Algorithm Theoretical Baseline Document for Sentinel-5 Precursor: HDO/H2O Total Column Retrieval

Jochen Landgraf, Tobias Borsdorff, Soumyajit Mandal
Andreas Schneider, Joost aan de Brugh, Remco A. Scheepmaker

document number : SRON-ESG-RP-2024-020
CI identification : CI-7430-ATBD
issue : 1.0.0
date : 2025-01-15
status : released

Document approval record

	digital signature
prepared:	Soumyajit Mandal Digitally signed by Soumyajit Mandal Date: 2025.01.16 11:10:22 +01'00'
checked:	
approved PM:	Digitally signed by Tobias Borsdorff Date: 2025.01.16 12:08:36 +01'00'
approved PI:	

Document change record

issue	date	item	comments
1.0.0	2025-01-15	All	Initial draft version

Contents

Document approval record 2
Document change record 3
List of Tables 5
List of Figures 6
1 Introduction 7
1.1 Identification 7
1.2 Purpose and objectives 7
1.3 Document overview 7
2 Applicable and reference documents 8
2.1 Applicable documents 8
2.2 Standard documents 8
2.3 Reference documents 8
2.4 Electronic references 10
3 Terms, definitions and abbreviated terms 11
3.1 Acronyms and abbreviations 12
4 Remote Sensing of HDO/H2O 14
5 Algorithm description 15
6 High level data product description 16
7 Performance and error analysis 18
7.1 Case A: cloud free conditions 21
7.2 Case B: clouds 22
7.3 Case C: photon trapping 24
7.4 Case D: aerosol 25
7.5 Case E: ice clouds 26
7.6 Conclusions 27
8 Sensitivity analysis 31
8.1 A priori HDO profile 31
8.2 A priori H_2O profile 32
8.3 A priori CH_4 profile 32
8.4 Temperature profile 32
8.5 Pressure profile 32
8.6 Radiometric offset 32
8.7 Slit function 32
8.8 Use of different H_2O spectroscopy 33
8.9 Spectral sampling 33
8.10 Conclusions 34
9 Feasibility 35
10 Validation 36
10.1 TCCON and NDACC-MUSICA 36
10.2 GOSAT 36
A Appendix: SWIR Pre-Processing 38

List of Tables

1 Overview of the HDO/H_2O data product. 17
2 Microphysical properties of water and ice clouds: $n(r)$ represents the size distribution type, r_{eff} and v_{eff} are the effective radius and variance of the size distribution, $n = n_r - in_i$ is the refractive index.The ice cloud size distribution follows a power-law distribution as proposed by [RD1]. 18
3 Summary of the different generic test cases A-F. 19
4 Summary of the sensitivity study. Mean systematic errors and standard deviations are derived for the surface albedos 0.03, 0.05 and 0.075 (45 scenes for case A and 4 scenes for case C after filtering) 31
5 Sensitivity to a change of water line parameters from HITRAN2008 to the parameters from Scheepmaker et al. (2013). 33
6 Mean systematic errors as a function of spectral sampling resolution. 34
7 HDO/H_2O computation times for two different resolution settings and fortran compilers. 35

List of Figures

1 Simulated reflectance spectrum for cloud free conditions, a surface albedo of 0.6 and SZA=$0°$, showing only the contributions from the water isotopes $H_2{}^{16}O$, HDO and $H_2{}^{18}O$. The window used for the HDO/H_2O retrieval is indicated with the red lines. 15
2 Overview of the HDO/H_2O processing scheme. 16
3 Left: atmospheric concentration profiles (bottom axis) and temperature profile (top axis) used as input for the model atmosphere. The concentrations are normalised to the concentration at ground level. Right: assumed profiles for the amount of HDO depletion (solid line, lower axis) and $H_2{}^{18}O$ depletion (dashed line, top axis). 19
4 Cloud free retrieval bias (Case A) for the total columns of H_2O (top left), HDO (top right) and δD (bottom right). The bottom left panel shows the relative difference in total CH_4 column between the weak and strong bands from the pre-processor two-band CH_4 cloud filter. 21
5 Statistical error estimates for Case A. Top left: H_2O column ($\sigma_{\mathrm{H_2O}}$). Top right: HDO column (σ_{HDO}). Bottom left: CH_4 column ($\sigma_{\mathrm{CH_4}}$). Bottom right: δD ($\sigma_{\delta\mathrm{D}}$). 22
6 Total column averaging kernels for H_2O (solid lines), HDO (dashed lines) and CH_4 (dash-dotted lines). Left: for a fixed SZA of $30°$. The colours indicate different surface albedos from 0.03 (blue) to 0.6 (red). Right: for fixed surface albedo of 0.03. The colours indicate different SZAs from $0°$ (blue) to $70°$ (red). 23
7 Retrieval bias as a function of cloud top height and cloud fraction (Case B, assuming $\tau_{\mathrm{cld}} = 5$) for the total columns of H_2O (top left), HDO (top right) and δD (bottom right). The bottom left panel shows the relative difference in total CH_4 column between the weak and strong bands from the pre-processor two-band CH_4 cloud filter. 23
8 Total column averaging kernels for H_2O (solid lines) and HDO (dashed lines) for a simulated atmosphere that contained a 1 km thick, $\tau_{\mathrm{cld}} = 50$, $z_{\mathrm{cld}} = 5$ km cloud layer. The colours indicate different cloud fractions from 0.0 (blue) to 1.0 (red). 24
9 Same as Figure 7, but now with the δD bias filtered against cloud contaminated data using a two-band methane cloud filter threshold of 6%. Note that this filter allows for all cloud fractions at a cloud top height of 1 km, and all cloud-free scenes (the coloured contours in the bottom-left panel), even though these scenes are not visible in the bottom-right panel. 25
10 Statistical error estimates for Case B, assuming $\tau_{\mathrm{cld}} = 50$. Top left: H_2O column ($\sigma_{\mathrm{H_2O}}$). Top right: HDO column (σ_{HDO}). Bottom left: CH_4 column ($\sigma_{\mathrm{CH_4}}$). Bottom right: δD ($\sigma_{\delta\mathrm{D}}$). 26
11 Retrieval bias as a function of surface albedo and cloud fraction (Case C, assuming $\tau_{\mathrm{cld}} = 2$ and $z_{\mathrm{cld}} = 5$ km) for the total columns of H_2O (top left), HDO (top right) and δD (bottom right). The bottom left panel shows the threshold values from the two-band CH_4 cloud filter. The δD bias has been masked against a threshold of 6%. 27
12 Retrieval bias as a function of surface albedo and aerosol optical thickness (Case D, assuming a sulfate-type aerosol layer in the boundary layer between 0–2 km) for the total columns of H_2O (top left), HDO (top right) and δD (bottom right). The bottom left panel shows the threshold values from the two-band CH_4 cloud filter. The δD bias has been masked against a threshold of 6%. 28
13 Same as Figure 12, but now for an urban-type aerosol layer between 4–5 km. 29
14 Retrieval bias as a function of surface albedo and cirrus optical thickness (Case E) for the total columns of H_2O (top left), HDO (top right) and δD (bottom right). The bottom left panel shows the threshold values from the two-band CH_4 cloud filter. The δD bias has been masked against a threshold of 6%. 30
15 Overall structure of the SWIR preprocessor. 38
16 CH_4 error of a non-scattering retrieval from the SWIR 2315–2324 nm spectral window for a water cloud with optical thickness of 5 as function of cloud height and cloud fraction (left panel, for more details see generic scenario B in Sec. 7) and for a cirrus cloud at 10 km height as function of surface albedo and cirrus optical thickness (right panel, for more details see generic scenario E in Sec. 7). 40
17 CH_4 two-band cloud filter for the cloud scenarios of Fig. 16. The methane cloud filter relies on non-scattering methane column retrieval from strong and weak absorption features at 2363-2373 nm and 2310-2315 nm, respectively. 41
18 H_2O two-band cloud filter for the cloud scenarios of Fig. 16. The filter relies on non-scattering methane column retrieval from strong and weak absorption features at 2367-2377 nm and 2329-2334 nm, respectively. 41

1 Introduction

1.1 Identification

This document describes the HDO/H_2O column retrieval algorithm from Sentinel-5 Precursor (S5P) measurements in the shortwave infrared (SWIR) spectral range between 2310 and 2340 nm. It is one of the deliverables of the ESA project 'Sentinel-5 P level 2 processor development' [AD1].

1.2 Purpose and objectives

The purpose of the document is to describe the theoretical baseline of the algorithm that is used for the operational processing of the HDO/H_2O column densities from S5P measurements in the SWIR spectral range. Input, output and ancillary data are described. Additionally, the performance of the algorithm is analyzed with respect to the expected calculation times and the data product uncertainty.

1.3 Document overview

The document is structured as follows: After this introduction, references are provided in Sec. 2 and Sec. 3 contains a list of abbreviations used in this document. Sec. 5 we describe how the HDO/$H_2{}^{16}O$ algorithm has been changed compared to the CO algorithm. The high level data product is described in Sec. 6. An analysis of the performance and estimated errors for various simulated measurement conditions is presented in Sec. 7, followed by a sensitivity analysis in Sec. 8. We discuss the computational effort of the algorithm in Sec. 9 and conclude with a validation plan in Sect. 10.

2 Applicable and reference documents

2.1 Applicable documents

[AD1] Sentinel-5P Level 2 Processor Development – Statement of Work.
source: ESA; **ref:** S5P-SWESA-GS-053; **date:** 2012.

2.2 Standard documents

[SD1] Space Engineering – Software.
source: ESA; **ref:** ECSS-Q-ST-80C; **date:** 2009.

[SD2] Space Product Assurance – Software Product Assurance.
source: ESA; **ref:** ECSS-E-ST-40C; **date:** 2009.

2.3 Reference documents

[RD1] A. J. Heymsfield and C. M. R. Platt; A Parameterization of the Particle Size Spectrum of Ice Clouds in Terms of the Ambient Temperature and the Ice Water Content. *J. Atmos. Sci.*; **41** (1984), 846; doi:10.1175/1520-0469(1984)041<0846:APOTPS>2.0.CO;2.

[RD2] Terms, definitions and abbreviations for TROPOMI L01b data processor.
source: KNMI; **ref:** S5P-KNMI-L01B-0004-LI; **date:** 2011.

[RD3] Terms and symbols in the TROPOMI algorithm team.
source: KNMI; **ref:** SN-TROPOMI-KNMI-049; **date:** 2012.

[RD4] I. M. Held and B. J. Soden; Robust Responses of the Hydrological Cycle to Global Warming. *Journal of Climate*; **19** (2006), 5686; doi:10.1175/JCLI3990.1.

[RD5] W. Dansgaard; Stable isotopes in precipitation. *Tellus*; **16** (1964) (4), 436.

[RD6] W. Dansgaard, S. J. Johnsen, J. Møller *et al.*; One Thousand Centuries of Climatic Record from Camp Century on the Greenland Ice Sheet. *Science*; **166** (1969) (3903), 377; doi:10.1126/science.166.3903.377.

[RD7] J. Jouzel, R. B. Alley, K. M. Cuffey *et al.*; Validity of the temperature reconstruction from water isotopes in ice cores. *J. Geophys. Res.*; **102** (1997), 26471; doi:10.1029/97JC01283.

[RD8] J. R Petit, J Jouzel, D Raynaud *et al.*; Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*; **399** (1999), 429; doi:10.1038/20859.

[RD9] D. H. Ehhalt, F. Rohrer and A. Fried; Vertical profiles of HDO/H2O in the troposphere. *J. Geophys. Res.*; **110** (2005), 13301; doi:10.1029/2004JD005569.

[RD10] M.J. Kurylo and S. Solomon; *Network for the Detection of Stratospheric Change: A Status and Implementation Report* (NASA, Upper Atmosphere Research Program and NOAA Climate and Global Change Program, Washington DC, 1990).

[RD11] 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; doi:10.1098/rsta.2010.0240.

[RD12] M. Schneider, S. Barthlott, F. Hase *et al.*; Ground-based remote sensing of tropospheric water vapour isotopologues within the project MUSICA. *Atmos. Meas. Tech. Disc.*; **5** (2012) (4), 5357; doi:10.5194/amtd-5-5357-2012.

[RD13] Vyacheslav I Zakharov, Ryoichi Imasu, Konstantin G Gribanov *et al.*; Latitudinal distribution of the deuterium to hydrogen ratio in the atmospheric water vapor retrieved from IMG/ADEOS data. *Geophys. Res. Lett.*; **31** (2004), 12104; doi:10.1029/2004GL019433.

[RD14] J. Worden, D. Noone, K. Bowman *et al.*; Importance of rain evaporation and continental convection in the tropical water cycle. *Nature*; **445** (2007), 528; doi:10.1038/nature05508.

[RD15] H Herbin, D Hurtmans, C Clerbaux *et al.*; H216O and HDO measurements with IASI/MetOp. *Atmos. Chem. Phys.*; **9** (2009), 9433.

[RD16] C. Frankenberg, K. Yoshimura, T. Warneke *et al.*; Dynamic Processes Governing Lower-Tropospheric HDO/H2O Ratios as Observed from Space and Ground. *Science*; **325** (2009), 1374; doi:10.1126/science.1173791.

[RD17] C. Frankenberg, D. Wunch, G. Toon *et al.*; Water vapor isotopologue retrievals from high-resolution GOSAT shortwave infrared spectra. *Atmospheric Measurement Techniques*; **6** (2013) (2), 263; doi:10.5194/amt-6-263-2013. URL `http://www.atmos-meas-tech.net/6/263/2013/`.

[RD18] H. Boesch, N. M. Deutscher, T. Warneke *et al.*; HDO/H_2O ratio retrievals from GOSAT. *Atmos. Meas. Tech.*; **6** (2013), 599; doi:10.5194/amt-6-599-2013.

[RD19] K. Yoshimura, C. Frankenberg, J. Lee *et al.*; Comparison of an isotopic atmospheric general circulation model with new quasi-global satellite measurements of water vapor isotopologues. *J. Geophys. Res.-Atmos.*; **116** (2011), D19118; doi:10.1029/2011JD016035.

[RD20] C. Risi, D. Noone, J. Worden *et al.*; Process-evaluation of tropospheric humidity simulated by general circulation models using water vapor isotopologues: 1. Comparison between models and observations. *J. Geophys. Res.-Atmos.*; **117** (2012), D05303; doi:10.1029/2011JD016621.

[RD21] C. Risi, D. Noone, J. Worden *et al.*; Process-evaluation of tropospheric humidity simulated by general circulation models using water vapor isotopic observations: 2. Using isotopic diagnostics to understand the mid and upper tropospheric moist bias in the tropics and subtropics. *J. Geophys. Res.-Atmos.*; **117** (2012), D05304; doi:10.1029/2011JD016623.

[RD22] 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. Envirom.*; **120** (2012), 70.

[RD23] GMES Sentinels 4 and 5 mission requirements document.
source: ESA; **ref:** EOP-SMA/1507/JL-dr; **date:** 2011.

[RD24] Algorithm Theoretical Baseline Document for Sentinel-5 Precursor: Carbon Monoxide Total Column Retrieval.
source: SRON; **ref:** SRON-S5P-LEV2-RP-002; **issue:** 2.4.0; **date:** 2022-07-12.

[RD25] J.M.J. Schepers, D.and aan de Brugh, Ph. Hahne, A. Butz *et al.*; LINTRAN v2.0: A linearised vector radiative transfer model for efficient simulation of satellite-born nadir-viewing reflection measurements of cloudy atmospheres. *J. Quant. Spectrosc. Radiat. Transfer*; **149** (2015), 247.

[RD26] S. Chandrasekhar; *Radiative transfer* (Dover Publications, Inc., New York, 1960).

[RD27] M. Hess and M. Wiegner; COP: a data library of optical properties of hexagonal ice crystals. *Appl. Opt.*; **33** (1994) (33), 7740; doi:10.1364/AO.33.007740.

[RD28] M. Hess, R. B. A. Koelemeijer and P. Stammes; Scattering matrices of imperfect hexagonal ice crystals. *Journal of Quantitative Spectroscopy and Radiative Transfer*; **60** (1998), 301 ; doi:DOI: 10.1016/S0022-4073(98)00007-7.

[RD29] A Marshak, A. Davis, W. Wiscombe *et al.*; The Verisimilitude of the Independent Pixel Approximation Used in Cloud Remote Sensing. *Remote Sens. Envirom.*; **52** (1995), 71.

[RD30] Instrument noise model for the Sentinel 5 SWIR bands.
source: Netherlands Insitute for Space Research, SRON; **ref:** SRON-TROPSC-TN-2011-002; **date:** 2011.

[RD31] U.S. Standard Atmosphere, 1976.
source: National Oceanic and Atmospheric Administration; **ref:** NOAA-S/T76-1562.

[RD32] Observation Techniques and Mission Concepts for Atmospheric Chemistry (CAMELOT).
source: European Space Agency; **ref:** 20533/07NL/HE.

[RD33] D. H. Ehhalt; Vertical profiles of HTO, HDO, and H2O in the troposphere. *NCAR Tech. Note NCAR-TN-STR-100*; (1974).

[RD34] M. Schneider, G. C. Toon, J.-F. Blavier *et al.*; H_2O and δD profiles remotely-sensed from ground in different spectral infrared regions. *Atmos. Meas. Tech.*; **3** (2010), 1599; doi:10.5194/amt-3-1599-2010.

[RD35] H. Craig; Isotopic Variations in Meteoric Waters. *Science*; **133** (1961), 1702; doi:10.1126/science.133.3465.1702.

[RD36] A. Galli, A. Butz, R. A. Scheepmaker *et al.*; CH_4, CO, and H_2O spectroscopy for the Sentinel-5 Precursor mission: an assessment with the Total Carbon Column Observing Network measurements. *Atmos. Meas. Tech.*; **5** (2012), 1387; doi:10.5194/amt-5-1387-2012.

[RD37] C. Frankenberg, P. Bergamaschi, A. Butz *et al.*; Tropical methane emissions: A revised view from SCIAMACHY onboard ENVISAT. *Geophys. Res. Lett.*; **35** (2008), 15811; doi:10.1029/2008GL034300.

[RD38] O. Schneising, M. Buchwitz, J. P. Burrows *et al.*; Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite - Part 2: Methane. *Atmos. Chem. Phys.*; **9** (2009), 443.

[RD39] 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.

[RD40] 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/H_2O measurements. *Atmos. Meas. Tech.*; **6** (2013) (4), 879; doi:10.5194/amt-6-879-2013. URL `http://www.atmos-meas-tech.net/6/879/2013/`.

[RD41] L. S. Rothman, I. E. Gordon, A. Barbe *et al.*; The HITRAN 2008 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transfer*; **110** (2009), 533; doi:10.1016/j.jqsrt.2009.02.013.

[RD42] J. M. Krijger, M. van Weele, I. Aben *et al.*; Technical Note: The effect of sensor resolution on the number of cloud-free observations from space. *Atmos. Chem. Phys.*; **7** (2007) (11), 2881; doi:10.5194/acp-7-2881-2007.

[RD43] Algorithm Theoretical Baseline Document for Sentinel-5 Precursor methane retrieval.
source: SRON; **ref:** SRON-S5P-LEV2-RP-001; **date:** 2014.

[RD44] F. Kasten and T. Young; Revised optical air mass tables and approximation formula. *Appl. Opt.*; **28** (1989), 4735.

2.4 Electronic references

There are no electronic references

3 Terms, definitions and abbreviated terms

Terms, definitions and abbreviated terms that are used in the development program for the TROPOMI L0 1b data processor are described in [RD2]. Terms, definitions and abbreviated terms that are used in development program for the TROPOMI L2 data processors are described in [RD3]. Terms, definitions and abbreviated terms that are specific for this document can be found below.

3.1 Acronyms and abbreviations

ADEOS	Advanced Earth Observing System
AIRS	Atmospheric Infrared Sounder
AOT	Aerosol Optical Thickness
AQUA	A NASA Earth Science satellite mission focussing on the Earth's water cycle
CTM	Chemical Transport Model
DFS	Degree of Freedom for Signal
S-LINTRAN	Scalar linearised Radiative Transfer Program for a Multi-Layered Plane-Parallel Medium
ECMWF	European Centre for Medium-Range Weather Forecasts
ERI	European Research Institute
ESRL	Earth System Research Laboratory
FFD	Fourier filter destriping
FFM	Fixed mask destriping
FTS	Fourier Transform Spectrometer
FTIR	Fourier Transform Infrared
FRESCO	Fast Retrieval Scheme for Clouds from the Oxygen A band
FWHM	Full Width Half Maximum
GCM	General Circulation Model
GMES	Global Monitoring for Environment and Security
GNIP	Global Network for Isotopes in Precipitation
GOSAT	Greenhouse gases Observing Satellite
IAGOS	In-service Aircraft for a Global Observing System
IASI	Infrared Atmospheric Sounding Interferometer
IMAP	Iterative Maximum A Posteriori
IMG	Interferometric Monitor for Greenhouse gases
IMLM	Iterative Maximum Likelihood Method
IRWG	Infrared Working Group
ISRF	Instrument Spectral Response Function
L1	Level-1
L2	Level-2
LER	Lambert-equivalent Reflectivity
LOS	Line of Sight
MACC	Monitoring Atmospheric Composition and Climate
MAPS	Measurement of Air Pollution from Satellites
MODIS	Moderate Resolution Imaging Spectroradiometer
MOPITT	Measurements of Pollution in the Troposphere
MOZAIC	Measurement of Ozone and Water Vapour on Airbus in-service Aircraft
NDACC	Network for the Detection of Atmospheric Composition Change
NOAA	National Oceanic and Atmospheric Administration
NPP	National Polar-orbiting Partnership
NRT	Near Real Time
OMI	Ozone Monitoring Instrument
PIFM	Practical Improved Flux Method
RemoTeC	Remote Sensing of Greenhouse Gases for Carbon Cycle Modelling
RMS	Root Mean Square

S5P	Sentinel-5 Precursor
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Chartography
SEOM-IAS	Scientific Exploitation of Operational Missions - Improved Atmospheric Spectroscopy Databases
SICOR	Shortwave Infrared CO Retrieval
SMOW	Standard Mean Ocean Water
SNR	Signal-to-Noise Ratio
SPEC	Standard Performance Evaluation Corporation
SSD	Spectral Sampling Distance
SWIR	Shortwave Infrared
SZA	Solar Zenith Angle
TCCON	Total Carbon Column Observing Network
TES	Tropospheric Emission Spectrometer
TM4	Transport Model 4
TM5	Transport Model 5
TOA	Top Of model Atmosphere
TROPOMI	Tropospheric Monitoring Instrument
VIIRS	Visible Infrared Imager Radiometer Suite
VZA	Viewing Zenith Angle
WFM-DOAS	Weighting Function Modified-Differential Optical Absorption Spectroscopy

4 Remote Sensing of HDO/H2O

The hydrological cycle is a key element in our understanding of climate change. Being the strongest natural greenhouse gas, water vapour plays an important role in atmospheric feedback mechanisms and related processes such as cloud formation [RD4]. Atmospheric general circulation models (GCMs) are used to model such processes and come to global climate change projections. The validity of such projections depends on the correct understanding of many interacting processes that control atmospheric humidity. This is where measurements of the isotopic composition of water vapour can play a crucial role, as many processes and different source regions leave a distinct isotopic signature on the water vapour, while their impact on the total humidity might be the same. Different water isotopologues, such as $H_2{}^{16}O$, $HD^{16}O$ (which will be denoted HDO in the following) and $H_2{}^{18}O$ have different equilibrium vapour pressures, which leads to a temperature dependent isotope fractionation, whenever phase changes occur [RD5]. The ratio $HDO/H_2{}^{16}O$ of water vapour is therefore dependent on the source region's location and temperature and the entire transport history of the air parcel, including all evaporation, condensation and mixing events. This makes measurements of the ratio $HDO/H_2{}^{16}O$ a valuable benchmark for the evaluation and further development of GCMs.

The first atmospheric measurements of $HDO/H_2{}^{16}O$ were performed in situ, e.g. from trapped air in ice cores [RD6, RD7, RD8], aircraft measurements [RD9] and precipitation (e.g. the Global Network for Isotopes in Precipitation (GNIP)). These measurements are sparse and infrequent, or measure the composition of water vapour indirectly (e.g. after a final condensation process in the case of precipitation measurements). Remote-sensing has made it possible to measure $HDO/H_2{}^{16}O$ in water vapour on a more frequent basis, like with ground-based networks of Fourier-Transform Spectrometers [RD10, RD11, RD12]. The first satellite-based measurements of $HDO/H_2{}^{16}O$ in the middle to high troposphere were performed by the Interferometric Monitor for Greenhouse gases (IMG) [RD13]), the Tropospheric Emission Spectrometer (TES) [RD14] and the Infrared Atmospheric Sounding Interferometer (IASI) [RD15]. The Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) was the first instrument to measure $HDO/H_2{}^{16}O$ almost globally with a high sensitivity near the surface [RD16], followed recently by the Greenhouse Gases Observing Satellite (GOSAT) [RD17, RD18]. The first comparisons of such datasets with GCMs have been performed, and show the clear potential of global measurements of $HDO/H_2{}^{16}O$ to further our understanding of the hydrological cycle [RD16, RD19, RD20, RD21].

With the TROPOMI instrument, we have the opportunity to extend and improve the existing global $HDO/H_2{}^{16}O$ datasets. Like SCIAMACHY, the TROPOMI SWIR measurements will benefit from a high sensitivity to $HDO/H_2{}^{16}O$ near the surface. Here we describe how we have adapted the TROPOMI CO retrieval algorithm to perform retrievals of $HDO/H_2{}^{16}O$. The high spatial resolution of TROPOMI, in combination with a wide swath, will ensure a large amount of data over cloud-free ground pixels. Due to the large difference in atmospheric abundance between HDO and $H_2{}^{16}O$, the measurement sensitivity, reflected in the averaging kernels, will be very different for HDO and $H_2{}^{16}O$. This makes the accurate retrieval of their ratio very challenging under conditions with elevated scattering layers such as clouds. We will describe how we use the non-scattering mode of the CO retrieval algorithm and pre-filter for the strongest cloudy conditions to reduce processing time. Due to the smaller ground pixels and shorter revisit times, TROPOMI increases the amount of near-surface cloud-free $HDO/H_2{}^{16}O$ data, which allows for studying spatial and temporal gradients of $HDO/H_2{}^{16}O$ with higher resolution. Like the CO algorithm, the $HDO/H_2{}^{16}O$ algorithm provides the column averaging kernels of both HDO and $H_2{}^{16}O$, which are essential for a proper comparison of the data to the isotope-enabled GCMs.

HDO/H2O level-2 requirements

To improve our present knowledge on HDO/H2O on a global scale, satellite measurements of the total HDO/H2O column are needed within an accuracy of $< 20\%$ and a precision with ≤ 10 % even for accounting for background abundance and low surface reflection in the shortwave infrared spectral range [RD22, RD23]. For the error budget, we assume that instrument and forward model errors contribute equally to the error budget and that all error terms add up quadratically. Herewith, both instrument and forward model errors must not exceed 8 %. These level-2 requirements should be considered as thresholds.

5 Algorithm description

The HDO/$H_2{}^{16}O$ retrieval algorithm is adapted from the CO retrieval approach outlined in the Carbon Monoxide ATBD (Algorithm Theoretical Basis Document, [RD24]). It has been updated to account for an extended scientific HDO/$H_2{}^{16}O$ total column data product from short-wave infrared (SWIR) measurements by the Tropospheric Monitoring Instrument (TROPOMI), including both clear-sky and cloudy scenes. In contrast to the CO full-physics retrieval approach, which relies on the retrieval of effective scattering parameters from the SWIR spectrum, the HDO/$H_2{}^{16}O$ algorithm employs a scattering-aware retrieval. The algorithm simultaneously infers trace gas column information, surface properties, and effective cloud parameters from the observations. This approach significantly enhances coverage compared to the previous clear-sky-only data product, particularly by including scenes over low clouds, which enables data retrieval over oceans where the albedo in the SWIR spectral range is too low for cloud-free conditions.

Key retrieval configurations include:

- Retrieval window: 2354–2374 nm (SWIR band).
- The total column density of the water isotopes $H_2{}^{16}O$, HDO, and $H_2{}^{18}O$ are fitted as separate species jointly with the CH_4 column density using the profile scaling approach.
- A higher internal sampling (0.01 cm^{-1}) is used to properly resolve the different absorption lines by line-by-line spectral simulations.

The HDO/$H_2{}^{16}O$ retrieval is implemented using the existing SICOR algorithm, and the data processing is foreseen as part of the offline (reprocessing) mode of the Sentinel-5 Precursor mission. Unlike the earlier version, which focused on clear-sky observations using strict cloud filters from the SWIR pre-processor, the updated algorithm uses a scattering-aware retrieval approach. This approach fits total columns of H_2O, HDO, $H_2{}^{18}O$, CH_4, and CO, together with the surface albedo, its linear dependence on wavelength, a spectral offset, and an offset on the reflectance. The inversion uses the same standard least squares approach as the CO prefit, in combination with the transmission radiative transfer model (see Appendix A for the details). The selection of the 2354–2374 nm retrieval window is based on the presence of relatively strong HDO absorption lines, which are at least partly separated from the stronger $H_2{}^{16}O$ lines.

In Fig. 1, we show a simulated reflectance spectrum for TROPOMI resolution at cloud-free conditions, showing only the contributions from the water isotopes $H_2{}^{16}O$, HDO, and $H_2{}^{18}O$.

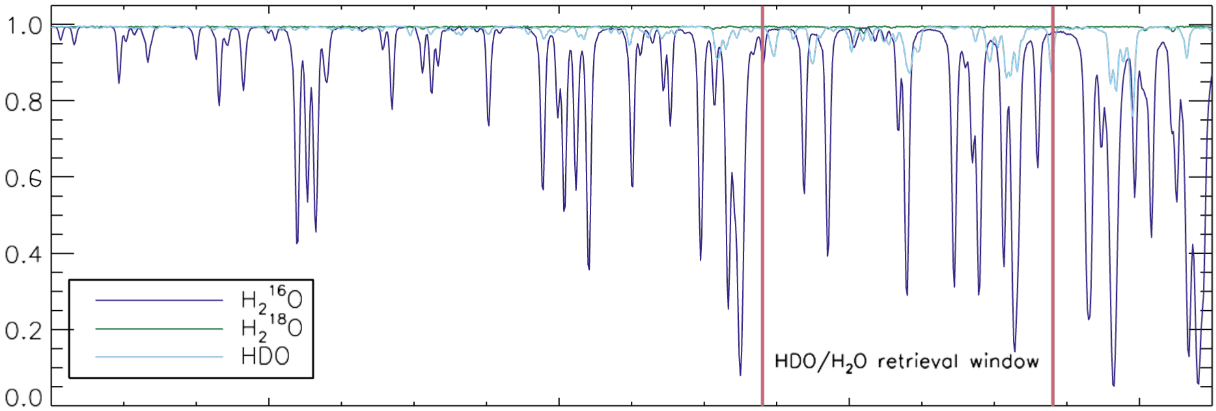

Figure 1: Simulated reflectance spectrum for cloud free conditions, a surface albedo of 0.6 and SZA=$0°$, showing only the contributions from the water isotopes $H_2{}^{16}O$, HDO and $H_2{}^{18}O$. The window used for the HDO/H_2O retrieval is indicated with the red lines.

The water isotopes $H_2{}^{16}O$ and $H_2{}^{18}O$ are fitted as separate species in order to take the ratio HDO/$H_2{}^{16}O$ without any interference from $H_2{}^{18}O$ absorption lines. From here on, we always assume that this separation has been made, and whenever we refer to "H_2O" we mean the primary isotope $H_2{}^{16}O$.

An overview of the HDO/H_2O processing scheme is shown in Fig. 2. The HDO/H_2O retrieval will be processed in offline mode (compared to near real-time mode for CO) with a higher internal sampling of 0.01 cm^{-1} (compared to 0.05 cm^{-1} for CO). As a result of the higher resolution, the processing time per spectrum is higher than the other non-scattering prefit retrievals. This is compensated by using the two-band methane cloud product from the SWIR pre-processor (see Appendix A) to pre-filter against clouds and aerosol. Therefore, only a small fraction of all the measurements will be retrieved. We will study the performance of the retrieval under the presence of clouds in Sec. 7.2, where we will also define preliminary cloud filter criteria. The computational effort of the retrieval will be discussed in Sec. 9.

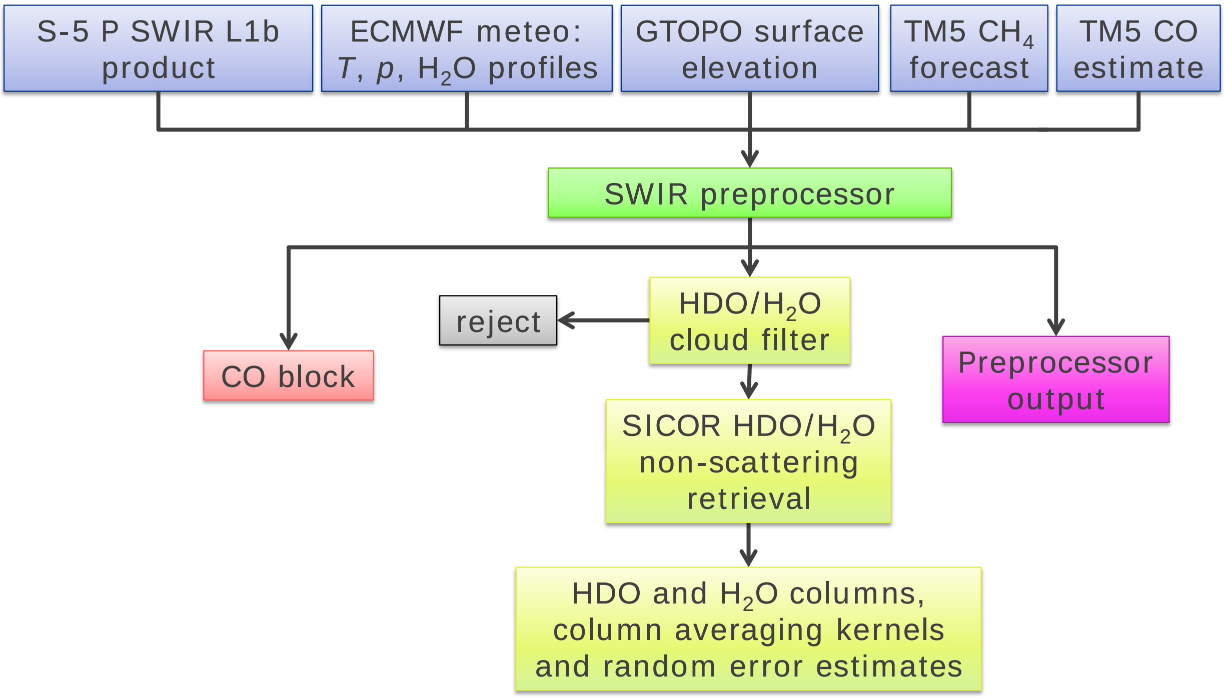

Figure 2: Overview of the HDO/H_2O processing scheme.

6 High level data product description

The output of the HDO/H_2O retrieval software will be analogous to the output of the CO retrieval software and Table 1 summarises the high level data product. To facilitate a posteriori cloud filtering (see Sec. 7.2), we also include the a priori, vertically integrated, H_2O and CH_4 column densities smoothed by the averaging kernels. Moreover, we provide the retrieved ratio HDO/H_2O in δD notation, relative to the Standard Mean Ocean Water (SMOW) abundance ratio of $R_s = 3.1153 \cdot 10^{-4}$:

$$\delta\mathrm{D}_{\mathrm{output}} = \left(\frac{c_{\mathrm{HDO}}}{c_{\mathrm{H2O}}} \frac{1}{R_s} - 1 \right) \cdot 1000\,‰. \tag{1}$$

Here, c_{HDO} and c_{H2O} are the retrieved total column densities of HDO and H_2O, respectively. Table 1 presents a list of all the variables that will be included in the HDO/H_2O data product (it is possible that this list will be extended in the future).

variable	description	data type	# of entries	units
LAT	latitude coordinates (centre, corners)	float	5	degree
LON	longitude coordinates (centre, corners)	float	5	degree
SZA	solar zenith angle at pixel centre	float	1	degree
VZA	viewing zenith angle at pixel centre	float	1	degree
AZI	azimuth angle relative to solar line of sight	float	1	degree
ALT	surface altitude	float	1	meter
TIME	Modified Julian Day (days since 1.1.2000 00:00)	double	1	days
HDO_VCD	vertically integrated HDO column density	float	1	molec/cm^2
HDO_VCD_SIGMA	retrieval noise on HDO_VCD (1-sigma error)	float	1	molec/cm^2
HDO_AK	HDO column averaging kernel	float	50	TBD
H2O_VCD	vertically integrated H_2O column density	float	1	molec/cm^2
H2O_VCD_SIGMA	retrieval noise on H2O_VCD (1-sigma error)	float	1	molec/cm^2
H2O_VCD_PRIOR	smoothed a priori H_2O column density	float	1	molec/cm^2
H2O_AK	H_2O column averaging kernel	float	50	TBD
CH4_VCD	vertically integrated CH_4 column density	float	1	molec/cm^2
CH4_VCD_SIGMA	retrieval noise on CH4_VCD (1-sigma error)	float	1	molec/cm^2
CH4_VCD_PRIOR	smoothed a priori CH_4 column density	float	1	molec/cm^2
CH4_AK	CH_4 column averaging kernel	float	50	TBD
DELTA_D	vertically integrated δD	float	1	‰
DELTA_D_SIGMA	retrieval noise on DELTA_D (1-sigma error)	float	1	‰
AK_ALT	altitude levels of averaging kernels	float	50	meter
AK_PRESS	pressure levels of averaging kernels	float	50	hPa
CHI2	χ^2 of spectral fit residuals	float	1	-
NITER	number of iterations	integer	1	-
CONV	convergence flag (0-4)	integer	1	-
FLAG	quality flag	integer	1	-

Table 1: Overview of the HDO/H_2O data product.

Table 2: Microphysical properties of water and ice clouds: $n(r)$ represents the size distribution type, r_{eff} and v_{eff} are the effective radius and variance of the size distribution, $n = n_r - in_i$ is the refractive index.The ice cloud size distribution follows a power-law distribution as proposed by [RD1].

	water clouds	ice clouds
$n(r)$	gamma	$(r/r_1)^{-3.85}$
r_{eff} [μm]	20	-
v_{eff}	0.10	-
n_r	1.28	1.26
n_i	$4.7 \cdot 10^{-4}$	$2.87 \cdot 10^{-4}$

7 Performance and error analysis

To assess the performance of the retrieval algorithm, we generated simulated measurements for various test cases using the S-LINTRAN version 2.0 radiative transfer model [RD25]. S-Lintran is a scalar plane-parallel radiative transfer model based on the discrete ordinate method [RD26], which accounts for multiple elastic scattering by particles (clouds and cirrus) and molecules, along with the interaction with the reflecting Earth surface. This version of LINTRAN does not account for the polarisation properties of light, which are of minor relevance for this study. The optical properties of water clouds are calculated using Mie theory with microphysical cloud properties provided in Table 2, while for ice clouds, the ray tracing model of Hess et al. [RD27, RD28] is employed, assuming hexagonal, columnar ice crystals randomly oriented in space. Cirrus and water clouds are described by their cloud top and base heights and optical thickness. Cirrus clouds are assumed to fully cover the observed ground scene, while water clouds can exhibit partial coverage, simulated using the independent pixel approximation [RD29]. The radiance spectra are superimposed with measurement noise using the TROPOMI noise model of [RD30]. For this, we assume an observed ground scene of 7×7 km^2 and a telescope aperture of 6×10^{-6} m^2, with the optical transmittance adjusted to achieve a signal-to-noise ratio of 120 in the continuum for a dark reference scene under specific conditions. The instrument noise, estimated at 346 electrons per signal, includes contributions from thermal background, dark current, readout noise, and analog-to-digital converter noise.

Generic scenarios

For all generic cases, which we describe in more detail below, we assume the same atmospheric model. We assume the US standard atmosphere [RD31] for the profiles of dry air density, pressure, water and CO. The CH_4 profile is taken from the European background profile of the TM4 + CHIMERE chemical transport model [RD32], interpolated to the same pressure grid and converted from mixing ratios to densities using the air densities from the US standard atmosphere. We separated the water profile into individual profiles for the three isotopic components with absorption features in the TROPOMI SWIR range: $H_2{}^{16}O$, $H_2{}^{18}O$ and HDO.

In Fig. 3, we show the (normalised) atmospheric H_2O, CO, CH_4 and air density profiles, the temperature profile and the profiles for the HDO and $H_2{}^{18}O$ depletion. For all cases A-F, the measurement spectra are simulated for a water vapor column of $4.79 \cdot 10^{22}$ molec./cm^2, a CH_4 column of $3.68 \cdot 10^{19}$ molec./cm^2, and a CO column of $2.38 \cdot 10^{18}$ molec./cm^2. In addition to this model atmosphere, we defined different cases to test the influence of clouds, aerosol, surface reflection and solar geometry. All simulations assumed a fixed sensor viewing zenith angle of 40° and an azimuthal difference between the sensor and the sun of 60°. These test cases are described below and an overview of the test cases is given in Table 3.

Case A is defined to study the performance of the algorithm under clear sky conditions. The ensemble contains simulated measurements without clouds and aerosol, but with a variable surface albedo (in the SWIR range) between 0.03–0.6 (in 10 non-equidistant steps i.e. A_s = 0.03, 0.05, 0.075, 0.10, 0.15, 0.20, 0.25, 0.3, 0.4, 0.5, 0.6) and a variable SZA between 0.0–70.0° in 14 steps of 5°.

Case B is defined to study the effect of clouds on the retrieval algorithm. The ensemble contains simulated measurements for a fixed cloud optical thickness (τ_{cld}) of 5, 10, 30 or 50 and a variable cloud top height between 1–8 km (in 7 steps of 1 km) and a variable cloud fraction between 0.0–1.0 (in 10 steps of 0.1). The clouds have a constant geometrical thickness of 1 km. The surface albedo is fixed at 0.05 and the SZA is fixed at 50°.

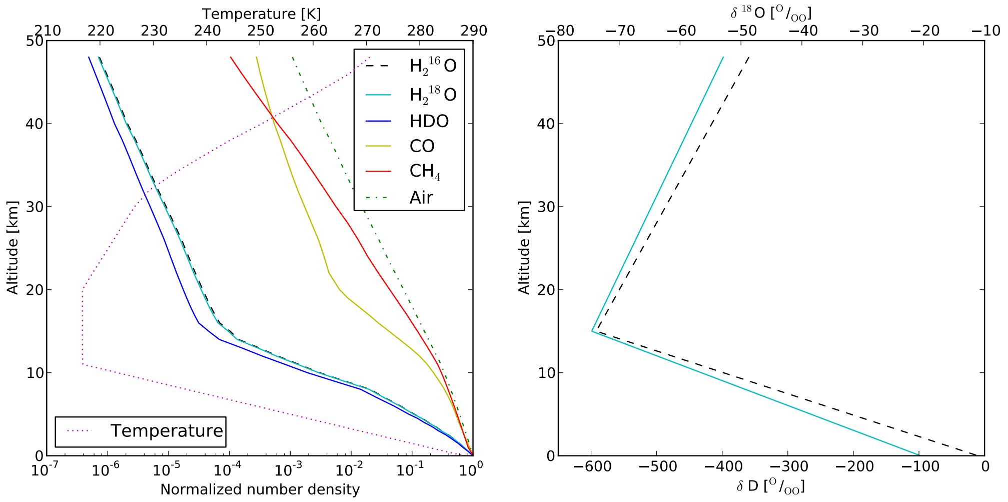

Figure 3: Left: atmospheric concentration profiles (bottom axis) and temperature profile (top axis) used as input for the model atmosphere. The concentrations are normalised to the concentration at ground level. Right: assumed profiles for the amount of HDO depletion (solid line, lower axis) and $H_2{}^{18}O$ depletion (dashed line, top axis).

Table 3: Summary of the different generic test cases A-F.

Case	Target of study	Variable X	Variable Y	Remarks
A	cloud free	surface albedo	SZA	
B	clouds	cloud top height	cloud fraction	for $\tau_{\mathrm{cld}} = 5, 10, 30, 50$
C	photon trapping	surface albedo	cloud fraction	for τ_{cld} = 2 & 5 and cloud top height = 2 & 5 km
D	aerosol	surface albedo	AOT	for a 0–2 km and 4–5 km aerosol layer
E	cirrus	surface albedo	cirrus optical depth COT	
F	multiple cloud layers	cloud fraction	COT	for A_S = 0.05, 0.1, 0.3, 0.6

Case C allows us to study the effect of "photon trapping". We have varied the surface albedo between 0.03–0.6 (in 10 steps) together with the cloud fraction (between 0.0–1.0 in 20 steps) for four different water clouds: a cloud optical thickness τ_{cld} of 2 and 5, combined with a cloud top height of 2 and 5 km. The geometrical cloud thickness was 1 km and the SZA was fixed at 50°.

Case D is to study the impact of different aerosol loads. We have varied the surface albedo between 0.03–0.6 (in 10 non equidistant steps, same as for Case A) together with the aerosol optical thickness (AOT) at 550 nm between 0.0–1.0 in 20 steps. We assumed two different scenarios: one scenario contained a sulfate-type aerosol in the boundary layer between 0–2 km, the other scenario contained an urban-type aerosol between 4–5 km. Both scenarios also contained a background aerosol layer with an AOT of 0.1 (at 2300 nm) at the surface, exponentially decreasing with altitude z with a factor of the form 0.9998^z with z in meters. The SZA was again fixed at 50°.

Case E is defined to study the impact of ice (i.e. cirrus) clouds. We have varied the surface albedo between 0.03–0.6 (in 10 steps) and the cirrus optical depth at 2300 nm between 0.0–1.0 in steps of 0.05. The cirrus cloud fraction was 100% for a layer between 9 and 10 km. The SZA was again fixed at 50°.

Case F is defined to study the effect of multiple layer clouds and the interaction with surface reflection. We have considered a cirrus cloud between 9 –10 km with optical depth between 0.0–1.0. and a water cloud between 2–3 km with a cloud fraction between 0.0–1.0 (in 10 steps of 0.1). The water cloud has an optical depth of 5. Moreover, no aerosols are present and we consider four different surface albedos of 0.05, 0.10, 0.30, and 0.60. The SZA is fixed at 50° and VZA is 40°.

In the input for these measurement simulations, the water profiles of $H_2{}^{16}O$, $H_2{}^{18}O$ and HDO were scaled with the isotopic abundance of the specific species from SMOW. Additionally, a realistic altitude-dependent depletion of HDO and $H_2{}^{18}O$ was assumed. For HDO we assumed a linear decrease from $\delta D = -100\,‰$ at the surface to $\delta D = -600\,‰$ at 15 km, followed by a linear increase to $\delta D = -400\,‰$ at the top of the atmosphere at an altitude of 48 km [RD33, RD9, RD34]. In this notation, δD denotes the abundance ratio between HDO and $H_2{}^{16}O$, relative to the SMOW abundance ratio:

$$\delta D = \left(\frac{[\mathrm{HDO}]/[\mathrm{H_2{}^{16}O}]}{R_s} - 1 \right) \cdot 1000\,‰. \tag{2}$$

An equivalent notation holds for $\delta^{18}O$. We further assumed that the concentration of $H_2{}^{18}O$ is related to the concentration of HDO according to the empirically determined ”global meteoric water line” [RD35]:

$$\delta D = 8 \cdot \delta^{18}O + 10\,‰. \tag{3}$$

Similar to the CO retrieval, we summarised the retrieval performance by considering separately the retrieval noise (or statistical error) $\sigma_{\delta D}$ and the bias $\Delta\delta D$. For the definition of the bias $\Delta\delta D$, we first determine the retrieved $\delta D_{\mathrm{retrieval}}$, for which we remove the noise on the total columns HDO (c_{HDO}) and $H_2{}^{16}O$ (c_{H2O}):

$$\delta D_{\mathrm{retrieval}} = \left(\frac{c_{\mathrm{HDO}} - \mathbf{G}_{\mathrm{HDO}}\mathbf{e}_y}{c_{\mathrm{H2O}} - \mathbf{G}_{\mathrm{H2O}}\mathbf{e}_y} \frac{1}{R_s} - 1 \right) \cdot 1000\,‰, \tag{4}$$

in which $\mathbf{G}_{\mathrm{HDO}}\mathbf{e}_y$ and $\mathbf{G}_{\mathrm{H2O}}\mathbf{e}_y$ are estimates of the noise on the total columns HDO and $H_2{}^{16}O$, using the relevant contributions of the gain matrix (see Sec. 5 in [RD24]). We need to compare $\delta D_{\mathrm{retrieval}}$ with $\delta D_{\mathrm{model}}$, where $\delta D_{\mathrm{model}}$ is δD of the “true” model atmosphere, smoothed with the total column averaging kernels of HDO ($\mathbf{A}_{\mathrm{HDO}}$) and $H_2{}^{16}O$ ($\mathbf{A}_{\mathrm{H2O}}$):

$$\delta D_{\mathrm{model}} = \left(\frac{\mathbf{A}_{\mathrm{HDO}}\rho_{\mathrm{HDO,true}}}{\mathbf{A}_{\mathrm{H2O}}\rho_{\mathrm{H2O,true}}} \frac{1}{R_s} - 1 \right) \cdot 1000\,‰. \tag{5}$$

Here, $\rho_{\mathrm{HDO,true}}$ and $\rho_{\mathrm{H2O,true}}$ represent the true atmospheric profiles of HDO and $H_2{}^{16}O$, respectively. Finally, the retrieval bias on δD is defined as:

$$\Delta\delta D = \delta D_{\mathrm{retrieval}} - \delta D_{\mathrm{model}}. \tag{6}$$

In the following figures, we also consider the bias in the retrieved H_2O and HDO columns, which give insight in the cause for the bias in δD. These biases are defined in the same way as for CO (Eq. 71 in [RD24]), namely by first removing the retrieval noise and then taking the difference with the model atmosphere that has been smoothed with the total column averaging kernel. Additionally, we show the relative difference between the retrieved methane columns from the two-band methane cloud filter. These values are passed through from the SWIR pre-processor and will be used as a cloud/aerosol pre-filter.

The retrieval noise (statistical error) estimates for the total columns H_2O, HDO and CH_4 are calculated in the retrieval algorithm in the same way as for the total column CO (Eq. 55 in [RD24]). The retrieval noise estimate for δD is found by applying the error propagation equation to Eq. (2), where we have replaced the abundances [HDO] and [H_2O] by the total columns C_{HDO} and C_{H2O}, respectively:

$$\sigma_{\delta D} = \frac{1000}{R_s} \sqrt{\left(\frac{\sigma_{\mathrm{HDO}}}{c_{\mathrm{H2O}}} \right)^2 + \left(\frac{\sigma_{\mathrm{H2O}} c_{\mathrm{HDO}}}{c_{\mathrm{H2O}}^2} \right)^2}. \tag{7}$$

Here we assume that the total column retrievals of HDO and H_2O are uncorrelated, which is a reasonable assumption knowing that the absorption features of HDO and H_2O are adequately separated (see Fig. 1).

7.1 Case A: cloud free conditions

First we consider the HDO/H_2O retrieval performance under cloud free conditions as a function of surface albedo and SZA. Figure 4 shows that the retrieval performs very well for the majority of the scenes, with $\Delta\delta$D less than 0.8 ‰. Only for the low surface albedos (0.03–0.05) the δD bias increases to a few per mil, due to a slightly more negative bias in H_2O compared to HDO.

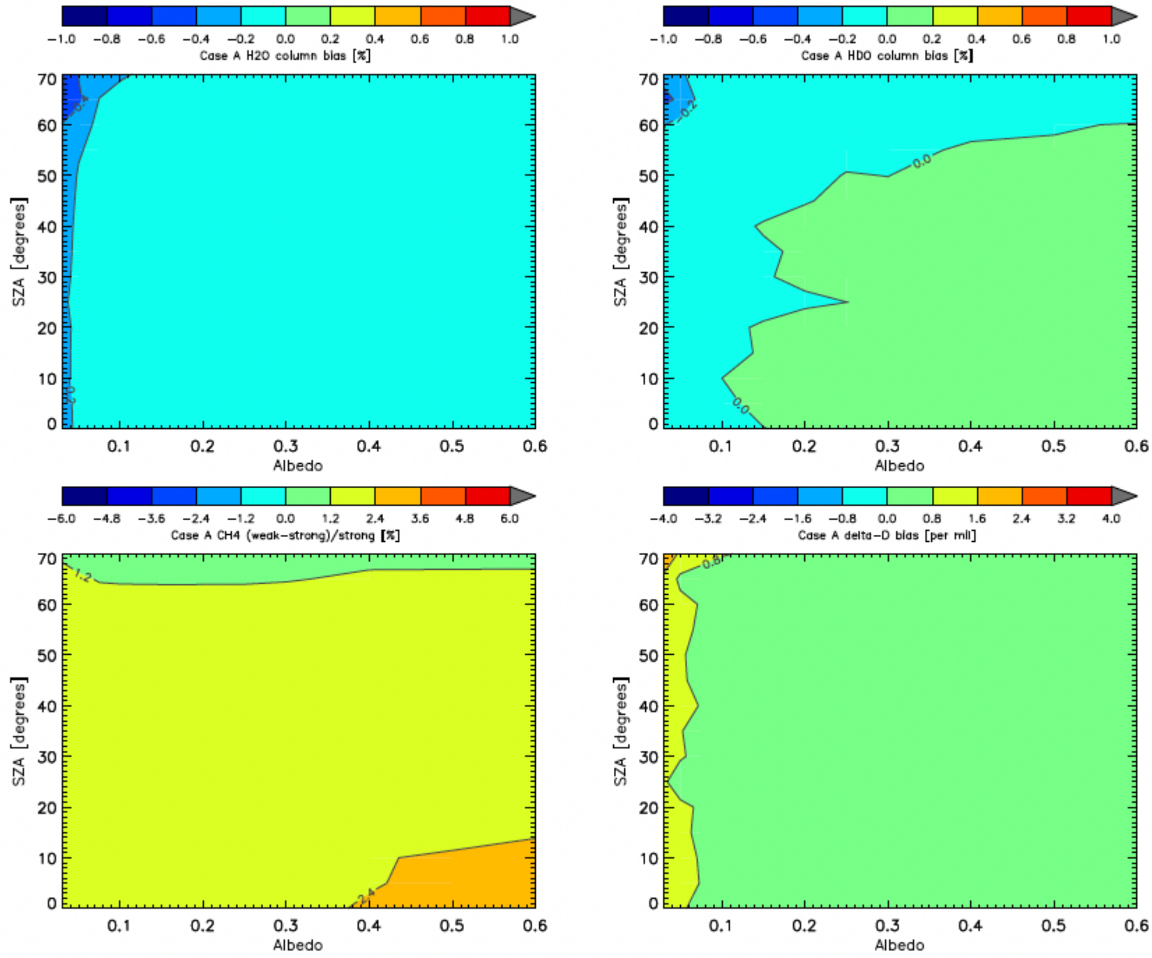

Figure 4: Cloud free retrieval bias (Case A) for the total columns of H_2O (top left), HDO (top right) and δD (bottom right). The bottom left panel shows the relative difference in total CH_4 column between the weak and strong bands from the pre-processor two-band CH_4 cloud filter.

The corresponding statistical error estimates are shown in Fig. 5. We find $\sigma_{\mathrm{H_2O}}$ and $\sigma_{\mathrm{CH_4}}$ values reaching maxima of 0.6–1.6% for the lowest surface albedos, depending on SZA. The estimated errors in the HDO total column, σ_{HDO}, are larger due to the weaker absorption features, resulting in absolute values for $\sigma_{\delta\mathrm{D}}$ of the order of 15–25‰ for the lowest surface albedos. For high surface albedo regions (such as deserts with an albedo of ~ 0.3 in the SWIR) typical values for $\sigma_{\delta\mathrm{D}}$ are 2–4‰.

The column averaging kernels for Case A are shown in Fig. 6. Due to the differences in line strengths between H_2O and HDO, their averaging kernels are significantly different at higher altitudes. The HDO averaging kernels are close to unity due to the weak HDO absorption features. The strong absorption features of H_2O, combined with pressure broadening near the surface, make the H_2O averaging kernels strongly decrease with altitude. However, since the scale height of water vapour is 1–2 km under most conditions, most of the water vapour resides near the surface, where the HDO and H_2O averaging kernels are similar. Higher atmospheric layers, where the sensitivities to HDO and H_2O are different, will contribute very little to the total columns. Nevertheless, the column averaging kernels are provided with every individual sounding and should

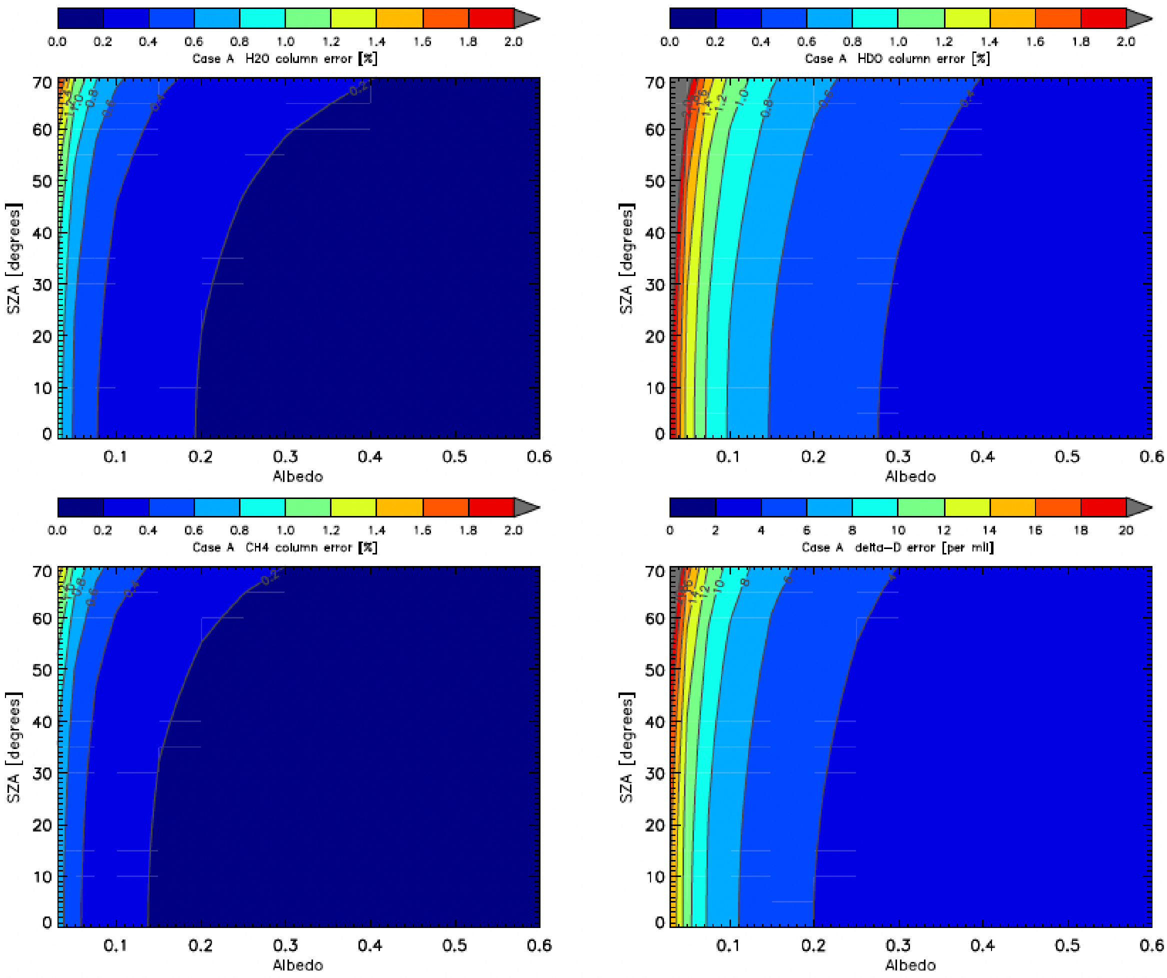

Figure 5: Statistical error estimates for Case A. Top left: H_2O column ($\sigma_{\mathrm{H_2O}}$). Top right: HDO column (σ_{HDO}). Bottom left: CH_4 column ($\sigma_{\mathrm{CH_4}}$). Bottom right: δD ($\sigma_{\delta\mathrm{D}}$).

be taken into account in every model-data intercomparison as we have described above.

7.2 Case B: clouds

In Fig. 7 we show the retrieval performance for different cloudy scenarios, assuming a cloud optical thickness of $\tau_{\mathrm{cld}} = 5$. For higher optical thicknesses ($\tau_{\mathrm{cld}} = 10$, 30 and 50), the figures look similar, but with slightly increased biases. The figure shows that with increasing cloud top height and cloud fraction we find increasingly negative biases in total column H_2O, HDO and CH_4 due to the clouds shielding parts of the lower atmosphere and thereby shortening the optical path-length. Since the atmospheric profile (Fig. 3) and column averaging kernel of H_2O (Fig. 6) are steeper than those of CH_4, the bias for H_2O is larger than for CH_4 for the same cloud conditions. Although the atmospheric profile of HDO is slightly steeper than that of H_2O (see Fig. 3) the bias in total column HDO is smaller than that of total column H_2O for the same cloud conditions, which can be explained by the steeper averaging kernel of H_2O. Since we take the ratio HDO/H_2O, we find a positive bias in δD, strongly increasing with cloud top height and cloud fraction (bottom right panel in Fig. 7).

In principle, applying the averaging kernels to the true or model atmospheric profiles should not lead to an bias. This could be seen in Fig. 4 for case A, where both the averaging kernels (through the forward model in the retrieval algorithm) and the simulated atmospheric profile assumed cloud free conditions. For case B, however, the averaging kernels are incorrect, since our forward model used for the HDO/H_2O retrieval does not take into account clouds. This can be seen by comparing Fig. 8, which shows typical H_2O and HDO column

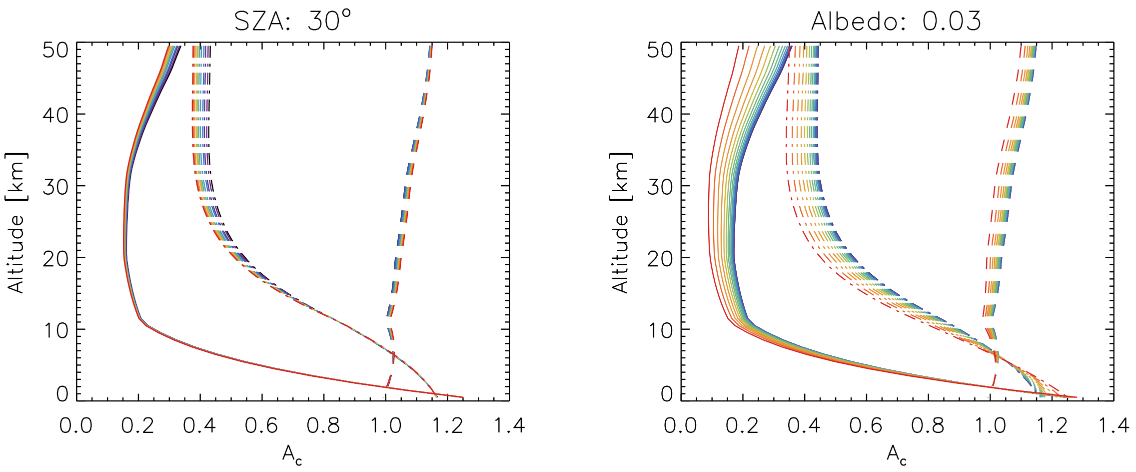

Figure 6: Total column averaging kernels for H_2O (solid lines), HDO (dashed lines) and CH_4 (dash-dotted lines). Left: for a fixed SZA of $30°$. The colours indicate different surface albedos from 0.03 (blue) to 0.6 (red). Right: for fixed surface albedo of 0.03. The colours indicate different SZAs from $0°$ (blue) to $70°$ (red).

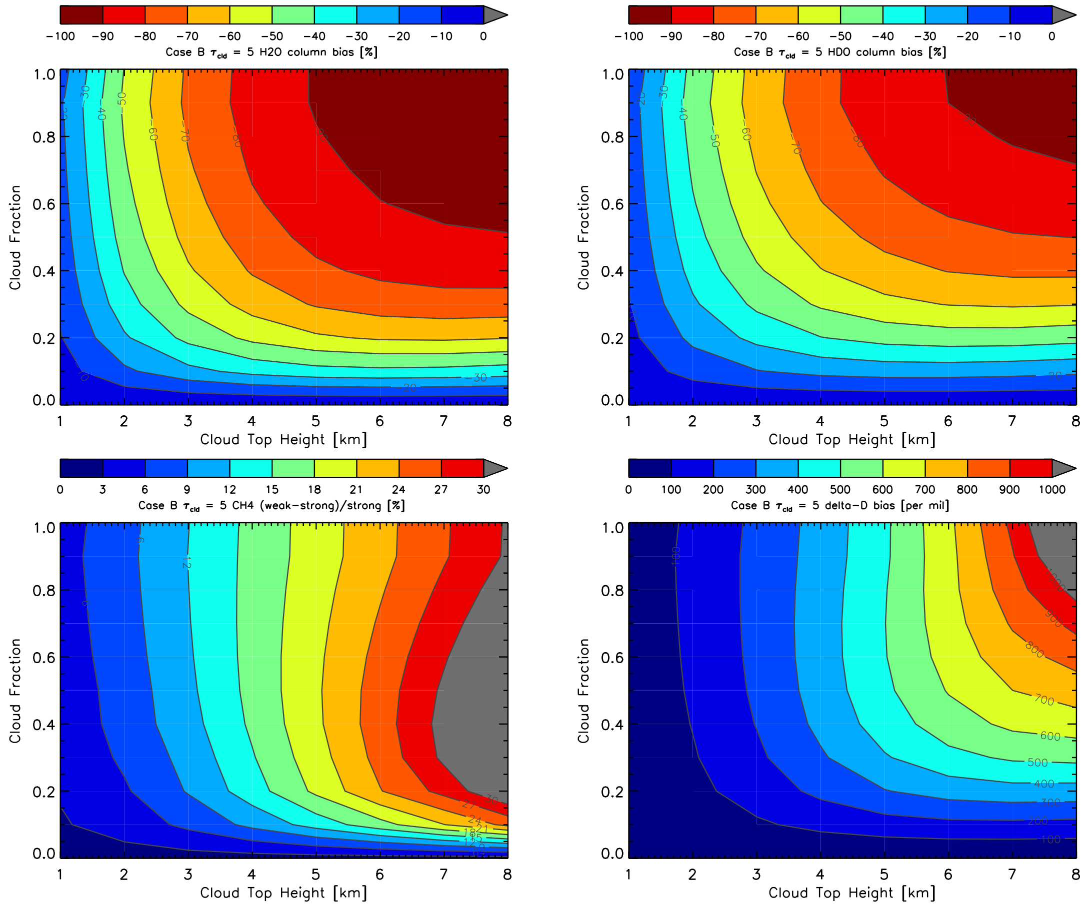

Figure 7: Retrieval bias as a function of cloud top height and cloud fraction (Case B, assuming $\tau_{\rm cld} = 5$) for the total columns of H_2O (top left), HDO (top right) and δD (bottom right). The bottom left panel shows the relative difference in total CH_4 column between the weak and strong bands from the pre-processor two-band CH_4 cloud filter.

averaging kernels for the lowest 12 km of an atmosphere including a $\tau_{\text{cld}} = 50$ cloud at an altitude of 5 km, to Fig. 6 in [RD24], which shows the CO column averaging kernels for a similar atmosphere for a retrieval that takes clouds into account. Since our forward model does not take clouds into account, Fig. 8 shows *no* strong drop in sensitivity around an altitude of 5 km. By applying the averaging kernels of Fig. 8 to the model atmosphere, in order to calculate the retrieval bias for cloudy conditions, we incorrectly assume too much sensitivity to the layers below the cloud, which contain most of the water vapour, resulting in a negative bias which will be strongest for H_2O.

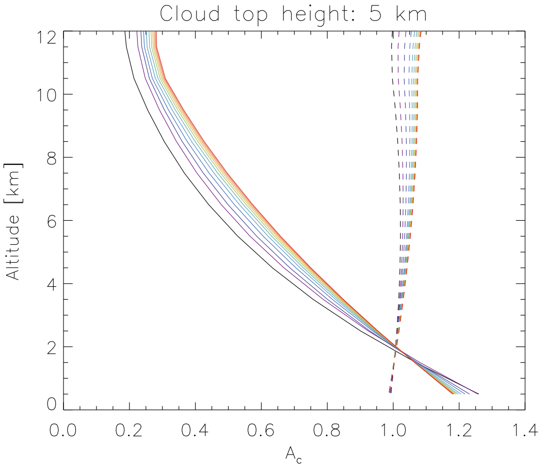

Figure 8: Total column averaging kernels for H_2O (solid lines) and HDO (dashed lines) for a simulated atmosphere that contained a 1 km thick, $\tau_{\text{cld}} = 50$, $z_{\text{cld}} = 5$ km cloud layer. The colours indicate different cloud fractions from 0.0 (blue) to 1.0 (red).

Since we do not account for clouds in the retrieval, and they lead to large retrieval biases, we need to filter the retrieval results for cloudy conditions. Figure 7 shows that we can use the two-band CH_4 cloud filter from the SWIR pre-processor for this. We have tested that for a relative difference in total CH_4 column between the weak and strong bands of $< 6\%$, we limit the retrieval bias in δD to a maximum of ~ 50–70 ‰ for all simulated measurements, while simultaneously reducing the computational effort of the retrieval (since cloud contaminated scenes will not be processed, see Sect. 9). Figure 9 (bottom-left panel) shows that applying such a filter allows only for ground scenes which are covered by less than 10–20% of higher-level clouds (cloud top heights of 1 km or more) or ground scenes which are covered by an arbitrary fraction of low-level clouds (cloud top heights up to 1 km). In the remainder of this appendix, we assume that cloud filtering is performed using the two-band CH_4 filter from the pre-processor with a threshold of 6%.

Finally, in Fig. 10 we show the statistical error estimates for case B, assuming $\tau_{\text{cld}} = 50$. The errors are largest for scenes covered with high-level clouds, as such clouds cover most of the absorption features caused by the atmospheric layers underneath. The errors are lowest for large fractions of low-level clouds, as the high albedos of such clouds enhance the reflectance signal, while there are still enough absorbers present above the clouds to give absorption features. Typical errors after applying the cloud filter are $\sigma_{\mathrm{H_2O}} \lesssim 0.8\%$ and $\sigma_{\text{HDO}} \lesssim 2\%$, resulting in single measurement errors of 10–20‰ for $\sigma_{\delta\text{D}}$ (for a surface albedo of 0.05 and SZA=50°).

7.3 Case C: photon trapping

Case C is designed to study the impact of “photon trapping”. When an optically thin cloud layer is considered above a surface with a high albedo, there is a chance that a photon from the surface scatters back and forth between the cloud and the surface (becomes trapped), before reaching the instrument, causing a path length increase. Other photons from the top of the atmosphere will directly be reflected off the cloud into the instrument causing a path length decrease. Figure 11 shows the induced retrieval biases by neglecting these effects in the retrieval, for the combination $\tau_{\text{cld}} = 2$ and $z_{\text{cld}} = 5$ km.

Figure 11 shows that the direct reflection off the clouds is most important for low albedo surfaces, which have the highest contrast between the bright clouds and the dark surface (visible as negative total column

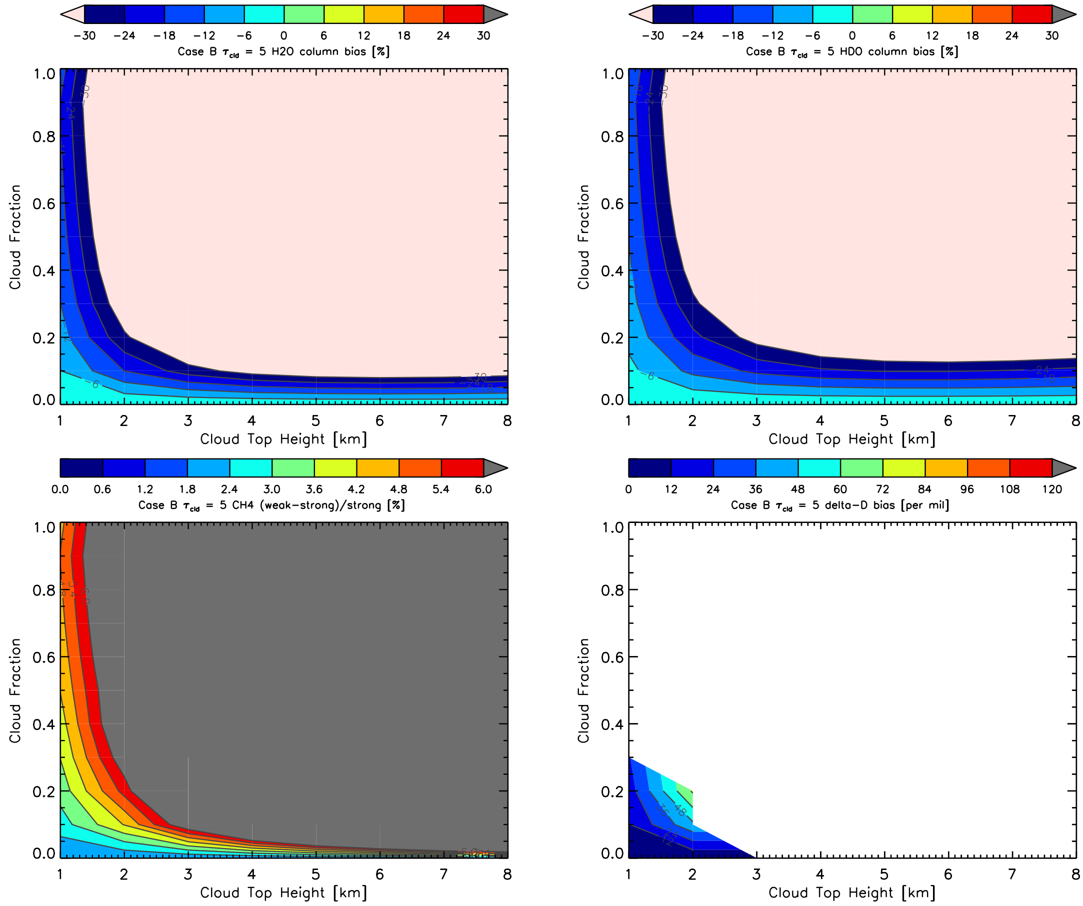

Figure 9: Same as Figure 7, but now with the δD bias filtered against cloud contaminated data using a two-band methane cloud filter threshold of 6%. Note that this filter allows for all cloud fractions at a cloud top height of 1 km, and all cloud-free scenes (the coloured contours in the bottom-left panel), even though these scenes are not visible in the bottom-right panel.

biases for H_2O and HDO and due to path length shortening). This effect becomes less for higher surface albedos, as the contribution from trapped photons and photons which reflected directly of the surface increases.

The more negative retrieval bias for H_2O compared to HDO results again in a positive retrieval bias for δD. A situation with thin, low clouds ($\tau_{\mathrm{cld}} = 2$, $z_{\mathrm{cld}} = 2$ km, not shown) is most challenging for the retrieval algorithm, as this leads to very efficient photon trapping, while the cloud filter is not very effective against low clouds. However, the retrieval still performs reasonably well ($\Delta\delta\mathrm{D} \lesssim 70\ ‰$) after the cloud filter has removed the retrievals for the largest cloud fractions and lowest albedos.

For $\tau_{\mathrm{cld}} = 2$ and $z_{\mathrm{cld}} = 5$ km (Fig. 11), more retrievals are removed by the cloud filter and $\Delta\delta$D stays below $36\ ‰$. The cases for $\tau_{\mathrm{cld}} = 5$ also perform very well (not shown), because the cloud filter was tuned to this optical thickness. Values for $\Delta\delta$D stay below $\sim 30\ ‰$ and the lowest biases are found for the lowest cloud fractions and the highest albedos.

The statistical error estimates for case C show a strong correlation with surface albedo (larger errors for lower albedos, similar to case A) and a weaker correlation with cloud fraction (larger errors for larger cloud fractions). After cloud filtering all values for $\sigma_{\delta\mathrm{D}}$ are in the range 2.2–19‰.

7.4 Case D: aerosol

With Case D we study the retrieval performance under two different types of aerosol loads. Figure 12 shows the simulated retrieval biases for a sulfate-type aerosol in the boundary layer between 0–2 km, as a function of aerosol optical thickness (AOT) and surface albedo. Light scattering in the boundary layer aerosol generally leads to an increase in the optical path length (positive biases for H_2O and HDO). For very low surface albedos,

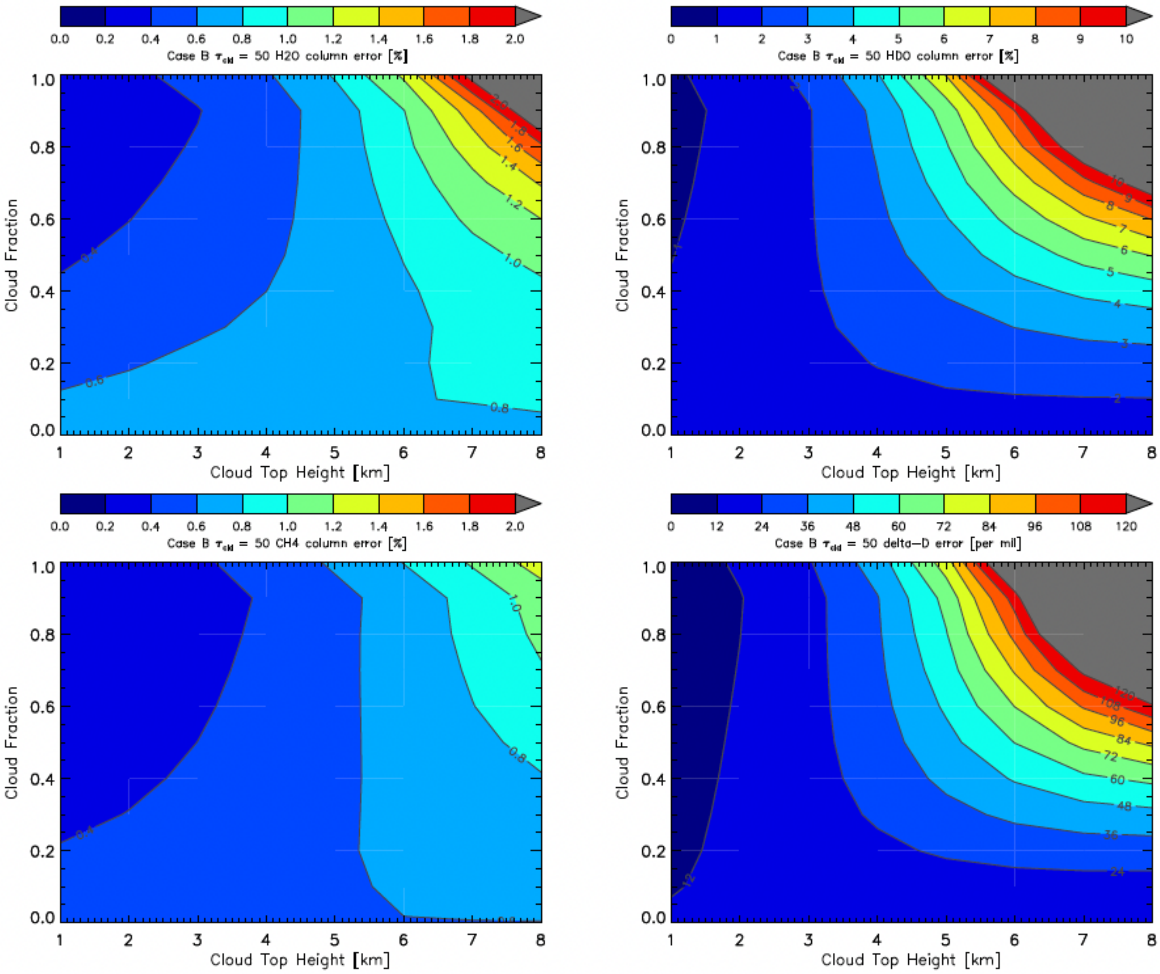

Figure 10: Statistical error estimates for Case B, assuming $\tau_{\mathrm{cld}} = 50$. Top left: H_2O column ($\sigma_{\mathrm{H_2O}}$). Top right: HDO column (σ_{HDO}). Bottom left: CH_4 column ($\sigma_{\mathrm{CH_4}}$). Bottom right: δD ($\sigma_{\delta\mathrm{D}}$).

however, direct reflection off the aerosol layer leads to a path length shortening. For low surface albedos we see that the bias for H_2O increases faster than for HDO, possibly due to the higher sensitivity of the retrieval to H_2O near the surface where also most of the water vapour resides. The resulting retrieval bias in δD is therefore highest for low surface albedos and high aerosol optical thicknesses. The two-band CH_4 cloud filter, however, is very effective at removing the highest δD biases and the remaining scenes all have biases of $\Delta\delta\mathrm{D} < 36‰$.

For the case of an urban-type aerosol layer between 4–5 km (Fig. 13) we see that the bias patterns of the boundary layer aerosol (Fig. 12) are somewhat enhanced. The two-band CH_4 cloud is also more strict: more of the higher albedo scenes are filtered out, restricting the bias in HDO of the remaining scenes to $\Delta\delta\mathrm{D} < 36‰$.

We find that the statistical error estimates do not strongly depend on AOT. Similar to case A, they primarily vary with surface albedo and reach peak values of $\sigma_{\delta\mathrm{D}} = 20‰$ for the lowest albedo.

7.5 Case E: ice clouds

With case E we study the retrieval performance under the presence of an optically thin cirrus layer (with a cloud fraction of 100%, $z_{\mathrm{cld}} = 10$ km and varying cirrus optical thickness). The retrieval biases are show in Fig. 14. For low albedos and high cirrus optical thicknesses we find negative biases for H_2O and HDO due to the direct reflection of light on the cirrus layer. Similar to the results from case C, the biases become positive for higher surface albedos due to the effect of photon trapping. The positive biases for HDO are higher than for H_2O, leading to large retrieval biases in δD for cirrus optical thicknesses (at 2300 nm) above 0.3. The two-band CH_4

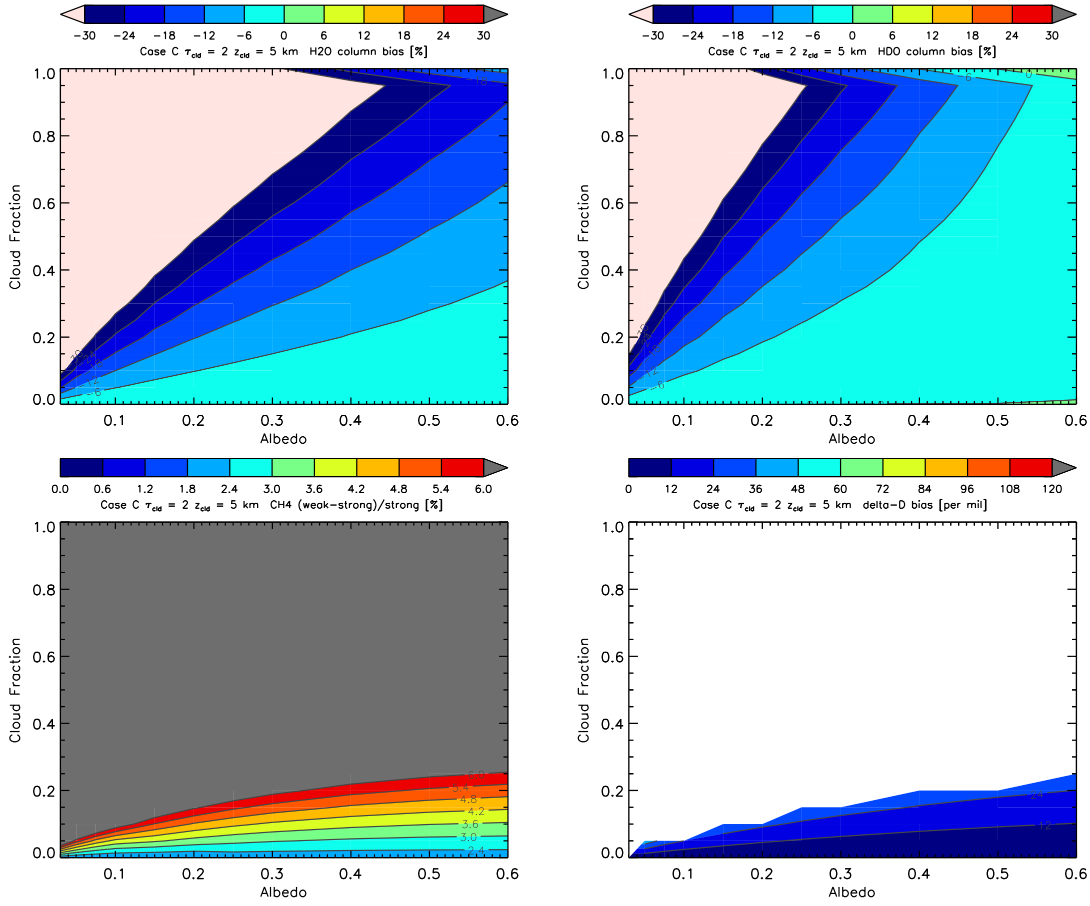

Figure 11: Retrieval bias as a function of surface albedo and cloud fraction (Case C, assuming $\tau_{\mathrm{cld}} = 2$ and $z_{\mathrm{cld}} = 5$ km) for the total columns of H_2O (top left), HDO (top right) and δD (bottom right). The bottom left panel shows the threshold values from the two-band CH_4 cloud filter. The δD bias has been masked against a threshold of 6%.

cloud filter, however, is very effective at removing cirrus-contaminated scenes: all scenes with a cirrus optical thickness above 0.1 are removed by the threshold of 6%. Only scenes with a relatively high albedo and a cirrus optical thickness below 0.1 are passed through, for which the retrieval bias in δD is restricted to $\Delta\delta\mathrm{D} < 12‰$.

The statistical error estimates (not shown) behave similarly to case A and D: they primarily vary with surface albedo and reach peak values of $\sigma_{\delta\mathrm{D}} = 20‰$ for the lowest albedo (after cloud filtering).

7.6 Conclusions

Having studied the performance of the HDO/H_2O retrieval for the generic test cases of Sec. 7, we can conclude that the retrieval performs well under cloud free conditions. The retrieval bias in δD ($\Delta\delta$D) will be less than 3‰, even for the lowest albedos, and the statistical errors ($\sigma_{\delta\mathrm{D}}$) vary from 15–25‰ for the lowest albedos to 2–4‰ for high albedos.

Under cloudy conditions, or conditions with aerosol, the retrieval performs less well and we generally find a positive bias in δD. To restrict this bias, we need strict filtering against clouds and aerosol using the two-band CH_4 cloud filter from the pre-processor, which additionally leads to a great reduction in the computational effort (see Sect. 9). By restricting the retrieval to cases with a two-band CH_4 threshold of less than 6%, we restrict the bias in δD to $\Delta\delta$D <50–70‰ for all simulated measurements. Averaging multiple single measurements over time and space will further reduce the statistical error, and will improve the accuracy to better than the maximum ~ 50–70 ‰, bringing the measurements within the requirements to study typical temporal and spatial gradients (which are of the order of 50–100‰).

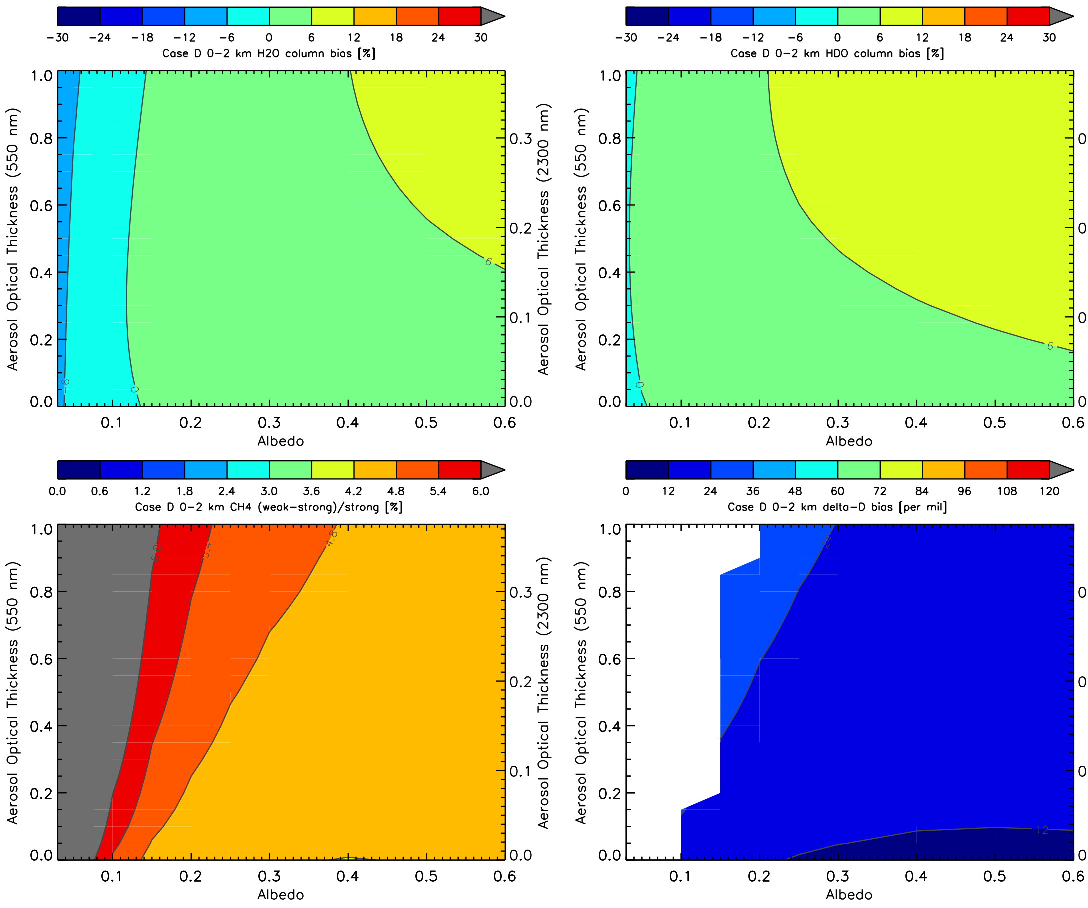

Figure 12: Retrieval bias as a function of surface albedo and aerosol optical thickness (Case D, assuming a sulfate-type aerosol layer in the boundary layer between 0–2 km) for the total columns of H_2O (top left), HDO (top right) and δD (bottom right). The bottom left panel shows the threshold values from the two-band CH_4 cloud filter. The δD bias has been masked against a threshold of 6%.

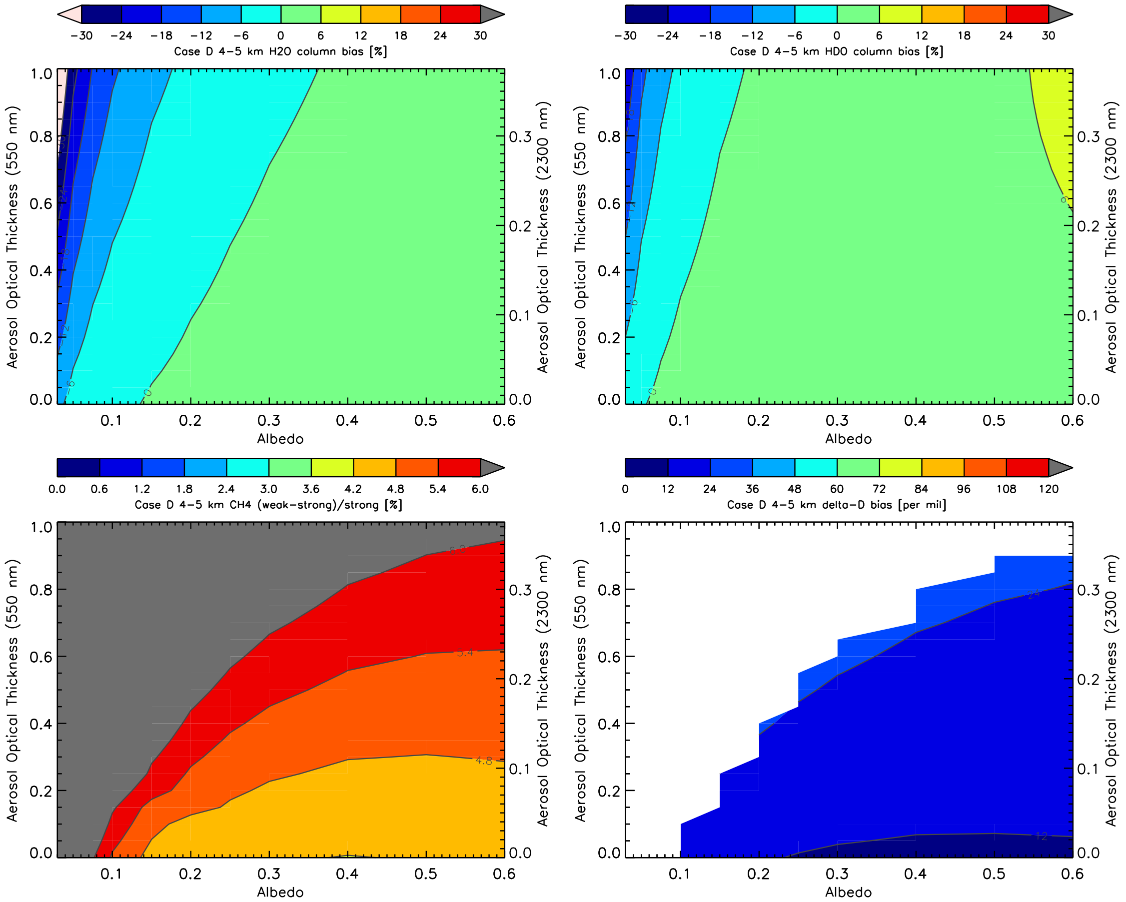

Figure 13: Same as Figure 12, but now for an urban-type aerosol layer between 4–5 km.

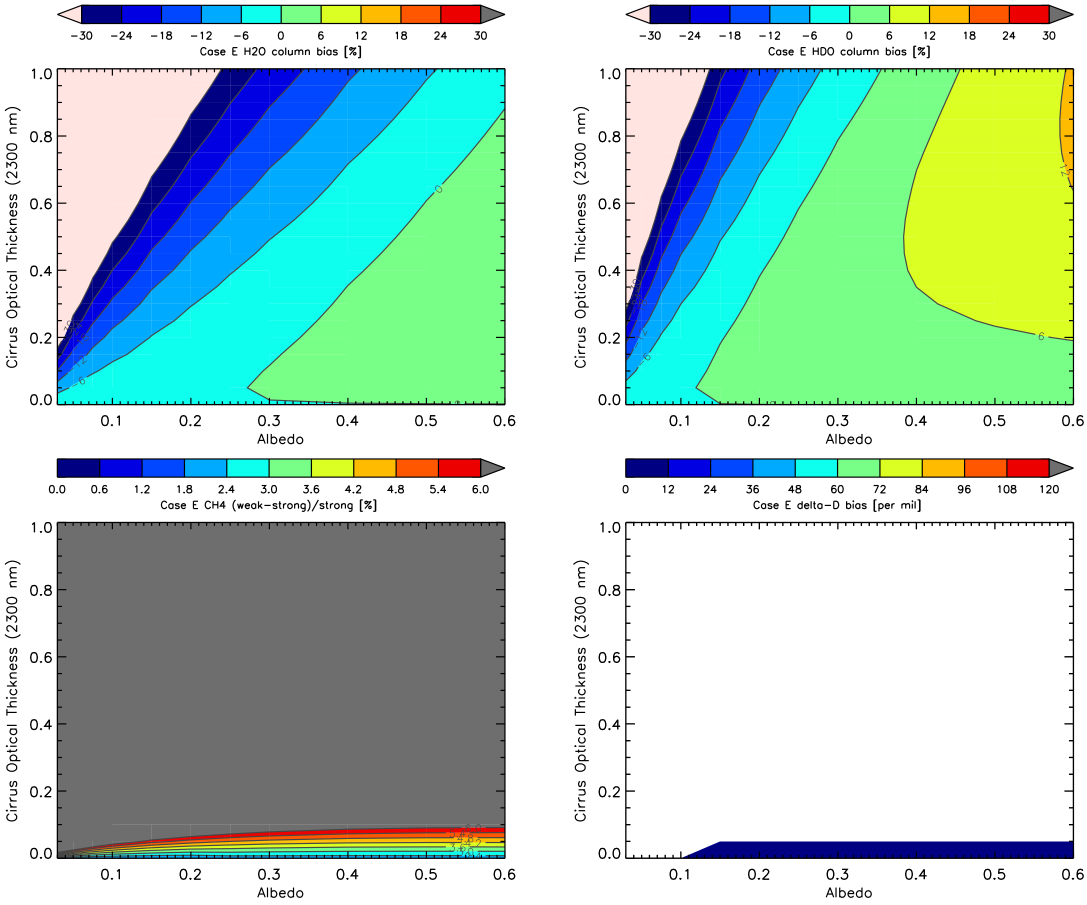

Figure 14: Retrieval bias as a function of surface albedo and cirrus optical thickness (Case E) for the total columns of H_2O (top left), HDO (top right) and δD (bottom right). The bottom left panel shows the threshold values from the two-band CH_4 cloud filter. The δD bias has been masked against a threshold of 6%.

Table 4: Summary of the sensitivity study. Mean systematic errors and standard deviations are derived for the surface albedos 0.03, 0.05 and 0.075 (45 scenes for case A and 4 scenes for case C after filtering).

A priori parameter	Systematic Error on δD [‰]	
	Case A	Case C
δD SMOW	0.74 ± 1.1	0.88 ± 1.2
δD -200‰	-0.29 ± 0.50	-0.38 ± 0.21
δD -400‰	-3.1 ± 3.1	-1.8 ± 1.7
δD -600‰	-7.3 ± 8.0	-4.7 ± 2.8
δD -700‰	-10 ± 12	-15 ± 20
H_2O profile $I_{\mathrm{H2O}} = 1$	6.8 ± 1.4	6.1 ± 1.7
H_2O profile $I_{\mathrm{H2O}} = 10$	-4.8 ± 1.9	-4.1 ± 2.4
CH_4 profile $I_{\mathrm{CH4}} = 1$	-1.3 ± 0.99	-1.5 ± 0.11
CH_4 profile $I_{\mathrm{CH4}} = 10$	6.2 ± 1.1	6.1 ± 0.16
Temp -0.5 K	-6.9 ± 0.74	-7.1 ± 0.14
Temp $+0.5$ K	7.0 ± 0.11	7.1 ± 0.16
Temp -1 K	-14 ± 0.74	-14 ± 0.27
Temp $+1$ K	14 ± 0.17	14 ± 0.27
Pressure $\times 0.99$	-4.5 ± 0.74	-4.8 ± 0.15
Pressure $\times 1.01$	4.5 ± 0.54	4.7 ± 0.17
Rad. offset +0.1%	-0.054 ± 0.11	-0.052 ± 0.035
Rad. offset +0.5%	-0.20 ± 0.16	-0.30 ± 0.25
ISRF FWHM -1%	0.36 ± 0.85	0.11 ± 0.29
ISRF FWHM $+1\%$	-0.33 ± 0.84	-0.13 ± 0.32

8 Sensitivity analysis

We performed various studies to test the sensitivity of the HDO/H_2O retrieval to the a priori assumptions and uncertainties in the instrument calibration. The default retrieval settings were the same as described in Sec. 5. Below we describe the impact of certain perturbations with respect to this default retrieval, for Case A (cloud free, varying surface albedo and SZA) and Case C (varying cloud fraction and surface albedo for a $\tau_{\mathrm{cld}} = 2$ cloud layer at $z_{\mathrm{cld}} = 5$ km). The case C retrievals were first filtered for clouds. The impact is expressed as a systematic error and standard deviation, where we define the systematic error as the mean difference in retrieved δD between the perturbed and the default retrieval for the lowest surface albedos only (0.03, 0.05 and 0.075). The results are summarised in Table 4.

8.1 A priori HDO profile

So far, we assumed the same atmospheric profile for the amount of HDO depletion (expressed as δD), for both the measurement simulations and the HDO/H_2O retrieval. Although the profile for the measurement simulations is based on a realistic atmosphere (see Sec. 7), the exact profile for every measurement will not be known. The H_2O input profile for the operational algorithm will come from ECMWF and will also be used for HDO without applying any additional depletion w.r.t. Standard Mean Ocean Water (SMOW). Therefore, we also performed retrievals were the δD profile was held constant with altitude at the SMOW value (i.e. $\delta\mathrm{D} = 0$‰ or no depletion w.r.t. Standard Mean Ocean Water) and -200, -400, -600 and -700‰, respectively. We find that this induces errors in the retrieved δD for low albedo scenes from as low as -0.29 ± 0.50‰ (for Case A, assuming a constant δD profile of -200‰) to as large as -15 ± 20‰ (for Case C, assuming a constant δD profile of -700‰, see Table 4). These errors are caused by errors in the retrieved HDO total column. Most realistic variations, however, are expected to be on the order of 0–200‰ relative to the standard δD profile, thus leading to errors well below 10‰.

8.2 A priori H_2O profile

We varied the shape of the a priori H_2O profile to estimate the induced errors related to uncertainties in the ECMWF H_2O fields. We used the same ensemble of perturbed profiles as were used for the CO sensitivity analysis, namely the relative profiles depicted in the left-hand panel of Fig. 22 in [RD24]. The perturbed profiles that underestimate the water abundance in the lower atmosphere, and overestimate it in the higher atmosphere, relative to the reference profile, induce positive errors on the retrieved HDO/H_2O ratio of up to $6.8 \pm 1.4‰$ (for the relative profile with index $I_{\mathrm{H2O}} = 1$). The perturbed profiles with the reversed shape (overestimation of the abundance in the lower atmosphere and underestimation in the higher atmosphere) lead also to reversed errors (up to -4.8 ± 1.9 for $I_{\mathrm{H2O}} = 10$). We find that the uncertainty of the ECMWF H_2O profile affects both the retrieved H_2O and HDO column densities, but in opposite directions, enhancing the impact on the ratio HDO/H_2O.

8.3 A priori CH_4 profile

We also varied the shape of the a priori assumption for the CH_4 profile, by performing retrievals with the perturbed profiles depicted in the right-hand panel of Fig. 22 in [RD24]. Perturbed CH_4 profiles that underestimate the CH_4 abundance in the lower atmosphere induce negative errors on δD (e.g. $-1.5 \pm 0.11‰$ for $I_{\mathrm{CH4}} = 1$, case C), and vice versa ($6.2 \pm 1.1‰$ for $I_{\mathrm{CH4}} = 10$, case A). In contrast to variations of the H_2O profile, we find that the uncertainty of the CH_4 profile affects the H_2O and HDO column densities in the same direction (although with stronger impact on HDO), partially mitigating the induced errors on the ratio HDO/H_2O.

8.4 Temperature profile

To test the impact of uncertainties in the a priori temperature profile from ECMWF, we have varied the temperature profile by ± 0.5 K and ± 1 K. This has a strong effect on the retrieved total column H_2O, while the total columns of HDO and CH_4 are not sensitive to temperature variations. We find that a perturbation of $+1$ K (-1 K) leads to a decrease (increase) in the retrieved H_2O column of 1.8%, inducing a systematic error in the retrieved HDO/H_2O ratio of $+14‰$ ($-14‰$). This error is constant for all surface albedos and SZAs, and scales linearly with the size of the temperature perturbation.

8.5 Pressure profile

The HDO/H_2O retrieval algorithm relies on an a priori surface pressure estimate from ECMWF, from which the atmospheric pressure profile is derived. To test the impact of inaccuracies in the ECMWF surface pressure, we have applied perturbations to this surface pressure of $\pm 1\%$. We find that these perturbations lead to systematic errors of about 0.5% in retrieved H_2O and 0.13% in HDO (but with reversed sign), together inducing errors of about 4.5‰ in the retrieved HDO/H_2O ratio.

8.6 Radiometric offset

Differences in the radiometric offset between irradiance spectra measured directly from the Sun and radiance spectra from the Earth's surface, could lead to offsets and spectral features in the reflectance spectra. The instrument requirement for the radiometric offset is 0.1% of the continuum level [AD1]. To test the impact of such a radiometric offset on the HDO/H_2O retrieval, we have applied an offset to the simulated Earth radiance spectra, before the radiance was converted to a reflectance (by division with the sun spectrum) and fed into the inversion scheme. We have used a radiometric offset of 0.1% and 0.5% of the maximum value in the retrieval window. We find that the retrieval of HDO/H_2O is not sensitive to such an offset (errors in δD less than 0.5‰). This can be explained by the fact that we fit for an offset in the measured reflectance spectra. Switching off this fit parameter induces errors of 0.4–1.0% in the retrieved H_2O and CH_4 columns (for an offset of 0.1%), which leads to a systematic error of $1.9 \pm 0.93‰$ in δD. Although an offset in radiance does not translate into the same offset in reflectance, fitting for an offset in reflectance seems sufficient to account for inaccuracies in the radiometric calibration of the instrument.

8.7 Slit function

We have also tested the impact of an inaccurate slit function (or Instrument Spectral Response Function, ISRF). For the default retrieval and measurement simulations we have assumed a Gaussian slit function with a

Table 5: Sensitivity to a change of water line parameters from HITRAN2008 to the parameters from Scheepmaker et al. (2013).

Isotope	Systematic Error			
	H_2O	HDO	CH_4	δD
$H_2{}^{16}O$	$8.0 \pm 0.55\%$	$-4.6 \pm 0.50\%$	$1.3 \pm 0.35\%$	-97 ± 1.5‰
HDO	$-0.069 \pm 0.036\%$	$-4.2 \pm 0.043\%$	$-0.040 \pm 0.037\%$	-34 ± 0.29‰
$H_2{}^{18}O$	$0.011 \pm 0.017\%$	$-0.11 \pm 0.0090\%$	$-0.029 \pm 0.017\%$	-0.97 ± 0.21‰
All	$7.9 \pm 0.53\%$	$-8.7 \pm 0.50\%$	$-1.3 \pm 0.37\%$	-128 ± 1.5‰

FWHM of 0.25 nm. We have tested the impact of perturbing this FWHM by $\pm 1\%$ and find that the induced systematic errors are strongly dependent on surface albedo and SZA. The largest errors are found for high albedos and low SZAs, and reach ± 3‰. The mean systematic errors for the lowest albedos are 0.36 ± 0.85‰. The sensitivity to the slit function is relatively large for the HDO/H_2O ratio, since it induces errors on both the retrieved HDO and H_2O column, but with opposite signs (which means they do not cancel in the HDO/H_2O ratio).

8.8 Use of different H_2O spectroscopy

Recent studies have shown that spectroscopic uncertainties of water can have a large impact on total column retrievals of CO [RD36], CH_4 [RD37, RD38], H_2O [RD39] and the HDO/H_2O ratio [RD40]. As a test of the possible impact that such uncertainties in the water line parameters can have on our retrieval of the HDO/H_2O ratio, we have repeated the retrieval of the simulated measurements of Case A, while replacing the line parameters of the water isotopes from the HITRAN2008 [RD41] values by the ones from [RD40]. Table 5 shows the induced systematic errors for replacing a single isotope at a time, and for replacing all modeled water isotopes simultaneously.

Table 5 shows that the retrieval of HDO/H_2O can be very sensitive to spectroscopic uncertainties, especially since HDO and H_2O can show sensitivities with opposite sign, that strengthen each other when taking the ratio (as can be seen from replacing the $H_2{}^{16}O$ parameters). The differences in spectroscopy between HITRAN2008 and [RD40] can lead to differences in δD of up to 128‰. Although we find that the differences do not depend on surface albedo or SZA, we cannot exclude a dependency with the total amount of water vapour that might lead to seasonal and latitudinal biases. Similar to the retrieval of CO [RD36], the HDO/H_2O retrieval will very likely benefit from a reassessment of the spectroscopic line parameters of water. Regardless of such reassessments, validation studies will be needed to select the best spectroscopy and to define corrections that might mitigate spectroscopy related biases.

8.9 Spectral sampling

The default HDO/H_2O retrieval works with a spectral sampling of $0.01\ \mathrm{cm}^{-1}$ for its internal line-by-line calculations. The same sampling has been used in the calculations of the measurement simulations, which means that possible biases due to this sampling setting will not be detectable. We have tested how sensitive the retrieval is to this sampling by first reprocessing the cloud-free measurement simulations (Case A) with a higher sampling (of $0.001\ \mathrm{cm}^{-1}$). We then retrieved these simulations using different, lower resolution, sampling settings, and compared the output to a retrieval that also used the higher sampling of $0.001\ \mathrm{cm}^{-1}$.

From this comparison we find that with our default sampling setting of $0.01\ \mathrm{cm}^{-1}$, we have a only very small systematic error in δD of 0.31 ± 0.11‰ and a systematic error in CH_4 of $-0.28 \pm 0.011\%$. The errors in H_2O and HDO are less than 0.05%. Table 6 shows that these errors increase for lower resolution sampling settings. For each sampling setting, the largest errors in δD were found for the highest signals (i.e. the highest surface albedos and lowest SZAs).

By reducing the sampling resolution, the processing time will reduce almost linearly (see Sect. 9 below, at the cost of introducing systematic errors in δD. It might be possible to reduce these systematic errors (this is ongoing work) by combining a lower sampling resolution with a triangular convolution scheme (see Sect. 5.4 in [RD24]). This could lead to a faster retrieval algorithm without loss of performance.

Table 6: Mean systematic errors as a function of spectral sampling resolution.

Sampling [cm^{-1}]	Systematic Error			
	H_2O [%]	HDO [%]	CH_4 [%]	δD [‰]
0.005	$-4.4 \pm 8.6 \times 10^{-4}$	$0.025 \pm 1.7 \times 10^{-5}$	$-3.0 \pm 0.76 \times 10^{-3}$	$3.7 \pm 7.1 \times 10^{-3}$
0.01	-0.037 ± 0.011	$2.2 \pm 1.5 \times 10^{-4}$	-0.28 ± 0.011	0.31 ± 0.11
0.02	-0.18 ± 0.051	0.038 ± 0.071	-1.2 ± 0.046	1.8 ± 0.55
0.03	-0.37 ± 0.11	0.11 ± 0.16	-2.5 ± 0.10	4.0 ± 1.3
0.04	-0.64 ± 0.18	0.35 ± 0.28	-3.9 ± 0.18	8.3 ± 2.3
0.05	-0.80 ± 0.26	0.48 ± 0.39	-5.4 ± 0.28	11 ± 3.2
0.1	-3.5 ± 0.71	1.1 ± 0.70	-11 ± 0.99	40 ± 5.8

8.10 Conclusions

As a result of our sensitivity studies, we find that the HDO/H_2O algorithm is most sensitive to uncertainties in the a priori temperature profile, followed by uncertainties in the a priori water and pressure profiles. The sensitivity to other a priori or instrument parameters is about an order of magnitude smaller (see Table 4). The uncertainties, however, should mostly be of a random nature, which means they should diminish when taking larger averages in time and space. Incorrect water spectroscopy could potentially induce large systematic uncertainties as well, but as mentioned above, reassessments of the spectroscopy and validation studies should be used to constrain these effects.

Table 7: HDO/H_2O computation times for two different resolution settings and fortran compilers.

Internal sampling (cm^{-1})	Time per spectrum (s)	
	Intel	gfortran
0.01	0.34083	0.41367 ($+21.371\%$)
0.03	0.12593	0.15068 ($+19.653\%$)

9 Feasibility

We have estimated the computational effort of the HDO/H_2O algorithm assuming a continuous data flow of 260 spectra/sec during 50 minutes of the 100 minute orbit period. This results in a processing request of 130 spectra/sec. We measured the average processing time per spectrum by processing the 187 measurement simulations from Case A, using a single CPU of a local computer. The local computer is an HP dc7800 SFF with an Intel® Core™2 Duo CPU E6750 at 2.66GHz and 6 GB RAM. For compilation we have used the Intel fortran compiler as well as the GNU gfortran compiler, both with optimisation flag O2. The results are summarised in Table 7 for two different internal sampling resolutions for the line-by-line calculations (0.01 and 0.03 cm^{-1}). Other parameter settings can also impact the processing time, such as the convergence and step-control settings, but these have been kept at their baseline values for the current study. Assuming the (slower) gfortran compiler and a sampling resolution of 0.01 cm^{-1}, the averaging computation time per spectrum is 0.41367 seconds (excluding file I/O). Reducing the sampling resolution by a factor 3, we gain a factor 2.7 in computational speed (0.15068 seconds).

A further reduction in computation time will result from rejecting (a priori) most cloud/aerosol contaminated measurements above land and oceans, and most cloud-free measurements above oceans. A priori cloud filtering takes place using the two-band CH_4 cloud filter from the pre-processor (see Appendix A). Based on a CH_4 threshold of 6%, Fig. 9 and the results from [RD42], we estimate that 20% of the measurements will pass the cloud filter. Assuming a land fraction of 30%, this results in a processing request of $0.20 \cdot 0.30 = 6.0\%$ of the measurements (i.e. 7.8 spectra/sec.), if only measurements above land would be considered. Above oceans we actually rely on the existence of low-level clouds, as the ocean surface itself is too dark to reflect a signal high enough to be measured reliably. These dark ocean measurements will be rejected (a priori) using the LER threshold value from the pre-processor. Using cloud information from one year of data from the MODIS instrument on board the Terra satellite, we estimate that about 14% of all the measurements above oceans are measured above (partially) cloudy scenes with a cloud top pressure higher than 900 hPa (i.e. a cloud top height less than ~ 1 km). These scenes are expected to come through the two-band CH_4 cloud filter with a threshold of 6% (see Fig. 9). These ocean scenes add another $0.14 \cdot 0.70 \cdot 130 = 13$ spectra/sec. to the processing request.

Combining land and ocean measurements, we come to a total processing request of 21 spectra/sec. To handle this data stream with a resolution of 0.03 cm^{-1}, 3 cores of the prescribed hardware are needed. If the higher accuracy of the 0.01 cm^{-1} sampling resolution is assumed, 9 cores would be needed. These numbers will be further reduced by considering up-to-date hardware, and/or by more strict cloud filtering using a lower two-band CH_4 threshold.

10 Validation

10.1 TCCON and NDACC-MUSICA

For the validation of HDO/H_2O product we can use the TCCON and NDACC-MUSICA networks of ground-based, direct-sun, Fourier-transform spectrometers. Both networks consist of 10–20 globally distributed ground stations (a number of stations are part of both networks), and measure a variety of gases, including H_2O and HDO [RD11, RD12]. The TCCON network uses micro-windows in the SWIR (around 1.6 μm and 2.2–2.4 μm) to measure column-averaged abundances, and the MUSICA network uses multiple micro-windows in the mid-IR (between 3.4–3.8 μm) to measure both column-averaged abundances and profiles.

The inclusion of the ratio HDO/H_2O is rather new for these networks, which means that the product is not as well calibrated and validated as other species. The TCCON retrievals of HDO, for example, are uncalibrated, because there have been no WMO-standard profiles of HDO measured over the TCCON stations. Caution is therefore recommended when using the data, but this situation is expected to improve with time.

10.2 GOSAT

Complimentary to the validation with the ground-based networks, an inter-comparison with HDO/H_2O measurements from the GOSAT satellite will be very useful [RD17, RD18]. The GOSAT HDO/H_2O measurements are taken with a Fourier-transform spectrometer around 1.56 μm and provide a completely independent set of global observations, also for areas where the density of ground-based FTS stations is low. It is important, however, that the GOSAT measurements are also validated with ground-based data, and that differences in vertical sensitivity are accounted for, e.g. by performing tests of applying the averaging kernels of GOSAT and TROPOMI to the same model H_2O and HDO profiles. The GOSAT HDO/H_2O measurements are currently available for the period 2009–2011. It is expected that this dataset will be extended with time, however, it is not guaranteed that GOSAT measurements will be available in the operational time frame of TROPOMI.

Acknowledgements

We would like to thank Michael Buchwitz and Thomas Krings (IUP, University Bremen, Germany) for their constructive comments that aided us to improve this document.

A Appendix: SWIR Pre-Processing

To optimise synergies between the different SWIR retrievals, i.e. the CO and HDO/H_2O retrieval described in this document and the CH_4 retrieval as described by [RD43], a SWIR pre-processing module is developed. Exploiting spectral information from the SWIR spectral ranges, the module provides the following auxiliary products:

- A cloud filter based on a non-scattering retrieval of the total amount of CH_4 and a corresponding column estimate from the TM5 forecast (CH_4 a priori cloud filter)
- A cloud filter using a non-scattering retrieval of the total amount of CH_4 in a strong and weak methane absorption band (CH_4 two-band cloud filter)
- A cloud filter using a non-scattering retrieval of the total amount of H_2O in a strong and weak water absorption band (H_2O two-band cloud filter)

Technically, the pre-processing module is implemented in the CO prototype software to reduce interfaces with the processing framework as indicated in Fig. 2 in [RD24]. However, functionally the module does not depend on the CO retrieval. The overall algorithm structure is summarised in Fig. 15. Via the CO interface to the processing framework, the TROPOMI measurements in the SWIR spectral range are allocated in combination with the meteo fields for temperature, pressure and water vapour abundance, information on the pixel's orography and the model forecast of the CH_4 and CO profiles. The different auxiliary products are processed in a hierarchical manner to optimise the computational effort. In a first step, only measurements are accepted with a continuum LER value exceeding a certain threshold. Due to this, only measurements are processed with an acceptable signal level. For example, clear sky ocean measurements are rejected by this inquiry due to the very low ocean surface reflection in the SWIR spectral range. Subsequently, the CH_4

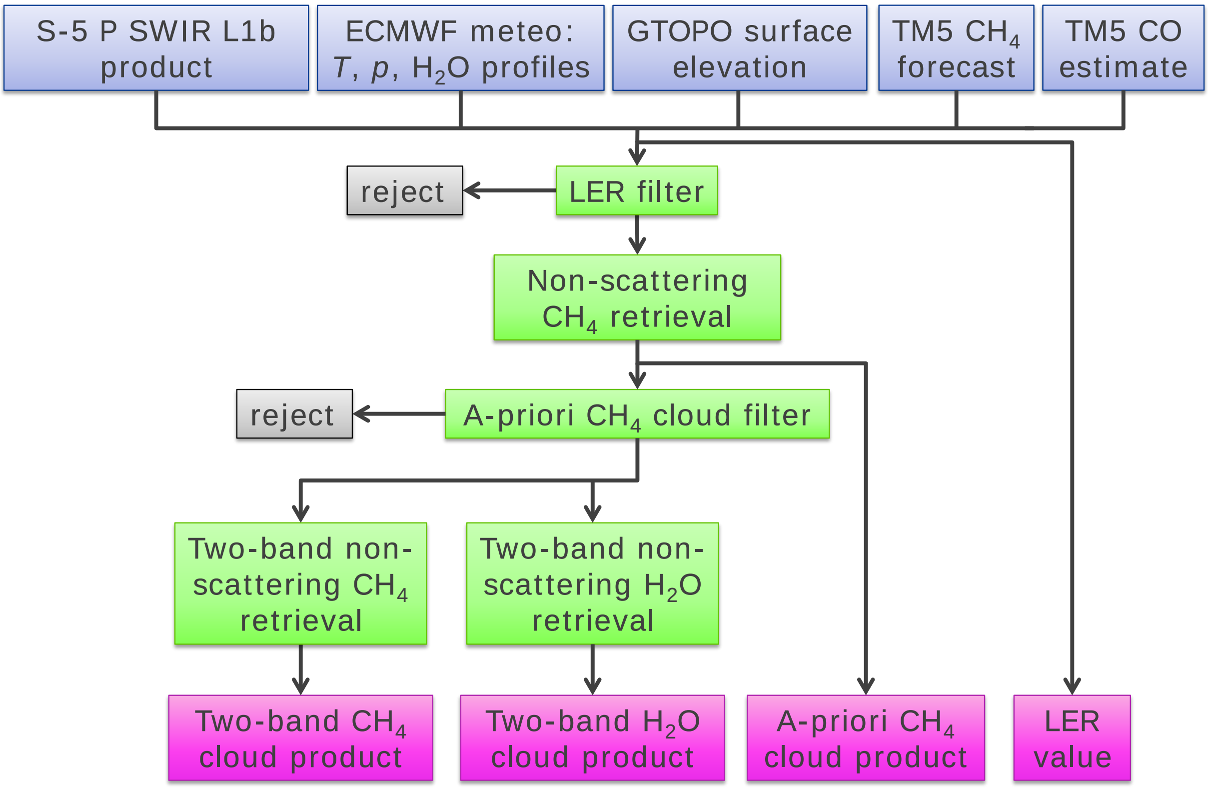

Figure 15: Overall structure of the SWIR preprocessor.

a priori filter is processed. A comparison of a non-scattering methane column retrieval with a model a priori methane estimate is used to reject spectral observations contaminated by high and optically thick clouds. This cloud filter is required for the SICOR CO retrieval, which is designed to process clear sky SWIR observation and measurements of scenes with low water clouds and optically thin cirrus and aerosol loads (see Sec. 5 in [RD24]). For the processing of CH_4 and HDO/H_2O an even stricter cloud screening is required. Therefore, the remaining auxiliary products, viz. the CH_4 and H_2O two-band cloud filters, are only processed for observations which pass the CH_4 a priori filter. The retrieval of the auxiliary products are described in more detail below.

Non-scattering total column retrieval

The cloud filters of the SWIR processing are based upon a non-scattering retrieval of a total column abundance from dedicated spectral windows in the SWIR spectral range. Here, we employ the inversion module as described in Sec.5.2 in [RD24] to infer the total column abundance of the relevant trace gases using the profile scaling approach together with the retrieval of an effective Lambertian surface albedo and a spectral shift of the forward model. Cloud properties are not retrieved from the measurement and so a regularisation of the least squares solution to maintain the numerical stability, as described in Sec. 5.2.3 in [RD24], is not required. Moreover, the forward model employs a radiative transmission model that ignores atmospheric scattering. Here, sunlight is reflected at the Earth surface into the satellite line of sight (LOS) and is attenuated by atmospheric absorption along its path. Using this approximation, the simulated radiance at the TOA $I^{\mathrm{TOA}}(\lambda)$ is given by:

$$I^{\mathrm{TOA}}(\lambda) = A_s(\lambda)\frac{\mu_o F_o}{\pi}\exp(-\frac{1}{\tilde{\mu}}\tau_{\mathrm{tot}}(\lambda))\,, \tag{8}$$

where A_s is the surface albedo, $\mu_o = \cos(\Theta_o)$ with the solar zenith angle Θ_o. This cosine is corrected for the sphericity of the Earth according to Kasten and Young (1989) [RD44]. F_o is the solar irradiance and

$$\frac{1}{\tilde{\mu}} = \frac{\mu_o + \mu_v}{\mu_o \mu_v} \tag{9}$$

is the air mass factor with $\mu_v = \cos(\Theta_v)$ and viewing zenith angle Θ_v. The total optical thickness τ_{tot} is given by

$$\tau_{\mathrm{tot}}(\lambda) = \sum_k \int \sigma_k(z,\lambda)\,\rho_k(z)\,dz\,, \tag{10}$$

where z indicates the altitude, index k represents the relevant absorbers CO, CH_4 H_2O and HDO, $\rho_k(z)$ is the concentration of absorber k at altitude z, $\sigma_k(z,\lambda)$ is the corresponding wavelength-dependent absorption cross sections.

In the following, we assume that the relative profile

$$\rho_k^{\mathrm{rel}} = \frac{\rho_k}{c_k} \tag{11}$$

of absorber k is constant, where

$$c_k = \int \rho_k(z)dz \tag{12}$$

is the column density of this absorber. So,

$$\frac{\partial \tau_{\mathrm{tot}}}{\partial c_k} = \frac{1}{c_k}\int \sigma_k(z,\lambda)\,\rho_k(z)\,dz \tag{13}$$

and thus the derivative with respect to the total column amount c_k of a trace gas k is given by

$$\frac{\partial I^{\mathrm{TOA}}}{\partial c_k} = -\frac{I^{\mathrm{TOA}}}{\tilde{\mu} c_k}\int \sigma_k(z)\rho_k(z)dz\,. \tag{14}$$

Finally, the derivative of I^{TOA} with respect to surface albedo A_s is

$$\frac{\partial I^{\mathrm{TOA}}}{\partial A_s} = \frac{\mu_o F_o}{\pi} \exp(-\frac{1}{\tilde{\mu}} \tau_{\mathrm{tot}}) \,. \tag{15}$$

To account for the spectral instrument response, the TOA radiance and its derivatives have to be convoluted with the ISRF (For details, see [RD24], specifically Eqs. 3 and Eqs. 5)

CH_4 a priori cloud filter

The methane a priori cloud filter relies on a non-scattering methane column retrieval. In the presence of high and optically thick clouds, the lightpath gets effectively shortened due to reflection of light by clouds. Thus assuming a clear sky model atmosphere, the retrieved methane total column is underestimated depending on cloud occurrence in the observed scene. Figure 16 shows the induced error in the non-scattering methane column retrieved from the spectral window 2315–2324 nm for two cloudy scenes. For a water cloud with an optical depth of 5, the methane error increases with cloud coverage and cloud height. In case of a cirrus at 10 km height, the non-scattering CH_4 column can be used to identify cloud cirrus contamination with optical depth > 0.5 for low and moderate surface reflection. However for bright surfaces, the lightpath shortening due the reflection of light by the cirrus is compensated by an enhancement of the lightpath because of multiple reflection of light between the cirrus layer and the surface. In this case, the non-scattering CH_4 column is less suited for cirrus detection. Keeping in mind that the current accuracy of the methane column forecast is in the order of 2–3 % using state-of-the-art chemical transport models (see Sec. 7.6 in [RD24]), the CH_4 a priori cloud filter is a powerful tool to screen measurement with respect to the presence of high and optically thick clouds.

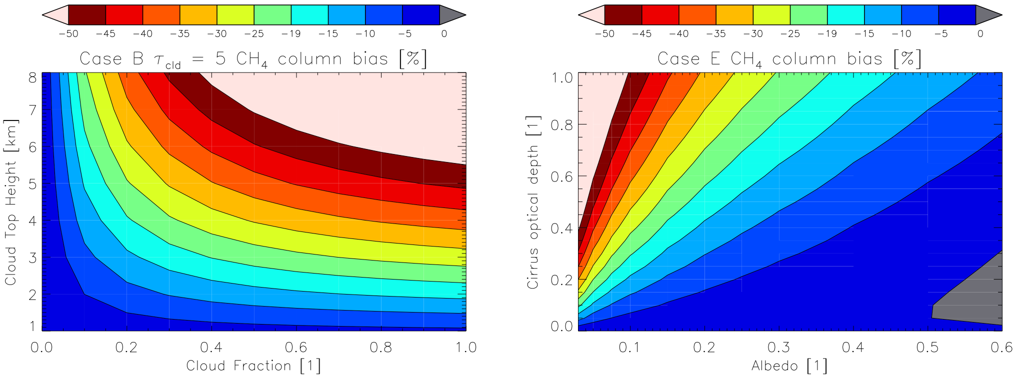

Figure 16: CH_4 error of a non-scattering retrieval from the SWIR 2315–2324 nm spectral window for a water cloud with optical thickness of 5 as function of cloud height and cloud fraction (left panel, for more details see generic scenario B in Sec. 7) and for a cirrus cloud at 10 km height as function of surface albedo and cirrus optical thickness (right panel, for more details see generic scenario E in Sec. 7).

CH_4 two-band cloud filter

Complementary to the CH_4 a priori filter, the SWIR pre-processor includes a cloud filter based on two methane non-scattering retrievals that utilise two different spectral bands with a strong and weak absorption band, respectively. This method relies on the fact that in a strong absorption band, photons along an enhanced light path contribute less to the total signal than in a weak absorption band. Thus, we expect that the non-scattering methane column, which are inferred from a weak absorption, exceeds the corresponding column retrieved from a strong absorption band. Moreover, the difference between the two columns indicates changes of the lightpath due to atmospheric scattering and so can be used as a cloud filter. In Fig. 17, this difference is depicted for the cloud scenarios of Fig. 16 using the strong CH_4 absorption at 2363-2373 nm and the weak absorption at 2310-2315 nm. The figure indicates that the methane two-band retrieval is well suited to detect cirrus cloud even above bright surfaces and complements the CH_4 a priori filter. Therefore, this cloud filter is particular valuable for cirrus screening needed by the CH_4 and HDO/H_2O data processing.

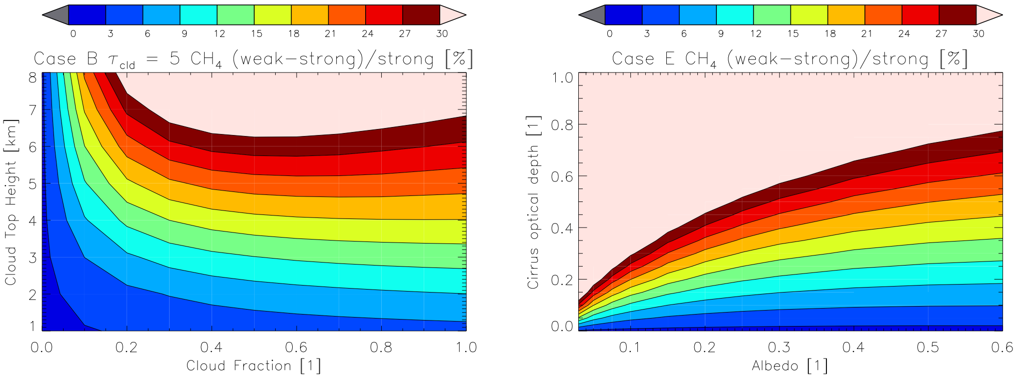

Figure 17: CH_4 two-band cloud filter for the cloud scenarios of Fig. 16. The methane cloud filter relies on non-scattering methane column retrieval from strong and weak absorption features at 2363-2373 nm and 2310-2315 nm, respectively.

H_2O two-band cloud filter

The concept of the two-band CH_4 cloud filter can be applied to H_2O as well. The two-band cloud filter does not rely on a-priori knowledge, so the poor a-priori knowledge of H_2O does not hamper the application of the two-band cloud filter for water. Compared to CH_4, H_2O absorption lines are narrower, resulting in a different sensitivity towards pressure broadening close to the surface. The H_2O two-band cloud filter is set up analogous to the CH_4 two-band cloud filter. Here, the window 2329-2334 nm contains weak H_2O absorption and the window 2367-2377 nm includes strong H_2O absorption features.

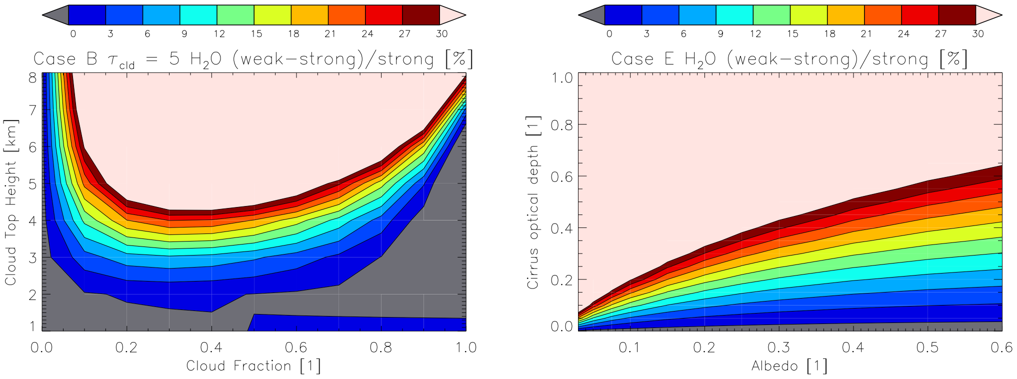

Figure 18: H_2O two-band cloud filter for the cloud scenarios of Fig. 16. The filter relies on non-scattering methane column retrieval from strong and weak absorption features at 2367-2377 nm and 2329-2334 nm, respectively.

Figure 18 shows the relative difference in the retrieved water columns using the weak and strong absorption bands. Compared to the CH_4 two-band cloud filter in Fig. 17, the H_2O two-band filter shows larger sensitivity to clouds and so can be used to compliment the other cloud filters.