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

Auxiliary User Manual (AUM)

Authors:
Tim Trent and Hartmut Bösch: Department of Physics and Astronomy, University of Leicester, Leicester, United Kingdom and National Centre for Earth Observation NCEO
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	31-Jan-2020	Initial internal draft for project team	T. Trent, H. Bösch M. Schneider, F. Khosrawi, C. Diekmann, , H. Sodemann	New document
Draft 1	4-Feb-2020	Document prepared for submission		Consolidated version
Draft 1.1	24-Feb-2020	Document amended for acceptance by ESA		Updates based on ESA comments
Draft 1.2	28-Feb-2020	Document prepared for submission		Co-author feedback
Draft 1.3	22-Jun-2020	Document updated		Addition of spectroscopic database
Draft 1.4	09-Sep-2021	Document updated		Addition of in situ and other intercomparison data sets
Draft 2	22- June-2022	Document updated		Final updates and release

Table of Content

1 Introduction

The Sentinel era of remote sensing represents a step change in our ability to monitor environmental change within the Earth system. The diversity of instruments being flown is also driving innovative research, leading to new novel state-of-the-art geophysical products. In order to successfully translate the generation of satellite products from research mode to operational production, all elements surrounding the workflow need accurate and concise documentation. Estimation from remote sensing platforms of geophysical parameters using modern algorithms use auxiliary data from numerous sources. With the increasing sophistication of these algorithms, coupled with technological improvements of our measurement systems, this complexity can be mirrored in the auxiliary data used.

1.1 Scope and Objectives

This document is the Auxiliary User Manual (AUM) for the European Space Agency (ESA) Sentinel-5p Innovation (S5p+I) project for stable water isotopologues. The objectives of this document are to describe:

- All individual auxiliary datasets used to run the University of Leicester Full Physics algorithm (UoL-FP). The structure of this document is designed to mirror the flow of data into the processor and includes information where the input datasets are processed into the internal format used by the algorithm. Information is also provided that maps variables within the internal format to their original source.

- Data used/involved in the verification of the retrieved output from the water vapour isotopologue processor.

- Detail secondary sources of data used for the validation, intercomparison and scientific exploitation of the TROPOspheric Monitoring Instrument (TROPOMI) stable water vapour isotopologues. These include in situ measurements from ground sites and campaigns, additional satellite products and water isotope enabled models.

This document is complimented by an online datapool which collects together source information regarding all the data products described in this document. In the case of large datasets (e.g. ECMWF meteorological fields, S5p L1b spectra) or 3rd party datasets (where the project does not have permission to redistribute), links are provided to where the original data can be found.

2 Data for the University of Leicester Full Physics Processor

This section provides the overview of data products used in relation to the workflow of the University of Leicester Full Physics (UoL-FP) algorithm. In total, there are six separate sources of data that are fed in to the UoL-FP workflow at different stages. These six sources are:

- TROPOMI measured shortwave infrared (SWIR) radiances from bands 7 and 8, level 2 (L2) and auxiliary products.
- European Centre for Medium-Range Weather Forecasts (ECMWF) 5th Reanalysis (ERA5) meteorological fields.
- Greenhouse gas (GHG) and GHG flux products from the Copernicus Atmosphere Monitoring Service (CAMS).
- Surface height information from the Shuttle Radar Topography Mission (SRTM).
- Land cover information from ESA Climate Change Initiative (CCI).
- Spectroscopic databases such as the high-resolution transmission molecular absorption database (HITRAN).

Prior to the launch of a retrieval, a pre-processing step is implemented. Within the UoL-FP algorithm, this stage is known as the Leicester Retrieval Preparation Toolset (LRPT) that essentially prepares the majority of the data described above for ingestion into the main retrieval code.

The first step selects which TROPOMI pixels are to be used within the retrievals by the sounding selector. Based on user-defined inputs that specify the date(s) and geographic region to be processed, pixels are identified using the TROPOMI Band 7 L1b file(s). Additional information on the scene is then gathered from the TROPOMI L2 cloud product, L2__NP_BND7. This product maps cloud properties as observed by the Visible Infrared Imaging Radiometer Suite (VIIRS) on-board the Suomi National Polar-orbiting Partnership (Suomi NPP) satellite which proceeds S5p by 3.5 minutes in the Local Time Ascending Node (LTAN). The computed cloud fraction is then used to remove cloudy TROPOMI scenes from entering the main retrieval code. A secondary static product filter is then used to remove ocean scenes, based on the ESA CCI LAND Cover Change (LCC) level (L4) product. This auxiliary product can be reprocessed to add or remove additional surface classification, such as lakes if needed. The sounding selector then produces a list of scenes for processing with a unique exposure identifier which are written to file (sounding.list) along with the full path to the L1b Band 7 file. At a later stage in the retrieval code when the spectra are read-in, bands 7 and 8 are read-in jointly to form a singular SWIR band. The formatting of the L1b file names allows for mapping to the corresponding L1b and L2 files. Additional information about the TROPOMI pixels are also written to the aux and ids_info files for late use by the retrieval code.

The next stage of the LRPT is to run the MetSampler to map all the state vector elements needed for each pixel along with supporting metadata. This data feed draws surface altitude from a static digital elevation mask (DEM) produced from SRTM data, surface temperature and pressure, temperature and water vapour profiles from ERA5, and greenhouse gas profile information (CO & CH4) from CAMS data. All profile data is sampled using the full ECMWF model resolution, before being mapped on the retrieval pressure level grid. All the state vector information is then written to the MET auxiliary file. This workflow is summarised in Figure 2-1.

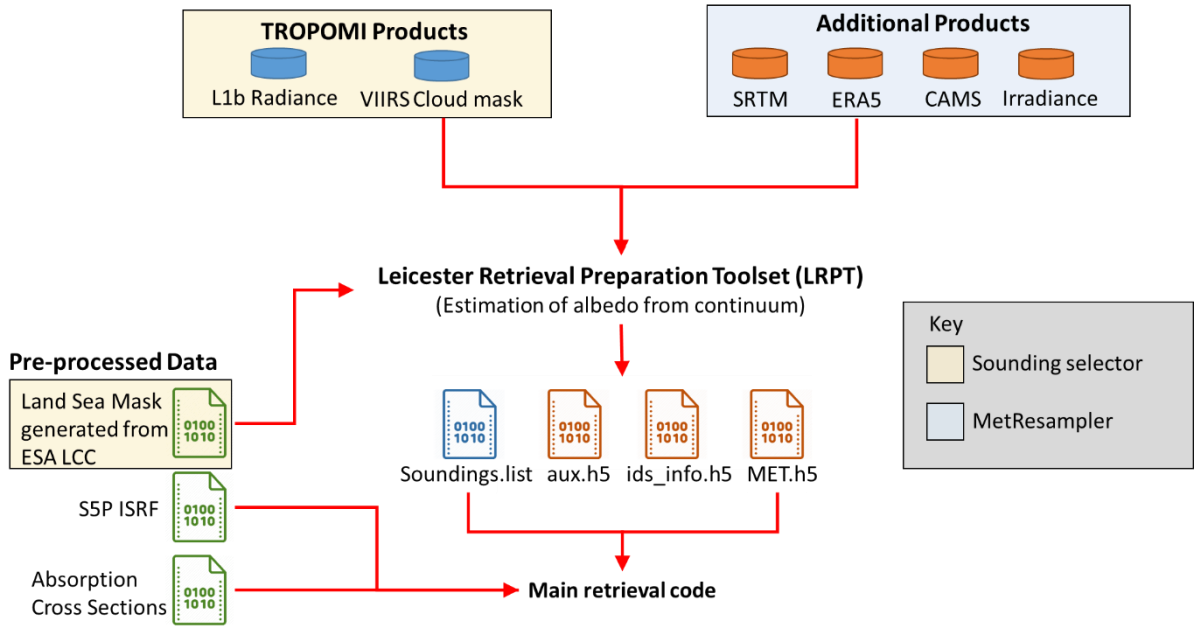

Figure 2-1: Workflow for data inputs used within the UoL-FP algorithm for the retrieval of water vapour isotopologues from the TROPOMI instrument on board Sentinel 5P. Datasets handled by the sounding selector and MetResampler modules of the LRPT are identified with coloured boxes.

To further describe the data flowing into the UoL-FP processor, they are now categorised as one of two types:

- Input Data Bases (IDB): These are data files that are current TROPOMI products and that are already described in various reference documents, or from third parties with similar levels of documentation (e.g. ECMWF).
- Auxiliary (AUX): These are files internally generated by the processing chain and used by the IDB as inputs. Some files are pre-processed from additional data sources and are used by all retrievals (non-scene dependent).

The allocation of these categories to the previously described data products is summarised by Table 2-1.

Table 2-1: Overview of auxiliary data used for the retrieval of water vapour isotopologues from TROPOMI.

ID	Name	Short Description	Format	Size	Static or Dynamic	Spatial Coverage	Generation Frequency	Validity	Source
IDB_1	S5P_OFFL_L1B_RA_<BAND_NUMBER>_<yyyymmddThhmmss>_<YYYYMMDDTHHMMSS>_<ooooo>_<cc>_<pppppp>_<YYYYMMDDTHHMMSS>.nc	S5P L1b SWIR radiances where: **product validity start time**: <yyyymmddThhmmss> **product validity stop time:** <YYYYMMDDTHHMMSS> **absolute orbit number:** <ooooo> **collection number:** <cc> **processor version number**: <pppppp> **production (start) time:** <YYYYMMDDTHHMMSS>	NetCDF	~1.5 Gb	D	1 orbit	~90 minutes	n/a	CEDA Archive
IDB_2	S5P_OFFL_L2__NP_BD7_<yyymmddThhmmss>_<YYYYMMDDTHHMMSS>_<ooooo>_<cc>_<pppppp>_<YYYYMMDDTHHMMSS>.nc	Propagated cloud information relevant for TROPOMI SWIR bands derived from observations made by the VIIRS instrument on NPP. File name format as for IDB_1	NetCDF	~150 Mb	D	1 orbit	~90 minutes	n/a	Sentinel-5P Pre-Operations Hub
IDB_3	srtm_dem.nc	SRTM digital elevation map	NetCDF	3.7 Gb	S	Global	n/a	Until revised	Shuttle Radar Topography Mission
IDB_4	<YYYYMM>_<Var>.nc	ECMWF Meteorological fields of surface and profile variables (Var) stored in monthly files.	NetCDF	1.6-3.4 Gb	D	Global	monthly	1 month	ECMWF ERA interim/ERA5
IDB_5	z_macc_l_jrc_<YYYYMM>_v10-S1NOAA_ra_ml_dm_ch4.nc	CAMS Greenhouse Gases Flux Inversions CH4 data product	NetCDF	~70 Mb	D	Global	n/a	2012-present	ECMWF CAMS GHG Flux inversions

IDB_6	co_<YYYYMM>_cams_6hr.nc	CAMS Near Real Time/Reanalysis CO profiles	NetCDF	1.7 Gb	D	Global	6 hourly	2012-present	ECMWF CAMS
IDB_7	s5p_mean_swir_iradiance.dat	Mean SWIR irradiance calculated from TROPOMI L1B_IR_SIR product	.dat	45 Kb	S	Global	n/a	Until revised	Sentinel-5P Pre-Operations Hub
IDB_8	<GAS_ID><meas_info>.hit	SEOM-IAS Spectroscopic parameters database 2.3 μm	.hit	1.4 – 4.3 Mb	S	Global	n/a	Until revised	SEOM-IAS_databse
AUX_1	<Job_id>.list.gz	List of cloud cleared S5p sounding IDs	.list	~500 Mb	D	n/a	per day	1 day	generated by LRPT
AUX_2	<Job_id>_aux.h5	Auxiliary data used by the UoL-FP processor. This includes a unique sounding id, radiance statistics and albedo information	hdf5	~10 Mb	D	Global	per day	1 day	generated by LRPT
AUX_3	<Job_id>_ids_info.h5	Geospatial information used by UoL-FP processor for S5p pixel being retrieved	hdf5	~250 Mb	D	Global	per day	1 day	generated by LRPT
AUX_4	<Job_id>_MET.h5	All state vector information to be used by UoL-FP processor	hdf5	1.5-3 Gb	D	Global	per day	1 day	generated by LRPT
AUX_5	lsm_cci_lcc_X0.05_deg_Y0.05_deg_2018.h5	Land sea mask derived from ESA land cover L4 product	hdf5	210 Mb	S	Global	n/a	Until revised	ESA Land Cover Project Site
AUX_6	<GAS_ID>_atm_<yyyymmdd>.abs	Pre-calculated absorption cross section file using either HITRAN, SEOM_IAS or JPL databases. Date on file corresponds to the generation date of the ABSCO file	.abs	~960 Mb	S	n/a	n/a	Until revised	HITRAN
AUX_7	s5p_<BAND_ID>_atp_<PIXEL_ID>_isrf.dat	Pre-formatted S5p ISRF information	.dat	~600 Kb	D	n/a	n/a	Until revised	S5P_OPER_AUX_ISRF___v3.0.0.zip

2.1 S5p L1b SWIR (Spectra Input Data Base 1)

TROPOMI SWIR spectra for bands 7 and 8 are used for the retrieval of water vapour isotopologues within this project. These are the official offline L1 files released by ESA and are openly available from Sentinel-5P Pre-Operations Data Hub, or via the Centre for Environmental Data Analysis (CEDA) with a user account. Further details regarding the file contents of the TROPOMI L1b radiance files can be found in Loots et al. 2017 and Vonk et al. 2018.

2.2 Cloud Information for S5p IFOV (Input Data Base 2)

Though the SWIR bands are split in two, bands 7 and 8, they both still have the same ground footprint. Therefore this project is able to utilise the official TROPOMI L2 VIIRS cloud product (L2__NP_BND7) to mask the both SWIR bands. Like the L1b spectra, the cloud product can be freely obtained via the Sentinel-5P Pre-Operations Data Hub. A subset of this archive is currently being uploaded on CEDA for this project and can be accessed in the same manner as the L1b files. Further details regarding the file contents of the VIIRS cloud mask product for TROPOMI can be found in Siddens 2016.

2.3 Surface Elevation Input (Input Data Base 3)

Information on the surface topography comes from the SRTM database. The DEM data has been processed and converted into netCDF format for use within the LRPT.

Table 2-2: Contents of digital elevation file, srtm_dem.nc.

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
lon	Longitude	https://dds.cr.usgs.gov/srtm/	Degrees East
lat	Latitude	https://dds.cr.usgs.gov/srtm/	Degrees North
data	Elevation	https://dds.cr.usgs.gov/srtm/	Metres

2.4 Meteorological a priori Information (Input Data Base 4)

ECMWF ERA5 provides the meteorological inputs to the retrieval processor. These are described in Hersbach and Dee, 2016a and 2016b. Meteorological fields are interpolated to the sounding date/time. Details of used variables are given in Section 2.10.

2.5 Source of CH_4 a priori Information (Input Data Base 5)

The methane data input uses a static field from CAMS (formally MACC II) (Marécal et al. 2015) which is propagated through time by applying a growth rate within the LRPT. The CH_4 profiles are interpolated to the sounding date/time. Details of used variables are given in Section 2.10.

2.6 Source of CO a priori Information (Input Data Base 6)

The carbon monoxide data input uses reanalysis/near-real-time (NRT) profiles from CAMS (formally MACC II) (Marécal et al. 2015). The CO profiles are interpolated to the sounding date/time by the LRPT. Details of used variables are given in Section 2.10

2.7 TROPOMI Mean Irradiance Spectra (Input Data Base 7)

The final input data source to the retrieval preprocessor is a precomputed irradiance spectrum for TROPOMI's SWIR bands 7 and 8. Measurements made by TROPOMI, taken from the L1B_IR_SIR product between 1st of July 2018 and 7th of February 2020, have been used to calculate the mean irradiance spectra. All inputs are quality flagged and filtered for bad values before being used in the final calculation (Figure 2-2). This spectra is used in the calculation of the albedo a priori (Auxiliary 2).

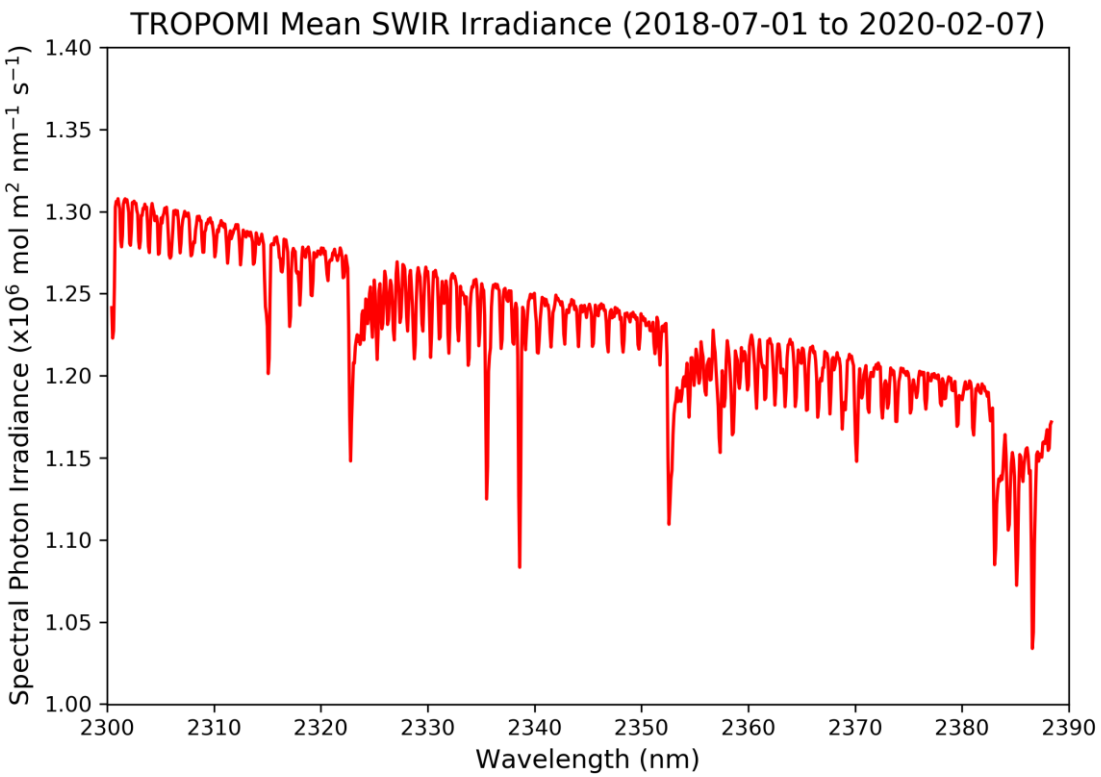

Figure 2-2: Mean SWIR irradiance spectrum calculated from TROPOMI measurements.

2.8 SEOM-IAS (Input Data Base 8)

The Scientific Exploitation of Operational Missions Improved Atmospheric Spectroscopy (SEOM-IAS) databaseS contain molecular absorption line parameters. This improved line parameter database of H2O, CH4 and CO absorption lines were produced in accordance to the user needs of teams working with the SWIR bands of the TROPOMI instrument (Birk et al. 2017).

Variable Names/Description	Source/Reference/Citation	Variable Units
Line position uncertainty	IBD_8	cm^{-1}
Line intensity uncertainty		cm^{-1}
Lower state energy uncertainty		1
Air broadening uncertainty		$cm^{-1}\,atm^{-1}$
Pressure-induced line shift uncertainty		$cm^{-1}\,atm^{-1}$
Air-broadening temperature exponent uncertainty		1
Speed-dependence of air broadening		$cm^{-1}\,atm^{-1}$
Speed-dependence of air broadening uncertainty		$cm^{-1}\,atm^{-1}$
Speed-dependence of pressure-induced line shift		$cm^{-1}\,atm^{-1}$
Speed-dependence of pressure-induced line shift uncertainty		$cm^{-1}\,atm^{-1}$
Frequency of velocity-changing collisions (Dicke-effect		$cm^{-1}\,atm^{-1}$
Rosenkranz line mixing		atm^{-1}
Rosenkranz line mixing uncertainty		atm^{-1}
Smith line mixing, quadratic pressure dependence of intensity		atm^{-2}
Smith line mixing, quadratic pressure dependence of intensity uncertainty		atm^{-2}
Smith line mixing, pressure dependence of pressure-induced shift		$cm^{-1}\,atm^{-2}$
Smith line mixing, pressure dependence of pressure-induced shift uncertainty		$cm^{-1}\,atm^{-2}$
Temperature dependence of pressure-induced line shift		$cm^{-1}\,atm^{-1}\,K^{-1}$

Temperature dependence of pressure-induced line shift uncertainty		$cm^{-1}\,atm^{-1}\,K^{-1}$
Temperature dependence of Rosenkranz line mixing	IDB_8	K^{-1}
Temperature dependence of Rosenkranz line mixing uncertainty		K^{-1}
Self broadening uncertainty		$cm^{-1}\,atm^{-1}$
Self broadening temperature exponent		1
Self broadening temperature exponent uncertainty		1
Self pressure-induced line shift (H2O only)		$cm^{-1}\,atm^{-1}$
Self pressure-induced line shift uncertainty (H2O only)		$cm^{-1}\,atm^{-1}$
Speed-dependence of self broadening (H2O only)		$cm^{-1}\,atm^{-1}$
Speed-dependence of self broadening uncertainty (H2O only)		$cm^{-1}\,atm^{-1}$
Temperature dependence of self pressure-induced line shift (H2O only)		$cm^{-1}\,atm^{-1}\,K^{-1}$
Temperature dependence of self pressure-induced line shift uncertainty (H2O only)		$cm^{-1}\,atm^{-1}\,K^{-1}$

2.9 Sounding List (Auxiliary 1)

The first auxiliary file contains a list of pixel identifiers that are fed to the retrieval code.

Table 2-3: Contents of sounding list file.

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
Exposure_ID	Unique identifier for specific S5p FOV/pixel	IDB_1	-
L1b file name	Filename including full path to corresponding S5p L1b file	IDB_1	-

2.10 UoL-FP Auxiliary Data (Auxiliary 2)

Auxilary data used by the UoL-FP retrieval processor. It should be noted that while the variables are only present for band 7, when the radiances are read in bands 7 and 8 are combined to form a singler band.

Table 2-4: Summary of the contents the auxiliary data used by UoL-FP.

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
Exposure_ID	Unique identifier for specific S5p FOV/pixel	IDB_1	-
albedo_B7	Estimated albedo (α)	$\alpha = \frac{\pi \overline{L}_{SWIR}}{\left(\overline{I}_{SWIR} cos(SZA)\right)}$ Where $\overline{L}_{SWIR}$ is the mean SWIR band radiance, $\overline{I}_{SWIR}$ is the mean SWIR and irradiance (IDB_7) and SZA is the solar zenith angle.	1
maxrad_B7	Maximum band radiance	IDB_1	mol.s-1.m-2.nm-1.sr-1
meanrad_B7	Mean band Radiance	IDB_1	mol.s-1.m-2.nm-1.sr-1
snr_B7	Mean band SNR calculated from pixel radiance noise (L1b variable radiance_noise)	IDB_1	1

2.11 Geospatial/Temporal Data (Auxiliary 3)

The third auxiliary file type contain temporal and geospatial information on the sounding/exposure indentifies in Auxiliary 1.

Table 2-5: Contents of file containing geospatial information used by UoL-FP.

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
Day	Day of sounding	IDB_1	DD
Exposure_ID	Unique identifier for specific S5p FOV/pixel	IDB_1	-

FP_vertex_lat_1	Lower left pixel corner latitude	IDB_1	degrees
FP_vertex_lat_2	Lower right pixel corner latitude	IDB_1	degrees
FP_vertex_lat_3	Upper right pixel corner latitude	IDB_1	degrees
FP_vertex_lat_4	Upper left pixel corner latitude	IDB_1	degrees
FP_vertex_lon_1	Lower left pixel corner longitude	IDB_1	degrees
FP_vertex_lon_2	Lower right pixel corner longitude	IDB_1	degrees
FP_vertex_lon_3	Upper right pixel corner longitude	IDB_1	degrees
FP_vertex_lon_4	Upper left pixel corner longitude	IDB_1	degrees
GroundPixelQuality	Quality assessment information for each ground pixel. Flag meanings are: no_error, solar_eclipse, sun_glint_possible, descending, night, geo_boundary_crossing, geolocation_error	IDB_1	-
Hour	Hour in day	IDB_1	hh
L1B_path	Path to L1b file	IDB_1	-
Lat	Pixel centre latitude	IDB_1	degrees
Lon	Pixel centre longitude	IDB_1	degrees
MeasurementQuality	Overall quality information for a measurement. Flag meanings are: no_error, proc_skipped, no_residual,	IDB_1	-

	saa, spacecraft_manoeuvre, sub_grp, irr_out_of_range, sub_group		
Minute	Minute of day	IDB_1	mm
Month	Month of sounding	IDB_1	MM
SAA	Solar azimuth angle	IDB_1	degrees
SZA	Solar zenith angle	IDB_1	degrees
Second	Second in day	IDB_1	ss.sss
Year	Year of sounding	IDB_1	YYYY
frac_days_since	Fractional days since beginning of YYYYMMDD	IDB_1	-
satAA	Satellite azimuth angle	IDB_1	degrees
satZA	Satellite zenith angle	IDB_1	degrees

2.12 State Vector Elements (Auxiliary 4)

The <Job_id>_MET.h5 auxiliary file contains all the state vector elements for the UoL-FP retrieval processor. Profiles are interpolated from the ERA5/CAMS gridded space to TROPOMI pixel time and location on sigma levels. Gas concentarions are then converted from kg/kg to ppm before being mapped on the retrieval grid. For water isotopologues there is an additional flag that will alter which a priori profile is used based on whether a ratio or profile retrieval is being performed. The approach here ahas been taken and adapted from Scheepmaker et al. (2016). 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):

$$\delta D = 8 \cdot \delta^{18}O + 10‰ \quad \text{Equation 1}$$

The H_2O profile is then used to convert δD and $\delta^{18}O$ to the HDO and H_2O^{18} profiles respectively.

Table 2-6: Contents of the <Job_id>_MET.h5 file used to populate the state vector components for UoL-FP.

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
Exposure_id	Unique identifier for specific S5p FOV/pixel	IDB_1	-
Lon	Pixel centre longitude	IDB_1	degrees
Lat	Pixel centre latitude	IDB_1	degrees
T700	700 hPa temperature	IDB_4	K
a_CH4	Methane a prior profiles	IDB_5	ppm
a_CO	Carbon Monoxide a priori profiles	IDB_6	ppm
a_H2O	Water vapour a priori profiles used for H_2O	IDB_4	ppm
a_HDO	Water vapour a priori profiles used for HDO	IDB_4	ppm
a_H2O18	Water vapour a priori profiles used for H2O18	IDB_4	ppm
a_P	Pressure a priori profiles	IDB_4	Pa
a_alt_levels	Altitude a priori profiles	IDB_4	m
a_g_levels	Acceleration due to gravity a priori profiles	IDB_4	ms^2
alt	Surface altitude from SRTM	IDB_3	m
psurf	Surface pressure within IFOV from ECMWF	IDB_4	Pa
ptropo	Cold point tropopause height (Reichler et al. 2003)	IDB_4	Pa

2.13 Land Sea Mask (Auxiliary 5)

The Land sea mask generated for S5p retrievals is based on the ESA land cover (LCC) CCI L4 product. Within the LCC files, each cell is either assigned to a different land type or a value that indicates whether it is a water body or snow/ice. Because no differentiation is given between ocean, lakes or rivers we calculate land fraction within the grid pixel. A script is used to generate an augmented product that is used to mask TROPOMI pixels post retrieval if necessary (Figure 2-3). In addition, a similar variable id calculated for snow and ice fraction. Currently, this is not used within the preprocessing. This auxillary product is produced at a userspecified gridded resolution, with a default resolution is 0.05x0.05.

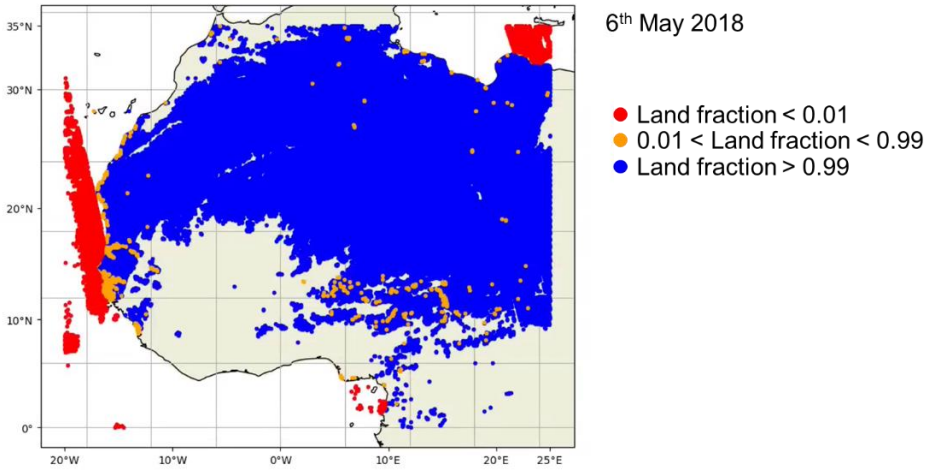

Figure 2-3: Cloud cleared TROPOMI pixels within the West African study region for 06/05/2018. Different colours have been assigned to each pixel beased on their land fraction. Default threshold values have been used, though these can be applied by the user post retrieval. The philosophy here is to not cast out any scenes that might be of interest later. This will also allow for different thresholds to be tested when assessing the quality of the product. Ocean pixels have also been filtered for sun glint (GroundPixelQuality, IBD_1).

Table 2-7: Summary of the contents of the file used for land/sea masking.

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
lons	Longitude	ftp://anon-ftp.ceda.ac.uk/neodc/esacci/land_cover/ (Defourney et al. 2015)	Degrees east
lats	Latitude		Degrees north
landFrac	Fractional land mass within grid cell		1
snowIceFrac	Fraction of snow and ice covered surfaces within grid cell		1

2.14 Absorption Cross Sections (Auxiliary 6)

ACOS/OCO-2 absorption coefficient (ABSCO) binary formatted data files used by forward model (Payne 2017), produced using the HITRAN Application Programming Interface (HAPI, Kochanov et al., 2016).

Table 2-8: Contents of the absorption cross section used for the forward modelling of TROPOMI top-of-atmosphere (TOA) radiances.

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
Gas_[Qabs]_Absorption	The absorption coefficients stored as a 4D table.	Payne 2017, Kochanov et al., 2016	-
Gas_Index	A string containing the 2 digit HITRAN index of the principal absorbing gas, equivalent to Qabs above.	Payne 2017, Kochanov et al., 2016	-
Pressure	The pressure is a dataset of size Npres representing the pressure in Pascals at each atmospheric level in the table.	Payne 2017, Kochanov et al., 2016	Pa
Temperature	Temperature is a 2D dataset of degrees Kelvin, of shape Npres x Ntemps It records the temperature grid point values, which might differ depending on pressure level.	Payne 2017, Kochanov et al., 2016	K
Wavenumber	The Wavenumber object is a dataset of size Nfreqs describing the frequency grid spacing.	Payne 2017, Kochanov et al., 2016	cm^{-1}

2.15 S5p Instrument Line Shape Functions (Auxiliary 7)

These files are internally formatted S5p instrument line shape (ILS) parameters based on the officially released ISRF-dataset product.

Table 2-9: Input variables within the native UoL-FP ILS files.

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
ILS_PIXELS	S5p spectral band index number	http://www.tropomi.eu/data-products/isrf-dataset	-
ILS_DELTA_LAMBDA_1	ILS offset from central pixel		nm
ILS_RESPONSE_1	ILS response		n/a

3 In Situ Data for Validation and Intercomparisons

This section describes the dataset used for the validation of and intercomparison with the TROPOMI XHDO/XH2O product.

3.1 MUSICA NDAAC

MUSICA (Multi-platform remote sensing of Isotopologues for investigating the Cycle of Atmospheric water) NDACC data is measured in the MIR (750 – 4200 cm-1) spectral region with high resolution (0.005 cm-1). Data from the groundbased-sites at Kiruna, Sweden and Karlsruhe, Germany are used for intercomparisons. MUSICA NDACC data can be downloaded at https://www-air.larc.nasa.gov/missions/ndacc/data.html?MUSICA=musica-list as HDF-file, exemplary:

groundbased_ftir.iso.h2o_kit001_kiruna_20190720t074920z_20200706t165819z_001.hdf

The MUSICA NDACC data is described in more detail in Schneider et al. 2017.

Table 3-1: Contents of MUSICA NDAAC files

Variable Names	Variable Description	Source/Reference /Citation	Variable Units
DATETIME	datetime (UT), defined relative to reference datetime of Jan. 1, 2000 at 0:00:00 UT which is equal to 0.00	Schneider et al. 2017	MJD2K
LATITUDE.INSTRUMENT	latitude north (decimal degrees) of the location of the instrument		deg
LONGITUDE.INSTRUMENT	longitude east (decimal degrees) of the location of the instrument		deg
ALTITUDE.INSTRUMENT	altitude of the location of the instrument		km
SURFACE.PRESSURE_INDEPENDENT	estimated surface pressure		hPa
SURFACE.TEMPERATURE_INDEPENDENT	estimated surface temperature		K

ALTITUDE	The retrieved target vmr profile as well as pressure and temperature profiles are reported with respect to the actual grid.		km
ALTITUDE.BOUNDARIES	Upper and lower boundaries of the layers for which partial columns are reported.		km
PRESSURE_INDEPENDENT	Total effective air pressures for vertical profile retrieval		hPa
TEMPERATURE_INDEPENDENT	Pressure/temperature profiles are taken from NCEP; between altitudes, linear interpolation along altitude axis is assumed.		K
COLUMN.PARTIAL_INDEPENDENT	Vertical profile of partial columns of air number densities (for conversion between VMR and partial column profile)		Zmolec2 cm-4
H2O.ISO.MIXING.RATIO.VOLUME_ABSORPTION.SOLAR	retrieved vertical profile of H2O.ISO ({H216O}, {H218O}, {HD16O}) from solar absorption measurements	values given in ppmv; retrieval algorithm PROFFIT96; HITRAN 12 spectroscopy + updates similar to Schneider et al., JQSRT 112, 465-474, 2011.	ppmv
H2O.ISO.COLUMN.PARTIAL_ABSORPTION.SOLAR	retrieved H2O.ISO ({H216O}, {H218O}, {HD16O}) (vertical) partial columns enclosed by reported altitude boundaries		Zmolec cm-2

H2O.ISO.COLUMN_ABSORPTION.SOLAR	Total column of of H2O.ISO ({H216O}, {H218O}, {HD16O})		Zmolec cm-2
H2O.ISO.COLUMN_ABSORPTION.SOLAR_UNCERTAINTY.RANDOM.STANDARD	random error covariance on the retrieved total vertical columns of H2O.ISO({H216O}, {H218O}, {HD16O})		Zmolec cm-2
H2O.ISO.COLUMN_ABSORPTION.SOLAR_UNCERTAINTY.SYSTEMATIC.STANDARD	systematic error covariance on the retrieved total vertical columns of H2O.ISO({H216O}, {H218O}, {HD16O})		Zmolec cm-2
ANGLE.SOLAR_ZENITH.ASTRONOMICAL	astronomical solar zenith angle		deg
ANGLE.SOLAR_AZIMUTH	using north as the reference plane and increasing clockwise (0 for north, increasing towards East)		deg
H2O.MIXING.RATIO.VOLUME_ABSORPTION.SOLAR	assumed vertical profile of H216O, in VMR units (optimally estimated H216O (type 1), see section 4.1, Schneider et al., AMT, 5, 3007-3027, 2012)		ppmv
H2O.COLUMN_ABSORPTION.SOLAR	assumed total vertical column of H216O (optimally estimated H216O (type 1), see section 4.1, Schneider et al., AMT, 5, 3007-3027, 2012)		ppmv

3.2 TCCON

The Total Carbon Column Observing Network (TCCON; [RD8]) is a network of ground-based Fourier Transform Spectrometers that record spectra in the NIR region from 3900 to 15500 cm−1. Data from the groundbased-sites at Burgos, Philippines; Darwin, Australia; Karlsruhe, Germany; Sodankylä, Finnish Lapland and Wollongong, Australia are used for

intercomparisons. TCCON data can be downloaded at https://tccondata.org/ as netcdf-file, exemplary:

ka20100419_20201031.public.nc

The TCCON data is described in more detail in Wunch et al. (2011).

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
zobs_km	Geometric Altitude (km)	https://tccon-wiki.caltech.edu Wunch, D., G. C. Toon, J.-F. L. Blavier, R. A. Washenfelder, J. Notholt, B. J. Connor, D. W. T. Griffith, V. Sherlock, and P. O. Wennberg (2011), The total carbon column observing network, Philosophical Transactions of the Royal Society - Series A: Mathematical, Physical and Engineering Sciences, 369(1943), 2087-2112, doi:10.1098/rsta.2010.0240. Available from: http://dx.doi.org/10.1098/rsta.2010.02	km
zmin_km	Pressure Altitude(km)		km
year	Year (e.g. (2009))		YYYY
xhdo_ppm_error	one-sigma precision		ppm
xhdo_ppm	0.2095*column_hdo/column_o2		ppm
xh2o_ppm_error	one-sigma precision		ppm
xh2o_ppm	0.2095*column_h2o/column_o2		ppm
time	Fractional days since 1970/1/1 00:00:00		days
pout_hPa	External Surface Pressure (hPa		hpa
long_deg	Longitude (deg.)		deg
lat_deg	Latitude (deg.)		deg
hour	Fractional UT Hour		–
day	Day of the year (1-366)		–
azim_deg	Solar Azimuth Angle (deg)		deg
asza_deg	Solar Zenith Angle (deg)		asza_deg

3.3 ULR Profiles

In situ measurements of δD and other meteorological variables during the L-WAIVE (Lacustrine-Water vApor Isotope inVentory Experiment) field campaign using a cavity ring-down laser spectrometer on an ultra-light aircraft are used for the impact assessment. This data is described in detail in Chazette et al, (2021) and is used here in 10 s temporal resolution.

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
delta_D	δD in water vapour	Chazette et al. 2021	‰
q	Specific humidity		g kg^{-1}
p	Ambient pressure		hPa
TA	Ambient temperature		°C
UU	Relative humidity		%
LAT	Latitude along flight path		°N
LON	Longitude along flight path		°E
ALT	Altitude along flight path		m agl

4 Data for Satellite Intercomparisons

This section to be completed in the second release

4.1 MUSICA IASI

The MUSICA IASI {H2O, δD} pair dataset used throughout this project is documented in Diekmann et al. (2021a) and archived in Diekmann et al. (2021b). The individual MUSICA IASI files are named in the form of:

IASI<instrument>_MUSICA_<version>_L2pp_H2Oiso_vs_<YYYYMMDD>_<overpass>_global.nc

with <instrument> indicating the included IASI sensors (A, B and C), <version> as the MUSICA IASI retrieval version, <YYYYMMDD> as the date of the included data and <overpass> as the overpass type (morning or evening).
Table 4-1 provides an overview of the geospatial metadata and the retrieved profiles relevant for this project. Table 4-2 lists flag variables that further characterize the MUSICA IASI {H2O, δD} pair data. This includes quality flags indicating observations with high quality. Table 4-3 shares the variables related to the averaging kernel matrices of the variables relevant for this project. XXX kernels? Uncertainties? {H2O, δD} pair data. The kernels are given for the water vapour proxy state vector [0.5*(ln(H2O) + ln(HDO); ln(HDO) – ln(H2O)], which represents a reliable proxy for variations in H2O and δD and is used during the MUSICA IASI retrieval. Further, the averaging kernel matrices are decomposed using a singular value decomposition and the respective components are then given in the MUSICA IASI netCDF4 files. The kernel matrix A can be reconstructed as follows:

	$A = UDV^T$	Equation 2

with U being the left vectors, V the right vectors and D the diagonal singular value matrix (see Table 4-1). Detailed information about the kernel decomposition and reconstruction are given in Schneider et al. (2021).

Table 4-1: MUSICA IASI variables that are relevant in the context of this project

Variable Names	Variable Description	Source/Reference/ Citation	Variable Units
observation_id	Observation identifier, unique within each file	Schneider et al., 2021 Diekmann et al., 2021a	-
lat	latitude		degree
lon	longitude		degree
time	UTC time		seconds since 2000-01-01 00:00:00

musica_nol	number of atmospheric grid levels		-
musica_altitude_levels	atmospheric altitude levels		m
musica_h2o	retrieved H2O state		ppmv
musica_h2o_apriori	a priori H2O profile		ppmv
musica_h2o_error	total H2O error (including noise and temperature errors)		ppmv
musica_deltad	retrieved δD state		‰
musica_deltad_apriori	a priori δD profile		‰
musica_deltad_error	total δD error		‰
musica_at	retrieved atmospheric temperature		K
musica_st	retrieved surface skin temperature		K
musica_pressure_levels	atmospheric pressure levels		Pa

4.2 GOSAT

Monthly level 2 GOSAT HDO/H2O files are made available to this project, with the algorithm described in Boesch et al., (2013). Unlike the IASI MUSICA data, these are scaler retrievals which operate on the whole profile at the same time, and as such do not produce additional outputs such as averaging kernels. Monthly files are provided in h5 format, split between land and sun glint observations. The relevant content detail are given in Table 4-2.

Table 4-2: Variable details for the contents of L2 GOSAT data files.

Variable Names	Variable Description	Source/Reference/ Citation	Variable Units
Lon	longitude	Boesch et al. 2013	degrees east
Lat	latitude		degrees north
GlintFlag	Flag indicating whether		1

	retrieval is done over sunglint ocean surfaces		
Year	Year of observation		years
Month	Month of observation		months
Day	Day of observation		days
Hour	Hour of observation		hours
Minute	Minute of observation		minutes
Second	Second of observation		seconds
SAA	Solar azimuth angle		degree
SZA	Solar zenith angle		degree
satAA	Satellite azimuth angle	Boesch et al. 2013	degree
satZA	Satellite zenith angle		degree
a_H2O	a prioiri water vapour profile		ppmv
a_HDO	a prioiri water vapour isotopologue profile (HDO)		ppmv
a_P	a prioiri pressure profile		Pa
a_T	a prioiri atmospheric temperature vapour profile		K
r_svsv_h2o_scale	Retrieved H2O profile scaler		1
r_svsv_hdo_scale	Retrieved HDO profile scaler		1
r_sver_h2o_scale	Retrieved H2O profile scaler error		1
r_sver_hdo_scale	Retrieved HDO profile scaler error		1
a_T	a prioiri atmospheric temperature profile		K

r_chi2_2	Retrieval χ^2 value	Boesch et al. 2013	1
r_rel_rms_2	Retrieval relative root mean square fit		1
r_num_div	Number of divergent steps in retrieval		1
r_num_iterations	Number of iterations for retrieval convergence		1
r_outcome	Retrieval outcome flag		1

5 Data for Model Intercomparisons

The datasets used for the model intercomparisons are introduced in this section.

5.1 COSMO$_{iso}$

Simulations with the isotope enabled COSMO model (COSMO$_{iso}$, Pfahl et al., 2012) are used for the model intercomparison. The two 7 day forecasts with a spatial resolution of 0.1°, explicit convection and ECHAM6-wiso boundary and initial data are initiated at 0 UTC 12 June 2019 and 0 UTC 15 June 2019, respectively. 1-hourly outputs of these simulations are used for the intercomparison.

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
QV	specific water vapour content	Available upon request: iris.thurnherr@uib.no	kg kg^{-1}
QV2H	specific water vapour content of HDO		kg kg^{-1}
P	pressure		Pa
PS	surface pressure		Pa
T	Temperature		K
HPBL	Height of boundary layer		m
HSURF	surface height		m
U	U-component of wind (grid eastward wind)		m s^{-1}
V	V-component of wind (grid northward wind)		m s^{-1}
W	vertical wind velocity		m s^{-1}
QR	specific rain content		kg kg^{-1}
QS	specific snow content		kg kg^{-1}
QI	Specific cloud ice		kg kg^{-1}

	water content		
QC	specific cloud liquid water content		kg kg^{-1}
TOT_PREC	total precipitation amount		kg m-2
RELHUM	Relative humidity		%
T_G	grid mean surface temperature		K
QVTAG	tagged (evaporated in the simulation domain) specific water vapour content of H2O		kg kg^{-1}
CLCT	total cloud cover		-
AEVAP_S	surface evaporation		kg m^{-2}
FR_LAND	land-sea fraction		-
lat	latitude		°N
lon	longitude		°E
rlat	rotated latitude		°
rlon	rotated longitude		°
rotated_pole	coordinates of the rotated North Pole		°N/°E

5.2 ICON-ART-iso

A full description of the model structure of ICON-ART-iso can be found in Eckstein et al. (2018) and a technical user guide is found in Eckstein (2017).
The here used ICON-ART-iso files are named in the form of:

icon-dev_motiv_r2b05_init20160620__210224_094310-ISO_DOM<domid>_ML_<outputstep>.nc

with <domid> indicating whether this file refers to the global parent domain (01) or the regional refined nested domain (02), here over West Africa. Table 5-1 provides an overview of the variables relevant for this project.

Table 5-1: ICON-ART-iso variables that are relevant in the context of this project

Variable Names	Variable Description	Source/Reference/ Citation	Variable Units
ncells	number of included grid points	Eckstein 2017 Eckstein et al. 2018	-
time	UTC time		YYYYMMDD.frac(hh)
clat	latitude		radian
clon	longitude		radian
z_mc	geometric height		m
qv	specific humidity of H_2O		kg kg-1
temp	air temperature		K
pres	air pressure		Pa
TRqvHDO	specific humidity of HDO		kg kg-1
TRqvH2Olnd	H_2O from land surface evaporation		kg kg-1
TRqvH2Oocn	H_2O from ocean evaporation		kg kg-1

6 Acronyms and Abbreviations

Acronym	Definition
ABSCO	ACOS/OCO-2 absorption coefficient
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
ESA	European Space Agency
GHG	Greenhouse Gas
HAPI	HITRAN Application Programming Interface
IASI	Infrared Atmospheric Sounding Instrument
ICON-ART-iso	Icosahedral Nonhydrostatic – Aerosols and Reactive Trace Gases - Isotopologues
ISRF	Instrument Spectral Response Function
L1b	Level 1b data product
L2	Level 2 data product
L4	Level 4 data product
LCC	Land Cover Change
LRPT	Leicester Retrieval Preparation Toolset
LSM	Land Sea Mask
LTAN	Local Time Ascending Node
L-WAIVE	Lacustrine-Water vApor Isotope inVentory Experiment
MIR	Mid infrared
NIR	Near Infrared
MUSICA	Multi-platform Remote Sensing of Isotopologues for Investigating the Cycle of Atmospheric Water
NDACC	Network for the Detection of Atmospheric Composition Change
NRT	Near Real Time
S5p	Sentinel 5 precursor
S5p+I	Sentinel-5p+Innovation
SRTM	Shuttle Radar Topography Mission
Suomi-NPP	Suomi National Polar-orbiting Partnership
SWIR	Shortwave Infrared
TCCON	Total Carbon Column Observing Network
TOA	Top-of-Atmosphere

TROPOMI	TROPOspheric Monitoring Instrument
UoL-FP	University of Leicester Full Physics
VIIRS	Visible/Infra-red Imager and Radiometer Suit

7 References

/**Birk et al. 2017**/ Birk, M., Wagner, G., Loos, J., Mondelain, D. and Campargue, A., 2017. ESA SEOM-IAS–Spectroscopic parameters database 2.3 µm region.

/**Boesch et al. 2013**/ 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

/**Chazette et al., 2021**/ Chazette, P., Flamant, C., Sodemann, H., Totems, J., Monod, A., Dieudonné, E., Baron, A., Seidl, A., Steen-Larsen, H.C., Doira, P. and Durand, A., 2021. Experimental investigation of the stable water isotope distribution in an Alpine lake environment (L-WAIVE). Atmospheric Chemistry and Physics, 21(14), pp.10911-10937.

/**Craig 1961**/ Craig, H., 1961. Standard for reporting concentrations of deuterium and oxygen-18 in natural waters. Science, 133(3467), pp.1833-1834.

/**Defourney et al. 2015**/ Defourney, P. et al. (2015): ESA Land Cover Climate Change Initiative (ESA LC_cci) data: Global Land Cover Maps, Version 2.0.7 via Centre for Environmental Data Analysis, 30/01/2020

/**Diekmann et al. 2021a**/ Diekmann, C. J. et al. (2021): The MUSICA IASI {H2O, δD} pair product, Earth Syst. Sci. Data Discuss. [preprint], DOI: 10.5194/essd-2021-87, in review

/**Diekmann et al. 2021b**/ Diekmann, C. J. et al. (2021): MUSICA IASI water isotopologue pair product (a posteriori processing version 2). Institute of Meteorology and Climate Research, Atmospheric Trace Gases and Remote Sensing (IMK-ASF), Karlsruhe Institute of Technology (KIT). DOI: 10.35097/415

/**Eckstein 2017**/ Eckstein, J. (2017): Atmospheric Models and Aircraft Measurements: Representativeness and Model Validation, Dissertation at Institute of Meteorology and Climate Research, Atmospheric Trace Gases and Remote Sensing (IMK-ASF), Karlsruhe Institute of Technology (KIT). DOI: 10.5445/IR/1000073299

/**Eckstein et al. 2018**/ Eckstein, J. et al. (2018): From climatological to small-scale applications: simulating water isotopologues with ICON-ART-Iso (version 2.3), Geosci. Model Dev., 11, 5113–5133, DOI: 10.5194/gmd-11-5113-2018

/**Hersbach and Dee, 2016a**/ Hersbach, H. and Dee, D., 2016. ERA5 reanalysis is in production, ECMWF Newsletter 147, ECMWF. Reading, UK.

/**Hersbach and Dee, 2016b**/ Hersbach, H. and Dee, D., 2016. ERA5 reanalysis is in production. ECMWF newsletter, 147(7), pp.5-6

/**Kochanov et al., 2016**/ 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

/**Loots et al. 2017**/ E. Loots, L. Babić, R. Braak, W. Dierssen, J. Kissi-Ameyaw, Q. Kleipool, J. Leloux, A. Ludewig, N. Rozemeijer, J. Smeets, G. Vacanti (2017): Algorithm theoretical basis document for the TROPOMI L01b data processor, source: S5P-KNMI-L01B-0009-SD

/**Marécal et al. 2015**/ Marécal, V., Peuch, V.H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A. and Chéroux, F., 2015. A regional air quality forecasting system over Europe: the MACC-II daily ensemble production.

/**Payne 2017**/ OCO (Orbiting Carbon Observatory) -2 ABSCO User Guide, Version 5, V. Payne, JPL, July 24 2017

/**Pfahl et al., 2012**/ Pfahl, S., Wernli, H. and Yoshimura, K., 2012. The isotopic composition of precipitation from a winter storm–a case study with the limited-area model COSMO iso. Atmospheric Chemistry and Physics, 12(3), pp.1629-1648.

/**Reichler et al. 2003**/ Reichler, T., Dameris, M. and Sausen, R., 2003. Determining the tropopause height from gridded data. Geophysical research letters, 30(20).

/**Scheepmaker et al. 2016**/ 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.

/**Schneider et al. 2017**/ Schneider, M., Borger, C., Wiegele, A., Hase, F., García, O. E., Sepúlveda, E., and Werner, M.: MUSICA MetOp/IASI {H_2O,δD} pair retrieval simulations for validating tropospheric moisture pathways in atmospheric models, Atmos. Meas. Tech., 10, 507–525, https://doi.org/10.5194/amt-10-507-2017, 2017

/**Siddens 2016**/ R Siddens (2016): S5P-NPP Cloud Processor ATBD, source: S5P-NPPC-RAL-ATBD-0001

/**Vonk et al. 2018**/ F. Vonk, N. Rozemeijer, Q. Kleipool, R. van Versendal (2018): Input/output data specification for the TROPOMI L01b data processor, source: S5P-KNMI-L01B-0012-SD.

/**Wunch et al. 2011**/ Wunch, D., G. C. Toon, J.-F. L. Blavier, R. A. Washenfelder, J. Notholt, B. J. Connor, D. W. T. Griffith, V. Sherlock, and P. O. Wennberg (2011), The total carbon column observing network, Philosophical Transactions of the Royal Society - Series A: Mathematical, Physical and Engineering Sciences, 369 (1943), 2087-2112, doi:10.1098/rsta.2010.0240. Available from: http://dx.doi.org/10.1098/rsta.2010.02.