Sentinel-5p+ Innovation (S5p+I) - Water Vapour Isotopologues (H2O-ISO)

Algorithm Theoretical Basis Document (ATBD)

Authors:
Tim Trent and Hartmut Bösch: School of Physics and Astronomy, University of Leicester, Leicester and National Centre for Earth Observation NCEO, United Kingdom
Matthias Schneider, Farahnaz Khosrawi, and Christopher Diekmann: Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe Institute of Technology, Karlsruhe, Germany
Harald Sodemann, Geophysical Institute, University of Bergen, Bergen, Norway

Change log

Version	Date	Status	Authors	Reason for change
Draft 0.1	24-Feb-2020	Initial internal draft for project team	T. Trent, H. Bösch	New document
Draft 1.0	6-April-2020	Sent to ESA		Consolidated version
Draft 1.1	29-April-2020	Document amended for acceptance by ESA		Updates based on ESA comments

Table of Contents

1 Purpose and objective 5

2 Document overview 5

3 References, terms and acronyms 6

3.1 Reference Documents 6
3.2 Applicable Documents 8
3.3 Terms, definitions and abbreviated term 8

4 TROPOMI instrument description 9

5 Introduction to the TROPOMI product 10

6 Algorithm description 16

6.1 The Leicester Retrieval Preparation Toolset 16
6.1.1 Cloud Masking 18
6.1.2 Ocean Pixel Masking 19
6.1.3 Addition of state vector a priori data. 19
6.2 Full Physics Algorithm 20
6.2.1 Overview 20
6.2.2 The Inverse Method 21
6.3 Forward Model 22
6.3.1 Solar Model 22
6.3.2 Radiative Transfer 22
6.4 Instrument Model 23
6.4.1 Instrument Noise 23
6.4.2 Pixel-Wavelength Mapping 23
6.4.3 Instrument Lineshape Function 24
6.4.4 Zero-level Offset 24
6.5 State Vector 24
6.5.1 Description of Stable Water Isotoplogue Retrievals 24
6.5.2 Methane Concentration 25
6.5.3 Carbon Monoxide Concentration 25
6.5.4 Surface Pressure 25
6.5.5 Temperature and Water Vapour 26
6.5.6 Surface Properties 27
6.5.7 Cross sections 27
6.5.8 A Prioiri Error covariance 27

7 Feasibility 28

8 Error analysis 28

9 Validation 28

10 Conclusion 28

1 Purpose and objective

The purpose of this document is to describe the theoretical basis and the implementation of the S5p+I Level-2 product algorithm for water isotopologues for TROPOMI. The document is maintained during the development phase of the data products in the context of S5p+I. Two updates of the document are planned.

2 Document overview

This document covers the algorithm theoretical basis for the parameters of the Level 2 products that are produced at the University of Leicester within the scope of the Sentinel-5p (S5p) innovation project. Section 1 provides the purpose and scope of the document. Section 2 describes the input and auxiliary data used and its flow through the pre-processor. Section 3 describes the processing and algorithm and section 4 gives the references for publications cited in the document.

3 References, terms and acronyms

3.1 Reference Documents

RD1	Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., Shepherd, T. G., Sherwood, S. C., Siebesma, A, P., Sobel, A. H., Watanabe, M. and Webb, M. J.: Clouds, ciculation and climate sensitivity, Nature Geosci., 8, 261-268, doi:10.1038/NGEO2398, 2015.
RD2	Fink, A. H., S. Pohle, J. G. Pinto, and P. Knippertz: Diagnosing the influence of diabatic processes on the explosive deepening of extratropical cyclones, Geophys. Res. Lett., 39, L07803, doi:10.1029/2012GL051025, 2012.
RD3	Evans, C., K.M. Wood, S.D. Aberson, H.M. Archambault, S.M. Milrad, L.F. Bosart, K.L. Corbosiero, C.A. Davis, J.R. Dias Pinto, J. Doyle, C. Fogarty, T.J. Galarneau, C.M. Grams, K.S. Griffin, J. Gyakum, R.E. Hart, N. Kitabatake, H.S. Lentink, R. McTaggart-Cowan, W. Perrie, J.F. Quinting, C.A. Reynolds, M. Riemer, E.A. Ritchie, Y. Sun, and F. Zhang: The Extratropical Transition of Tropical Cyclones. Part I: Cyclone Evolution and Direct Impacts, Mon. Wea. Rev., 145, doi:10.1175/MWR-D-17-0027.1, 2017.
RD4	Noone, D.: Pairing Measurements of the Water Vapor Isotope Ratio with Humidity to Deduce Atmospheric Moistening and Dehydration in the Tropical Midtroposphere, J. Climate, 25, 4476–4494, doi:10.1175/JCLI-D-11-00582.1, 2012
RD5	Sherwood, S. C., Bony, S., and Dufresne, J.-L.: Spread in climate model sensitivity traced back to atmospheric convective mixing, Nature, 505, 37-42, doi:10.1038/nature12829, 2014.
RD6	Gat, J.: Oxygen and hydrogen isotopes in the hydrologic cycle, Annu. Rev. Earth Planet. Sci., 24, 225–262, 1996.
RD7	Hurley, J. V., and J. Galewsky: A last-saturation diagnosis of subtropical water vapor response to global warming, Geophys. Res. Lett., 37, L06702, doi:10.1029/2009GL042316, 2010.
RD8	González, Y., Schneider, M., Dyroff, C., Rodríguez, S., Christner, E., García, O. E., Cuevas, E., Bustos, J. J., Ramos, R., Guirado-Fuentes, C., Barthlott, S., Wiegele, A., and Sepúlveda, E.: Detecting moisture transport pathways to the subtropical North Atlantic free troposphere using paired H2O-δD in situ measurements, Atmos. Chem. Phys., 16, 4251-4269, doi:10.5194/acp-16-4251-2016, 2016
RD9	Field, R.D., Kim, D., LeGrande, A.N., Worden, J., Kelley, M., and Schmidt, G.A.: Evaluating climate model performance in the tropics with retrievals of water isotopic composition from Aura TES, Geophys. Res. Lett., doi:10.1002/2014GL060572, 2014
RD10	Parker, R., Boesch, H., Cogan, A., et al., Methane Observations from the Greenhouse gases Observing SATellite: Comparison to ground-based TCCON data and Model Calculations, Geophys. Res. Lett., doi:10.1029/2011GL047871, 2011
RD11	Boesch, H., Toon, G. C., Sen, B., Washenfelder, R. A., Wennberg, P. O., Buchwitz, M., de Beek, R., Burrows, J. P., Crisp, D., Christi, M., Connor, B. J.,

	Natraj, V., and Yung, Y. L.: Space-based near-infrared CO2 measurements: Testing the Orbiting Carbon Observatory retrieval algorithm and validation concept using SCIAMACHY observations over Park Falls, Wisconsin, J. Geophys. Res., 111, D23302, doi:10.1029/2006JD007080, 2006.
RD12	Boesch, H., D. Baker, B. Connor, D. Crisp, and C. Miller, Global characterization of CO2 column retrievals from shortwave-infrared satellite observations of the Orbiting Carbon Observatory-2 mission, Remote Sensing, 3 (2), 270-304, 2011
RD13	Boesch, H., Deutscher, N. M., Warneke, T., Byckling, K., Cogan, A. J., Griffith, D. W. T., Notholt, J.,Parker, R. J., and Wang, Z.: HDO/H2O ratio retrievals from GOSAT, Atmos. Meas. Tech., 6, 599-612, https://doi.org/10.5194/amt-6-599-2013, 2013
RD14	Trent, T.; Boesch, H.; Somkuti, P.; Scott, N.A. Observing Water Vapour in the Planetary Boundary Layer from the Short-Wave Infrared. Remote Sens. 2018, 10, 1469
RD15	Somkuti, P., Bösch, H., Feng, L., Palmer, P.I., Parker, R.J. and Quaife, T., 2020. A new space-borne perspective of crop productivity variations over the US Corn Belt. *Agricultural and Forest Meteorology*, *281*, p.107826
RD17	R Siddens (2016): S5P-NPP Cloud Processor ATBD, source: S5P-NPPC-RAL-ATBD-0001
RD18	Landgraf, J., aan de Brugh, J., SCHEEPMAKER, R., BORSDORFF, T., HOUWELING, S. and HASEKAMP, O., 2018. Algorithm Theoretical Baseline Document for Sentinel-5 Precursor: Carbon Monoxide Total Column Retrieval.
RD19	D. Crisp, B. M. Fisher, C. O'Dell, C. Frankenberg, R. Basilio, H. Bösch, L. R. Brown, R. Castano, B. Connor, N. M. Deutscher, A. Eldering, D. Griffith, M. Gunson, A. Kuze, L. Mandrake, J. McDuffie, J. Messerschmidt, C. E. Miller, I. Morino, V. Natraj, J. Notholt, D. M. O'Brien, F. Oyafuso, I. Polonsky, J. Robinson, R. Salawitch, V. Sherlock, M. Smyth, H. Suto, T. E. Taylor, D. R. Thompson, P. O. Wennberg, D. Wunch, and Y. L. Yung, Atmos. Meas. Tech., 5, 687-707, 2012.
RD20	Rodgers C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific Publishing, 2000
RD21	Connor, B. J., H. Boesch, G. Toon, B. Sen, C. Miller, and D. Crisp, Orbiting Carbon Observatory: Inverse method and prospective error analysis, J. Geophys. Res., 113, D05305, doi:10.1029/2006JD008336, 2008.
RD22	Spurr, R. J. D., Kurosu, T. P., and Chance, K. V.: A linearized discrete ordinate radiative transfer model 940 for atmospheric remote-sensing retrieval, J. Quant. Spectrosc. Radiat. Transfer, 68, 689–735, doi:10.1016/S0022-4073(00)00055-8, 2001.
RD23	Spurr, R. J. D., Simultaneous derivation of intensities and weighting functions in a general pseudo-spherical discrete ordinate radiative transfer treatment, J. Quant. Spectrosc. Rad. Tran., 75(2), 129–175, 2002
RD24	O'Dell, C.W.: Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation, J. Geophys. Res., 115, D10 206, doi:10.1029/2009JD012803, 2010
RD25	Somkuti, P., Boesch, H., Natraj V., & Kopparla, P. (2017). Application of a PCA-Based Fast Radiative Transfer Model to XCO2 Retrievals in the Shortwave Infrared. *J. Geophys. Res. Atmos.*, 122, 10477–10496. https://doi.org/10.1002/2017JD027013

RD26	Nakajima, T. and Tanaka, M.: Algorithms for radiative intensity calculations in moderately thick atmospheres using a truncation approximation. J. Quant. Spectrosc. Radiat. Transfer, 40, 51–69, doi:10.1016/0022-4073(88)90031-3, 1988.
RD27	Natraj, V. and Spurr, R. J. D.: A fast linearized pseudo-spherical two orders of scattering model to account for polarization in vertically inhomogeneous scattering absorbing media, J. Quant. Spectrosc. Radiat. Transfer, 107, 263–293, doi:10.1016/j.jqsrt.2007.02.011, 2007
RD28	Scheepmaker, R.A., Hu, H., Borsdorff, T., Frankenberg, C., Risi, C., Hasekamp, O., Aben, I. and Landgraf, J., 2016. HDO and H2O total column retrievals from TROPOMI shortwave infrared measurements. Atmospheric Measurement Techniques, 9(8), pp.3921-3937
RD29	R.V. Kochanov, I.E. Gordon, L.S. Rothman, P. Wcislo, C. Hill, J.S. Wilzewski, HITRAN Application Programming Interface (HAPI): A comprehensive approach to working with spectroscopic data, J. Quant. Spectrosc. Radiat. Transfer 177, 15-30 (2016) DOI: 10.1016/j.jqsrt.2016.03.005

3.2 Applicable Documents

AD1	TROPOMI Instrument and Performance Overview. source: KNMI; ref: S5p-KNMI-L2-0010-RP; issue: 0.10.0; date: 2014-03-15.
AD2	T. Trent, H. Boesch, M. Schneider, F. Khosrawi, C.Diekmann and H. Sodemann (2020), Sentinel-5p+ Innovation (S5p+I) - Water Vapour Isotopologues (H2O-ISO) Auxiliary User Manual (AUM), URL: https://s5pinnovationh2o-iso.le.ac.uk/wp-content/uploads/2020/02/aum_s5p-i_iso_version_1.2.pdf
AD3	ECMWF Technical Notes, MARS User Guide, User Support, Operations Department, 2008

3.3 Terms, definitions and abbreviated term

Term	Definition
AUM	Auxiliary User Manual
CAMS	Copernicus Atmospheric Monitoring Service
CCI	Climate Change Initiative
CEDA	Centre for Environmental Data Analysis
ECMWF	European Centre for Medium Range Forecasting
ERA5	ECMWF 5th Reanalysis
EE8	Earth Explorer 8
ESA	European Space Agency
FTS	Fourier Transform Spectrometer
GHG	Greenhouse Gas
GOSAT	Greenhouse gases Observing Satellite

ILS	Instrument Line shape Function
IFOV	Instantaneous Field of View
L1b	Level 1b data product
L2	Level 2 data product
L4	Level 4 data product
LCC	Land Cover Change
LRPT	Leicester Retrieval Preparation Toolset
LSI	Low Streams Interpolation
LSM	Land Sea Mask
OCO	Orbiting Carbon Observatory
PCA	Principle Component Analysis
ROI	Region-of-Interest
RT	Radiative Transfer
RTE	Radiative Transfer Equation
S5p	Sentinel 5 precursor
S5p+I	Sentinel-5p+Innovation
SIF	Solar Induced Fluorescence
SRTM	Shuttle Radar Topography Mission
Suomi-NPP	Suomi National Polar-orbiting Partnership
SWIR	Shortwave Infrared
TOA	Top-of-Atmosphere
TROPOMI	TROPOspheric Monitoring Instrument
UoL-FP	University of Leicester Full Physics
VIIRS	Visible/Infra-red Imager and Radiometer Suit
VSMOW	Vienna Standard Mean Ocean Water
WCRP	World Climate Research Programme

4 TROPOMI instrument description

A description of the TROPOMI instrument and performance, referred to from all ATBDs, can be found in: TROPOMI Instrument and Performance Overview [AD1].

5 Introduction to the TROPOMI product

Atmospheric moisture is a key factor for the redistribution of heat in the atmosphere. Clouds and water vapour control radiative heating and cooling, whereas condensation and evaporation of water determines where latent heating (or latent heat consumption) takes place. The heating patterns feed then back on the atmospheric circulation, thereby causing additional evaporation/condensation and impacting on the distribution of water vapour and clouds, which in turn modify the latent and radiative heating patterns of the atmosphere. This strong coupling between atmospheric circulation and moisture pathways is one of the grand challenges of atmospheric research identified by the World Climate Research Programme (WCRP). It is responsible for most climate feedback mechanisms (e.g. RD1) and controls the evolution of severe weather events (e.g. RD2, RD3).

Water isotopologues can make a unique contribution for better understanding this coupling. The ratio between the heavy and light isotopologues contains information about the water source as well as the transport and small-scale in-cloud transformation processes. This way isotopologue ratios “tag” water and its phase change conditions, i.e. the isotopologues are directly linked to the moist processes of interest. These ratios can be modelled as well as measured and they have a large potential for investigating deficits in modelling the moisture pathways and thus for constraining climate feedback mechanisms and for improving weather forecasts. In the boundary layer the isotopologues can help to characterize boundary layer mixing (e.g. [RD4]), which would allow to better constrain the low cloud feedback in climate models [RD5]. Water originating from the land is more enriched in heavy isotopes than the ambient water vapor, i.e. the isotopologues can theoretically be used to distinguish the water evaporated over the ocean from the water that has been recycled over land [RD6]. In the free troposphere the isotopologue data reveal source regions and transport pathways as well as mixing on different scales [RD7], [RD8] and in-cloud transformation and transport processes (e.g. [RD9]).

The algorithm setup of HDO/H_2O from TROPOMI is based on the University of Leicester Full Physics (UoL-FP) processor which has been used to retrieve XCH_4, XCO_2, HDO/H_2O, planetary boundary layer (PBL) XH_2O and solar induced fluorescence (SIF) from a number of satellite instruments operating in the shortwave infrared [RD10, RD11, RD12, RD13, RD14,RD15]. The retrieval use a state structure that represents the atmosphere, surface and instrument. The state vector of our retrieval consists of scaling factors for H_2O, HDO, temperature, CO and CH_4 profiles, surface albedo. A 3rd order polynomial is also retrieved which describes the shift and stretch of the dispersion relation. In total, our retrieval state vector presently consists of 11 (or 49 for full a profile retrieval) elements for land cases as shown in Table 5-2. In contrast to GOSAT retrievals of XCO2 or PBL XH2O where the O_2 A band is used to retrieve aerosol and cirrus extinction coefficients, the scattering effects in the retrieved H_2O/HDO are cancelled out when ratioing similar to the CH_4 proxy retrieval [RD10]. Therefore, we do not include measurements O_2 A band from TROPOMI in our scheme. The setup of retrievals from the UoL-FP all use the same common frame work which is illustrated by Figure 5-1, with greater detail is given in Section 6.

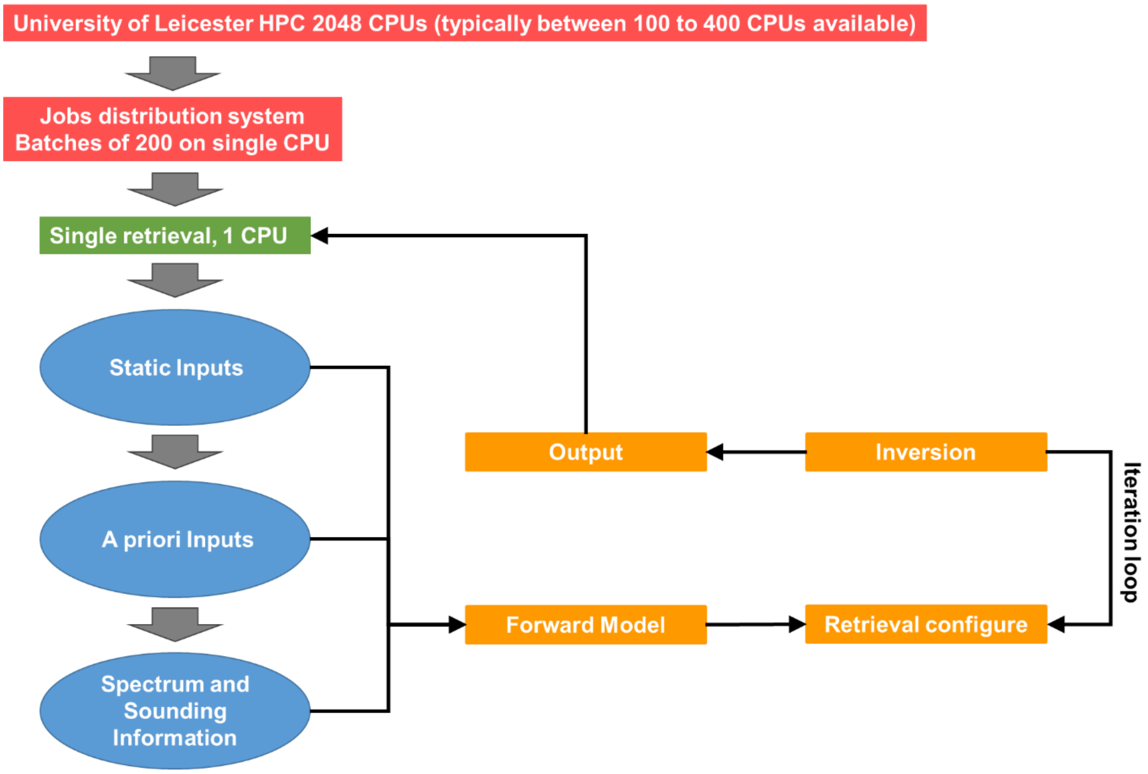

Figure 5-1: Schematic demonstrating retrieval outline within UoL-FP.

Output from the UoL-FP are aggregated into netCDF format following current naming conventions in used by existing TROPOMI products (Table 5-1).

Table 5-1: Convention for output filenames. All components use an underscore to separate them with the exception of the file extension, which uses a period. Indices for characters begin at 0. This convention is adapted from Table 2 in the Sentinel-5 precursor/TROPOMI Level 2 product user manual for cloud properties showing the components of an S5p product file name.

Start	End	Length	Meaning/Value
0	3	3	Mission name **S5P**
4	8	4	Processing stream **OFFL**
9	19	10	Product identifier **L2__H2O_IS**
20	35	15	Start of granule in UTC as “**YYYYMMDDTHHMMSS**”. The “T” is a fixed character
36	51	15	End of the granule in UTC as “**YYYYMMDDTHHMMSS**”. The “T” is a fixed character
52	57	5	Orbit number
58	60	2	Collection number
61	67	6	Processor version number as “**MMmmpp**”, with “MM” the major version number, “mm”the minor version number, and “pp” the patch level.

68	83	15	The time of processing for this granule in UTC as “**YYYYMMDDTHHMMSS**”. The “T”is a fixed character.
84	86	2	The file name extension **nc**

Data output structure is based on the current file structure used in existing C3S and CCI projects, with updates for the new water isotopologue variables and adoption of SI units in compliance with existing TROPOMI products (Table 5-2).

Table 5-2: Variables output to prototype TROPOMI L2 H2O isotopologue product.

Name	Type	Dimensions	Units	Description
solar_zenith_angle	float	n	degree	Angle between line of sight to the sun and local vertical
sensor_zenith_angle	float	n	degree	Angle between the line of sight to the sensor and the local vertical
time	double	n	seconds since 1970-01-01 00:00:00	Measurement time
longitude	float	n	degrees_east	Centre longitude
latitude	float	n	degrees_north	Centre latitude
pressure_levels	float	n, m	hPa	Vertical altitude coordinate in pressure units as used for averaging kernels
pressure_weight	float	n, m		Pressure weights as used for averaging kernels
xh2o	float	n	kg/kg	Retrieved column-averaged dry-air mole fraction of

				atmospheric water vapour (XH_2O)
xhdo	float	n	kg/kg	Retrieved column-averaged dry-air mole fraction of atmospheric deuterium hydrogen monoxide (XHDO).
x_delta_D	float	n	1	Retrieved column-averaged dry-air mole fraction of Deuterium ratio to VSMOW
xh2o_uncertainty	float	n	kg/kg	Statistical uncertainty of XH_2O in ppm (1σ)
xhdo_uncertainty	float	n	kg/kg	Statistical uncertainty of XHDO in ppm (1σ)
x_delta_D_uncertainty	float	n	1	Statistical uncertainty of Deuterium ratio (1σ)
retrieved_h2o_scaler	float	n	1	Retrieved H_2O scaler for h2o_profile_apriori
retrieved_h2o_scaler_uncertainty	float	n	%	Uncertainty of retrieved H_2O scaler for h2o_profile_apriori
retrieved_hdo_scaler	float	n	1	Retrieved HDO scaler for h2o_profile_apriori

retrieved_h2o_scaler_uncertainty	float	n	%	Uncertainty of retrieved HDO scaler for h2o_profile_apriori
xh2o_scaler_averaging_kernel	float	n, m	1	XH$_2$O averaging kernel (a profile = vector for each single observation). Quantifies the altitude sensitivity of the XH$_2$O retrieval
xhdo_scaler_averaging_kernel	float	n, m	1	XHDO averaging kernel (a profile = vector for each single observation). Quantifies the altitude sensitivity of the XH$_2$O retrieval
exposure_id	char	n, 22	n/a	Exposure identification number of the sounding
surface_altitude	float	n	metres	Altitude is the (geometric) height above the geoid, which is the reference geopotential surface
surface_altitude_stdev	float	n	metres	Standard deviation of the surface elevation within the area of the GOSAT sounding, as derived from

				the SRTM database
surface_air_pressure_apriori	float	n	hPa	A-priori surface pressure value
surface_air_pressure_apriori_std	float	n	hPa	A-priori surface pressure standard deviation
air_temperature_apriori	float	n, m	K	Air temperature is the bulk temperature of the air, not the surface (skin) temperature
h2o_profile_apriori	float	n, m	kg/kg	A-priori mole fraction profile of atmospheric H_2O
hdo_profile_apriori	float	n, m	kg/kg	A-priori mole fraction profile of atmospheric HDO
r_outcome	int	n	1	Retrieval outcome flag
r_chi2	float	n	1	Retrieval $\chi 2$ value
r_num_div	int	n	1	Number of divergent steps in the retrieval
retr_flag	int	n	1	Retrieval type flag (0 = land, 1 = glint)

6 Algorithm description

6.1 The Leicester Retrieval Preparation Toolset

The preparation of data for use in the retrieval is processed through the Leicester Retrieval Preparation Toolset (LRPT). The LRPT has heritage at the University of Leicester (UoL) in the production of GOSAT XCO2, XCH4, HDO/H_2O, PBL XH2O and SIF L2 products. Further adaptations have been made to include the OCO-2, TANSAT and Microcarb missions. For S5p, a number of TROPOMI products are used as inputs into the retrieval, which are matched at the orbit level. These consist of two types: the L1b radiance and L2 cloud products. With eight available bands, the LRPT driver file can take a list of the required bands that the LRPT will loop over. For the purposes of stable water vapour isotopologues we use the both bands 7 and 8 (L1B_RA_BD7, L1B_RA_BD8 respectively) that cover shortwave infrared (SWIR) region and use the corresponding TROPOMI L2 VIIRS cloud mask product, L2__NP_BD7 [RD17].

Once matches between TROPOMI L1B and VIIRS L2 files have been made, the individual pixels are screened by a series of filters to product the final set that will be passed to the retrieval processor. The first step is to apply a region of interest (ROI) filter, this is to allow for focused study regions to be processed. However, the default setting is global which currently removes pixels below 60°S. The second stage is to apply a cloud filter which is constructed from the TROPOMI L2__NP_BD7 product. The final filter applies tests for surface type (land or water bodies), solar zenith angle and valid data ranges for geo-temporal TROPOMI variables. An example of this process is shown in Figure 6-1.

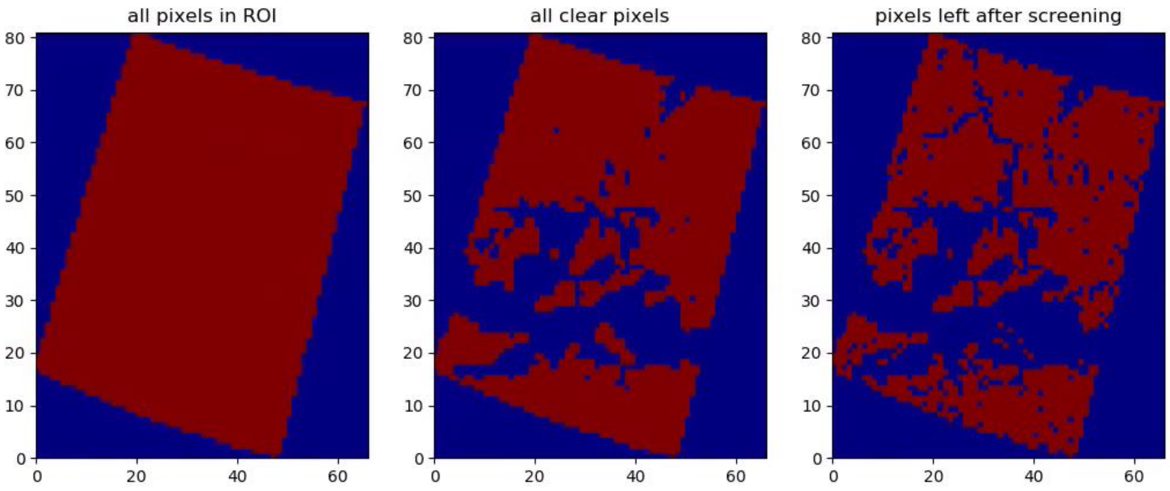

Figure 6-1: Example of the three filtering steps applied to TROPOMI pixels within a 5°x5° region of interest around KIT. Red indicates a valid TROPOMI pixel at each stage of filtering, where blue points have no data.

A list of unique string IDs are generated for the remaining TROPOMI pixels with the format:

<orbit_number>_<across_track_location>_ <along_track_position>

The unique IDs are then used to subset all variables needed within the full retrieval processor. When multiple bands are being used, all band radiances are combined to form single spectra per pixel. Albedo is added as an additional variable at this point for filtered pixels (Section 6.5.6). Once collated, the observed radiances and all variables that describe the time, location, viewing geometry, surface properties of the measurement are written to the soundings, aux and ids_info files, details of which can be found in the Auxiliary User manual (AUM) [AD2]. A schematic illustrating the data flow of the LRPT is shown in Figure 6-2.

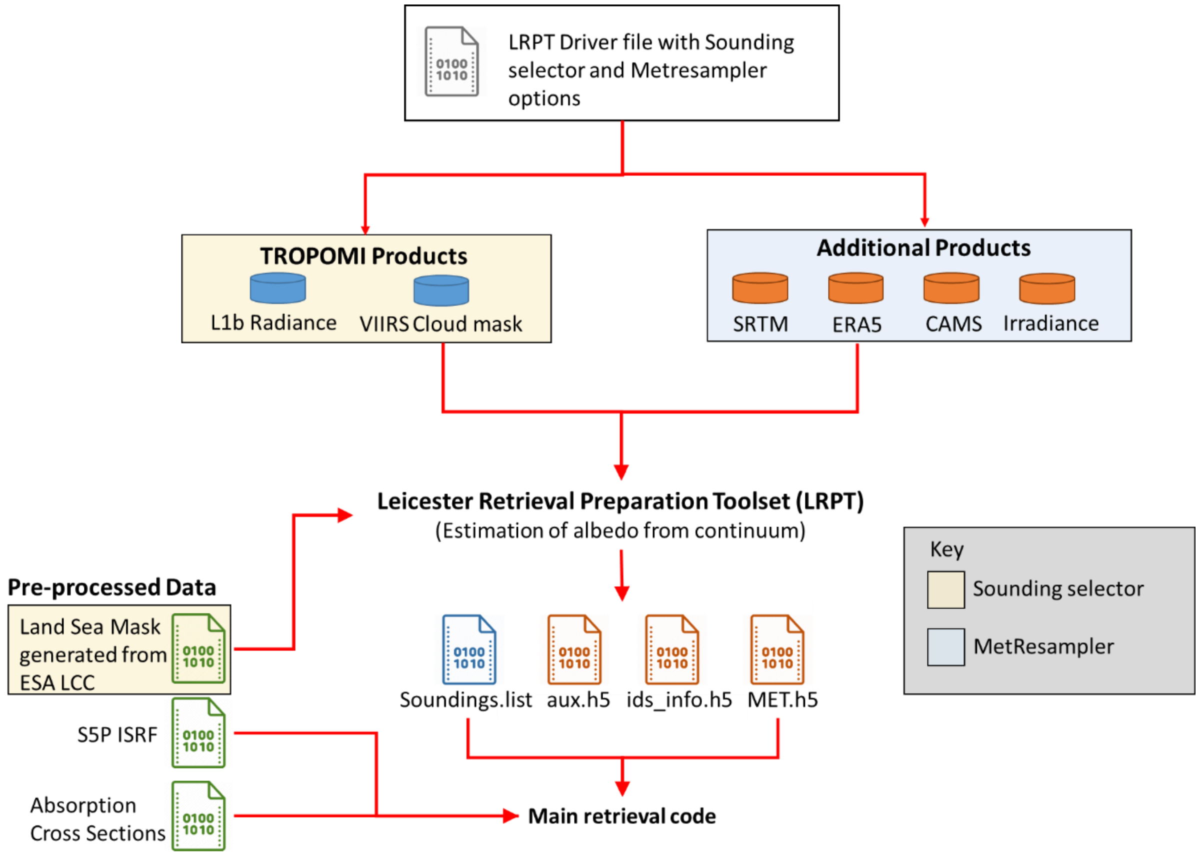

Figure 6-2: Flow of data in and out of the LRPT. Different shaded boxes highlight at which stage data is used, i.e. SoundingSelector or MetResampler. Further details of the different files can be found in the AUM.

6.1.1 Cloud Masking

The L2__NP_BD7 cloud product stores information on the VIIRS cloud mask within each TROPOMI pixel, split into four categories, confident cloudy, probably cloudy, probably clear and confident clear. The product also includes information at four different scales (x1, x1.2, x1.5, x2) of the instantaneous-field-of-view (IFOV). It has been shown that 99% of an area equal 4 times that of a ground pixel should be confidently clear as a first pass filter. Therefore, the x2 scaled cloud information is used for each pixel. If strict cloud filtering (default) is required, then only confident clear VIIRS pixels, otherwise for more relaxed cloud filtering we will also consider probably clear VIIRS pixels also. The cloud fraction for each TROPOMI pixel (cfc) is then calculated from:

$$cfc = \frac{number\ of\ clear\ VIIRS\ pixels}{total\ number\ of\ VIIRS\ pixels} \qquad \text{Equation 1}$$

A prescribed cloud fraction threshold is then used to determine whether the TROPOMI pixel is clear or cloudy. The default threshold flags everything over 1% as cloudy, which can remove up to 97% TROPOMI pixels of a scene [RD18]. An example cloud fraction is shown in Figure 6-3.

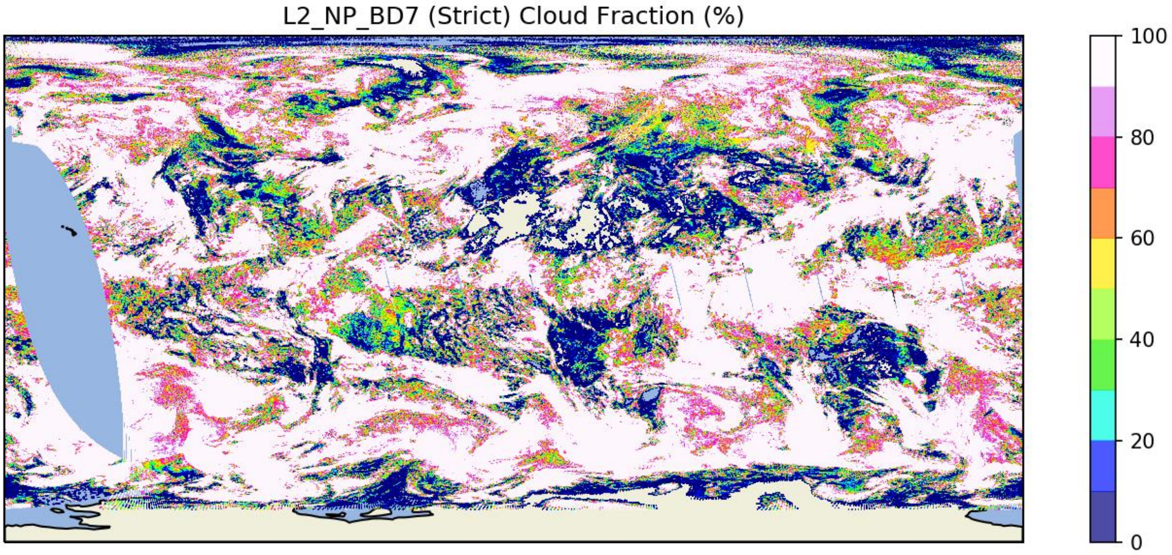

Figure 6-3: Example of cloud fraction calculated from the TROPOMI L2__NP_BD7 VIIRS cloud mask product. Pixels with no cloud cover (i.e. 0%) are transparent.

6.1.2 Ocean Pixel Masking

Retrievals of stable water vapour isotopologues are currently only performed over land. To classify each pixel a land sea mask describing whether a pixel is (permeant) ice free land, or ocean is mapped onto the TROPOMI swath from a pre-computed gridded auxiliary land fraction product. The land fraction is calculated from the ESA CCI Land cover change product on a 0.05 x 0.05 degree grid (Figure 6-4). As for the cloud fraction, a threshold is used to mask water bodies. The current default is to mask pixels with less than 99% land fraction, however, this can be controlled from the LRPT driver file.

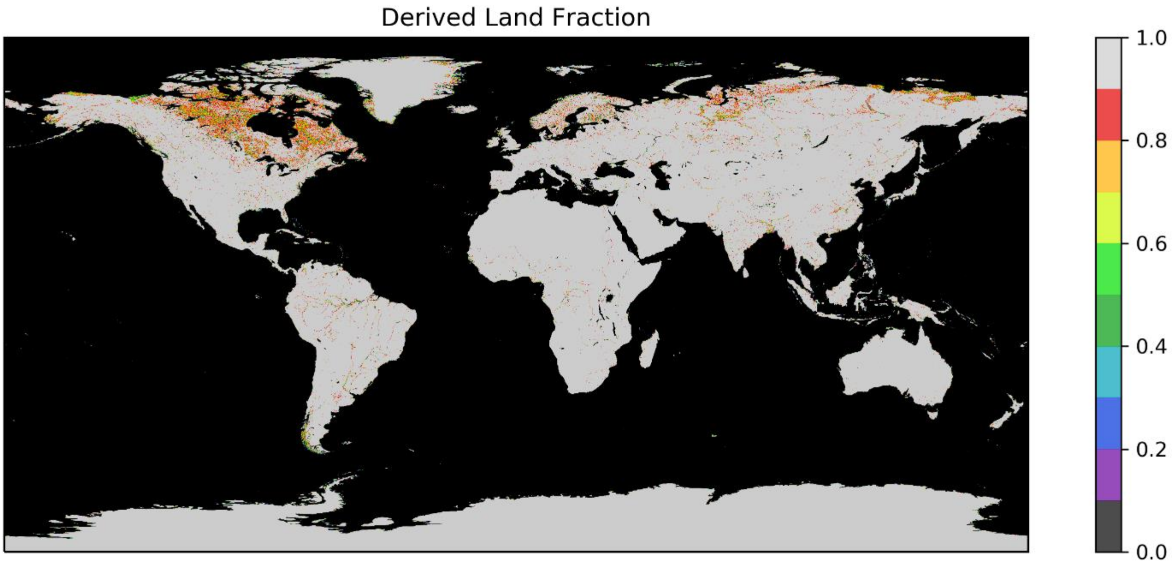

Figure 6-4: Land fraction calculated from the ESA CCI land cover product for 2018.

6.1.3 Addition of state vector a priori data.

The final stage of the LRPT runs the MetResampler which extracts all the additional input data needed for the retrieval. Meteorological and reanalysis regularly gridded fields (surface pressure, water vapour, temperature, methane, carbon monoxide) are mapped to TROPOMI pixel footprints using trilinear interpolation longitude, latitude and time information. First, the differences are found between the spatio-temporal coordinates:

$$lon_d = {(lon_p - lon_1)}\big/{(lon_1 - lon_0)}$$ Equation 2

$$lat_d = {(lat_p - lat_1)}\big/{(lat_1 - lat_0)}$$ Equation 3

$$time_d = {(time_p - time_1)}/{(time_1 - time_0)} \qquad \text{Equation 4}$$

where the subscripts *d* and *p* are the difference and pixel values respectively, subscripts 0 and 1 are the gridded coordinates below and above the pixel values respectively. Next values are interpolated along the longitude axis:

$$V(lon_0, lat_0) = V(lon_0, lat_0, time_0)(1 - lon_d) + V(lon_1, lat_0, time_0)lon_d \qquad \text{Equation 5}$$

$$V(lon_0, lat_1) = V(lon_0, lat_0, time_1)(1 - lon_d) + V(lon_1, lat_0, time_1)lon_d \qquad \text{Equation 6}$$

$$V(lon_1, lat_0) = V(lon_0, lat_1, time_0)(1 - lon_d) + V(lon_1, lat_1, time_0)lon_d \qquad \text{Equation 7}$$

$$V(lon_1, lat_1) = V(lon_0, lat_1, time_1)(1 - lon_d) + V(lon_1, lat_1, time_1)lon_d \qquad \text{Equation 8}$$

Where *V* is the variable being interpolated. Now the variable is interpolated along the latitude axis:

$$V(lat_0) = V(lon_0, lat_0)(1 - lat_d) + V(lon_1, lat_0)lat_d \qquad \text{Equation 9}$$

$$V(lat_1) = V(lon_0, lat_1)(1 - lat_d) + V(lon_1, lat_1)lat_d \qquad \text{Equation 10}$$

Finally, the variable is interpolated along the time axis:

$$V\left(lon_p, lat_p\right) = V(lat_0)(1 - time_d) + V(lat_1)time_d \qquad \text{Equation 11}$$

where $V\left(lon_p, lat_p\right)$ is the variable interpolated to the TROPOMI pixel position. For profile variables (e.g. water vapour, temperature) this is performed at each sigma level. When collected these variables are then stored in the MET output files.

6.2 Full Physics Algorithm

6.2.1 Overview

The UoL-FP retrieval algorithm is based on the algorithm that was initially developed to retrieve XCO2 from a simultaneous fit of the near-infrared O_2 A Band spectrum at 0.76 µm and the CO_2 bands at 1.61 and 2.06 µm as measured by the NASA OCO instrument. While the algorithm was developed to retrieve XCO2 from OCO and OCO-2 observations, it was designed to be adaptable to analyse data from other instruments for algorithm testing and validation. The algorithm has already been successfully used to:

- retrieve XCO2 and XCH4 from SCIAMACHY, OCO-2, TanSat or GOSAT [RD10, RD19, RD11],
- retrieve HDO/H2O, PBL XH2O and SIF from GOSAT [RD13, RD14, RD15], and
- perform retrievals simulations for many missions including EE8 CarbonSat, Sentinel 5 or CO2M.

The fundamental approach of the UoL-FP retrieval algorithm is to use an iterative retrieval scheme based on Bayesian optimal estimation to estimate a set of atmospheric/surface/instrument parameters, referred to as the state vector $\boldsymbol{x}$, from measured, calibrated spectral radiances. UoL-FP employs a forward model to describe the physics of the measurement process and to relate measured radiances to the state vector $\boldsymbol{x}$. It consists of a radiative transfer (RT) model coupled to a model of the solar spectrum to calculate the monochromatic spectrum of light that originates from the sun, passes through the atmosphere, reflects from the Earth's surface or scatters back from the atmosphere, exits at the top of the atmosphere and enters the instrument. The calculated top of atmosphere (TOA) radiances are then passed through the instrument model to simulate the measured radiances at the appropriate spectral resolution.

6.2.2 The Inverse Method

Wrapped around the forward model is an inverse method which extracts information from the measurement. A detailed description of the inverse method can be found in Rodgers [RD20] and Connor et al., [RD21]. An iterative inverse method is employed using the Levenberg-Marquardt modification of the Gauss-Newton method to find the estimate of the state vector $\hat{\boldsymbol{x}}$ with the maximum a posteriori probability, given the measurement y:

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + [(1+\lambda)\boldsymbol{S}_a^{-1} + \boldsymbol{K}_i^T \boldsymbol{S}_\varepsilon^{-1} \boldsymbol{K}_i\,)\;]^{-1}[\boldsymbol{K}_i^T S_\varepsilon^{-1}(\boldsymbol{y} - \boldsymbol{F}(x_i)) - \boldsymbol{S}_a^{-1}(\boldsymbol{x}_i - \boldsymbol{x}_a)] \qquad \text{Equation 12}$$

where $\boldsymbol{x_a}$ is the a priori of the state vector. $\boldsymbol{S_a}$ and $\boldsymbol{S_\varepsilon}$ indicate the covariances of the state vector and the measurement respectively. The weighting function (known as Jacobians) ***K*** gives the linear change of the spectrum per change in state vector $\partial \boldsymbol{y}/\partial \boldsymbol{x}$. The update step of the state vector's *i*th iteration from $\boldsymbol{x}_i$ to $\boldsymbol{x}_{i+1}$ can be adjusted by the L-M factor λ.

Once the iterative retrieval process has converged to a solution the error covariance $(\hat{\boldsymbol{S}})$ and averaging kernel matricies $(\boldsymbol{A})$

$$\hat{\boldsymbol{S}} = (\boldsymbol{K}^T \boldsymbol{S}_\varepsilon^{-1} \boldsymbol{K} + \boldsymbol{S}_a^{-1})^{-1} \qquad \text{Equation 13}$$

$$\boldsymbol{A} = \frac{\delta\hat{\boldsymbol{x}}}{\delta\boldsymbol{x}} = \hat{\boldsymbol{S}}\boldsymbol{K}^T\boldsymbol{S}_\epsilon^{-1}\boldsymbol{K}$$ Equation 14

The target column, e.g. XH2O, is inferred by averaging the retrieved target profile, weighted by the pressure weighting function, $(\boldsymbol{h})$, such that:

$$X_{target} = \boldsymbol{h}^T\hat{\boldsymbol{x}}_{target}$$ Equation 15

The associated column averaging kernel for a level j is then given by

$$\left(\boldsymbol{a}_{target}\right)_j = \frac{\delta X_{target}}{\delta \boldsymbol{u}_i}\frac{1}{\boldsymbol{h}_j} = (\boldsymbol{h}^T\boldsymbol{A})_j\frac{1}{\boldsymbol{h}_j}$$ Equation 16

and the variance of X_{target} by

$$\sigma_{X_{target}} = \boldsymbol{h}^T\hat{\boldsymbol{S}}\boldsymbol{h}$$ Equation 17

The main parameters for the characterization of the X_{target} retrieval that are calculated by the retrieval algorithm are the a posteriori X_{target} retrieval error given by the square root of the variance $\left(\sigma_{X_{target}}\right)$ and the column averaging kernel $\left(\boldsymbol{a}_{target}\right)$.

6.3 Forward Model

6.3.1 Solar Model

The monochromatic TOA spectrum calculated by the RT code is multiplied with a synthetic solar spectrum, which is calculated with an algorithm based on an empirical list of solar line parameters (G. Toon, private communication). The solar line list covers the range from 550 to 15,000 cm-1 and is derived from FTS solar spectra: Atmospheric Trace Molecule Spectroscopy (ATMOS), MkIV balloon spectra for the range 550–5650 cm^{-1}, and Kitt Peak ground-based spectra for 5000–15,000 cm^{-1}. The solar model includes both disk center and disk integrated line lists.

6.3.2 Radiative Transfer

The radiative transfer module (RTM) encapsulates all the physics associated with the modification of the solar radiation during its passage through the atmosphere and reflection by the surface. The UoL-FP algorithm uses LIDORT [RD22, RD23] to solve the radiative transfer equation (RTE). LIDORT is a linearized discrete ordinate radiative transfer (RT) model that generates radiances and Jacobians (derivatives of the radiance with respect to atmospheric and surface properties) simultaneously. The Jacobians are computed by an internal perturbation analysis of the complete discrete ordinate solution to the RTE. To avoid tens of thousands of computationally expensive RT calculations per forward model run, we adopt two options for an approximate approach: the “Low Streams Interpolation” (LSI), which is described fully in [RD24] or the Principal Component Analysis (PCA) method detailed in RD25. Both produces radiance errors typically less than a tenth of a percent. Rather than performing full-accuracy calculations with a large number of angular streams at all monochromatic wavelengths, such calculations are only performed at a few tens of wavelengths. Very fast, low accuracy calculations are performed at all the monochromatic wavelengths; these are combined with the small number of high accuracy calculations to provide an estimate of the Stokes vector at each monochromatic point. Monochromatic RT calculations are made through combination with scattering model(s). These include a fast single-scattering model [RD26] the LIDORT scalar multiple-scattering model [RD22], and a second-order-of-scattering polarization model called 2OS [RD27]. The UoL-FP retrieval setup provides large flexibility in choosing different RT setups or to apply a no-scattering forward model.

6.4 Instrument Model

The instrument model convolves the monochromatic radiance spectrum with the instrument lineshape function (ILS). As described in [RD11], the instrument model can also simulate continuum intensity scaling, zero-level offsets and channeling effects. The instrument model performs these actions and is described below. Additionally, the steps taken to calibrate the measured spectra are also described.

6.4.1 Instrument Noise

Spectrally-resolved estimates of instrument noise are provided for each TROPOMI pixel within the L1B product.

6.4.2 Pixel-Wavelength Mapping

The dispersion gives the pixel-wavelength mapping and consists of fitting a 3rd order polynomial to the TROPOMI nominal wavelength provided in the L1B files:

Equation 18

$$v_i = W + \sum_{k=3}^{n} a_k i$$

Where v for detector wavenumber for pixel i, W is the band start wavenumber and a is the kth degree term in the polynomial fit. While the initial the pixel wavelength mapping is estimated by fitting the polynomial to the TROPOMI L1B data, adjustment is required during the retrieval process (Figure 6-5).

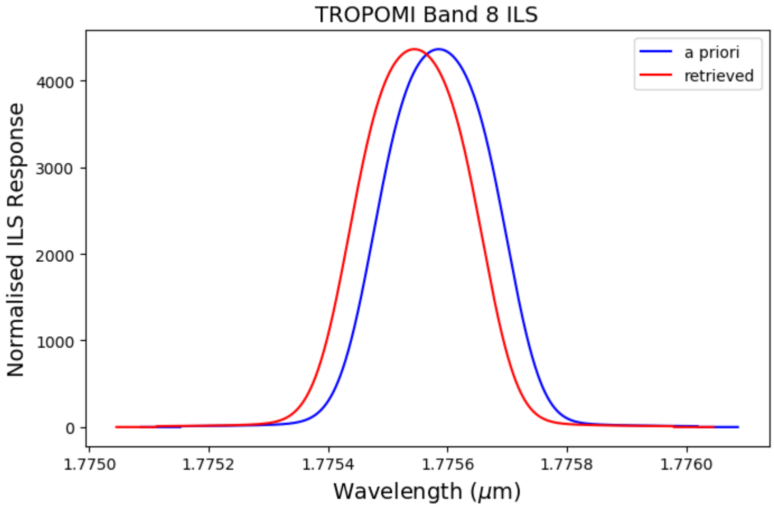

Figure 6-5: Example of ILS shift applied to TROPOMI ILS during UoL-FP optimisation.

6.4.3 Instrument Lineshape Function

To describe the response of the instrument to light with wavelength an instrument line shape function (ILS) is used, which is obtained from http://www.tropomi.eu/data-products/isrf-dataset.

6.4.4 Zero-level Offset

UoL-FP can retrieve an additive zero-level offset parameter which can allow to mitigate calibration issues. This is an optional parameter and is not retrieved in the HDO/H_2O configuration.

6.5 State Vector

6.5.1 Description of Stable Water Isotoplogue Retrievals

The retrievals for HDO and H_2O use a state structure that represents the atmosphere, surface and instrument. The state structure presently consists of 11 or 49 elements depending if H_2O and HDO profiles are retrieved (Table 6-1).

Table 6-1: State vector components for water vapour isotopologue retrievals.

Description	Parameters	Number of Elements
Water Vapour (H2O)	Scaler or profile	1 or 20
Methane	Scaler	1
Carbon Monoxide	Scaler	1
Water Vapour Isotope (HDO)	Scaler or profile	1 or 20
Temperature	Scaler	1
Albedo	Slope and offset	2
Dispersion	Polynomial fit	4
Total		**11 or 49**

6.5.2 Methane Concentration

The a priori profile of CH_4 is acquired from the MACC-II Reanalysis (S1-NOAA-EXT) with the stratosphere replaced by ACE-FTS satellite data assimilated into the TOMCAT model. These atmospheric profiles are then interpolated to the desired latitude, longitude and time.

6.5.3 Carbon Monoxide Concentration

The a priori profile of CO is acquired from the CAMS Reanalysis atmospheric profiles interpolated to the desired latitude, longitude and time.

6.5.4 Surface Pressure

It is important that the surface pressure is well constrained since this can have a large effect on the spectral lines retrieved. The European Centre for Medium-Range Weather Forecasts (ECMWF) is an assimilation model that uses observations from surface buoy and satellite measurements [AD3]. ECMWF ERA5 provide atmospheric profiles of pressure, temperature and specific humidity on a 0.25 degree by 0.25 degree global grid with 137 levels. Given the latitude, longitude and altitude of a site of interest the surface pressure can be determined from these profiles. ECMWF provide potential data for the lowest level of the same grid, which can be used to find the geopotential height of each grid point level as:

$$height = Potential/g \qquad \text{Equation 19}$$

where gravity, g, is calculated as a function of latitude and approximate altitude. Taking the four surrounding grid points of the site of interest, the pressure, P, at the site altitude can be found for each grid point by using the hydrostatic equation:

$$P = P_0 e^{\left(-\frac{z}{z_0}\right)} \qquad \text{Equation 20}$$

where P_0 is the pressure of the grid point level lower than the site altitude, z is the difference in altitude between the grid point level and the site altitude and z_0 is the scale height defined:

$$z_0 = \frac{RT}{Mg} \qquad \text{Equation 21}$$

where R is the ideal gas constant, T is the average temperature across the differential, M is the Molar mass of wet air and g is the gravitational acceleration as a function of latitude and altitude. The Molar mass of wet air can be calculated by:

$$M = \rho_d(1 - SH) + \rho_w SH \qquad \text{Equation 22}$$

where SH is the ECMWF specific humidity, ρ_d is the dry air mass and ρ_w is the mass of wet air. The site altitude can be obtained from a global digital elevation model with a horizontal grid spacing of 3 arc seconds from the Shuttle Radar Topography Mission (SRTM) provided by the U.S. Geological Survey. Note, that in the case that the site altitude is lower than the lowest level of a grid point the pressure is calculated with respect to the lowest level, where the temperature and molar mass are extrapolated downwards based on the lapse rate and gradient of the 5 lowest levels above, respectively. The surface pressure for the site can then be resolved by interpolating the pressures with latitude, longitude and time.

6.5.5 Temperature and Water Vapour

The ECMWF specific humidity data can be used to generate water vapour volume mixing ratio profiles using the equation:

$$H2O_{vmr} = 10^6 \left(\frac{SH}{\left({}^{R_d}/_{R_w}\right) - \left(SH\left(\left({}^{R_d}/_{R_w}\right) - 1\right)\right)} \right) \qquad \text{Equation 23}$$

where R_d and R_w are the gas constants in dry and wet air respectively. ECMWF also provides temperature profiles which, along with the $H2O_{VMR}$, are interpolated with latitude, longitude and time to the specific observation. For water isotopologue profiles we have adapted the approach from [RD28]. For HDO, a δD profile is first constructed with a surface value of -100 ‰ which linearly decreases to -600 ‰ at the tropopause, increasing (linearly) to -400 ‰ at the TOA. For H_2O^{18} the relationship between $\delta^{18}O$ and δD (“global meteoric water line”, Craig 1961:

$$\delta^{18}O = \frac{(\delta D - 10\ \text{‰})}{8} \qquad \text{Equation 24}$$

The isotopologue profile is then calculated thus:

$$H2O_{iso} = \frac{\left({}^{\delta_{iso}}/_{1000}\right) + 1}{H2O_{vmr}} \qquad \text{Equation 25}$$

where $H2O_{iso}$ and δ_{iso} are the isotopologue and delta value (e.g. HDO, δD) respectively. This relationship relies on the assumption the fractionation, i.e. the Vienna Standard Mean Ocean Water (VSMOW) standard, is applied to the absorption cross sections.

6.5.6 Surface Properties

The estimate of the a priori albedo (α) is calculated from the spectral continuum from the TROPOMI L1B using the reflectivity:

$$\alpha = \frac{\pi \bar{L}_{SWIR}}{(\bar{I}_{SWIR} \cos(SZA))} \quad \text{Equation 26}$$

where SZA is the solar zenith angle and $\bar{L}_{SWIR}$ and $\bar{I}_{SWIR}$ are the observed TROPOMI solar irradiance and radiance respectively. The retrieval uses two albedo parameters for each spectral band, giving the albedo for the centre wavelength of the band and the slope of the albedo.

6.5.7 Cross sections

The retrieval algorithm uses pre-calculated absorption cross-sections for H_2O (without HDO), HDO, CH_4 and CO. These are tabulated for 70 pressure levels and 17 temperature values on a spectral grid of 0.01 cm^{-1}. This format of the cross-section tables is based on the ABSCO version 5.0 from NASA JPL and are calculated using the HITRAN Application Programming Interface [RD29].

6.5.8 A Prioiri Error covariance

In the case of scaler retrievals for trace gases, a covariance value equal to a standard deviation of 0.32 is used for H_2O, HDO, CO and CH_4. The covariance for mean albedo is essentially unconstraint and the slope a priori error is set so that the uncertainty at the band edges equals 50%.

7 Feasibility

This section will be completed in later revisions of this document.

8 Error analysis

This section will be completed in later revisions of this document.

9 Validation

This section will be completed in later revisions of this document.

10 Conclusion

This section will be completed in later revisions of this document.