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

Validation Report (VR)

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

European Space Agency
Agence spatiale européenne

Change log

Version	Date	Status	Author	Reason for change
Draft 0.1	**2020-08-29**	initial internal draft for project team	M. Schneider, A.N. Röhling, F. Khosrawi, H. Bösch, T. Trent, H. Sodemann	New document
Draft 1.0	**2020-10-02**	Document prepared for submission	“	Consolidated Version
Draft 1.1	**2020-11-11**	Revised document	“	Updates based on ESA comments
Draft 1.2	**2021-07-23**	Reworked document	“	Updated data product V0.9.5
Draft 1.3	**2021-09-29**	Document prepared for submission	“	Consolidated Version

Contents

1 Purpose and objective 5
2 Document overview 5
3 References, terms and acronyms 6
3.1 References 6
3.2 Applicable Documents 8
3.3 Terms, definitions and abbreviated terms 8
4 Product requirements 9
4.1 Theoretical characterization of ratio remote sensing data 9
4.2 Requirements on sensitivity, bias and noise 12
5 Reference Measurements 13
5.1 Ground based monitoring networks 14
5.1.1 NDACC-IRWG .... 15~~15~~~~14~~
5.1.2 TCCON 16
5.2 Satellite measurements 20
5.2.1 MUSICA IASI 20
5.2.2 GOSAT 22
5.3 Field campaigns 22
5.4 Discussion on the vertical representativeness of reference data 23
5.5 Modelling support 23
6 Validation approach 24
6.1 Spatial and temporal collocation with reference observations 24
6.2 Recommendations for TROPOMI data usage 25
6.3 Status of validation 26
6.4 Dependence on influence quantities 26
6.5 Bias 31
6.6 Dispersion 33
6.7 Short-scale and short-term variability 33
6.8 Geographical pattern 34
6.9 Comparison to MUSICA IASI references 35
6.10 Bias Correction 37

7 Conclusions 39

1 Purpose and objective

The purpose of this document is to describe the validation operations of the S5P+I Level-2 products. The document is maintained during the development phase and the lifetime of the data products. Updates and new versions will be issued in case of changes in the processing chains or for novel validation exercises.

2 Document overview

In Section 3 we present the literature references and a list of the acronyms used throughout this document. Section 4 discusses the requirements on the new TROPOMI XHDO/XH2O product. This comprises the need for a detailed analytic description of the product characteristics as well as a discussion of the uncertainty levels required for a scientific usefulness of the product. Section 5 presents the data products used as references in this validation study. Their main characteristics (coverage, reliability, representativeness) are discussed. Section 6 displays the comparisons of the TROPOMI XHDO/XH2O product with the reference data. Section 7 concludes with the main validation results.

3 References, terms and acronyms

3.1 References

[RD1] Schneider, M., Wiegele, A., Barthlott, S., González, Y., Christner, E., Dyroff, C., García, O. E., Hase, F., Blumenstock, T., Sepúlveda, E., Mengistu Tsidu, G., Takele Kenea, S., Rodríguez, S., and Andrey, J.: Accomplishments of the MUSICA project to provide accurate, long-term, global and high-resolution observations of tropospheric {H2O,δD} pairs – a review, Atmos. Meas. Tech., 9, 2845-2875, doi:10.5194/amt-9-2845-2016, 2016.

[RD2] von Clarmann, T., Degenstein, D. A., Livesey, N. J., Bender, S., Braverman, A., Butz, A., Compernolle, S., Damadeo, R., Dueck, S., Eriksson, P., Funke, B., Johnson, M. C., Kasai, Y., Keppens, A., Kleinert, A., Kramarova, N. A., Laeng, A., Langerock, B., Payne, V. H., Rozanov, A., Sato, T. O., Schneider, M., Sheese, P., Sofieva, V., Stiller, G. P., von Savigny, C., and Zawada, D.: Overview: Estimating and reporting uncertainties in remotely sensed atmospheric composition and temperature, Atmos. Meas. Tech., 13, 4393–4436, https://doi.org/10.5194/amt-13-4393-2020, 2020.

[RD3] 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.

[RD4] Schneider, M., Barthlott, S., Hase, F., González, Y., Yoshimura, K., García, O. E., Sepúlveda, E., Gómez-Peláez, A., Gisi, M., Kohlhepp, R., Dohe, S., Blumenstock, T., Wiegele, A., Christner, E., Strong, K., Weaver, D., Palm, M., Deutscher, N. M., Warneke, T., Notholt, J., Lejeune, B., Demoulin, P., Jones, N., Griffith, D. W. T., Smale, D., and Robinson, J.: Ground-based remote sensing of tropospheric water vapour isotopologues within the project MUSICA, Atmos. Meas. Tech., 5, 3007-3027, doi:10.5194/amt-5-3007-2012, 2012.

[RD5] Barthlott, S., Schneider, M., Hase, F., Blumenstock, T., Kiel, M., Dubravica, D., García, O. E., Sepúlveda, E., Mengistu Tsidu, G., Takele Kenea, S., Grutter, M., Plaza-Medina, E. F., Stremme, W., Strong, K., Weaver, D., Palm, M., Warneke, T., Notholt, J., Mahieu, E., Servais, C., Jones, N., Griffith, D. W. T., Smale, D., and Robinson, J.: Tropospheric water vapour isotopologue data (H216O, H218O, and HD16O) as obtained from NDACC/FTIR solar absorption spectra, Earth Syst. Sci. Data, 9, 15-29, doi:10.5194/essd-9-15-2017, 2017.

[RD6] Schneider, M., Hase, F., Blumenstock, T.: Ground-based remote sensing of HDO/H 2 O ratio profiles: introduction and validation of an innovative retrieval approach, Atmos. Chem. Phys., Vol. 6, 4705-4722, doi:10.5194/acp-6-4705-2006, 2006.

[RD7] Worden, J., Bowman, K., Noone, D., Beer, R., Clough, S., Eldering, A., Fisher, B., Goldman, A., Gunson, M., Herman, R., Kulawik, S. S., Lampel, M., Luo, M., Osterman, G., Rinsland, C., Rodgers, C., Sander, S., Shephard, M., and Worden, H.: Tropospheric Emission Spectrometer observations of the tropospheric

HDO/H2O ratio: Estimation approach and characterization, J. Geophys. Res., 111, D16309, https://doi.org/10.1029/2005JD006606, 2006.

[RD8] Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J.,Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 369, 2087–2112, doi:10.1098/rsta.2010.0240, 2011.

[RD9] Schneider, M. and Hase, F.: Optimal estimation of tropospheric H2O and δD with IASI/METOP, Atmos. Chem. Phys., 11, 11 207–11 220, https://doi.org/10.5194/acp-11-11207-2011, 2011.

[RD10] Lacour, J.-L., Risi, C., Clarisse, L., Bony, S., Hurtmans, D., Clerbaux, C., and Coheur, P.-F.: Mid-tropospheric δD observations from IASI/MetOp at high spatial and temporal resolution, Atmos. Chem. Phys., 12, 10 817–10 832, https://doi.org/10.5194/acp-12-10817-2012, 2012.

[RD11] Frankenberg, C., Wunch, D., Toon, G., Risi, C., Scheepmaker, R., Lee, J.-E., Wennberg, P., and Worden, J.: Water vapor isotopologue retrievals from high-resolution GOSAT shortwave infrared spectra, Atmos. Meas. Tech., 6, 263–274, doi:10.5194/amt-6-263-2013, 2013.

[RD12] 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, doi:10.5194/amt-6-599-2013, 2013.

[RD13] Schneider, A., Borsdorff, T., aan de Brugh, J., Aemisegger, F., Feist, D.G., Kivi, R., Hase, F., Schneider, M., and Landgraf, J.: First data set of H2O/HDO columns from the Tropospheric Monitoring Instrument (TROPOMI), Atmos. Meas. Tech., 13, 85–100, https://doi.org/10.5194/amt-13-85-2020, 2020.

[RD14] Hamperl, J., Capitaine, C., Santagata, R., Dherbecourt, J., Melkonian, J., Godard, A., Raybaut, M., Régalia, L., Grouiez, B., Blouzon, F., Geyskens, N., Evesque, C., Chazette, P., Totems, J., and Flamant, C.: WaVIL: a Differential Absorption LIDAR for Water Vapor and Isotope HDO Observation in the Lower Troposphere - Instrument Design, Optical Sensors and Sensing Congress, OSA Technical Digest, paper LM4A.4, Optical Society of America, 2020.

[RD15] Schneider, M., Ertl, B., Diekmann, C. J., Khosrawi, F., Weber, A., Hase, F., Höpfner, M., García, O. E., Sepúlveda, E., and Kinnison, D.: Design and description of the MUSICA IASI full retrieval product, Earth Syst. Sci. Data Discuss. [preprint], https://doi.org/10.5194/essd-2021-75, in review, 2021.

[RD16] Diekmann, C. J., Schneider, M., Ertl, B., Hase, F., García, O., Khosrawi, F., Sepúlveda, E., Knippertz, P., and Braesicke, P.: The MUSICA IASI {H_2O, δD} pair product, Earth Syst. Sci. Data Discuss. [preprint], https://doi.org/10.5194/essd-2021-87, in review, 2021.

[URL1] MUSICA web page. URL https://www.imk-asf.kit.edu/english/musica.php. Date: 2020-09-29

3.2 Applicable Documents

[AD1] Requirements Baseline Document (RB). https://s5pinnovationh2o-iso.le.ac.uk/wp-content/uploads/2019/11/RBD_S5pI_H2O_ISO_v1.1.pdf. Date: 2020-09-29

[AD2] Algorithm Theoretical Basis Document (ATBD). https://s5pinnovationh2o-iso.le.ac.uk/wp-content/uploads/2021/02/S5P-I_ISO_ATBD_Version1.3.pdf. Date: 2020-09-29

[AD3] Auxiliary User Manual (AUM). https://s5pinnovationh2o-iso.le.ac.uk/wp-content/uploads/2020/06/aum_s5p-i_iso_version_1.3.pdf. Date 2021-08-13

3.3 Terms, definitions and abbreviated terms

The most important symbols, acronyms and abbreviations related to the data product of this document are described in this subsection.

GOSAT	Greenhouse Gases Observing Satellite
IASI	Infrared Atmospheric Sounding Interferometer
MUSICA	Multi-platform remote Sensing of Isotopologues for investigating the Cycle of Atmospheric water
NDACC	Network for the Detection of Atmospheric Composition Change
S5P	Sentinel-5p
TCCON	Total Carbon Column Observing Network
MOTIV	Moisture Transport pathways and Isotopologues in water Vapour
TEDDY	Testing isotopologues as Diabatic heating proxy for atmospheric Data analyses
TROPOMI	Tropospheric Monitoring Instrument
WAVIL	Water Vapor Isotope Lidar
NIR	Near Infrared
MIR	Mid infrared

4 Product requirements

For a proper usage of any remote sensing product as well as for a concise definition of the requirements on the product an analytic method for documenting the product's characteristics is essential. This means an analytic method for describing the uncertainty of the product, but also for documenting the principle characteristics of its representativeness. Representativeness here means the sensitivity of the retrieved product to real atmospheric variabilities. For trace gas remote sensing products this representativeness is documented by the averaging kernel and reporting these averaging kernels is strongly recommended for remote sensing products ([RD2], actually it is indispensable for a proper usage of the data obtained by retrievals that use a priori information).

Trace gas ratio remote sensing products like HDO/H2O are very sophisticated remote sensing data products and an analytic formulation of the corresponding averaging kernels need special care. In the first part of this section we present a formalism for analytically documenting the characteristics of total column trace gas ratio remote sensing products (averaging kernels and uncertainties). We think that it is essential to document the representativeness of the TROPOMI HDO/H2O product using such a method. In the second part of this section we discuss the characteristics (sensitivity and uncertainty) required for a scientific usefulness of the product.

4.1 Theoretical characterization of ratio remote sensing data

Trace gas ratio remote sensing products are very sophisticated remote sensing data products. In the context of the European Research Council project MUSICA (MUlti-platform remote Sensing of Isotopologues for the investigation of the Atmospheric water Cycle), an analytic method for characterising such products has been developed [RD1]. The MUSICA project dealt with the remote sensing of water isotoplogue ratio profiles (using ground-based NDACC and space-based IASI observations).

The principle idea of the MUSICA method is to characterise trace gas ratio products by a transformation into the logarithmic scale, because in a logarithmic scale we can write the ratio as a difference:

$$\ln\left[\frac{\hat{x}_{\mathrm{HDO}}}{\hat{x}_{\mathrm{H2O}}}\right] = \ln\hat{x}_{\mathrm{HDO}} - \ln\hat{x}_{\mathrm{H2O}}$$

Let the operator M be a diagonal matrix with entries corresponding to the retrieved H2O and HDO mixing ratio profiles. The logarithmic scale-averaging kernel $\mathrm{A}_{\mathrm{log}}$ can then be calculated from the linear scale-averaging kernel $\mathrm{A}_{\mathrm{lin}}$ by:

$$\mathrm{A}_{\mathrm{log}} = \mathrm{M}^{-1}\,\mathrm{A}_{\mathrm{lin}}\,\mathrm{M},$$

with the $2N$ x $2N$ diagonal matrix:

$$\mathrm{M} = \begin{pmatrix} \hat{x}_{\mathrm{H2O},1} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{x}_{\mathrm{H2O},N} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \hat{x}_{\mathrm{HDO},1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \hat{x}_{\mathrm{HDO},N} \end{pmatrix}.$$

Using the 2N x 2N transformation matrix P:

$$\mathrm{P} = \begin{pmatrix} +0.5 & \cdots & 0 & +0.5 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & +0.5 & 0 & \cdots & +0.5 \\ -1.0 & \cdots & 0 & +1.0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -1.0 & 0 & \cdots & +1.0 \end{pmatrix},$$

We can then calculate the proxy state kernel $\mathrm{A_p}$ as:

$$\mathrm{A_p} = \mathrm{P}\,\mathrm{A_{log}}\,\mathrm{P}^{-1}.$$

The proxy state consists of the sum of the logarithms ($0.5[\ln \hat{x}_{\mathrm{H2O}} + \ln \hat{x}_{\mathrm{HDO}}]$) and the difference of the logarithms ($\ln \hat{x}_{\mathrm{HDO}} - \ln \hat{x}_{\mathrm{H2O}}$). The former is a proxy of humidity and the latter a proxy of the HDO/H2O ratio (for more details please refer to [RD1]).

Here we extend this MUSICA formalism to total column ratio remote sensing products. A column averaging kernel can be calculated from a linear profile averaging kernel $\mathrm{A_{lin}}$ by:

$$a^T = s^T\,\mathrm{H}\,\mathrm{A_{lin}}\,\mathrm{H}^{-1},$$

with s^T being a row vector with entry 1.0 everywhere and H being a 2N x 2N diagonal matrix:

$$\mathrm{H} = \begin{pmatrix} h_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & h_N & 0 & \cdots & 0 \\ 0 & \cdots & 0 & h_1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & h_N \end{pmatrix}$$

Here h_i is the number of dry air molecules within the layer represented by altitude level i.

The proxy kernel for column ratio products can now be written as:

$$a_p^T = \mathrm{P_X}\, \mathrm{M_X^{-1}}\, \mathrm{H_X^{-1}}\, a^T \mathrm{H\,M\,P^{-1}},$$

with $\mathrm{H_X}$, $\mathrm{M_X}$, and $\mathrm{P_X}$ being the following 2 x 2 matrices:

$$\mathrm{H_X} = \begin{pmatrix} \sum_{i=1}^{i=N} h_i & 0 \\ 0 & \sum_{i=1}^{i=N} h_i \end{pmatrix}$$

$$\mathrm{M_X} = \begin{pmatrix} \hat{X}_{\mathrm{H2O}} & 0 \\ 0 & \hat{X}_{\mathrm{HDO}} \end{pmatrix}$$

$$\mathrm{P_X} = \begin{pmatrix} +0.5 & +0.5 \\ -1.0 & +1.0 \end{pmatrix}.$$

Here $\hat{X}_{\mathrm{H2O}}$ and $\hat{X}_{\mathrm{HDO}}$ are the retrieved dry air column averaged mixing ratios of H2O and HDO, respectively.

In analogy, we can write the error covariance for the proxy state as:

$$\mathrm{S_{X,p}} = \mathrm{P_X}\, \mathrm{M_X^{-1}}\, \mathrm{H_X^{-1}}\, \mathrm{S_X}\, [\mathrm{P_X}\, \mathrm{M_X^{-1}}\, \mathrm{H_X^{-1}}]^T,$$

with $\mathrm{S_X}$ being the error covariance of the retrieved $\hat{X}_{\mathrm{H2O}}$ and $\hat{X}_{\mathrm{HDO}}$ total column products.

Most water vapour resides in the lower troposphere meaning that the total water vapour column is strongly linked to the water vapour amount in the lower troposphere. In order to characterise the representativeness of the retrieved total column data we introduce a parameter that documents the sensitivity of the product to the lower tropospheric variabilities. This can be simply achieved by summing up the lower tropospheric entries of a_p^T and normalization to the tropospheric column by scaling with $\mathrm{M_X}\, \mathrm{H_X}\, \mathrm{M_{LTX}^{-1}}\, \mathrm{H_{LTX}^{-1}}$ (with $\mathrm{M_{LTX}}$ and $\mathrm{H_{LTX}}$ of being the lower tropospheric counterparts of $\mathrm{M_X}$ and $\mathrm{H_X}$). We call this parameter the lower tropospheric sensitivity parameter "*sens(LT)*". Here and in the following we consider as the lower troposphere the altitudes from the surface to 2.5 km above sea level.

The representativeness of a total column integrated HDO/H2O data product should be documented by means of total column proxy kernels. The parameter "*sens(LT)*" is a very useful measure for the sensitivity of total column product on the lower troposphere.

4.2 Requirements on sensitivity, bias and noise

We estimate the uncertainty needed for a very detailed detection of fluxes between the surface and the atmosphere to be below 5 ‰, whereby at the same time a horizontal resolution of a few hundred meters would be required. However, it is rather unrealistic to achieve such low uncertainty levels together with the required high horizontal resolution from space-based remote sensing techniques.
From our experiences with using water vapour isotopologue data for scientific applications (e.g. within the projects MUSICA, MOTIV, and TEDDY), we conclude that the S5-P column integrated HDO/H_2O (or δD) product should have a systematic and random uncertainty of less than 50 ‰. This is the threshold for being able to identify different water cycle processes (see for instance figures in [RD3]). Our target will be an uncertainty of 10-20 ‰, which is similar to the uncertainty of the MUSICA IASI and MUSICA NDACC δD remote sensing products and has shown to be sufficient for investigating a variety of atmospheric water cycle processes.
TROPOMI can complement the global free tropospheric water vapour isotopologue MUSICA IASI data with information on the lower troposphere (the total column averaged ratio mainly represents the ratio in the lower troposphere, where most water vapour resides). However, it is important to theoretically demonstrate the lower tropospheric sensitivity of the TROPOMI total column integrated HDO/H2O ratio product. In the previous subsection, we developed a theoretical framework for this purpose and the parameter "*sens(LT)*" can reveal this lower tropospheric sensitivity.

5 Reference Measurements

The TROPOMI water vapour isotopologue product is a column integrated product. The isotopologue retrievals are conducted over land surfaces with global coverage and will be available for a period starting in May 2018. For the development and validation (including impact studies) of the TROPOMI isotopologue product, we need reference data sets with the following characteristics:

1. They should cover the period between July 2018 and June 2019 (better: until June 2020).
2. They should be representative for different climate zones over land.
3. They should be sensitive to water vapour throughout the troposphere.
4. They should be traceable to fiducial references.

Figure 1 gives an overview on the reference products that fulfil some of these requirements. The MUSICA NDACC (Network for the Detection of Atmospheric Composition Change) reference data is representative for the whole troposphere, calibrated with respect to fiducial references. However, only at three stations (Kiruna, Karlsruhe and Izaña) are data currently available for the era 2018-2020 of TROPOMI observations. TCCON (Total Carbon Column Observing Network) data are also representative for the whole troposphere and available for the whole TROPOMI observation era; although, the TCCON isotopologue ratio data are not calibrated with respect to fiducial references. The same is true for GOSAT space-based observations. MUSICA IASI space-based products are calibrated with respect to fiducial references, available during the TROPOMI era on global scale; however, they are only representative for the free troposphere (thus not well representative for the boundary layer). The aircraft-based in-situ observations (IGP/LEMON) represent a fiducial reference, but they are only available for short campaign periods, for a limited geographical region, and only cover the lowermost four kilometres of the troposphere.

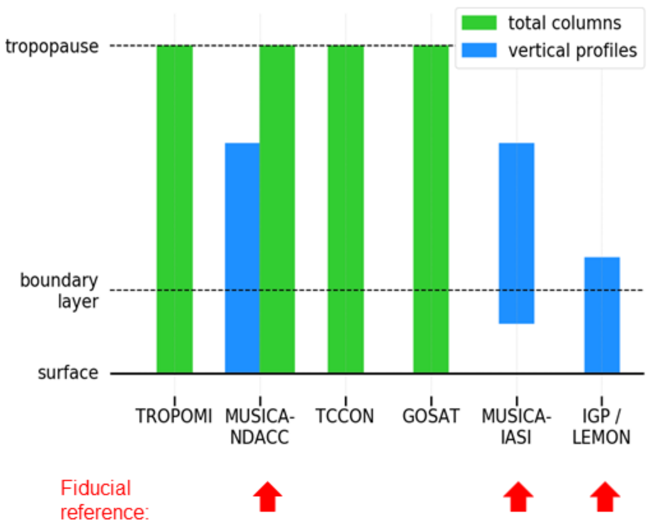

Figure 1: Vertical representativeness of TROPOMI data and of the associated data. Green collour means column integrated data and blue means data containing some information about the vertical distribution. Note: generally, more than 50 % of the water vapour total column resides in the boundary layer.

In summary, there is no reference data product that fulfils all the requirements. Instead, we will exploit all these different reference data together in order to be able to perform a successful validation of the TROPOMI product.
Figure 2 gives an overview on the geographical location of the different reference observation sites. The reference data are obtained in Northern and Central Europe as well as in the tropics of the Western Pacific and Southern Australia.

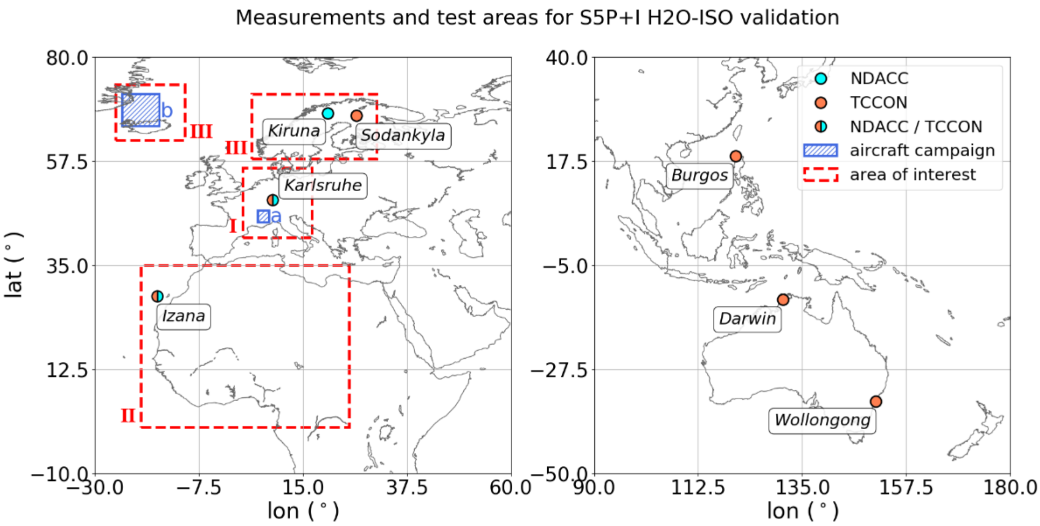

Figure 2: Collection of the test areas and observations defined for the TROPOMI validation and impact assessment. The cyan and orange dots indicate the chosen NDACC and TCCON stations, respectively. The regions over Central Europe (red, I), West Africa (red, II) and Iceland/Scandinavia (red, III) serve as areas of interest and will be provided with model data. Additionally, the observations from the aircraft campaigns LEMON (blue, a) and IGP (blue, b) might be useful for validation.

5.1 Ground based monitoring networks

For this study, we use data produced from high-resolution solar absorption spectra measured within the framework of the two ground-based networks NDACC (Network for the Detection of Atmospheric composition Change) and TCCON (Total Carbon Column Observing Network). At the sites we consider the spectra that are recorded with the FTIR (Fourier Transform InfraRed) spectrometer Bruker IFS 125 HR. This instrument is currently the most robust and stable high-resolution Fourier Transform Spectrometer (FTS) commercially available to acquire solar absorption spectra in the NIR (near infrared) and MIR (mid infrared) regions.

5.1.1 NDACC-IRWG

The Infrared Working Group of the Network for the Detection of Atmospheric Composition Change (NDACC-IRWG) is part of the GAW program (Global Atmosphere Watch) of the World Meteorological Organization (WMO). The network consists of about 20 globally distributed ground-based FTIR stations. Standard NDACC instruments measure in the MIR (750 – 4200 cm^{-1}) spectral region using InSb and MCT detectors, which allow obtaining profile information about several trace gases. In the framework of the MUSICA project a retrieval for the water vapour isotopologues $H2^{16}O$, $H2^{18}O$ and $HD^{16}O$ has been developed [RD4] [RD5]. This MUSICA NDACC retrieval consists of an optimal estimation of the vertical distribution of water vapour and the ratios between the different water vapour isotopologues. For achieving the optimal estimation of the highly varying water vapour concentrations and their ratios, the use of a logarithmic scale during the inversion procedure is decisive [RD6] [RD7].
Figure 3 depicts the typical averaging kernels for MUSICA NDACC data given in the {H2O, δD, D-excess} proxy state (more details see [RD5]). The H2O data can be retrieved with a typical Degree Of Freedom for Signal (DOFS) of about 3, the ratio HDO/H2O (or δD) with a typical DOFS of 1.5-2.0, and D-excess (defined as $\delta D- 8*\delta^{18}$) with a DOFS of clearly below 1.0. This means that the data enables us to distinguish (to some extent) the HDO/H2O as present in the lower troposphere from the ratio in the free troposphere.

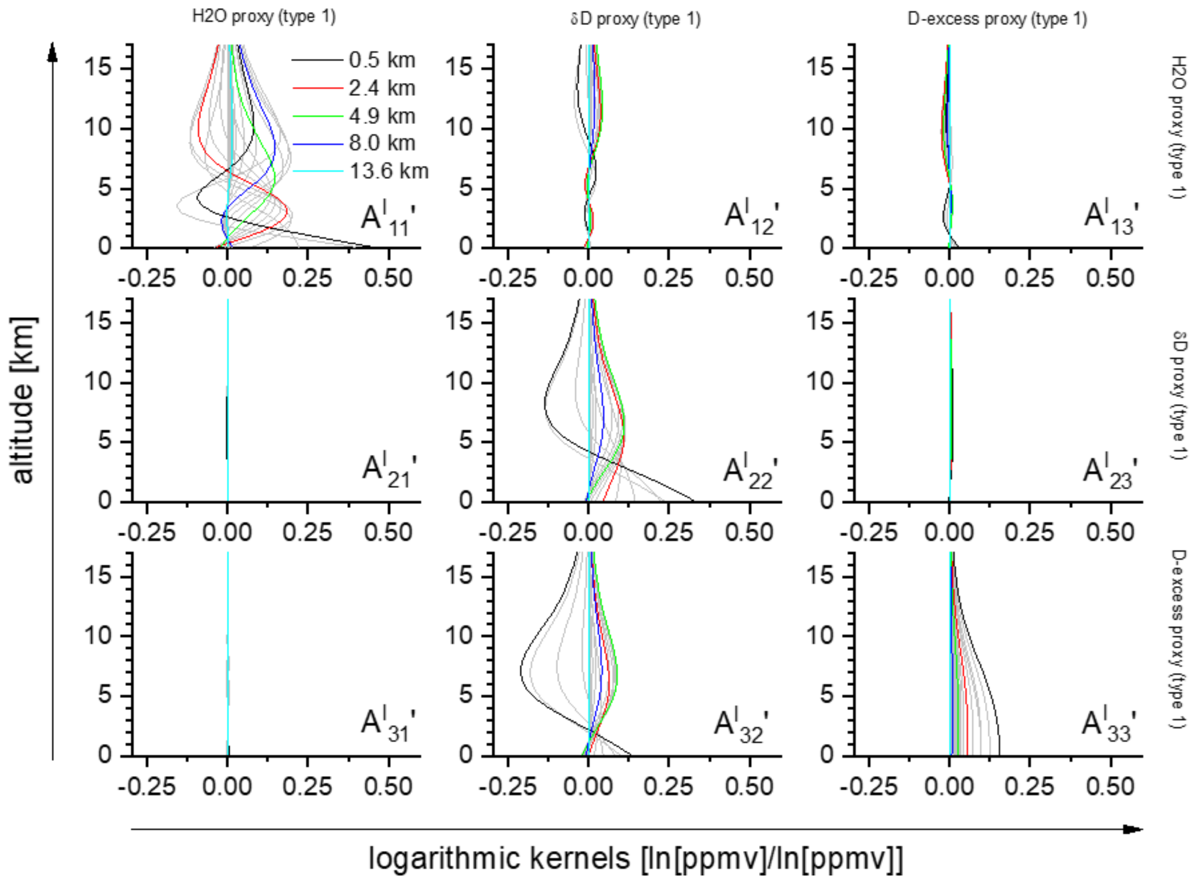

Figure 3: Example of typical MUSICA NDACC averaging kernels. Relevant for this study are the averaging kernel plots representing the {H2O, δD} proxy state, i.e. the blocks: Al11’, Al12’, Al21’ and Al22’ (Figure adopted from [RD5]).

The MUSICA NDACC data have been calibrated with respect to fiducial standards using six aircraft profiles [RD1]. This calibration constrained the column integrated δD (XδD) bias within 15 ‰ and the column integrated H2O (XH2O) bias within 15 %. In a recent update we have calibrated the MUSICA NDACC XH2O data with respect to coinciding TCCON XH2O (recall that TCCON XH2O is calibrated with respect to fiducial references). The calibration factors are 0.89 for XH2O (as well as for XHDO in order to maintain the MUSICA NDACC XδD calibration).

MUSICA NDACC data are available for 11 stations during a continuous time-period until the end of 2014. Since the end of the MUSICA funding period no further MUSICA processing of the continuously measured NDACC spectra has been possible (lack of funding). In the context of this TROPOMI validation project we have been able to perform MUSICA NDACC retrievals at the three NDACC stations operated by the Karlsruhe Institute of Technology (Kiruna, Karlsruhe and Izaña), which is important given the unique water vapour isotopologue reference data quality offered by the MUSICA NDACC data.

Stations with MUSICA NDACC data used for this TROPOMI validation are listed in Table 1 (please note Izaña data are not used because there are no TROPOMI water isotopologues available for the ocean scenes around the island of Tenerife).

Table 1: Summary of the NDACC FTIR stations used for TROPOMI validation. Location, coordinates, altitude above sea level in meter, and spectral resolution~~, retrieval code.~~ .

FTIR Station	Location	Coordinates	Altitude [m. a.s.l.]	Spectral resolution [cm^{-1}]
Karlsruhe	Germany	49.10°N 8.439°E	119	0.005
Kiruna	Sweden	67.8°N 20.4°E	420	0.005
Izaña	Tenerife	28.3°N 16.5°E	2370	0.005

5.1.2 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. Two detectors, an extended InGaAs and silicon (Si) detector cover the spectral range from 3900 to 15500 cm^{-1} in dual acquisition mode giving column-averaged abundances of several trace gases, including XH2O and XHDO. The retrievals use vertical profiles from meteorological reanalyses as a priori. These profiles are then scaled during the retrieval procedure (no profile retrievals!). The XHDO/XH2O ratio is then calculated a posteriori (after the retrieval process). This is a rather different retrieval procedure as the one used by MUSICA NDACC.

We found a systematic bias of the TCCON XδD values with respect to the MUSICA NDACC XδD values of -65 ‰. This is in line with the findings of [RD13]. Because TCCON XH2O is calibrated with respect to fiducial references, we can assign this bias to -6.5 % in TCCON XHDO. Consequently, our TCCON bias correction consists in applying a calibration factor of 0.935 to the TCCON XHDO data.

Figure 4 shows an example of the good agreement between MUSICA NDACC and TCCON XH2O and XδD after applying the calibration factors as described above. By transferring the MUSICA NDACC standard to TCCON we can use MUSICA NDACC and TCCON XδD in a consistent manner, i.e. use TCCON as an XδD reference where no MUSICA NDACC data are available. The TCCON stations used for this TROPOMI validation study can be found in Table 2.

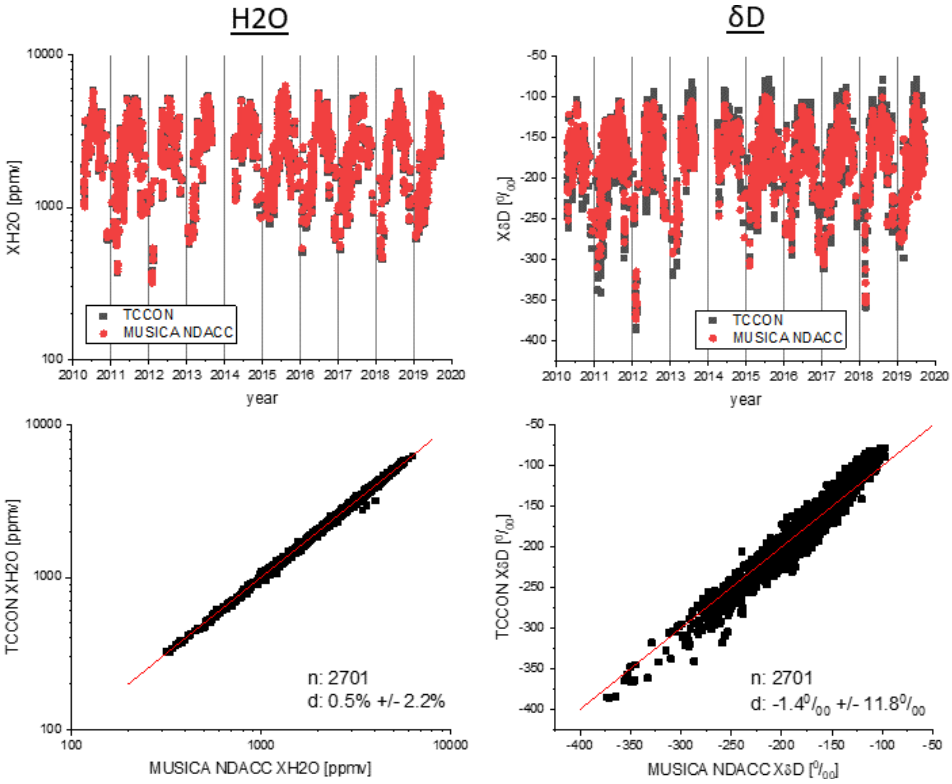

Figure 4: Comparison of the MUSICA NDACC and TCCON XH2O and XδD data after transferring the MUSICA NDACC fiducial XδD standard to the TCCON XδD data (taking data from the Karlsruhe site as an example).

Table 2: Summary of the TCCON FTIR stations used for TROPOMI validation. Location, coordinates, altitude above sea level in metres, and spectral resolution~~, and retrieval code~~.

FTIR Station	Location	Coordinates	Altitude [m. a.s.l.]	Spectral resolution [cm^{-1}]
Burgos	Philippines	18.53°N 120.65°E	35	0.02
Darwin	Australia	12.42°S 130.89°E	30	0.02
Karlsruhe	Germany	49.10°N 8.439°E	119	0.02
Sodankylä	Finnish Lapland	67.37°N 26.63°E	188	0.02
Wollongong	Australia	34.41°S 150.88°E	30	0.02

TCCON provides the XH2O and XHDO data together with column averaging kernels, which inform about the sensitivity of the total column retrieval on H2O or HDO changes at different altitudes (change of number of molecules in the retrieved total column due to change of number of molecules at a certain altitude level). If the XH2O and XHDO averaging kernels were identical, the a posteriori calculation of XHDO/XH2O would be straightforward. However, actually there are generally small differences between the XH2O and XHDO kernels. The left panels of Fig. 5 represent typical TCCON XH2O and XHDO kernels (a^T from Section 4.1) for the lowermost 10 km (where almost all atmospheric water vapour resides). Cross-kernels (i.e. information on the effect of atmospheric H2O variability on XHDO and of atmospheric HDO variabilities on XH2O) are not provided, and for the further calculation we assume them to be negligible. The right panels of Fig. 5 depict the proxy column averaging kernels introduced in Section 4.1. as a_p^T. The HDO/H2O (or δD) proxy column averaging kernel is shown as the bottom right chart of the right panels. It reveals how a change in the HDO/H2O ratio at a certain altitude layer affects the retrieved total column ratio (XHDO/XH2O). The shape of this kernel strongly depends on the vertical distribution of H2O and on the thickness of the layers for which it is calculated. Since most water resides in the lowermost troposphere, these proxy kernels have a maximum in the lower troposphere.

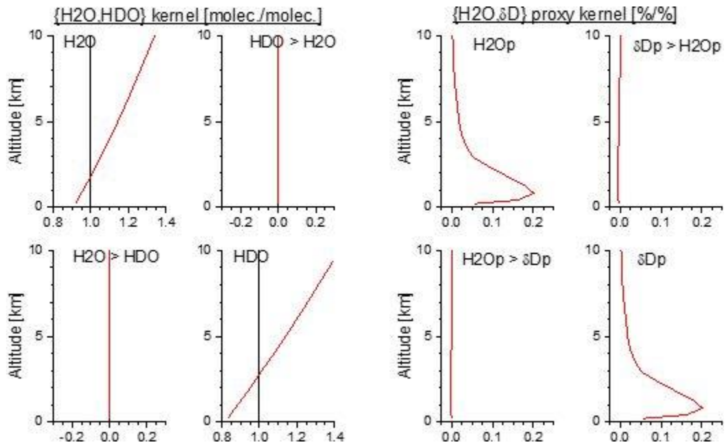

Figure 5: Typical column averaging kernels for TCCON. Left: kernels and cross kernels for H2O and HDO; Right: kernels and cross-kernels for the H2O and δD proxy states (0.5*(ln[H2O]-ln[HDO]) and ln[H2O]-ln[HDO], respectively).

In Figure 6, we compare HDO/H2O proxy column averaging kernels obtained for different reference data sets (and from an exemplary TROPOMI observation). The left panels show typical examples of the proxy column kernel. Please note that the proxy kernels depend on the vertical distribution of H2O, which is prescribed for TCCON and can be retrieved in the case of MUSICA NDACC and MUSICA IASI. For MUSICA NDACC and TCCON the kernel values show good sensitivities in the lowermost troposphere, indicating the good sensitivity for lower tropospheric HDO/H2O variabilities. This is confirmed by the right panel of Fig. 6, which depicts a time series of the "*sens(LT)*" parameters obtained for the Karlsruhe site for the different data products. For both MUSICA NDACC and TCCON the values of the parameter "*sens(LT)*" are very close to 1.0 (black and red dots, respectively), meaning that both datasets are well sensitive to lower tropospheric HDO/H2O variabilities. For MUSICA IASI (blue dots) "sens(LT)" can be significantly smaller than 1.0, meaning that the MUISCA IASI product has rather limited sensitivity on lower tropospheric HDO/H2O ratios. We have also calculated the "sens(LT)" values for TROPOMI (green dots). Here the values are again close to 1.0, although not that close as for MUSICA NDACC and TCCON. We conclude that the TROPOMI total column integrated HDO/H2O ratio product has a reasonable sensitivity on lower tropospheric HDO/H2O ratios. In this regard, it can well complement the respective MUSICA IASI satellite product, which has only limited sensitivity on the lower tropospheric HDO/H2O ratios.

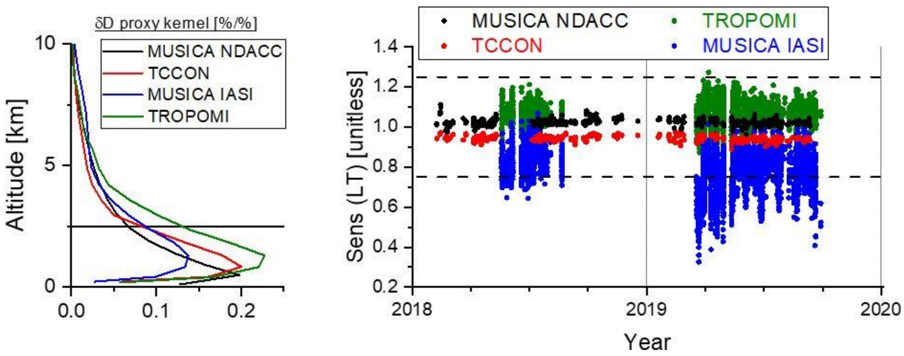

Figure 6: Left: example of δD proxy column averaging kernels for the reference data MUSICA NDACC, TCCON, MUSICA IASI) obtained at Karlsruhe in May 2018. Right: time series of lower tropospheric sensitivity (*sens(LT)* parameter) calculated for the reference data sets at the location of Karlsruhe.

5.2 Satellite measurements

Currently there are two sensors that can be used for the measurement of tropospheric water vapour isotopologues from space. This is IASI (Infrared Atmospheric Sounding Interferometer), which measures high resolution (0.5 cm^{-1}) thermal nadir spectra (645-2760 cm^{-1}) aboard three different Metop satellites and GOSAT (Greenhouse gas Observing SATellite), which measures surface reflected short-waver infrared solar spectra (in the region from 700-1320 cm^{-1} and with a spectral resolution of 0.2 cm^{-1}). First IASI water vapour isotoplogue ratio retrievals have been presented by [RD9] and [RD10]. First GOSAT water isotopologue retrievals have been presented by [RD11] and [RD12].

5.2.1 MUSICA IASI

In addition to the MUSICA NDACC retrieval, the MUSICA project developed a water vapour isotopologue retrieval for IASI spectra. It uses the broad spectral window between 1190 and 1400 cm^{-1} and simultaneously estimates the vertical profiles of H2O, HDO/H2O, N2O, CH4, HNO3 and atmospheric temperature retrieval ([RD1] [RD15], [RD16]). Figure 7 depicts typical averaging kernels of the retrieved atmospheric composition state. Concerning water vapour and its isotopologues, we typically get a DOFS of 5-6 for H2O and 1-2 for HDO/H2O (or δD). We can also observe that these thermal nadir products generally lack sensitivity close to ground (due to reduced thermal contrast). This issue has to be analysed in more detail if we want to use the MUSICA IASI data as validation reference for a total column integrated HDO/H2O TROPOMI product (recall most water resides in the lowermost troposphere, i.e. close to the surface). As for the MUSICA NDACC and the TCCON product we calculate HDO/H2O (or δD) proxy column averaging kernels. A typical proxy kernel is shown as blue line in the left panel of Fig. 6. The lack of lower tropospheric sensitivity of the MUSICA IASI HDO/H2O product is documented by the right panel of Fig. 6. The blue dots represent the parameter "*sens(LT)*" for the MUSICA IASI δD proxy product. It is generally significantly below 1.0. Concerning the Karlsruhe site (whose time series is shown in Fig. 6) the parameter "*sens(LT)*" is in particularly low in winter. Nevertheless, during summer many observations with *sens(LT)>0.75* also exist.

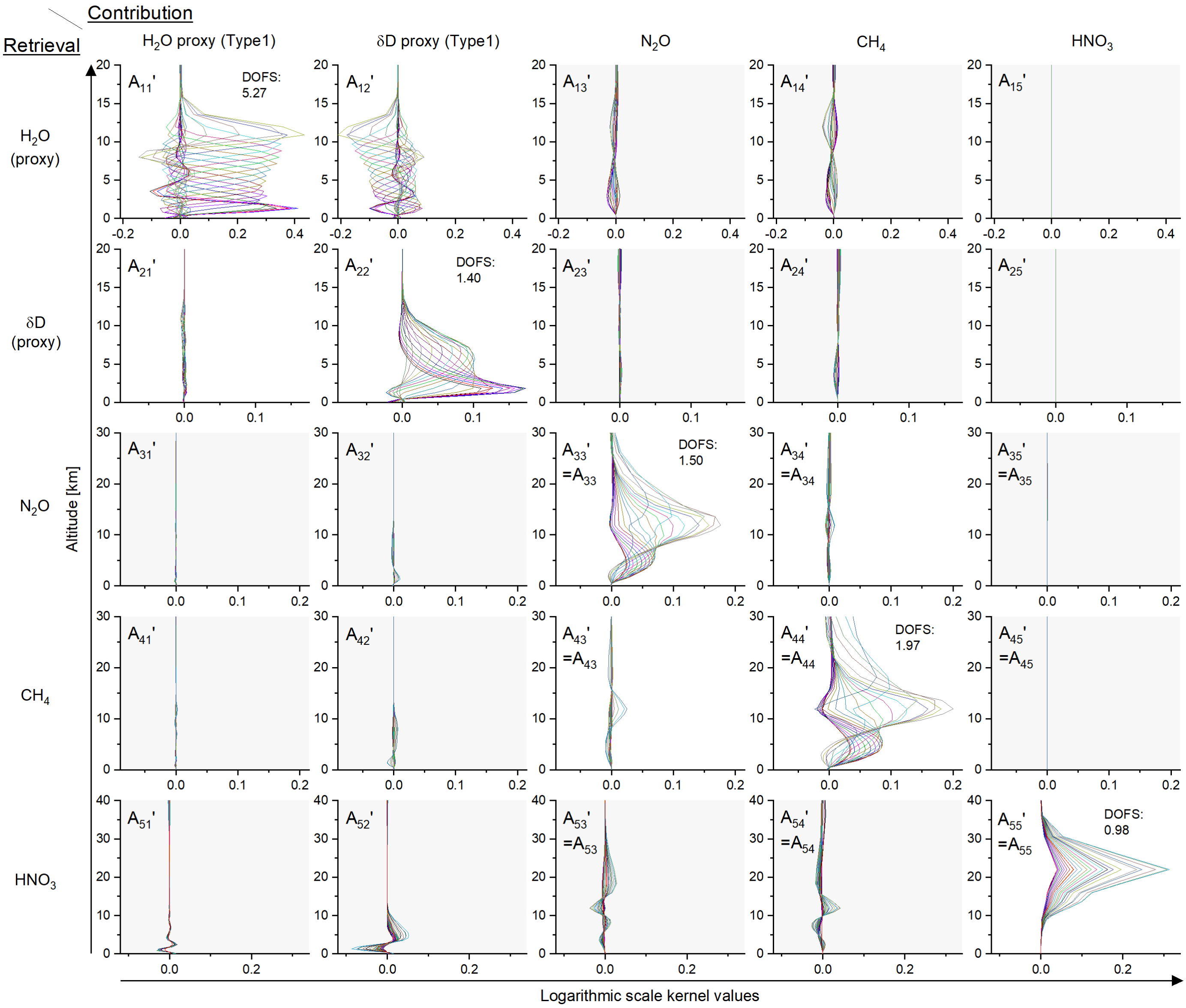

Figure 7: Examples of typical MUSICA IASI averaging kernels. Relevant for this study are the averaging kernel plots representing the {H2O, δD} proxy state, i.e. the blocks: A11', A12', A21' and A22'.

In Fig. 8 the MUSICA IASI total column integrated HDO/H2O ratios are compared to the respective TCCON data (collocation within 50 km and 4 hours). The left panel shows the comparisons for all coincidences fulfilling the collocation criteria. The collour code highlights the "*sens(LT)*" value. For *sens(LT)*>0.75 we observe a good agreement between MUSICA IASI and TCCON. This is resumed in the right panel of Fig. 8 showing only the coincidences with MUSICA IASI data, for which *sens(LT)*>0.75. With this analysis we can demonstrate that the MUSICA IASI data are reasonably sensitive to the lower troposphere (where most water resides), and that the column integrated MUSICA IASI HDO/H2O ratio compares reasonably well to the TCCON total column integrated HDO/H2O ratios in this case. The scatter achieved is only 15 ‰, which is very similar to the scatter observed for the comparison between MUSICA NDACC and TCCON. Nevertheless, there is a significant negative bias of the MUSICA IASI data with respect to the TCCON data. This makes sense, and indicates that the MUSICA IASI data are more influenced by free tropospheric ratios (which are generally lower than the boundary layer ratios) than the TCCON data (please recall that the *sens(LT)* value of the used MUSICA IASI data is still significantly below 1.0).

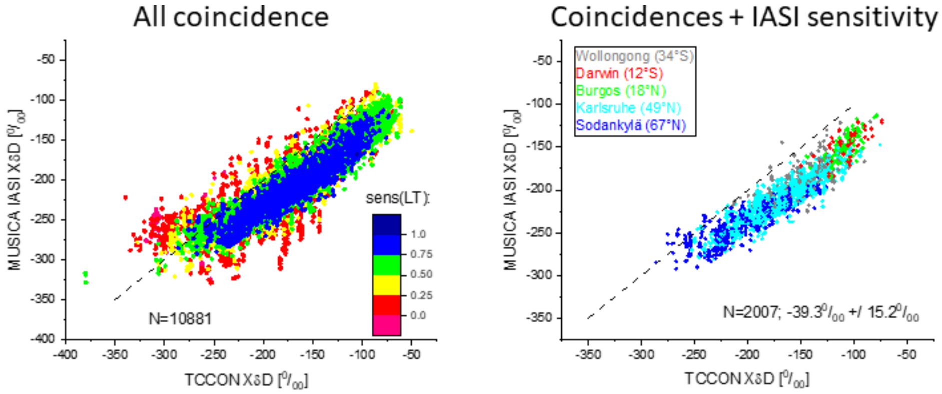

Figure 8: Comparison of MUSICA IASI and TCCON XδD. Left: all data points for which the temporal and spatial coincidence criteria are fulfilled. Right: only data points where the MUSICA IASI product is well sensitive to lower tropospheric XδD.

IASI offers a very dense coverage. The three currently orbiting IASI instruments measure almost 4 million individual nadir spectra per 24 hours and offer six times per day quasi-global coverage. The MUSICA IASI product has been generated on global scale for the whole time-period between Oct. 2014 and Dec. 2019 (retrievals for 2020 are currently processed). The processing is limited to cloud free scenes. Currently the MUSICA IASI data set comprehends trace gas profile generated from more than 1 billion individual IASI spectra.

5.2.2 GOSAT

GOSAT data are not used for this validation study, since GOSAT data provide only sparse data coverage.

5.3 Field campaigns

The objective of the WaVIL (Water Vapor Isotope Lidar) project is the development of a compact, transportable differential absorption lidar for measuring the concentration of water vapor and its isotope HDO with high vertical and temporal resolution in the lower troposphere with an unprecedented accuracy [RD14]. The WaVIL data was not used for this validation study due to the very limited number of collocated TROPOMI observations and its limited vertical representativeness (only up to 3km above ground). Nevertheless, we plan to use it at a later stage of the project in the context of the impact study.

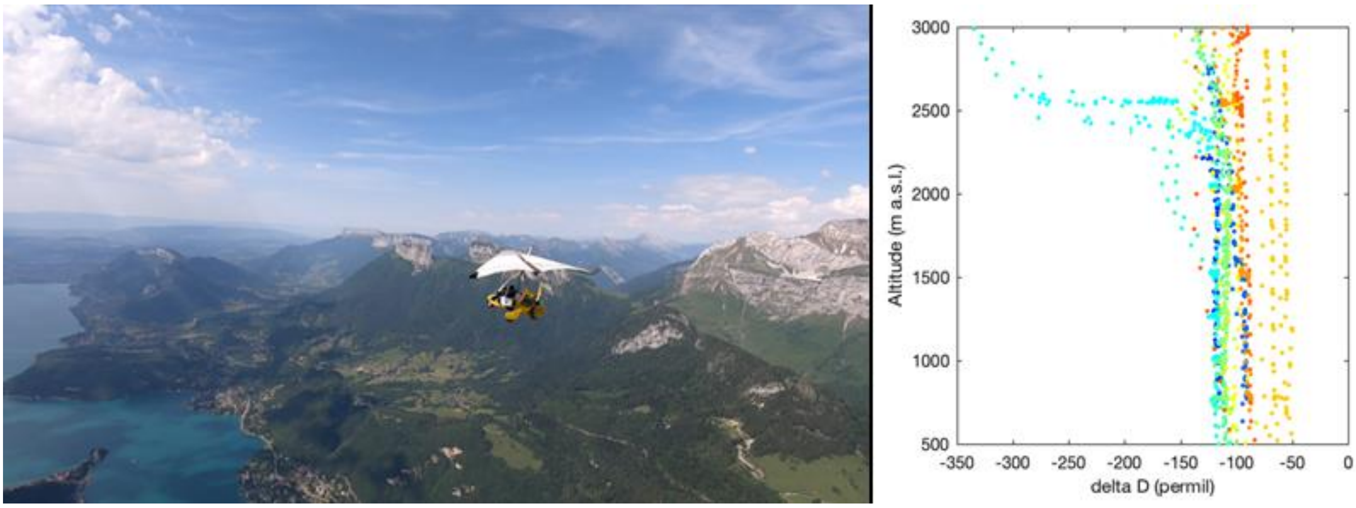

Figure 9: The WAVIL 2019 field campaign. Left: Example of a WAVIL measurement activity; Right: Overview of all measured WAVIL δD profiles.

5.4 Discussion on the vertical representativeness of reference data

We expect a total column HDO/H2O ratio product to mainly reflect the ratio of HDO/H2O in the lower troposphere (where most atmospheric water vapour resides). For this reason, we require the validation references to also have sensitivity to lower tropospheric HDO/H2O. As discussed in Sections 5.1 and 5.2 this condition is achieved by the NDACC MUSICA and TCCON ground-based remote sensing references, but not comprehensively by the MUSICA IASI reference data, which have best sensitivity in the free troposphere.

We recommend to use the references of the ground-based networks as primary references for the validation of TROPOMI total column XHDO/XH2O ratios. Data from these ground-based networks are well representative for total column variabilities. Nevertheless, the MUSICA IASI reference data are a valuable secondary validation reference for supporting the validation results obtained by the primary reference data.

5.5 Modelling support

The isotope-enabled limited-area models COSMOiso will be used for high-resolution isotope simulations in the context of the impact study. No model data are used for this validation study.

6 Validation approach

For the validation of the TROPOMI XHDO/XH2O (or XδD) column integrated data we use the ground-based direct sun measurements made within the NDACC (MUSICA NDACC) and TCCON networks at seven different stations as the primary reference. These stations represent different climate zones (polar, mid-latitudinal, subtropical, and tropical) of the northern and southern hemisphere. Comparisons to these primary references are used for a detailed investigation of the TROPOMI quality (bias, dispersion, dependence on solar zenith angle, albedo, and total atmospheric water vapour content). The MUSICA IASI XδD product is our secondary validation reference and nicely confirms the validation results obtained from the ground-based references. The availability of the ground-based MUSICA NDACC and TCCON reference data is listed in Table 3.
The period used for this validation study extends from May 2018 to August 2020. We are investigating L2 V0.9.5, the latest TROPOMI XδD product version developed at the University of Leicester. We call this in the following the prototype product. The only difference of this prototype product to the V1.0 product (which will be made available to the user in the very near future) is the quality assurance parameter. For the prototype we have the classes qa=0 (not for scientific use) and qa=1 (data of good quality). For the V1.0 product there wil be in addition qa=2 (data of best quality). This validation report will has contributed to identify the conditions for qa=2, see ATBDv1.4 Chapter 7.4 [AD2].
For this validation report we assess the quality of the TROPOMI prototype data that fulfill the qa=1 criteria.

Table 3: Ground-based station, FTIR data availability, Code used for retrieving the FIR data and responsible team.

FTIR Station	Ground-based data availability	Code	Team
Karlsruhe	Apr 2010 – Apr 2021	PROFFIT	KIT-ASF
Kiruna	Mar 1996 – Apr 2021	PROFFIT	KIT-ASF; IRF Kiruna
Sodankylä	May 2009 - Oct 2020	GFIT	FMI
Burgos	Mar 2017 - Mar 2020	GFIT	NIES Tsukuba; U. of Wollongong
Karlsruhe	Apr 2010 – Oct 2020	GFIT	KIT-ASF
Darwin	Aug 2005 – Mar 2020	GFIT	U. of Wollongong
Wollongong	Jun 2008 – Mar 2020	GFIT	U. of Wollongong

6.1 Spatial and temporal collocation with reference observations

As collocation criteria we use 50 km (horizontal collocation of satellite ground pixel and ground-based station) and Δt = 3 h. The spatial and temporal collocation criteria are consistent among one another assuming a mean wind velocity of about 5 m/s. Lowering those criteria did not lead to a significant change in bias or dispersion.

Figure 10 shows the available TROPOMI data in a radius of 50 km and Δt = 3 h around the NDACC and TCCON stations with quality assurance criteria (qa = 1). Furthermore, in order to use comparable XδD total column values, TROPOMI pixels were removed if they differ in surface height by more than 250 m from the corresponding station height. When applying the aforementioned spatial and temporal collocation filters the data as depicted in Fig. 10 remain. Depending on the location there are currently between 14,567 and 61,465 pixel or 88 and 241 days, respectively with available coincident TROPOMI XδD data for the period used in this from May 2018 to August 2020. Throughout the rest of this document, the qa=1 filtered data for collocation within 3 hours and 50 km will be referred to as standard filtered.

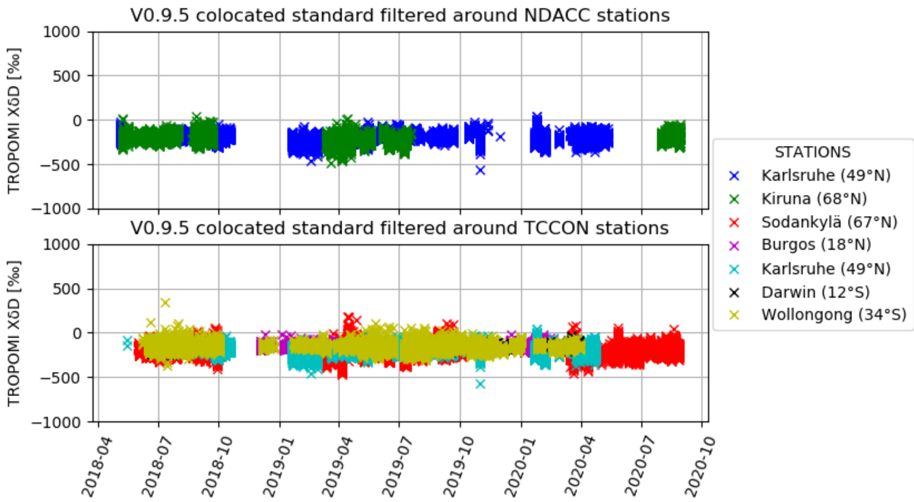

Figure 10: Coincident TROPOMI XδD L2 V0.9.5 data within a collocation radius of 50 km and and Δt=3 h around NDACC (upper panel) and TCCON sites (lower panel) for qa=1 filtered data.

6.2 Recommendations for TROPOMI data usage

The associated Algorithm Theoretical Basis Document (ATBD) [AD2] describes the Leicester TROPOMI XδD product. For the L2 version V0.9.5 filtering for cloud cover is already included in the dataset, hence no additional cloud filtering is needed. For the L2 product version V1.0 the data user will be able to choose between a quality assurance (qa) value of 0, 1 and 2. This qa-value can be easily applied and identifies data of different quality levels. Data with a qa-value of 1 passed the 1st stage quality filter and are suitable for scientific use (this is the filter for which we perform the validations study). In addition, we show in the following sections that setting limits with respect to XH2O and/or albedo values could further increase the data quality significantly. In this context, the outcome of this validation study will be used for defining the conditions for qa=2 (best quality data). Furthermore, in Sect. 6.10 we recommend a method for correcting the observed bias in the TROPOMI data.

6.3 Status of validation

This validation report refers to the University of Leicester TROPOMI L2 XδD product (Version 0.9.5, a prototype data product). The study investigates the prototype data product for the time period from May 2018 to August 2020 as shown in Fig. 10 and Table 3. Data quality is investigated after applying the aforementioned collocation- and qa=1 filters. The validation is performed with the difference in XδD, for which we use the difference of the TROPOMI XδD product with the collocated reference XδD product.

6.4 Dependence on influence quantities

We have investigated whether the quality of the XδD TROPOMI product depends on different influence parameters. Thereby, we found significant dependence on three parameters: the solar zenith angle, the retrieved surface albedo, and the retrieved total water vapour amount.

The dependence on the solar zenith angle is depicted in Fig. 11. As expected, we observed that for the polar sites Kiruna and Sodankylä some pronounced outliers remain for solar zenith angles above 55° when using the recommended qa=1.

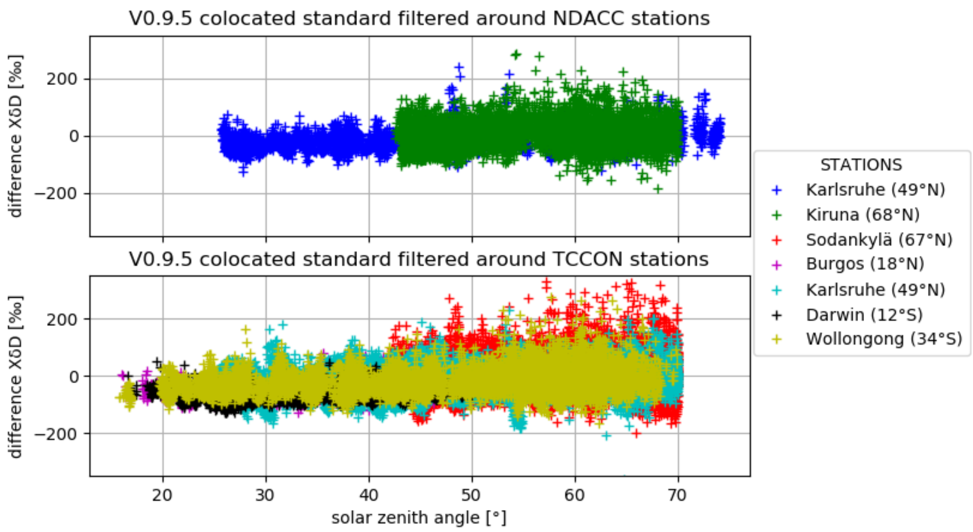

Figure 11: Dependence of the difference between TROPOMI and ground-based FTIR XδD [‰] on the satellite solar zenith angle (SZA) for single pixel measurements.

Figure 11 shows all individual data points (in total more than 200,000), and a few outliers might affect the visual impression. In order to present the main statistical characteristics of the data we calculate the mean difference and the standard deviation around the mean difference for data binned according to the solar zenith angle. For each station we determine 10 solar zenith angle bins. We require that each bin represents 10% of the data. The mean differences in XδD for the different stations and solar zenith angle bins are depicted on the left panel of Fig. 12. On the right panel of Fig. 12 we show the respective standard deviations. Both mean difference and standard deviations seem to change systematically with increasing solar zenith angle. Higher SZA mean lower signal to noise ratio in the measured radiances, which in turn causes larger uncertainties in the data product. In this context the observed increase in standard deviation is within expectations and in line with Fig. 19.The highest standard deviation values are found for the polar sites Kiruna and Sodankylä for solar zenith angles above 55°.

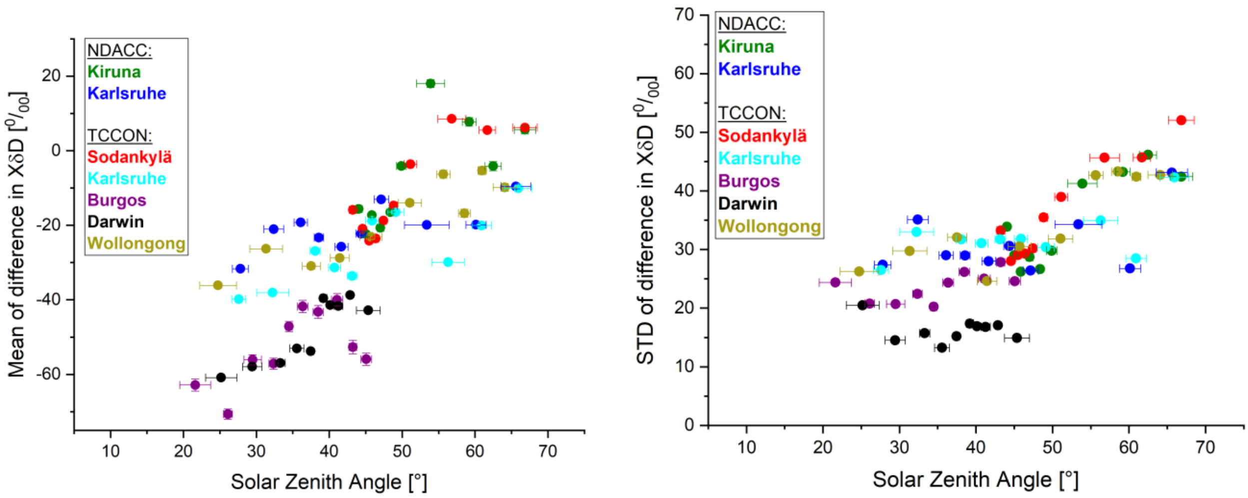

Figure 12: Same as Fig. 11 but for the XδD differences binned according to the solar zenith angle (each bin contains 10% of the data from a certain station). Left: mean difference for each bin; Right: standard deviation for each bin.

Figure 13 shows the dependence of the difference of the TROPOMI product with respect to the ground-based references on the retrieved short-wave infrared (SWIR) surface albedo. After applying the qa-filter we still observe an increased scatter for albedo values below about 0.07, especially at the high latitude sites Sodankylä (red) and Kiruna (green).

Figure 14 shows the statistics of the dependency of the differences in XδD on the SWIR albedo. We observe a clear change of the mean difference with the SWIR albedo (left panel of Fig. 14). For low albedo the mean difference seems to approximate 0‰. However, we also observe clearly different mean difference for the two tropical sites Burgos and Darwin. Concerning the standard deviation, the dependency on the albedo is clearly evident. For high albedo values the standard deviation is generally below 30‰, whereas it can reach 50‰ for very low albedos. Lower albedo mean lower signal to noise ratio in the measured radiances, which in turn causes larger uncertainties in the data product. In this context the observed increase in standard deviation is within expectations and in line with Fig. 19. At the tropical sites Burgos and Darwin we observe a significantly lower standard deviation than at the other stations.

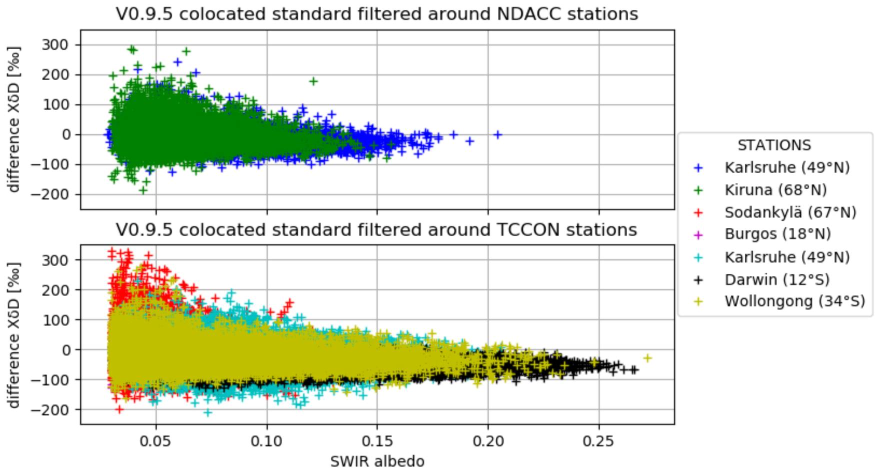

Figure 13: Dependence of the difference between TROPOMI and ground-based FTIR XδD [‰] on the satellite SWIR albedo for single pixel measurements.

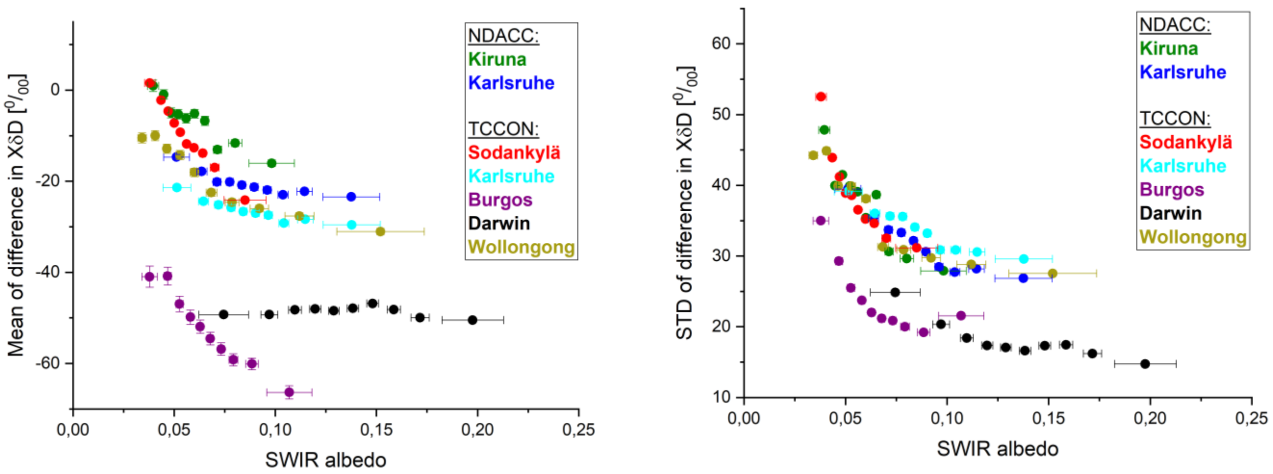

Figure 14: Same as Fig. 13, but for the XδD differences binned according to the SWIR albedo (each bin contains 10% of the data from a certain station). Left: mean difference for each bin; Right: standard deviation for each bin.

Figure 15 reveals a clear dependency of the data quality with the total atmospheric water vapour content and Fig. 16 provides detailed insight into this dependency. The largest standard deviations between the TROPOMI product and the references occur for XH2O values of about smaller than 1500 ppmv (right panel of Fig. 16). For high XH2O values the standard deviation is much smaller (below 30‰ for XH2O above 3500 ppmv), which explains the distinct behaviour for the tropical sites Burgos and Darwin as observed in the right panel of Fig. 14. The left panel of Fig. 16 reveals a systematic dependency of the mean difference on the total atmosphere water vapour content (XH2O). The mean difference is most negative at large XH2O values and is close to zero for a dry atmosphere (XH2O below 1500). The distinct behaviour for the tropical sites Burgos and Darwin cannot be traced back to a specific factor yet and requires further investigation.

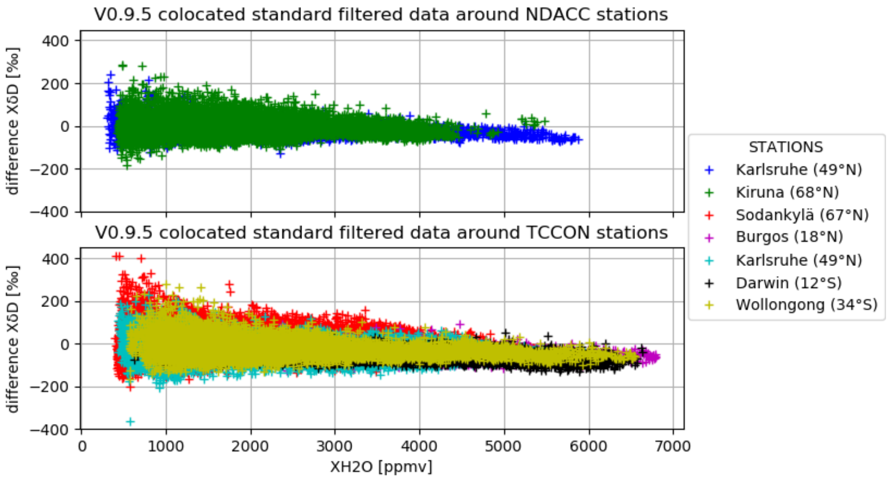

Figure 15: Dependence of the difference between TROPOMI and ground-based FTIR XδD data in ‰ on the satellite XH2O in ppmv (total column) for single pixel measurements.

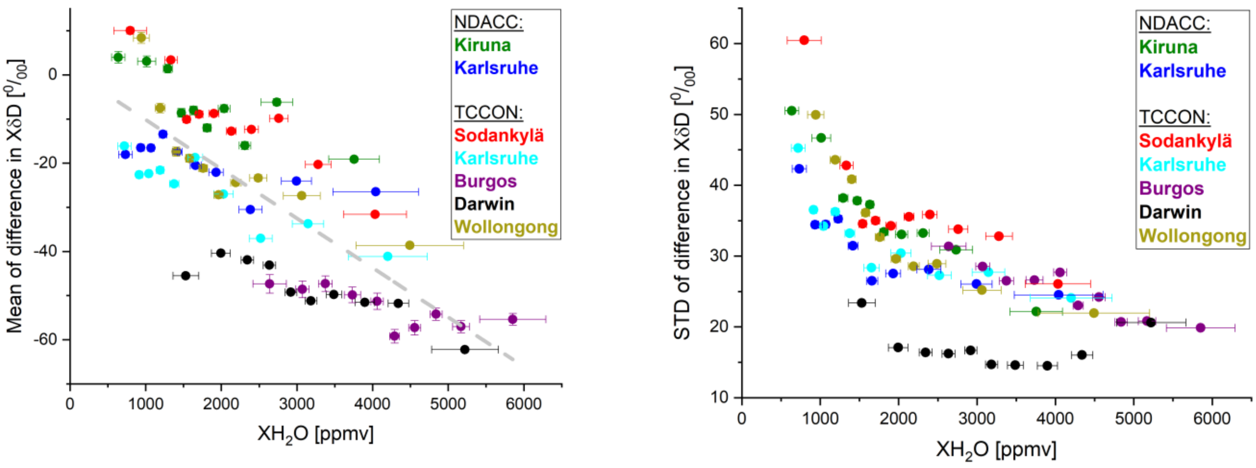

Figure 16: Same as Fig. 15, but for the XδD differences binned according to XH2O (each bin contains 10% of the data from a certain station). Left: mean difference for each bin; Right: standard deviation for each bin. The grey dashed line shows the regression line obtained from a linear least squares fit.

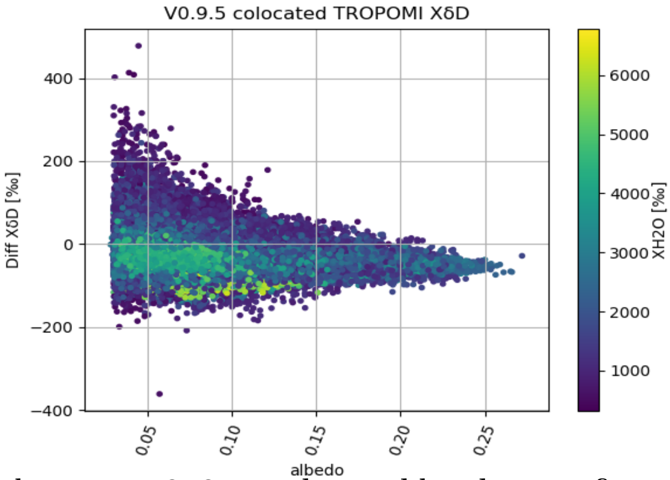

Figure 17: Difference between TROPOMI and ground-based FTIR XδD against SWIR albedo as dependent on atmospheric water vapour content.

Albedo and XH2O seem to most strongly affect the TROPOMI XδD data quality. Figure 17 shows that the highest differences between TROPOMI and FTIR XδD data occur when XH2O and the albedo are simultaneously low.

When plotting the difference between TROPOMI and ground-based FTIR XδD against the XδD TROPOMI estimated noise error we observe the greatest differences when the estimated errors are large (Fig. 18). However, the estimated XδD TROPOMI noise error represents the 1-σ error and can be well evaluated by the statistical analysis of Fig. 19. The right panel of Fig. 19 depicts the standard deviations of the difference for the data binned with respect to the estimated XδD TROPOMI noise error. We see a clear correlation between the estimated error and the standard deviation. This means that a part of the differences between the TROPOMI and the reference XδD data are well understood, as they are expected from the noise error. However, the estimated XδD TROPOMI noise error is generally smaller than the standard deviation between the observations. This is also expectable, because the reference data are also subject to uncertainties. Furthermore, an imperfect collocation between the two remote sensing systems (TROPOMI and ground-based FTIRs) and small difference in the vertical sensitivities (the averaging kernels, see discussion in the context of Fig. 6) affect the standard deviation.

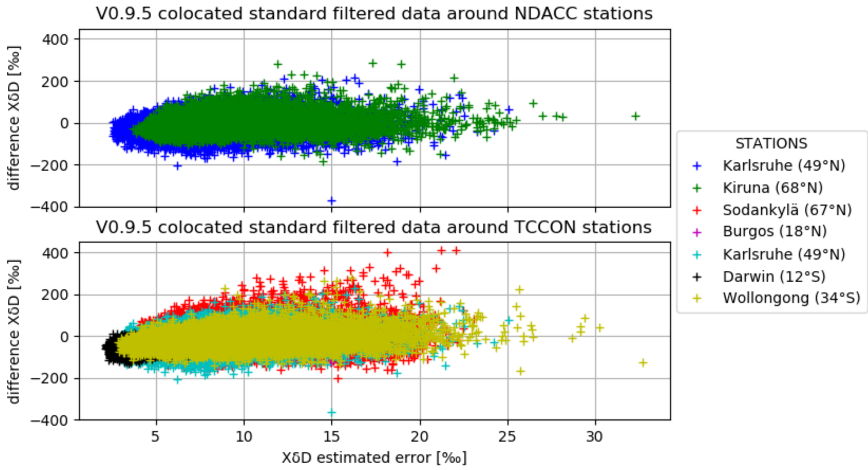

Figure 18: Difference between TROPOMI XδD and ground-based FTIR XδD against TROPOMI XδD estimated error for single pixel measurements.

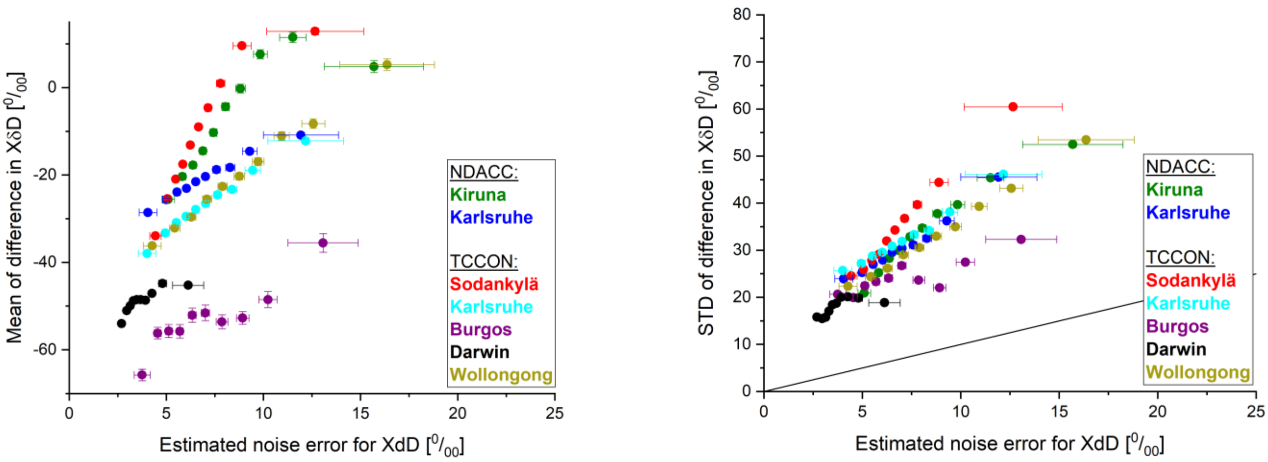

Figure 19: Same as Fig. 18, but for the XδD differences binned according to noise error estimated for the TROPOMI XδD product (each bin contains 10% of the data from a certain station). Left: mean difference for each bin; Right: standard deviation for each bin. The black line in the right panel shows the one-to-one diagonal.

6.5 Bias

We define the bias as the mean difference between the TROPOMI product and the reference data (TROPOMI - Reference). For comparison to the MUSICA NDACC data we found a negative bias of about -17.3 ‰. Table 4 provides more detailed information on the calculated bias values. For the TCCON data we determine a

negative mean bias of about -21.0 ‰. The mean biases at Burgos and Darwin are distinctly higher than at the other stations with -52.7 ‰ and -48.7 ‰, respectively.
We calculate the statistical uncertainty of the bias values as the standard error of the mean difference, which we define as the 1-σ standard deviation around the mean divided by the square root of the number of data used to calculate the mean bias. In summary, we get a bias and uncertainty of the bias for MUSICA NDACC data of -17.3 +/- 0.1 ‰ and using TCCON of -21.0 +/- 0.1 ‰.
Figure 20 depicts the bias for the different stations. We observe strong differences between the bias at high and low latitude stations. The lowest bias is observed at the high latitude station Kiruna with -6.9 +/- 0.3 ‰, the highest bias at the tropical site Burgos with -52.7 +/- 0.5 ‰ yet. This observation is due to the dependence of the mean difference in TROPOMI's XδD on the XH2O values, which are much larger in the tropics than at high latitudes.

Table 4: Number of collocated single pixel measurements and daily means (given in brackets) for standard filtered data. Mean bias, uncertainty of the mean bias and standard deviation for collocated data.

FTIR Station	# of collocated standard filtered pixel (days)	Mean bias [‰]	Uncertainty of mean bias [‰]	StdD of difference [‰]
Karlsruhe	46,950 (164)	-20.6 (-16.7)	0.2 (1.9)	31.9 (24.2)
Kiruna	14,567 (88)	-6.9 (6.6)	0.3 (3.0)	37.8 (27.7)
NDACC	**61,517 (252)**	**-17.3 (-8.6)**	**0.1 (1.8)**	**33.9 (27.8)**
Sodankylä	61,265 (241)	-10.2 (12.2)	0.2 (2.5)	39.6 (38.6)
Burgos	2,269 (85)	-52.7 (-41.1)	0.5 (3.2)	25.4 (29.8)
Karlsruhe	41,856 (170)	-26.5 (-21.9)	0.2 (2.3)	33.8 (29.4)
Darwin	22,437 (134)	-48.7 (-54.0)	0.1 (1.6)	18.3 (18.9)
Wollongong	15,987 (224)	-19.7 (-13.4)	0.3 (2.1)	36.8 (31.5)
TCCON	**143,814 (854)**	**-21.0 (-14.8)**	**0.1 (1.2)**	**36.5 (36.0)**
ALL SITES	**205,331 (1,106)**	**-21.1 (-15.1)**	**0.1 (1.1)**	**36.5 (36.8)**

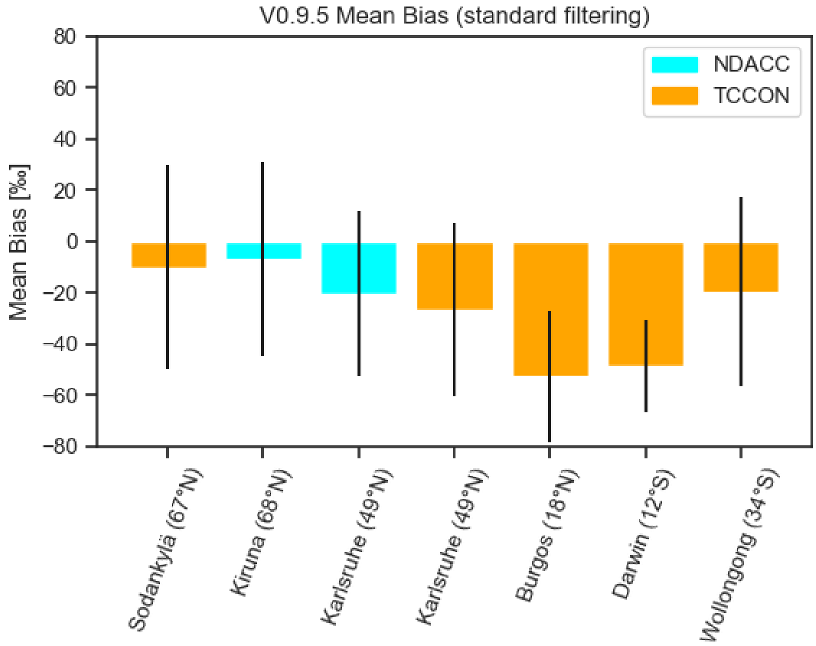

Figure 20: Mean difference between TROPOMI XδD and the references of MUSICA NDACC and TCCON at the 7 stations within the period between May 2018 and August 2020 arranged from highest to lowest latitude. The black lines indicate the standard deviation around the mean differences.

6.6 Dispersion

We define the dispersion between the TROPOMI product and the reference data as the 1-σ standard deviation of their difference (TROPOMI - Reference). For the comparison to MUSICA NDACC data this dispersion is 33.9 ‰ (Table 4). From the comparison to the TCCON data we get a dispersion of 36.5 ‰. The dispersion values are lowest for the tropical stations, where XH2O is highest, i.e. this observation is in agreement with Figs.15 and 16.

6.7 Short-scale and short-term variability

Figure 21 depicts the correlation for single pixels and daily means of standard filtered data. It shows that the TROPOMI data are in reasonable overall agreement with correlative ground-based measurements from both monitoring networks. We provide correlation plots based on all individual data points (N=205331, left) and on daily mean data (N=1106, right). When averaging over several individual pixels (i.e. daily mean data), the random uncertainty of TROPOMI data is reduced, resulting in a slightly better overall correlation from 0.72 to 0.74.

The left side of Fig. 21 depicts the correlation between the ground-based FTIR data and the TROPOMI data for the different sites for all individual observations (N=205331), i.e. it represents very short-scale and short –term signals. Despite some outliers we get a reasonable correlation (r=0.72) for this very short-term and short-scale signals.

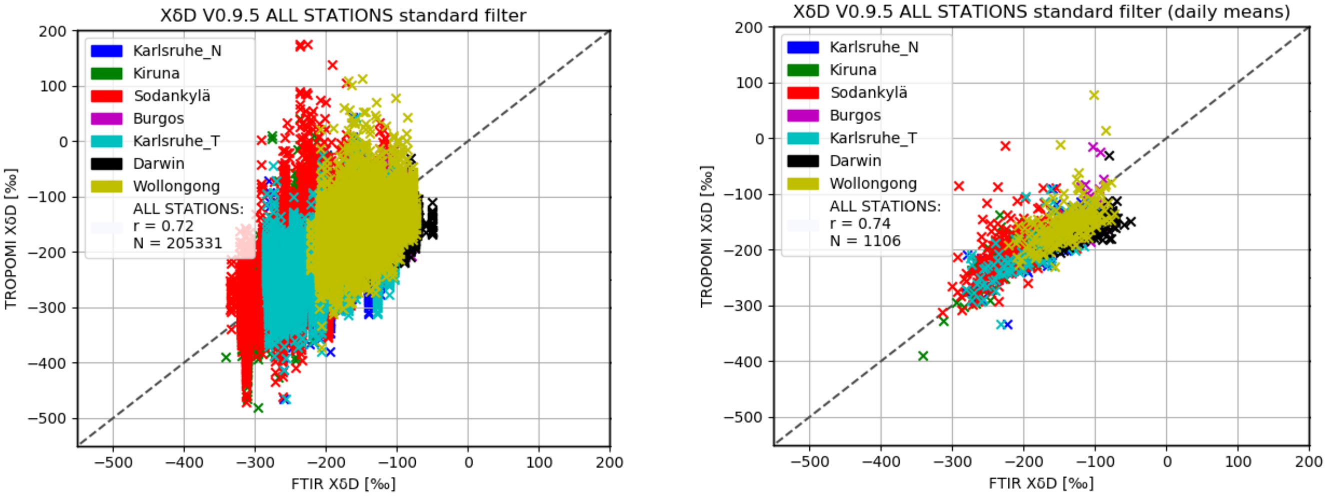

Figure 21: Correlation of FTIR and TROPOMI XδD single pixel measurements (left) and daily means (right) with standard filtering around the FTIR stations.

6.8 Geographical pattern

Each bar in the mosaic plot of Fig. 22 represents the weekly average of the XδD difference (weekly biases) of standard filtered TROPOMI and FTIR XδD data. We observe biases at the higher latitude stations Sodankylä and Kiruna that fluctuate from positive and negative XδD differences, with the most positive values in the middle of summer. This is in line with the high standard deviation of the relative difference seen for the high latitude stations (Table 4). No TROPOMI and ground-based FTIR data is available during the polar night for these stations, as no direct sun or surface reflected solar measurements can be taken during this period. High negative weekly biases are observed at the Darwin site with the most negative values close to -80 ‰. At mid and low latitude stations, we only observe weak fluctuation from positive to negative XδD differences, which documents that the TROPOMI short-term variability is in good agreement to the variability as observed by the FTIR systems.

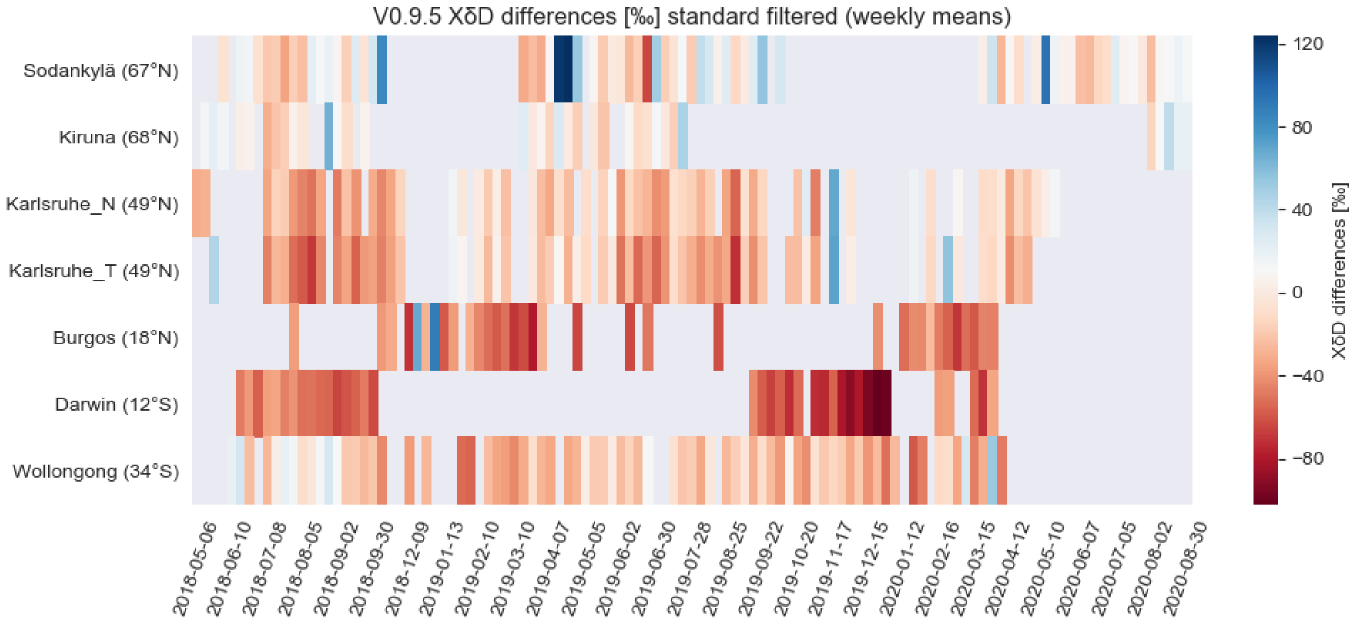

Figure 22: Weekly relative biases between collocated TROPOMI and FTIR XδD data. Spatial collocation with r = 50 km and Δt = 3 h around the satellite overpass was used. The stations are sorted from highest to lowest latitude. Karlsruhe_N (49°N) is the NDACC MUSICA and Karlsruhe_T (49°N) is the TCCON FTIR site. Time periods were no collocated data is available are indicated in grey (e.g. the period from 2018-09-30 to 2019-08-25 at the Darwin site).

Figures 20-22 allow some discussion on geographical patterns. For the polar stations Kiruna and Sodankylä, we observe only small biases, whereas at the tropical sites of Burgos and Darwin the XδD bias has high negative values. On the other hand, at these tropical sites the dispersion is lower than at mid-latitudinal or polar sites (Table 4). At polar sites the dispersion is relatively high, because there atmospheric water vapour content and albedo are generally low and SZA tends to be high. These are the most difficult conditions for high quality observations (see Figs. 11-17).

6.9 Comparison to MUSICA IASI references

In this section we complement the validation exercise with comparison to MUSICA IASI data. As shown in Sect. 5.2. the MUSICA IASI data are mostly sensitive to the free troposphere. They contain only limited information on the lower tropospheric HDO/H2O ratios and thus on the total column integrated ratios. We can filter MUSICA IASI data with a reasonable sensitivity on the lower troposphere by using data only for *sens(LT)*>0.75 (see details on the "*sens(LT)*" value in Section 4.1 and the discussion in the context of Fig. 6). We use the MUSICA IASI data here as secondary validation reference in order to show the consistency (and complementary) between the MUSICA IASI and the TROPOMI product. For this reason we limit the comparison to a subset of arbitrarily selected but fully representative data points (fully representative for different seasons and geographical locations). We compare about 5000 individual TROPOMI observations made in the vicinity of a polar site (Sodankylä/Kiruna), a mid-latitudinal site (Karlsruhe), a subtropical site

(Wollongong) and two tropical sites (Burgos and Darwin). For collocation with IASI we use the spatial collocation criteria of 50 km (same as in the previous sections), with a slightly relaxed temporal collocation criterion from 3 to 4 hours. We only use TROPOMI data that are standard filtered.
Figure 23 depicts the comparison between the TROPOMI and the MUSICA IASI data. The data pairs are collour-coded according to the locations (here Sodankylä means all data measured in northern Scandinavia around Sodankylä or Kiruna). The left panel shows as an overview the correlation plot. We find no significant bias (mean difference) between both data set and a dispersion (1-σ standard deviation around the mean difference) of about 27 ‰. While the dispersion is in the range of the dispersion found for the comparison between the TROPOMI and the ground-based network data, the bias is significantly smaller and by about 25 ‰ more positive. However, we have to consider that the IASI data, although filtered for sensitivity with respect to the lower troposphere, tend to be more sensitive in the free troposphere than in the lower troposphere. This has been shown in the right panel of Fig. 6: while the filtered MUSICA IASI data are for an *sens(LT)* value between 0.75 and 0.85, the TROPOMI data better represent the lower troposphere (*sens(LT)* values of mostly between 0.9 and 1.1).
Similar to the comparison to the ground-based network data we found a dependence of the dispersion on the total atmospheric water vapour content (XH2O, right panel of Fig. 23). The dispersion between TROPOMI and MUSICA IASI is in particularly large for XH2O values that are smaller than 2000 ppmv.

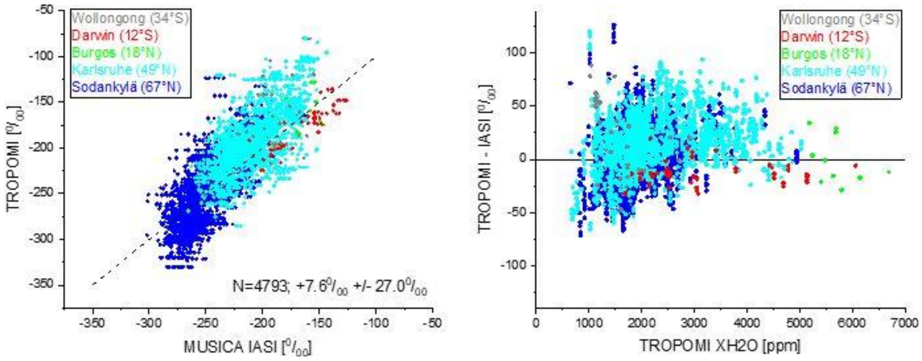

Figure 23: Comparison of TROPOMI and MUSICA IASI XδD with the latter fulfilling the lower tropospheric sensitivity criterion. Data belonging to different locations are marked by different collours. Left: correlation plot; Right: Dependence of the TROPOMI-IASI difference on TROPOMI XH2O.

The comparison between MUSICA IASI and TROPOMI suggests a good consistency between both data sets. The fact that TROPOMI shows a negative bias with respect to the ground-based FTIR references (see e.g. Fig 20), but a slightly positive bias with respect to the MUSICA IASI data, is in agreement with the different sensitivities of the TROPOMI and the ground-based FTIR data sets, on the one hand, and the MUSICA IASI data set, on the other hand. While the TROPOMI and the ground-based FTIR data sets are well sensitive to the lower troposphere and have similar sensitivities for the rest of the troposphere, the MUSICA AISI data set is mainly sensitive in the free troposphere and has only a limited lower tropospheric sensitivity. Due to their different sensitivities the TROPOMI and the MUSICA IASI HDO/H2O data well complement each another and we foresee interesting potential for a synergetic use of both data sets.

6.10 Bias Correction

The investigation of the prototype TROPOMI XδD data product reveals a bias with respect to the ground-based FTIR references that depends on the atmospheric water vapour content. This dependency becomes most evident in the left panel of Fig. 16. A linear least squares fit of these dependency yields the following coefficients: slope: -0.0112 ‰; intercept: +1.03 ‰; correlation coefficient: -0.77. The respective regression line is depicted as dashed grey line in the figure.

We can use this apparent dependency for correcting the bias by adding the following value to the original TROPOMI XδD data: -(-0.0112‰*XH2O + 1.03‰). Here XH2O is the XH2O value given in ppmv and retrieved for the TROPOMI observation. This regression line is depicted in Fig. 16 (dashed line in the left panel). It is determined by a linear least squares fit using all the binned data points as shown in this figure (here we bin the data of individual pixels, not the daily mean data).

Table 5 and Fig. 24 resume the bias and dispersion we achieve after applying the bias correction (Table 5 and Fig. 24 show the same as Table 4 and Fig. 20, but for the bias corrected data instead of the original data). As aforementioned the binning in Fig. 16 is made using individual pixel data (not daily mean data). For this reason the correction is very effective in removing the mean difference (the bias) calculated from the individual pixel data and less effective in removing the bias calculated from daily mean data (see Table 5, mean bias for statistics with all individual pixels is about +2.5‰, whereas the mean bias for statistics with daily mean data is about +13‰).

Table 5: Number of collocated single pixel measurements and days (given in brackets) for bias corrected data. Mean bias (bias corrected), uncertainty of the mean bias (bias corrected and standard deviation (bias corrected for collocated data.

FTIR Station	# of collocated standard filtered pixel (days)	Mean $bias_{corr}$ [‰]	Uncertainty of mean $bias_{corr}$ [‰]	$StdD_{corr}$ of difference [‰]
Karlsruhe	46,950 (164)	-0.9 (8.5)	0.0 (0.7)	32.5 (23.2)
Kiruna	14,567 (88)	13.0 (24.6)	0.1 (2.6)	37.5 (28.1)
NDACC	**61,517 (252)**	**2.4 (14.1)**	**0.0 (0.9)**	**34.3 (26.1)**
Sodankylä	61,265 (241)	13.4 (32.7)	0.1 (2.1)	38.4 (35.6)
Burgos	2,269 (85)	-7.2 (7.6)	0.2 (0.8)	26.3 (28.9)
Karlsruhe	41,856 (170)	-6.4 (3.8)	0.0 (0.3)	33.3 (25.2)
Darwin	22,437 (134)	-14.3 (-10.5)	0.1 (0.9)	18.6 (19.1)
Wollongong	15,987 (224)	2.8 (11.6)	0.2 (0.8)	35.4 (29.8)
TCCON	**143,814 (854)**	**2.0 (12.4)**	**0.0 (1.2)**	**35.3 (31.0)**
ALL SITES	**205,331 (1,106)**	**1.9 (12.6)**	**0.0 (0.2)**	**35.2 (31.3)**

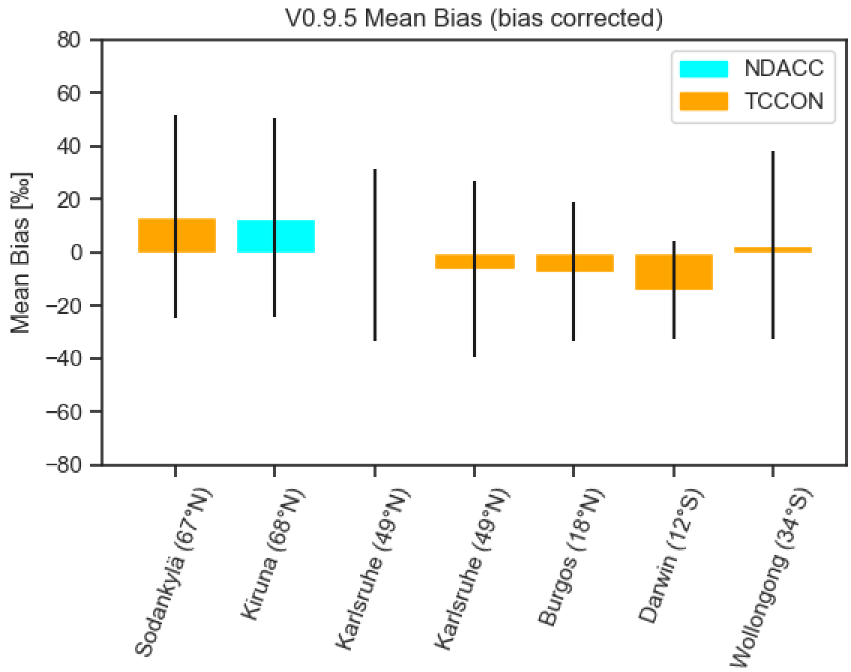

Figure 24: TROPOMI XδD validation results against MUSICA NDACC and TCCON data at 7 stations within the period between May 2018 and August 2020 arranged from highest to lowest latitude after the XH2O dependent bias correction. The mean bias for the NDACC site in Karlsruhe is at -0.9 and therefore so low that no cyan bar can be displayed.

The bias correction removes a large part of the inconsistencies between the high and low latitude sites. In the original data the inconsistencies in the bias for the high latitude site Sodankylä and the low latitude site Burgos was about 42 ‰ (bias of -10.2 ‰ and -52.7 ‰ for the high and low latitude site, respectively). In the corrected data the inconsistency in the bias between these two sites is reduced to about 20 ‰ (bias of +13.4 ‰ and -7.2 ‰ for the low and high latitude site, respectively).

7 Conclusions

The validation of the new TROPOMI L2 V0.9.5 XδD prototype product with seven ground-based FTIR sites demonstrates the good quality of the product developed by the University of Leicester. The TROPOMI XδD mean bias with respect to the ground-based FTIR data is about -21 ‰. The dispersion between the individual TROPOMI XδD data and the ground-based FTIR references is about 35 ‰ when considering all data (including data corresponding to observations at very dry conditions). If we exclude very dry conditions, e.g. by requiring XH2O values larger than 1500 ppmv the dispersion can be significantly reduced and remains generally below 30 ‰.

We can assume that the dispersion between TROPOMI and the reference datasets is due to: (1) an air parcel mismatch between TROPOMI and the reference experiments (which can be largely excluded, because we found no reduced dispersion by imposing stricter collocation criteria), (2) uncertainties in the ground-based references of about 10-15 ‰ (see [RD1]), (3) different vertical sensitivities of the TROPOMI and the reference data products, and (4) uncertainties in the TROPOMI data. If we consider a 10-15 ‰ uncertainty of the reference data and assume an uncertainty of another 15 ‰ due the different vertical sensitivities, the 35 ‰ dispersion means that the uncertainty of the TROPOMI data is about 25-30 ‰. The assumption that the error in the TROPOMI data is the main reason for the dispersion is consistent with Fig. 19 (right panel). Here, we observe a clear correlation between the theoretically estimated TROPOMI noise error and the dispersion (the standard deviation). We conclude that the uncertainty estimated for the TROPOMI data is slightly above the target requirement of the precision (set to 20 ‰ in Sect. 4.2). Nevertheless, this estimated uncertainty of 30‰ is for an individual observation (single TROPOMI pixel).In this context, the target value of 20 ‰ can be achieved by calculating averages over several individual pixels.

Furthermore, we have investigated a water vapour dependent bias correction. We found that it widely reduces the bias and most importantly it removes the inconsistency in the bias between low and high latitudes to a large part.