrieval windows.

3. The complex refractive index $m = m_r - i m_i$, which is assumed independent of wavelength within a retrieval window.

From the aerosol size distribution, refractive index, and the number of particles, the aerosol scattering optical thickness $\tau_{\text{sca,aer},k}$ and absorption optical thickness $\tau_{\text{abs,aer},k}$ are calculated for each layer k

$$\tau_{\text{sca,aer},k} = \sigma_{\text{sca,aer}} h(z_k) \Delta z_k \quad (65)$$

$$\tau_{\text{abs,aer},k} = \sigma_{\text{abs,aer}} h(z_k) \Delta z_k \quad (66)$$

 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Date : 17 May 2019	Page 35/91
--	-------------------------------	---	---------------

where $\sigma_{\text{sca,aer}}$ and $\sigma_{\text{abs,aer}}$ are the aerosol scattering and absorption cross-sections, respectively. They are obtained by:

$$\sigma_{\text{sca,aer}} = \sum_{i=1}^I K_{\text{sca},i}(m) r_i n_{\text{aer}}(r_i) v(r_i) \quad (67)$$

$$K_{\text{sca},i}(m) = \int_{\Delta \ln r_i} \frac{C_{\text{sca}}(r, m)}{v(r)} d \ln r, \quad (68)$$

where v denotes particle volume and C_{sca} is the Mie scattering cross section of a spherical particle of radius r and refractive index m . An analogous expression holds for the aerosol absorption cross sections. The kernels $K_{\text{sca},i}$ are pre-calculated for I size bins with representative radius r_i , $i = 1, \dots, I$ according to Dubovik et al., 2006 [RD53] and are stored in a lookup table as function of aerosol size parameter $x = 2\pi r/\lambda$, real refractive index, and imaginary refractive index. The values for the actual aerosol characteristics are obtained by linear interpolation from the tabulated values.² Similar expressions hold for the absorption cross-section and the aerosol scattering phase function.

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

$$\tau_{\text{abs},k} = \tau_{\text{abs,mol},k} + \tau_{\text{abs,aer},k} \quad (69)$$

$$\tau_{\text{sca},k} = \tau_{\text{sca,mol},k} + \tau_{\text{sca,aer},k} \quad (70)$$

$$P_k(\Theta) = \frac{\tau_{\text{sca,mol},k} P_{\text{mol}}(\Theta) + \tau_{\text{sca,aer},k} P_{\text{aer}}(\Theta)}{\tau_{\text{sca},k}}. \quad (71)$$

The vertically integrated absorption and scattering optical thickness τ_{abs} and τ_{sca} are the respective sums over all K_{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 $\omega_k = \tau_{\text{sca},k}/\tau_k$.

For multiple scattering calculations, the scattering phase function is expanded in L generalized spherical functions, where expansion coefficient α^l with index l is given by:

$$\alpha_k^l = \frac{2l+1}{2} \int_{-1}^1 P_{0,0}^l(\cos \Theta) P_k(\cos \Theta) d \cos \Theta. \quad (72)$$

Here, $P_{0,0}^l$ is element (1,1) of the generalized spherical function matrix (e.g. [RD54]). For RemoTeC-S5, we use $L = 16$ as baseline.

6.8.2 Physics-based forward model

The physics-based line-by-line forward model comprises three contributions, single and multiple scattering contribution of the backscattered radiances I_{ss} and I_{ms} , respectively and the contribution I_{SIF} from solar induced fluorescence (SIF) at the earth surface, namely

$$I = I_{\text{ss}} + I_{\text{ms}} + I_{\text{SIF}}. \quad (73)$$



Each component is simulated separately, where singly scattered light is simulated on a line-by-line spectral sampling and the multiple scattering simulations employ the linear- k method [RD29]. Finally, the fluorescence contribution is simulated for a non-scattering atmosphere.

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

$$I_{\text{ss}} = \frac{F_0 \tilde{\mu}}{4\pi \mu_v} \sum_{k=1}^{K_{RT}} \left\{ \omega_k P_k \left[1 - e^{-\tau_k/\tilde{\mu}} \right] e^{-\tau_i^{k-1}/\tilde{\mu}} \right\} + A \frac{\mu_0 F_0}{\pi} e^{\tau/\tilde{\mu}} \quad (74)$$

where μ_0 is the cosine of the solar zenith angle, μ_v is the cosine of the viewing zenith angle, $\tilde{\mu}$ is the effective air-mass direction cosine from Eq. 13. A is the Lambertian surface albedo with its spectral dependence parametrized in Eq. 14

² The lookup table contains values for spheres (Mie calculations) as well as for spheroids with a pre-defined axis ratio distribution [RD53]. The RemoTeC algorithm baseline consider only spherical particles.

 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 36/91
--	------------------------	--

and

$$\tau_{k_1}^{k_2} = \sum_{k=k_1}^{k_2} \tau_k . \quad (75)$$

The incoming solar irradiance can optionally be taken from a solar reference spectrum or be derived from the measured solar irradiance by deconvolution as described for the non-scattering retrieval (section 6.9). The derivatives of I_{ss} with respect to $\tau_{sca,k}$, $\tau_{abs,k}$, ω_k , P_k and A can be calculated in an analytical manner,

$$\frac{\partial I_{ss}}{\partial \tau_k} = \frac{F_0}{4\pi\mu_v} \left\{ \omega_k P_k e^{-\tau_k/\bar{\mu}} e^{-\tau_1^{k-1}/\bar{\mu}} - \sum_{k'=k+1}^{K_{RT}} \omega_{k'} P_{k'} [1 - e^{-\tau_{k'}/\bar{\mu}}] e^{-\tau_1^{k'-1}/\bar{\mu}} \right\} - A \frac{\mu_0 F_0}{\pi \bar{\mu}} e^{-\tau/\bar{\mu}} \quad (76)$$

$$\frac{\partial I_{ss}}{\partial P_k} = \frac{F_0 \bar{\mu}}{4\pi\mu_v} \omega_k [1 - e^{-\tau_k/\bar{\mu}}] e^{-\tau_1^{k-1}/\bar{\mu}} \quad (77)$$

$$\frac{\partial I_{ss}}{\partial \omega_k} = \frac{F_0 \bar{\mu}}{4\pi\mu_v} P_k [1 - e^{-\tau_k/\bar{\mu}}] e^{-\tau_1^{k-1}/\bar{\mu}} \quad (78)$$

$$\frac{\partial I_{ss}}{\partial A} = \frac{\mu_0 F_0}{\pi} e^{-\tau/\bar{\mu}} \quad (79)$$

with the deduced derivatives

$$\frac{\partial I_{ss}}{\partial \tau_{sca,k}} = \frac{\partial I_{ss}}{\partial \tau_k} + \frac{1 - \omega_k}{\tau_k} \frac{\partial I_{ss}}{\partial \omega_k} \quad (80)$$

$$\frac{\partial I_{ss}}{\partial \tau_{abs,k}} = \frac{\partial I_{ss}}{\partial \tau_k} - \frac{\omega_k}{\tau_k} \frac{\partial I_{ss}}{\partial \omega_k} . \quad (81)$$

The computation of the multiply scattered radiation involves the solution of the plane-parallel radiative transfer equation. For this purpose, we use the LINTTRAN V2.0 model [RD55], which provides the TOA radiance as well as its derivatives with respect to $\tau_{abs,k}$, $\tau_{sca,k}$, α_k^l , and A for each layer k of the model atmosphere using the forward adjoint perturbation theory. LINTTRAN is described in more detail in Appendix A.

Based on derivatives of the radiance with respect to the optical parameters τ_{abs} , τ_{sca} , and P or its expansion coefficients α^l , the derivatives with respect to a physical parameter x can be calculated in a straightforward manner by

$$\frac{\partial I_{ss}}{\partial x} = \sum_{k=1}^{K_{RT}} \left[\frac{\partial I_{ss}}{\partial \tau_{sca,k}} \frac{\partial \tau_{sca,k}}{\partial x} + \frac{\partial I_{ss}}{\partial \tau_{abs,k}} \frac{\partial \tau_{abs,k}}{\partial x} + \frac{\partial I_{ss}}{\partial P_k} \frac{\partial P_k}{\partial x} \right] \quad (82)$$

$$\frac{\partial I_{ms}}{\partial x} = \sum_{k=1}^{K_{RT}} \left[\frac{\partial I_{ms}}{\partial \tau_{sca,k}} \frac{\partial \tau_{sca,k}}{\partial x} + \frac{\partial I_{ms}}{\partial \tau_{abs,k}} \frac{\partial \tau_{abs,k}}{\partial x} + \sum_{l=0}^M \frac{\partial I_{ss}}{\partial \alpha_k^l} \frac{\partial \alpha_k^l}{\partial x} \right] . \quad (83)$$

In order to avoid time consuming multiple scattering calculations on the line-by-line grid, we aim at reducing the number of spectral calculations, following the linear- k approach of Hasekamp and Butz, 2008 [RD29]. For this purpose, we consider the intensity I_{ms} as a function of total absorption optical thickness τ_{abs} and its normalized vertical distribution \mathbf{n}


$$I_{ms}(\lambda) = I_{ms}(\tau_{abs}(\lambda), \mathbf{n}(\lambda)) . \quad (84)$$

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 \mathbf{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

$$\tau_{abs,k}(\lambda) = n_k(\lambda) \tau_{abs}(\lambda) . \quad (85)$$

For a vertically homogeneous atmosphere, I_{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. [RD56]). Finally, the

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 37/91
---	------------------------	--

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 [RD57]. 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. [RD58]).

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_{abs}^q and corresponding vertical distributions \mathbf{n}^q , with $q = 1, \dots, Q$. From the reference calculations $I_{\text{ms}}(\tau_{\text{abs}}^q, \mathbf{n}^q)$, we obtain the multiply scattered intensity vector $I_{\text{ms}}(\lambda)$ at any wavelength λ in the spectral range under consideration with absorption optical thickness $\tau_{\text{abs}}(\lambda)$ 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}^q used in the reference calculations, by employing the linear approximation:

$$I_{\text{ms}}(\tau_{\text{abs}}^q, \mathbf{n}(\lambda)) \approx I_{\text{ms}}(\tau_{\text{abs}}^q, \mathbf{n}^q) + \frac{\partial I_{\text{ms}}}{\partial \mathbf{n}} [\mathbf{n}(\lambda) - \mathbf{n}^q], \quad (86)$$

where the derivatives with respect to \mathbf{n} follow from the forward adjoint perturbation theory [RD59]. This yields the intensity vector of the multiply scattered radiation at the grid points τ_{abs}^q , corrected for the actual vertical distribution. To obtain the intensity I_{ms} at $\tau_{\text{abs}}(\lambda)$, 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 $\tau_{\text{abs}}(\lambda)$ and the two neighboring points. 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.:

$$\tau_{\text{grid}}^q = \tau_{\text{min}} \cdot \left(\frac{\tau_{\text{max}}}{\tau_{\text{min}}} \right)^{\frac{q-1}{Q-1}}. \quad (87)$$

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. For the numerical implementation we use $Q = 20$ grid points each of the three band, but this number can be updated during the commissioning phase. Figure 7 illustrates the accuracy of the linear- k method for the S5 NIR-2 band. For more information on the linear- k method we refer to [RD29].

The forward model also accounts for solar induced fluorescence (SIF) emitted by vegetation surfaces by

$$I_{\text{SIF}} = F_s(\lambda) e^{-\tau_{\text{abs}}(\lambda)/\mu_v} \quad (88)$$

as proposed by Frankenberg et al., 2012 [RD60]. Here $F_s(\lambda)$ is the SIF radiance emitting from the earth's surface. This approach ignores scattering of SIF-emitted light by the atmosphere. Furthermore, we use a linear spectral dependence of $F_s(\lambda)$, namely

$$F_s(\lambda) = f_0 + f_1 (\lambda - \lambda_0) \quad (89)$$

where $\lambda_0 = 755 \text{ nm}$. Here the coefficient f_0 is the fluorescence signal at the reference wavelength, whereas f_1 describes its spectral dependence. Obviously, this approximation means the following derivatives of the line-by-line radiance with respect to the SIF parameters,

$$\frac{\partial I}{\partial f_0} = \frac{\partial I_{\text{SIF}}}{\partial f_0} = e^{-\tau_{\text{abs}}(\lambda)/\mu_v} \quad (90)$$

$$\frac{\partial I}{\partial f_1} = \frac{\partial I_{\text{SIF}}}{\partial f_1} = (\lambda - \lambda_0) e^{-\tau_{\text{abs}}(\lambda)/\mu_v} \quad (91)$$

Finally, the spectral degradation to the sensor spectral resolution and the calculation of derivatives for spectral shift parameters is the same as for the non-scattering retrieval (section 6.3.2), providing us the the forward model \mathbf{F} for the physics based retrieval.

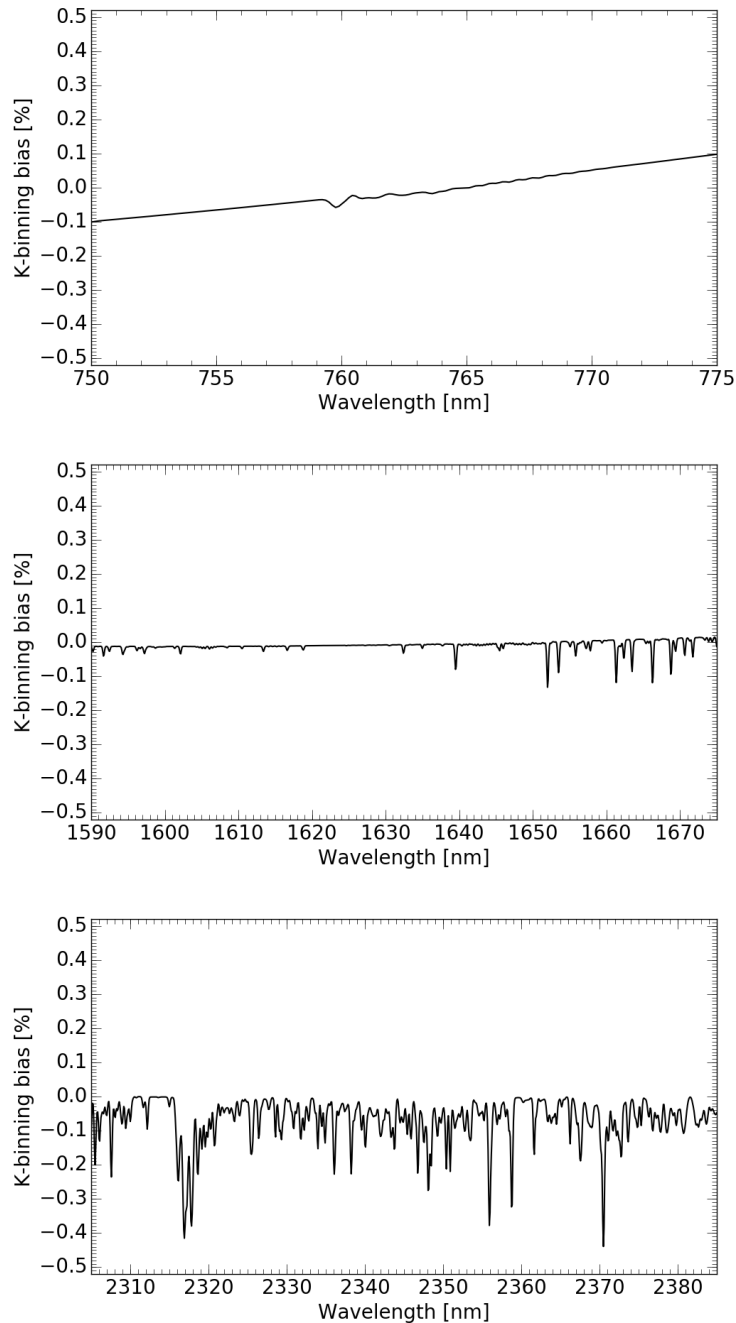


Figure 7: Relative error of the difference between a spectrum calculated using the linear- k method with respect to line-by-line calculations for the NIR-2 band (upper panel), SWIR-1 band (middle panel) and SWIR-3 band (lower panel). The spectra have been convolved with a Gaussian spectral response function with a Full Width at Half Maximum (FWHM) in Table 1. For the calculations a boundary layer aerosol was used with an optical thickness of 0.3 at 765 nm. Furthermore, we used a solar zenith angle (SZA) of 50° and a viewing zenith angle of 0° .



 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Date : 17 May 2019	Page 39/91
--	------------------------	---	---------------

Table 6: State vector elements and their a priori and first guess values for the physics-based three band retrieval. The state vector of a two band retrieval is defined by a subset, omitting the CO₂ total column and all entries related to the SWIR-1 band. The symbol L₁ and L₀ indicates first and zero order Tikhonov regularization, ∅ indicates no side-constraint for these parameters.

Number of elements	State vector element	A priori and first guess value	regularization
$K_{\text{ret}} = 12$	CH ₄ sub-columns	CTM/DEM	L ₁
1	Aerosol parameter N_{aer} (number column)	$\tau_{\text{aer}} = 0.1$ at 765 nm	L ₀
1	Aerosol parameter α (size parameter)	4.0	L ₀
1	Aerosol parameter z_{aer} (central height of Gaussian height distribution)	5000 m	L ₀
1	CO ₂ total column	CTM/DEM	∅
1	H ₂ O total column	CTM/DEM	∅
1	CO total column	CTM/DEM	∅
3	Lambertian surface albedo in NIR-2, 2nd order spectral dependence	maximum of measured reflectance	∅
3	Lambertian surface albedo in SWIR-1, 2nd order spectral dependence	maximum of measured reflectance	∅
3	Lambertian surface albedo in SWIR-3, 2nd order spectral dependence	maximum of measured reflectance	∅
1	Spectral shift of earth radiances in NIR-2 (higher orders optional)	zero	∅
1	Spectral shift of earth radiances in SWIR-1 (higher orders optional)	zero	∅
1	Spectral shift of earth radiances in SWIR-3 (higher orders optional)	zero	∅
1	Spectral shift of solar irradiances in NIR-2 (higher orders optional)	zero	∅
1	Spectral shift of solar irradiances in SWIR-1 (higher orders optional)	zero	∅
1	Spectral shift of solar irradiances in SWIR-3 (higher orders optional)	zero	∅
2	Fluorescent emission and 1st order spectral dependence	tbd	∅


6.8.3 Physics-based inversion

The physics-based inversion retrieves a set of atmospheric parameters from the S5 measurements. As for the non-scattering retrieval, the retrieved atmospheric parameters are chained into a state vector \mathbf{x} , and the S5-L1B radiances are chained into the measurement vector \mathbf{y} . The physics-based forward model \mathcal{F} links the state vector to the measurement vector. Thus, the equation to be inverted for \mathbf{x} is the same as Eq. (26), but now with the full-physics forward model \mathcal{F} .

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \mathbf{b}) + \mathbf{e}_y + \mathbf{e}_f. \quad (92)$$

Here \mathbf{e}_y and \mathbf{e}_f represents the measurement noise and the forward model error, respectively. For the physics-based three band retrieval, the state vector comprises the parameters listed in Table 6, the state vector elements of a two band retrieval are same but with all SWIR-1 related elements omitted.

The physics-based inversion employs the regularization scheme described in Sect. 6.3.3 for the vertical profile of CH₄. Furthermore, aerosol-related parameters are also constrained to improve the stability. For instance, if the aerosol column tends to very low values, the model loses sensitivity to the aerosol size parameter and the aerosol height. A

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 40/91
---	------------------------	--

side-constraint is necessary to keep the inversion stable. The aerosol-related parameters are directly regularized, using a zeroth derivative matrix \mathbf{L}_0 , which only scales the parameters to control the regularization strength. The regularization matrix has $K_{\text{ret}} + 2$ rows, constraining $K_{\text{ret}} - 1$ first derivatives and 3 aerosol parameters.

$$\mathbf{W} = \begin{bmatrix} \frac{\gamma_j}{C_{j1}^{\text{ref}}} & -\frac{\gamma_j}{C_{j2}^{\text{ref}}} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & \frac{\gamma_j}{C_{j2}^{\text{ref}}} & -\frac{\gamma_j}{C_{j3}^{\text{ref}}} & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & \frac{\gamma_j}{C_{j3}^{\text{ref}}} & -\frac{\gamma_j}{C_{j4}^{\text{ref}}} & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & 0 & \cdots & \gamma_{a1} & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \gamma_{a2} & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \gamma_{a3} & 0 & 0 & \cdots \end{bmatrix} \quad (93)$$

Note that j refers to CH_4 for the S5 physics-based retrieval.

The ancillary parameter vector \mathbf{b} comprises all non-retrieved parameters required for evaluating the forward model. The most important ones are:

- Temperature vertical profile at $K_{\text{atm}} = 72$ layers of the cross-section vertical grid.
- Pressure vertical profile at $K_{\text{atm}} = 72$ layers of the cross-section vertical grid.
- The aerosol complex refractive index, fixed at $m = 1.4 - 0.01i$ for the NIR-2 and $m = 1.47 - 0.008i$ for the SWIR-1 and SWIR-3 bands.
- The width w_0 of the Gaussian aerosol height distribution, fixed at 2000 m.
- Molecular absorption cross sections σ_j .
- Solar irradiance spectrum F_0 .
- Instrument spectral response function (ISRF) s .
- Solar and viewing observation geometry μ_0 and μ_v and the relative azimuth angle φ .

The measurement vector \mathbf{y} comprises the S5 radiance measurements of the NIR-2 band (755-773 nm), the SWIR-3 band (2305-2385 nm), and in case of a three band retrieval the SWIR-1 band (1593-1654 nm).

Given the state vector \mathbf{x} , the measurements \mathbf{y} , and the forward model \mathcal{F} of this section, the inversion procedure aims at estimating the state vector by inverting Eq. (92) iteratively. This follows the same procedure as in Sect. 6.3.3, using the iterative approach with a step size reduction system to increase the stability.

The final state vector $\mathbf{x}_{\text{ret}} = \mathbf{x}_N$ is related to the true state vector and to the prior via

$$\mathbf{x}_{\text{ret}} = \mathbf{A}\mathbf{x}_{\text{true}} + (\mathbf{I} - \mathbf{A})\mathbf{x}_a + \mathbf{G}\mathbf{e}_y + \mathbf{G}\mathbf{e}_f. \quad (94)$$


with the gain matrix \mathbf{G} and the averaging kernel \mathbf{A} as defined in Eq. (35) and (36), respectively. The covariance matrix \mathbf{S}_x of the retrieval noise $\mathbf{G}\mathbf{e}_y$ is given by

$$\mathbf{S}_x = \mathbf{G}\mathbf{S}_y\mathbf{G}^T. \quad (95)$$

For the S5 mission, the operational methane product is the column averaged dry air mole fraction of methane, XCH_4 . This quantity is obtained from the methane entries of the retrieved state vector through

$$\text{XCH}_4 = \frac{C_{\text{CH}_4}}{C_{\text{air,dry}}} \quad (96)$$

with $C_{\text{CH}_4} = \sum_{i=1}^{K_{\text{ret}}} \mathbf{x}_{i,\text{ret}}$, assuming that the CH_4 subcolumns are the first twelve state vector parameters. $C_{\text{air,dry}}$ represents the dry air column calculated from the surface pressure and water vapor profile, both obtained from a

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001
		Version : 3.1 Page
		Date : 17 May 2019 41/91

meteorological model (required as input). The retrieval noise $\varepsilon_{\text{XCH}_4}$ is the error estimate that is given as output together with XCH_4 . It reads

$$\varepsilon_{\text{XCH}_4} = \frac{\sqrt{\sum_{i=1}^{K_{\text{ret}}} \sum_{j=1}^{K_{\text{ret}}} S_{x,i,j}}}{C_{\text{air,dry}}} \quad (97)$$

where the retrieval covariance matrix is summed over all elements related to the CH_4 sub-columns.

For validation and application purposes, it is important to realize that the provided data product XCH_4 represents an estimate of the true total column of CH_4 due to the prior contribution in Eq. (94). The actual vertical sensitivity of the retrieval is described by

$$\mathbf{a}_k = \frac{C_{\text{CH}_4,\text{ret}}}{C_{\text{CH}_4,k}} \sum_{i=1}^{K_{\text{ret}}} \sum_y \mathbf{G}_{iy} \mathbf{K}_{yk} = \sum_{i=1}^{K_{\text{ret}}} \mathbf{A}_{ik} \quad (98)$$

and is referred to as the column averaging kernel [RD61]. It is also part of the physics-based retrieval output. The definition of the column averaging kernel is set such that it is unitless and a value of one represents perfect sensitivity.

An important diagnostic tool is the residual χ^2 , which is defined as the difference between the measured and fit spectrum normalized to the instrument noise estimate.

$$\chi^2 = \frac{(\mathbf{y} - \mathcal{F}(\mathbf{x}_{\text{ret}}))^T \mathbf{S}_y^{-1} (\mathbf{y} - \mathcal{F}(\mathbf{x}_{\text{ret}}))}{N_y - \text{DFS}} \quad (99)$$

with N_y is the number of measurements and

$$\text{DFS} = \sum_i A_{ii} \quad (100)$$

is the degrees of freedom for signal. If there are no forward model errors, χ^2 reaches a value around one.

The S5 L2 CH_4 product also includes a QA-value. This is a number ranging from zero to one indicating the reliability of the product. As baseline, the QA value is only based on the physics-based retrieval and is calculated as follows:

- If the retrieval did not converge, the QA-value is set to 0.
- If a side parameter (e.g. aerosol height) tends to unphysical values during the inversion, the QA-value is set to 0.4.
- Otherwise, if the deconvolution of the solar spectrum failed (reverting to the model solar spectrum), the QA-value is set to 0.8.
- If nothing of the above happened, the QA-value is set to 1.

6.9 Deconvolution solar spectrum



In Sects. 6.3.2 and 6.8.2, it was mentioned that as an alternative of using an a-priori model solar spectrum, the solar spectrum can be deduced from S5 irradiance measurements. We assume that there is only one irradiance measurement available during the execution of the processor, so the deconvolution of the solar spectrum is executed at initialization.

Analogous to Eq. (24), we can simulate the solar measurement $\mathbf{F}_{0,\text{meas}}$ by

$$\mathbf{F}_{0,\text{meas}}(\lambda_i) = \mathbf{S}_{\text{sun}} \mathbf{F}_0(\lambda) \quad (101)$$

where \mathbf{S}_{sun} is the ISRF matrix of the solar measurement analogous to Eq. (24). Before executing the deconvolution, the solar measurement is spectrally shifted 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 S5 irradiance measurements are compared with the model solar spectrum convolved with the ISRF,

$$\delta\lambda_s = \max_{\delta\lambda_s} \rho(\mathbf{F}_{0,\text{meas}}(\lambda + \delta\lambda_s), \mathbf{F}_0^*(\lambda)) \quad (102)$$

 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 42/91
--	------------------------	--

where $\rho(X, Y)$ is the Pearson's correlation coefficient.

After correcting the spectral shift, the deconvolution is executed by inverting Eq. 101, where $\mathbf{F}_{0,\text{meas}}$ is the measurement (\mathbf{y}), the deconvolved solar spectrum \mathbf{F}_0 represents the state to be retrieved (\mathbf{x}) and the ISRF matrix S_{sun} is the Jacobian (\mathbf{K}). This is a linear problem, so no iterative approach is required. However, this problem is underdetermined because the measurement vector contains fewer values than the state to be retrieved.

Van Deelen et al. (2007) [RD62] showed that the least squares minimum length solution, which minimizes the length of the solution vector as a side constraint, is of sufficient accuracy to simulate earth radiance measurements of the GOME mission. Following this approach, we obtain

$$\mathbf{F}_0 = \mathbf{S}_{\text{sun}}^T (\mathbf{S}_{\text{sun}} \mathbf{S}_{\text{sun}}^T)^{-1} \mathbf{F}_{0,\text{meas}}, \quad (103)$$

which contains significant noise contributions. However, this noise is in the null-space of the ISRF and will affect the simulation only little after the final convolution of the earth radiance spectrum with its corresponding ISRF. Moreover, to mitigate edge effects, we execute the deconvolution on a slightly extended window (~ 3 nm extra on both sides).

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. For further details and for an overview of the benefits of this method see [RD62] and [RD63].

As a correction of a spectral shift in the solar spectrum ($\delta\lambda_s$) is involved in the deconvolution process, the baseline is not to fit $\delta\lambda_s$ again during the retrievals if the deconvolved solar spectrum is used.

6.10 A posteriori data filtering and bias correction

Considering XCH_4 retrievals from current and past missions, the data processing can be divided in several main steps: a pre-processing screening is applied, which aims to filter out corrupted level 1 spectra and exposures where clouds are within the instrument FOV. After the actual retrieval, which is described in detail in this document, an additional data filtering is applied based on level-2 data quality flags. Finally, the quality of the post-screened data product can be improved further using empirical bias corrections due to instrument malfunction and retrieval algorithm artifacts. Obviously, this last processing step depends on both the performance of the instrument and the retrieval algorithm. Here, we summarize the RemoTeC XCH_4 bias correction as it is derived to correct the GOSAT GHG-CCI data product of ESA's Climate Change Initiative (CCI) (<http://www.esa-ghg-cci.org/>). For GOSAT, the RemoTeC CH_4 physics-based products is quality filtered after processing, based on data quality flags defined in [RD64]. Subsequently, the level-2 product is bias-corrected utilizing correlations of the difference between the RemoTeC XCH_4 product and collocated TCCON measurements with retrieval parameters. It was found that the error in XCH_4 correlates with α , the reciprocal of the retrieved aerosol size parameter, namely

$$\text{XCH}_4^{\text{corr}} = \text{XCH}_4 (a + b\alpha) \quad (104)$$

with empirically determined constants a and b . Other algorithms use different correction schemes, showing that the bias correction not only corrects for instrumental effects on the retrieval. For S5, we expect much larger validation data set than available for GOSAT, which may allow to apply alternative correction schemes based e.g. on a trained neural network (see e.g. [RD65]). Obviously, details of a correction scheme cannot be determined in advance but a processing scheme should consider these corrections as an add-on in a later phase of the operational processing.

6.11 Detailed Algorithm Setup

In the previous sections, we described in detail the theoretical baseline of the RemoTeC algorithm with the overall flow diagram in Fig. 4. This section focuses on the S5 specific software implementation, which results in an operational efficient and stable software. The current RemoTeC software separates the software initialization from the actual ground-pixel based retrieval and so allows for performance optimization within an overall processing framework. Here, we distinguish between the initialization, whose computation can be parallelized, from a serial initialization. Finally, the actual (non-scattering and full-physics) retrievals are performed per ground scene in a separate software module, for which parallel computing is an important mean for performance optimization. Figure 8 summarizes the overall software structure with the three main modules, which we discuss below in more detail.

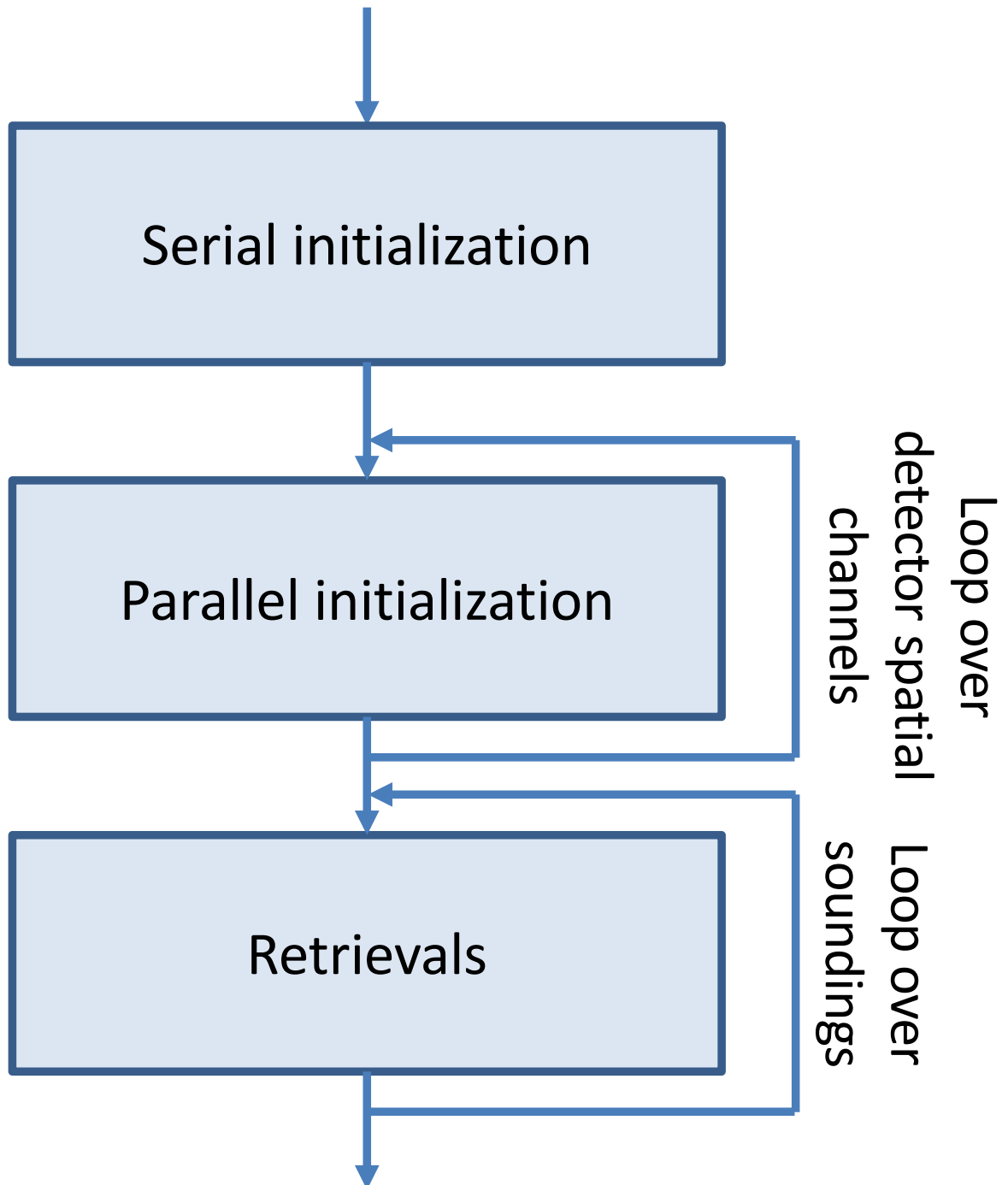


Figure 8: RemoTeC algorithm architecture for initialization and pixel processing optimized for parallel computing. Here, two parallelization options are foreseen: first, the initialization of the algorithm with respect to the detector spatial channel index of the measurements and second the actual retrievals with respect to the pixels to be processed within a data granule.

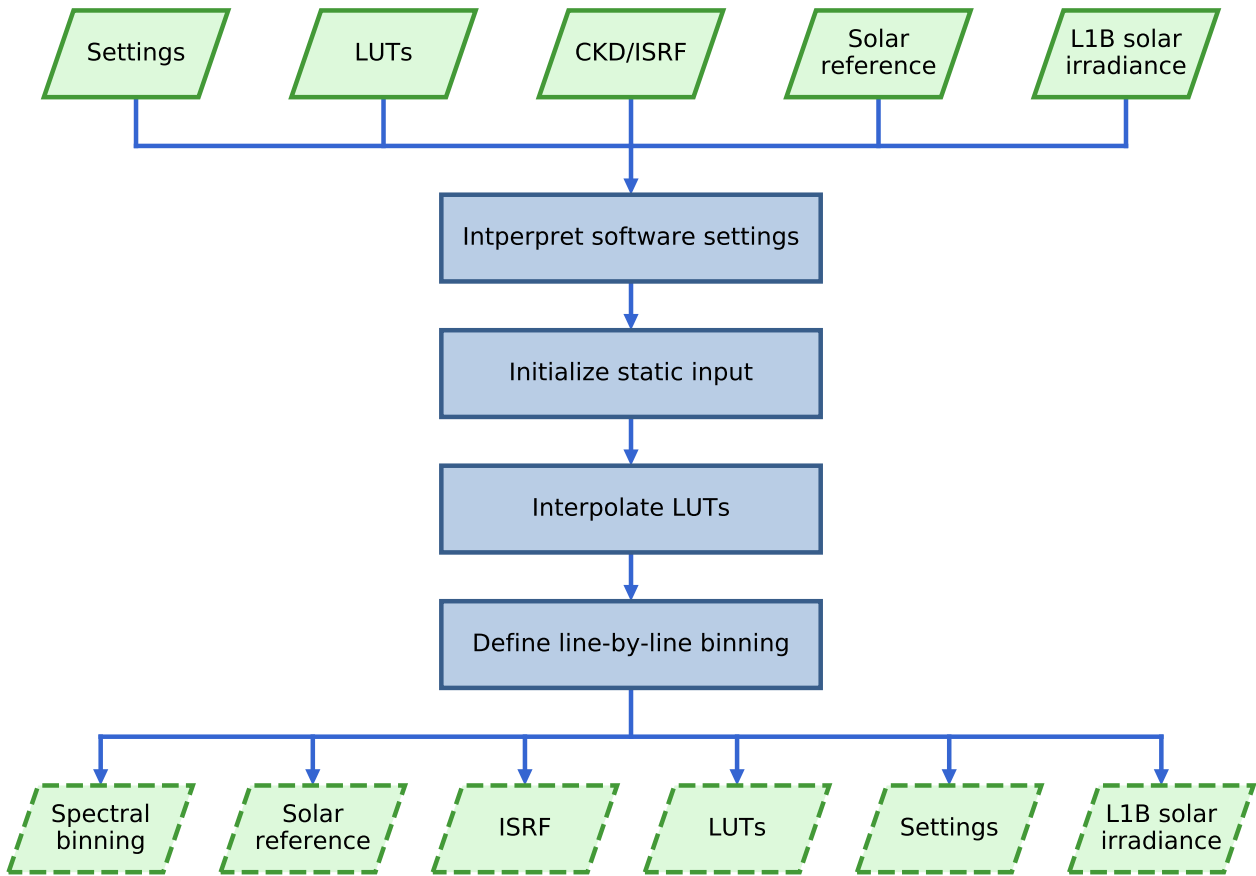


Figure 9: Serial initialization, which is executed once per processor call. The solid green boxes indicate static input, the dashed green boxes indicate output to the remaining RemoTeC software provided by a pointer to the internal memory.

Table 7: Static input of the serial algorithm initialization.

Parameter	Symbol	Source
Algorithm settings	-	S5L2PP
Absorption cross sections CH ₄ , H ₂ O, CO ₂ and CO	$\sigma_{\text{CH}_4}, \sigma_{\text{H}_2\text{O}}, \sigma_{\text{CO}_2}, \sigma_{\text{CO}}$	LUT
Reference solar spectrum	$F_{0,\text{ref}}$	LUT
Instrument spectral response function	S	CKD
Irradiance spectrum NIR/SWIR-1/SWIR-3	F_0	L1b (Coregistered)
Hybrid pressure coefficient A, B	$A_{\text{hyb}}, B_{\text{hyb}}$	CAMS / ECMWF

6.11.1 Serial Initialization

The serial initialization provides retrieval input, which has to be allocated once per processing call. The numerical effort is minor and any parallelization of the software is not considered. The module requires static input data, indicated by the solid green data boxes in Fig. 9. It comprises the retrieval input settings, lookup tables for molecular absorption cross sections and aerosol optical properties, the ISRF as part of the instrument calibration key data and a solar reference spectrum. Furthermore, the module defines the model internal line-by-line spectral grid. All data created by the initialization, which are indicated by the green dashed data boxes in the figure, are accessible for the remaining software via pointers to the internal memory. The static input of this algorithm component is summarized in Tab. 7. The module also provides the static output of Tab. 8, which is the part of the S5 L2 output product that is applicable for all retrievals.

Table 8: Output data of the static initialization.

Parameter	Symbol	Destination
Model atmosphere grid	ATM	S5 L2 CH ₄ product
SIF wavelengths	SIF- λ	S5 L2 CH ₄ product

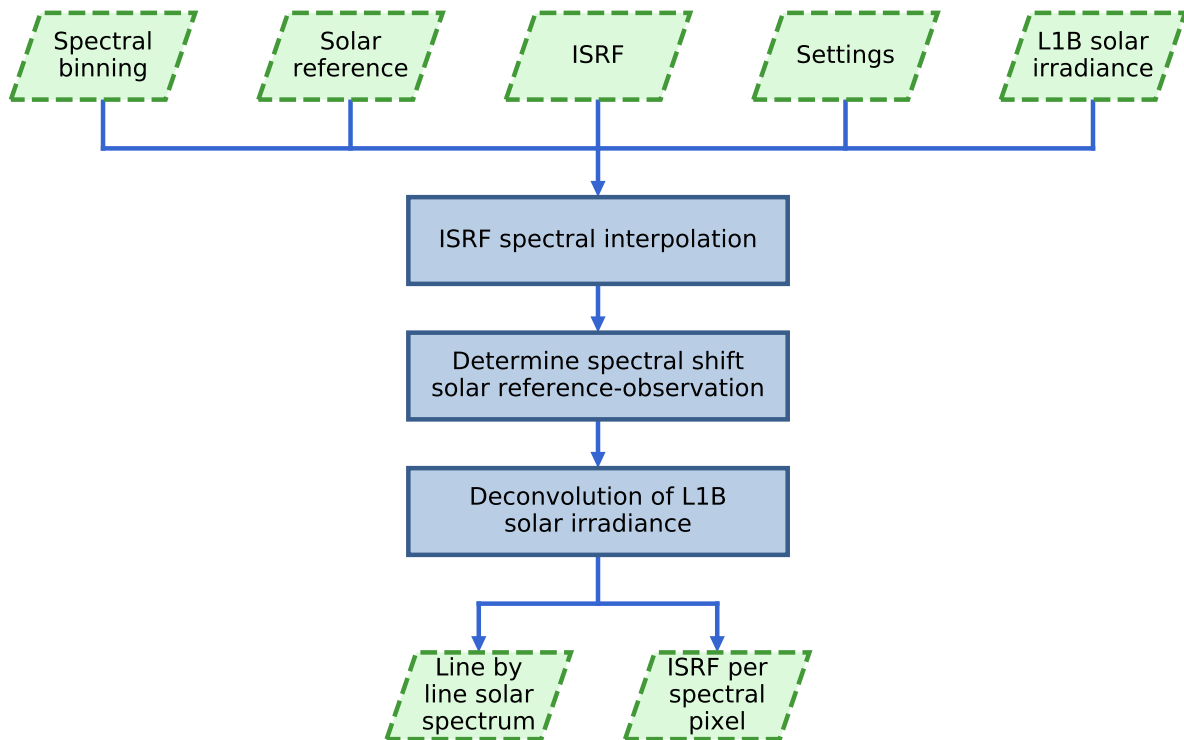


Figure 10: Parallel initialization: Based on input from the internal memory, the ISRF from the calibration key data is adapted to each spectral pixel. Moreover, the spectral shift between solar reference spectrum and observation is determined. This is needed as input for the subsequent de-convolution of the solar spectrum.

6.11.2 Parallel initialization

Part of the algorithm initialization can be optimized by parallel computing. It includes independent initialization steps for the different positions of the ground pixel within the across-track swath. For example, the calibration key data of the ISRF has to be adapted for each spectral and spatial pixel along the swath. Here, the swath position of a ground pixel is an instrument specific property that can be indicated by the detector spatial channel index. Similarly, the measured solar spectrum varies with swath position. Therefore, the spectral deconvolution approach to estimate the solar line-by-line spectrum has to be performed for each swath position and so corresponding calculations can be parallelized.

This parallel initialization does not receive any input from the calling framework nor does it provide output to it. Data transfer from the serial initialization or to the retrieval part of RemoTeC is managed by transfer of pointers to the internal memory.

6.11.3 Retrievals

The S5 specific retrievals, based on individual ground pixel observations, are performed by a separate module and can be computed parallel for all observations of the data granule to be processed. The parallelization over ground pixels is essential to satisfy the performance requirements. The software architecture of the module is depicted in Fig. 11 and reflects the algorithm design in Fig. 4. Here, several input data stem from the initialization and are provided via pointers


	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001
		Version : 3.1 Page
		Date : 17 May 2019 46/91

Table 9: Dynamic input to the CH₄ retrieval algorithm. Metimage input as specified in Tab. 2.

Parameter	Symbol	Source
Radiance spectrum NIR/SWIR-1/SWIR-3	I	L1b (Coregistered)
Solar geomtry*	ϑ_0, φ_0	L1b (Coregistered)
Viewing geometry*	ϑ_v, φ_v	L1b (Coregistered)
Surface elevation of pixel	z_{surf}	DEM**
surface classification	f_{surf}	DEM**
Surface pressure	p_{surf}	CAMS / ECMWF**
CH ₄ , CO ₂ and CO mixing ratio profile	$r_{\text{CH}_4}, r_{\text{CO}_2}, r_{\text{CO}}$	CAMS**
Surface elevation	$z_{\text{surf,ECMWF}}$	ECMWF**
Specific humidity profile	q	ECMWF**
Temperature profile	T	ECMWF**
10 m wind	(u, v)	ECMWF**

* The S5 RemoTeC algorithm assumes spatially collocated measurements and thus common solar and viewing geometry for all three bands

** Parameters are provided by the S5 L2 AUX product after regridding.

to the internal memory. It includes retrieval settings, the line-by-line solar irradiances, the corresponding spectral grid, the ISRF and lookup tables for aerosols and molecular absorption. Table 9 gives an overview of the dynamic input, which includes the L1B radiances, the surface elevation from the digital elevation map (DEM), the atmospheric trace gas profiles from the chemical transport model (CTM) and the S5 L2 METimage cloud product. The RemoTeC algorithm requires spatially and temporally collocated Earth radiances with common solar and viewing geometry for the different bands. After a first data filtering as discussed in Sect. 6.2, the non-scattering retrievals are performed for spectral windows and target species in Tab. 3. Subsequently the proxy data filter B selects the data for which the proxy data is computed, and cloud filter C refines data for the physics-based retrieval. As shown in Fig. 11, the S5 CH₄ product is provided in parts. As a consequence, part of the output may have valid data and another part of the output will have missing data, depending on the filters A, B and C. The entire output product, including missing data, will be transferred to the framework. The most important output is summarized in Tab. 10.

Both the non-scattering and physics based retrievals are performed by the same software tool, where specific settings switch between the different retrieval approaches. Figure 12 shows the architecture of this software element. First, RemoTeC derives the measurements for the selected fitting windows. Next, the a-priori state is constructed, which includes aerosols if it is the physics-based retrieval. Setting up the state and measurement vector initializes the iterative inversion. The forward model subsequently performs line-by-line radiative transfer simulations and convolves these with the ISRF. Having these calculation at hand, the goodness of the forward simulation is estimated by a χ^2 test and the Levenberg-Marquardt parameter is adjusted to control the step-size of the inversion. In case convergence is not yet achieved, the state vector is updated using the approach of Sects. 6.3.3 and 6.8.3 and forward model parameters are adjusted accordingly to start the next iteration step. In case of convergence, a generic retrieval output is written, from which the calling routine can extract the required fields.

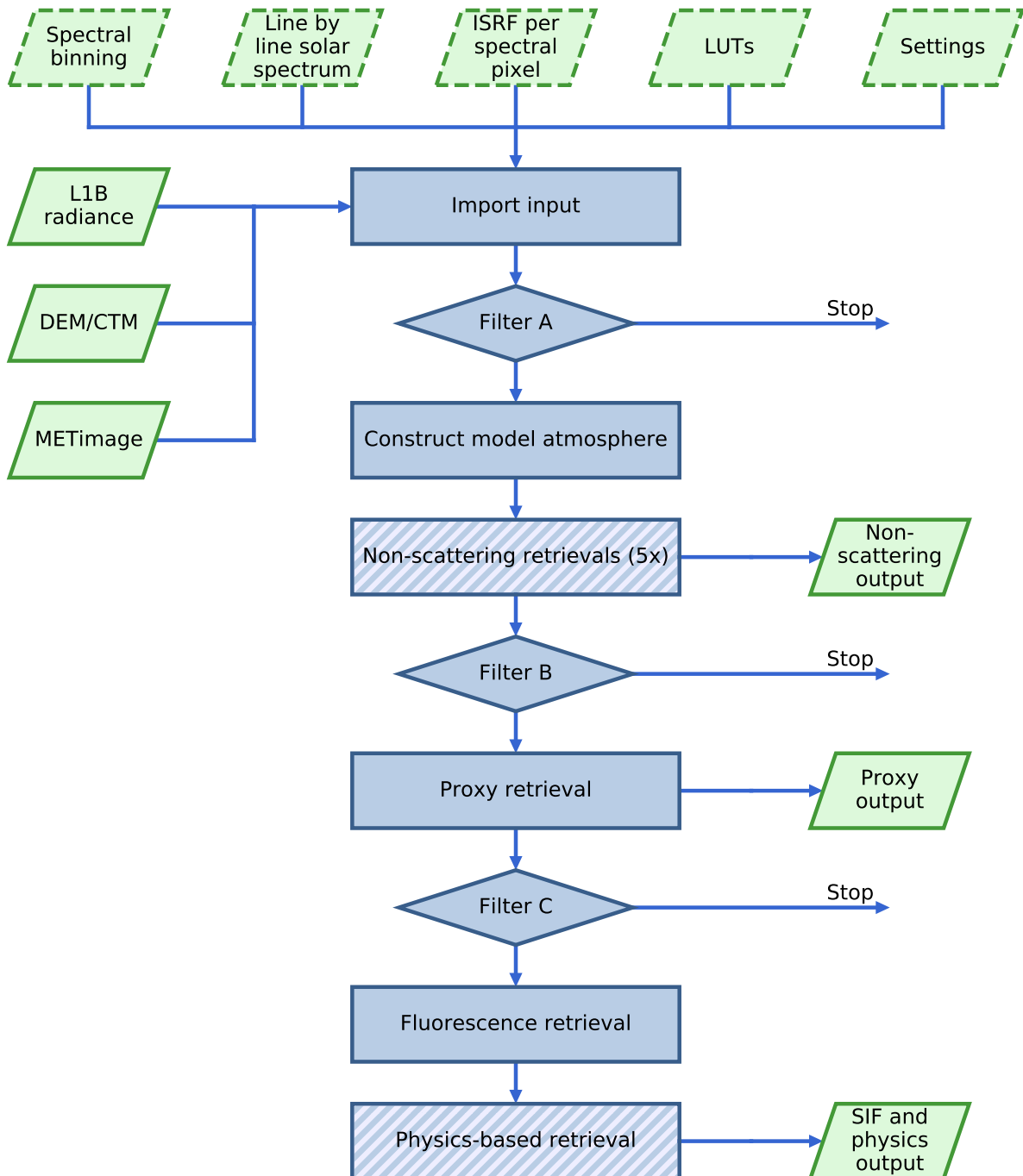


Figure 11: Software architecture of the S5 methane L2 algorithm. The shaded boxes indicated two specific calls of the retrieval model (non-scattering retrieval and physics-based retrieval), which is explained in more detail in Fig. 12.

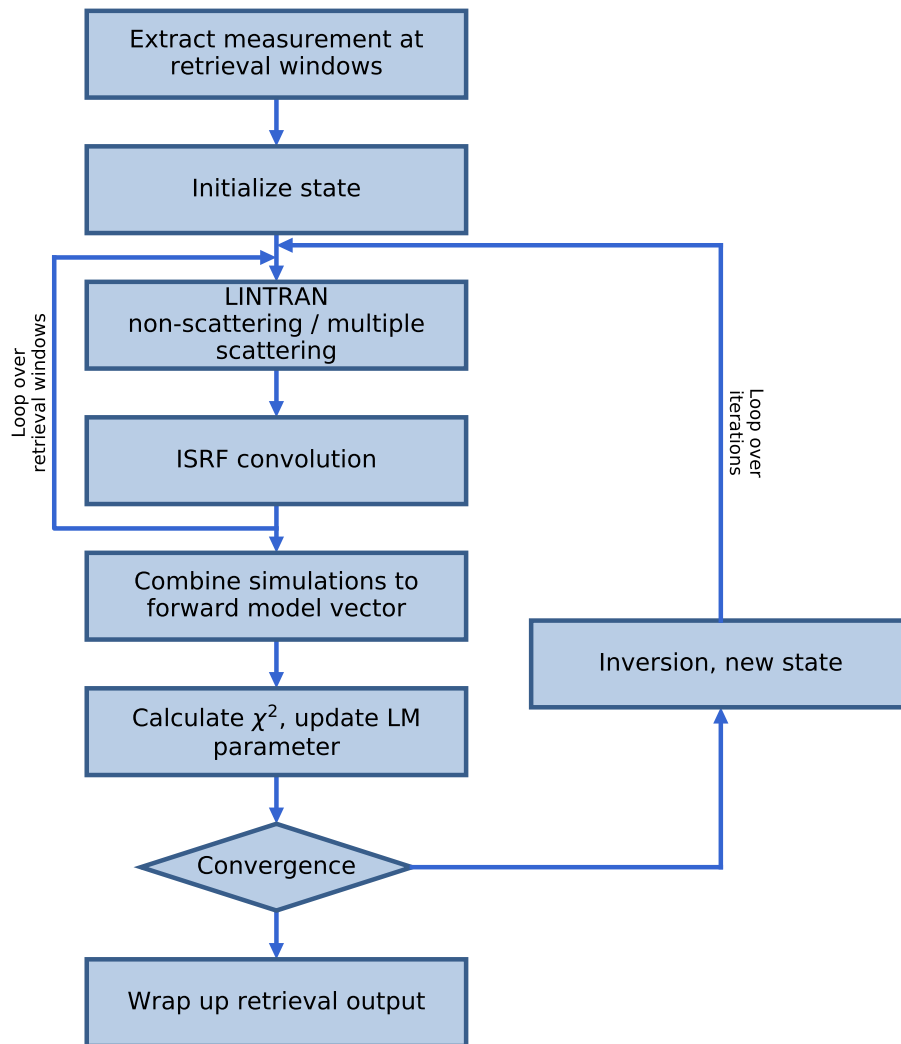


Figure 12: Software architecture of the non-scattering and physics-based retrieval.

Table 10: Dynamic output data.

Parameter	Symbol	Destination
Proxy CH ₄ dry air column mixing ratio	$XCH_{4,proxy}$	S5 L2 CH ₄ product
Proxy XCH ₄ noise estimate	$e_{XCH_{4,proxy}}$	S5 L2 CH ₄ product
Solar induced fluorescence	F_s	S5 L2 CH ₄ product
Physics-based CH ₄ dry air column mixing ratio	$XCH_{4,physics}$	S5 L2 CH ₄ product
Physics-based XCH ₄ noise estimate	$e_{XCH_{4,physics}}$	S5 L2 CH ₄ product
CH ₄ column averaging kernel	a	S5 L2 CH ₄ product
H ₂ O total column	C_{H_2O}	S5 L2 CH ₄ product
CO ₂ total column	C_{CO_2}	S5 L2 CH ₄ product
CO total column	C_{CO}	S5 L2 CH ₄ product
Aerosol particle column, size parameter, layer height	N_{aer}, α, z_{aer}	S5 L2 CH ₄ product
NIR/SWIR-1/SWIR-3 surface albedo	A	S5 L2 CH ₄ product
NIR/SWIR-1/SWIR-3 spectral shifts	$\delta_{sun}, \delta_{earth}$	S5 L2 CH ₄ product
Number of iterations	N_{iter}	S5 L2 CH ₄ product
NIR/SWIR-1/SWIR-3 χ^2 of the spectral fit	χ^2	S5 L2 CH ₄ product
Total reduced χ^2 of the spectral fit	χ_{tot}^2	S5 L2 CH ₄ product



 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Date : 17 May 2019	Page 50/91
--	------------------------	---	---------------

Table 11: Parameters a , b and N of the S5 instrument noise model in Eq. 105 at the reference wavelength λ_{ref} .

Band	λ_{ref} (nm)	a ($\text{cm}^2 \text{ s sr nm/ph}$)	b	N
NIR-2	755	$6.47 \cdot 10^{-9}$	187	36
SWIR-1	1630	$2.27 \cdot 10^{-8}$	193	9
SWIR-3	2344	$7.00 \cdot 10^{-8}$	212	3

Table 12: Reference scenario on which the generic ensembles are based.

Property	Value
CH ₄	0.641 mol m ⁻²
CO ₂	143 mol m ⁻²
H ₂ O	783 mol m ⁻²
CO	0.0394 mol m ⁻²
A_{NIR}	0.25
$A_{\text{SWIR-1}}$	0.3
$A_{\text{SWIR-3}}$	0.2
Aerosol optical depth (765 nm)	0.1
Aerosol center height	5 km
Aerosol geometric thickness	2 km
Solar zenith angle	50°
Viewing zenith angle	0°

7 Error Analysis

In this section, we test the RemoTeC-S5 algorithm performance with respect to instrument related errors and estimate the retrieval sensitivity to uncertainties in instrument parameters and atmospheric input. We compare the induced methane errors with the requirements on the methane data product as summarized in Sect. 4.3. The analysis is performed for both the proxy and the physics-based product, where for the latter we consider the three-band retrieval (i.e. NIR-2, SWIR-1 and SWIR-3) as baseline.

To account for noise on individual radiance measurements I , we employ ESA's Sentinel 5 noise model for Earth shine observations,

$$\text{SNR}(\lambda) = \sqrt{N} \frac{aI}{\sqrt{aI+b^2}} \quad (105)$$

(pers. comm. H. Weber, ESA, May 12, 2017) where a and b are instrument specific parameters and N represents the spectral binning factor. The Earth radiance I is given in photons/(cm² s sr nm). The full wavelength dependence of the parameters is applied in the noise model, whereas Tab. 11 reports only representative values of a and b at a certain reference wavelength λ_{ref} per band for conciseness.

The retrieval performance is characterized by the retrieval noise $\varepsilon_{\text{XCH}_4}$, see Eq. (97), and the bias b_{XCH_4} , which is defined as the difference between the retrieved XCH₄ without retrieval noise and the truth.

7.1 Sensitivity to instrument errors

7.1.1 Generic ensembles

In first instance we considered an ensemble of generic measurement simulations that refer to the atmospheric reference scenario described in Tab. 12. It includes an atmospheric aerosol layer at 5 km altitude with a moderate aerosol optical depth of 0.1 at 765 nm. The corresponding profiles of CH₄, CO₂, H₂O and CO are displayed in Fig. 13.

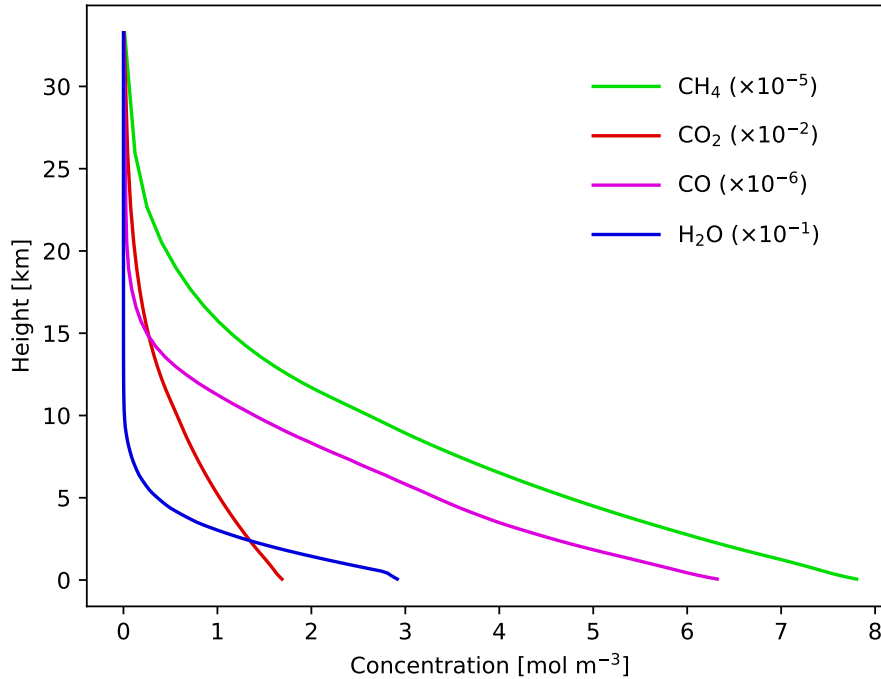


Figure 13: Vertical profiles of CH₄, CO₂, H₂O and CO used for the generic ensembles.

Table 13: Interdependence of surface albedo in the three bands used for the generic ensemble for increasing scene brightness.

A_{NIR}	$A_{\text{SWIR-1}}$	$A_{\text{SWIR-3}}$
0.15	0.1	0.05
0.2	0.2	0.125
0.25	0.3	0.2
0.3	0.4	0.275
0.35	0.5	0.35
0.4	0.6	0.425
0.45	0.7	0.5

7.1.2 Noise analysis

To estimate the XCH₄ precision due to the instrument noise, we described the atmospheric aerosols in the measurement simulation consistently with the forward model of the physics-based retrieval. To study the XCH₄ precision as a function of scene brightness, we assume an interdependence of the Lambertian albedo in the NIR-2, SWIR-1 and SWIR-3 which is appropriate to most surface types except for ice and snow. For the first generic ensemble, we coupled the surface albedos in the three bands as indicated in Tab. 13 and consider the SWIR-1 albedo as the independent surface parameter to be varied. Besides the surface albedo, also the solar zenith angle strongly governs the scene brightness and the XCH₄ precision for clear sky observations.

Figure 14 depicts the XCH₄ precision due to measurement noise as a function of both parameters. Here, the proxy XCH₄ product is more noisy than the physics based product, because the retrieval method relies on two column products retrieved from the SWIR-1 spectral range using less measurement points than the physics-based retrieval. Obviously, both retrievals show largest noise propagation for low sun and dark surfaces due to the low radiance signal

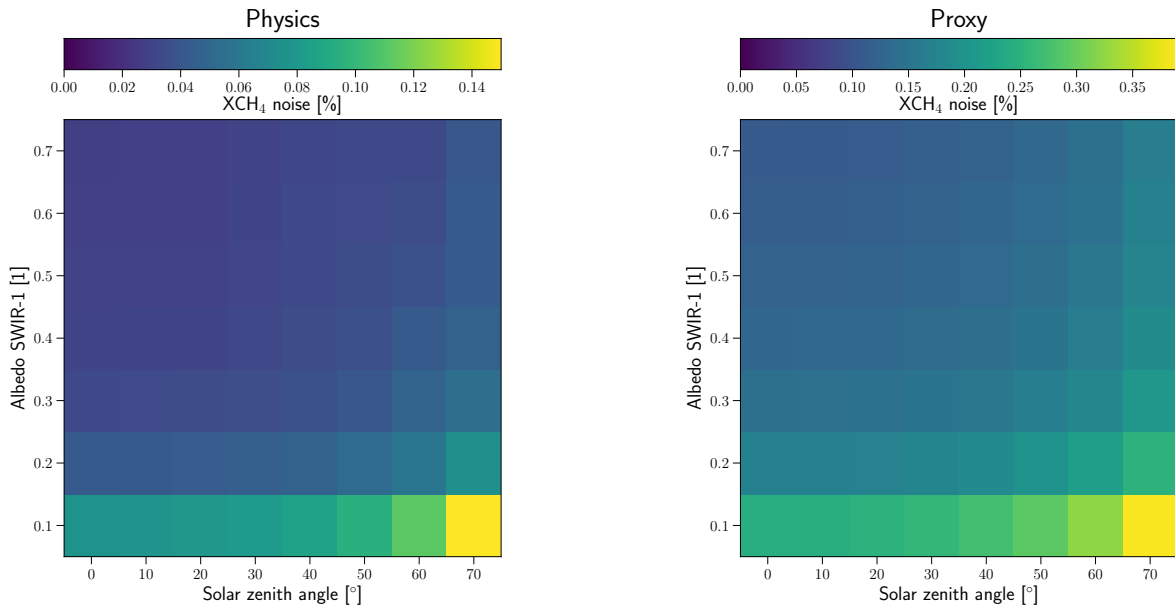


Figure 14: XCH₄ precision of the physics-based and proxy retrieval product for the reference scenario but for varying surface albedo and solar zenith angle. Surface albedos in NIR-2, SWIR-1 and SWIR-3 are interdependent according to Tab. 13.

and therefore the low SNR, but in all cases the expected XCH₄ precision stays clearly within the requirement of the product of $\leq 1\%$.

Besides the estimated retrieval precision, the column averaging kernel from Eq. (98) is also essential for interpretation of the methane product. To illustrate the variability of the column averaging kernel, these are shown for the lowest albedo, $(A_{\text{NIR}}, A_{\text{SWIR-1}}, A_{\text{SWIR-3}}) = (0.15, 0.1, 0.05)$, with varying solar zenith angle in Fig. 15. It shows that the methane retrieval is slightly more sensitive to CH₄ in the lower atmosphere than in the higher atmosphere, which is caused by atmospheric scattering and pressure-induced broadening of absorption lines.

7.1.3 ISRF distortion

To investigate the sensitivity of the S5 XCH₄ product to knowledge errors of the ISRF, we distorted the full-width half-maximum (FWHM) of the instrument spectral response function in the three spectral bands between $\pm 1\%$ in the measurement simulation and use the unperturbed ISRF in the retrieval. The induced XCH₄ biases are shown in Fig. 16.

A wider ISRF reduces the depth of absorption lines and so leads to an underestimation of the retrieved XCH₄ for a positive ISRF stretch in the SWIR bands. Also for the NIR-2 band, a stretched ISRF yields more shallow absorption lines of the oxygen A band and hence causes a shortening of the effective light path estimated by the retrieval. Consequently, this leads to an overestimation of the CH₄ concentration, which explains the corresponding positive XCH₄ bias in Fig. 16. Here, the product is mostly sensitive to a distortion in the SWIR-3, where the most methane information is coming from, with biases up to 0.3%. The proxy retrieval is clearly sensitive to the ISRF distortion in the SWIR-1, especially for an ISRF squeeze with an additional XCH₄ bias up to -0.9% .³ The non-scattering retrievals underestimate both the CH₄ and CO₂ columns for a positive ISRF stretch, where the error sensitivity is stronger for CO₂ than for CH₄. Hence, the ratio approach overestimates XCH₄ for a stretched ISRF in the SWIR-1.

³ Note that the undistorted case for $\delta = 0$ has a bias of -0.153% , which has to be subtracted from the proxy biases in Fig. 16.

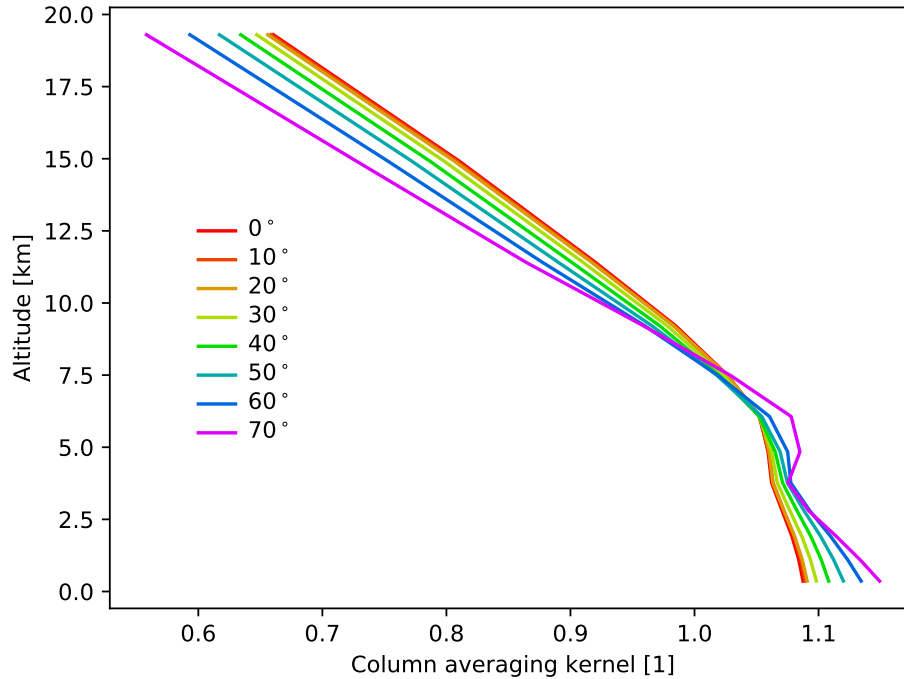


Figure 15: Averaging kernel as a function of solar zenith angle for a surface albedo of $(A_{\text{NIR}}, A_{\text{SWIR-1}}, A_{\text{SWIR-3}}) = (0.15, 0.1, 0.05)$.

7.1.4 Additive and multiplicative radiometric offset

To evaluate the XCH_4 sensitivity with respect to the radiometric calibration, we studied the effect of an additive and a multiplicative radiometric offset. First, we introduced a constant radiometric offset between $\pm 0.1\%$ of the continuum. Here, a positive additive offset fills absorption lines, and so one may expect an effect similar to a positive ISRF stretch. However, the results in Fig. 17 indicates that the proxy retrieval is remarkably resilient to a radiometric offset, meaning that the effect of a radiometric offset on both the CH_4 and CO_2 non-scattering retrievals cancel out in the proxy product. The physics-based retrieval shows the expected results with the largest effect for the SWIR-3.

Second, we applied a multiplicative offset between $\pm 2\%$ to the simulated measurement spectrum. This radiometric error can be mitigated well by fitting an effective Lambertian surface albedo. Figure 18 indicates that the proxy product is not sensitive at all to a multiplicative offset (biases $< 0.001\%$ after subtracting the overall bias of 0.153% of the proxy retrieval) because it can be effectively described by the surface albedo of the employed non-scattering radiative transfer model. However, the full-physics retrieval cannot fully compensate for the offset by an effective surface albedo because of the simulated interaction of atmospheric scattering with surface reflection. Here, the NIR-2 band is mostly affected by aerosols and Rayleigh scattering and so a multiplicative offset in the NIR-2 leads to the largest (but still small) biases of $< 0.25\%$.

7.1.5 Spectral alignment

Both the physics-based and proxy algorithms fit a constant spectral offset as an element of the state vector and so makes the retrieved XCH_4 insensitive to spectrally constant misalignment in the individual bands. Hence, in this section we evaluate the XCH_4 sensitivity with respect to a spectral stretch of the measurement in each band. To this end, we assume that the wavelength λ_t on which the measurements are truly simulated and λ_i , which is assigned to the measurement, are related by

$$\lambda_t = \lambda_i + S (\lambda_i - \lambda_a) \quad (106)$$

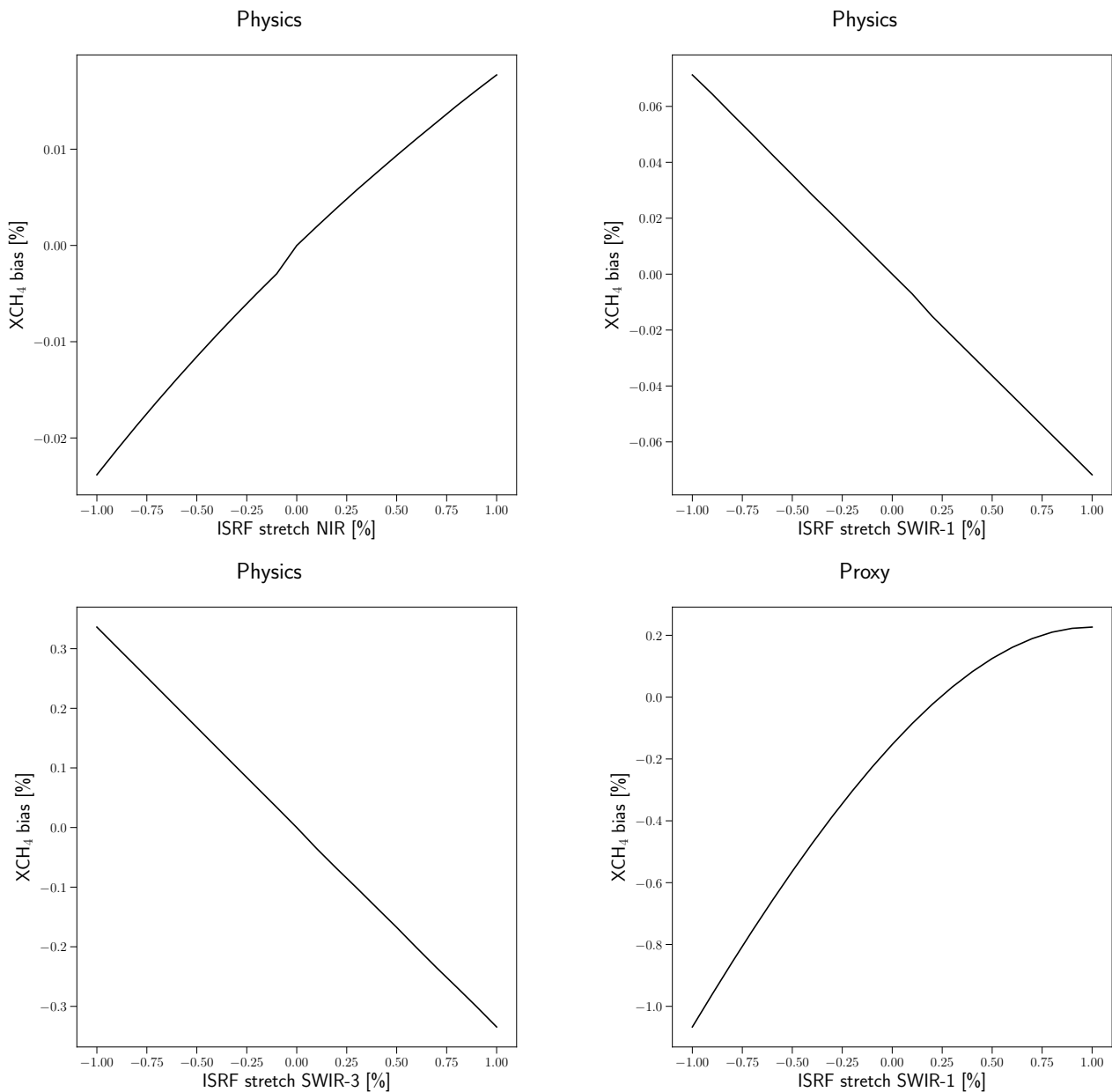


Figure 16: XCH₄ retrieval bias due to distortions of the FWHM of the ISRF: (upper left) physics-based retrieval ISRF distortion in the NIR-2 band, (upper right) same but for the ISRF distortion in the SWIR-1 band, (lower left) same but for the ISRF distortion in the SWIR-3 band, (lower right) same but for the proxy product. Simulations are performed for the reference scenario in Tab. 12.

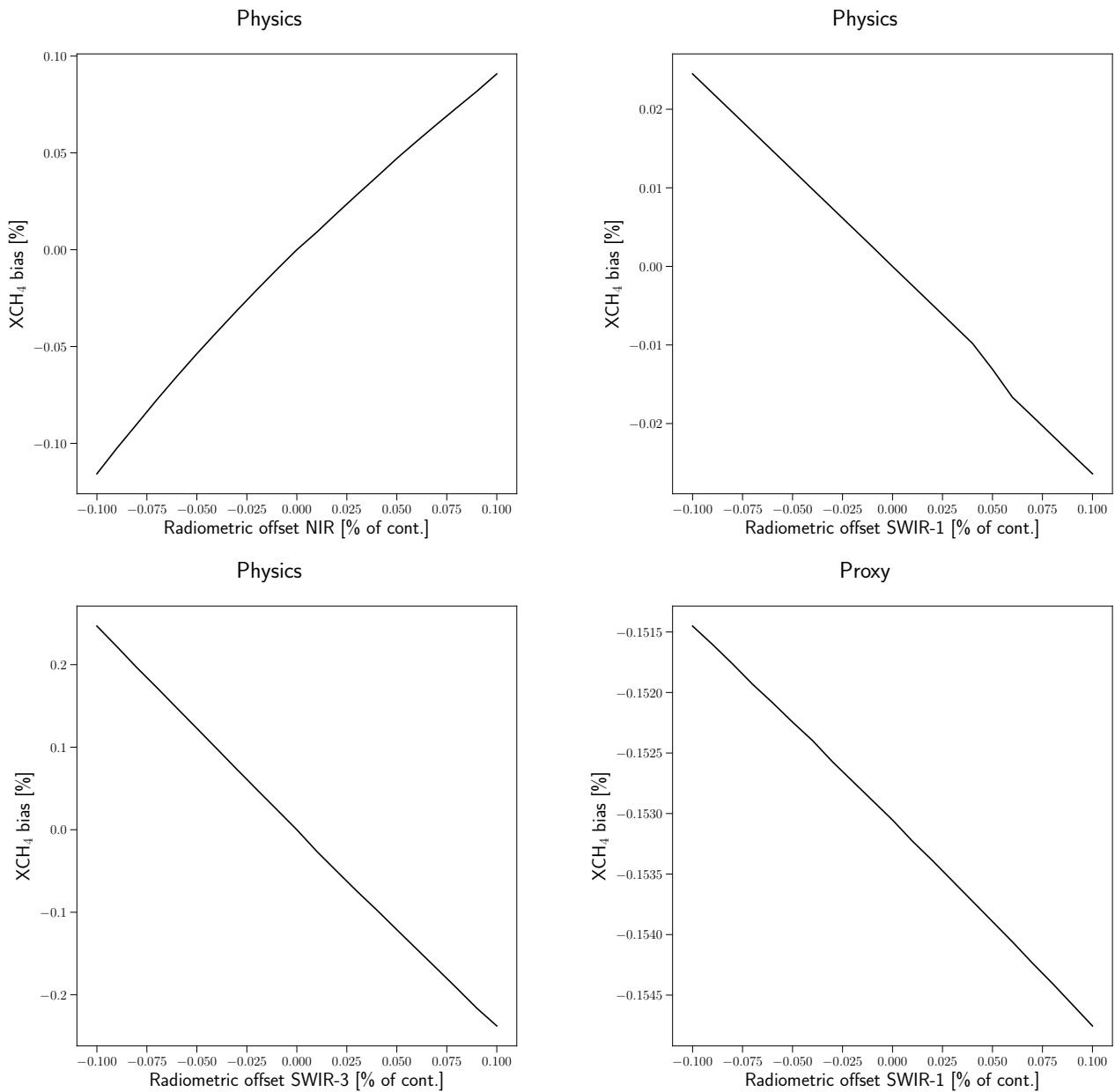


Figure 17: Same as Fig. 16 but for a varying additive radiometric offset.

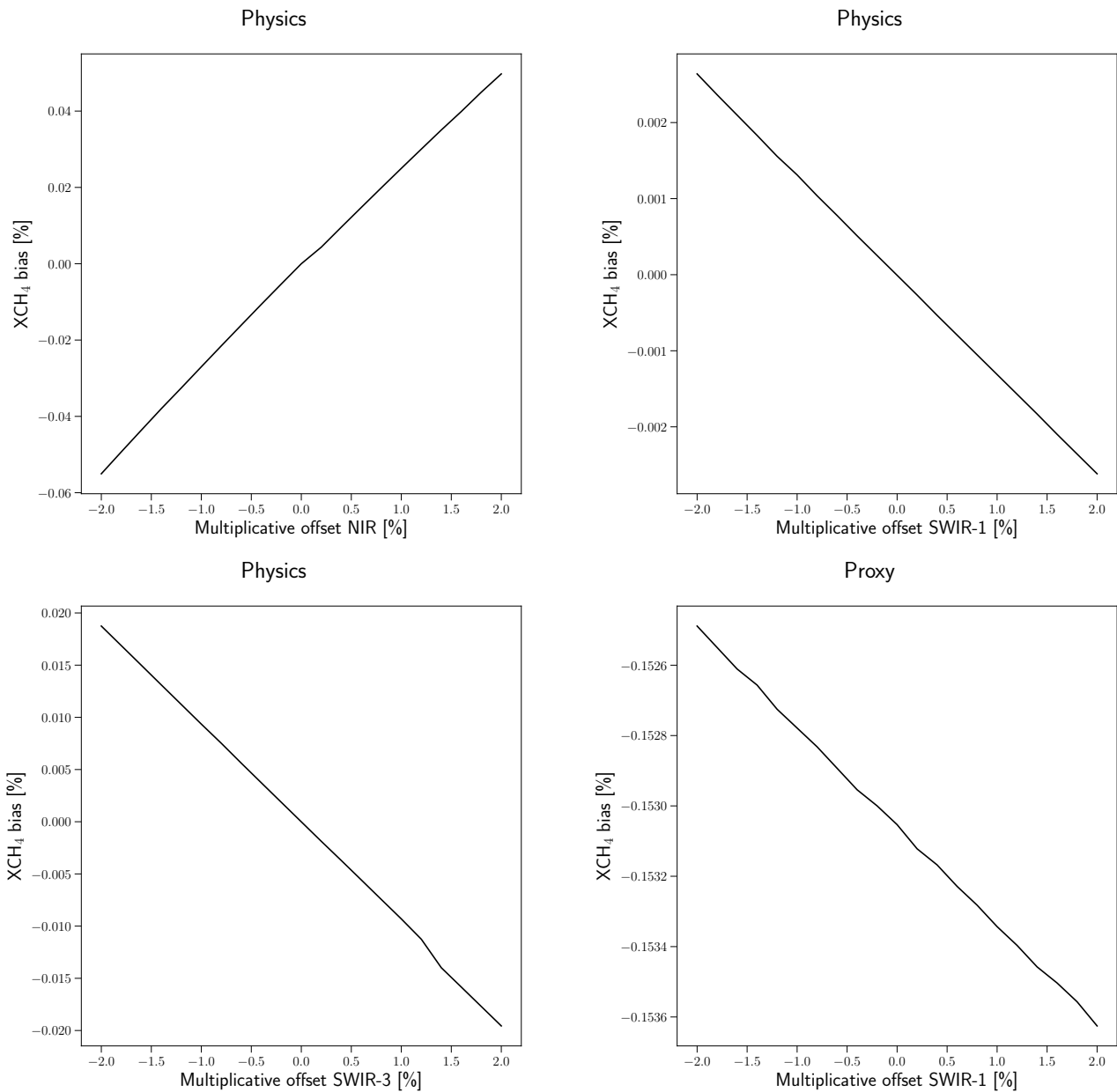


Figure 18: Same as Fig. 16 but for a varying multiplicative radiometric offset.

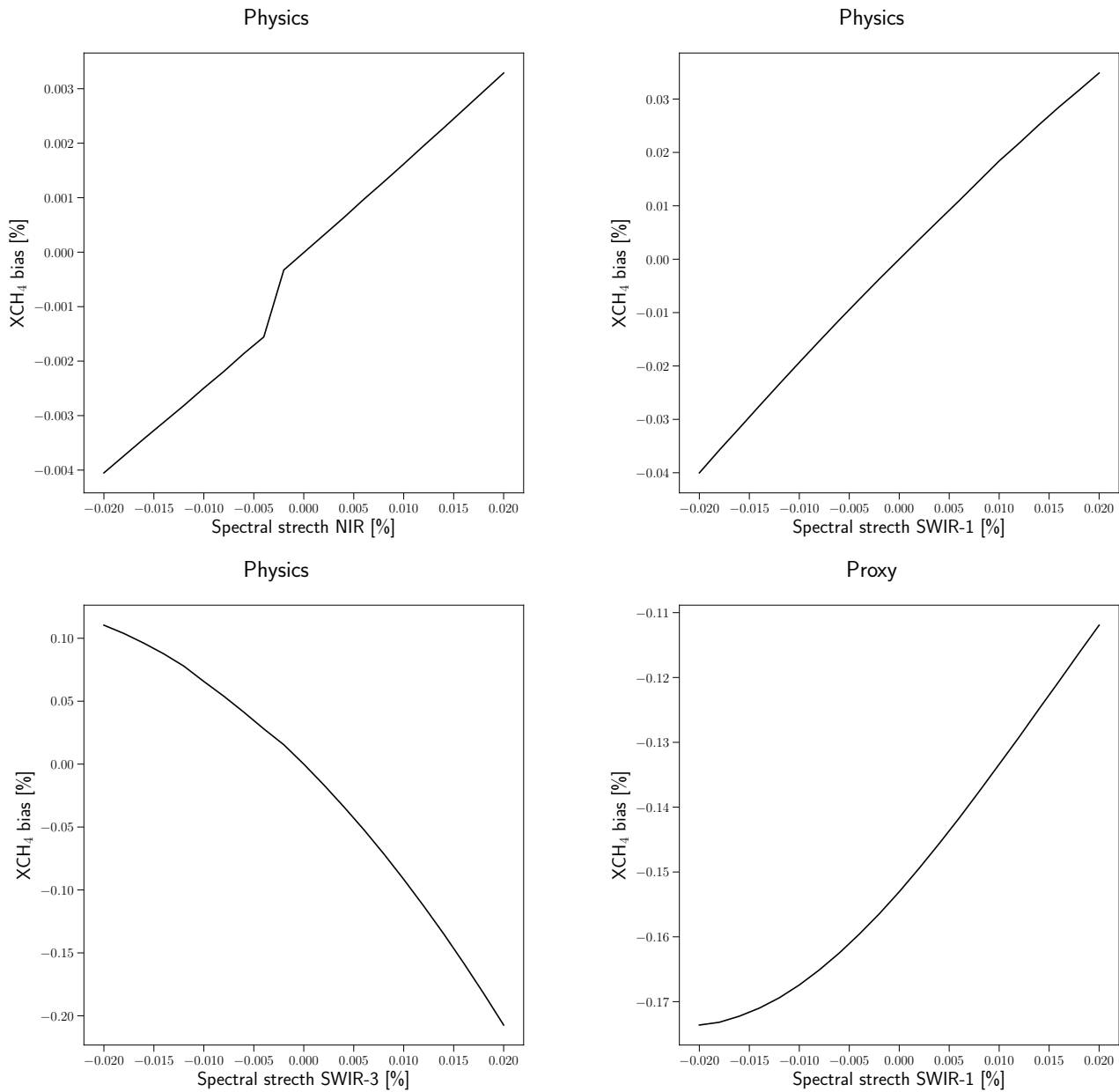




Figure 19: Same as Fig. 16 but as a function of the spectral stretch parameter S .

where λ_a is a reference wavelength, which only affects a constant zeroth order spectral offset, and S described a spectral stretch ($S > 0$) or squeeze ($S < 0$). The XCH₄ bias induced by this misalignment per spectral band is shown in Fig. 19 with S in the range $\pm 2 \times 10^{-4}$. Considering a spectral band width of 100 nm, this corresponds to a difference in the spectral misregistration between the left and right edge of the band of ± 0.02 nm.

Overall, the XCH₄ biases are small with a maximum absolute bias < 0.04 % for a spectral squeeze in the NIR-2 and SWIR-1, and < 0.2 % in the SWIR-3. It is noticeable that the XCH₄ biases for corresponding spectral misalignments in the SWIR-1 and SWIR-3 spectral bands are opposite in sign which indicates that the bias depends not only on the presence of spectral methane absorption but also on the strength, shape and relative position of different absorption lines in the spectrum.

 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 58/91
--	------------------------	--

7.1.6 Spectrometer stray light

Spectrometer stray light is the light in the instrument that was not intended by its design and so refers to the real propagation of light in an instrument in contrast to the ideal performance. Here, the stray light level is one of the most critical specifications of an instrument. Reasons for stray light can be manifold, e.g. diffuse reflection at surfaces, irregular groove spacing of a grating and reflection at lens surfaces. To define stray light in a mathematical sense, we started from a general description of the measurement process by

$$I_{ij}^{\text{meas}} = \iint dx d\lambda K_{ij}(x, \lambda) I_{\text{bl}}(x, \lambda) . \quad (107)$$

Here, I_{ij}^{meas} represents the simulated measurement by a two-dimensional detector at pixel (i, j) , where we assigned detector index i to the spectral dimension and index j to the spatial dimension across flight direction. $I_{\text{bl}}(x, \lambda)$ is the geophysical spectral radiance field at a swath position x and at wavelength λ . The instrument kernel $K_{ij}(x, \lambda)$ describes the optical transmission of light through the instrument. Simulating a radiometrically well-calibrated instrument, the instrument kernel must be normalized,

$$\iint dx d\lambda K_{ij}(x, \lambda) = 1 . \quad (108)$$

Analogously, we described the performance of an ideal instrument without any stray light by

$$I_{ij}^{\text{ideal}} = \iint dx d\lambda K_{ij}^{\text{ideal}}(x, \lambda) I_{\text{bl}}(x, \lambda) \quad (109)$$

with the normalized instrument kernel K_{ij}^{ideal} .

It is common to factorize K_{ij}^{ideal} by

$$K_{ij}^{\text{ideal}}(x, \lambda) = \varphi_j(x) s_i(\lambda) \quad (110)$$

with the point spread function φ_j and the ISRF s_i . Both φ and s are normalized to one. To simulate S5 measurements, we assumed a geometrical projection of the instrument entrance slit on the ground for φ , namely

$$\varphi_j(x) = \begin{cases} \frac{1}{\Delta x} & \text{for } x_j - 0.5\Delta x < x < x_j + 0.5\Delta x \\ 0 & \text{else} \end{cases} \quad (111)$$

Here x_j is the centre of the ground pixel and $\Delta x = 7$ km is the spatial sampling distance in swath direction.

We considered stray light I_{ij}^{stray} as an additive contribution to the actually measured signal

$$I_{ij}^{\text{meas}} = I_{ij}^{\text{ideal}} + I_{ij}^{\text{stray}} \quad (112)$$

with

$$I_{ij}^{\text{stray}} = \iint dx d\lambda K_{ij}^{\text{stray}}(x, \lambda) I_{\text{bl}}(x, \lambda) \quad (113)$$

and the stray light kernel K_{ij}^{stray} , respectively.

Obviously, the stray light contribution in Eq. (112) violates the assumption of a well calibrated instrument and so Eq. (112) needs to be renormalized. For this purpose, we considered the norm of the stray light kernel

$$\kappa = \iint dx d\lambda K_{ij}^{\text{stray}}(x, \lambda), \quad (114)$$

which is also known as the total internal scattering (TIS) of the spectrometer. Hence, the renormalization yields

$$I_{ij} = \frac{1}{1 + \kappa} (I_{ij}^{\text{ideal}} + I_{ij}^{\text{stray}}) . \quad (115)$$

To determine the stray light kernel K_{ij}^{stray} , we considered stray light measurements of the TROPOMI on-ground calibration. Figure 20 shows pronounced TROPOMI instrument specific features of purely spatial and spectral stray

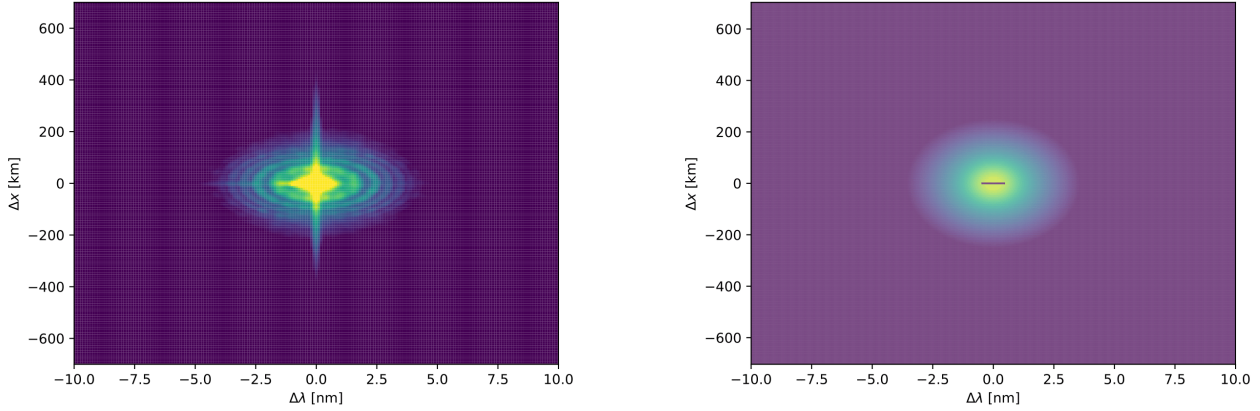


Figure 20: TROPOMI stray light kernel measured during the instrument on-ground calibration campaign (left). Stray light kernel used in this study (right).

light, probably due to the telescope and the immersed grating of the instrument, respectively. We omitted these features and fitted the function

$$S(\delta\lambda, \delta x) = \frac{A}{B + \left(\left(\frac{\delta\lambda}{\Delta\lambda} \right)^2 + \left(\frac{\delta x}{\Delta x} \right)^2 \right)^{\frac{g}{2}}} \quad (116)$$

to the remaining data. Here, $\delta\lambda$ and δx are the difference in wavelength and swath position between the detector pixel assigned wavelength and the wavelength and swath position of the radiance. $\Delta\lambda$ and Δx are the sampling distances of the detector in the spectral and spatial domains. For the fitted parameters $B = 130$ and $g = 2.23$, we obtained the best match to the measurements for TROPOMI. Subsequently, we scaled parameter A such that the TIS amounts to 0.2%. The corresponding stray light kernel is shown in the right panel of Fig. 20. Here, we assumed that the stray light contribution on the spectral measurements, which is typically in the order of several percent of the radiometric measurement, is significantly reduced due to a correction in the Level 0 to level 1 data processing. Furthermore, for simplicity reasons we assumed that the shape of the stray light kernel remains after correction.

Finally, we defined a spatial and spectral ensemble of line-by-line spectra at the top of the model atmosphere, which describes a realistic variation of scene brightness within the instrument swath. Here, we focused on a measurement ensemble of mixed cloudy and clear sky conditions for a data granule of $500 \times 1000 \text{ km}^2$ over Europe with a $1 \times 1 \text{ km}^2$ pixel size using MODIS Aqua products of cloud optical thickness, cloud fraction and cloud top height collocated to the granule. The cloud coverage is depicted in Fig. 21. For the simulations, we considered periodic boundary conditions at the swath edges. Meteorological data like pressure and water vapor concentration is adapted from ECMWF output, CO , CH_4 and CO_2 abundances are taken from TM5 and Carbontracker simulations. All data are collocated on the granule grid using the multi-instrument preprocessor MIPrep [RD66]. For the entire ensemble, we presumed a constant low surface albedo of 0.10 in the NIR-2 and 0.05 in the SWIR-1 and SWIR-3 and a boundary layer aerosol of an optical thickness $\tau = 0.1$. The surface elevation is adapted from a digital elevation map [RD36, RD67]. The ensemble choices maximize the contrast between clouds and clear sky scenes. Additionally, we analyzed a corresponding data ensemble of purely clear sky measurements with little radiometric variation within a swath.

To simulated S5 measurements, we employed the independent pixel approximation to account for fractional cloud coverage for the $1 \times 1 \text{ km}^2$ pixels and averaged seven spectra in flight direction. Here, XCH_4 can only be inferred from clear sky pixels, using calculated spectra without and with stray light accordingly to Eqs. (109) and (113). This leads to an estimate of the stray light induced error on XCH_4 per spectral band.

Figure 22 shows that stray light in the SWIR-1 and SWIR-3 spectral band causes a negative bias on the retrieved XCH_4 because methane absorption lines are filled up by stray light. In the NIR-2, however, the opposite is true due to the reasons explained in Sect. 7.1.3. Hence, the effects of stray light in the NIR and SWIR spectral range cancel in part, depending on band and retrieval settings. Obviously, the proxy retrieval is only sensitive to stray light in the SWIR-1 band and retrievals of CH_4 and CO_2 are affected in a similar way. Therefore, part of the stray light induced error is compensated in the proxy ratio as illustrated in Fig. 22.

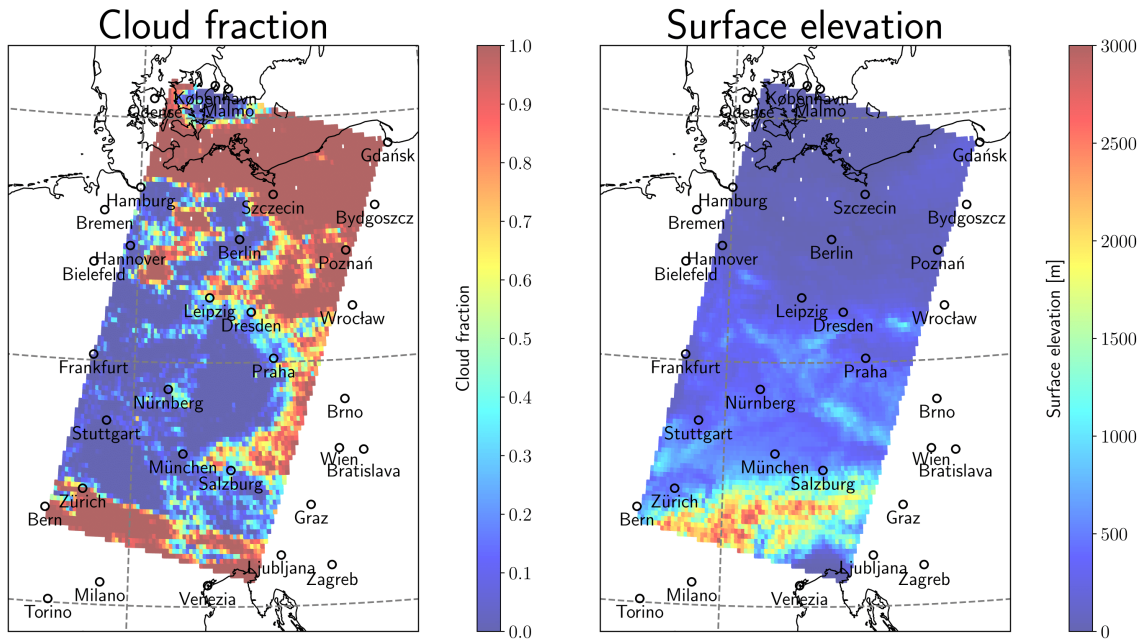


Figure 21: (Left panel) MODIS cloud fraction for the stray light data granule, (right panel) surface elevation.

band	mean bias [ppb]	standard deviation [ppb]	data yield [%]
NIR-2	1.73 (0.96)	2.03 (1.79)	100 (100)
SWIR-1	-1.93 (-0.86)	1.35 (1.28)	69 (85)
SWIR-3	4.90 (-2.95)	1.95 (1.00)	100 (100)
proxy	-1.71 (-0.90)	1.1 (0.19)	100 (100)

Table 14: Mean XCH₄ bias, standard deviation and data yield for the stray light analysis using the ensemble of Fig. 21. Corresponding numbers for a purely clear sky ensemble are given in parenthesis.

Figure 23 shows the corresponding distribution function of the XCH₄ biases with a non-Gaussian shape and mean biases, standard deviations and the data yields summarized in Tab. 14. Overall, for the assumed TIS of 0.2% the biases are well within the requirement. Here, the data quality of the physics-based retrieval is clearly affected by stray light in the SWIR-3 band with a mean bias of about 5 ppb for the cloudy scenario, where SWIR-1 stray light reduces mainly the data yield with about 30% of non convergence. Figure 23 also depicts the bias distributions for the corresponding clear sky ensemble. Obviously, the effect of stray light is significantly reduced, since the surrounding of a ground scene in swath direction is less bright for the clear sky ensemble than for the cloudy ensemble and so less stray light is received by the corresponding detector pixel. Overall, this halves the stray light effect. Finally, we emphasize that the stray light analysis in this section is based on the assumption of a very effective stray light correction with a remaining TIS of 0.2%. The bias effect for larger stray light contributions can be estimated in first order by linear scaling of the derived errors.

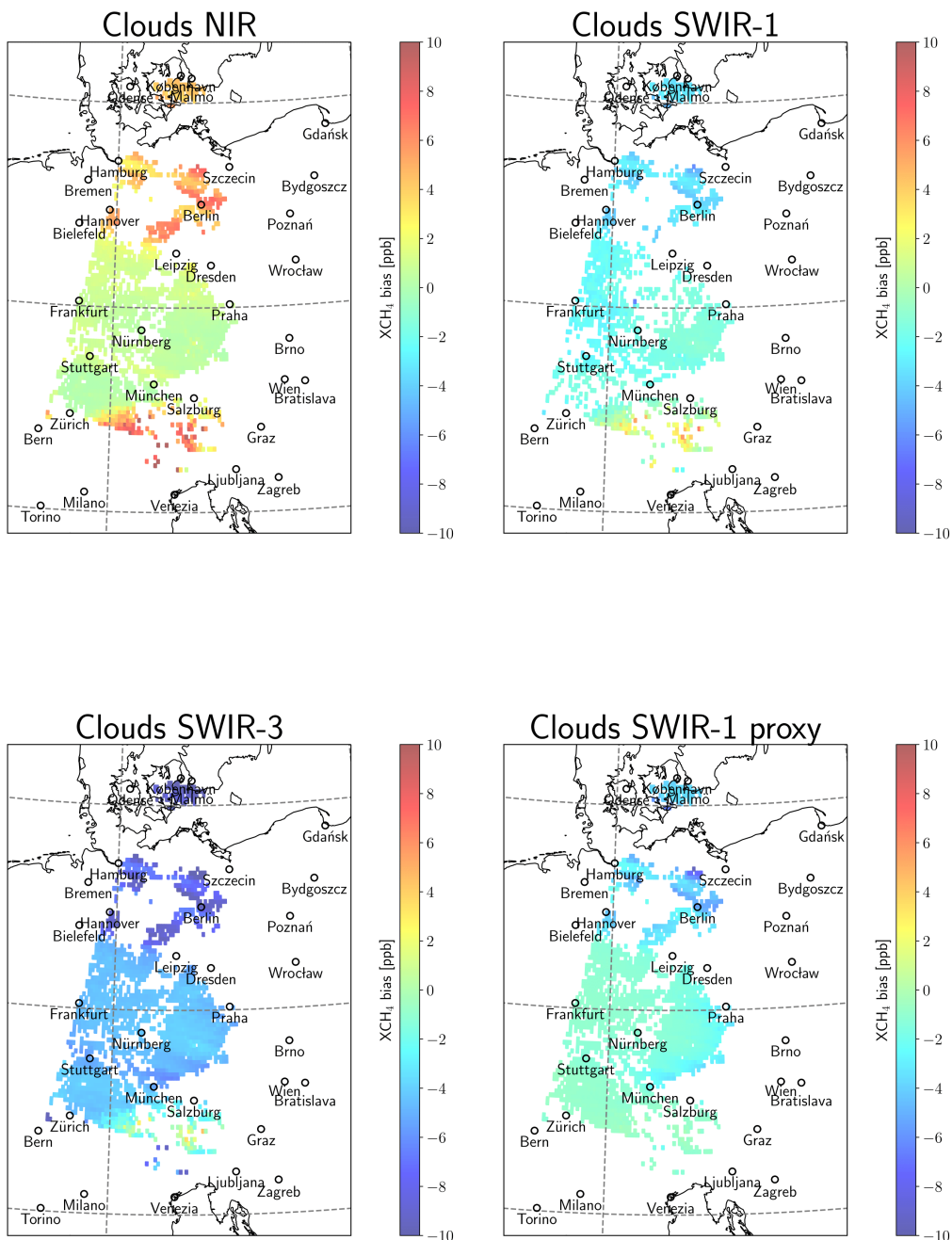


Figure 22: Retrieval bias of the physics-based and proxy XCH₄ product for stray light in the individual bands.

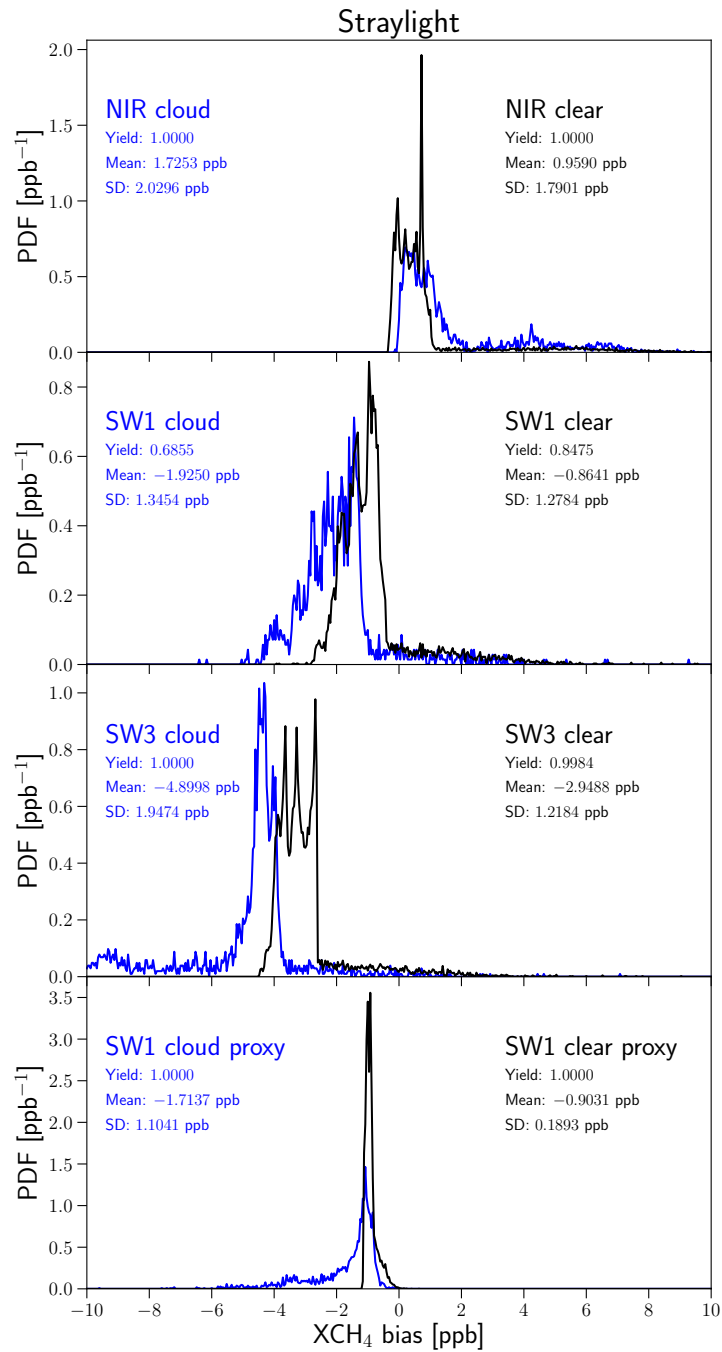


Figure 23: Bias distribution function due to stray light in the different bands of the physics-based retrieval and the proxy retrieval.

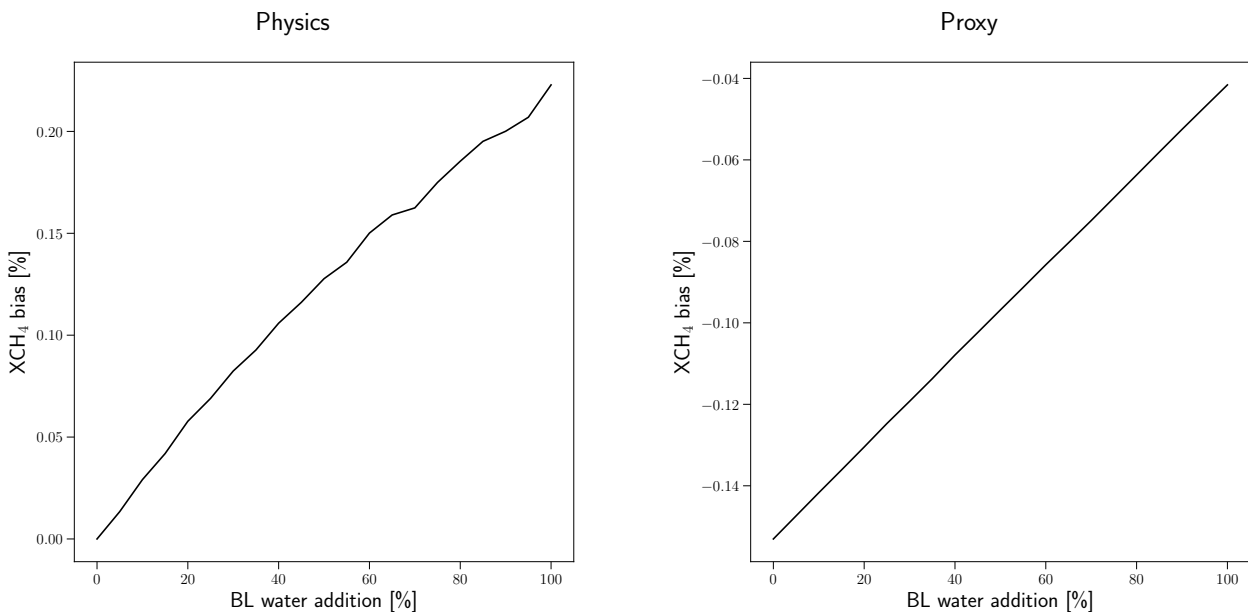


Figure 24: XCH₄ retrieval bias of the physics-based and proxy products with varying added H₂O in the boundary layer.

7.2 Sensitivity to atmospheric input

7.2.1 Profiles of atmospheric trace gases

To evaluate the bias sensitivity of the CH₄ retrieval to the vertical profiles of H₂O, CH₄ or CO₂, we simulated spectra of the reference scenario in Sect. 7.1.1, where we enhanced the trace gas concentrations in the planetary boundary layer. Because the retrievals are performed with the a-priori profiles equal to the reference scenario, the induced inconsistency in the profile shapes causes systematic retrieval errors. In our analysis of these biases, we enhanced the boundary layer concentrations of H₂O, CH₄ and CO₂ by up to 100 %, 20 % and 10 % respectively, which increase the corresponding total columns up to 59.48 %, 4.37 % and 2.12 % for the three trace gases respectively.

Figures 24, 25, and 26 show the corresponding retrieval sensitivity. Generally, we see a similar sensitivity of the physics-based and proxy CH₄ products, with biases all well within the requirement. An exception is the perturbation of the CO₂ profile, where the physics-based retrieval is hardly sensitive to. This is because the CO₂ absorption lines are predominantly outside the spectral fitting windows of the retrieval. The proxy retrieval, however, is fully dependent on an accurate prior knowledge of CO₂ and therefore has a bias similar to the a-priori error in the total column of CO₂. This important error source of the proxy retrievals is discussed in detail by Schepers et al., 2012 [RD16] for the GOSAT proxy product.

7.2.2 Temperature offset

Next, we shifted the entire temperature profile of the atmospheric input by $\pm 2^\circ\text{C}$, while the retrieval is performed with the original temperature profile of the reference scenario. As the absorption cross sections are temperature-dependent, a wrong knowledge of the atmospheric temperature causes a retrieval bias as depicted in Fig. 27.

Here, the proxy retrieval shows a linear sensitivity to the temperature offset, caused by a different temperature-dependence of the CH₄ and CO₂ cross sections. The physics-based retrieval is most resilient towards temperature shifts with biases mostly below 0.1 %. However, these retrievals suffer from stability issues, requiring a lot more iterations to converge (up to $N_{\text{iter}} = 70$) than usual ($N_{\text{iter}} = 10 - 15$). Due to a maximum number of iterations in the operational processor, a few retrievals do not provide valid output. This iteration limit was ignored in order to show the results in Fig. 27.

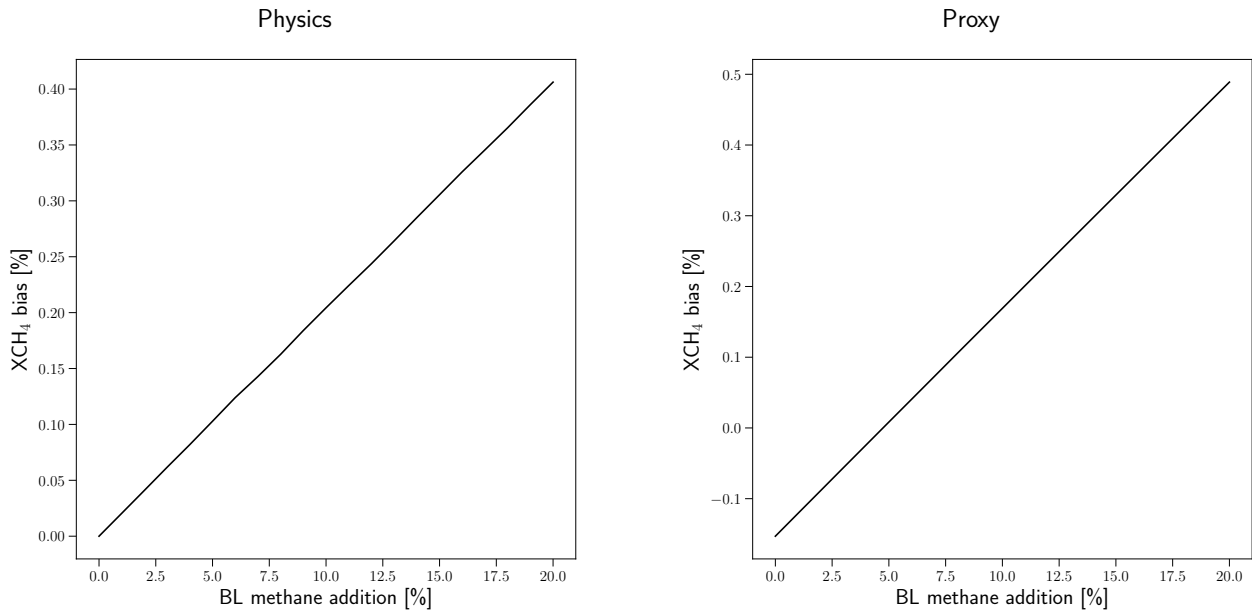


Figure 25: Same as Fig. 24 but with varying added CH₄ in the boundary layer.

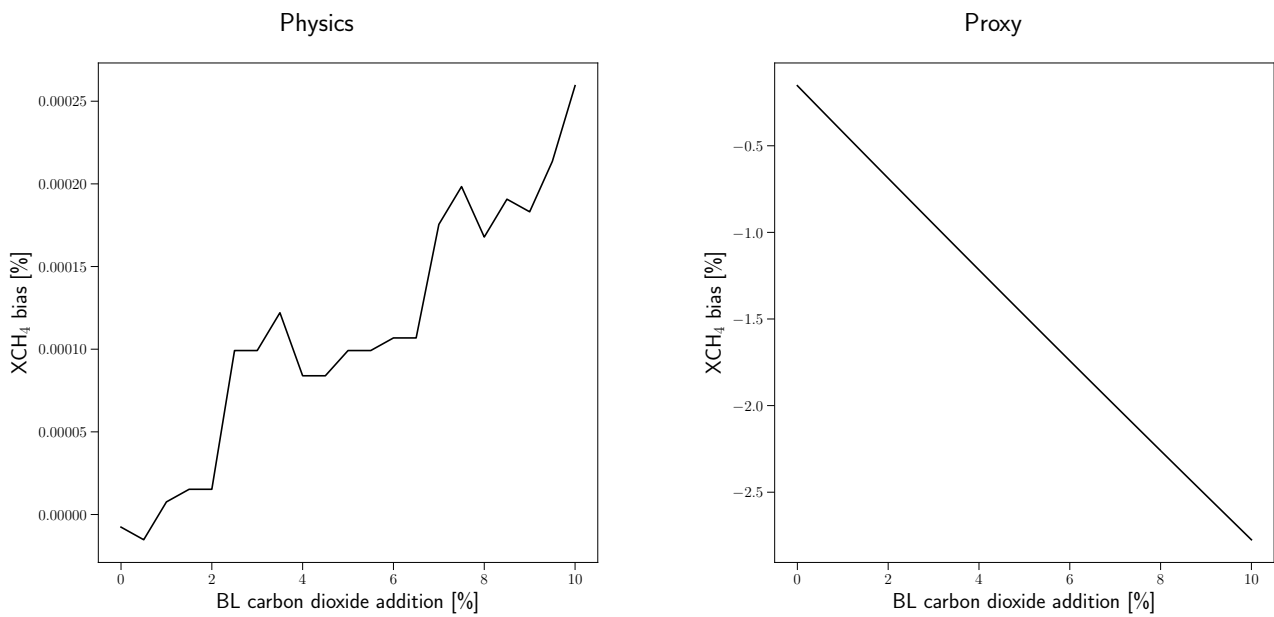


Figure 26: Same as Fig. 24 but with varying added CO₂ in the boundary layer.

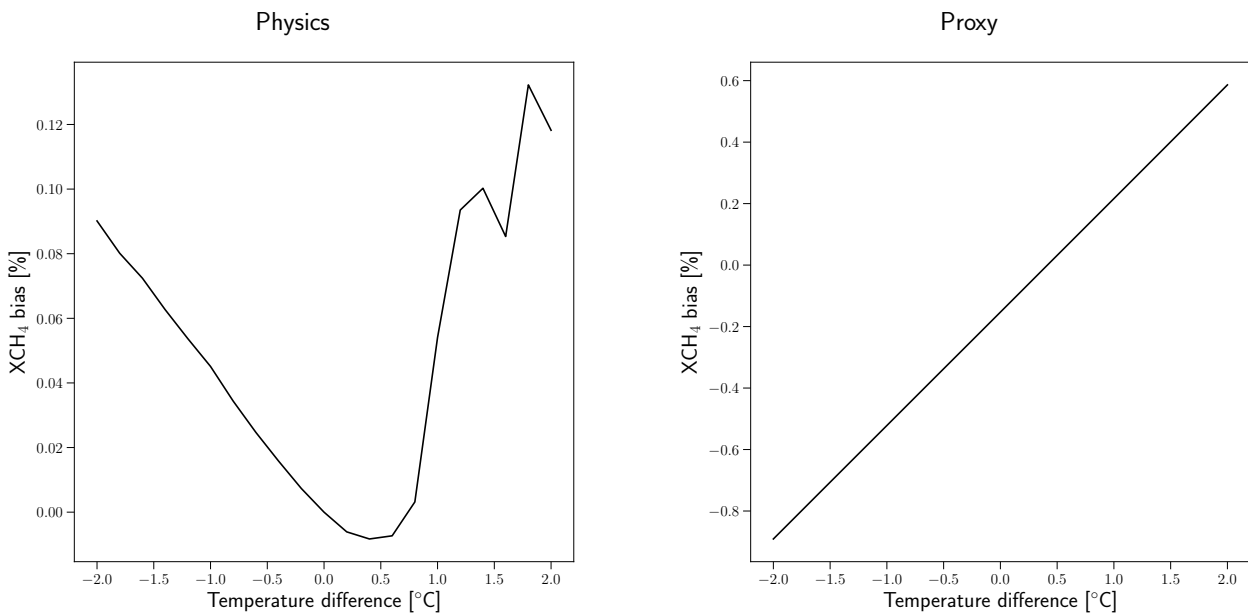


Figure 27: Retrieval bias of the physics-based and proxy XCH₄ products with varying shifts of the atmospheric temperature.

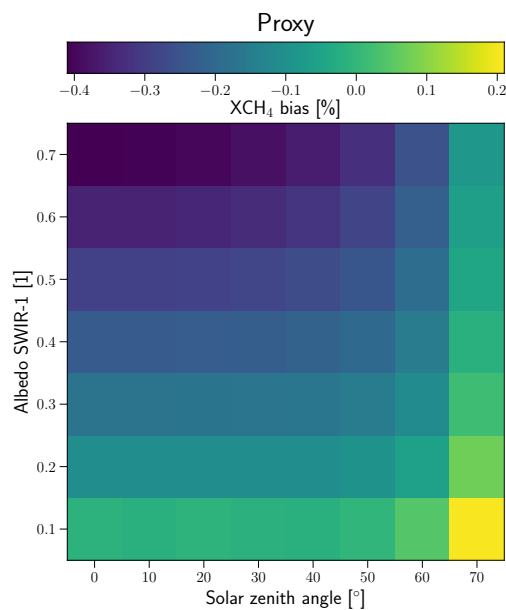


Figure 28: Retrieval bias of the physics-based and proxy XCH₄ product with varying surface albedo and solar zenith angle.

7.2.3 Aerosol induced errors

Finally, we investigated XCH₄ biases induced by atmospheric aerosols. First, we evaluated the error due to aerosols for the generic ensemble of Sect. 7.1.1, where aerosols are described in a consistent manner in the measurement simulation as well as in the physics-based retrieval. This obviously leads to a vanishing bias of the physics-based XCH₄ product, whereas the proxy CH₄ product shows a clear bias as indicated in Fig. 28. The non-scattering retrievals of both the CO₂ and CH₄ column are subject to aerosol induced errors that do not fully cancel out in the proxy ratio. Here the remaining errors on XCH₄ vary between -0.4% and +0.2% depending on solar zenith angle and surface albedo.

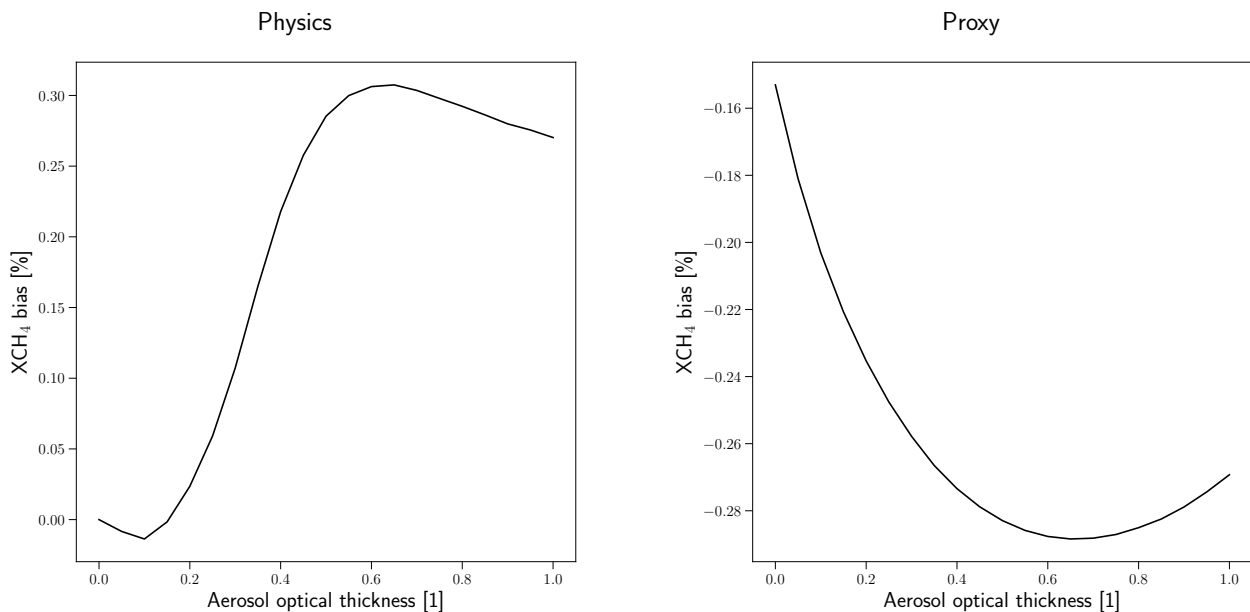


Figure 29: Retrieval bias of the physics-based and proxy XCH₄ product with varying boundary layer aerosol optical depth at 765 nm.

To test the sensitivity of the physics-based retrieval in more detail, we added a second layer of aerosols besides the background aerosols that already exists in the reference scenario. The layer is centered at 2 km altitude, has a geometric thickness of 500 m with varying optical thickness. The retrieval forward model adjusts only a single aerosol layer and so is not fully consistent with the multi-layer aerosol model of the measurement simulation. Figure 29 illustrates the induced biases, where the aerosol induced error of both products are similar and well within the required accuracy.

Finally, we evaluated the sensitivity of the CH₄ algorithm on a global ensemble of simulated spectra consisting of land-only, clear sky scenes on a 2.8° × 2.8° latitude × longitude grid. This ensemble is to a large extent identical to the one used by Butz et al. 2012 [RD68]. It contains realistic aerosol and cirrus loaded scenes for four days, one per season. The treatment of aerosols and cirrus in the simulations are far more complex than in the retrieval forward model, where only one effective aerosol type is considered as explained in Sect. 6.8.1. For the simulations, the aerosol physical properties and vertical distributions are derived from the global aerosol model ECHAM5-HAM [RD52] for five different chemical species and on a superposition of 7 log-normal size distributions. The aerosol optical thickness is derived from MODIS observations [RD69]. Furthermore, the simulations contain cirrus with optical thickness and vertical distribution based on CALIOP measurements [RD70]. Finally, the surface albedo in the NIR-2 is taken from the MODIS land albedo product in the 841–876 nm channel and the surface albedo in the SWIR-1 is taken from the MODIS land albedo product in the 1628–1652 nm channel. For the albedo in the SWIR-3, the SCIAMACHY surface albedo product at 2350 nm is used [RD71]. For simplicity, we assumed a constant surface albedo within the bands in our simulations. We refer to Figs. 2 and 3 of Butz et al. 2012 [RD68] for the geographical distribution of the total optical thickness and surface albedo, respectively, used in the simulations. The measurements are simulated for the nadir viewing direction and a solar zenith angle that is representative for an S5P overpass.

Figure 30 depicts the global bias distribution for the four seasons and Fig. 31 reflects the corresponding bias distribution. The physics retrieval shows biases on regional scales above bright surfaces which reaches 3 % in April with elevated aerosol layers. Also for dark surfaces, biases can be substantial leading to mean biases of up to 14 ppb with standard deviations of several percent. As part of an operational data processing, retrievals with these large biases are filtered out a posteriori using appropriate filter criteria. This improves the bias characteristics at the cost of data coverage of the product.

The proxy retrieval seems to outperform the physics-based retrieval with little regional and global biases. However, the weakness of the proxy retrieval are not covered by the analysis of the global ensemble, which is its dependence on the a-priori knowledge of CO₂ and the enhanced noise propagation. Overall both the physics-based and proxy retrieval are

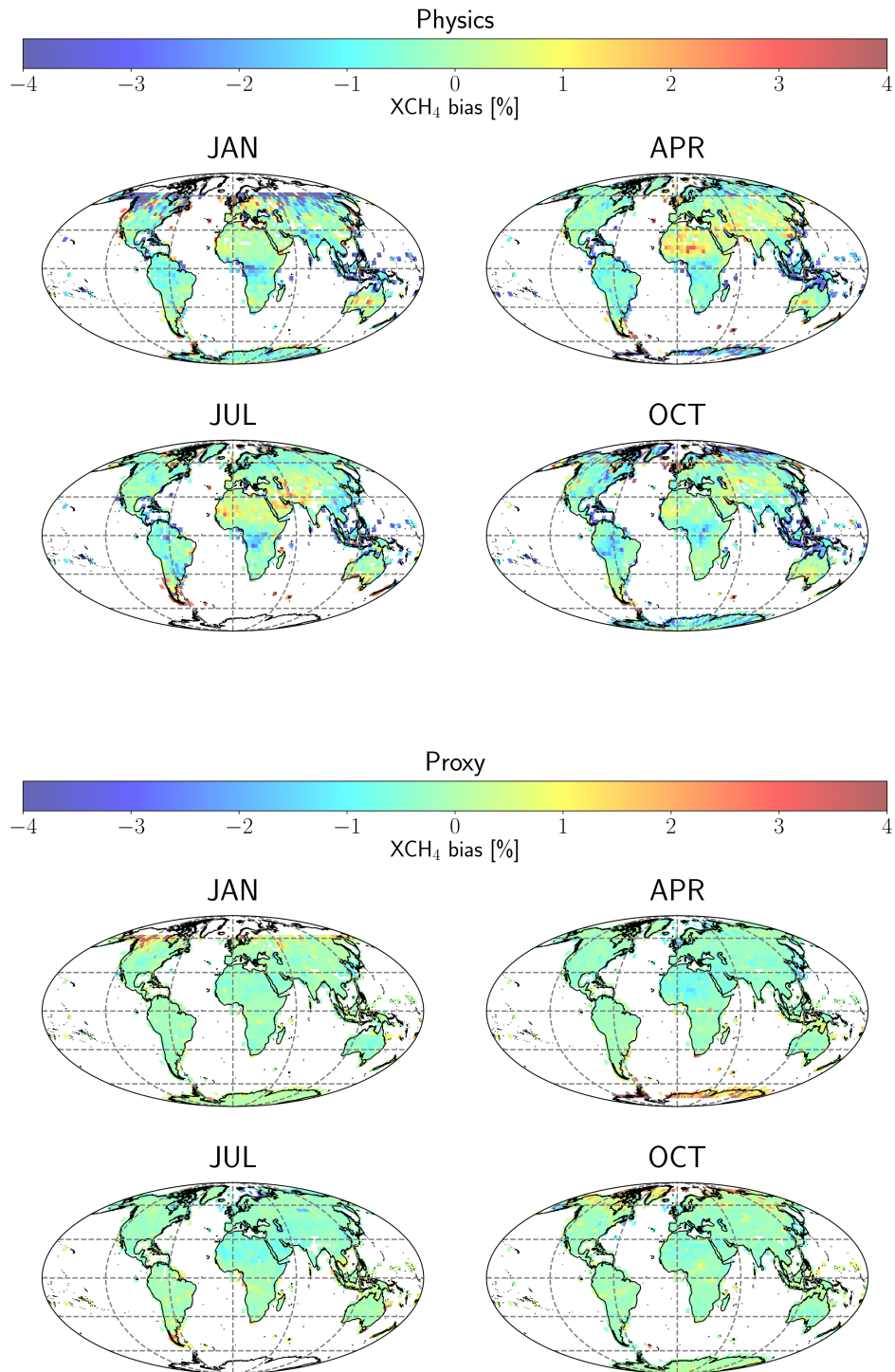


Figure 30: Retrieval bias of the physics-based and proxy XCH₄ products for the global ensemble. No filters apart from the convergence check are applied.

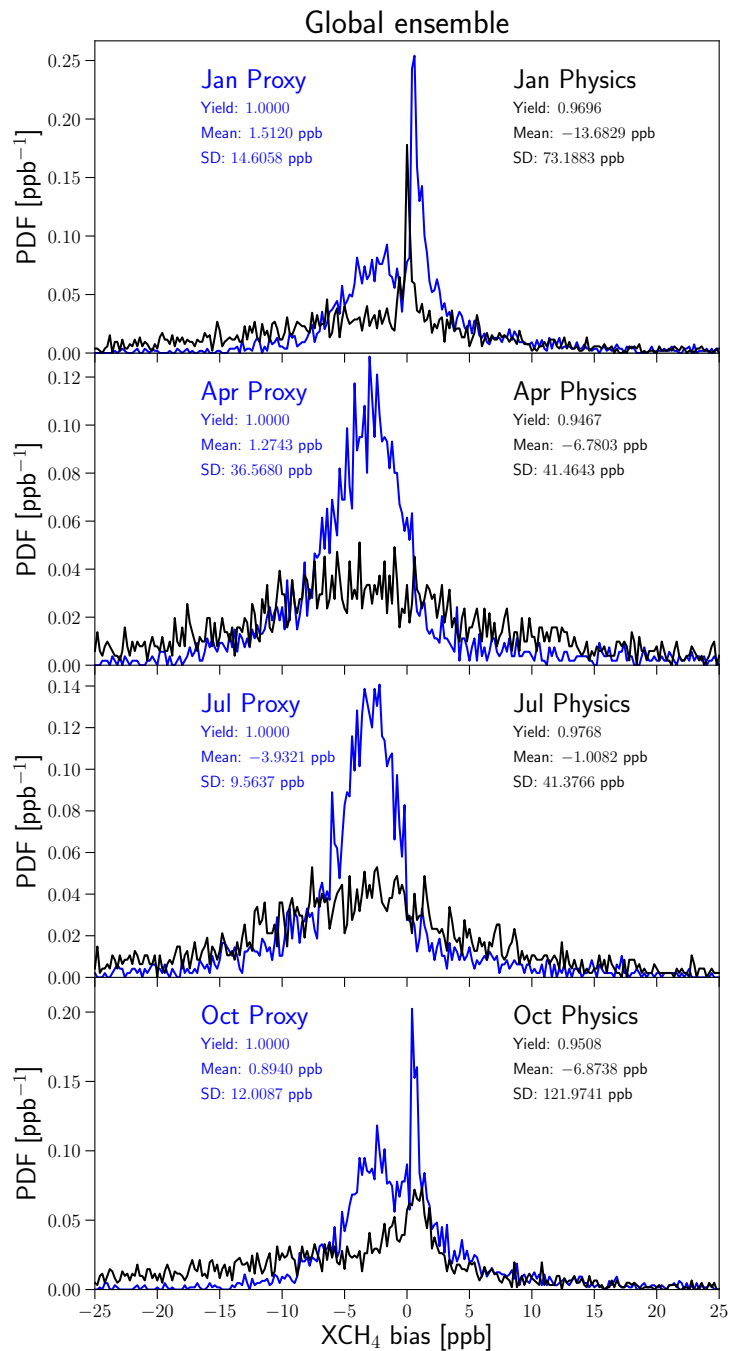


Figure 31: Bias distribution function of Fig. 30 for the physics-based retrieval and the proxy retrieval

complementary in their performance characteristics, which justifies that both are included in the S5 methane data product.

8 Validation

8.1 Ground based

8.1.1 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 CO₂ and CH₄ [RD42, RD72]. TCCON is a network of inter-calibrated ground-based Fourier transform spectrometers that measure the absorption in the NIR/SWIR of direct sunlight by trace gas species such as CO₂, CH₄, CO, HDO, etc. These measurements are thus much less influenced by atmospheric scattering by cirrus and aerosols than satellite observations of backscattered/reflected sunlight. Also for S5 retrievals of XCH₄ the TCCON network will be the prime source of validation data. TCCON XCH₄ measurements have been calibrated and validated against the WMO-standard of in-situ measurements using dedicated aircraft campaigns of XCH₄ profiles and their resulting accuracy have been estimated to 0.4% (2 σ value) [RD42]. Table 15 shows an overview of the existing and some expected future TCCON stations. An important limitation of the TCCON network for validation of satellite retrievals of greenhouse gas concentrations is the limited albedo range that is covered by the TCCON stations. Figure 32 shows the surface albedo range at 2300 nm for the different TCCON stations. It can be seen that surface albedos > 0.20 are not covered by the TCCON network except for the Caltech and Four Corners stations. However, the latter 2 stations may not be very well suited for validation purposes because of local sources (Caltech) or mountainous area (Four Corners). Globally, over 20% of the land surfaces have an albedo >0.2, up to 0.7. Therefore, there is a strong need for validation (TCCON) stations that cover higher surface albedos [RD73, RD30]. This is very important as these scenes pose a strong challenge for the XCH₄ retrieval algorithm because of the large sensitivity to aerosols. Also TCCON stations at the challenging very low albedo conditions are lacking to some degree. An interesting development in this respect is the development of somewhat cheaper spectrometers complementary to the TCCON high spectral resolution FTIRs [RD74, RD75]. Clearly, for these systems to be useful for S5P XCH₄ validation their accuracy needs to be high enough. But they offer the potential of deployment in the now poorly sampled areas including high (and low) albedo regions. Finally, it is important to note that a delivery of TCCON data with 3 month after sensing is essential for timely validation of the S5 CH₄ data product.

Table 15: Overview of TCCON stations (information from 2017)

station	latitude [degree]	longitude [degree]	height [m]	period	comment
Ammyeondo	36.5382	126.3311	30	since Aug 2014	
Arrival Heights	-77.83	166.663	250	-	too southerly
Ascension Island	-7.9165	-14.3325	10	since May 2012	
Burgos	18.5325	120.6496	35	since Mar. 2017	
Bialystok	53.23	23.025	180	since Mar. 2009	
Bremen	53.1037	8.84952	30	since Jan. 2005	
CalTech	34.136	-118.127	230	since Sept 2012	
Darwin	-12.4245	130.892	30	since Aug. 2005	
Dryden	34.95	-117.83	692	since July 2013	from JPL
Eureka	80.05	-86.42	610	since July 2010	too northerly
Four Corners	36.7975	-108.48	1643	since Mar. 2011	new, to be moved temporarily to Man-aus
Garmisch	47.476	11.063	743	since July 2007	
Lzaña	28.3	-16.5	2370	since May 2007	too high
Karlsruhe	49.1	8.438	111	since Apr 2010	

Continued on next page

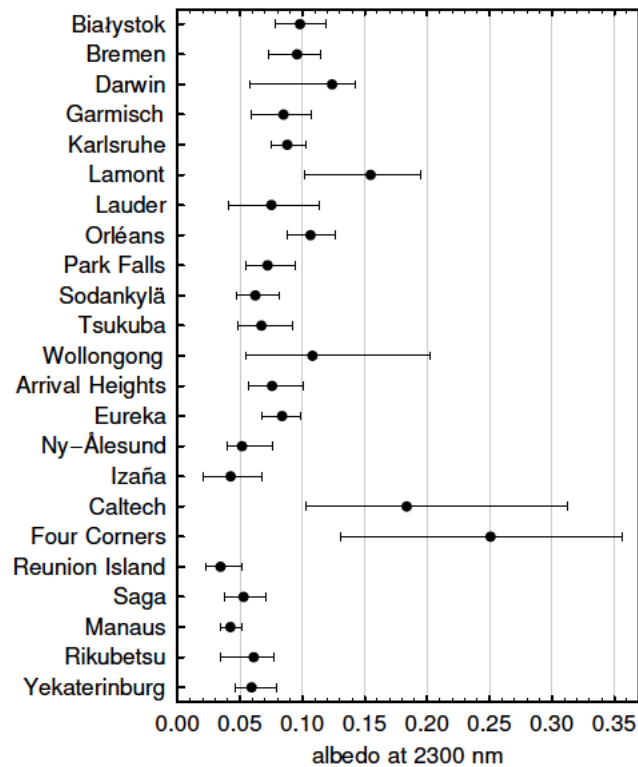




Figure 32: Range of the central 80% of the land albedo values in an area of about $500 \times 500 \text{ km}^2$ around each TCCON station, except Ascension Island, for which no land data is available. A dot indicates the median value. Sea values are excluded, but they may contaminate coastal land.

Table 15 – Continued from previous page

station	latitude [degree]	longitude [degree]	height [m]	period	comment
Lamont	36.604	-97.486	320	since July 2008	
Lauder	-45.038	169.684	370	since June 2004	
Manaus	-3.1	-60.02	36	future	from Four Corners
Ny Ålesund	78.9232	11.923	20	since Mar. 2005	too northerly
Orléans	47.97	2.113	130	since Aug. 2009	
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	new
Rikubetsu	43.4567	143.766	380	since Nov 2013	
Saga	33.241	130.288	7	since June 2011	new
Sodankylä	67.368	26.633	179	since Feb. 2009	
Tsukuba	36.0513	140.122	31	since Dec. 2008	
Wollongong	-34.406	150.879	30	since June 2008	
Zugspitze	47.42	10.98	2960	2012	too high
Yekaterinburg	57.0383	59.545	300	2010	


 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 71/91
--	------------------------	--

8.1.2 In Situ Measurements

In situ measurements of the full vertical CH₄ profile will be extremely useful as they allow the column averaging kernel to be applied and hence the retrieved XCH₄ from S5 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 [RD76]. The 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. The 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 CO₂ and CH₄ (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 CH₄ [RD42]. Another possible application of AirCore would be to use it in areas where there are no TCCON validation sites.

8.2 Satellite Intercomparison

GOSAT-2 will be launched in 2018 and will provide column integrated XCH₄ with sensitivity to the ground making use of the 1.6 μm channel. GOSAT retrieval can make use of both the proxy method as well as the physics based method such as described here (see also [RD16]). The RemoTeC algorithm, is continuously being applied to level-1 data set of the precursor mission GOSAT and will be used also for scientific data interpretation of its successor. Comparing both GOSAT mission, GOSAT-2 has the advantage that it also includes a spectral band around 2300 nm, like S5P, and hence will be very interesting to validate/compare to S5.

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 72/91
---	------------------------	--

9 Feasibility

The RemoTeC algorithm is been used successfully to process GOSAT and OCO-2 data (e.g. [RD64] and [RD77]). Moreover, it is implemented in the Sentinel 5 Precursor operational framework as well as in the SRON scientific data processing framework using a GRID processing infrastructure, which is a distributed infrastructure of compute clusters and storage systems. During the preparation phase of the S5P mission, both framework implementations were successfully tested, where processing of 300 S5-P test orbits on 1000 cores of the SURFsara GRID [RD78] showed good scalability of the software for the employed infrastructure. Here MPI is used for parallelization on single processing nodes. Overall, the RemoTeC heritage demonstrates the readiness level and maturity of the algorithm and software implementation. The RemoTeC software is written in Fortran 90 with a well-defined interface to a framework written in C.

9.1 Estimated Computational Effort

For the retrieval of methane columns, the runtime performance of the RemoTeC algorithm was tested extensively for the operational processor of the Sentinel 5 mission using synthetic measurements with varying atmospheric, surface and viewing conditions. Here, the algorithm is based on the S5 NIR and SWIR-3 measurements. For the estimate of the computation effort of RemoTeC-S5, we considered two different processing lines: (1) retrieval of the proxy product only, and (2) retrieval of the physics-based three band data product in combination with the proxy product. The computation time is estimated as the averaged single sounding processing time for the global ensemble described in Sect. 7.2.3.

For a single sounding and using the GNU Fortran compiler with optimization flag O3, a typical S5 CH₄ retrieval takes about 0.82 seconds for the proxy retrieval and 11.4 seconds for the combined proxy and physics-based retrieval, respectively, on a single state-of-the-art PC (HP Z230 SFF D1P35AV with a Intel(R) Xeon(R) CPU E3-1226 v3 @ 3.30GHz processor and a cache size of 8192 KB). The initialization of the software requires 10 seconds. The hardware computer benchmark for floating point performance to about SPECfp2006 = 60, which is estimated from the specification of the similar processor Intel(R) Xeon(R) CPU E3-1230 v3 with SPECfp2006 =72.9 scaled by a factor 0.83 to account for the performance difference between both processors (see also <https://www.spec.org/cpu2006/results/res2014q2/cpu2006-20140407-29279.pdf>). Assuming the swath sampling of the S5 instruments, and orbit consists of about 700000 measurements and takes 6090 seconds. to data filtering, the number of measurements for which a methane retrieval is performed is significantly less:

- Only measurements over land are used, which mean about 30% of the remaining data.
- Only cloud free measurements (fraction of confidently clear pixels) are considered. This leads to a reduction of 80%. (Tests with the RAL Suomi NPP processor on real VIIRS data show that at a global scale 22% of the S5 pixels have a fraction of confidently clear pixels).

Thus, on average 7 measurements per second will be performed that are useful for methane retrieval. To keep track with the data flow of useful S5 measurements, we estimate that roughly 90 and 7 threads are needed for the two processing lines. The required RAM memory is about 5 GB. It is important to note that the estimated computation time does not include any computation cost for the initialization, which depends on the choice of data granules to be processed. Moreover, the estimated processing time is based on a mean performance and may very significantly when individual data granules are considered.

9.2 Inputs

The RemoTeC-S5 input and output is already discussed in detail in Sec. 6.11 and is summarized in the Tables 16 to 19. We distinguish between static and dynamic fields. Here, static fields are data arrays which do not change during the processing of a data granule, whereas dynamic fields have different values for the individual ground pixel processing. It is important to realize that the RemoTeC-S5 implementation requires spatially collocated Earthshine radiances. Hence, we assumed that a spatial resampling of the measurement is applied to the L1B data before providing the data to RemoTeC. The algorithm specific settings are provided via a cumulative string, which is unraveled by the RemoTeC software. Appendix B gives an overview of all relevant setting parameters.

9.2.1 Static Inputs

Table 16: Static input data of RemoTeC-S5


Parameter	Symbol	Dimension	Physical Units	Source	M/O ^a	C/F/U ^b
Length of algorithm settings	CFG- <i>N</i>	1	-	S5 L2 CH4 CFG	M	U
Algorithm settings	CFG- <i>S</i>	CFG- <i>N</i>	-	S5 L2 CH4 CFG	M	U
Level of breakpoint output	FWK- <i>L</i>	1	-	S5 L2 CH4 CFG	O	U
Number of swath positions SWIR-1 ^c	N_{sp}	1	-	S5 L2 AUX ISRF	M	U
Number of wavelength pixels NIR-2a ^d	N_{N2A}	1	-	S5 L2 AUX ISRF	M	U
Number of wavelength pixels NIR-2 ^d	N_{NR2}	1	-	S5 L2 AUX ISRF	M	U
Number of wavelength pixels SWIR-1 ^d	N_{SW1}	1	-	S5 L2 AUX ISRF	M	U
Number of wavelength pixels SWIR-3 ^d	N_{SW3}	1	-	S5 L2 AUX ISRF	M	U
Irradiance wavelength NIR-2a	IR- λ_{N2A}	$N_{sp} \times N_{N2A}$	nm	S5 L1B irradiance product	M	U
Irradiance spectrum NIR-2a	IR- I_{N2A}	$N_{sp} \times N_{N2A}$	mol/(m ² s nm)	S5 L1B irradiance product	M	U
Irradiance noise spectrum NIR-2a	IR- e_{N2A}	$N_{sp} \times N_{N2A}$	mol/(m ² s nm)	S5 L1B irradiance product	M	U
Pixel mask of irradiance spectrum NIR-2a	IR- p_{N2A}	$N_{sp} \times N_{N2A}$	-	S5 L1B irradiance product	M	U
Irradiance wavelength NIR-2	IR- λ_{NR2}	$N_{sp} \times N_{NR2}$	nm	S5 L1B irradiance product	M	U
Irradiance spectrum NIR-2	IR- I_{NR2}	$N_{sp} \times N_{NR2}$	mol/(m ² s nm)	S5 L1B irradiance product	M	U
Irradiance noise spectrum NIR-2	IR- e_{NR2}	$N_{sp} \times N_{NR2}$	mol/(m ² s nm)	S5 L1B irradiance product	M	U
Pixel mask of irradiance spectrum NIR-2	IR- p_{NR2}	$N_{sp} \times N_{NR2}$	-	S5 L1B irradiance product	M	U
Irradiance wavelength SWIR-1	IR- λ_{SW1}	$N_{sp} \times N_{SW1}$	nm	S5 L1B irradiance product	M	U
Irradiance spectrum SWIR-1	IR- I_{SW1}	$N_{sp} \times N_{SW1}$	mol/(m ² s nm)	S5 L1B irradiance product	M	U
Irradiance noise spectrum SWIR-1	IR- e_{SW1}	$N_{sp} \times N_{SW1}$	mol/(m ² s nm)	S5 L1B irradiance product	M	U
Pixel mask of irradiance spectrum SWIR-1	IR- p_{SW1}	$N_{sp} \times N_{SW1}$	-	S5 L1B irradiance product	M	U
Irradiance wavelength SWIR-3	IR- λ_{SW3}	$N_{sp} \times N_{SW3}$	nm	S5 L1B irradiance product	M	U
Irradiance spectrum SWIR-3	IR- I_{SW3}	$N_{sp} \times N_{SW3}$	mol/(m ² s nm)	S5 L1B irradiance product	M	U
Irradiance noise spectrum SWIR-3	IR- e_{SW3}	$N_{sp} \times N_{SW3}$	mol/(m ² s nm)	S5 L1B irradiance product	M	U
Pixel mask of irradiance spectrum SWIR-3	IR- p_{SW3}	$N_{sp} \times N_{SW3}$	-	S5 L1B irradiance product	M	U
Number of wavelength difference in ISRF definition NIR-2a	ISRF- $N_{\delta\lambda, N2A}$	1	-	S5 L2 AUX ISRF	M	U

Continued on next page

Table 16 – Static input - Continued from previous page

Parameter	Symbol	Dimension	Physical Unit	Source	M/O ^a	C/F/U ^b
Wavelength difference in ISRF definition NIR-2a	ISRF- $\delta\lambda_{N2A}$	ISRF- $N_{\delta\lambda,N2A}$	nm	S5 L2 AUX ISRF	M	U
Instrument spectral response function NIR-2a	ISRF- S_{N2A}	$N_{sp} \times N_{N2A} \times$ ISRF- $N_{\delta\lambda,N2A}$	nm ⁻¹	S5 L2 AUX ISRF	M	U
Number of wavelength difference in ISRF definition NIR-2	ISRF- $N_{\delta\lambda,NR2}$	1	-	S5 L2 AUX ISRF	M	U
Wavelength difference in ISRF definition NIR-2	ISRF- $\delta\lambda_{NR2}$	ISRF- $N_{\delta\lambda,NR2}$	nm	S5 L2 AUX ISRF	M	U
Instrument spectral response function NIR-2	ISRF- S_{NR2}	$N_{sp} \times N_{NR2} \times$ ISRF- $N_{\delta\lambda,NR2}$	nm ⁻¹	S5 L2 AUX ISRF	M	U
Number of wavelength difference in ISRF definition SWIR-1	ISRF- $N_{\delta\lambda,SW1}$	1	-	S5 L2 AUX ISRF	M	U
Wavelength difference in ISRF definition SWIR-1	ISRF- $\delta\lambda_{SW1}$	ISRF- $N_{\delta\lambda,SW1}$	nm	S5 L2 AUX ISRF	M	U
Instrument spectral response function SWIR-1	ISRF- S_{SW1}	$N_{sp} \times N_{SW1} \times$ ISRF- $N_{\delta\lambda,SW1}$	nm ⁻¹	S5 L2 AUX ISRF	M	U
Number of wavelength difference in ISRF definition SWIR-3	ISRF- $N_{\delta\lambda,SW3}$	1	-	S5 L2 AUX ISRF	M	U
Wavelength difference in ISRF definition SWIR-3	ISRF- $\delta\lambda_{SW3}$	ISRF- $N_{\delta\lambda,SW3}$	nm	S5 L2 AUX ISRF	M	U
Instrument spectral response function SWIR-3	ISRF- S_{SW3}	$N_{sp} \times N_{SW3} \times$ ISRF- $N_{\delta\lambda,SW3}$	nm ⁻¹	S5 L2 AUX ISRF	M	U
Number of wavelengths reference solar spectrum	SUN- N	1	-	S5 L2 CH4 LUT	M	U
Wavelengths reference solar spectrum	SUN- λ	SUN- N	nm	S5 L2 CH4 LUT	M	U
Reference solar spectrum	SUN- I	SUN- N	mol/(m ² s nm)	S5 L2 CH4 LUT	M	U
Number of hybrid layers of meteorological input	ATM- N_{met}	1	-	S5 L2 AUX product (CAMS)	M	U
Hybrid pressure coefficient A at layer interface for meteorological input	ATM- A_{met}	ATM- $N_{met} + 1$	Pa	S5 L2 AUX product (CAMS)	M	U
Hybrid pressure coefficient B at layer interface for meteorological input	ATM- B_{met}	ATM- $N_{met} + 1$	1	S5 L2 AUX product (CAMS)	M	U

Continued on next page



S5L2PPP
Methane ATBD

Reference : SRON-ESA-S5L2PPP-ATBD-001
Version : 3.1
Date : 17 May 2019

Page
74/91

Table 16 – Static input - Continued from previous page

Parameter	Symbol	Dimension	Physical Unit	Source	M/O ^a	C/F/U ^b
Number of hybrid layers of CH ₄ input	ATM- N_{CH_4}	1	-	S5 L2 AUX product (CAMS)	M	U
Hybrid pressure coefficient A at layer interface for CH ₄ input	ATM- A_{CH_4}	ATM- $N_{CH_4} + 1$	Pa	S5 L2 AUX product (CAMS)	M	U
Hybrid pressure coefficient B at layer interface for CH ₄ input	ATM- B_{CH_4}	ATM- $N_{CH_4} + 1$	1	S5 L2 AUX product (CAMS)	M	U
Number of hybrid layers of CO ₂ input	ATM- N_{CO_2}	1	-	S5 L2 AUX product (CAMS)	M	U
Hybrid pressure coefficient A at layer interface for CO ₂ input	ATM- A_{CO_2}	ATM- $N_{CO_2} + 1$	Pa	S5 L2 AUX product (CAMS)	M	U
Hybrid pressure coefficient B at layer interface for CO ₂ input	ATM- B_{CO_2}	ATM- $N_{CO_2} + 1$	1	S5 L2 AUX product (CAMS)	M	U
Number of hybrid layers of CO input	ATM- N_{CO}	1	-	S5 L2 AUX product (CAMS)	M	U
Hybrid pressure coefficient A at layer interface for CO input	ATM- A_{CO}	ATM- $N_{CO} + 1$	Pa	S5 L2 AUX product (CAMS)	M	U
Hybrid pressure coefficient B at layer interface for CO input	ATM- B_{CO}	ATM- $N_{CO} + 1$	1	S5 L2 AUX product (CAMS)	M	U
Number of real refractive indices	AER- N_{m_r}	1	-	S5 L2 CH4 LUT	M	U
Number of imaginary refractive indices	AER- N_{m_i}	1	-	S5 L2 CH4 LUT	M	U
Number of scattering angles	AER- N_{Θ}	1	-	S5 L2 CH4 LUT	M	U
Number of effective radii aerosols	AER- $N_{r_{eff}}$	1	-	S5 L2 CH4 LUT	M	U
Number of scattering matrix elements	AER- N_s	1	-	S5 L2 CH4 LUT	M	U
Aerosol reference wavelength	AER- λ	1	nm	S5 L2 CH4 LUT	M	U
Real refractive indices	AER- m_r	AER- N_{m_r}	1	S5 L2 CH4 LUT	M	U
Imaginary refractive indices	AER- m_i	AER- N_{m_i}	1	S5 L2 CH4 LUT	M	U
Scattering angles	AER- Θ	AER- N_{Θ}	°	S5 L2 CH4 LUT	M	U
Effective radii aerosols	AER- r_{eff}	AER- $N_{r_{eff}}$	μm	S5 L2 CH4 LUT	M	U
Ellipsoid extinction coefficient	AER- $\sigma_{ext,el}$	AER- $N_{m_i} \times$ AER- $N_{m_r} \times$ AER- $N_{r_{eff}}$	a.u.	S5 L2 CH4 LUT	M	U
Sphere extinction coefficient	AER- $\sigma_{ext,sph}$	AER- $N_{m_i} \times$ AER- $N_{m_r} \times$ AER- $N_{r_{eff}}$	a.u.	S5 L2 CH4 LUT	M	U
Ellipsoid absorption coefficient	AER- $\sigma_{abs,el}$	AER- $N_{m_i} \times$ AER- $N_{m_r} \times$ AER- $N_{r_{eff}}$	a.u.	S5 L2 CH4 LUT	M	U

Continued on next page



S5L2PPP
Methane ATBD


Reference : SRON-ESA-S5L2PPP-ATBD-001
Version : 3.1
Date : 17 May 2019

Page
75/91

Table 16 – Static input - Continued from previous page

Parameter	Symbol	Dimension	Physical Unit	Source	M/O ^a	C/F/U ^b
Sphere absorption coefficient	$AER-\sigma_{abs,sph}$	$AER-N_{m_i} \times AER-N_{m_r} \times AER-N_{r_{eff}}$	a.u.	S5 L2 CH4 LUT	M	U
Ellipsoid scattering matrix	$AER-F_{el}$	$AER-N_s \times AER-N_{m_i} \times AER-N_{m_r} \times AER-N_{r_{eff}} \times AER-N_{\Theta}$	1	S5 L2 CH4 LUT	M	U
Sphere scattering matrix	$AER-F_{sph}$	$AER-N_s \times AER-N_{m_i} \times AER-N_{m_r} \times AER-N_{r_{eff}} \times AER-N_{\Theta}$	1	S5 L2 CH4 LUT	M	U
Number of pressures cross sections	$XS-N_p$	1	-	S5 L2 CH4 LUT	M	U
Number of temperatures cross sections	$XS-N_T$	1	-	S5 L2 CH4 LUT	M	U
Pressures cross sections	$XS-p$	$XS-N_p$	Pa	S5 L2 CH4 LUT	M	U
Temperatures cross sections	$XS-T$	$XS-N_T \times XS-N_p$	K	S5 L2 CH4 LUT	M	U
Number of NR2 wavenumbers cross sections	$XS-N_{\nu_{NR2}}$	1	-	S5 L2 CH4 LUT	M	U
NIR-2 wavenumbers cross sections	$XS-\nu_{NR2}$	$XS-N_{\nu_{NR2}}$	cm^{-1}	S5 L2 CH4 LUT	M	U
Cross sections O ₂ NIR-2	$XS-\sigma_{O_2,NR2}$	$XS-N_p \times XS-N_T \times XS-N_{\nu_{NR2}}$	cm^2	S5 L2 CH4 LUT	M	U
Number of SWIR-1 wavenumbers cross sections	$XS-N_{\nu_{SW1}}$	1	-	S5 L2 CH4 LUT	M	U
SWIR-1 wavenumbers cross sections	$XS-\nu_{SW1}$	$XS-N_{\nu_{SW1}}$	cm^{-1}	S5 L2 CH4 LUT	M	U
Cross sections H ₂ O SWIR-1	$XS-\sigma_{H_2O,SW1}$	$XS-N_p \times XS-N_T \times XS-N_{\nu_{SW1}}$	cm^2	S5 L2 CH4 LUT	M	U
Cross sections CH ₄ SWIR-1	$XS-\sigma_{CH_4,SW1}$	$XS-N_p \times XS-N_T \times XS-N_{\nu_{SW1}}$	cm^2	S5 L2 CH4 LUT	M	U
Cross sections CO ₂ SWIR-1	$XS-\sigma_{CO_2,SW1}$	$XS-N_p \times XS-N_T \times XS-N_{\nu_{SW1}}$	cm^2	S5 L2 CH4 LUT	M	U
Number of SWIR-3 wavenumbers cross sections	$XS-N_{\nu_{SW3}}$	1	-	S5 L2 CH4 LUT	M	U
SWIR-3 wavenumbers cross sections	$XS-\nu_{SW3}$	$XS-N_{\nu_{SW3}}$	cm^{-1}	S5 L2 CH4 LUT	M	U
Cross sections H ₂ O SWIR-3	$XS-\sigma_{H_2O,SW3}$	$XS-N_p \times XS-N_T \times XS-N_{\nu_{SW3}}$	cm^2	S5 L2 CH4 LUT	M	U

Continued on next page



S5L2PPP
Methane ATBD

Reference : SRON-ESA-S5L2PPP-ATBD-001
Version : 3.1
Date : 17 May 2019
Page 76/91

Table 16 – Static input - Continued from previous page

Parameter	Symbol	Dimension	Physical Unit	Source	M/O ^a	C/F/U ^b
Cross sections CH ₄ SWIR-3	XS- $\sigma_{\text{CH}_4, \text{SW3}}$	XS- $N_p \times \text{XS-}N_T \times$ XS- N_{VSW3}	cm ²	S5 L2 CH4 LUT	M	U
Cross sections CO SWIR-3	XS- $\sigma_{\text{CO}, \text{SW3}}$	XS- $N_p \times \text{XS-}N_T \times$ XS- N_{VSW3}	cm ²	S5 L2 CH4 LUT	M	U

^a M mandatory input, O optional input.

^b C copy to output, F use for filtering U use for processing.

^c RemoTeC-S5 requires that all spatial (detector spatial channel) information in collocated to the SWIR-1

^d Spectral information (ISRF, radiance and irradiance) are given on a grid with common band specific size

The NIR, SWIR-1 and SWIR-3 L1B irradiance product is assumed to be spatially collocated to the SWIR-1 pixels. This is accomplished by applying the co-registration algorithm [RD79] with the NIR and SWIR-3 as source and the SWIR-1 as target. Obviously, the irradiance product has only a detector spatial channel dimension but does not have a geo-location and does not have a scanline dimensions. Therefore, the co-registration is based on the geo-locations of the L1b radiance products of the same bands and subsequently integrated over the scanline dimension. The spectral dimension of the radiance, irradiance and ISRF input is band but not product specific.

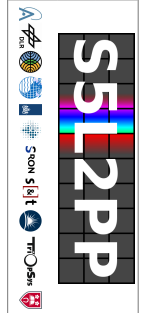
The static input contains the algorithm configuration parameters as character string. This contains the contents of a settings file, which will be interpreted by the algorithm. The content of these settings is explained in Appendix B.

9.2.2 Dynamic Inputs

Table 17: Dynamic input of RemoTeC-S5.

Parameter	Symbol	Dimension	Physical Unit	Source	M/O ^a	C/F/U ^b
Framework processing quality flags	FWK-PQF _{in}	1	-	S5 L2 CH4 CFG	O	C
Swath position index of the sounding ^d	i_{sp}	1	-	S5 L1B radiance product (band 4)	M	U
Radiance wavelength NIR-2a	RA- λ_{N2A}	N_{N2A}	nm	S5 L1B radiance product (band 3b)	M	P
Radiance spectrum NIR-2a	RA- I_{N2A}	N_{N2A}	mol/(m ² s sr nm)	S5 L1B radiance product (band 3b)	M	P
Radiance noise spectrum NIR-2a	RA- e_{N2A}	N_{N2A}	mol/(m ² s sr nm)	S5 L1B radiance product (band 3b)	M	P
Pixel mask of radiance spectrum NIR-2a	RA- p_{N2A}	N_{N2A}	-	S5 L1B radiance product (band 3b)	M	P
Radiance wavelength NIR-2	RA- λ_{NR2}	N_{NR2}	nm	S5 L1B radiance product (band 3c)	M	P
Radiance spectrum NIR-2	RA- I_{NR2}	N_{NR2}	mol/(m ² s sr nm)	S5 L1B radiance product (band 3c)	M	P
Radiance noise spectrum NIR-2	RA- e_{NR2}	N_{NR2}	mol/(m ² s sr nm)	S5 L1B radiance product (band 3c)	M	P

Continued on next page



S5L2PP
Methane ATBD

Reference : SRON-ESA-S5L2PP-ATBD-001
Version : 3.1
Date : 17 May 2019
Page 77/91

Table 17 – Dynamic input data - Continued from previous page

Parameter	Symbol	Dimension	Physical Unit	Source	M/O ^d	C/F/U ^b
Pixel mask of radiance spectrum NIR-2	RA- p_{NR2}	N_{NR2}	-	S5 L1B radiance product (band 3c)	M	P
Radiance wavelength SWIR-1	RA- λ_{SW1}	N_{SW1}	nm	S5 L1B radiance product (band 4)	M	U
Radiance spectrum SWIR-1	RA- I_{SW1}	N_{SW1}	mol/(m ² s sr nm)	S5 L1B radiance product (band 4)	M	U
Radiance noise spectrum SWIR-1	RA- e_{SW1}	N_{SW1}	mol/(m ² s sr nm)	S5 L1B radiance product (band 4)	M	U
Pixel mask of radiance spectrum SWIR-1	RA- p_{SW1}	N_{SW1}	-	S5 L1B radiance product (band 4)	M	U
Radiance wavelength SWIR-3	RA- λ_{SW3}	N_{SW3}	nm	S5 L1B radiance product (band 5)	M	U
Radiance spectrum SWIR-3	RA- I_{SW3}	N_{SW3}	mol/(m ² s sr nm)	S5 L1B radiance product (band 5)	M	U
Radiance noise spectrum SWIR-3	RA- e_{SW3}	N_{SW3}	mol/(m ² s sr nm)	S5 L1B radiance product (band 5)	M	U
Pixel mask of radiance spectrum SWIR-3	RA- p_{SW3}	N_{SW3}	-	S5 L1B radiance product (band 5)	M	U
Solar zenith angle ^d	GEO- θ_0	1	°	S5 L1B radiance product (band 4)	M	U
Solar azimuth angle ^d	GEO- φ_0	1	°	S5 L1B radiance product (band 4)	M	U
Viewing zenith angle ^d	GEO- θ_v	1	°	S5 L1B radiance product (band 4)	M	U
Viewing azimuth angle ^d	GEO- φ_v	1	°	S5 L1B radiance product (band 4)	M	U
Latitude pixel center	GEO- Θ_{center}	1	°	S5 L1B radiance product (band 4)	M	U
Longitude pixel center	GEO- Φ_{center}	1	°	S5 L1B radiance product (band 4)	M	C
Latitude pixel corners	GEO- Θ_{corner}	4	°	S5 L1B radiance product (band 4)	M	C
Longitude pixel corners	GEO- Φ_{corner}	4	°	S5 L1B radiance product (band 4)	M	C
Time of measurement	GEO- t	7	TMDHMSM ^c	S5 L1B radiance product L1B	M	C
Mean surface elevation of pixel	GEO- z_{surf}	1	m	S5 L2 AUX product (GMTED2010) ^e	M	U
Standard deviation of surface elevation	GEO- $\sigma_{z_{surf}}$	1	m	S5 L2 AUX product (GMTED2010) ^e	M	F
Surface classification	GEO- f_{surf}	1	-	S5 L2 AUX product (GMTED2010) ^e	M	F
Surface pressure meteorological model	ATM- $p_{surf,met}$	1	Pa	S5 L2 AUX product (ECMWF) ^e	M	U
Surface elevation model	ATM- z_{surf}	1	m	S5 L2 AUX product (ECMWF) ^e	M	U
Specific humidity profile	ATM- q	ATM- N_{met}	kg/kg	S5 L2 AUX product (ECMWF) ^e	M	U
Temperature profile	ATM- T	ATM- N_{met}	K	S5 L2 AUX product (ECMWF) ^e	M	U
Eastward ten-meter wind	ATM- u	1	m/s	S5 L2 AUX product (ECMWF) ^e	M	U
Northward ten-meter wind	ATM- v	1	m/s	S5 L2 AUX product (ECMWF) ^e	M	U
Surface pressure CH ₄ profile	ATM- p_{surf,CH_4}	1	Pa	S5 L2 AUX product (CAMS) ^e	M	U
CH ₄ mixing ratio profile	ATM- r_{CH_4}	ATM- N_{CH_4}	1	S5 L2 AUX product (CAMS) ^e	M	U
Surface pressure CO ₂ profile	ATM- p_{surf,CO_2}	1	Pa	S5 L2 AUX product (CAMS) ^e	M	U

Continued on next page

Table 17 – Dynamic input data - Continued from previous page

Parameter	Symbol	Dimension	Physical Unit	Source	M/O ^a	C/F/U ^b
CO ₂ mixing ratio profile	ATM- r_{CO_2}	ATM- N_{CO_2}	1	S5 L2 AUX product (CAMS) ^e	M	U
Surface pressure CO profile	ATM- $p_{\text{surf,CO}}$	1	Pa	S5 L2 AUX product (CAMS) ^e	M	U
CO mixing ratio profile	ATM- r_{CO}	ATM- N_{CO}	1	S5 L2 AUX product (CAMS) ^e	M	U
METimage FOV averaged radiance band 4	MET- I_{b4}	4	W/ (m ² sr μm)	S5 L2 HET	O	F
METimage standard deviation radiance band 4	MET- $\sigma_{I_{b4}}$	4	W/ (m ² sr μm)	S5 L2 HET	O	F
METimage FOV averaged radiance band 9	MET- I_{b9}	4	W/ (m ² sr μm)	S5 L2 HET	O	F
METimage standard deviation radiance band 9	MET- $\sigma_{I_{b9}}$	4	W/ (m ² sr μm)	S5 L2 HET	O	F
Number of METimage pixels confidently cloudy	MET- $N_{c\text{-cld}}$	4	-	S5 L2 CLA	O	F
Number of METimage pixels probably cloudy	MET- $N_{p\text{-cld}}$	4	-	S5 L2 CLA	O	F
Number of METimage pixels probably clear	MET- $N_{p\text{-clr}}$	4	-	S5 L2 CLA	O	F
Number of METimage pixels confidently clear	MET- $N_{c\text{-clr}}$	4	-	S5 L2 CLA	O	F
Ice cloud fraction	MET- f_{ice}	1	1	S5 L2 CLA	O	F
Ice cloud optical thickness	MET- τ_{ice}	1	1	S5 L2 CLA	O	F
Ice cloud top pressure	MET- p_{ice}	1	Pa	S5 L2 CLA	O	F
Ice cloud effective radius	MET- r_{ice}	1	μm	S5 L2 CLA	O	F
Liquid cloud fraction	MET- f_{liq}	1	1	S5 L2 CLA	O	F
Liquid cloud optical depth	MET- r_{liq}	1	1	S5 L2 CLA	O	F
Liquid cloud top pressure	MET- p_{liq}	1	Pa	S5 L2 CLA	O	F
Liquid cloud effective radius	MET- r_{liq}	1	μm	S5 L2 CLA	O	F
Coverage of valid METimage data	MET- N	1	-	S5 L2 CLA	O	F

^a M mandatory input, O optional input.

^b C copy to output, F use for filtering, U use for processing.



^c Year, month, day, hour, minute, second, millisecond.

^d The S5 RemoTeC algorithm assumes spatially common solar and viewing geometry common swath index.

^e Parameters are provided by the S5 L2 AUX product after regridding.

^f CFG means configuration.

The NIR-2a, NIR-2, SWIR-1 and SWIR-3 L1B radiance product of the dynamical input is assumed to be spatially collocated to the SWIR-1 pixels. This means that the coregistration algorithm [RD79] is executed with the SWIR-1 as target and the NIR-2a, NIR-2 and SWIR-3 as source. Therefore, the spatial coordinates (latitude, longitude) of the footprint and the solar and viewing geometry is only required for the SWIR-1 band. The METimage input data, used for data filtering, are classified as optional input. In case data are not available, the data selection will be based on internal non-scattering retrievals as described in Sect. 6.6.

 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Date : 17 May 2019	Page 80/91
--	-------------------------------	---	---------------

9.3 Outputs

9.3.1 Static Output

Table 18: Output data of the static initialization

Parameter	Symbol	Dimension	Physical Unit	Range	Destination
String length of RemoTeC version	VER- <i>N</i>	1	-	1 – 10	S5 L2 CH ₄ product
RemoTeC version	VER- <i>S</i>	VER- <i>N</i>	-	NA	S5 L2 CH ₄ product
Number of vertical layers	MDL- <i>N</i>	1	-	1 – 40	S5 L2 CH ₄ product
Number of wavelengths for SIF output	SIF- <i>N</i>	1	-	2 – 6	S5 L2 CH ₄ product
Wavelengths for SIF output	SIF- λ	SIF- <i>N</i>	nm	730 – 760	S5 L2 CH ₄ product

9.3.2 Dynamic Output


Table 19: Dynamic output of RemoTeC. Processing Quality Flags are specified separately in Tab. 20.

Parameter	Symbol	Dimension	Physical Unit	Range	Destination
Pressure grid	MDL- <i>p</i>	MDL- <i>N</i>	Pa	$0 - 1.1 \cdot 10^5$	S5 L2 CH ₄ product
altitude grid	MDL- <i>z</i>	MDL- <i>N</i>	m	$0 - 1.0 \cdot 10^5$	S5 L2 CH ₄ product
Dry air column	AP- <i>C</i> _{air}	1	mol / m ²	$3 \cdot 10^5 - 4 \cdot 10^5$	S5 L2 CH ₄ product
A-priori CH ₄ profile	AP- <i>c</i> _{CH₄}	MDL- <i>N</i>	mol / m ²	0 – 0.1	S5 L2 CH ₄ product
A-priori CO ₂ profile	AP- <i>c</i> _{CO₂}	MDL- <i>N</i>	mol / m ²	0 – 20	S5 L2 CH ₄ product
A-priori O ₂ column	AP- <i>C</i> _{O₂}	1	mol / m ²	$7 \cdot 10^4 - 8 \cdot 10^4$	S5 L2 CH ₄ product
A-priori H ₂ O column	AP- <i>C</i> _{H₂O}	1	mol / m ²	$0 - 5 \cdot 10^3$	S5 L2 CH ₄ product
Solar induced fluorescence	SIF- <i>F</i>	SIF- <i>N</i>	mol / (m ² s sr nm)	$\pm 0 - 10^{-7}$	S5 L2 CH ₄ product
Non-scattering O ₂ column	NS- <i>C</i> _{O₂}	1	mol / m ²	$0 - 1 \cdot 10^5$	S5 L2 CH ₄ product
Non-scattering H ₂ O column strong	NS- <i>C</i> _{H₂O,s}	1	mol / m ²	$0 - 1 \cdot 10^4$	S5 L2 CH ₄ product
Non-scattering H ₂ O column weak	NS- <i>C</i> _{H₂O,w}	1	mol / m ²	$0 - 1 \cdot 10^4$	S5 L2 CH ₄ product
Non-scattering CH ₄ column strong	NS- <i>C</i> _{CH₄,s}	1	mol / m ²	0 – 1	S5 L2 CH ₄ product
Non-scattering CH ₄ column weak (and proxy)	NS- <i>C</i> _{CH₄,w}	1	mol / m ²	0 – 1	S5 L2 CH ₄ product
Non-scattering CO ₂ column	NS- <i>C</i> _{CO₂}	1	mol / m ²	0 – 300	S5 L2 CH ₄ product
Non-scattering CH ₄ column averaging kernel	NS- <i>a</i> _{CH₄}	MDL- <i>N</i>	1	0 – 3	S5 L2 CH ₄ product
Non-scattering CO ₂ column averaging kernel	NS- <i>a</i> _{CO₂}	MDL- <i>N</i>	1	0 – 3	S5 L2 CH ₄ product
Albedo non-scattering CH ₄ retrieval	NS- <i>A</i> _{CH₄}	1	1	0 – 1	S5 L2 CH ₄ product
Albedo non-scattering CO ₂ retrieval	NS- <i>A</i> _{CO₂}	1	1	0 – 1	S5 L2 CH ₄ product

Continued on next page

Table 19 – Dynamic Output - Continued from previous page

Parameter	Symbol	Dimension	Physical Unit	Range	Destination
Proxy XCH ₄	PR-XCH ₄	1	ppb	1750 – 1850	S5 L2 CH ₄ product
Proxy XCH ₄ precision	PR- e_{XCH_4}	1	ppb	0 – 40	S5 L2 CH ₄ product
Physics XCH ₄	PH-XCH ₄	1	ppb	1750 – 1850	S5 L2 CH ₄ product
Physics XCH ₄ precision	PH- e_{XCH_4}	1	ppb	0 – 40	S5 L2 CH ₄ product
Physics CH ₄ column averaging kernel	PH- a_{CH_4}	MDL- N	1	0 – 3	S5 L2 CH ₄ product
H ₂ O column	PH- C_{H_2O}	1	mol / m ²	0 – 1 · 10 ⁴	S5 L2 CH ₄ product
CO ₂ column	PH- C_{CO_2}	1	mol / m ²	0 – 300	S5 L2 CH ₄ product
CO column	PH- C_{CO}	1	mol / m ²	0 – 0.1	S5 L2 CH ₄ product
Aerosol size parameter	PH- α	1	-	1 – 5	S5 L2 CH ₄ product
Aerosol particle column	PH- N_{aer}	1	mol / m ²	0 – 10 ⁻⁶	S5 L2 CH ₄ product
Aerosol layer height	PH- z_{aer}	1	m	0 – 15000	S5 L2 CH ₄ product
NIR-2 surface albedo	PH- A_{NR2}	1	1	0 – 1	S5 L2 CH ₄ product
SWIR-1 surface albedo	PH- A_{SW1}	1	1	0 – 1	S5 L2 CH ₄ product
SWIR-3 surface albedo	PH- A_{SW3}	1	1	0 – 1	S5 L2 CH ₄ product
NIR-2 spectral shift	PH- $\delta\lambda_{e,NR2}$	1	nm	±0 – 0.02	S5 L2 CH ₄ product
SWIR-1 spectral shift	PH- $\delta\lambda_{e,SW1}$	1	nm	±0 – 0.02	S5 L2 CH ₄ product
SWIR-3 spectral shift	PH- $\delta\lambda_{e,SW3}$	1	nm	±0 – 0.02	S5 L2 CH ₄ product
NIR-2 spectral shift sun	PH- $\delta\lambda_{s,NR2}$	1	nm	±0 – 0.02	S5 L2 CH ₄ product
SWIR-1 spectral shift sun	PH- $\delta\lambda_{s,SW1}$	1	nm	±0 – 0.02	S5 L2 CH ₄ product
SWIR-3 spectral shift sun	PH- $\delta\lambda_{s,SW3}$	1	nm	±0 – 0.02	S5 L2 CH ₄ product
Temperature offset	PH- ΔT	1	K	±0 – 5	S5 L2 CH ₄ product
Number of iterations	PH- N_{iter}	1	-	0 – 30	S5 L2 CH ₄ product
NIR-2 χ^2 of the spectral fit	PH- χ^2_{NR2}	1	-	tbd	S5 L2 CH ₄ product
SWIR-1 χ^2 of the spectral fit	PH- χ^2_{SW1}	1	-	tbd	S5 L2 CH ₄ product
SWIR-3 χ^2 of the spectral fit	PH- χ^2_{SW3}	1	-	tbd	S5 L2 CH ₄ product
Total reduced χ^2 of the spectral fit	PH- χ^2	1	-	tbd	S5 L2 CH ₄ product
Total degrees of freedom for signal	PH-DFS	1	-	tbd	S5 L2 CH ₄ product
CH ₄ degrees of freedom for signal	PH-DFS _{CH₄}	1	-	1 – 2	S5 L2 CH ₄ product
Aerosol degree of freedom for signal	PH-DFS _{aer}	1	-	1 – 3	S5 L2 CH ₄ product
Number of spectral pixels in NIR-2 fitting window	PH- N_{NR2}	1	-	≤ IR- N_{NR2}	S5 L2 CH ₄ product
Number of spectral pixels in SWIR-1 fitting window	PH- N_{SW1}	1	-	≤ IR- N_{SW1}	S5 L2 CH ₄ product
Number of spectral pixels in SWIR-3 fitting window	PH- N_{SW3}	1	-	≤ IR- N_{SW3}	S5 L2 CH ₄ product
Processing quality flag	FWK-PQF	1	-	any	S5 L2 CH ₄ product

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 82/91
---	------------------------	--

The processing quality flag (PQF) is an important element of the dynamic algorithm output, which indicates the status, errors, data filtering and additional warnings to evaluate the retrieval quality. The PQFs are summarized in Tab. 20.

Table 20: 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	F	Scene flagged as cloudy by internal cloud filter protocol.
metimage_cloud_filter	F	Scene flagged as cloudy from METimage input data.
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.
metimage_scene_heterogeneity_filter	F	Scene flagged as cloud from METimage input data.
metimage_cirrus_filter	F	Scene flagged as cirrus-contaminated from dedicated cirrus METimage 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.
sun_glint_warning	W	Sun glint possibility warning.
snow_ice_warning	W	Scene is over snow or ice.
pixel_level_input_data_missing	W	Dynamic auxiliary input data (e.g.. METimage) is missing for this ground pixel. A fallback option is used.
altitude_consistency_warning	W	Large, but not too large, difference between model altitude and scene altitude value.
deconvolution_warning	W	Failed deconvolution irradiance spectrum (not specific for ground pixel, but specific for detector spatial channel index).
side_retrieval_failure_warning	W	A retrieval of a side product that is not used for filtering failed.
input_spectrum_warning	W	Few, but not too few, non-flagged spectral pixels in both radiance and irradiance to perform retrieval.

9.3.3 Breakpoint Output



 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Date : 17 May 2019	Page 83/91
--	--------------------------------------	---	---------------

Table 21: Breakpoint output of RemoTeC.

Parameter	Symbol	Dimension	Physical Unit	Range	Destination
wavelengths NIR-2	RES- λ_{NR2}	IR- N_{NR2}	nm	755 – 773	S5 L2 CH ₄ product
measured spectrum NIR-2	RES- $y_{meas,NR2}$	IR- N_{NR2}	mol/(m ² s sr nm)	0 – 10 ⁻⁶	S5 L2 CH ₄ product
noise spectrum NIR-2	RES- $e_{meas,NR2}$	IR- N_{NR2}	mol/(m ² s sr nm)	0 – 10 ⁻⁸	S5 L2 CH ₄ product
modeled spectrum NIR-2	RES- $y_{mod,NR2}$	IR- N_{NR2}	mol/(m ² s sr nm)	0 – 10 ⁻⁶	S5 L2 CH ₄ product
wavelengths SWIR-1	RES- λ_{SW1}	IR- N_{SW1}	nm	1590 – 1675	S5 L2 CH ₄ product
measured spectrum SWIR-1	RES- $y_{meas,SW1}$	IR- N_{SW1}	mol/(m ² s sr nm)	0 – 10 ⁻⁶	S5 L2 CH ₄ product
noise spectrum SWIR-1	RES- $e_{meas,SW1}$	IR- N_{SW1}	mol/(m ² s sr nm)	0 – 10 ⁻⁸	S5 L2 CH ₄ product
modeled spectrum SWIR-1	RES- $y_{mod,SW1}$	IR- N_{SW1}	mol/(m ² s sr nm)	0 – 10 ⁻⁶	S5 L2 CH ₄ product
wavelengths SWIR-3	RES- λ_{SW3}	IR- N_{SW3}	nm	2305 – 2385	S5 L2 CH ₄ product
measured spectrum SWIR-3	RES- $y_{meas,SW3}$	IR- N_{SW3}	mol/(m ² s sr nm)	0 – 10 ⁻⁶	S5 L2 CH ₄ product
noise spectrum SWIR-3	RES- $e_{meas,SW3}$	IR- N_{SW3}	mol/(m ² s sr nm)	0 – 10 ⁻⁸	S5 L2 CH ₄ product
modeled spectrum SWIR-3	RES- $y_{mod,SW3}$	IR- N_{SW3}	mol/(m ² s sr nm)	0 – 10 ⁻⁶	S5 L2 CH ₄ product

Table 21 lists additional optional output that is meant for investigation on the retrieval process. In the default run, this output is not written out.

Appendices

A Appendix A: 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 [RD16, RD55] and its latest version 2.0 is already used for the operational processing of the S5P CH₄ 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 S . In its forward formulation, the transport equation is written as

$$\hat{\mathbf{L}} \circ \mathbf{I} = \mathbf{S}_0, \quad (117)$$

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 [RD80, RD81].

$$\hat{\mathbf{L}} = \hat{\mathbf{L}}_e - \hat{\mathbf{L}}_s \quad (118)$$


with

$$\hat{\mathbf{L}}_e = \left[\mu \frac{\delta}{\delta z} + \beta_e(z) \right] \quad (119)$$

and

$$\hat{\mathbf{L}}_s = \int_{4\pi} d\tilde{\Omega} \left\{ \frac{\beta_s(z)}{4\pi} P(z, \tilde{\Omega}, \Omega) + \rho(\tilde{\Omega}, \Omega) \delta(z) \Theta(\mu) |\mu| \Theta(-\tilde{\mu}) |\tilde{\mu}| \right\}. \quad (120)$$

Here z denotes the vertical coordinate running from the surface to the top of the atmosphere, Ω 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

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 84/91
---	------------------------	--

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. (117) is the unidirectional solar source S_0 , incident at the top of the atmosphere,

$$S_0 = u_0 \delta(z - z_{\text{TOA}}) \delta(\Omega - \Omega_0) F_0, \quad (121)$$

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 $\Omega_0 = (-u_0, \varphi_0)$ (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 [RD82, RD83] through an inner product:

$$\mathcal{F} = \langle \mathbf{R}, \mathbf{I} \rangle. \quad (122)$$

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

$$\langle \mathbf{a}, \mathbf{b} \rangle = \int dz \int d\Omega a(z, \Omega) b(z, \Omega). \quad (123)$$

Considering a downward viewing instrument measuring intensity at the top of the atmosphere in viewing direction Ω_v , the response function is given by

$$R(z, \Omega) = \frac{1}{F_0} \delta(\Omega - \Omega_v) \delta(z - z_{\text{TOA}}). \quad (124)$$

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. (117). 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. (118) as shown in Schepers et al.(2014) [RD84]. Here, the order of scattering is defined as follows. The source of Eq. (117) 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

$$\hat{L}_e \circ \mathbf{I}_n = \mathbf{S}_n \quad (125)$$

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

$$(\hat{L}_e - \hat{L}_s) \circ \mathbf{I}_{n+} = \mathbf{S}_n \quad (126)$$

where for $n = 0$, $\mathbf{I}_{0+} = \mathbf{I}$ and it will be exactly Eq. (117). Combining Eqs. (125) and (126) gives.

$$\hat{L}_e \circ \mathbf{I}_{(n+1)+} = \hat{L}_s \circ \mathbf{I}_{n+} \quad (127)$$

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

$$\hat{L}_e \circ \mathbf{I}_{n+1} = \hat{L}_s \circ \mathbf{I}_n \quad (128)$$

And substituting the left hand side with Eq.(125) for $n + 1$ gives



$$\hat{L}_s \circ \mathbf{I}_n = \mathbf{S}_{n+1} \quad (129)$$

Using Eqs.(125), (129) and (122), any finite order of scattering can be solved analytically.

$$\mathcal{F}_n = \langle \mathbf{R}, (\hat{L}_e^{-1} \hat{L}_s)^n \hat{L}_e^{-1} \mathbf{S}_0 \rangle \quad (130)$$

Here, \hat{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{L}_e or \hat{L}_s can be obtained by straight forward differentiation.

Higher order of scattering are calculated by solving Eq.(126) 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

 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 85/91
--	------------------------	--

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.

$$\mathcal{F}_{n+} = \langle \mathbf{R}, \mathbf{I}_{n+} \rangle \quad (131)$$

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. [RD82, RD83]). 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

$$\hat{\mathbf{L}}^\dagger \circ \mathbf{I}^\dagger = \mathbf{R} \quad (132)$$

Here $\hat{\mathbf{L}}^\dagger$ represents the adjoint radiative transfer operator [RD83, RD80, RD81]. Within the forward-adjoint perturbation theory, the adjoint intensity field represents the importance of sources within the atmosphere for a given observation \mathcal{F} [RD85, RD86]. This follows from the definition of the adjoint operator $\hat{\mathbf{L}}^\dagger$ with

$$\mathcal{F} = \langle \mathbf{R}, \mathbf{I} \rangle = \langle \hat{\mathbf{L}}^\dagger \circ \mathbf{I}^\dagger, \mathbf{I} \rangle = \langle \mathbf{I}^\dagger, \hat{\mathbf{L}} \circ \mathbf{I} \rangle = \langle \mathbf{I}^\dagger, \mathbf{S} \rangle \quad (133)$$

The third equal sign uses the definition of an adjoint operator and the other equal signs use Eqs. (122), (132) and (117).

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 [RD87, RD80]

$$\frac{\partial \mathcal{F}}{\partial x} = -\langle \mathbf{I}^\dagger, \hat{\mathbf{L}}' \mathbf{I} \rangle + \langle \mathbf{I}^\dagger, \mathbf{S}' \rangle + \langle \mathbf{R}', \mathbf{I} \rangle, \quad (134)$$

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

$$\hat{\mathbf{L}}' = \lim_{\Delta x \rightarrow 0} \frac{\Delta \hat{\mathbf{L}}}{\Delta x} \quad (135)$$

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. [RD88, RD89, RD81]). Summarizing the use of the forward-adjoint perturbation theory for the linearization of our problem is justified by the fact that the inner product in Eq. (134) can be calculated in a straight forward manner, which provides the derivatives of \mathcal{F} with respect to the optical parameters $\tau_{\text{abs},k}$, $\tau_{\text{sca},k}$, 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.

	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001
		Version : 3.1 Page
		Date : 17 May 2019 86/91

B Algorithm settings

This appendix describes the user settings for RemoTeC-S5. They are provided in an ASCII format to be read in by the S5 framework and passed to RemoTeC as a character string via the framework interface. RemoTeC parses and interprets the string content. The core of the RemoTeC algorithm is programmed for several missions and so parts of the user settings are prescribed for the S5 mission. This section describes only those setting parameters, which are relevant in the context of this ATBD. Changes of other prescribed setting parameters are not recommended and result in a performance, which is not compliant with the described RemoTeC-S5 algorithm.

B.1 Filter settings

As part of the S5 CH₄ processing protocol up to seven retrievals are performed. At three designated points in the protocol, there is a filter step that, if triggered, stops the execution. All filter thresholds are part of the user settings.

Table 22: Filter settings used in RemoTeC.

Tag	Description
min_signal_sw3	Minimum intensity in SWIR-3 (T_{A1}).
max_sza	Maximum solar zenith angle (T_{A2}).
max_frac_con_cloudy	Maximum fraction of confidently cloudy pixels from METImage (T_{A3}).
max_liquid_cloud_parameter	Maximum liquid cloud parameter (T_{A4}).
max_o2_nonscat_bias ^a	Maximum difference between retrieved O ₂ from non-scattering retrieval and a-priori O ₂ for proxy retrieval. (T_{B1}).
min_frac_con_clear	Minimum fraction of confidently clear pixels from METImage (T_{C1}).
max_b9_ifov	Maximum cirrus reflectance in inner field of view from METImage (T_{C2}).
max_dif_b9	Maximum difference in cirrus reflectance between fields of view from METImage (T_{C3}).
max_stdv_b4	Maximum scene heterogeneity from METImage (T_{C4}).
max_dif_weakstring_h2o	Maximum two-band water cloud filter parameter in case backup cloud filter is used (T_{C5}).
max_dif_weakstring_ch4	Maximum two-band methane cloud filter parameter in case backup cloud filter is used (T_{C6}).
max_o2_nonscat_bias ^a	Maximum difference between retrieved O ₂ from non-scattering retrieval and a-priori O ₂ for full-physics retrieval. (T_{C7}).

^a These duplicate tag names can be distinguished by the `namelist` section. For the proxy retrieval, thresholds are discussed in section `filter_b` and the corresponding thresholds for the full-physics retrieval are presented in section `filter_c`.

B.2 Fluorescence retrieval settings

The fluorescence retrieval of RemoTeC-S5 differs in many aspects from the trace gas retrievals of the algorithm and so has its specific settings, which are described in Tab. 23.

Table 23: Fluorescence retrieval settings used in RemoTeC.

Tag	Description
fluor%reso	Internal sampling of the band (always NIR) as wavenumber interval (cm ⁻¹).
fluor%wave_start	Retrieval window starting wavenumber (cm ⁻¹).

Continued on next page



 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Date : 17 May 2019	Page 87/91
--	------------------------	---	---------------

Table 23 – Fluorescence retrieval settings - Continued from previous page

Tag	Description
fluor%wave_stop	Retrieval window ending wavenumber (cm^{-1}).
fluor%wavenumber_shift_range	Range for which spectral shifts can be fit adequately (cm^{-1}).
fluor%wavenumber_shift_sampling	Sampling for spectral shift retrieval (cm^{-1}).
fluor%order_albedo	Polynomial order for wavenumber dependence of surface albedo.
fluor%order_fluorescence	Polynomial order for wavenumber dependence of fluorescence.
fluor%order_shift	Polynomial order for wavenumber dependence of spectral shift.
fluor%niter	Number of iterations in fluorescence retrieval.
fluor%outputwave_fluorescence	Wavenumbers at which fluorescence output will be provided (cm^{-1}).

B.3 Atmospheric gridding

All trace gas retrievals share the same atmospheric grid. Therefore, settings on the atmospheric grid are listed in a specific field of the settings. These comprise of the following.

Table 24: Atmospheric settings used in RemoTeC.

Tag	Description
grid%flag	Flag for equidistant pressure grid or equidistant altitude grid.
grid%nlayers	Number of layers for profile retrieval.
grid%nrt	Number of layers in radiative transfer model per retrieval layer.
grid%natm	Number of layers for cross section calculation per radiative-transfer layer.

B.4 Trace gas retrieval settings

All trace gas retrievals (non-scattering and physics-based) have their own retrieval-specific settings. Many of the retrieval settings are specific for a retrieval window. The physics-based retrieval has three retrieval windows while the non-scattering retrievals rely only on one window. Therefore, there are three window settings for the physics-based retrieval and only one for each of the non-scattering retrievals. Besides that, each retrieval also has settings unspecific to a window. The following settings are specific to a retrieval, but not specific to a window.

Table 25: Window unspecific retrieval settings used in RemoTeC.

Tag	Description
flag%temp	Flag for fitting temperature offset.
flag%fit	Flag for fitting radiance or reflectance.
flag%oceanglint	Flag for allowing ocean glint retrievals.

For each retrieval window, the following settings are added.

Table 26: Window settings used in RemoTeC.

Tag	Description
reso	Internal sampling of the band as wavenumber interval cm^{-1} .
wave_start	Retrieval window starting wavenumber (cm^{-1}).
wave_stop	Retrieval window ending wavenumber (cm^{-1}).

Continued on next page



 	S5L2PP Methane ATBD	Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Date : 17 May 2019	Page 88/91
--	------------------------	---	---------------

Table 26 – Window settings - Continued from previous page

Tag	Description
ntau	Number of K-bins (only relevant for physics-based retrieval).
ntau_2nd_max	Number of K-bins in second dimension (only relevant for physics-based retrieval).
type_x_reg	Regularization strengths of absorbers involved in this window.
albflag	Polynomial order in which surface albedo is fitted.
iofflag	Flag for fitting intensity offset.
sps0flag	Flag for fitting spectral shift, for polynomial orders 0.
sps1flag	Flag for fitting spectral shift, for polynomial orders 1.
sps2flag	Flag for fitting spectral shift, for polynomial orders 2.

The physics-based retrieval includes atmospheric scattering for which additional settings are included.

Table 27: Additional settings for physics-based retrieval used in RemoTeC.

Tag	Description
reg	Regularization strength for aerosol parameters.
tau_ref	A-priori aerosol optical thickness at 765 nm.
alt1	A-priori aerosol height (m).
alt2	Aerosol geometric thickness (m).
rm	Real refractive index per window.
fim	Imaginary refractive index per window.
reff	A-priori size parameter.

C Usage of L1b-flagging at framework level

The S5 Level-2 prototype processor uses the flag information in the L1b product as follows:

1. For processing-critical flags that apply to a ground pixel, an error flag for the ground pixel will be set in the L2 output product, and processing for the ground pixel is skipped.
2. For flags that apply to a ground pixel that potentially impact processing in a non-critical way, a warning flag for the ground pixel is set in the L2 output product, and processing for the ground pixel is performed.
3. For processing-critical flags that apply to a spectral pixel, the processing will assume a missing value for the spectral pixel and apply an algorithm specific evaluation of the missing values for each ground pixel
4. Flags that do not impact processing are ignored.

Table 28 lists details on the L1b flag usage (note that the .

Flag name	Possible values	Meaning	Action	L2 flag-ging
ground_pixel_quality	See below	Quality assessment information for each ground pixel		
	no_error	No error	Use pixel	
	solar_eclipse	Potentially affected by solar eclipse	Use pixel	Warning
	sun_glint_possible	Potentially affected by sun glint	Use pixel	Warning
	ascending	Pixel measured in ascending node (dlatitude/dt > 0)	Use pixel	
	night	Pixel measured on night side (SZA > 90)	Use pixel	
	geo_boundary_crossing	Pixel crosses a geo-boundary	Use pixel	
measurement_quality	geolocation_error	Error in geolocation assignment	Do not use pixel	Error
	See below	Overall quality information for a single measurement i.e. one value for each scanline		
	no_error	No error	Use pixel	
	proc_skipped	Processing skipped	Do not use pixel	Error
	no_residual	Undefined meaning (leftover from TROPOMI)	Use pixel	
	saa	measurement in South Atlantic Anomaly	Use pixel	Warning
	spacecraft_manoeuvre	Measurement during spacecraft manoeuvre	Use pixel	Warning
spectral_channel_quality	sub_grp	Undefined meaning (leftover from TROPOMI)	Use pixel	
	irr_out_range	Irradiance out of range	Do not use pixel	Error
	sub_group	Undefined meaning (leftover from TROPOMI)	Use pixel	
	See below	Quality assessment information for each spectral pixel	*	*
	no_error	No error	Use pixel	
missing	Missing pixel	*	*	

detector_column_ - qualification	bad_pixel	Bad pixel	*	*
	processing_error	Processing error	*	*
	saturated	Saturated pixel	*	*
	transient	Transient pixel	*	*
	rts	Pixel affected by RTS	*	*
	underflow	Underflow error	*	*
	See below	Qualification flag for a full column indicating detector column type or state		
	no_qualification	(Normal column)	Use pixel	
	skipped	Skipped column	Do not use pixel (i.e. ground pixel that includes this column)	Error
	uvn_odd	Odd Column number	Use pixel	
uvn_prepost	Pre/postscan column	Do not use pixel (i.e. ground pixel that includes this column)	Error	
uvn_overscan	Overscan column	Do not use pixel (i.e. ground pixel that includes this column)	Error	
detector_row_ qualification	swir_adc0	SWIR ADC0 used (TBC)	Use pixel	
	swir_adc1	SWIR ADC1 used (TBC)	Use pixel	
	swir_adc2	SWIR ADC2 used (TBC)	Use pixel	
	swir_adc3	SWIR ADC3 used (TBC)	Use pixel	
	See below	Qualification flag indicating row type or state.	**	**
	no_qualification	(Normal row)	Use pixel	
	uvn_ror	Read-out register row	**	**
	uvn_dump	Dump row	**	**
	uvn_covered	Covered row	**	**
	uvn_overscan	Overscan row	**	**
uvn_higain	Hi-gain row	TBD	TBD	
swir_reference	SWIR reference row	**	**	
gen_transition	Transition row	**	**	
gen_non_illuminated	Non illuminated row	**	**	
spectral_calibration_ - quality	Several	Spectral calibration quality assessment information for each ground pixel.	Do not use spectral calibration in case any flag set, instead use nominal calibration	Warning
quality_level	0-1	Overall quality assessment information for each spectral pixel	*Use pixel in case quality exceeds configurable threshold	*

 	<p style="text-align: center;">S5L2PP Methane ATBD</p>	<p>Reference : SRON-ESA-S5L2PP-ATBD-001 Version : 3.1 Page Date : 17 May 2019 91/91</p>
--	--	---

overall_quality_flag	Several values	Overall quality flag for product	Use pixels in product	Copy quality flag to output
-----------------------------	----------------	----------------------------------	-----------------------	-----------------------------

Table 28: Usage of L1b flags by algorithm

*: Setting of this flag implies a “missing” spectral pixel that is excluded from processing. The following rules apply for processing of a ground pixel in relation to one or more missing spectral pixels: **: Flagged columns are treated as a set of missing spectral pixels for each ground pixel in the corresponding scanline. Warning and error criteria for the occurrence of such missing spectral pixels have been defined in the previous foot note.

RemoTeC does not apply any filtering on the amount of consecutive missing spectral pixel but uses a posteriori filter criteria for overall data quality control.