The SPARC water vapour assessment II:
Profile-to-profile and climatological comparisons of stratospheric δD(H$_2$O) observations from satellite

Charlotta Högberg[1], Stefan Lossow[2], Farahnaz Khosrawi[2], Ralf Bauer[3], Kaley A. Walker[3], Patrick Eriksson[4], Donal P. Murtagh[4], Gabriele P. Stiller[2], Jörg Steinwagner[5], and Qiong Zhang[1]

[1]Stockholm University, Department of Physical Geography, Svante-Arrhenius-väg 8, 10691 Stockholm, Sweden.
[2]Karlsruhe Institute of Technology, Institute for Meteorology and Climate Research, Hermann-von-Helmholtz-Platz 1, 76344 Leopoldshafen, Germany.
[3]University of Toronto, Department of Physics, 60 St. George Street, Toronto, Ontario M5S 1A7, Canada.
[4]Chalmers University of Technology, Department of Earth and Space Sciences, Hörsalsvägen 11, 41296 Göteborg, Sweden.
[5]Max-Planck-Institute for extraterrestrial Physics, Gießenbachstraße 1, 85748 Garching, Germany

Correspondence: Charlotta Högberg (charlotta.hogberg@natgeo.su.se)

Version: Thursday 17th January, 2019 11:38 CET

Abstract. Within the framework of the second SPARC (Stratosphere-troposphere Processes And their Role in Climate) water vapour assessment (WAVAS-II), we have evaluated five data sets of δD(H$_2$O) obtained from observations by Odin/SMR (Sub-Millimetre Radiometer), Envisat/MIPAS (Environmental Satellite/Michelson Interferometer for Passive Atmospheric Sounding) and SCISAT/ACE-FTS (Science Satellite/Atmospheric Chemistry Experiment-Fourier Transform Spectrometer) using profile-to-profile and climatological comparisons. These comparisons aimed to provide a comprehensive overview of typical uncertainties in the observational database that can be considered in the future in observational and modelling studies. Our primary focus is on stratospheric altitudes, but results for the upper troposphere and lower mesosphere are also shown. There are clear quantitative differences in the measurements of the isotopic ratio, mainly in regard to comparisons between the SMR data set and both the MIPAS and ACE-FTS data sets. In the lower stratosphere, the SMR data set shows a higher depletion in δD than the MIPAS and ACE-FTS data sets. The differences maximise close to $50\,\text{hPa}$ and exceed $200\,\text{per mille}$. With increasing altitude, the biases decrease. Above $4\,\text{hPa}$, the SMR data set shows a lower depletion in δD than the MIPAS data sets, occasionally exceeding $100\,\text{per mille}$. Overall, the δD biases of the SMR data set are driven by HDO biases in the lower stratosphere and by H$_2$O biases in the upper stratosphere and lower mesosphere. In between, in the middle stratosphere, the biases in δD are the result of deviations in both HDO and H$_2$O. These biases are attributed to issues with the calibration, in particular in terms of the sideband filtering, and uncertainties in spectroscopic parameters. The MIPAS and ACE-FTS data sets agree rather well between about $100\,\text{hPa}$ and $10\,\text{hPa}$. The MIPAS data sets show less depletion below about $15\,\text{hPa}$ (up to about $30\,\text{per mille}$), due to differences in both HDO and H$_2$O. Higher up the picture is reversed, and towards the upper stratosphere the biases increase. This is driven by increasing biases in H$_2$O and on occasion the differences in δD exceed $80\,\text{per mille}$. Below $100\,\text{hPa}$, the differences between the MIPAS and ACE-FTS data sets are even larger. In the climatological

comparisons, the MIPAS data sets continue to show less depletion in δD than the ACE-FTS data sets below 15 hPa during all seasons, with some variations in magnitude. The differences between the MIPAS and ACE-FTS data have multiple causes, such as differences in the temporal and spatial sampling (except for the profile-to-profile comparisons), cloud influence, vertical resolution, and the microwindows and spectroscopic database chosen. Differences between data sets from the same instrument are typically small in the stratosphere. Overall, if the data sets are considered together, the differences in δD among them in key areas of scientific interest (e.g. tropical and polar lower stratosphere, lower mesosphere and upper troposphere) are too large to draw robust conclusions on atmospheric processes affecting the water vapour budget and distribution, e.g. the relative importance of different mechanisms transporting water vapour into the stratosphere.

1 Introduction

Water vapour is one of the most important trace constituents in the Earth's atmosphere. In the upper troposphere and lower stratosphere, water vapour is the most important greenhouse gas. A large part of the predicted global warming is a result of different feedback processes induced by greenhouse gas emissions, where a significant contribution is related to an increased amount of water vapour in the troposphere due to increased average global temperatures. A warmer climate can hold more water vapour in the atmosphere and the main source for water vapour in the troposphere comes from evaporation, in a large part originating from the oceans. Changes in the troposphere will also influence higher altitudes, and it has been shown that the amount of stratospheric water vapour increases with increasing tropospheric temperature due to reduced freeze-drying in the tropical tropopause layer (TTL; Gettelman et al., 2009). The strength of the stratospheric water vapour feedback, which implies that an increase of water vapour in the lower stratosphere in turn leads to an even warmer climate, is estimated to be $0.3\,\mathrm{Wm^{-2}}$ for a 1 K temperature anomaly at 500 hPa (Dessler et al., 2013). As a main component of polar stratospheric clouds (PSCs), water vapour also plays a crucial role in the ozone chemistry in the middle atmosphere. The heterogeneous reactions that take place on the surface of the PSC particles cause the ozone depletion observed during winter and spring in the polar lower stratosphere. The formation of PSCs also plays a role in dehydration in the polar regions during winter and spring, where large PSC particles containing ice crystals can sediment due to gravitational effects and permanently remove water vapour from the stratosphere (Kelly et al., 1989). Further water vapour is the primary source for the hydrogen radicals OH, H, and HO_2, which also contributes to the loss of ozone within auto-catalytic cycles. They dominate the ozone budget in the lower stratosphere and at altitudes above 50 km (Brasseur and Solomon, 2005). Beyond that, water vapour is a valuable tool to diagnose dynamical processes in the stratosphere and mesosphere (e.g. Mote et al., 1996; Seele and Hartogh, 1999; Lossow et al., 2009).

A major source of stratospheric water vapour is the transport from the troposphere, which occurs mainly through the TTL. Slow ascent, accompanied by large horizontal motions, is thought to be the most important pathway (Brewer, 1949; Fueglistaler et al., 2009). The amount of water vapour entering the stratosphere is controlled by the cold point temperature along the air parcel trajectories, which have a seasonal variation. Lower temperatures during the boreal winter time result in drier air entering the stratosphere. In the boreal summer the situation is reversed, and due to higher temperatures moister air enters the

stratosphere. This signal is transported upwards by the upwelling branch of the Brewer-Dobson circulation and is traceable up to about 30 km in the isolated tropical pipe region, above which it mixes out. It is known as the tape recorder signal (Mote et al., 1996). Another pathway into the stratosphere is the convective lofting of ice (Moyer et al., 1996). Typically 3.5 to 4.0 ppmv of water vapour are transported from the troposphere into the stratosphere (Kley et al., 2000). Within the stratosphere, the

5 in situ oxidation of methane is a major source for water vapour. In the lower stratosphere typically 1.5 to 1.7 water vapour molecules are produced from one methane molecule. With increasing altitude, the production rate increases. In the upper stratosphere about 2.2 water vapour molecules are produced from one methane molecule (Le Texier et al., 1988; Frank et al., 2018). A minor source of water vapour is the oxidation of molecular hydrogen, which is of importance only in the upper stratosphere and lower mesosphere (Wrotny et al., 2010). The reaction with $O(^1D)$ to form hydrogen radicals is the major sink

10 of water vapour in the stratosphere. With increasing altitude, the destruction by photolysis becomes more important. Overall, in the stratosphere the amount of water vapour increases with altitude. Around the stratopause, a maximum is found. In the mesosphere the amount typically decreases, as there is no major source.

More than 99.7% of water vapour exists in the form of $H_2^{16}O$. There are several minor isotopologues, like $H_2^{18}O$ (0.20%), $H_2^{17}O$ (0.037%), and $HD^{16}O$ (0.03%). Although found in low abundance, the minor isotopologues can provide information on

15 the process history of air parcels from their isotopic ratios relative to the main isotopologue, $H_2^{16}O$ (hereafter named H_2O). In this regard, $HD^{16}O$ (hereafter named HDO) is most interesting, as the isotopic ratio typically exhibits pronounced variations. The isotopic ratio between HDO and H_2O is typically given in the δD notation (Eq. 1), which describes the relative deviation of deuterium (D) to hydrogen (H) with respect to the reference ratio Reference = VSMOW = 155.76×10^{-6} (Vienna Standard Mean Ocean Water, Hagemann et al. (1970)).

20 $$\delta D(H_2O) \quad = \quad \left(\frac{R_{sample}}{R_{reference}} - 1 \right) \cdot 1000\,\text{‰} \tag{1}$$

$$R_{sample} \quad = \quad \left(\frac{[D]}{[H]} \right)_{sample} \equiv \left(\frac{[HDO]}{2 \cdot [H_2O]} \right)_{sample} \tag{2}$$

Hereafter we refer to $\delta D(H_2O)$ simply as δD. In Eq. (2), two approximations are made: (a) that the deuterium content in a sample is dominated by the contribution from $HD^{16}O$ and that the contributions from the other deuterium bearing isotopologues are negligible, (b) that the hydrogen content essentially comes from H_2O. In the upper troposphere and tropopause

25 region, the isotopic ratio is primarily determined by condensation and evaporation processes, as a consequence of the vapour pressure isotope effect. The heavier isotopologue, i.e. HDO, has a lower vapour pressure than H_2O, leading to changes of the isotopic ratio whenever a phase change occurs. In the stratosphere, the oxidation of methane causes an increase in the isotopic ratio, since methane is not depleted in the heavier isotopologues to the same extent as water vapour during the transport from the troposphere to the stratosphere.

30 Given these influences, the isotopic ratio can be used to investigate the relative importance of different processes that contribute to the transport of water vapour from the troposphere to the stratosphere (Moyer et al., 1996; Nassar et al., 2007; Payne

et al., 2007; Sayres et al., 2010; Steinwagner et al., 2010; Eichinger et al., 2015). If air dehydrates to the saturation mixing ratio as it slowly ascends through the TTL, undergoing a Rayleigh fractionation process, a δD value around -900 permille would be expected near the tropopause. Observations, however, exhibit δD values between -700 per mille and -500 per mille (e.g. Moyer et al., 1996; Johnson et al., 2001; Webster and Heymsfield, 2003). This indicates that non-Rayleigh processes like the convective lofting of ice particles and their subsequent sublimation must occur.

Within the framework of the second SPARC water vapour assessment, we present here a comprehensive comparison of δD data sets obtained from satellite observations. These data sets are evaluated from the upper troposphere to the lower mesosphere with a primary focus on the stratosphere. In the comparison, we focus on satellite observations made since the new millennium. In that time, three satellite instruments have provided information on stratospheric δD: Odin/SMR (Sub-Millimetre Radiometer), Envisat/MIPAS (Environmental Satellite/Michelson Interferometer for Passive Atmospheric Sounding) and SCISAT/ACE-FTS (Science Satellite/Atmospheric Chemistry Experiment-Fourier Transform Spectrometer). The satellites were launched in 2001, 2002, and 2003, respectively. While Envisat ceased its operation in 2012, the other two instruments are still performing observations. In an earlier work, Lossow et al. (2011) compared HDO data observed by these three instruments. They found a good agreement in the MIPAS and ACE-FTS data sets in the stratosphere, while the SMR data set showed a low bias. This bias could be explained by uncertainties in the spectroscopic parameters. Among the data sets, a high degree of consistency in the latitudinal distribution of HDO was found. For δD there are no published comparisons between these satellite observations. Instead, the data sets were analysed individually, typically focusing on the variability of dehydration in the tropical tropopause layer and lower stratosphere. Nassar et al. (2007) used ACE-FTS observations from 2004 and 2005 to examine the dehydration in the TTL. They found δD values between -700 per mille and -600 per mille and a seasonal variation that is more obvious in the Northern Hemisphere. Steinwagner et al. (2010) used MIPAS observations from 2002 to 2004 and showed a tape recorder signal in δD, corroborating the dominant role of the slow ascent for the transport of water vapour from the troposphere to the stratosphere. Later, Randel et al. (2012) evaluated ACE-FTS data from 2004 to 2009 in this regard and found a coherent tape recorder signal only up to about $20\,\text{km}$. This observational discrepancy remains unresolved up to now.

In this work, we present coincident profile-to-profile and climatological comparisons of δD. For a better attribution and discussion of the issues in the isotopic ratio we also show the corresponding HDO and H_2O results. In the next section we describe in detail the individual data sets. In Sect. 3 our approach for the profile-to-profile and climatological comparisons is outlined. In Sect. 4 we present the results which will be summarised and discussed in Sect. 5.

2 Data sets

In this work we consider five data sets. From the SMR observations one data set is derived. Based on different retrieval versions two data sets each are obtained from the MIPAS and ACE-FTS observations. Data from the older retrieval versions have been already used in previous studies (Steinwagner et al., 2010; Randel et al., 2012) and therefore provide context for the more recent retrieval results.

2.1 Odin/SMR

Odin is a Swedish-led satellite that is dedicated to both aeronomy and astronomy observations. Launched on 20 February 2001, it uses a sun-synchronous orbit with equator-crossing times of about 6 LT (local time) and 18 LT on the descending and ascending node, respectively. Two instruments are deployed aboard the satellite; one of them is the Sub-Millimetre Radiometer (SMR). It measures the thermal emission at the atmospheric limb using a 1.1 m telescope. The instrument consists of five radiometers that cover several frequency bands between 486 GHz and 581 GHz and around 119 GHz (Frisk et al., 2003). For the detection of the measured signal, either one acousto-optical spectrograph or one of two autocorrelators are used. The SMR observations typically cover the latitude range between 82.5°S and 82.5°N based on measurements along the orbital track. Since 2004, some measurements off the orbital track have been performed to obtain full latitudinal coverage from pole to pole, mainly during boreal winter.

The HDO and H_2O information used in the present work is retrieved from emission lines centred at 490.597 GHz and 488.491 GHz, respectively (Urban et al., 2007). These lines are always measured with the 495A2 radiometer that can be tuned to measure a maximum bandwidth of 0.8 GHz within a frequency range from 486 GHz to 504 GHz. As the spectral separation of the two emission lines is larger than the maximum bandwidth, HDO and H_2O information cannot be obtained at the same time; consequently, no δD information on a single profile basis is available. Typically the HDO and H_2O observations are performed in an alternating manner, with one orbit measuring HDO, followed by one orbit measuring the H_2O emission line. Since SMR has a multitude of measurement targets and modes, the HDO and H_2O bands considered here are not observed on a daily basis. Until 25 April 2007, the typical observation frequency was 3 – 4 days days per month. After that, the astronomy observations ceased, leading to an increased observation frequency of 8 – 9 days days a month. The band is typically observed in the altitude range between 7 and 110 km with an effective vertical sampling of 3 km. Such scan takes about 140 seconds, which corresponds to a horizontal sampling of 1 scan per 1000 km.

In the comparisons, we consider data that was derived at the Chalmers University of Technology in Gothenburg, Sweden using retrieval version 2.1 (Jones et al., 2009; Lossow et al., 2011). The retrieval of HDO and H_2O profiles is based on the Optimal Estimation Method (OEM, Rodgers, 2000) using spectroscopic data from the Verdandi database (Eriksson, 1999). Both HDO and H_2O information can be retrieved in the altitude range between 20 and 70 km with a vertical resolution that is very close to the vertical sampling (Urban et al., 2007). The precision of a single HDO scan is best at around 30 km with about 20%. Towards the lower and upper boundaries, the precision degrades to values beyond 50% on average. For H_2O, the single-scan precision is for a large part of the covered altitude range better than 10%. At the profile boundaries, the precision is typically in the order of 30%.

Earlier comparisons have shown that the SMR HDO data exhibit a dry bias in the stratosphere (Lossow et al., 2011). A dry bias has been also observed in H_2O in the upper stratosphere and lower mesosphere (Hegglin et al., 2013). In this altitude region a positive drift is currently being investigated.

The SMR data are screened according to three parameters, namely the retrieval quality flag, the measurement response, and a χ^2 flag. The quality flag marks if a retrieved profile should be used in any scientific analysis based on a number of

retrieval diagnostics (i.e. convergence, cost function, regularisation of the retrieval and the retrieved pointing offset). Here we considered only data with the two best quality categories, i.e. 0 (all diagnostics are fine) and 4 (all diagnostics are fine except the value of the final regularisation parameter). The measurement response describes the relative contribution of observational and a priori information to the retrieved data. We chose a minimum measurement response of 70% to limit the influence of the a priori data on our results. The χ^2 flag is an additional criterion to earlier studies aiming to improve the screening of unreasonable data. This flag combines the fit quality of the retrieved values towards the measurement and that of the true state towards the a priori. In general, the first part contributes the most. For HDO, only data with a χ^2 flag ranging from 0.03 and 0.8 have been taken into consideration; for H_2O, data in the range between 0.03 and 0.6 have been considered. Application of this criterion rejects about 3% of the HDO profiles and 4% of the H_2O profiles. Scientifically usable data dates back to November 2001. Earlier data are not recommended due to problems with the instrument pointing. After consultation with the SMR data team, we only included data until May 2009 in the comparisons, due to a combination of calibrational issues, instrument drifts, and also some processing issues. In general, only limited HDO data are currently available after April 2011 due to a frequency drift, which is typically less than 10% on a yearly basis relative to the number of available profiles in 2008. This drift is basically handled by the retrieval scheme; however, larger frequency shifts trigger indirect forward model problems that cause an incorrect representation of the line shape. In total, 77,000 HDO profiles and 83,000 H_2O profiles are available for the comparisons presented here.

2.2 Envisat/MIPAS

The Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) was a cooled high-resolution Fourier transform spectrometer aboard Envisat (Environmental Satellite). This satellite was launched on 1 March 2002 and performed observations until 8 April 2012, when communication with it broke down. Envisat orbited the Earth 14 times a day on a polar, sun-synchronous orbit with an altitude of about 790 km. The equator crossing times were 10 LT and 22 LT for the descending and ascending node, respectively. MIPAS measured thermal emission at the atmospheric limb in the wavelength range between 4.1 and 14.6 μm ($685 - 2410$ cm^{-1}) covering all latitudes. MIPAS observations from July 2002 to March 2004, which is referred to as the full resolution period of MIPAS, were included in the comparison. During this period, measurements used a spectral resolution of 0.035 cm^{-1} (un-apodised) and covered the altitude range between 6 and 68 km, with a vertical sampling of 3 (up to 42 km) to 8 km. A whole scan took 76 s, corresponding to a horizontal sampling of 1 scan per 530 km. Later MIPAS observations used a coarser spectral resolution due to an instrument problem, and also the vertical sampling pattern was changed (Fischer et al., 2008).

We use two sets of HDO and H_2O data in the comparisons, namely retrieval version 5 and 20. Both data sets are retrieved with the IMK/IAA processor, which has been developed in cooperation between the "Institut für Meteorologie und Klimaforschung" (IMK) in Karlsruhe (Germany) and the "Instituto de Astrofísica de Andalucía" (IAA) in Granada (Spain). The retrieval approach for both data sets is the same; the only difference is the calibration of the spectra provided by the European Space Agency (ESA). Retrieval version 5 is based on calibration version 3 (Lossow et al., 2011), while retrieval version 20 uses calibration version 5. The retrieval of HDO and H_2O used spectral information from 14 microwindows located

between 6.7 and 8.0 μm (1250 – 1483 cm^{-1}) (Steinwagner et al., 2010). It should be noted that this H$_2$O data set is different from the nominal IMK/IAA H$_2$O data set (see studies by e.g. Schieferdecker et al., 2015 or Lossow et al., 2017) which is based on spectral information of 12 microwindows between 6.3 and 12.6 μm (796 – 1579 cm^{-1}). As the SMR retrieval, the MIPAS retrieval uses a non-linear least square approach. To avoid unphysical oscillations in the retrieved profiles, a first-order

5   Tikhonov-type regularisation (Tikhonov and Arsenin, 1977) is employed. Spectroscopic data are taken from a compilation set-up especially designed for the MIPAS mission (Flaud et al., 2003), which considers for this retrieval data from the updated version of HITRAN-2000 (High Resolution Transmission, Rothman et al., 2003. HDO information can be retrieved between about 10 and 60 km with a typical single scan precision of 20% below 50 km and around 100% at 60 km. In the upper tropo-sphere and lower stratosphere, the vertical resolution is typically about 5 km. Towards higher altitudes, the resolution degrades

10   and in the upper stratosphere and lower mesosphere it is in the order of 8 to 10 km. According to Steinwagner et al. (2007), the HDO accuracy is 0.1 ppbv in the lower stratosphere and increases up to 0.2 ppbv in the upper stratosphere. These estimations are, however, only based on a single tropical profile in January. For H$_2$O, retrievals are possible to somewhat higher altitudes than for HDO. The single scan precision is in general within 10%. The vertical resolution is worse than for HDO up to the middle stratosphere (about 6 km), while it is better in the upper stratosphere and lower mesosphere (about 8 km). For δD,

15   Steinwagner et al. (2007) estimated an accuracy between 100 and 150 per mille.

The MIPAS data are screened according to the visibility flag and the averaging kernel diagonal criterion. The visibility flag marks retrieved data below the lowermost useful tangent altitude. This lowermost altitude is often determined by clouds, whose inference is derived using a cloud index (Spang et al., 2004). Clouds are rigorously screened, resulting in a bias towards cloud-free situations. The averaging kernel diagonal indicates the measurement contribution to the retrieved data, similar to

20   the measurement response used to screen the SMR data. Data with a diagonal value of less than 0.03 are discarded to ensure at least a minimum of measurement information. In addition, data above the uppermost tangent altitude are not considered any further. Overall, the v5 data set consists of more than 460,000 simultaneous observations of HDO and H$_2$O with a nearly daily coverage. More than 480,000 profiles comprise the v20 data set.

2.3   SCISAT/ACE-FTS

25   ACE-FTS (Atmospheric Chemistry Experiment - Fourier Transform Spectrometer) is one of three instruments aboard the Canadian SCISAT (or SCISAT-1) satellite (Bernath et al., 2005). It was launched on 12 August 2003 into a high inclination orbit with an altitude of 650 km. This orbit provides a latitudinal coverage from 85°S to 85°N, but is optimised for observations at high and mid-latitudes. Like MIPAS, ACE-FTS performs observations in the infrared. The instrument covers the wavelength range between 2.3 and 13.3 μm (750 – 4400 cm^{-1}) with a high spectral resolution of 0.02 cm^{-1}. The observations are based

30   on the solar occultation technique. ACE-FTS scans the Earth's atmosphere 30 times a day (15 sunrises and 15 sunsets) from about 5 to 150 km. The vertical sampling varies with altitude, ranging from about 1 km in the middle troposphere, to 3 to 4 km at around 20 km, and 6 km in the upper stratosphere and mesosphere.

The ACE-FTS retrieval is based on an unconstrained, non-linear, least-squares, global-fitting technique (Boone et al., 2005, 2013) In the comparisons we considered data from two retrieval versions, i.e. from the well-validated version 2.2 and the newer

version 3.5. The HDO retrieval of version 2.2 (Nassar et al., 2007) is based on spectral information from 24 microwindows in two separated wavelength intervals. One interval ranges from 3.7 to 3.8 μm ($2612 - 2673\,\mathrm{cm}^{-1}$) and the other ranges from 6.7 to 7.1 μm ($1402 - 1498\,\mathrm{cm}^{-1}$). This yields HDO profiles covering altitudes from 5.5 to 37.5 km. The single-scan precision is in general better than 10% below 30 km. At the top of the profiles, the precision typically amounts to 25%. In retrieval version 3.5,
5   the number of microwindows was increased to 26. In addition, the 2 wavelength intervals were extended, i.e. they range from 3.7 to 4.0 μm ($2493 - 2673\,\mathrm{cm}^{-1}$) and from 6.6 to 7.2 μm ($1383 - 1511\,\mathrm{cm}^{-1}$). As a result of these changes, HDO information can be retrieved at higher altitudes, typically up to 49.5 km. The single-scan precision is rather similar to version 2.2. At 49.5 km, the precision is about 50%. The H_2O retrieval of version 2.2 uses 60 microwindows, providing profile information up to 89.5 km (Carleer et al., 2008). As for the HDO retrieval, the microwindows are separated in 2 wavelength intervals, ranging
10   from 5.0 to 7.3 μm (1362–$2000\,\mathrm{cm}^{-1}$) and from 10.3 to 10.5 μm (953–$974\,\mathrm{cm}^{-1}$). In version 3.5, the microwindows were optimised once more. Using 54 microwindows between 3.3 and 10.7 μm (937–$2993\,\mathrm{cm}^{-1}$), retrieval version 3.5 yields again an extension of the upper altitude limit where H_2O information can be retrieved. The single-scan precision of H_2O profiles is within 5% for both retrieval versions. The spectroscopic data employed in the ACE-FTS retrieval is taken from HITRAN-2004 (Rothman et al., 2005), including some more recent updates as detailed in Boone et al. (2013). The vertical resolution is
15   the same for all data sets and corresponds to the vertical sampling outlined above. Like for MIPAS, δD can be derived on a single-profile basis.

    The screening of the ACE-FTS depends on the retrieval version. For version 2.2 the "data issues page" https://databace. scisat.ca/validation/data_issues_table.php is taken into account, which lists problematic occultations. Besides the occultations flagged with "do not use" we also discard occultations that are marked with the label "use with caution". Data after September
20   2010 are not used due to problems with the input pressure and temperature data used in the retrieval. In version 3.5 these problems are resolved and data until the end of 2014 are used in this study, which is the last year considered in the water vapour assessment activity. The screening of this data set uses the flag system described in detail by Sheese et al. (2015). We remove all profiles that are flagged with a value between 3 and 7 at any altitude. These numbers indicate either outliers, a lack of data to perform the outlier analysis as well as instrument or processing errors. In total, 22000 simultaneous observations of
25   HDO and H_2O for version 2.2 are available for comparison. For the longer version 3.5 data set these are slightly more than 40000 profiles.

3   Approach

The present quality assessment of δD, HDO and H_2O data focuses primarily on the stratosphere, although we use data for the upper troposphere and lower mesosphere where available. For the comparisons provided in this study we use different
30   approaches that are based on coincident measurements as well as monthly and seasonal averages.

    A set of simultaneous HDO and H_2O observations can in general be combined to a δD product, as for example an average, in two different ways:

(1) Calculate δD from individual HDO and H_2O profiles and subsequently derive the δD product of interest. Here we denote this approach as "individual".

(2) Derive first the product of interest separately for HDO and H_2O and subsequenly calculate δD. Here we refer to this approach as "separate".

In general these two approaches are not commutative in a mathematical sense and will yield different results. While we prefer the "individual" approach, the results presented in this work are based on the "separate" approach. The main motivation for this is consistency, as the SMR observations do not allow the derivation of δD on a single profile basis (see Sect. 2.1). In the supplement to this paper we show δD results from MIPAS and ACE-FTS that are based on the "individual" approach. Those results are also compared to the results that we show here in the main part of the manuscript to provide estimates of the differences caused by the different approaches.

For the comparisons, the profiles of the individual data sets are interpolated on a common regular pressure grid, consisting of 32 levels per pressure decade and ranging from 421 to $0.1\,\mathrm{hPa}$. For MIPAS and ACE-FTS, a consistent set of HDO and H_2O observations is used. If an HDO observation is not available (due to a failed retrieval or screening) then the simultaneous H_2O profile is not used in the comparisons and vice versa.

To handle data that could potentially negatively influence the comparison results, we defined further screening criteria in addition to the standard screening of the individual data sets outlined in Sect. 2. The screening criteria are based on intervals for the volume mixing ratios of HDO and H_2O and the δD isotopic ratio. The intervals are listed in Tab. 1. They are fairly large, targeting the most obvious outliers that passed the previous screenings. Profiles that exhibit data points outside these intervals are discarded. For HDO and H_2O this typically concerns only a handful of profiles. Solely, for the ACE-FTS v2.2 data set the number of affected profiles is higher, i.e. 30 for HDO and 50 for H_2O. Considerably more profiles are screened for δD. The percentage ranges from around 0.5% for the MIPAS data sets to 1% for ACE-FTS v3.5 and 1.8% for ACE-FTS v2.2. Note that HDO and H_2O can have negative volume mixing ratios due to measurement noise that propagates through the retrieval. These volume mixing ratios are not removed; however, in combination they can cause isotopic ratios below the theoretical limit of $-1000‰$.

In the following, we describe the approaches and considerations for the different comparisons. First we discuss the profile-to-profile comparisons, then the seasonal comparisons, and finally the monthly comparisons.

3.1 Profile-to-profile comparisons

3.1.1 Coincidences

In the profile-to-profile comparisons, we consider in general observations from two data sets as coincident when they meet the following criteria: (1) a spatial separation of less than $1000\,\mathrm{km}$, (2) a temporal separation that does not exceed $24\,\mathrm{h}$ and a separation of less than $5°$ both in (3) geographical and (4) equivalent latitude. In comparisons between the two MIPAS data sets or the two ACE-FTS data sets the exact same observations are used (see Tab. 2). For the equivalent latitude criterion, we

use a stratospheric average value derived from MERRA (Modern Era Retrospective-Analysis for Research and Applications, Rienecker et al., 2011) reanalysis data. In cases where multiple coincidences are found for one observation, the coincidence closest in space is used, given the small local time variation of stratospheric water vapour (Haefele et al., 2008). In the tropo- sphere and higher up in the mesosphere, the variation can be substantial; however, these regions are not the main focus of this study. Additionally, we consider only unique coincidences, i.e. once an observation is considered as a coincidence it is not used any further as coincidence for other observations.

3.1.2 Consideration of differences in the vertical resolution

The vertical resolution of the different HDO and H_2O data sets varies as described in Sect. 2. In particular, the MIPAS HDO and H_2O data sets have a lower vertical resolution compared to the other data sets. These differences only require consideration in the comparisons at altitudes where the vertical distribution of the parameter in question is highly structured. This especially concerns the hygropause region in the lowermost stratosphere. The maximum of HDO and H_2O around the stratopause is relatively broad. This makes differences in the vertical resolution a smaller issue than for the hygropause region, yet a direct comparison may still not be appropriate.

The differences in the vertical resolution in the HDO and H_2O comparisons to MIPAS are accounted for by the method of Connor et al. (1994). Using the averaging kernel $\mathbf{A}$ and the a priori profile $\boldsymbol{x}_{\mathrm{a\ priori}}$ from the MIPAS retrieval, the higher vertically resolved SMR and ACE-FTS data ($\boldsymbol{x}_{\mathrm{high}}$) can be degraded onto the MIPAS vertical resolution as follows:

$$\boldsymbol{x}_{\mathrm{deg}} = \boldsymbol{x}_{\mathrm{a\ priori}} + \mathbf{A} \cdot (\boldsymbol{x}_{\mathrm{high}} - \boldsymbol{x}_{\mathrm{a\ priori}}). \tag{3}$$

The degraded data $\boldsymbol{x}_{\mathrm{deg}}$ can subsequently be compared directly to the lower resolved MIPAS data. The comparisons between SMR and ACE-FTS are performed directly, as these data sets have a very similar vertical resolution in the altitude range where they overlap.

Also, δD shows pronounced structures in its vertical distribution, for example in the tropopause region (Webster and Heyms- field, 2003; Nassar et al., 2007). Thus, comparisons of individual δD profiles between the MIPAS and ACE-FTS data sets (see Supplement) should again consider the differences in the vertical resolution. In our case a direct application of Eq. 3 is not possible since δD is no retrieved quantity for any of the data sets and consequently no averaging kernels exist. Alternatively the vertical resolution differences in HDO and H_2O can be considered prior to the calculation of the individual δD profiles. Another solution is the generation of appropriate kernels using triangle or Gaussian functions (Dupuy et al., 2009; Lossow et al., in preparation). In the supplement we employ the first alternative.

3.1.3 Bias determination

The bias $\bar{b}(P,z)$ between two coincident data sets for a given parameter P (i.e. HDO, H_2O or δD) at an altitude z is derived as follows:

$$\bar{b}(P,z) = \frac{1}{n_c(P,z)} \cdot \sum_{i=1}^{n_c(P,z)} b_i(P,z). \tag{4}$$

5   In the equation $n_c(P,z)$ represents the number of coincident measurements and $b_i(P,z)$ denotes the individual differences between those. We have considered these differences in absolute terms:

$$b_i(P,z) = b_{i,\mathrm{abs}}(P,z) = x_i(P,z)_1 - x_i(P,z)_2 \tag{5}$$

and also in relative terms:

$$b_i(P,z) = b_{i,\mathrm{rel}}(P,z) = \frac{x_i(P,z)_1 - x_i(P,z)_2}{[x_i(P,z)_1 + x_i(P,z)_2]/2}. \tag{6}$$

10    $x_i(P,z)_1$ and $x_i(P,z)_2$ are the individual observations of the two data sets that are compared with each other. For the relative bias, the average over both data sets is used as reference. This follows the convenience argument employed by Randall et al. (2003), given the a priori knowledge that the data sets have large uncertainties, especially for HDO and δD.

   Before we derive the overall bias $\bar{b}(P,z)$, we perform an additional screening on individual biases $b_i(P,z)$ that are obvious outliers and would skew the bias estimates. This screening employs the median and median absolute deviation (MAD; e.g. Jones et al., 2012), which are more robust quantities in terms of larger outliers. We discarded individual biases that were outside the following interval: $\langle \mathrm{median}[b_i(P,z)] \pm 10 \cdot \mathrm{MAD}[b_i(P,z)] \rangle$ with $i = 1, ..., n_c(P,z)$. A factor of 10 for the median absolute deviation corresponds to a factor of about 7.5 for the standard deviation considering a normally distributed set of data. Thus the screening is relatively weak and aims to detect the most prominent outliers.

20   For the δD bias, the terminology above applies only for the individual approach, for which results are shown in the Supplement. In the separate approach, we calculate first average profiles of HDO and H_2O (again described by P) for each data set separately from the coincident set of data:

$$\bar{x}(P,z) = \frac{1}{n_c(P,z)} \cdot \sum_{i=1}^{n_c(P,z)} x_i(P,z). \tag{7}$$

In this step we consider exactly the same data points of the data sets that are compared for a given parameter, i.e. if a data point does not exist in one data set (due to missing coverage or screening) the corresponding data point in the other data set is also not considered. The average HDO and H_2O profiles of a given data set are subsequently combined to an average δD profile, following Eq. 1:

$$\overline{x}(\delta D, z) = \left[\frac{\overline{x}(HDO, z)}{2 \cdot VSMOW \cdot \overline{x}(H_2O, z)} - 1 \right] \cdot 1000. \tag{8}$$

The absolute δD bias between two coincident data sets is then calculated as:

$$\overline{b}(\delta D, z) = \overline{x}(\delta D, z)_1 - \overline{x}(\delta D, z)_2. \tag{9}$$

For the relative bias, we follow the approach outlined in Eq. 6:

$$\overline{b}(\delta D, z) = \frac{\overline{x}(\delta D, z)_1 - \overline{x}(\delta D, z)_2}{[\overline{x}(\delta D, z)_1 + \overline{x}(\delta D, z)_2]/2}. \tag{10}$$

3.1.4 More considerations

The results we present for the profile-to-profile comparisons are based on all available coincidences. No separation into specific seasons or latitude band has been considered, thus neglecting these dependencies in Eqs. 4 to 10. Furthermore only biases that are based on at least 20 coincident observations are taken into consideration to avoid spurious result. This concerns mostly the lower and upper altitude boundaries where comparisons are possible.

3.2 Comparisons of seasonally averaged latitude cross-sections

3.2.1 Data binning

To provide an overview how the data sets compare as function of time and space, we consider latitude cross sections for different seasons. Accordingly, we average the complete data sets for a given parameter for all seasons (i.e. MAM, JJA, SON, and DJF, represented by t) and latitude bins ϕ of 10°(centred at 85°S, 75°S, ..., 75°N and 85°N):

$$\overline{s}(P, t, \phi, z) = \frac{1}{n_o(P, t, \phi, z)} \cdot \sum_{i=1}^{n_o(P, t, \phi, z)} x_i(P, t, \phi, z). \tag{11}$$

Here, $n_o(P,t,\phi,z)$ describes the total number of observations $x_i(P,t,\phi,z)$ that fall into a specific bin. Before these data are averaged, they are screened using again the median and median absolute difference. Here we use the interval $\langle \text{median}[x_i(P,t,\phi,z)] \pm 7.5 \cdot \text{MAD}[x_i(P,t,\phi,z)] \rangle$ with $i = 1,...,n_o(P,t,\phi,z)$. This screening is a bit more strict than for the profile-to-profile comparisons, as typically less data fall into individual bins. Besides the average, we also calculate the corresponding standard error $\epsilon(P,t,\phi,z)$:

$$\epsilon(P,t,\phi,z) = \sqrt{\frac{1}{n_o(P,t,\phi,z) \cdot [n_o(P,t,\phi,z)-1]} \cdot \sum_{i=1}^{n_o(P,t,\phi,z)} [x_i(P,t,\phi,z) - \overline{s}(P,t,\phi,z)]^2}. \tag{12}$$

Averages that are based on fewer than 20 individual observations or are less than its associated standard error in absolute terms are not considered any further.

Within the separate framework, the seasonally averaged latitude cross sections for δD are calculated from the corresponding results of HDO and H_2O:

$$\overline{s}(\delta D,t,\phi,z) = \left[\frac{\overline{s}(\text{HDO},t,\phi,z)}{2 \cdot \text{VSMOW} \cdot \overline{s}(H_2O,t,\phi,z)} - 1\right] \cdot 1000. \tag{13}$$

The corresponding standard error is calculated according to the Gaussian error propagation:

$$\epsilon(\delta D,t,\phi,z) = \frac{500}{\text{VSMOW}} \cdot \sqrt{\left[\frac{\epsilon(\text{HDO},t,\phi,z)}{\overline{s}(H_2O,t,\phi,z)}\right]^2 + \left[\frac{\overline{s}(\text{HDO},t,\phi,z) \cdot \epsilon(H_2O,t,\phi,z)}{\overline{s}(H_2O,t,\phi,z)^2}\right]^2}. \tag{14}$$

Given that this information is not available from all data sets, this calculation assumes no error covariance between the HDO and H_2O data to keep consistency.

3.2.2 Summarising the results

For a summary of these comparisons, we provide some additional quantities by combining the results from all individual latitude bands. Results from at least 6 latitude bands are required, otherwise the combined quantity is discarded. As a first quantity, we consider the average bias $\overline{b}_\phi(P,t,z)$ of the latitude cross sections from two data sets:

$$\overline{b}_\phi(P,t,z) = \frac{1}{n_\phi(P,t,z)} \cdot \sum_{i=1}^{n_\phi(P,t,z)} [\overline{s}(P,t,\phi_i,z)_1 - \overline{s}(P,t,\phi_i,z)_2]. \tag{15}$$

The subscripts at the end of the variables refer to the two data sets. Besides the average, we look also at the de-biased standard deviation:

$$\sigma_\phi(P,t,z) = \sqrt{\frac{1}{n_\phi(P,t,z)-1} \cdot \sum_{i=1}^{n_\phi(P,t,z)} \left[\overline{s}(P,t,\phi_i,z)_1 - \overline{s}(P,t,\phi_i,z)_2 - \overline{b}_\phi(P,t,z) \right]^2}. \tag{16}$$

The de-biased standard deviation is generally interpreted as measure of the combined precision of the two data sets that are compared (von Clarmann, 2006). In this specific case, it describes the combined precision of the seasonally averaged latitude cross section from two data sets. As a last step, we consider the correlation coefficient $r_\phi(P,t,z)$ between the latitudinal cross sections from two data sets:

$$r_\phi(P,t,z) = \frac{\sum_{i=1}^{n_\phi(P,t,z)} \left[\overline{s}(P,t,\phi_i,z)_1 - \overline{s}_\phi(P,t,z)_1 \right] \cdot \left[\overline{s}(P,t,\phi_i,z)_2 - \overline{s}_\phi(P,t,z)_2 \right]}{\sqrt{\sum_{i=1}^{n_\phi(P,t,z)} \left[\overline{s}(P,t,\phi_i,z)_1 - \overline{s}_\phi(P,t,z)_1 \right]^2} \cdot \sqrt{\sum_{i=1}^{n_\phi(P,t,z)} \left[\overline{s}(P,t,\phi_i,z)_2 - \overline{s}_\phi(P,t,z)_2 \right]^2}} \tag{17}$$

with

$$\overline{s}_\phi(P,t,z)_1 = \frac{1}{n_\phi(P,t,z)} \cdot \sum_{i=1}^{n_\phi(P,t,z)} \overline{s}(P,t,\phi_i,z)_1 \quad \text{and} \tag{18}$$

$$\overline{s}_\phi(P,t,z)_2 = \frac{1}{n_\phi(P,t,z)} \cdot \sum_{i=1}^{n_\phi(P,t,z)} \overline{s}(P,t,\phi_i,z)_2, \tag{19}$$

which is the average of the seasonally averaged latitude cross section over all latitude bins. No error estimates are considered in the calculation of the correlation coefficient, since the latitudinal cross sections from two data sets are expected to be highly correlated and not simply correlated by chance.

3.2.3 More considerations

These comparisons do not take into account the differences in the vertical resolution between MIPAS and the other data sets. Obtaining the convolution data to adapt to the vertical resolution of non-coincident data sets is not a trivial procedure. Risi et al. (2012) and Schieferdecker (2015) describe methods of doing so. Since the data averaging tends to reduce differences in the vertical resolution, we decided not to consider this aspect any further.

3.3 Comparisons of monthly averaged profiles in the tropics

In the final comparison we consider the tropics, a key region in particular for lower stratospheric water vapour. We focus on February, April, August, and October which are the months the ACE-FTS observations typically provide tropical coverage. In

recent years, there has also been some very limited coverage in May and November due to the shift of the SCISAT orbit, but these months are not considered here. The calculation of the monthly averaged profiles follows the same scheme as outlined by Eqs. 11 to 14, except that only one latitude bin (15°S – 15°N) is considered and that the seasonal dependence is replaced by a monthly dependence. Screening before and after the average calculation is the same as used for the seasonally averaged

5 latitude cross sections (see Sect. 3.2). Again, no degradation of the ACE-FTS and SMR data sets onto the vertical resolution of the MIPAS data set is considered.

4 Results

4.1 Profile-to-profile comparison

In Fig. 1 we show an example of the absolute δD (upper row), HDO (middle row) and H_2O (bottom row) profiles from the

10 coincident SMR and ACE-FTS v3.5 data sets derived according to Eqs.7 and 8. In the lower stratosphere the SMR δD values reach -750 per mille, while the values in the ACE-FTS profile only reach -550 per mille. Since the ACE-FTS values are in the range of expected δD values in the lower stratosphere (Moyer et al., 1996; Johnson et al., 2001; Nassar et al., 2007), the SMR values seem to be too low. A separate consideration of the profiles separately with regard to the processes contributing to the transport of water vapour into the stratosphere would yield different conclusions. The values of the SMR profile would

15 indicate a larger contribution of the slow ascent, while the ACE-FTS profile indicate a larger influence of ice-lofting related to convective processes. With increasing altitude, the bias between the data sets become smaller. The different gradients in δD suggest different interpretation of the contribution from methane oxidation. Close to 1 hPa the agreement between the two data sets looks quite good, however for the wrong reason since there are large biases in both HDO and H_2O. Over the entire altitude range we find a more or less constant bias in HDO, while in H_2O the agreement becomes better with decreasing altitude. In the

20 lower stratosphere the biases in δD arise from the biases in HDO. The corresponding figures for the comparisons of the other data sets are provided in the Supplement.

In contrast in Fig. 2, directly the biases derived from the profile-to-profile comparisons (according to the description in Sect.3.1) are shown and all data set combinations are considered. The upper row shows the results for δD, in the middle row the HDO results are considered, while in the lower row the focus is on H_2O. In the left panels the absolute biases are shown,

25 while the right panels show the relative biases. Table 2 summarises the characteristics of the comparisons between the different data sets in terms of their overlap period, covered latitude range, number of coincidences, and the average separation in time and space.

The comparisons are typically based on several thousand coincidences and cover almost all latitudes and a time period of multiple years. The only exceptions are the comparisons between the MIPAS and ACE-FTS data sets. Given that the MIPAS

30 data sets end in March 2004 and the ACE-FTS observations effectively start in the second half of February 2004, there is only a very limited overlap period. During the short overlap period, the majority of ACE-FTS observations occurred in March at northern polar latitudes. Overall, these comparisons cover only latitudes from 51°N to 83°N, and most of the coincidences are concentrated near 70°N. The number of coincident profiles vary between 300 and 400, depending upon which of the MIPAS

and ACE-FTS data sets are compared with each other. Also, the average separation in time and longitude are larger for the comparisons between the MIPAS and ACE-FTS data sets than for any of the other comparisons.

The largest deviations in δD are found in the comparisons to the SMR data set. Between the SMR and MIPAS data sets, the absolute bias ranges from -230 per mille near the 50 hPa region to almost 150 per mille in the 1 hPa region. In relative terms, the corresponding numbers are 40% at 100 hPa and -40% in the stratopause region. The biases change signs slightly below 4 hPa. The comparisons of the SMR data set with the coincident ACE-FTS profiles show similar biases to the ones described above, with a peak deviation of -250 per mille around 60 hPa. The biases decrease in size to -100 per mille at 20 hPa. Above 15 hPa, the biases to the ACE-FTS v3.5 data set are smaller than the bias to the ACE-FTS v2.2 data set. For example, at 4 hPa the biases amount to -55 per mille (15%) and -30 per mille (7%), respectively. Higher up, the ACE-FTS retrievals come closer to their upper limit and the biases to the SMR data set exhibit distinct variations. The four comparisons between the different MIPAS vs. ACE-FTS data sets show in general a good agreement in δD. The biases are typically within ± 30 per mille between about 100 hPa and 4 hPa, corresponding to biases within $\pm 10\%$ in relative terms. Below and above this altitude range, the biases deteriorate. Comparisons between data sets from the same instrument typically exhibit very small biases. Exceptions occur primarily towards the upper and lower boundaries of the data sets.

Close to 60 hPa, the HDO comparisons to the SMR data set exhibit biases of about -0.4 to -0.3 ppbv (-125% to -90%). Towards higher altitudes the biases decrease in size; in the altitude range between 10 and 0.1 hPa, the biases are within $\pm 25\%$. The comparisons between the ACE-FTS and MIPAS data sets show a good agreement with deviations in the range of $\pm 10\%$ in the lower and middle stratosphere. Biases with a larger size are only observed below 100 hPa and above 4 hPa. Also, for H_2O, the comparisons indicate very good agreement between the different MIPAS and ACE-FTS data sets (typically within $\pm 10\%$). The comparisons to the SMR data set show the best agreement between about 50 and 10 hPa, where the biases to the MIPAS and ACE-FTS data sets do not exceed -0.5 ppmv (-10%). Here, the biases typically have a larger size in the comparison to the MIPAS data sets than the ACE-FTS data sets. Higher up, the deviations increase to about -25%. In volume mixing ratio terms, this corresponds to biases between -2 and -1 ppmv. Again, the size of the deviations is larger in the comparisons to the MIPAS data sets.

4.2 Comparisons of seasonally averaged latitude cross-sections

In these sections, the comparison results from the seasonally averaged latitude cross sections are presented. They aim to provide an overview of how the data sets compare as function of time and space. We start with an overview of all altitudes (Fig. 3) followed by three examples focusing on cross sections at 100 hPa, 10 hPa and 1 hPa (Figs. 4 to 6). These three levels have been chosen in order to cover the atmosphere from the tropopause region to the stratopause. After that a summary of the results is presented (Figs. 7 to 9), as described in Sect. 3.2.2.

4.2.1 Examples

All altitudes

Figure 3 shows the latitude-pressure cross-sections of δD for JJA for all data sets considered in the comparisons. During this season PSCs occur over the Antarctic and a pronounced latitudinal gradient in δD can be expected due to atmospheric dynamics. On top of the climatological tropopause large depletion in δD is found in the MIPAS and ACE-FTS data sets, especially in the tropics and Antarctic region. An exception here is the SMR data set which shows the depletion in δD at a higher altitude. Since this is the lower limit of the SMR measurements, the structure in δD is not that pronounced and the derived values at these altitudes are more uncertain compared to the MIPAS and ACE-FTS measurements. In the Antarctic, in the PSC areas we expect at this time of the year an influence on δD due to phase transitions related to the formation of ice particles and removal of ice particles through sedimentation. The differences found between the MIPAS and ACE-FTS climatologies would lead however to different interpretations of PSC influence on δD. An increase in δD is found with increasing altitude and accompanied by a latitudinal variation. Higher δD values are found over the high latitudes because of "older air" which had more time for methane oxidation (Stiller et al., 2012; Haenel et al., 2015). Higher δD values are also seen in the Antarctic region in the middle and higher stratosphere due to downwelling of older air within the polar vortex. In the data sets there is a decrease in δD in the lower mesosphere, but with variations between the data sets. Larger differences are seen for the other seasons (see figures in the Supplement). These differences also would lead to different conclusions concerning the processes involved when considering the data sets individually. In the tropical upper troposphere there are large differences related to the lower limit of the MIPAS observations and the influence of clouds. Probably the ACE-FTS data sets show a more realistic and clearer picture of the δD distribution at these altitudes due to the fact that the ACE-FTS instrument can measure at lower altitudes compared to the MIPAS instrument.

100 hPa

Figure 4 shows the latitudinal cross sections of δD (left column), HDO (centre column), and H_2O (right column) for the 100 hPa level. The different rows consider the results for the different seasons MAM, JJA, SON, and DJF. Dashed lines indicate the standard error of the seasonal averages (see Eq. 12 and 14). 100 hPa corresponds to the level of the tropical tropopause where the temperature largely determines the transport rate of water vapour into the stratosphere. It is therefore an interesting altitude for comparisons. Both the MIPAS and ACE-FTS data sets cover this altitude, even though it is close to the lower limit of possible retrievals for both instruments. The SMR observations provide only HDO results at this altitude and therefore are not shown in this comparison.

The ACE-FTS data sets show typically the highest depletion in δD in the tropics. The v3.5 data set indicates consistently lower δD values than the v2.2 data set. Towards higher latitudes, the δD values generally increase. A pronounced decrease is observed in the Antarctic in JJA and SON. For the MIPAS data sets, the latitudinal structure in MAM resembles a "W", with a maximum in the inner tropics and minima at subtropical latitudes. This structure mainly originates from a pronounced deviation in H_2O. Polewards of 60°latitude, the δD values vary less. A similar picture is visible in DJF. Unlike in the ACE-FTS

data sets, the MIPAS data sets do not show any decrease in the Antarctic in JJA and SON. In both seasons, the MIPAS data sets indicate a maximum in the northern tropics and lower δD values towards the Arctic. In JJA, the δD values derived in the high Antarctic actually indicate the lowest depletion compared to all other latitudes. Overall, deviations of 100 per mille and beyond are observed occasionally. For HDO also differences are observed. The latitude dependence of the data sets is more

5   coherent than for δD, with some obvious exceptions observed both in the tropics and the higher latitudes. In JJA and SON, the ACE-FTS data sets show a distinct drop in the volume mixing ratios polewards of 60°S as well, which is not captured by the MIPAS data sets. Also, the latitude dependence is in general more consistent among the data sets for H_2O than for δD. The best agreement between the MIPAS and ACE-FTS data sets is found in SON. In particular, the MIPAS v20 data set shows a pronounced deviation at 30°S and 30°N in MAM and DJF. Like spikes, the amount of water vapour increases prominently by

10   0.5-1 ppmv in a small latitude range. This clearly influences the latitude dependence observed in δD, resulting in the "W"-shaped latitude dependence described above. A similar, but weaker, behaviour can be observed in the other seasons because the agreement with the other data sets is better, making it less obvious. In DJF, the MIPAS v5 data set shows a pronounced deviation from the ACE-FTS data sets in H_2O in the tropics and subtropics. The differences observed in the Antarctic in JJA and SON for δD and HDO can also be observed in H_2O, but are less pronounced. Be reminded that the MIPAS H_2O data sets

15   included here are specially retrieved versions with a lower vertical resolution adjusted to the HDO resolution, as described in Sect. 2.2.

10 hPa

At 10 hPa (Fig. 5), in the middle stratosphere, all three satellites provide data for comparisons. The situation is clearly different from that at 100 hPa. The overall latitude dependence in δD is smoother and more evenly distributed compared to 100 hPa. The

20   most pronounced difference in δD is that the SMR data set shows between 50 to 100 per mille more depletion than the MIPAS and ACE-FTS data sets, with little latitude dependence. In δD, the MIPAS and ACE-FTS data sets exhibit smaller deviations in some regions. Consistently, for all seasons, the ACE-FTS v2.2 data set shows less depletion in δD than the ACE-FTS v3.5 and the two MIPAS data sets for all latitudes and during all seasons. The difference between the ACE-FTS v2.2 and the MIPAS data sets is typically about 25 per mille in the tropical regions, and often increasing somewhat towards the middle latitudes.

25   Poleward of 60°S in SON the ACE-FTS v2.2 and v3.5 data sets show a distinct increase from -450 to -350 per mille. A corresponding feature is also visible in both HDO with increases from 1 to 1.25 ppbv and H_2O from 5.5 to 6.5 ppmv. This structure is not apparent in either the MIPAS or SMR data sets. For HDO, latitude dependence is very consistent among the data sets, as observed for δD. There is again a low bias in the SMR data set, which is approximately 0.15 ppbv and rather independent from season and latitude. In H_2O, such a difference is not obvious. Instead, the SMR data set agrees quantitatively

30   rather well with the ACE-FTS data sets. The MIPAS data sets typically exhibit the highest volume mixing ratios. Towards polar latitudes, this high bias often increases, occasionally exceeding 0.5 ppmv. In the tropics, the spread among the data sets varies between 0.2 and 0.5 ppmv, roughly.

1 hPa

In the stratopause region at $1\,hPa$ (Fig. 6), the volume mixing ratio of water vapour in the middle atmosphere reaches its maximum. Data for comparison at this altitude are available from the SMR and MIPAS data sets. The ACE-FTS v3.5 data set has very limited coverage of this altitude level and is thus omitted here. Similar to the previous altitudes, there are pronounced deviations between the data sets. In δD, the sign of the bias between the SMR and MIPAS data sets is reversed compared to the $10\,hPa$ level. The SMR data set shows typically $25\,per\,mille$ to $100\,per\,mille$ less depletion than the MIPAS data sets. The bias is more dependent on season than latitude. The largest deviations are visible in SON and DJF, while in JJA the bias is smallest. In JJA and SON, there is a tendency for the bias to be largest in the Arctic. The MIPAS v20 data set shows consistently larger δD values than the v5 data set, but the deviation is on the order of only a few $per\,mille$. The cross sections exhibit rather similar structures in both HDO and H_2O, but the SMR observations show lower volume mixing ratios than the MIPAS data sets for all seasons and latitudes. In HDO, the deviation varies between 0.1 and $0.2\,ppbv$ and for H_2O, between 0.7 and $2.0\,ppmv$. While for HDO the biases show only a small seasonal dependence, the biases in H_2O minimise in JJA. In general, the biases are a bit larger in the polar regions than in the tropics.

4.2.2   Summary of the seasonal comparisons

Figures 7 to 9 show summary plots for the comparisons of the seasonal cross sections (see Sect. 3.2.2) that were exemplarily described for three pressure levels in Fig. 4 to 6. The left column shows the biases between the data sets averaged over all latitudes, while the middle column shows the corresponding (de-biased) standard deviations. In the right column, the correlation coefficients between the cross sections are shown. Figure 7 presents the results for δD, Fig. 8 those for HDO, and Fig. 9 the results for H_2O, respectively. Each figure shows the four different seasons in different rows. The summary figures extend the results obtained from the profile-to-profile comparisons shown in Fig. 2. This is particularly valuable for the comparisons between the MIPAS and ACE-FTS data sets, which were only possible with a very limited temporal and spatial overlap in the profile-to-profile approach.

δD

In general, the δD biases among the different data sets averaged over all latitudes (Fig. 7) yield very similar results to those derived from the profile-to-profile comparisons (Fig. 2). Some prominent deviations are visible in the comparisons between the SMR and MIPAS data sets as well as in the comparisons between those from MIPAS and ACE-FTS. For the latter, this concerns primarily altitudes below $100\,hPa$. The profile-to-profile comparisons in Fig. 2 indicate a low bias of the MIPAS data sets, while here (Fig. 7) a clear high bias is found. For the comparison between the SMR and MIPAS data sets, the approaches yield differing results at $50\,hPa$ as well as in the upper stratosphere and lower mesosphere. At about $50\,hPa$, the negative biases are larger than in the profile-to-profile comparisons. The δD bias above $0.3\,hPa$ in Fig. 2 approaches zero, while here (Fig. 7) the biases continue to increase for all seasons except for JJA. In general, a clear seasonal dependence is observed in the upper stratosphere and lower mesosphere for the comparison between the SMR and MIPAS data sets. At $0.6\,hPa$, the δD biases are

roughly 75 per mille in MAM while in SON, DJF, and the profile-to-profile comparisons they amount to 125–150 per mille. In JJA, the biases are here around 40 per mille. During this season, the bias between the SMR and MIPAS v5 data sets switches signs at about 0.3 hPa. Slightly below 0.1 hPa it amounts to about −50 per mille. This behaviour is not observed during other seasons.

5   The de-biased standard deviations show a characteristic altitude dependence. Between 20 and 0.5 hPa, the deviations are typically smaller than 25 per mille. The lowest values occur in the comparisons among the data sets from the same instrument (typically less than 10 per mille). Below 20 hPa, the standard deviations increase significantly and occasionally exceed 100 per mille at altitudes lower than 100 hPa. Again, estimates for the comparisons among data sets from the same instrument are smaller, but they are still substantial. Above 0.5 hPa, the comparisons between the SMR and MIPAS data sets also exhibit a

10   pronounced increase in the de-biased standard deviations, maximising in MAM and SON. Values exceeding 100 per mille are also observed close to 0.1 hPa. The correlation coefficients reflect in many ways the picture found for the de-biased standard deviation. Very high correlation coefficients (close to 1) are found between about 20 and 2 hPa. Lower correlations are visible at altitudes where one of the data sets compared reaches one of its vertical boundaries, like for the SMR data set around 50 hPa, the ACE-FTS v2.2 data set around 4 hPa (especially in SON), or for the ACE-FTS v3.5 data set close to 1 hPa. The compar-

15   isons between the MIPAS and ACE-FTS data sets show a distinct reduction of the correlations in a layer around 130 hPa, where the values turn negative in all seasons. In the lower mesosphere (above 0.3 hPa), the correlation coefficients between the SMR and MIPAS data sets turn negative as well.

HDO

As for δD, the HDO bias results from the seasonal means averaged over all latitudes (Fig. 8) and the profile-to-profile com-

20   parisons (Fig. 2) are quite similar. Differences among these two approaches are visible again below 100 hPa concerning all comparisons. Higher up, above 1 hPa, the biases between the SMR and MIPAS data sets are rather consistent with altitude in JJA, while during the other seasons and in the profile-to-profile comparisons they approach zero. The de-biased standard deviations show a similar picture, qualitatively, to those from δD. Typically, the values in the middle and upper stratosphere are smaller than 0.07 ppbv. Below, the values increase again, in particular below 100 hPa (exceeding 0.5 ppbv). Also, in the

25   lower mesosphere, the comparisons between the SMR and MIPAS data sets exhibit increasing estimates. Close to 0.1 hPa, the standard deviations amount to almost 0.35 ppbv in MAM, 0.2 ppbv in SON, and around 0.15 ppbv in JJA and DJF. The correlation coefficients show the highest values also in the stratosphere, in particular in MAM, where values larger 0.9 are found between 80 and 1.5 hPa. In JJA and SON, a clear reduction (occasionally lower than 0.3) is found between 100 and 30 hPa in most comparisons (except between those data sets from the same instrument). This behaviour coincides with increases in the

30   de-biased standard deviations.

H$_2$O

The bias results for H$_2$O summarised in Fig. 9 resemble those from the profile-to-profile comparison shown in Fig. 2, similarly as for δD and HDO. Again, differences are visible below 100 hPa among the two comparison approaches. The comparisons

between the MIPAS and ACE-FTS data sets exhibit larger biases around 2 hPa. They amount to more than 0.5 ppmv (Fig. 9), while they do not exceed 0.3 ppmv in the profile-to-profile comparisons (Fig. 2). The biases relative to the SMR data set show a seasonal variation around 1 hPa and the smallest estimates are observed in JJA. The de-biased standard deviations are typically smaller than 0.25 ppmv between 20 and 2 hPa. In this altitude range, the values are around 0.1 ppmv for the comparison among the different data sets from the same instrument. Lower down, the de-biased standard deviations increase again. In JJA and SON, a local maximum is observed around 40 hPa, coinciding with a distinct local minimum in the correlation coefficients, which is also observed in the comparisons for HDO. In δD, this behaviour is not as obvious, indicating some cancelation of the problems in HDO and H_2O. At 200 hPa, the de-biased standard deviations exceed 1 ppmv. In the lower mesosphere, the increase is more moderate (up to 0.7 ppmv). The highest correlations in the latitudinal distribution are found in about the same altitude region as δD and HDO. A distinct local minimum (with negative values) is observed in SON between 1 and 0.3 hPa. Similarly, a local minimum is found around 100 hPa in MAM and DJF. Be reminded again that the MIPAS H_2O data sets are based on special retrievals, different to the nominal data sets (see Sect. 2.2).

4.3 Monthly averaged profiles in the tropics

Figure 10 shows the tropical (15°S to 15 °N) monthly mean profiles of δD (left column), HDO (middle column), and H_2O (right column) for February, April, August, and October (different rows). The ACE-FTS observations, which focus on middle and high latitudes, typically cover the tropics only during these months. For this reason, the monthly averaged data are shown since they give a more appropriate picture of the isotopic ratio and the corresponding water vapour profiles in this region instead of the seasonal averages used in the previous section. Our primary focus here is on the lower stratosphere. In this region, HDO and H_2O exhibit the tape recorder signal manifesting itself in additional extrema in the vertical distribution above the hygropause. For δD, the existence of a tape recorder is under debate, as described in the Introduction.

Lower stratosphere

In February, both MIPAS data sets capture the tape recorder structure in δD with the first minimum of -670 per mille near 100 hPa, followed by the maximum of -580 per mille propagating from the previous season at an altitude of 70 hPa. Above, the corresponding second minima from previous winter of -600 per mille at 35 hPa can be seen. In the HDO and H_2O data, however, the MIPAS data sets do not capture the tape recorder structure in the lower stratosphere. Instead, the minimum associated with the hygropause is rather wide, due to the rather low vertical resolution in the dry phase of the tape recorder. In δD the ACE-FTS v3.5 shows a clear first minimum of -700 per mille at an altitude of approximately 100 hPa. This minimum is wider in altitude than for the MIPAS data sets. Higher up there is no structure of any tape recorder, except a small tendency to another minimum visible around 20 hPa. This is slightly more obvious for the ACE-FTS v2.2 data set, with a weak maximum at about 40 hPa and a second minimum close to 30 hPa. The first minimum around 100 hPa indicates approximately 50 per mille less depletion than for the v3.5 data set. This behaviour is observed during all months considered here. In contrast to the wide minimum observed in the MIPAS data sets, the ACE-FTS HDO and H_2O data sets exhibit a distinct hygropause. A weak local maximum is found close to 50 hPa and a second minimum slightly above 30 hPa, both in HDO and H_2O. This structure is

most pronounced in H_2O for the v3.5 data set. The SMR δD data deviate from the other data sets, in particular in terms of the absolute values. The lower limit of this data set is close to 50 hPa. There, it seems to display the tape recorder maximum (around -700 per mille) and the second minimum is presumably located at 30 hPa (about -720 per mille). As seen in the previous comparisons, the SMR HDO data has a low bias of about 0.2 ppbv relative to the other data sets in the lower stratosphere.

5 At the lower limit, the mixing ratio approaches zero and actually turns negative. However, the tape recorder structure is well visible in these data. The altitudes of the maximum and second minimum are quite similar to those observed in the ACE-FTS data sets. In H_2O, these extrema are located slightly higher than in the ACE-FTS data.

In April, the first δD minimum is clearly located higher in the MIPAS (around 75 hPa) than in the ACE-FTS data sets, which show this at about the same altitude as in February. The MIPAS data sets show a rather constant isotopic ratio between 50 hPa

10 and 10 hPa. Weak tape recorder structures are visible in the ACE-FTS data sets, but the extrema are located higher in altitude for the v3.5 data set. The SMR data set shows pronounced structures with a maximum slightly below 30 hPa and a minimum at 20 hPa. In HDO and H_2O, the tape recorder structures are clearly visible in all data sets with differences in the absolute mixing ratios and the altitudes of the extrema. In particular, the maximum is located at a lower altitude in the MIPAS data sets. They show this at 50 hPa, while in the other data sets the maximum is located at about 40 hPa.

15 For δD, the situation in August is rather similar to that in April. In HDO, the deviations between the SMR and the other data sets are larger at the hygropause compared to April. Also, the hygropause is higher in altitude in the ACE-FTS than the MIPAS data sets. Likewise, the tape recorder structures in H_2O differ among the data sets. They are most pronounced in the SMR data set, but the maximum occurs at a lower altitude than in the other data sets. The second minimum has the highest location in the ACE-FTS data sets.

20 There are pronounced differences in δD between the MIPAS and ACE-FTS data sets in October in the lower stratosphere. Both HDO and H_2O differences contribute to the bias in δD. The first minimum is observed at slightly below 100 hPa in the ACE-FTS data sets while it is located much higher, i.e. around 50 hPa, in the MIPAS data sets. Above, clear tape recorder structures are visible in all data sets from both instruments, but the shift in altitude remains.

Altitudes above 10 hPa

25 Focusing on altitudes above 10 hPa, the SMR data set shows lower depletion in δD than the MIPAS data sets above about 3 hPa, as also observed in the seasonal comparisons. The biases vary from month to month and maximise in the lower mesosphere in February. The SMR and ACE-FTS v3.5 data sets show a good agreement between 3 and 1 hPa. Good agreement in HDO between the SMR and MIPAS data sets is observed in the altitude range from 1 to 0.2 hPa in February. Above 0.2 hPa, the SMR data set shows typically higher mixing ratios than the MIPAS data sets; elsewhere, the low bias already seen in the previous

30 figure (i.e. 2 and 8) is here also visible. Between 7 and 1.5 hPa the MIPAS data sets show slightly higher HDO volume mixing ratios than the ACE-FTS data sets. At the upper limit of the ACE-FTS v3.5 data set, i.e. close to 0.9 hPa, it shows higher mixing ratios of 0.1 to 0.15 ppbv (largest in February and October) relative to the MIPAS data sets. In terms of H_2O, the data from the three instruments start to deviate above 10 hPa. The lowest mixing ratios are observed for the SMR data set. Up to about 0.5 hPa, the ACE-FTS data sets show lower mixing ratios than the MIPAS data sets (up to 0.75 ppmv), while above the

picture is reversed. The altitude where the middle atmospheric water vapour maximum is found differs among the data sets from the three instruments. At 0.1 hPa, the MIPAS v5 data set shows a lower H_2O mixing ratio than the v20 data set during all months. Likewise, the ACE-FTS v2.2 data set shows a lower amount of H_2O than the v3.5 data set at this altitude.

5   Discussion and conclusion

We have assessed the quality of satellite $\delta D(H_2O)$ observations from the upper troposphere to the lower mesosphere using profile-to-profile and climatological comparisons. We find clear quantitative differences in the isotopic ratio. These quantitative differences are largest with comparisons to the SMR data set. The MIPAS and ACE-FTS data sets agree rather well, with exceptions close to the vertical limit where observations can be made by these instruments, i.e. below 100 hPa and in the upper stratosphere.

For the profile-to-profile comparisons (Fig. 2), in the lower stratosphere the SMR data set shows significantly higher δD depletions than the MIPAS and ACE-FTS data sets. However, we note that this is close to the SMR lower limit, which has larger uncertainties. The biases maximise in size close to 50 hPa and exceed -200 per mille. In the upper stratosphere and lower mesosphere, the picture is reversed, with the SMR data set instead showing less depletion than the MIPAS data sets. The biases switch signs close to 4 hPa. Above, the biases increase to almost 150 per mille in the lower mesosphere, where arguably the MIPAS data sets, close to their upper retrieval limit, also become more uncertain. Qualitatively, the same picture is found in the seasonal comparisons (averaged over all latitudes), with some quantitative variations among the seasons, in particular in the lower mesosphere. The δD biases of the SMR data set can be related to biases in both HDO and H_2O.

For HDO, the SMR data set exhibits a low bias (Figs. 2 and 8) throughout all seasons. In the lower stratosphere, the biases are larger than 0.25 ppbv and then decrease with increasing altitude to become close to zero at 0.2 hPa. For H_2O, on the other hand, the SMR data set compares rather well with the MIPAS and ACE-FTS data sets in the lower stratosphere (Figs. 2 and 9), while in the upper stratosphere and lower mesosphere, distinct low biases are visible. The maximum H_2O bias for the SMR comparisons in the upper stratosphere varies between 1 and 2 ppmv, depending on season. Overall, the δD biases of the SMR data set are driven by HDO in the lower stratosphere, while H_2O is the clear driver in the upper stratosphere and lower mesosphere. In between, in the middle stratosphere, the biases in δD are a combination of deviations in both HDO and H_2O. Biases in both HDO and H_2O exist around 4 hPa where the δD biases of the SMR data set are close to zero compared to the MIPAS and ACE-FTS data sets, because the HDO and H_2O biases cancel each other out.

5.1   SMR results

The biases in the SMR data set can be attributed to both, uncertainties in the spectroscopic parameters used for the retrieval (they are larger than in the infrared region where MIPAS and ACE-FTS perform measurements) as well as to issues in the instrument calibration. In terms of spectroscopic parameters uncertainties in the line broadening, the temperature dependence exponent and the line intensity are of interest (Janssen, 1993). As described by Lossow et al. (2011) a 5% uncertainty in the line broadening parameter of the HDO emission line translates to biases of up to 0.05 ppbv in the retrieved data. For the

H_2O emission line the same uncertainty can cause biases in the retrieval results of up to 0.7 ppmv in the lower stratosphere. Around 5 hPa only small biases are visible while in the lower mesosphere the uncertainty amounts to biases between 0.3 and 0.4 ppmv. Combined to δD, the uncertainty of the line broadening parameters corresponds to biases of 30 per mille in the lower stratosphere and to 60 to 70 per mille in the lower mesosphere. For the temperature dependence exponent a 10% uncertainty for the HDO emission line results in biases of 0.01 ppbv in the retrieved results. In terms of H_2O the biases are about 0.1 ppmv, except in the lower stratosphere where they can be as large as 0.2 ppmv. Combining these results for δD yields biases smaller than 20 per mille for the assumed uncertainty in the temperature dependence exponent. For these two spectroscopic parameters higher uncertainties cannot be ruled out, increasing the bias estimates provided here. With regard to the line intensity the uncertainty is about 2% for the HDO and H_2O emission lines. This translates to biases of 0.025 ppbv for HDO and 0.1 ppmv for H_2O. In terms of δD these biases actually cancel out each other.

Issues with instrument calibration relate primarily to the sideband filtering, given that SMR is operated as a single sideband receiver (Frisk et al., 2003; Lossow et al., 2007). If nominal settings would be achieved side band leakages smaller than 2% are expected. The latest leakage estimates vary up to 5% and exhibit some time dependence. In principle, the maximum suppression of the sideband filter seems to not have changed significantly. Instead the filter reference length is the largest source of concern. This reference length controls at which frequencies the sideband filtering is optimal and those have apparently shifted over time. The uncertainty of the filter reference length is estimated to be 4 microns. For HDO this uncertainty corresponds to biases of up to 0.07 ppbv in the lower stratosphere, clearly being an important contribution. Higher up, the influence is only 0.01 ppbv. For H_2O the uncertainty of the filter reference length results in biases of up to 1 ppmv in the lower stratosphere. Around 10 hPa the influence is less than 0.2 ppmv but increases above up to 0.5 ppmv in the lower mesosphere. In terms of δD the uncertainty of the filter reference length translates to biases of 30 per mille in the lower stratosphere and 50 per mille in the lower mesosphere. Around 7 hPa a negligible influence is observed.

The uncertainties in the sideband leakage are associated with the prominent positive drifts which are observed in the SMR H_2O results (Khosrawi et al., 2018). Thus, over time, the low H_2O biases in this altitude range decrease and likewise the biases in δD. This has an impact on the results from the profile-to-profile and seasonal comparisons relative to MIPAS. While the profile-to-profile comparisons cover only the years from 2002 to 2004, the climatological comparisons include SMR data until 2009. In MAM and JJA in particular, reduced biases of δD and H_2O are visible in the upper stratosphere and lower mesosphere in the climatological comparisons relative to the profile-to-profile comparisons.

Besides the sideband filtering, there are also uncertainties arising from the autocorrelator calibration. They are part of ongoing studies. Overall, the uncertainties in the spectroscopic parameters and in the instrument calibration have the potential to explain the observed biases of the SMR data. In the lower mesosphere parts of the observed biases may actually be attributed to the MIPAS data sets, which are here also close to the upper vertical limit where retrievals are possible.

5.2 MIPAS and ACE-FTS results

The profile-to-profile comparisons (Fig. 2) between the MIPAS and ACE-FTS data sets exhibit the best agreement in the 100 to 5 hPa region. Most of the time the MIPAS data sets show a small high bias in δD, up to about 25 per mille (10%). Also

for the HDO and H_2O data, the agreement in the same altitude region is rather good. But due to the short overlap time and the small latitude coverage, these results have limited value. The climatological latitude cross sections include the entire data sets and thus allow a more conclusive picture. In these comparisons (Fig. 7), the MIPAS and ACE-FTS δD data sets agree best between 100 and 10 hPa. Even though the biases are low in this region, the de-biased standard deviations roughly amount up

5 to 50 per mille (larger in JJA). This indicates that there are biases in the latitudinal distribution between the data sets which cancel out in the average. Also, smaller biases exist in HDO (roughly around 0.1 ppbv) and H_2O (up to about 0.5 ppmv, larger in DJF), in particular in the lower part of this altitude range. They partly cancel out each other in the calculation of the δD ratio and in the average over all latitudes, and therefore result in a relatively good agreement in δD.

Above 10 hPa, the δD biases between the MIPAS and ACE-FTS data sets increase, in particular in the comparisons to the

10 ACE-FTS v2.2 data set. Here, the MIPAS data sets show a higher depletion. The de-biased standard deviations are, however, relatively low (largely within 20 per mille) and the correlation coefficients are relatively high, indicating only small variations over all latitudes. These δD biases are primarily driven by high biases in the MIPAS H_2O data (up to 0.8 ppmv) peaking at about 2 hPa. Below 100 hPa, the MIPAS and ACE-FTS data sets also show larger deviations. The bias estimates from the seasonal (Fig. 7) and profile-to-profile comparisons (Fig. 2) differ significantly. In the profile-to-profile comparisons, the

15 MIPAS δD data indicate a 100 per mille (20%) higher depletion than the ACE-FTS δD close to 200 hPa; in the climatological comparisons, on the other hand, the result is the opposite and the MIPAS data sets show significantly less depletion at 200 hPa. In both comparisons the biases are driven by differences in HDO and H_2O. In the profile-to-profile comparisons, the HDO biases are about 0.3 – 0.4 ppbv (60 – 90%), for H_2O they are within $\pm25\%$ at this altitude. For the climatological comparisons the HDO bias at 200 hPa is in the range of 0.25–1 ppbv depending on season, largest bias in MAM and least in SON and DJF.

20 The reasons for the biases between the MIPAS and ACE-FTS data sets are manifold: differences in temporal and spatial sampling, cloud influence, vertical resolution, the choice of microwindows and spectroscopic databases as well as a less-than perfect MIPAS H_2O product used here. This product is retrieved jointly with HDO in the attempt to match the vertical resolution of the latter (see Sect. 2.2). The differences in the temporal and spatial sampling between the MIPAS and ACE-FTS observations can affect the results of climatological comparisons. This concerns, on one hand, the different measurement

25 periods with the MIPAS observations providing data from July 2002 until March 2004 and the ACE-FTS observations starting in February 2004 and lasting for many years.

In this context, changes in variability, like trends, QBO influence, and the occurrence of sudden stratospheric warmings between these different time periods play a role. For example, according to the SWOOSH (Stratospheric Water and Ozone Satellite Homogenized, Davis et al., 2016) database, H_2O has increased from the MIPAS to the ACE-FTS observation period

30 at almost all latitudes and stratospheric altitudes. This does not help the discussion of biases among the data sets, as the MIPAS data exhibit higher H_2O mixing ratios in most of the stratosphere, in particular around 50 and 2 hPa (Fig. 9). On the other hand, the sampling issue concerns the actual sampling within a given time and latitude bin. The MIPAS satellite utilises a sun-synchronous orbit that provides coverage of almost all latitudes on a daily basis. In contrast, the latitudinal coverage of ACE-FTS observations varies throughout the year. The SCISAT orbit is optimised for high and middle latitudes, yielding

35 the bulk of observations in these regions. In this regard, the MIPAS observations provide a rather complete coverage of the

individual seasons in the tropics. For the ACE-FTS orbit, the seasonal means for MAM are based on observations in April, JJA on observations in August, SON on observations in October, and DJF on observations in February, respectively. This may explain some of the tropical biases observed in Figs. 4 and 5 and is more appropriately handled in the monthly comparisons in Sect. 4.3.

5 Besides the actual sampling, the variability within a given time and latitude bin plays a decisive role for the sampling bias (Toohey et al., 2013). One prominent example for this is the wintertime lower stratosphere in the Antarctic that is influenced by dehydration induced by PSCs. Such dehydration indications are visible in all data sets in both HDO and H_2O, but with various strengths. It is smallest in the MIPAS data and largest in the ACE-FTS (and SMR) data sets. This behaviour can be explained by sampling biases in combination with large variability. While the MIPAS data cover all winter months in the Antarctic, the

10 ACE-FTS coverage depends on the actual month. In June and July, the ACE-FTS observations reach only up to $60°S$ and $70°S$, respectively. It is only in August that the observations cover latitudes polewards of $70°S$, where the influence of PSCs and the associated dehydration is substantial, leading to a pronounced signal in the seasonal average for the ACE-FTS data sets. The other explanation for the different representation of dehydration in the data sets relates to differences in cloud influence on the MIPAS and ACE-FTS observations, which actually is another aspect of a sampling bias. While the observations of the

15 two instruments are performed in the infrared, they use different measurement techniques. The MIPAS instrument measured thermal emission at the atmospheric limb. The solar occultation technique used by the ACE-FTS instrument provides a stronger signal, which results in a lower sensitivity to cloud presence compared to the MIPAS observations (in comparison, the SMR observations exhibit the smallest cloud influence, as they are performed in the microwave region). In the presence of PSCs, many MIPAS observations actually have to be screened (see Section 2.2), weakening the dehydration signal. The other region

20 where these differences in the cloud influence between the MIPAS and ACE-FTS data sets are important is in the troposphere. Again, more MIPAS observations are screened, which can be expected to result in a dry bias. This is observed in H_2O (Fig. 9), but not in HDO (Fig. 8).

 In addition to this basic cloud influence, the behaviour of HDO and H_2O differs in the presence of clouds due to differences in the vapour pressure. Consequently, deviations in the depletion can be assumed, given the differences in the cloud influence

25 between the MIPAS and ACE-FTS data sets. For example, if there are no clouds present, the dehydration near the tropical tropopause should more reflect Rayleigh fractionation and result in isotopic ratios close to $-900\,per\,mille$. In the presence of clouds, the depletion should be lower. Given that the MIPAS data sets represent more cloud-free conditions, a higher depletion relative to the ACE-FTS data sets is expected but not apparent (Fig. 10). This may relate to differences in the vertical resolution among the data sets. As described in Sect. 3.1.2, the differences in the vertical resolution of the MIPAS and ACE-FTS data

30 sets can cause biases if they are not considered as with the climatological comparisons. Around the hygropause, the vertical resolution of the MIPAS data is roughly between 5 and $6\,km$, while that of the ACE-FTS data sets is between about 2 and $4\,km$. This degrades the comparisons around the hygropause and the upper troposphere compared to the core region of the stratosphere. It offers also a potential explanation why the profile-to-profile comparisons yield a better agreement, in particular in the upper troposphere where the differences in the vertical resolution among the data sets were considered, even though

35 the overlap is limited. The H_2O bias between the MIPAS and ACE-FTS data sets, peaking at $2\,hPa$ in the climatological

comparisons, is also influenced by differences in the vertical resolution among these data sets, in particular in the upper part. This part is close to the altitude where the middle atmospheric maximum in the water vapour is typically found (roughly at 0.8 hPa or 50 km). The influence is roughly of the order of 0.1 ppmv. However, since the MIPAS data sets used in these comparisons are worse resolved than the nominal MIPAS H_2O product (Schieferdecker et al., 2015; Lossow et al., 2017) and

5 than the ACE-FTS data sets, they should exhibit a low bias opposite to what is observed. Hence, it can be expected that the actual biases are even larger. In addition, there are differences in the vertical resolution of the MIPAS HDO and H_2O data sets. These affect the resulting δD values and consequently cause biases, varying from tens of per mille around the hygropause to a negligible impact in the middle and upper stratosphere.

Even though the MIPAS and ACE-FTS observations are both performed in the infrared region, there are differences in the
10 microwindows that are used to retrieve the HDO and H_2O information as described in Sects. 2.2 and 2.3. These differences are larger for H_2O than for HDO. The MIPAS retrievals for this special H_2O product use the same microwindows as for HDO, i.e. between 6.7 and 8.0 µm, while the ACE-FTS H_2O retrievals employ information from a much wider spectral region. In addition, there is also a small difference in the spectroscopic databases used in the retrieval. The MIPAS retrievals employ a special version of the HITRAN-2000 spectroscopic database, while the ACE-FTS retrievals use HITRAN-2004 with some
15 updates. For HDO, this difference in the spectroscopic database explains a high bias of up to 0.01 ppbv for the MIPAS data sets (Lossow et al., 2011). In terms of H_2O the different choice of the spectroscopic database results in a low bias of 0.2 ppmv to 0.3 ppmv for the MIPAS data sets at altitudes below 10 hPa. Above 10 hPa, the influence of the different spectroscopic databases gets smaller with increasing altitude and becomes zero close to 0.1 hPa. This means that this aspect is not a valid explanation for the high biases around 50 and 2 hPa observed in the MIPAS data sets. Here, the different spectral information
20 used in the MIPAS and ACE-FTS H_2O retrievals may serve as a potential solution for the observed biases. For δD, the different choice of the spectroscopic database yields a high bias for the MIPAS data sets between 15 and 30 per mille. Such biases are indeed often observed in the lower part of the stratosphere (see Figs. 2 and 7), but in the upper part of the stratosphere the picture is typically reversed.

Beyond that, a characteristic of the MIPAS retrieval is responsible for the obvious problems of these data (in particular for
25 v20) in the subtropics shown in Fig. 4. This is due to the handling of the a priori data in the retrieval, which here switches as function of latitude.

The δD comparisons of data sets from the same instrument typically show smaller deviations than those of data sets from other instruments. The largest biases are observed again below 100 hPa and stem from both HDO and H_2O. Between the MIPAS data sets, more pronounced differences are also visible in the lower mesosphere. These biases are due to deviations in
30 H_2O. As described in Sect. 2.2, the MIPAS data sets differ only in their calibration, while the retrieval itself was not changed. In contrast, the differences between the ACE-FTS data sets are due to changes in the instrument characterisation, choice of microwindows, and the temperature-pressure retrieval, as detailed by Waymark et al. (2013).

5.3 Comparisons in the tropical lower stratosphere

Finally, the tropical comparisons exhibit in all data sets structures characteristic of a tape recorder in δD, HDO, and H_2O. However, there are clear quantitative differences in the isotopic/mixing ratios as well as the altitudes and the magnitude of the structures. Overall, the consistency in HDO and H_2O is better than in δD. In February, the MIPAS data sets show a wide hygropause in HDO and H_2O. This is a consequence of the low temperatures in the TTL during this period, resulting in low emissions and a reduced vertical resolution of the retrieved data relative to the other months. Yet in δD, the first maximum appears well resolved. This might be a result of the different vertical resolutions of the MIPAS HDO and H_2O data, or simply a favourable interference (by coincidence) of HDO and H_2O characteristics. Differences in the vertical resolution between the MIPAS and ACE-FTS data certainly contribute to the altitude differences of the tape recorder structures. However, those are also visible among the ACE-FTS data sets that have the same resolution but differ in retrieval details and the time period considered. In addition, the tape recorder structures appear to be more pronounced in the SMR and MIPAS data sets compared to the ACE-FTS data sets. This relates to the differences in the δD tape recorder amplitude between the MIPAS and ACE-FTS data sets as derived from the previous studies by Steinwagner et al. (2010) and Randel et al. (2012). We will address this further in a subsequent study (Lossow et al., in preparation).

5.4 Concluding remarks

There are pronounced δD differences among the different data sets in key areas of scientific interest. Individual analyses of the data sets will draw different conclusions concerning atmospheric processes while their combination yield results with little significance and large uncertainties. This concerns for example the absolute amount of δD, its seasonal variation and vertical structure in the tropical lower stratosphere (relative importance of different processes for the transport of water vapour into the stratosphere), the vertical gradient in the stratosphere (role of methane oxidation) or the influence of clouds on δD (phase transitions and transport). Some of these difference issues we can resolve as the absolute amount of δD in the tropical lower stratosphere (too large depletion in the SMR data set), the vertical gradient in the stratosphere (too large gradient in the SMR data set) or influence of clouds (MIPAS data sets need to screen more cloud influenced observations and thus cover a different atmospheric scenario). The other issues remain part of active research.

The reliability of the different data sets will besides the aspects mentioned so far, principally depend on the type of application or the time, latitude and altitude of the application. Given the data currently available, fewer systematic errors were identified for the ACE-FTS data sets. Version 3.5 is an improvement over version 2.2 and covers a larger altitude range and time period. For the MIPAS data sets the differences in vertical resolution of the HDO and H_2O products used for the δD calculation cause the largest uncertainty (the products for itself are fine in principle). The largest issues occur for the SMR data set, primarily related to the characterisation of the sideband filtering and uncertainties in spectroscopic parameters. Despite this, the SMR data set shows a good consistency in e.g. the latitudinal cross sections and can correspondingly be used in applications where this aspect is relevant.

Data availability. The MIPAS data are available on the following website: https://www.imk-asf.kit.edu/english/308.php. The SMR data can be accessed on the following website: http://amazonite.rss.chalmers.se:8280/OdinSMR/searchl2. The ACE-TFS data can be downloaded from the following website: https://databace.scisat.ca.

Author contributions. The study was designed by CH and SL with contributions from the WAVAS-II core members GPS, KAW and FK. CH analysed the data and wrote the manuscript with contributions from SL and FK. Satellite data used in this study were provided by KAW, RB, PE, DPM and GPS. All co-authors helped with the interpretation of the results and provided valuable comments on the manuscript.

Competing interests. The authors declare that they have no conflict of interests.

Acknowledgements. This work is funded by the SNSB project Dnr 88/11 "Atmospheric modelling using space-based observations of stable water isotopes". Odin is a Swedish-led satellite project funded jointly by the Swedish National Space Board (SNSB), the Canadian Space Agency (CSA), the National Technology Agency of Finland (Tekes), and the Centre National d'Etudes Spatiales (CNES) in France. The Swedish Space Corporation has been the prime industrial constructor. Since April 2007, Odin has been a third-party mission of ESA. We would like to thank the European Space Agency (ESA) for making the MIPAS level-1b data set available. The Atmospheric Chemistry Experiment (ACE), also known as SCISAT, is a Canadian-led mission mainly supported by the Canadian Space Agency (CSA) and the Natural Sciences and Engineering Research Council of Canada (NSERC). C. Högberg thanks the Bolin Centre for Climate Research for financial support to visit KIT, Karlsruhe to work on this project. S. Lossow was funded by the DFG Research Unit "Stratospheric Change and its Role for Climate Prediction" (SHARP) under contract STI 210/9-2. He thanks the International Meteorological Institute (IMI) hosted by the Department of Meteorology at Stockholm University for financing multiple visits to Stockholm to work on this topic. We would like to thank the anonymous referee, David Johnson and the editor Karen Rosenlof for helpful comments on the manuscript. We want to express our gratitude to SPARC and WCRP (World Climate Research Programme) for their guidance, sponsorship, and support of the WAVAS-II programme.

References

Bernath, P. F., McElroy, C. T., Abrams, M. C., Boone, C. D., Butler, M., Camy-Peyret, C., Carleer, M., Clerbaux, C., Coheur, P.-F., Colin, R., DeCola, P., DeMazière, M., Drummond, J. R., Dufour, D., Evans, W. F. J., Fast, H., Fussen, D., Gilbert, K., Jennings, D. E., Llewellyn, E. J., Lowe, R. P., Mahieu, E., McConnell, J. C., McHugh, M., McLeod, S. D., Michaud, R., Midwinter, C., Nassar, R., Nichitiu, F.,

5   Nowlan, C., Rinsland, C. P., Rochon, Y. J., Rowlands, N., Semeniuk, K., Simon, P., Skelton, R., Sloan, J. J., Soucy, M.-A., Strong, K., Tremblay, P., Turnbull, D., Walker, K. A., Walkty, I., Wardle, D. A., Wehrle, V., Zander, R., and Zou, J.: Atmospheric Chemistry Experiment (ACE): Mission overview, Geophysical Research Letters, 32, L15S01, https://doi.org/10.1029/2005GL022386, 2005.

Boone, C. D., Nassar, R., Walker, K. A., Rochon, Y., McLeod, S. D., Rinsland, C. P., and Bernath, P. F.: Retrievals for the Atmospheric Chemistry Experiment Fourier-Transform Spectrometer, Applied Optics, 44, 7218 – 7231, 2005.

10   Boone, C. D., Walker, K. A., and Bernath, P. F.: Version 3 retrievals for the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS), pp. 103 – 127, The Atmospheric Chemistry Experiment ACE at 10: A Solar Occultation Anthology, A. Deepak Publishing, Hampton, Virginia, USA, 2013.

Brasseur, G. and Solomon, S.: Aeronomy of the middle atmosphere, Springer, ISBN-10 1-4020-3284-6, P.O. Box 17, 3300 AA Dordrecht, The Netherlands, 2005.

15   Brewer, A. W.: Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere, Quarterly Journal of the Royal Meteorological Society, 75, 351 – 363, https://doi.org/10.1002/qj.49707532603, 1949.

Carleer, M. R., Boone, C. D., Walker, K. A., Bernath, P. F., Strong, K., Sica, R. J., Randall, C. E., Vömel, H., Kar, J., Höpfner, M., Milz, M., von Clarmann, T., Kivi, R., Valverde-Canossa, J., Sioris, C. E., Izawa, M. R. M., Dupuy, E., McElroy, C. T., Drummond, J. R., Nowlan, C. R., Zou, J., Nichitiu, F., Lossow, S., Urban, J., Murtagh, D. P., and Dufour, D. G.: Validation of water vapour profiles from the

20   Atmospheric Chemistry Experiment (ACE), Atmospheric Chemistry & Physics Discussions, 8, 4499 – 4559, 2008.

Connor, B. J., Siskind, D. E., Tsou, J. J., Parrish, A., and Remsberg, E. E.: Ground-based microwave observations of ozone in the upper stratosphere and mesosphere, Journal of Geophysical Research, 99, 16 757 – 16 770, https://doi.org/10.1029/94JD01153, 1994.

Davis, S. M., Rosenlof, K. H., Hassler, B., Hurst, D. F., Read, W. G., Vömel, H., Selkirk, H., Fujiwara, M., and Damadeo, R.: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: a long-term database for climate studies, Earth System Science

25   Data, 8, 461 – 490, https://doi.org/10.5194/essd-8-461-2016, 2016.

Dessler, A. E., Schoeberl, M. R., Wang, T., Davis, S. M., and Rosenlof, K. H.: Stratospheric water vapor feedback, Proceedings of the National Academy of Science, 110, 18 087 – 18 091, https://doi.org/10.1073/pnas.1310344110, 2013.

Dupuy, E., Walker, K. A., Kar, J., Boone, C. D., McElroy, C. T., Bernath, P. F., Drummond, J. R., Skelton, R., McLeod, S. D., Hughes, R. C., Nowlan, C. R., Dufour, D. G., Zou, J., Nichitiu, F., Strong, K., Baron, P., Bevilacqua, R. M., Blumenstock, T., Bodeker, G. E., Borsdorff,

30   T., Bourassa, A. E., Bovensmann, H., Boyd, I. S., Bracher, A., Brogniez, C., Burrows, J. P., Catoire, V., Ceccherini, S., Chabrillat, S., Christensen, T., Coffey, M. T., Cortesi, U., Davies, J., de Clercq, C., Degenstein, D. A., de Mazière, M., Demoulin, P., Dodion, J., Firanski, B., Fischer, H., Forbes, G., Froidevaux, L., Fussen, D., Gerard, P., Godin-Beekmann, S., Goutail, F., Granville, J., Griffith, D., Haley, C. S., Hannigan, J. W., Höpfner, M., Jin, J. J., Jones, A., Jones, N. B., Jucks, K., Kagawa, A., Kasai, Y., Kerzenmacher, T. E., Kleinböhl, A., Klekociuk, A. R., Kramer, I., Küllmann, H., Kuttippurath, J., Kyrölä, E., Lambert, J., Livesey, N. J., Llewellyn, E. J., Lloyd, N. D., Mahieu,

35   E., Manney, G. L., Marshall, B. T., McConnell, J. C., McCormick, M. P., McDermid, I. S., McHugh, M., McLinden, C. A., Mellqvist, J., Mizutani, K., Murayama, Y., Murtagh, D. P., Oelhaf, H., Parrish, A., Petelina, S. V., Piccolo, C., Pommereau, J., Randall, C. E., Robert, C., Roth, C., Schneider, M., Senten, C., Steck, T., Strandberg, A., Strawbridge, K. B., Sussmann, R., Swart, D. P. J., Tarasick, D. W., Taylor,

J. R., Tétard, C., Thomason, L. W., Thompson, A. M., Tully, M. B., Urban, J., Vanhellemont, F., Vigouroux, C., von Clarmann, T., von der Gathen, P., von Savigny, C., Waters, J. W., Witte, J. C., Wolff, M., and Zawodny, J. M.: Validation of ozone measurements from the Atmospheric Chemistry Experiment (ACE), Atmospheric Chemistry & Physics, 9, 287 – 343, https://doi.org/10.5194/acp-9-287-2009, 2009.

5 Eichinger, R., Jöckel, P., and Lossow, S.: Simulation of the isotopic composition of stratospheric water vapour – Part 2: Investigation of HDO / H_2O variations, Atmospheric Chemistry & Physics, 15, 7003 – 7015, https://doi.org/10.5194/acp-15-7003-2015, 2015.

Eriksson, P.: Microwave radiometric observations of the middle atmosphere: Simulations and inversions, Ph.D. thesis, Chalmers University of Technology, School of Electrical and Computational Engineering, Göteburg, Sweden, 1999.

Fischer, H., Birk, M., Blom, C., Carli, B., Carlotti, M., von Clarmann, T., Delbouille, L., Dudhia, A., Ehhalt, D., Endemann, M., Flaud, J. M.,
10 Gessner, R., Kleinert, A., Koopman, R., Langen, J., López-Puertas, M., Mosner, P., Nett, H., Oelhaf, H., Perron, G., Remedios, J., Ridolfi, M., Stiller, G. P., and Zander, R.: MIPAS: An instrument for atmospheric and climate research, Atmospheric Chemistry & Physics, 8, 2151 – 2188, https://doi.org/10.5194/acp-8-2151-2008, 2008.

Flaud, J. M., Piccolo, C., Carli, B., Perrin, A., Coudert, L. H., Teffo, J. L., and Brown, L. R.: Molecular line parameters for the MIPAS (Michelson Interferometer for Passive Atmospheric Sounding) experiment, Atmospheric and Oceanic Optics, 16, 172 – 182, 2003.

15 Frank, F., Jöckel, P., Gromov, S., and Dameris, M.: Investigating the yield of H_2O and H_2 from methane oxidation in the stratosphere, Atmospheric Chemistry & Physics, 18, 9955 – 9973, https://doi.org/10.5194/acp-18-9955-2018, 2018.

Frisk, U., Hagström, M., Ala-Laurinaho, J., Andersson, S., Berges, J.-C., Chabaud, J.-P., Dahlgren, M., Emrich, A., Florén, H.-G., Florin, G., Fredrixon, M., Gaier, T., Haas, R., Hirvonen, T., Hjalmarsson, Å., Jakobsson, B., Jukkala, P., Kildal, P. S., Kollberg, E., Lassing, J., Lecacheux, A., Lehikoinen, P., Lehto, A., Mallat, J., Marty, C., Michet, D., Narbonne, J., Nexon, M., Olberg, M., Olofsson, A. O. H., Olof-
20 sson, G., Origné, A., Petersson, M., Piironen, P., Pons, R., Pouliquen, D., Ristorcelli, I., Rosolen, C., Rouaix, G., Räisänen, A. V., Serra, G., Sjöberg, F., Stenmark, L., Torchinsky, S., Tuovinen, J., Ullberg, C., Vinterhav, E., Wadefalk, N., Zirath, H., Zimmermann, P., and Zimmermann, R.: The Odin satellite. I. Radiometer design and test, Astronomy and Astrophysics, 402, L27 – L34, https://doi.org/10.1051/0004-6361:20030335, 2003.

Fueglistaler, S., Dessler, A. E., Dunkerton, T. J., Folkins, I., Fu, Q., and Mote, P. W.: Tropical tropopause layer, Reviews of Geophysics, 47,
25 RG1004, https://doi.org/10.1029/2008RG000267, 2009.

Gettelman, A., Birner, T., Eyring, V., Akiyoshi, H., Bekki, S., Brühl, C., Dameris, M., Kinnison, D. E., Lefevre, F., Lott, F., Mancini, E., Pitari, G., Plummer, D. A., Rozanov, E., Shibata, K., Stenke, A., Struthers, H., and Tian, W.: The tropical tropopause layer 1960 – 2100, Atmospheric Chemistry & Physics, 9, 1621 – 1637, 2009.

Haefele, A., Hocke, K., Kämpfer, N., Keckhut, P., Marchand, M., Bekki, S., Morel, B., Egorova, T., and Rozanov, E.: Diurnal changes in
30 middle atmospheric H_2O and O_3: Observations in the Alpine region and climate models, Journal of Geophysical Research, 113, D17 303, https://doi.org/10.1029/2008JD009892, 2008.

Haenel, F. J., Stiller, G. P., von Clarmann, T., Funke, B., Eckert, E., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., and Reddmann, T.: Reassessment of MIPAS age of air trends and variability, Atmospheric Chemistry & Physics, 15, 13 161–13 176, https://doi.org/https://doi.org/10.5194/acp-15-13161-2015, 2015, 2015.

35 Hagemann, R., Nief, G., and Roth, E.: Absolute isotopic scale for deuterium analysis natural waters. Absolute D/H ratio for SMOW, Tellus, 22, 712 – 715, 1970.

Hegglin, M. I., Tegtmeier, S., Anderson, J., Froidevaux, L., Fuller, R., Funke, B., Jones, A., Lingenfelser, G., Lumpe, J., Pendlebury, D., Remsberg, E., Rozanov, A., Toohey, M., Urban, J., Clarmann, T., Walker, K. A., Wang, R., and Weigel, K.: SPARC Data Initia-

tive: Comparison of water vapor climatologies from international satellite limb sounders, Journal of Geophysical Research, 118, 11 824, https://doi.org/10.1002/jgrd.50752, 2013.

Janssen, M. A.: Atmospheric remote sensing by microwave radiometry, ISBN 0-471-62891-3, Wiley-Interscience Publication, 1993.

Johnson, D. G., Jucks, K. W., Traub, W. A., and Chance, K. V.: Isotopic composition of stratospheric water vapor: Implications for transport, Journal of Geophysical Research, 106, 12 219 – 12 226, https://doi.org/10.1029/2000JD900764, 2001.

Jones, A., Urban, J., Murtagh, D. P., Eriksson, P., Brohede, S., Haley, C., Degenstein, D., Bourassa, A., von Savigny, C., Sonkaew, T., Rozanov, A., Bovensmann, H., and Burrows, J.: Evolution of stratospheric ozone and water vapour time series studied with satellite measurements, Atmospheric Chemistry & Physics, 9, 6055 – 6075, 2009.

Jones, A., Walker, K. A., Jin, J. J., Taylor, J. R., Boone, C. D., Bernath, P. F., Brohede, S., Manney, G. L., McLeod, S., Hughes, R., and Daffer, W. H.: Technical Note: A trace gas climatology derived from the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) data set, Atmospheric Chemistry & Physics, 12, 5207 – 5220, https://doi.org/10.5194/acp-12-5207-2012, 2012.

Kelly, K. K., Tuck, A. F., Murphy, D. M., Proffitt, M. H., Fahey, D. W., Jones, R. L., McKenna, D. S., Loewenstein, M., Podolske, J. R., Strahan, S. E., Ferry, G. V., Chan, K. R., Vedder, J. F., Gregory, G. L., Hypes, W. D., McCormick, M. P., Browell, E. V., and Heidt, L. E.: Dehydration in the lower Antarctic stratosphere during late winter and early spring, 1987, Journal of Geophysical Research, 94, 11 317 – 11 357, https://doi.org/10.1029/JD094iD09p11317, 1989.

Khosrawi, F., Lossow, S., Stiller, G. P., Rosenlof, K. H., Urban, J., Burrows, J. P., Damadeo, R. P., Eriksson, P., García-Comas, M., Gille, J. C., Kasai, Y., Kiefer, M., Nedoluha, G. E., Noël, S., Raspollini, P., Read, W. G., Rozanov, A., Sioris, C. E., Walker, K. A., and Weigel, K.: The SPARC water vapour assessment II: comparison of stratospheric and lower mesospheric water vapour time series observed from satellites, Atmospheric Measurement Techniques, 11, 4435 – 4463, https://doi.org/10.5194/amt-11-4435-2018, 2018.

Kley, D., Russell, J. M., and Philips, C.: Stratospheric Processes and their Role in Climate (SPARC) - Assessment of upper tropospheric and stratospheric water vapour, SPARC Report 2, WMO/ICSU/IOC World Climate Research Programme, Geneva, 2000.

Le Texier, H., Solomon, S., and Garcia, R. R.: The role of molecular hydrogen and methane oxidation in the water vapour budget of the stratosphere, Quarterly Journal of the Royal Meteorological Society, 114, 281 – 295, https://doi.org/10.1002/qj.49711448002, 1988.

Lossow, S., Urban, J., Eriksson, P., Murtagh, D. P., and Gumbel, J.: Critical parameters for the retrieval of mesospheric water vapour and temperature from Odin/SMR limb measurements at 557 GHz, Advances in Space Research, 40, 835 – 845, 2007.

Lossow, S., Urban, J., Schmidt, H., Marsh, D. R., Gumbel, J., Eriksson, P., and Murtagh, D. P.: Wintertime water vapor in the polar upper mesosphere and lower thermosphere: First satellite observations by Odin submillimeter radiometer, Journal of Geophysical Research, 114, D10 304, https://doi.org/10.1029/2008JD011462, 2009.

Lossow, S., Steinwagner, J., Urban, J., Dupuy, E., Boone, C. D., Kellmann, S., Linden, A., Kiefer, M., Grabowski, U., Höpfner, M., Glatthor, N., Röckmann, T., Murtagh, D. P., Walker, K. A., Bernath, P. F., von Clarmann, T., and Stiller, G. P.: Comparison of HDO measurements from Envisat/MIPAS with observations by Odin/SMR and SCISAT/ACE-FTS, Atmospheric Measurement Techniques, 4, 1855 – 1874, https://doi.org/10.5194/amt-4-1855-2011, 2011.

Lossow, S., Khosrawi, F., Nedoluha, G. E., Azam, F., Bramstedt, K., Burrows, J. P., Dinelli, B. M., Eriksson, P., Espy, P. J., García-Comas, M., Gille, J. C., Kiefer, M., Noël, S., Raspollini, P., Read, W. G., Rosenlof, K. H., Rozanov, A., Sioris, C. E., Stiller, G. P., Walker, K. A., and Weigel, K.: The SPARC water vapour assessment II: comparison of annual, semi-annual and quasi-biennial variations in stratospheric and lower mesospheric water vapour observed from satellites, Atmospheric Measurement Techniques, 10, 1111 – 1137, https://doi.org/10.5194/amt-10-1111-2017, 2017.

Lossow, S., Khosrawi, F., Kiefer, M., Walker, K. A., Bertaux, J.-L., Blanot, L., Russell, J. M., Remsberg, E. E., Gille, J. C., Sugita, T., Sioris, C. E., Dinelli, B. M., Papandrea, E., Raspollini, P., García-Comas, M., Stiller, G. P., von Clarmann, T., Dudhia, A., Read, W. G., Nedoluha, G. E., Damadeo, R. P., Zawodny, J. M., Weigel, K., Rozanov, A., Azam, F., Bramstedt, K., Noël, S., Burrows, J. P., Sagawa, H., Kasai, Y., Urban, J., Eriksson, P., Murtagh, D. P., Hervig, M., Högberg, C., Hurst, D. F., and Rosenlof, K. H.: The SPARC water vapour assessment II: Profile-to-profile comparisons of stratospheric and lower mesospheric water vapour data sets obtained from satellites, in preparation.

Lossow, S., Högberg, C., Khosrawi, F., Stiller, G. P., Bauer, R., Walker, K. A., Eriksson, P., Murtagh, D. P., Eichinger, R., Kellmann, S., Linden, A., Kiefer, M., Glatthor, N., von Clarmann, T., and Zhang, Q.: On the discrepancies in the annual variation of δD-H_2O in the tropical lower stratosphere observed between the MIPAS and ACE-FTS satellite data sets, in preparation.

Mote, P. W., Rosenlof, K. H., McIntyre, M. E., Carr, E. S., Gille, J. C., Holton, J. R., Kinnersley, J. S., Pumphrey, H. C., Russell, J. M., and Waters, J. W.: An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor, Journal of Geophysical Research, 101, 3989 – 4006, https://doi.org/10.1029/95JD03422, 1996.

Moyer, E. J., Irion, F. W., Yung, Y. L., and Gunson, M. R.: ATMOS stratospheric deuterated water and implications for troposphere-stratosphere transport, Geophysical Research Letters, 23, 2385 – 2388, https://doi.org/10.1029/96GL01489, 1996.

Nassar, R., Bernath, P. F., Boone, C. D., Gettelman, A., McLeod, S. D., and Rinsland, C. P.: Variability in HDO/H_2O abundance ratios in the tropical tropopause layer, Journal of Geophysical Research, 112, D21 305, https://doi.org/10.1029/2007JD008417, 2007.

Payne, V. H., Noone, D., Dudhia, A., Piccolo, C., and Grainger, R. G.: Global satellite measurements of HDO and implications for understanding the transport of water vapour into the stratosphere, Quarterly Journal of the Royal Meteorological Society, 133, 1459 – 1471, https://doi.org/10.1002/qj.127, 2007.

Randall, C. E., Rusch, D. W., Bevilacqua, R. M., Hoppel, K. W., Lumpe, J. D., Shettle, E., Thompson, E., Deaver, L., Zawodny, J., Kyrö, E., Johnson, B., Kelder, H., Dorokhov, V. M., König-Langlo, G., and Gil, M.: Validation of POAM III ozone: Comparisons with ozonesonde and satellite data, Journal of Geophysical Research, 108, 4367, https://doi.org/10.1029/2002JD002944, 2003.

Randel, W. J., Moyer, E., Park, M., Jensen, E., Bernath, P., Walker, K., and Boone, C.: Global variations of HDO and HDO/H_2O ratios in the upper troposphere and lower stratosphere derived from ACE-FTS satellite measurements, Journal of Geophysical Research, 117, D06 303, https://doi.org/10.1029/2011JD016632, 2012.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, Journal of Climate, 24, 3624 – 3648, https://doi.org/10.1175/JCLI-D-11-00015.1, 2011.

Risi, C., Noone, D., Worden, J., Frankenberg, C., Stiller, G., Kiefer, M., Funke, B., Walker, K., Bernath, P., Schneider, M., Wunch, D., Sherlock, V., Deutscher, N., Griffith, D., Wennberg, P. O., Strong, K., Smale, D., Mahieu, E., Barthlott, S., Hase, F., García, O., Notholt, J., Warneke, T., Toon, G., Sayres, D., Bony, S., Lee, J., Brown, D., Uemura, R., and Sturm, C.: Process-evaluation of tropospheric humidity simulated by general circulation models using water vapor isotopologues: 1. Comparison between models and observations, Journal of Geophysical Research, 117, D05 303, https://doi.org/10.1029/2011JD016621, 2012.

Rodgers, C. D.: Inverse methods for atmospheric soundings: Theory and practice, ISBN 981-02-2740-X, World Scientific Publishing Co. Pte. Ltd., 2000.

Rothman, L. S., Barbe, A., Benner, D. C., Brown, L. R., Camy-Peyret, C., Carleer, M. R., Chance, K., Clerbaux, C., Dana, V., Devi, V. M., Fayt, A., Flaud, J.-M., Gamache, R. R., Goldman, A., Jacquemart, D., Jucks, K. W., Lafferty, W. J., Mandin, J.-Y., Massie, S. T.,

Nemtchinov, V., Newnham, D. A., Perrin, A., Rinsland, C. P., Schroeder, J., Smith, K. M., Smith, M. A. H., Tang, K., Toth, R. A., Vander Auwera, J., Varanasi, P., and Yoshino, K.: The HITRAN molecular spectroscopic database: edition of 2000 including updates through 2001, Journal of Quantitative Spectroscopy and Radiative Transfer, 82, 5 – 44, https://doi.org/10.1016/S0022-4073(03)00146-8, 2003.

Rothman, L. S., Jacquemart, D., Barbe, A., Chris Benner, D., Birk, M., Brown, L. R., Carleer, M. R., Chackerian, C., Chance, K., Coudert, L. H., Dana, V., Devi, V. M., Flaud, J.-M., Gamache, R. R., Goldman, A., Hartmann, J.-M., Jucks, K. W., Maki, A. G., Mandin, J.-Y., Massie, S. T., Orphal, J., Perrin, A., Rinsland, C. P., Smith, M. A. H., Tennyson, J., Tolchenov, R. N., Toth, R. A., Vander Auwera, J., Varanasi, P., and Wagner, G.: The HITRAN 2004 molecular spectroscopic database, Journal of Quantitative Spectroscopy and Radiative Transfer, 96, 139 – 204, https://doi.org/10.1016/j.jqsrt.2004.10.008, 2005.

Sayres, D. S., Pfister, L., Hanisco, T. F., Moyer, E. J., Smith, J. B., St. Clair, J. M., O'Brien, A. S., Witinski, M. F., Legg, M., and Anderson, J. G.: Influence of convection on the water isotopic composition of the tropical tropopause layer and tropical stratosphere, Journal of Geophysical Research, 115, D00J20, https://doi.org/10.1029/2009JD013100, 2010.

Schieferdecker, T.: Variabilität von Wasserdampf in der unteren und mittleren Stratosphäre auf der Basis von HALOE/UARS und MIPAS/Envisat Beobachtungen, Ph.D. thesis, Karlsruhe Institute of Technology, 2015.

Schieferdecker, T., Lossow, S., Stiller, G. P., and von Clarmann, T.: Is there a solar signal in lower stratospheric water vapour?, Atmospheric Chemistry & Physics, 15, 9851 – 9863, https://doi.org/10.5194/acp-15-9851-2015, 2015.

Seele, C. and Hartogh, P.: Water vapor of the polar middle atmosphere: Annual variation and summer mesosphere conditions as observed by ground-based microwave spectroscopy, Geophysical Research Letters, 26, 1517 – 1520, https://doi.org/10.1029/1999GL900315, 1999.

Sheese, P. E., Boone, C. D., and Walker, K. A.: Detecting physically unrealistic outliers in ACE-FTS atmospheric measurements, Atmospheric Measurement Techniques, 8, 741 – 750, https://doi.org/10.5194/amt-8-741-2015, 2015.

Spang, R., Remedios, J. J., and Barkley, M. P.: Colour indices for the detection and differentiation of cloud types in infra-red limb emission spectra, Advances in Space Research, 33, 1041 – 1047, https://doi.org/10.1016/S0273-1177(03)00585-4, 2004.

Steinwagner, J., Milz, M., von Clarmann, T., Glatthor, N., Grabowski, U., Höpfner, M., Stiller, G. P., and Röckmann, T.: HDO measurements with MIPAS, Atmospheric Chemistry & Physics, 7, 2601 – 2615, https://doi.org/10.5194/acp-7-2601-2007, 2007.

Steinwagner, J., Fueglistaler, S., Stiller, G. P., von Clarmann, T., Kiefer, M., Borsboom, P., van Delden, A., and Röckmann, T.: Tropical dehydration processes constrained by the seasonality of stratospheric deuterated water, Nature Geoscience, 3, 262 – 266, https://doi.org/10.1038/ngeo822, 2010.

Stiller, G. P., von Clarmann, T., Haenel, F., Funke, B., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., Lossow, S., and López-Puertas, M.: Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period, Atmospheric Chemistry & Physics, 12, 3311 – 3331, https://doi.org/10.5194/acp-12-3311-2012, 2012.

Tikhonov, A. N. and Arsenin, V. Y.: Solutions of ill-posed problems, ISBN 0470991240, Winston Publishing, New York, 1977.

Toohey, M., Hegglin, M. I., Tegtmeier, S., Anderson, J., Añel, J. A., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Kasai, Y., Krüger, K., Kyrölä, E., Neu, J. L., Rozanov, A., Smith, L., Urban, J., Clarmann, T., Walker, K. A., and Wang, R. H. J.: Characterizing sampling biases in the trace gas climatologies of the SPARC Data Initiative, Journal of Geophysical Research, 118, 11 847 – 11 862, https://doi.org/10.1002/jgrd.50874, 2013.

Urban, J., Lautié, N., Murtagh, D. P., Eriksson, P., Kasai, Y., Lossow, S., Dupuy, E., de La Noë, J., Frisk, U., Olberg, M., Le Flochmoën, E., and Ricaud, P.: Global observations of middle atmospheric water vapour by the Odin satellite: An overview, Planetary and Space Science, 55, 1093 – 1102, https://doi.org/10.1016/j.pss.2006.11.021, 2007.

von Clarmann, T.: Validation of remotely sensed profiles of atmospheric state variables: strategies and terminology, Atmospheric Chemistry & Physics, 6, 4311 – 4320, https://doi.org/10.5194/acp-6-4311-2006, 2006.

Waymark, C., Walker, K. A., Boone, C. D., and Bernath, P. F.: ACE-FTS version 3.0 data set: Validation and data processing update, Annals of Geophysics, 56, https://doi.org/10.4401/ag-6339, 2013.

5   Webster, C. R. and Heymsfield, A. J.: Water Isotope Ratios D/H, $^{18}O/^{16}O$, $^{17}O/^{16}O$ in and out of clouds map dehydration pathways, Science, 302, 1742 – 1746, https://doi.org/10.1126/science.1089496, 2003.

Wrotny, J. E., Nedoluha, G. E., Boone, C., Stiller, G. P., and McCormack, J. P.: Total hydrogen budget of the equatorial upper stratosphere, Journal of Geophysical Research, 115, D04 302, https://doi.org/10.1029/2009JD012135, 2010.

Variable	Interval	Altitude range
δD	-10000‰– 10000‰ -1250‰– 250‰ -5000‰– 5000‰	1000 hPa – 70 hPa 70 hPa – 1 hPa 1 hPa – 0.1 hPa
HDO	-5 ppmv – 15 ppmv	70 hPa – 0.1 hPa
H_2O	-20 ppmv – 50 ppmv	70 hPa – 0.1 hPa

Table 1. Interval screening performed prior to the comparisons. Profiles that exhibited data points outside these intervals were discarded.

Comparison	Overlap period	Latitude range in deg	Number of coincidences	Average distance separation in km	Average time separation in h	Average latitude separation in deg	Average equivalent latitude separation in deg	Average longitude separation in deg
SMR v2.1 vs. MIPAS v5	Jul 2002 – Mar 2004	87°S – 89°N	11564 13107	296 296	12.1 12.1	1.3 1.3	1.9 1.9	3.5 3.5
SMR v2.1 vs. MIPAS v20	Jul 2002 – Mar 2004	87°S – 87°N	11840 13365	294 293	12.2 12.1	1.3 1.3	1.9 1.9	3.5 3.5
SMR v2.1 vs. ACE-FTS v2.2	Feb 2004 – May 2009	85°S – 86°N	2934 3146	489 492	11.8 11.8	2.3 2.3	2.1 2.2	9 8.8
SMR v2.1 vs. ACE-FTS v3.5	Feb 2004 – May 2009	85°S – 86°N	2972 3194	489 494	12 11.8	2.2 2.3	2.1 2.2	9.3 9.1
MIPAS v5 vs. MIPAS v20	Jul 2002 – Mar 2004	87°S – 89°N	451619	0	0	0	0	0
MIPAS v5 vs. ACE-FTS v2.2	Feb 2004 – Mar 2004	51°N – 83°N	384	552	17.5	2.2	2.2	16.5
MIPAS v5 vs. ACE-FTS v3.5	Feb 2004 – Mar 2004	51°N – 83°N	364	567	17.7	2.2	2.3	17.2
MIPAS v20 vs. ACE-FTS v2.2	Feb 2004 – Mar 2004	51°N – 83°N	373	577	17.7	2.2	2.3	18.1
MIPAS v20 vs. ACE-FTS v3.5	Feb 2004 – Mar 2004	51°N – 83°N	352	584	17.8	2.2	2.3	18.5
ACE-FTS v2.2 vs. ACE-FTS v3.5	Feb 2004 – Sep 2010	85°S – 87°N	19555	0	0	0	0	0

Table 2. Overview of the profile-to-profile comparisons and their coincidence characteristics. For comparisons to the SMR data set the characteristics vary slightly depending upon if the comparison considers HDO (upper row) or H_2O (lower row). The determination of the overlap period and latitude range uses the data from both data sets.
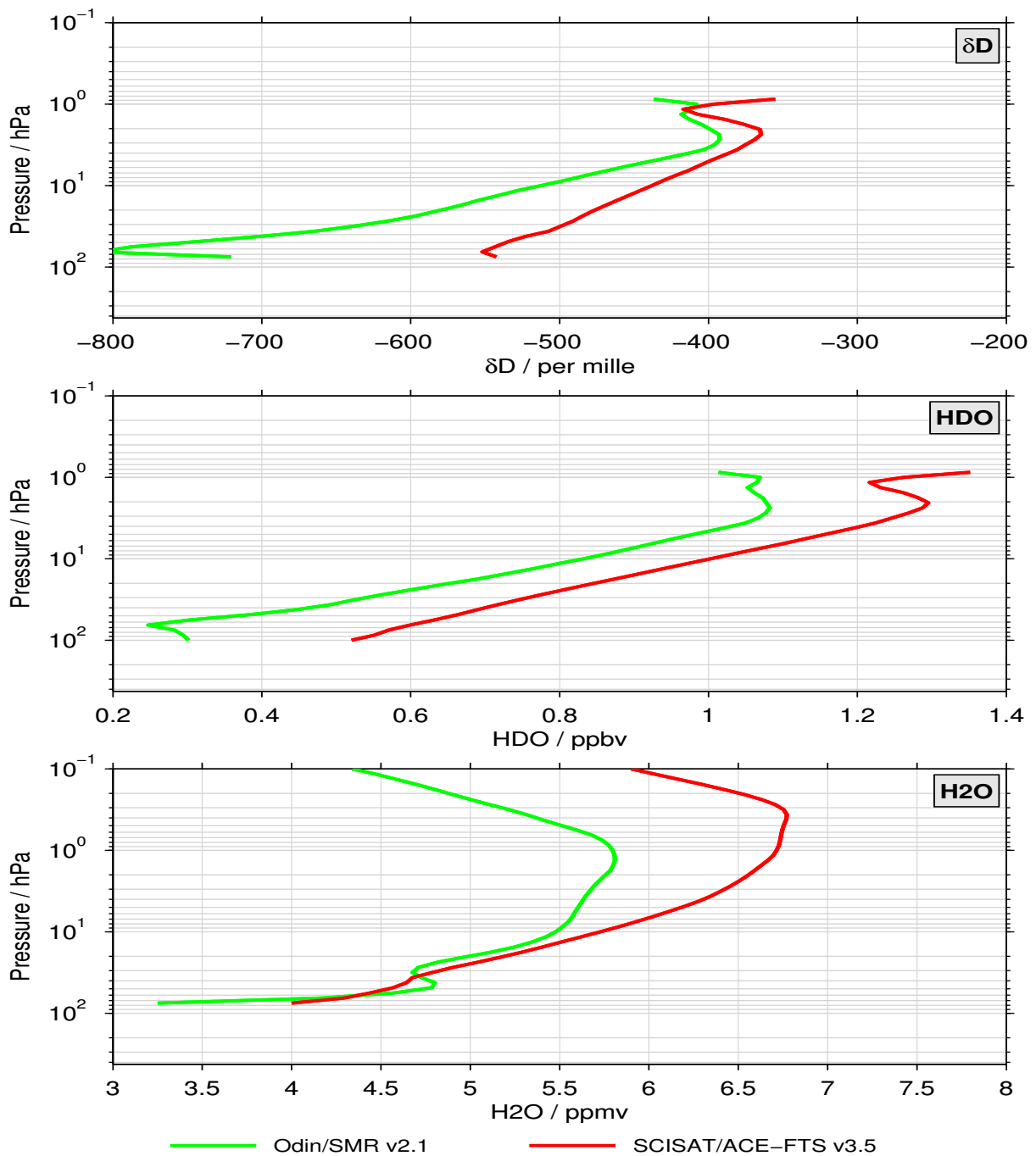
Figure 1. Profile-to-profile comparisons between the SMR v2.1 and ACE-FTS v3.5 data sets considering absolute profiles of δD (top panel), HDO (middle panel) and H_2O (lower panel).
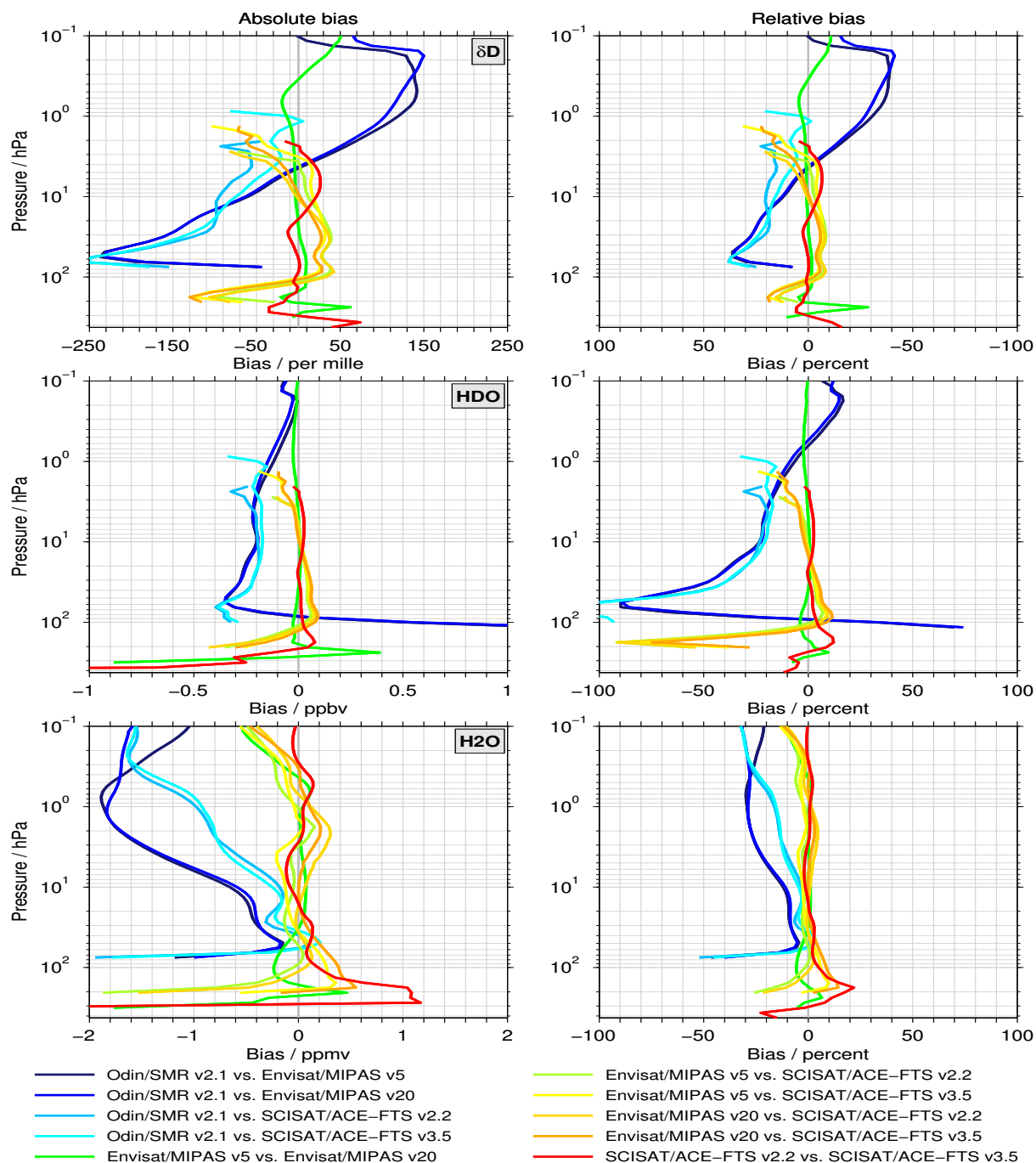
Figure 2. Profile-to-profile comparisons between all data sets for δD (top row), HDO (middle row) and H₂O (bottom row). The left panels show the biases in absolute terms, while in the right panel the relative biases are shown. Please note that for the relative bias in δD the x-axis has been reversed for visual consistency with the absolute bias.
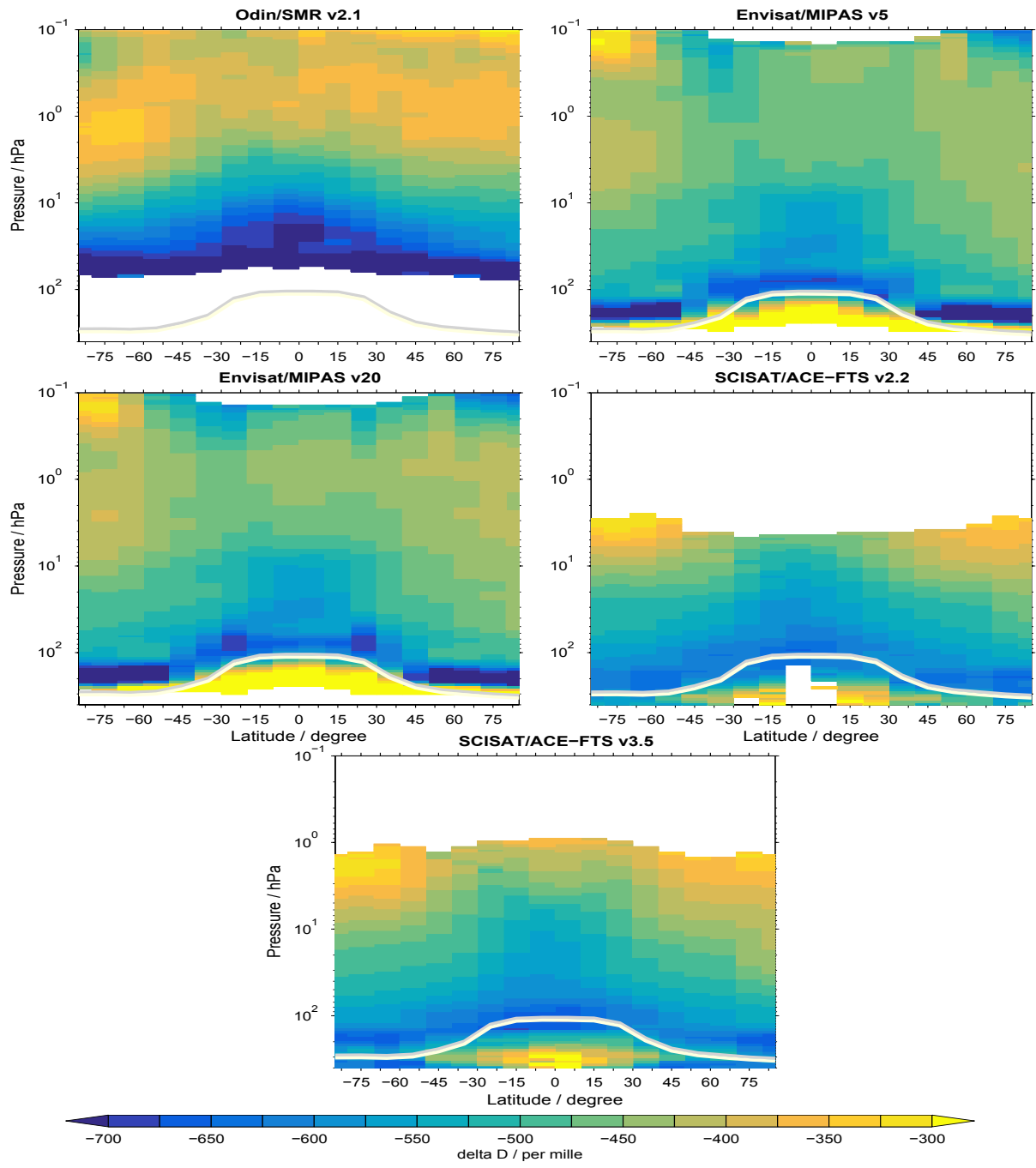
Figure 3. Climatologies of δD for the five data sets considered in this study and for the season JJA.
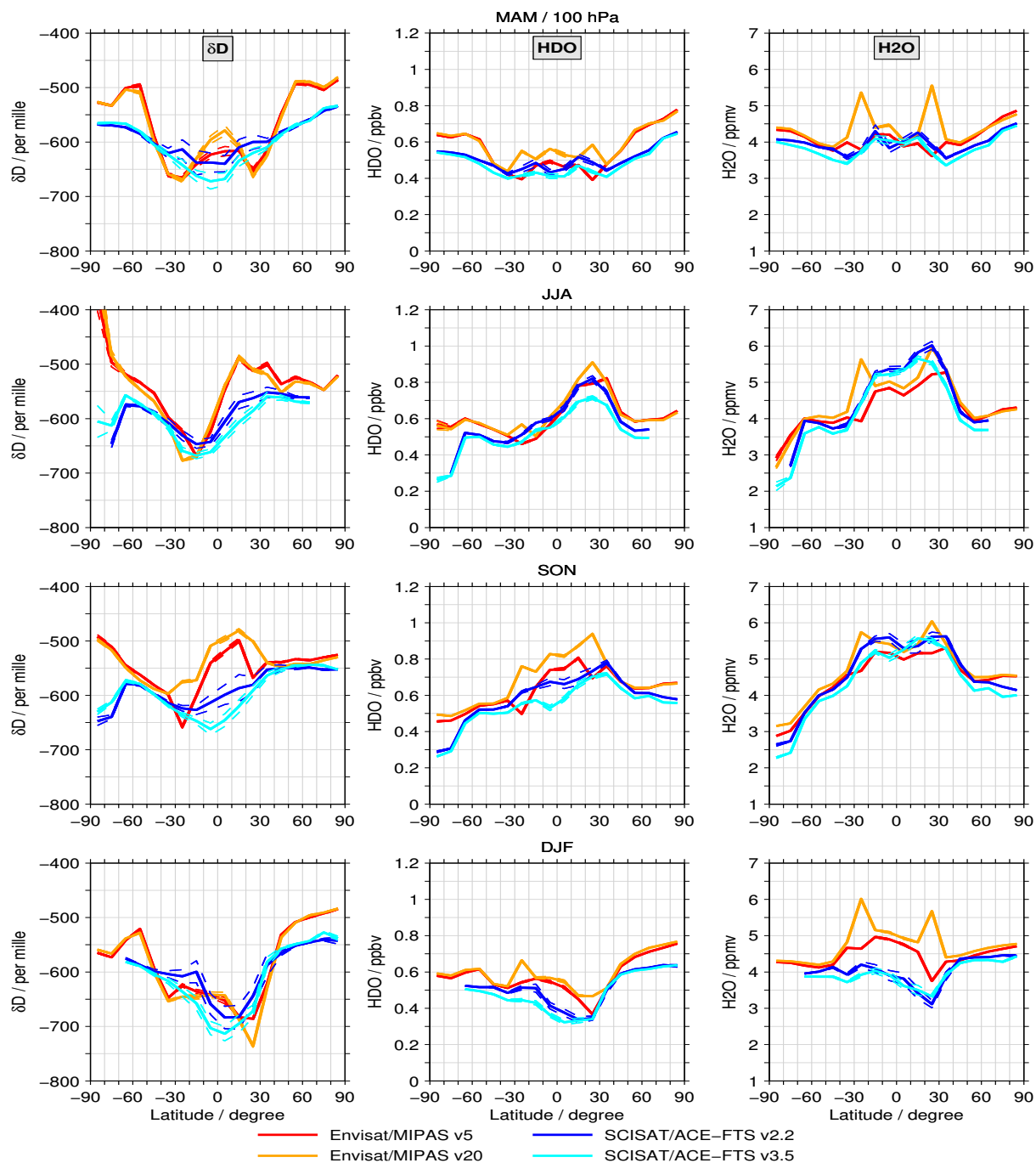
Figure 4. Latitude cross sections at 100 hPa from the different data sets for δD (left column), HDO (centre column) and H_2O (right column) for all seasons (different rows). The thin dashed line show the standard error of the binned data.
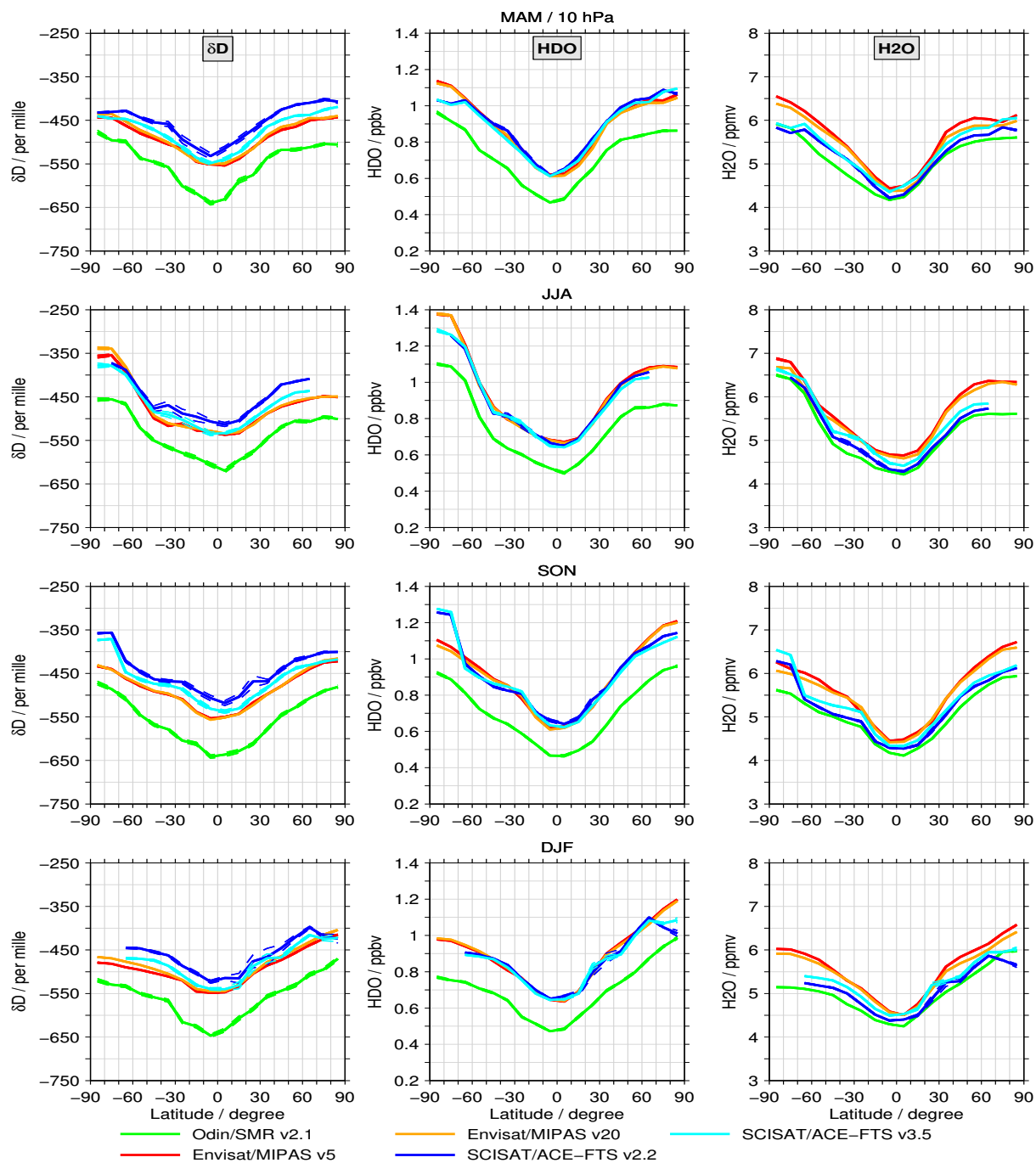
Figure 5. As Figure 4 but considering the 10 hPa level.
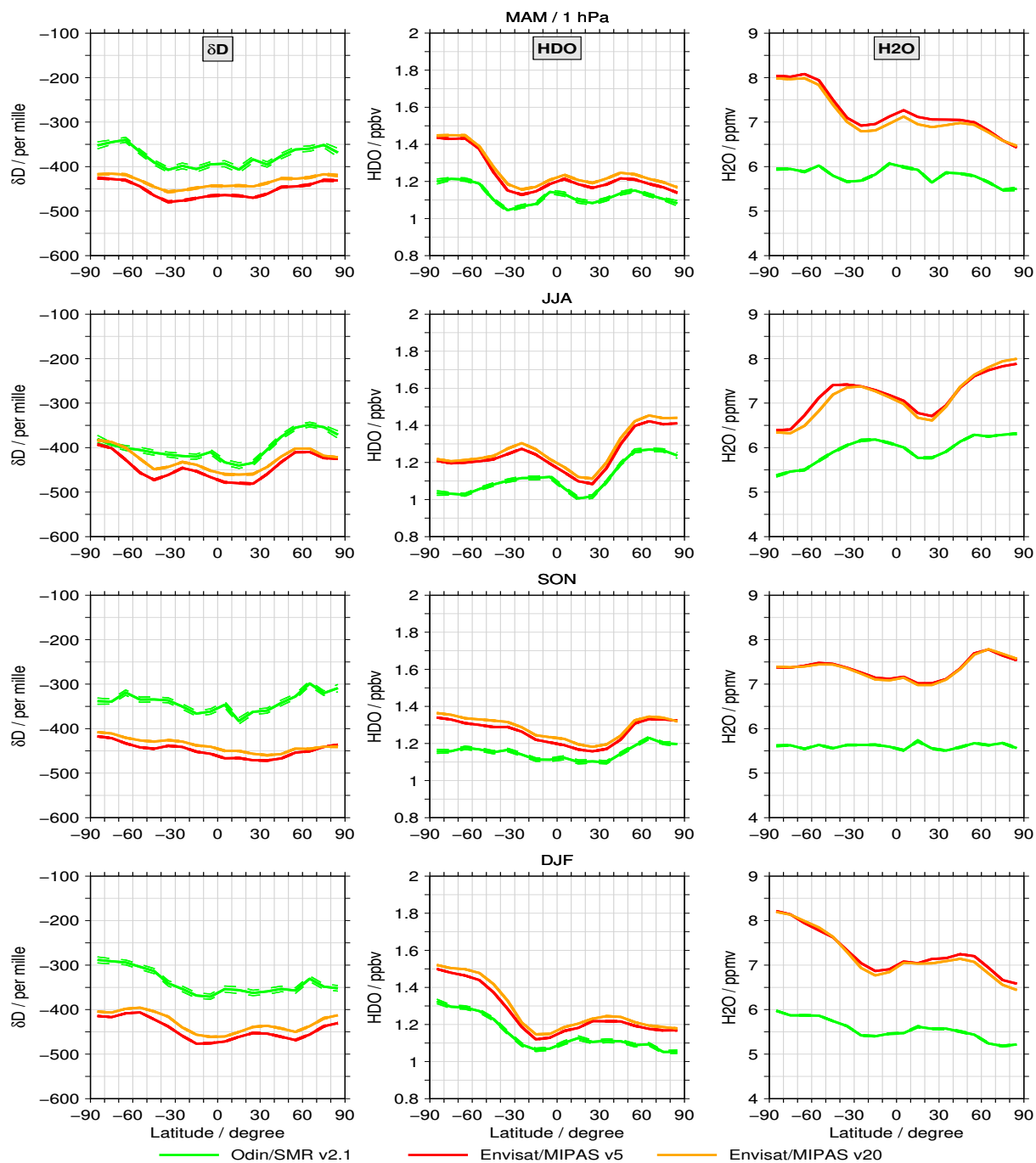
Figure 6. As Figures 4 and 5 but here for the 1 hPa level.
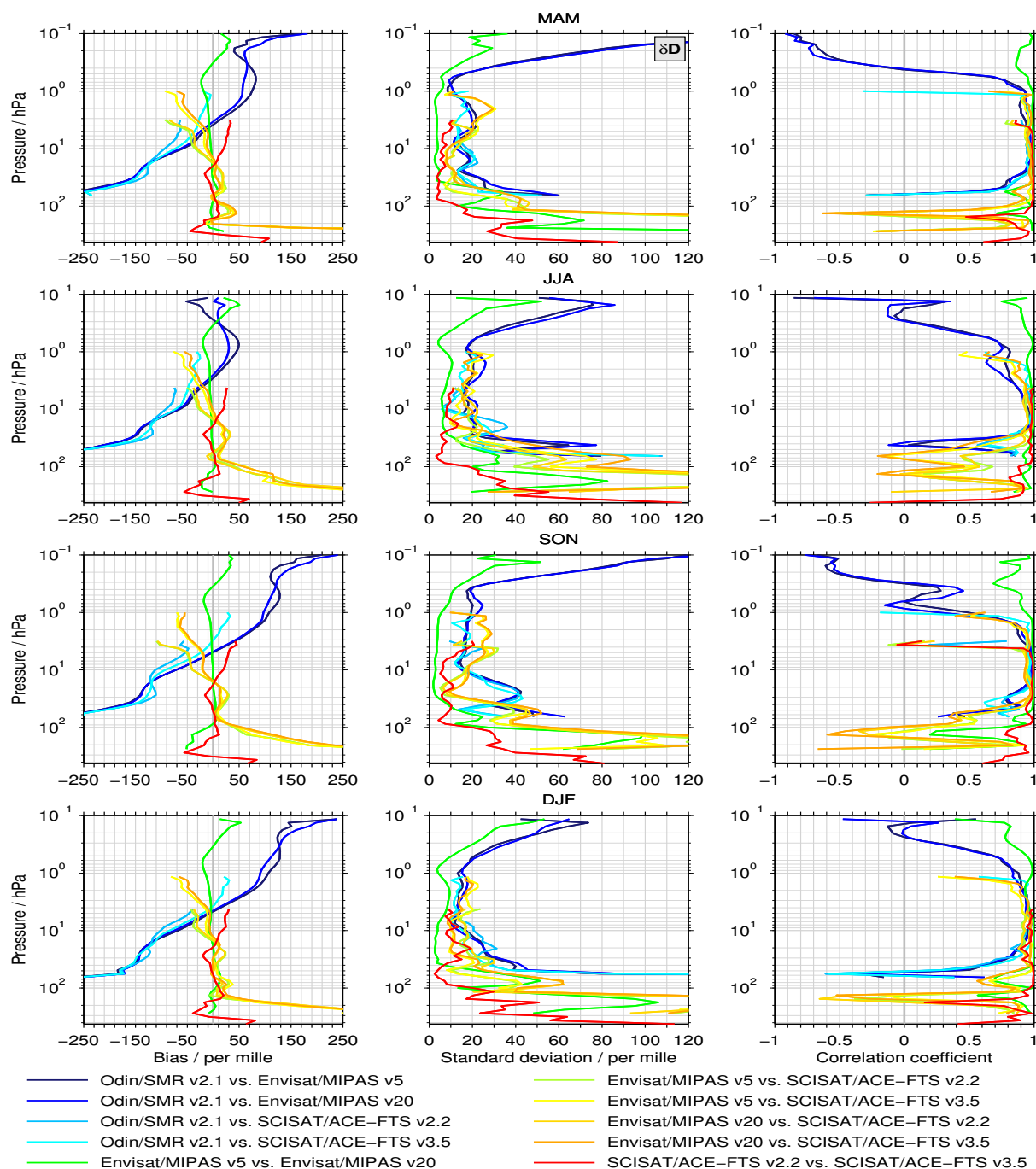
Figure 7. Summary of the seasonal comparisons that were shown exemplarily in Fig. 4 to 6. Here δD is considered. The left panels show the biases averaged over all latitudes for the individual seasons (different rows). The corresponding de-biased standard deviation are shown in the centre panels and in the right panels the correlations between the latitudinal cross sections are given.
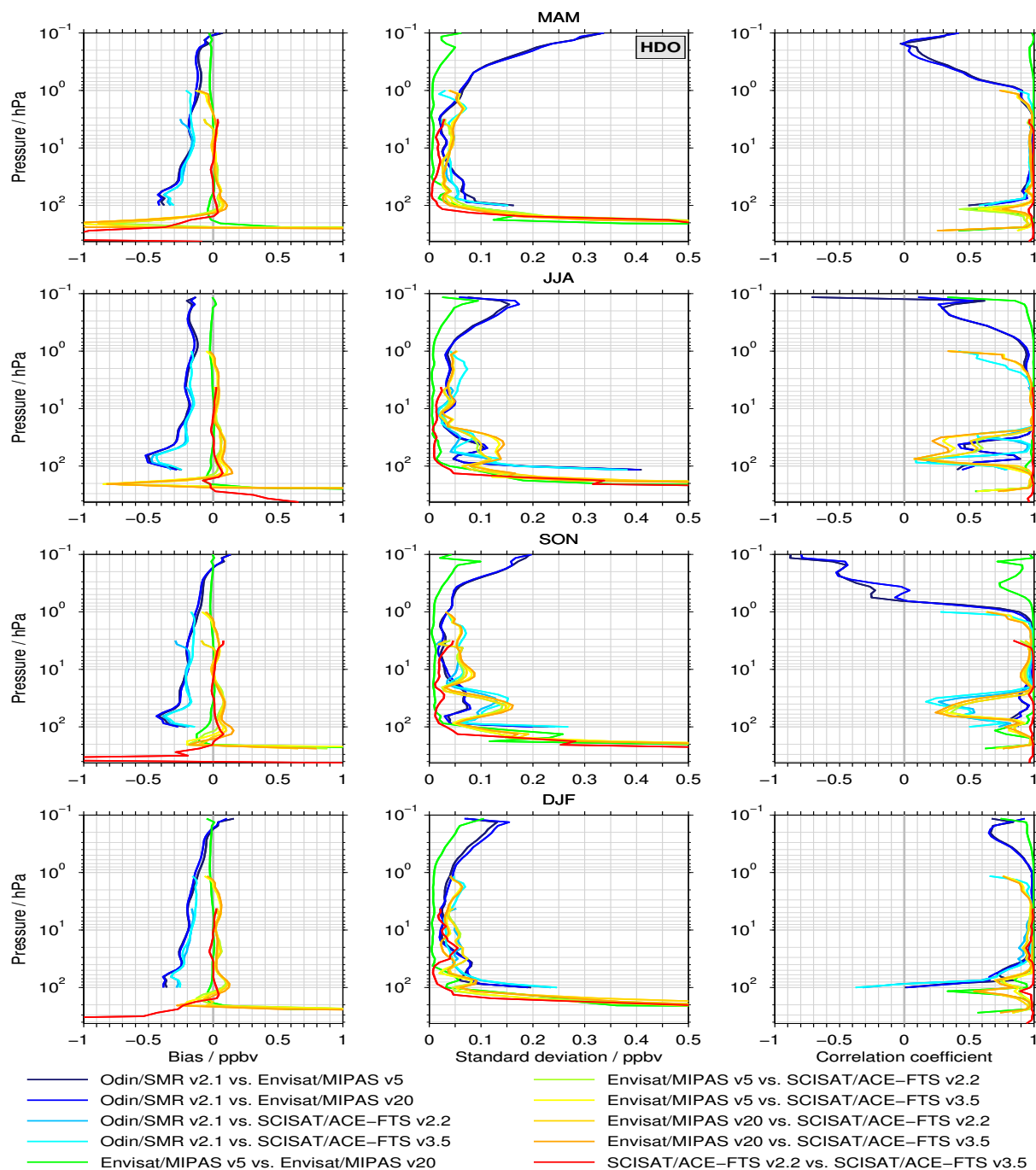
Figure 8. As Fig. 7 but here for HDO.
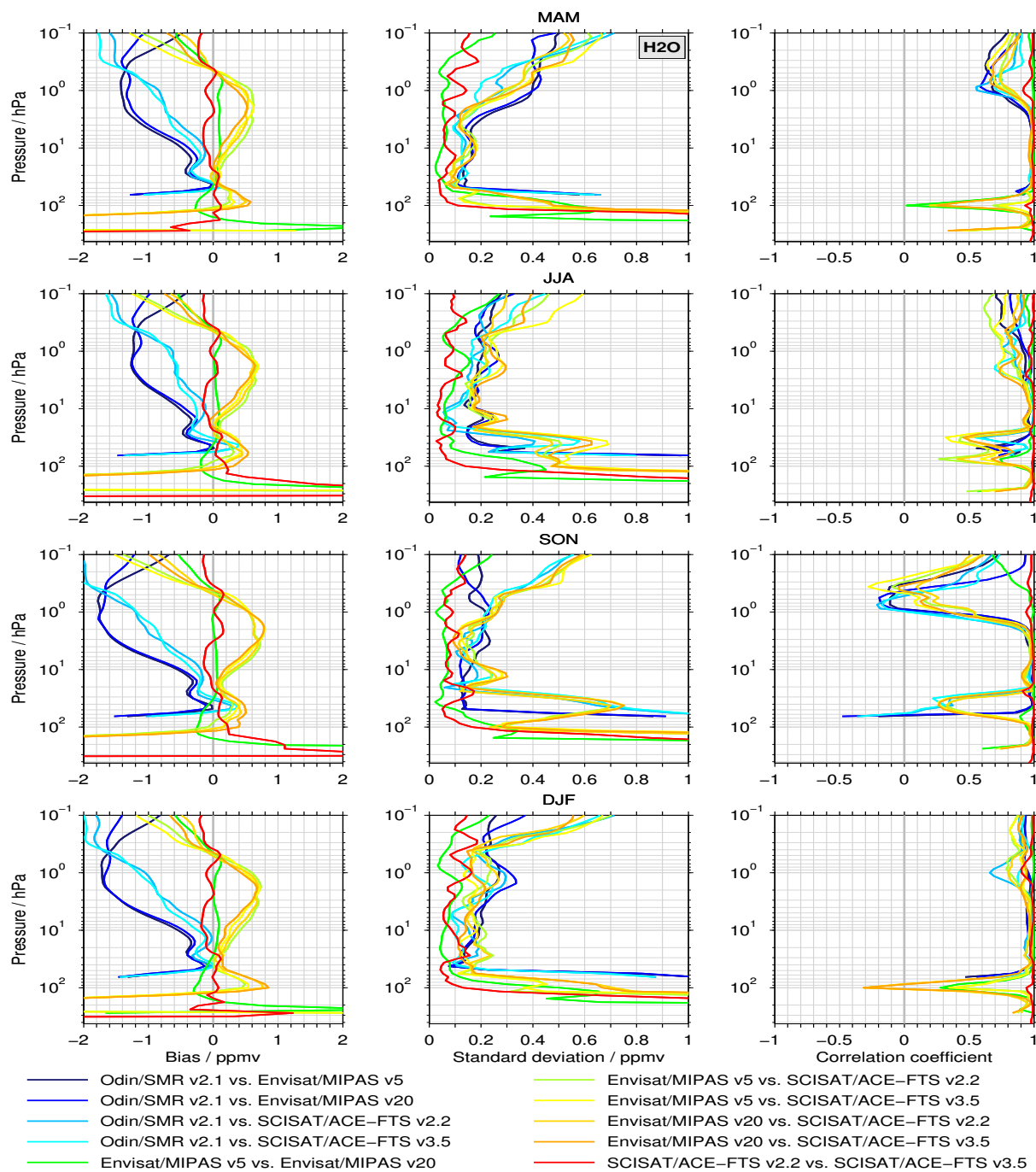
Figure 9. As Fig. 7 and 8 but here for H_2O.
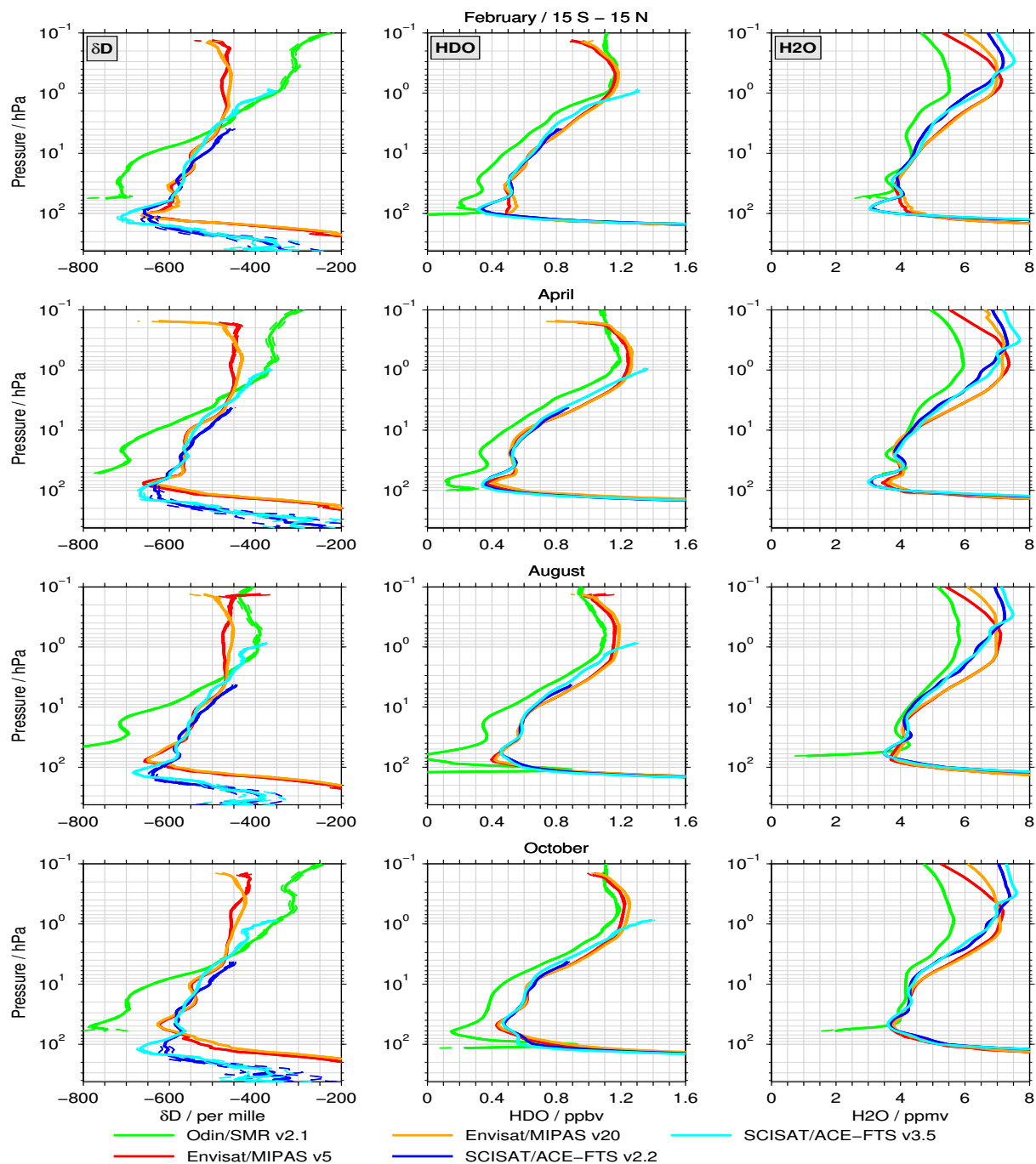
Figure 10. Tropical mean (15°S – 15°N) profiles for δD (left column), HDO (middle column) and H$_2$O (right column) for February, April, August and October given in different rows.