University of Wollongong

Research Online

6-3-2012

Process-evaluation of tropospheric humidity simulated by general circulation models using water vapor isotopologues: 1. comparison between models and observations

Camille Risi
University Of Colorado At Boulder

David Noone
University Of Colorado At Boulder

John Worden
California Institute of Technology

Christian Frankenberg
California Institute of Technology

Gabriele Stiller
IMK-A-SF, Karlsruhe Institute of Technology

Follow this and additional works at: https://ro.uow.edu.au/scipapers

Part of the Life Sciences Commons, Physical Sciences and Mathematics Commons, and the Social and Behavioral Sciences Commons

See next page for additional authors

Recommended Citation

Process-evaluation of tropospheric humidity simulated by general circulation models using water vapor isotopologues: 1. comparison between models and observations

Abstract

[1] The goal of this study is to determine how H2O and HDO measurements in water vapor can be used to detect and diagnose biases in the representation of processes controlling tropospheric humidity in atmospheric general circulation models (GCMs). We analyze a large number of isotopic data sets (four satellite, sixteen ground-based remote-sensing, five surface in situ and three aircraft data sets) that are sensitive to different altitudes throughout the free troposphere. Despite significant differences between data sets, we identify some observed HDO/H2O characteristics that are robust across data sets and that can be used to evaluate models. We evaluate the isotopic GCM LMDZ, accounting for the effects of spatiotemporal sampling and instrument sensitivity. We find that LMDZ reproduces the spatial patterns in the lower and mid troposphere remarkably well. However, it underestimates the amplitude of seasonal variations in isotopic composition at all levels in the subtropics and in midlatitudes, and this bias is consistent across all data sets. LMDZ also underestimates the observed meridional isotopic gradient and the contrast between dry and convective tropical regions compared to satellite data sets. Comparison with six other isotope-enabled GCMs from the SWING2 project shows that biases exhibited by LMDZ are common to all models. The SWING2 GCMs show a very large spread in isotopic behavior that is not obviously related to that of humidity, suggesting water vapor isotopic measurements could be used to expose model shortcomings. In a companion paper, the isotopic differences between models are interpreted in terms of biases in the representation of processes controlling humidity.

Keywords

Disciplines

Life Sciences | Physical Sciences and Mathematics | Social and Behavioral Sciences

Publication Details

Authors

Camille Risi, David Noone, John Worden, Christian Frankenberg, Gabriele Stiller, Michael Kiefer, Bernd Funke, Kaley Walker, Peter Bernath, Matthias Schneider, Debra Wunch, Vanessa Sherlock, Nicholas M. Deutscher, David W. Griffith, Paul O. Wennberg, Kimberly Strong, Dan Smale, Emmanuel Mahieu, Sabine Barthlott, Frank Hase, Omar Garcia, Justus Notholt, Thorsten Warneke, Geoffrey Toon, David Sayres, Sandrine Bony, Jeonghoon Lee, Derek Brown, Ryu Uemura, and Christophe Sturm

Process-evaluation of tropospheric humidity simulated by general circulation models using water vapor isotopologues: 1. Comparison between models and observations

Camille Risi,[1,2] David Noone,[1] John Worden,[3] Christian Frankenberg,[3] Gabriele Stiller,[4] Michael Kiefer,[4] Bernd Funke,[5] Kaley Walker,[6] Peter Bernath,[7] Matthias Schneider,[4,8] Debra Wunch,[9] Vanessa Sherlock,[10] Nicholas Deutscher,[11,12] David Griffith,[11] Paul O. Wennberg,[9] Kimberly Strong,[6] Dan Smale,[10] Emmanuel Mahieu,[13] Sabine Barthlott,[4] Frank Hase,[4] Omaira García,[8] Justus Notholt,[12] Thorsten Warneke,[12] Geoffrey Toon,[3] David Sayres,[14] Sandrine Bony,[2] Jeonghoon Lee,[3,15] Derek Brown,[1] Ryu Uemura,[16] and Christophe Sturm[17]

[1] The goal of this study is to determine how H_2O and HDO measurements in water vapor can be used to detect and diagnose biases in the representation of processes controlling tropospheric humidity in atmospheric general circulation models (GCMs). We analyze a large number of isotopic data sets (four satellite, sixteen ground-based remote-sensing, five surface in situ and three aircraft data sets) that are sensitive to different altitudes throughout the free troposphere. Despite significant differences between data sets, we identify some observed HDO/H_2O characteristics that are robust across data sets and that can be used to evaluate models. We evaluate the isotopic GCM LMDZ, accounting for the effects of spatiotemporal sampling and instrument sensitivity. We find that LMDZ reproduces the spatial patterns in the lower and mid troposphere remarkably well. However, it underestimates the amplitude of seasonal variations in isotopic composition at all levels in the subtropics and in midlatitudes, and this bias is consistent across all data sets. LMDZ also underestimates the observed meridional isotopic gradient and the contrast between dry and convective tropical regions compared to satellite data sets. Comparison with six other isotope-enabled GCMs from the SWING2 project shows that biases exhibited by LMDZ are common to all models. The SWING2 GCMs show a very large spread in isotopic behavior that is not obviously related to that of humidity, suggesting water vapor isotopic measurements could be used to expose model shortcomings. In a companion paper, the isotopic differences between models are interpreted in terms of biases in the representation of processes controlling humidity.

[1]Department of Atmospheric and Oceanic Sciences and Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, Colorado, USA.

[2]LMD/IPSL, CNRS, Paris, France.

[3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.

[4]IMK-ASF, Institute for Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany.

[5]Instituto de Astrofísica de Andalucía, Granada, Spain.

[6]Department of Physics, University of Toronto, Toronto, Ontario, Canada.

[7]Department of Chemistry, University of York, York, UK.

[8]Izaña Atmospheric Research Centre, Agencia Estatal de Meteorología, Santa Cruz de Tenerife, Spain.

[9]Division of Geological and Planetary Sciences and Engineering and Applied Science, California Institute of Technology, Pasadena, California, USA.

[10]National Institute of Water and Atmospheric Research, Lauder, New Zealand.

[11]Centre for Atmospheric Chemistry, University of Wollongong, Wollongong, New South Wales, Australia.

[12]Institute of Environmental Physics, University of Bremen, Bremen, Germany.

[13]Institute of Astrophysics and Geophysics, University of Liège, Liège, Belgium.

[14]School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA.

[15]Korea Polar Research Institute, Incheon, South Korea.

[16]Faculty of Science, University of the Ryukyus, Nishihara, Japan.

[17]Department of Geology and Geochemistry, Stockholm University, Stockholm, Sweden.

1. Introduction

[2] Despite continuous improvements in climate models, uncertainties in the predicted magnitude of climate change and associated feedbacks remain high [*Randall et al.*, 2007]. Processes controlling tropical and subtropical tropospheric humidity are involved both in the water vapor and cloud feedbacks. The former is one of the largest feedbacks in magnitude [e.g., *Soden and Held*, 2006], while the latter are the largest source of spread in climate change projections [*Bony and Dufresne*, 2005; *Bony et al.*, 2006]. Atmospheric general circulation models (GCMs) must therefore simulate the processes that control tropospheric humidity correctly for their climate change predictions to be credible.

[3] Tropical and subtropical tropospheric humidity results from a subtle balance between different processes: large-scale radiative subsidence [e.g., *Sherwood*, 1996; *T. Schneider et al.*, 2006; *Folkins and Martin*, 2005], detrainment of condensate from convective clouds and its subsequent evaporation [e.g., *Wright et al.*, 2009], evaporation of the falling precipitation [e.g., *Folkins and Martin*, 2005] and lateral mixing [e.g., *Zhang et al.*, 2003]. In models, an approximately correct humidity simulation could arise from compensating errors in the representation of these processes. Thus humidity observations alone are insufficient for verifying that all relevant processes are properly represented in the models.

[4] The stable isotopic composition of water vapor changes due to fractionation during phase changes. Measurements of the isotopologues of water vapor can thus provide complementary information on the water budget when combined with humidity because they record the integrated history of phase changes within a given air mass [*Dansgaard*, 1964]. Several studies have shown the value of the water vapor isotopic composition to investigate moistening and dehydrating processes in the tropical troposphere, such as condensate detrainment in the upper troposphere [*Moyer et al.*, 1996; *Kuang et al.*, 2003; *Webster and Heymsfield*, 2003; *Nassar et al.*, 2007; *Bony et al.*, 2008; *Steinwagner et al.*, 2010], precipitation evaporation in the lower troposphere [*Worden et al.*, 2007] and dehydration pathways and mixing of air masses [*Galewsky et al.*, 2007; *Galewsky and Hurley*, 2010; *Risi et al.*, 2010b]. Several studies have also highlighted the value of the water isotopic composition to evaluate convective parameterizations [*Bony et al.*, 2008; *Risi et al.*, 2010a; *Lee et al.*, 2009]. Here (and in a companion paper [*Risi et al.*, 2012], hereafter P2), we explore the added value of measurements of water vapor isotopologues to detect and understand biases in the representation of processes controlling tropospheric humidity in climate models, compared to the information which can be inferred from measurements of specific humidity alone.

[5] As a first step, we synthesize a large number of isotopic data sets (four satellite, sixteen ground-based remote-sensing, five surface in situ and three aircraft) and use them to evaluate the spatiotemporal isotopic distribution in the GCM LMDZ (Laboratoire de Météorologie Dynamique-Zoom). We focus on model strengths and weaknesses which can be reliably diagnosed from the ensemble of data, given limitations imposed by current deficiencies in remote-sensing measurement calibration and validation. Then, we compare LMDZ with six other isotopic GCMs from the SWING2 (Stable Water INtercomparison Group phase 2) project, to investigate whether the shortcomings evidenced in LMDZ are common to other models. We characterize the difference in the simulated isotopic composition between GCMs and assess whether isotopic measurements can discriminate between models in their representation of processes controlling humidity. In P2, the differences in the simulated isotopic composition between SWING2 models is exploited to understand the causes for humidity biases.

[6] We present the LMDZ GCM and the SWING2 database in section 2, and the various data sets and the model-data comparison methodology in section 3. In section 4, we extract features that are the most robust across the different data sets and we use LMDZ to understand the differences between data sets. In section 5, we use the different data sets to evaluate the spatiotemporal isotopic ratio distribution in LMDZ. In section 6, we compare the different SWING2 models. We conclude and propose perspectives for future work in section 7.

2. Model Simulations

2.1. The LMDZ4 Model and Its Control Simulation

[7] LMDZ4 [*Hourdin et al.*, 2006] is the atmospheric component of the Institut Pierre-Simon Laplace coupled model (IPSL-CM4) [*Marti et al.*, 2005] used in CMIP3 (Coupled Model Intercomparison Project) [*Meehl et al.*, 2007]. It is used here with a resolution of 2.5° in latitude, 3.75° in longitude and 19 vertical levels. The physical package includes the Emanuel convective scheme [*Emanuel*, 1991; *Emanuel and Zivkovic-Rothman*, 1999] and a statistical cloud scheme [*Bony and Emanuel*, 2001]. Water vapor and condensate are advected using a second order monotonic finite volume advection scheme [*Van Leer*, 1977; *Hourdin and Armengaud*, 1999]. The isotopic version of LMDZ is described in detail by *Risi et al.* [2010c].

[8] To compare with various data sets that have been collected since 1965, LMDZ is forced by observed sea surface temperatures (SST) and sea ice following the AMIP (Atmospheric Model Inter-comparison Project) protocol [*Gates*, 1992] from 1965 to 2009. The year 2010 is forced by NCEP (National Center for Environmental Prediction) SST [*Kalnay et al.*, 1996] because the AMIP SSTs were not yet available. We checked that the SST data set used has little impact on the simulations, by comparing LMDZ outputs forced by AMIP and NCEP SSTs for 2009: root mean square errors on monthly outputs are lower than 0.5 K for SST and lower than 10‰ for tropospheric δD. Horizontal winds at each vertical level are nudged by ECMWF reanalyses [*Uppala et al.*, 2005] as detailed by *Risi et al.* [2010c], forcing the simulations toward the actual meteorology and hence enabling direct comparison with observations on a daily basis.

2.2. SWING2 Models

[9] We compare eight simulations by seven other GCMs participating in the SWING2 inter-comparison project (http://people.su.se/~cstur/SWING2/). Some of them are nudged by reanalyses, while some of them are not (i.e. they are free-running) (Table 1). Since daily outputs are not available, we cannot compare these models directly to the data as rigorously as we can for LMDZ. Instead, we compare

Table 1. List of the SWING2 Models Used in This Study and Their Respective Simulations[a]

GCM	Reference	Simulations
GISS modelE	*Schmidt et al.* [2007]	free-running or nudged by NCEP
ECHAM4	*Hoffmann et al.* [1998]	nudged by ECMWF
LMDZ4	*Risi et al.* [2010c]	free-running or nudged by ECMWF
GSM	*Yoshimura et al.* [2008]	nudged by NCEP
CAM2	*Lee et al.* [2007]	free-running
Hadley	*Tindall et al.* [2009]	free-running
MIROC	*Kurita et al.* [2011]	free-running

[a]"Free-running" refers to standard AMIP-style simulations [*Gates*, 1992] forced by observed sea surface temperatures, and whose winds are not nudged.

all models to LMDZ as a common reference. One of the SWING2 models is a slightly different version of LMDZ (as described by *Risi et al.* [2010c]), where the second order advection scheme [*Van Leer*, 1977] was replaced by a simple upstream scheme (P2).

3. Data Sets and Comparison Methodology

[10] We focus on evaluating the HDO/H$_2$O ratio as quantified by the variable δD in ‰: $\delta D = (\frac{R}{R_{SMOW}} - 1) \cdot 1000$, where R is the HDO/H$_2$O ratio of the water vapor and R_{SMOW} is the Vienna Standard Mean Ocean Water (VSMOW) isotopic ratio [*Craig*, 1961]. To evaluate the simulated three-dimensional water vapor δD distribution from the surface to tropopause, we combine various data sets that sample different parts of the atmosphere. We use several satellite data sets which provide a global coverage (Table 2): SCIAMACHY (a short-wave infra-red spectrometer) mainly sensitive to the lower troposphere; TES (a nadir-viewing thermal infrared spectrometer) mainly sensitive to the mid-troposphere; ACE-FTS (an infrared solar-occultation instrument); and MIPAS (a limb infrared sounder). Both ACE-FTS and MIPAS are sensitive in the upper troposphere. In addition, we use ground-based remote-sensing data sets derived from mid-infrared and near-infrared solar absorption spectra acquired at 14 stations (Tables 3 and 4) and in situ measurements made at the surface and by aircraft (Table 5).

[11] Remote-sensing measurements of δD are a very new development. Absolute measurement calibration is dependent on accurate spectroscopy, while retrieval validation requires in situ profiling capability. Currently, both are lacking. In the interim a comparison methodology that is insensitive to absolute calibration uncertainties (i.e. characterization of spatiotemporal variability) is necessary. While it is possible that the observed variability is erroneous, the use of multiple data sets helps to ensure that the conclusions we draw are robust. Measurement principles, observing geometries as well as spectroscopic regions differ widely from data set to data set, hence, common errors are unlikely.

[12] Since each remote-sensing system of δD has its own sensitivity and is subject to sampling biases, we follow a model-to-satellite approach to estimate what the instruments would observe if measuring the model fields. First, to take into account the spatiotemporal sampling of the data, we collocate the outputs with the data at the daily scale. Second, to take into account the sensitivities of the different remote-sensing instruments to the true state, we apply the appropriate averaging kernels to the model outputs. Averaging kernels define the sensitivity of the retrieval at each level to the true state at each level.

3.1. SCIAMACHY

[13] The SCIAMACHY (Scanning Imaging Absorption Spectrometer for Atmospheric Chartography) instrument on board ENVISAT (European Space Agency environmental research satellite) measures short-wave infrared spectra from reflected sunlight from which precipitable water and δD integrated over the entire atmospheric column are retrieved [*Frankenberg et al.*, 2009]. It is mainly sensitive to the lower troposphere, since about 90% of the atmospheric water is found below 500 hPa and 55% below 800 hPa. The data are currently available from 2003 to 2005. The spatial footprint of SCIAMACHY δD is 120 km (across track) and 30 km (along-track) [*Frankenberg et al.*, 2009]. Precision for a single measurement has been estimated as 40–100‰, but statistical uncertainty in the mean is reduced by averaging in space and time [*Frankenberg et al.*, 2009] (Appendix B).

[14] To avoid potential isotopic biases related to the presence of clouds or sampling of an incomplete atmospheric column, we discarded all retrievals associated with a cloud fraction higher than 10% or with a retrieved precipitable

Table 2. The Different Data Sets of Water Vapor Isotopic Composition Retrieved By Satellite Instruments That We Used[a]

Data Set	Reference	Level	Spatial Coverage or Location	Period	Precision	Footprint	Comparison Methodology
SCIAMACHY	*Frankenberg et al.* [2009]	total column, mainly sensitive in the boundary layer	global	2003–2005	40‰–100‰, reduced by averaging	120 × 20 km	collocation
TES	*Worden et al.* [2007]	600 hPa	global	2004–2008	10–15‰, reduced by averaging	5.3 × 8.5 km	collocation, convolution with kernels
ACE	*Nassar et al.* [2007]	down to 500 hPa	global, but small number of measurements	2003–2008	about 50‰, reduced by averaging	limb measurement	collocation, smoothing
MIPAS	*Steinwagner et al.* [2010]	down to 300 hPa	global	September 2002– March 2004	about 50‰, reduced by averaging	limb measurement	collocation, convolution with kernels

[a]The period indicates that used in our analysis, which can be shorter than the total data set.

Table 3. Sites of the NDACC Network for Which δD Profiles Up to 10 km Have Been Retrieved as Part of the MUSICA Project[a]

Data Set	Spatial Coverage or Location	Altitude (m)	Period
Arrival heights (Antarctica)	77.82° S, 166.65°E	250	November 2002–December 2010, no winter measurements
Lauder (New Zealand)	45.05°Ss, 169.68°E	370	December 2001–December 2010
Wollongong (Australia)	34.41°S, 150.88°E	30	August 2007–December 2010
Izaña (Canary Islands)	28.30°N, 16.48°W	2367	June 1999–December 2010
Jungfraujoch (Switzerland)	46.6°N, 8.0°E	3580	April 1996–December 2010
Karlsruhe (Germany)	49.0°N, 8.38°E	116	May 2010–December 2010
Kiruna (Sweden)	67.84°N, 20.41°E	419	May 1996–December 2010
Eureka (Canada)	80.05°N, 86.42°W	610	August 2006 to December 2010, no winter measurements

[a]The period indicates that used in our analysis, which can be shorter than the total data set. For all these data sets, model outputs were both collocated and transformed by averaging kernels.

water differing from ECMWF reanalyses by more than 10%, selecting only about one third of the measurements [*Risi et al.*, 2010b]. The clear-sky sampling bias is further discussed in section 5.1. We sampled LMDZ-iso daily outputs coincident with observations and re-gridded the data on the LMDZ grid (2.5° in latitude × 3.75° in longitude).

3.2. TES

[15] The TES (Tropospheric Emission Spectrometer) instrument [*Worden et al.*, 2006, 2007] on board the Aura satellite measures high-resolution thermal infrared spectra from which the lower mid-tropospheric δD are retrieved. The footprint is 5.3 km × 8.5 km and the precision is about 10–15‰ (Appendix B2). We use the F04 version over the 2004–2008 period. The degree of freedom for signal (DOFS) is a scalar metric indicating the sensitivity of the measurement to the true state. We selected only retrievals for which the DOFS was higher than 0.5, to ensure a significant sensitivity to the true state [*Worden et al.*, 2006].

[16] A recent calibration study (see section 3.7) suggests that the raw TES δD is biased high [*Worden et al.*, 2006]. This may arise from spectroscopic errors. Following *Worden et al.* [2010] and *Lee et al.* [2011], we corrected the HDO data by reducing it by approximately 4% (depending on the averaging kernels, Appendix A). This correction leads to a decrease of δD of about 40‰ in the tropics on average and of about 25‰ in dry subtropical regions, so it significantly affects zonal gradients. Since this correction results from only one calibration campaign, there remains significant uncertainty in the absolute value of the TES δD and on its meridional gradients. We re-interpolated the TES data on the LMDZ grid and analyzed results at 600 hPa where the HDO sensitivity is a maximum.

[17] To mimic the TES temporal and geographic sampling pattern, we sampled LMDZ-iso daily outputs coincident with observations. Due to limited instrument sensitivity and vertical resolution, the TES retrieval at each level reflects the δD over a broader range of altitudes, and is sensitive to the a priori information. These effects are represented by the averaging kernels, which depend on geographical location and atmospheric state, including the presence of clouds. To make the closest possible comparison, we apply the same averaging kernels to the simulated profiles. To do so, we calculated monthly mean averaging kernels for each LMDZ grid box, and applied these kernels together with the a priori constraint to the model outputs [*Worden et al.*, 2006] (Appendix C2). In doing so, we neglect the day-to-day variability of the averaging kernels, such as that related to clouds. We determined that calculating monthly mean averaging kernels over total sky conditions or clear sky only conditions lead to differences less than 6‰ on kernel-weighted δD (Appendix C2), consistent with similar results for other chemical species [*Aghedo et al.*, 2011]. On average, applying the averaging kernels leads to a δD increase by up to 10‰ in convective regions and by up to 30‰ in dry subtropical regions, demonstrating the importance of accounting for the sensitivity of the TES retrievals for a rigorous model-data comparison.

3.3. ACE

[18] The ACE-FTS (Atmospheric Chemistry Experiment Fourier Transform Spectrometer) instrument on board the ACE satellite measures δD profiles from the stratosphere to about 400 hPa depending on cloud cover [*Nassar et al.*, 2007]. As an occultation sounder, it has better vertical resolution than nadir sounders, but a coarser horizontal resolution. We use the v2.2_HDO_Update over the 2003–2008 period. We discarded measurements with errors in H_2O and HDO higher than the retrieved values. This leads to a slight bias toward measurements when H_2O content is higher. In

Table 4. Sites of the TCCON Network for Which Total Column δD Have Been Retrieved[a]

Data Set	Reference	Spatial Coverage or Location	Period	Precision
Ny Alesund (Norway)	*Deutscher et al.* [2010]	78.923°N, 11.923°E	March 2005–August 2010	40‰, no winter measurements
Bremen (Germany)	*Messerschmidt et al.* [2010]	53.104°N, 8.850°E	March 2005–December 2010	35‰
Park Falls (United States)	*Washenfelder et al.* [2006]	45.94°N, 90.27°W	June 2004 to December 2009	35‰
Lamont (United States)	*Washenfelder et al.* [2006]	36.60°N, 97.49°W	July 2008 to December 2009	15‰
Pasadena (United States)	*Washenfelder et al.* [2006]	34.20°N, 118.189°W	July 2007 to June 2008	15‰
Darwin (Australia)	*Deutscher et al.* [2010]	12.43°S, 130.89°E	2004–2009	7‰
Wollongong (Australia)	*Deutscher et al.* [2010]	34.41°S, 150.88°E	2008–2009	5‰
Lauder (New Zealand)	*Wunch et al.* [2010]	45.05°S, 169.68°E	2004–2009	8‰

[a]The period indicates that used in our analysis, which can be shorter than the total data set. For all these data sets, model outputs were both collocated and transformed by averaging kernels. Precision estimates as described in Appendix B7.

Table 5. Summary of the Different in Situ Isotopic Data Sets Used[a]

Data Set	Reference	Level	Spatial Coverage or Location	Period	Precision	Measurement Method
GNIP-vapor at Vienna	IAEA web site	surface	48.25°N, 16.37°E	2001–2003	undocumented	undocumented
GNIP-vapor at Ankara	IAEA web site	surface	39.95°N, 32.88°E	2001–2002	undocumented	undocumented
GNIP-vapor at Manaus	IAEA web site	surface	3.12°S, 60.02°W	1978–1980	undocumented	undocumented
Sampling at Rehovot	Angert et al. [2008]	surface	31.9°N, 34.65°E	December 1997, November 1998	1‰	cryogenic extraction
Sampling at Saclay	Risi et al. [2010c]	surface	48.73°N, 2.17°E	September 1982 to September 1984	1‰	cryogenic extraction
Southern Ocean surface samples	Uemura et al. [2008]	surface	Southern Ocean	January 2004	1‰	cryogenic extraction
Picarro in Hawaii	Johnson et al. [2011]	surface at 680 hPa	19.53°N, 155.57°W	10 October–6 November 2008	5 to 10 ‰	Laser spectrometry with Picarro
Aircraft in Nebraska	Ehhalt et al. [2005]	profiles between 1.5 km and 9.2 km	41.83°N, 101.67°W	1965–1967	1‰	cryogenic extraction
Aircraft in California, near Santa Barbara	Ehhalt et al. [2005]	profiles between 15 m to 9.2 km	34.0°N, 125.0°W	1966–1967	1‰	cryogenic extraction
Aircraft in California, above the Death Valley	Ehhalt et al. [2005]	profiles between −30 m and 8.9 km	36.0°N, 117.0°W	1966–1967	1‰	cryogenic extraction
Aircraft during CR-AVE	Sayres et al. [2010]	profiles between 475 hPa and 70 hPa	near Costa-Rica (1.4°S–29.7°N, 95.2°W–78.0°W)	27 January to 2 February 2006	17‰	Laser spectrometry with ICOS
Aircraft during TC4	Sayres et al. [2010]	profiles between 475 hPa and 70 hPa	near Costa-Rica (same region)	3 to 13 August 2007	17‰	Laser spectrometry with ICOS
Aircraft during TC4	Sayres et al. [2010]	profiles between 450 hPa and 70 hPa	near Costa-Rica (same region)	3 to 13 August 2007	50‰	Laser spectrometry with Hoxotope

[a]For model-data comparison, model outputs were collocated with each data set at the daily scale.

addition, we applied a 3 median average deviation filter to remove outliers. We checked that this method does not distort the mean and median [*Jones et al.*, 2011]. The precision for individual measurement varies from about 20‰ in the tropics to 60‰ in midlatitudes (Appendix B3).

[19] Given the low number of solar occultation measurements per day, we re-gridded the data on a 10° × 100 hPa latitude/height grid. We sampled LMDZ-iso daily outputs coincident with observations. ACE does not use optimal estimation, and averaging kernels are not computed. To take into account the vertical resolution of the data, we applied a triangular kernel of base 3 km to the model outputs [*Dupuy et al.*, 2009].

3.4. MIPAS

[20] The MIPAS instrument on board the ENVISAT satellite is a limb sounder measuring δD profiles down to about 300 hPa [*Payne et al.*, 2007; *Steinwagner et al.*, 2007, 2010], at 10:00 am and 10:00 pm local time. We use the V3O_HDO_5 data between September 2002 and March 2004. We discarded data with the visibility flag equal to zero and with diagonal elements of the averaging kernels lower than 0.03. This rejects about 70% of the data in tropics at 13 km. The precision is about 60‰ in the tropics and 150‰ in midlatitudes.

[21] We sampled LMDZ-iso daily outputs coincident with observations. Based on 216 MIPAS profiles collected during three days representative of different seasons, we calculated 21 averaging kernels representative of different tropopause heights from 7 km to 17 km in bins of 0.5 km (more details in Appendix C2). We applied these representative averaging kernels to model outputs depending on the observed tropopause height. As for TES, applying the kernel is crucial for the model-data comparison, as the convolution increases the δD by up to 300‰ in the tropical upper troposphere.

[22] Biases related to incomplete scanning of the atmospheric column when clouds are present are discussed in Appendix B5, and are shown not to significantly affect our results.

3.5. Ground-Based FTIR at MUSICA Sites

[23] High resolution ground-based Fourier Transform Infrared (FTIR) spectrometers have been measuring solar absorption spectra in the mid-infrared region (750–4200 cm^{-1}) since the 1990s at about 15 globally distributed sites that are part of the Network for the Detection of Atmospheric Composition Change (NDACC, www.acd.ucar.edu/irwg [*Kurylo and Zander*, 2000]). In the mid-infrared spectral region, there are several spectral microwindows with well-isolated and strong H_2O and HDO signatures.

[24] In the framework of the project MUSICA (Multiplatform remote-sensing of Isotopologues for investigating the Cycle of Atmospheric water, www.imk-asf.kit.edu/english/musica), a dedicated water isotopologue retrieval algorithm is applied. It consists in a simultaneous optimal estimation of H_2O and HDO as well as δD [*M. Schneider et al.*, 2006, 2010b]. With this retrieval technique, tropospheric H_2O and δD column abundances and profiles with a modest vertical resolution can be produced from the NDACC spectra. In this paper we use the ground-based MUSICA data version v101220_Ca.0. These data are retrieved applying signatures

in the 2650–3050 cm^{-1} spectral region [Schneider et al., 2010a].

[25] The uncertainties are estimated in detail by Schneider et al. [2010a]. Concerning column-integrated data, the precision is about 5‰ and biases can reach 100‰. Concerning profile data, the precision is about 10–25‰ and biases can reach 25–50‰ (Appendix B6). The most important source for systematic biases as well as for the random errors in the profiles are uncertainties when modeling the shape of the high resolution absorption lines.

[26] The ground-based MUSICA data version v101220_Ca.0 is currently available for 8 FTIR NDACC sites: Eureka (Canadian Arctic), Kiruna (Northern Sweden), Karlsruhe (Germany), Jungfraujoch (Switzerland), Izaña (Canary Island), Wollongong (Australia), Lauder (New Zealand), and Arrival Heights (Antarctica) (Table 3). We sampled the LMDZ H$_2$O and HDO profiles coincident with the observations and applied the averaging kernels and a priori profiles corresponding to each measured profile (Appendix C3).

3.6. Ground-Based FTIR at TCCON Sites

[27] The Total Carbon Column Observing Network (TCCON) [Wunch et al., 2010, 2011] is a network of very high quality ground-based FTIR systems recording solar absorption spectra in the near infrared spectral region (3800–9000 cm^{-1}). In the near infrared there are strong and well isolated H$_2$O absorption signatures but HDO signatures are significantly weaker than the interfering absorption of H_2O and CH_4. We use data obtained at the TCCON sites Ny Alesund (Spitzbergen Island), Bremen (Germany), Park Falls, Lamont and Pasadena (United States), Lauder (New-Zealand) and Darwin and Wollongong (Australia) (Table 4). δD is inferred from the retrieved total columns of H$_2$O and HDO. The TCCON HDO data have not been evaluated for spectroscopic errors. Note that total column δD derived from measurements on TCCON and NDACC sites have different error characteristics and sensitivities, due to different spectroscopic errors and retrieval methodologies.

[28] We sampled the LMDZ H$_2$O and HDO profiles coincident with the observations and estimated the model-equivalent columns using the averaging kernels and a priori profiles, before calculating total column δD (Appendix C3). Averaging kernels were parameterized as a function of solar zenith angle alone. The uncertainty associated with using these kernels compared with using individual kernels is lower than 3‰ (Appendix C3).

3.7. In Situ Surface Measurements

[29] Two kinds of surface vapor measurements are used in this study. First, we use δD values from vapor samples obtained by cryogenic sampling: samples collected at GNIP (Global Network for Isotopes in Precipitation) vapor stations in Vienna, Ankara, Manaus (as in work by Risi et al. [2010c]), daily samples collected at Rehovot, Israel [Angert et al., 2008] and at Saclay, France (described by Risi et al. [2010c]), and samples collected during cruises in January 2004 by Uemura et al. [2008]. δD was measured by mass spectrometers with precision better than 1‰, and tied to the absolute scale using reference standards.

[30] Second, we use continuous data collected by a Picarro instrument in Hawaii at about 680 hPa [Johnson et al., 2011; Noone et al., 2011]. δD values have been corrected to match

values obtained by laboratory analysis of whole air vapor samples simultaneously collected with flasks. This data has been used for estimating the bias correction applied to the TES data [Worden et al., 2010] (section 3.2). We selected only the Picarro data during the nighttime, which represents the free troposphere [Worden et al., 2010]. Since LMDZ cannot see any land in the grid point of Hawaii, the model results at 680 hPa are representative of the free troposphere only, so we discarded observed daytime data that are representative of boundary layer air.

[31] For each measurement, we sampled LMDZ outputs coincident with observations and re-gridded all data on the model grid.

3.8. In Situ Aircraft Measurements

[32] Three aircraft data sets are used in this study. The first is data collected between 1–2 km and 9–11 km in Nebraska, around Santa Barbara and in Death Valley [Ehhalt, 1974]. Samples were collected by cryogenic sampling and analyzed on a mass spectrometer. Although the precision of the mass spectrometer is 1‰, some large errors can arise from the contamination in the sampling tubes. The data have recently been corrected for this effect but some unquantified and potentially large errors may remain especially in the upper troposphere [Ehhalt et al., 2005]. We sampled LMDZ outputs coincident with observations.

[33] The second data set was collected by the ICOS instrument [Sayres et al., 2009] during the Costa Rica Aura Validation Experiment (CR-AVE) campaign near Costa Rica in winter 2006, and the third was collected by the ICOS and Hoxotope instruments [St. Clair et al., 2008] during the Tropical Composition, Cloud and Climate Coupling (TC4) campaign in the same region in summer 2007. These data sets are both described by Sayres et al. [2010]. We applied the same data quality-filtering and processing as in work by Sayres et al. [2010], including screening of potentially contaminated data. The measurement precisions are about 17‰ for ICOS and 50‰ for Hoxotope. Both ICOS and Hoxotope were calibrated through laboratory experiments. We sampled LMDZ outputs coincident with observations and re-gridded all data on the model grid. We show data only in grid boxes where data was sampled during both campaigns to calculate seasonal variations and representative annual means.

4. Comparison Between Data Sets

[34] In this section, we compare the different data sets to extract the most robust features. Then we use LMDZ to understand and quantify the sources of differences between the data sets.

4.1. Robust Features Among Data Sets

[35] Figure 1 synthesizes the data as zonal, annual means at different altitudes throughout the free troposphere. We show in situ data at the surface; SCIAMACHY, MUSICA and TCCON data for total column δD; TES, MUSICA, Hawaii and Ehhalt [1974] data at 600 hPa; ACE, MIPAS, MUSICA, CR-AVE/TC4 and Ehhalt [1974] data at 350 hPa; and ACE, MIPAS and CR-AVE/TC4 at 250 hPa.

[36] The different data sets show large differences in δD (Figures 1a–1e). To better visualize the meridional gradients, we subtract the tropical average (Figures 1f–1j). We find that
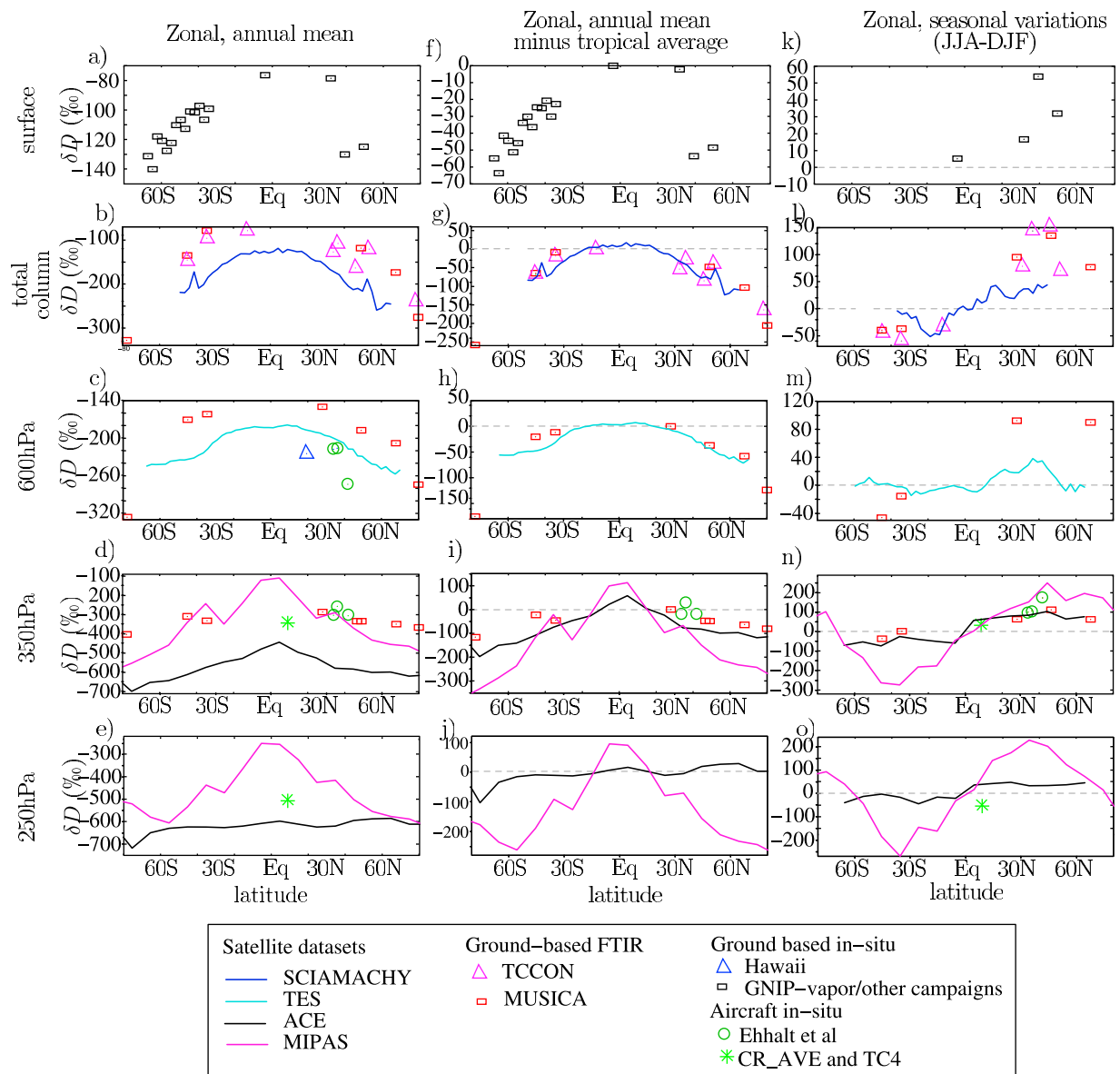
Figure 1. Synthesis of all data sets used in this study. Zonal, annual mean observed δD at different levels. (a) At the surface, using in situ measurements (black squares). (b) Total column (mainly sensitive to the boundary layer), using the SCIAMACHY satellite data set (blue line) and ground-based FTIR measurements from MUSICA (red squares) and TCCON (pink triangles) networks. MUSICA data at Izaña and Jungfraujoch were removed since these high altitude stations cannot sample the total column. (c) At 600 hPa, using the TES satellite data set (cyan line), ground-based FTIR measurements from MUSICA network (red squares), in situ measurements from Hawaii (at 680 hPa: blue triangle) and samples collected by aircraft by *Ehhalt* [1974] (green circles). (d) At 350 hPa (actually average between 320 and 380 hPa), using the MIPAS (purple line) and ACE (black line) satellite data sets, ground-based FTIR measurements from MUSICA network (red squares) and in situ measurements collected by aircraft by *Ehhalt*, [1974] (green circles) and during the CR-AVE and TC4 campaigns (green stars; the data has first been re-gridded on the model grid). (e) At 250 hPa (actually average between 220 and 280 hPa), using the same data sets as in Figure 1d. (f–j) Same as Figures 1a–1e, but the tropical average (30°S–30°N) has been subtracted to better visualize the meridional gradients. (k–o) Same as Figures 1a–1e but for seasonal variations (June–July August minus December–January–February).

the meridional gradient is qualitatively robust across data sets up to 350 hPa: in all data sets and at all levels, δD decreases poleward. This is qualitatively predicted from a simple Rayleigh distillation associated with decreasing temperature toward the poles and at higher altitudes (i.e. the temperature effect given by *Dansgaard* [1964]). For example, Figure 1g shows that the meridional gradients observed in total column δD are very consistent between

SCIAMACHY, MUSICA and TCCON data sets. The meridional gradient from 0 to 60° increases with altitude: about −50‰ at the surface, between −80 and −100‰ in the lower and mid troposphere (ground-based FTIR, SCIAMACHY, TES) and between −120 and −350‰ at 350 hPa (ACE and MIPAS). The weaker meridional gradient at the surface than at higher altitude can be explained by the evaporative recycling near the surface which partly counter-acts the temperature effects as air masses move poleward [e.g., *Werner et al.*, 2001; *Noone*, 2008]. Further, since the tropopause height decreases with latitude, the water vapor reaches low δD values at lower altitudes in high latitudes.

[37] Figures 1k–1o show the June–July August (JJA) minus December–January–February (DJF) differences at all levels. At all levels and in all data sets, δD is higher in summer than in winter in the subtropics and in midlatitudes. In all data sets, the seasonality reaches its maximum between about 30 and 50°. The amplitude varies between data sets and levels: 20–60‰ at the surface, 20–50‰ in the lower troposphere in SCIAMACHY and TES, 100–150‰ in the lower troposphere in ground-based FTIR, 100–200‰ from *Ehhalt* [1974], 50–100‰ in the upper troposphere in MIPAS and about 200‰ in the upper troposphere in MIPAS. Note that at Wollongong (marker at 34.41°S), the JJA-DJF variations at 600 and 350 hPa are weak, but this is due to the fact that the seasonal cycle in δD at Wollongong is shifted by two months. April–May–June (AMJ) minus October–November–December (OND) variations are −60‰ and −43‰ at 600 and 350 hPa respectively, which are more consistent with the other data sets. For remote-sensing data sets with averaging kernels, we checked that this seasonality was not simply an artifact of the instrument sensitivity: by applying the kernels to a constant δD profile with the averaging kernels, no such seasonality appears in the kernel-weighted profiles. To summarize, despite differing amplitudes amongst data sets, the sign of the δD seasonality is very robust. Therefore, δD seasonality is a robust observed property that can be used to evaluate models.

[38] While the equator-to-pole gradient and the seasonal differences are robust features, there are large differences in absolute values and variation magnitudes. The possible reasons for this are explored below.

4.2. Understanding Data Set Differences Using LMDZ

[39] The differences between the data sets can be explained by (1) spatiotemporal sampling, (2) instrument sensitivity, and (3) systematic biases in each data set. These three sources of differences are difficult to quantify directly since the data coverage is insufficient to quantify the effect of spatiotemporal sampling, and since vertical profiles through the troposphere are not available to explore the effect of instrument sensitivity. Therefore, we use LMDZ to quantify these three sources of differences, as explained in Appendix D. In doing so, we assume that LMDZ simulates the spatiotemporal δD patterns sufficiently well to quantify the spatiotemporal sampling and instrument sensitivity effects. We decompose the difference between each pair of data sets ($\Delta\delta D_{obs}$) as:

$$\Delta\delta D_{obs} = \Delta\delta D_{colloc} + \Delta\delta D_{convol} + \Delta\delta D_{errors}$$

where $\Delta\delta D_{colloc}$ is the effect of spatiotemporal sampling as predicted by LMDZ, $\Delta\delta D_{convol}$ is the effect of instrument sensitivity as predicted by LMDZ and $\Delta\delta D_{errors}$ is the combined effect of systematic biases in the data sets, of possible problems in the spatiotemporal patterns simulated by LMDZ and of sub-daily or sub_grid sampling that our collocation does not resolve (Appendix D). Results of this decomposition are shown in Table 6.

[40] At most stations, δD measured by ground-based FTIR is higher than measured by SCIAMACHY or by TES. Spatiotemporal sampling and instrument sensitivity effects are sometimes positive or negative, and cannot explain this systematic difference. The difference is thus likely due to systematic biases in the data or problems in simulated patterns. Except for Jungfraujoch (high altitude) and Eureka, Ny Alesund and Arrival Heights (high latitude), the third term is very consistent (between 30‰ and 87‰) across the 11 other FTIR stations which span various climate conditions. This suggests that the systematic difference between ground-based FTIR and satellite data is mainly due to systematic biases in the data. The δD in SCIAMACHY and TES might be too low, or that of ground-based FTIR too high.

[41] Aircraft measurements by *Ehhalt* [1974] have a consistently lower δD than TES at 600 hPa. This is explained partly by spatiotemporal sampling and by the TES instrument sensitivity. Overall, some systematic biases make δD 50–75‰ higher in aircraft data than in TES on two of the three sites. This aircraft data has also a systematically higher δD than ACE by 280 to 325‰, which is also mainly due to measurement biases in one of the data sets. In contrast, aircraft data measured during CR-AVE and TC4 usually have a lower δD than ACE at 250 hPa, which could be due to a systematic bias of about 100‰ in ACE or to systematic differences between clear sky conditions (which ACE requires) and cloudy conditions (which the aircraft may sample) that LMDZ does not resolve. Finally, MIPAS has a systematically higher δD than ACE. This difference is mainly (47 to 91%) due to the difference in instrument sensitivity, which can be taken into account in the model through convolution. However, the remaining 9–53% could be due to systematic biases in one of the data sets, or to different clear-sky sampling biases [*Lossow et al.*, 2011].

[42] If one were to assume that aircraft in situ data provide calibrated δD values, then the data by *Ehhalt* [1974] and ICOS are inconsistent, since the former has higher δD than ACE and the latter has lower δD than ACE, even after accounting for spatiotemporal effects. This points to systematic biases in the data, to problems in the δD patterns simulated by LMDZ, or even possibly to problems in the δD patterns observed by ACE. Therefore, even though we are using some in situ data, we remain cautious with all absolute values and we will thus only focus on spatiotemporal variations that are consistent across all data sets.

[43] A similar decomposition approach can be used to understand the differences of meridional gradient and seasonality between data sets. In particular, in the upper troposphere, the ACE and MUSCICA data exhibits two-four times smaller seasonality than in MIPAS in the subtropics and midlatitudes. This is explained mainly by the instrument sensitivity (e.g. 80% of the MIPAS-ACE difference in the Northern Hemisphere, 95% of the MIPAS-MUSICA data

Table 6. Differences of Annual, Zonal Mean δD Between Pairs of Data Sets ($\Delta\delta D_{obs} = \delta D_{i,obs,data} - \delta D_{j,obs,data}$), and Their Relative Contributions[a]

Data Sets	Location	Level	Total Difference ($\Delta\delta D_{obs}$) (‰)	Spatiotemporal Sampling Effect ($\Delta\delta D_{colloc}$) (‰)	Instrument Sensitivity Effect ($\Delta\delta D_{convol}$) (‰)	Residual ($\Delta\delta D_{errors}$) (‰)
MUSICA-TCCON	Lauder	total column	11	1	1	9
MUSICA-TCCON	Wollongong	total column	17	−3	9	11
MUSICA-SCIAMACHY	Arrival Heights	total column	−154	−24	−3	−127
MUSICA-SCIAMACHY	Lauder	total column	74	7	−2	69
MUSICA-SCIAMACHY	Wollongong	total column	106	19	−2	90
FTIR-SCIAMACHY	Izaña	total column	−25	−83	−4	62
MUSICA-SCIAMACHY	Jungfraujoch	total column	−77	−93	4	12
MUSICA-SCIAMACHY	Karlsruhe	total column	97	19	9	70
MUSICA-SCIAMACHY	Eureka	total column	−42	−63	3	17
TCCON-SCIAMACHY	Lauder	total column	63	6	−3	61
TCCON-SCIAMACHY	Wollongong	total column	89	22	−11	77
TCCON-SCIAMACHY	Darwin	total column	54	3	−4	61
TCCON-SCIAMACHY	Pasadena	total column	38	−53	7	84
TCCON-SCIAMACHY	Oklahoma	total column	72	5	−14	81
TCCON-SCIAMACHY	Wisconsin	total column	46	−13	−9	67
TCCON-SCIAMACHY	Bremen	total column	48	26	−7	30
TCCON-SCIAMACHY	Ny Alesund	total column	−24	−15	−7	−2
MUSICA-TES	Lauder	600 hPa	63	24	−30	69
MUSICA-TES	Wollongong	600 hPa	57	37	−26	45
MUSICA-TES	Izaña	600 hPa	46	−13	−11	70
MUSICA-TES	Karlsruhe	600 hPa	42	13	−34	52
MUSICA-TES	Kiruna	600 hPa	50	15	−19	54
MUSICA-TES	Eureka	600 hPa	−30	−53	64	87
Ehhalt-TES	Nebraska	600 hPa	−56	−92	−14	51
Ehhalt-TES	Santa Barbara	600 hPa	−17	−78	−13	75
Ehhalt-TES	Death Valley	600 hPa	−11	11	−10	−12
Ehhalt-ACE	Nebraska	350 hPa	282	74	0	208
Ehhalt-ACE	Santa Barbara	350 hPa	278	11	0	268
Ehhalt-ACE	Death Valley	350 hPa	325	65	0	260
ICOS-ACE	CR-AVE/TC4 region	350 hPa	139	163	0	−24
ICOS-ACE	CR-AVE/TC4 region	250 hPa	−27	73	0	100
MIPAS-ACE	tropics	350 hPa	281	2	255	24
MIPAS-ACE	global	350 hPa	254	7	232	15
MIPAS-ACE	tropics	250 hPa	269	−1	126	144
MIPAS-ACE	global	250 hPa	187	4	96	87

[a]Included are effect of spatiotemporal sampling ($\Delta\delta D_{colloc} = \Delta\delta D_{i,colloc,model} - \Delta\delta D_{j,colloc,model}$), effect of instrument sensitivity ($\Delta\delta D_{convol} = \Delta\delta D_{i,convol,model} - \Delta\delta D_{j,convol,model}$) and residual (i.e., due to measurement errors or problems in the spatiotemporal distribution simulated by LMDZ: $\Delta\delta D_{errors} = \Delta\delta D_{i,error,data} - \Delta\delta D_{j,error,data} + \epsilon_i - \epsilon_j$). See Appendix D for a detailed explanation of these different terms and notations. When we compare two satellite data sets, we average both data sets over the same band of latitude noted in the Location column: 30°S–30°N for the tropics and 80°S–80°N for global mean. When we compare a satellite data set with a ground-based or aircraft data set, the ground-based/aircraft observations are averaged over the region of observation indicated in the Location column, and the satellite observation is averaged zonally at the same latitude as the ground-based or aircraft observations. This makes the δD difference in the Total Difference column consistent with what we can see on Figure 1. Ground-based stations at which no satellite data are available are not shown (e.g., no TES data at Arrival Heights).

between Izaña and Eureka). Similarly, ACE and MUSICA also exhibit a meridional gradient that is two-three times smaller than in MIPAS. The instrument sensitivity explains about 30% of the MIPAS-ACE difference and up to 70% of the MIPAS-MUSICA difference. This is consistent with the small sensitivity of the MUSICA retrievals to δD in the upper troposphere [*M. Schneider et al.*, 2006], suggesting that MUSICA, and thus maybe also ACE, may underestimate the meridional gradient and seasonality of δD.

5. Model-Data Comparison

[44] In this section, we compare LMDZ simulations to the different data sets, and focus on model-data differences that are robust across data sets. Before this comparison, we summarize below the different sources of model-data differences.

5.1. Sources of Model-Data Differences

[45] In addition to the biases in the data found in section 4.2, some sources of model-data differences can arise from our comparison methodology.

[46] The effect of spatiotemporal sampling is taken into account by collocating model outputs with the data at the daily scale. The root-mean square error associated with the spatiotemporal sampling effect on zonal mean δD is lower than 5‰ for ground-based FTIR data that have a high measurement frequency, but is about 10–20‰ for SCIAMACHY, TES and ACE satellite data sets, and up to 50‰ for MIPAS (Figure 2, black). This shows the importance of taking this effect into account in the model-data comparison. Spatiotemporal sampling effects usually increases (decreases) δD in regions of large-scale ascent (descent) for SCIAMACHY, decreases δD in ACE and in MIPAS at 350 hPa, and has little coherent effect for other
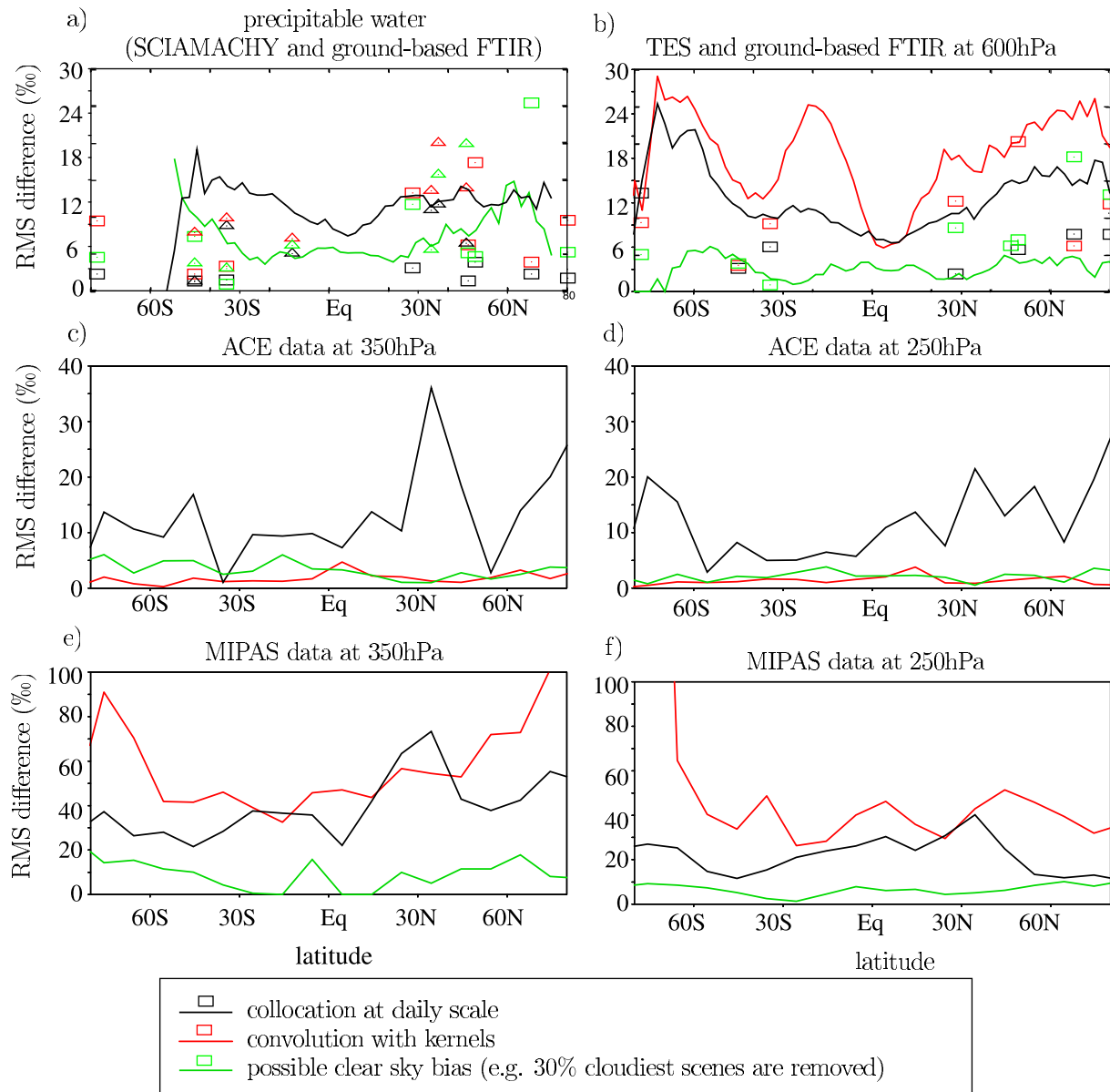
Figure 2. Effects of spatiotemporal sampling (black), instrument sensitivity (red) and clear-sky sampling biases (green) that need to be taken into account when comparing LMDZ outputs to the different data sets. These effects correspond to an estimation of the sources of errors in the zonal monthly means observed by different data sets compared to the truth. (a) Total column δD observed by SCIAMACHY (lines), ground-based MUSICA (squares) and TCCON (triangles), (b) δD at 600 hPa observed by TES (lines) and MUSICA (squares), (c) δD at 350 hPa observed by ACE, (d) same as Figure 2c but at 250 hPa, (e) δD at 350 hPa observed by MIPAS, and (f) same as Figure 2d but at 250 hPa. The effect of spatiotemporal sampling is calculated by the root mean square difference of δD between monthly mean raw model outputs and monthly mean collocated outputs. The effect of instrument sensitivity is calculated by the root mean square difference between monthly mean collocated and kernel-weighted model outputs and monthly mean collocated outputs. An upper bound for the effect of clear-sky sampling is calculated as the root mean square differences in δD between collocated and kernel-weighted model outputs and collocated and kernel-weighted model outputs after removing the 30% cloudiest scenes.

data sets. Our collocation method ignores spatial variations at small scales that could lead to differences between δD in a small instrument footprint and δD in the 2.5° × 3.75° GCM grid box. We also ignore sub-daily temporal variability. Some satellites sample the atmosphere once or twice a day at the same local time every day, which may have a systematic

effect on δD. These additional sources of model-data differences are difficult to quantify with a GCM.

[47] The effect of instrument sensitivity is taken into account by applying the averaging kernels to the model outputs. The root-mean square error associated with this effect is about 10–30‰ for TES and larger than 40‰ for
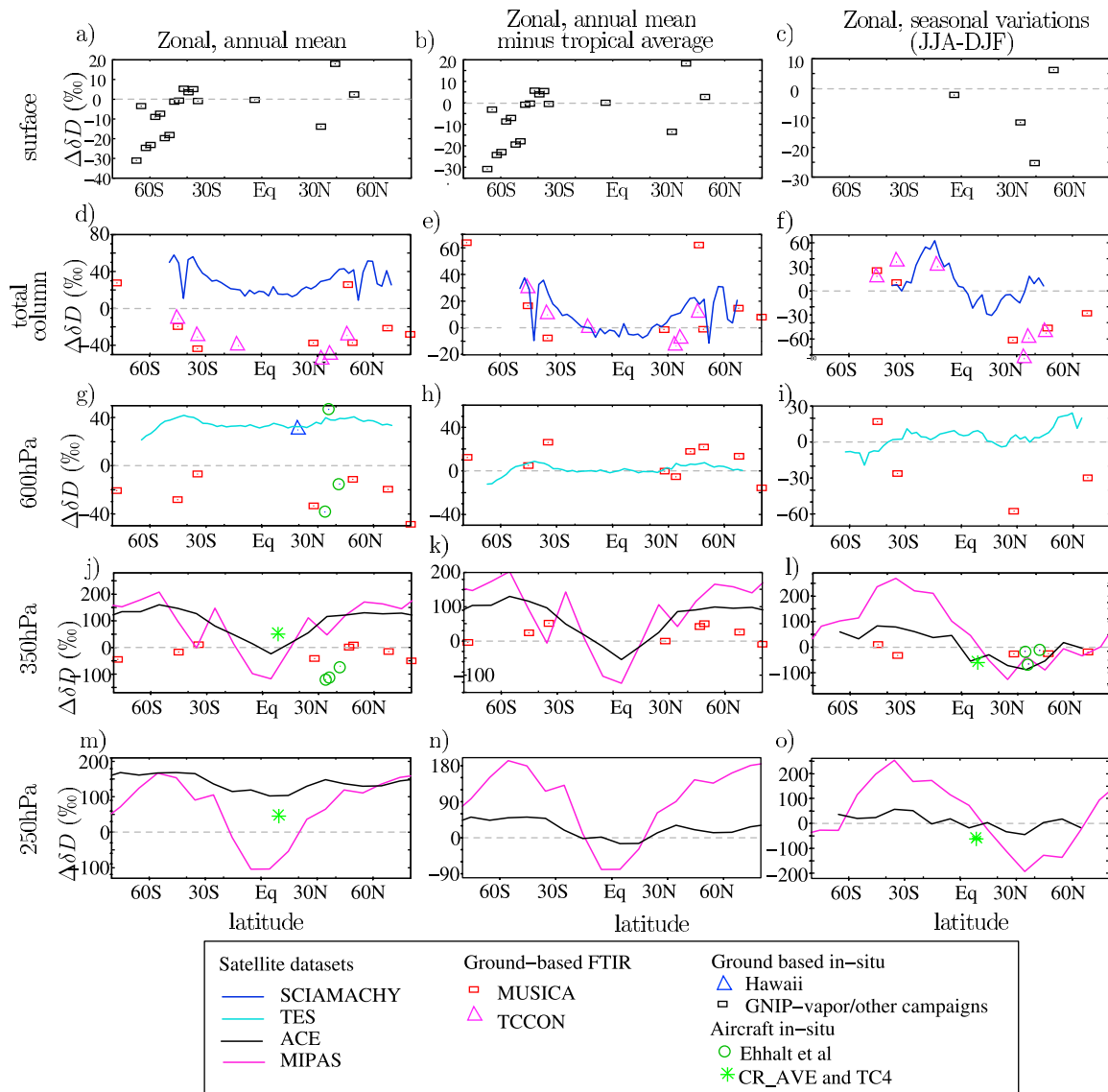
Figure 3. Same as Figure 1 but for model-data differences. All LMDZ outputs have been collocated and applied kernels with each of the corresponding data set.

MIPAS, showing the importance of taking this effect into account in the model-data comparison (Figure 2, red). Instrument sensitivity effects increase δD in deep convective regions, in the subtropics and higher latitudes for TES, and strongly increases δD for MIPAS (by about 250‰ in annual tropical average). An additional source of model-data difference can arise if the atmospheric conditions (especially the presence of clouds) in the data and the model are sufficiently different to affect the kernels used in the convolution. We estimated this effect for TES (Appendix C1) and show that it is small (6‰).

[48] The remote-sensing instruments used in this study preferentially sample clear-sky conditions. This sampling effect is taken into account by the collocation if the model simulates clouds exactly at the right place and time. If not, then the clear-sky bias exhibited by the data will be underestimated by the collocated outputs. To estimate an upper bound for this source of uncertainty, we performed a test in

which we rejected the cloudiest 30% of scenes among all collocated outputs. Then we examined the difference between monthly δD with and without this additional cloud mask. We assume that the arbitrary 30% threshold is sufficiently high to give an upper bound estimate for the cloud effect. The root-mean square errors associated with this effect are about 5–10‰ in SCIAMACHY, TES, ACE and MIPAS, and can reach up to 20‰ at some ground-based FTIR stations (Figure 2, green). Hereafter we will thus focus on signals that are larger than those values. The clear-sky sampling bias effect increases (decreases) δD in large-scale ascent (descent) regions in lower tropospheric measurements, and has little coherent effect in the upper-troposphere.

5.2. Meridional Gradient

[49] Figures 3a, 3d, 3g, 3j, and 3m show model-data differences at different levels. The model-data agreement is quite good at the surface, with model-data differences within
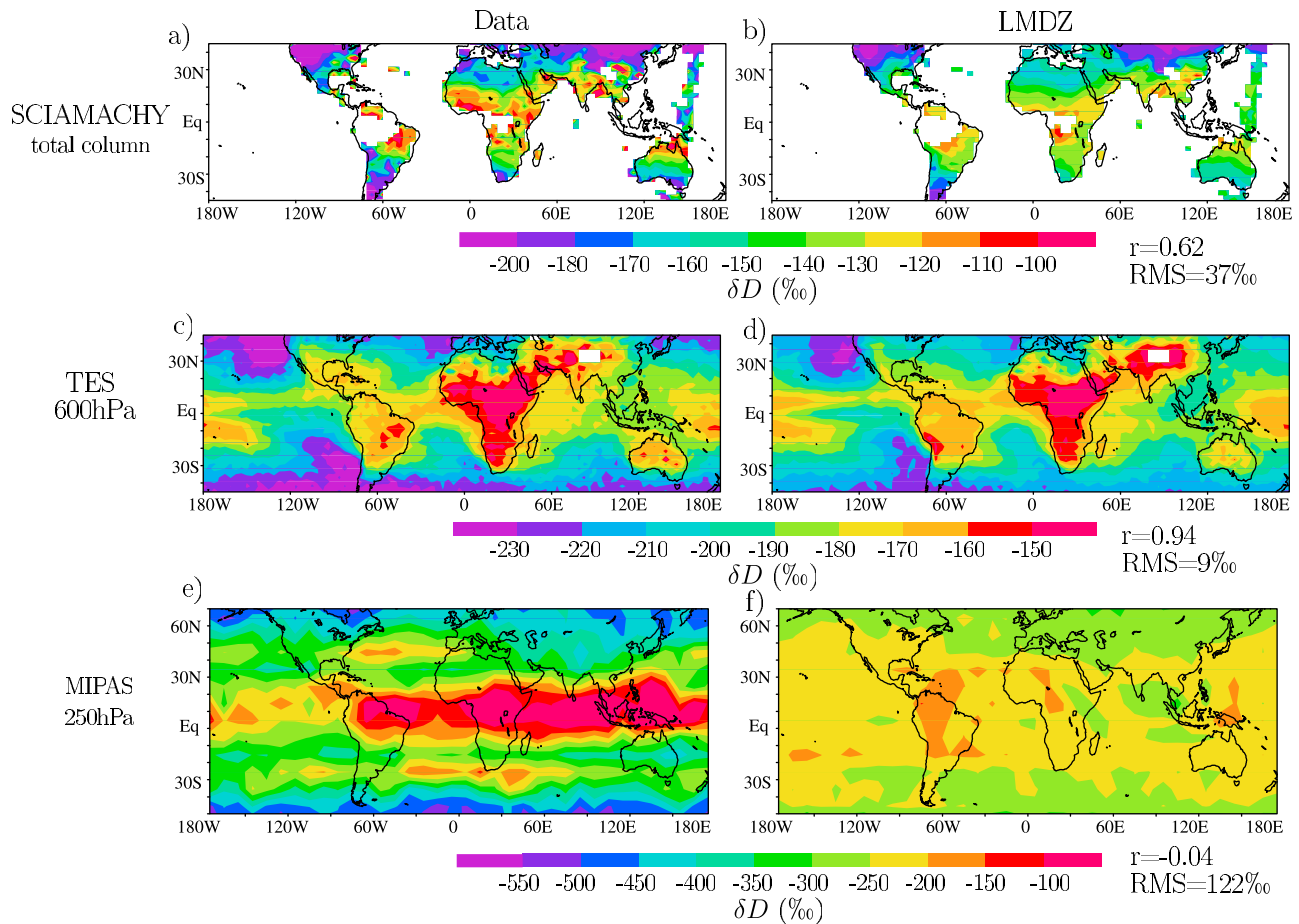
Figure 4. Annual mean δD at different levels (a, c, and e) measured by satellite instruments and (b, d, and f) simulated by LMDZ: Figure 4a shows total column δD measured by SCIAMACHY, and Figure 4b shows δD at 600 hPa measured by TES and δD at 350 hPa measured by MIPAS. Correlation coefficients and RMS differences are calculated for model-data comparisons within 45°S–45°N.

30‰ (Figure 3a). Model-data differences increase with altitude, reaching 200‰ in the upper troposphere compared to several data sets. There are systematic offsets between the model and the data, and the signs of the offsets differ between data sets. For example, simulated total column δD is higher than observed by SCIAMACHY, but lower than observed by ground-based FTIR (Figure 3d). This arises from systematic errors in the data sets (section 4.2), so we focus on spatiotemporal variations.

[50] Figures 3b, 3e, 3h, 3k and 3n show the model-data differences to which we subtracted the annual tropical average. The simulated meridional gradient is slightly too strong at the surface in mid and high latitudes. In contrast, in the free troposphere, simulated meridional gradients are too weak compared to all satellite and ground-based FTIR data sets. For example, the model-data difference for total column δD is about 20‰ higher in the midlatitudes (i.e. 45°N or 45°S) than in the tropics compared to both SCIAMACHY and ground-based FTIR (Figure 3e). The meridional gradient between the equator and midlatitudes is thus about 30% weaker in LMDZ than in the data. Similarly at 350 hPa, the model-data difference is 70‰ (respectively 300‰) higher in midlatitude and in the subtropics than in the tropics compared to ACE (respectively MIPAS) (Figures 3k and 3n).

[51] The fact that LMDZ underestimates the meridional gradient throughout the troposphere but not at the surface suggests that an overestimated evaporative recycling along poleward trajectories is not responsible for the model bias. Rather, the bias could be due to overestimated vertical mixing (which transports high δD values upward) in midlatitudes, to overestimated mixing between the tropics and midlatitudes (which smoothes the gradient), or to underestimated convective detrainment of condensate in the tropics (which increase upper tropospheric δD in the tropics [*Moyer et al.*, 1996] (also P2).

[52] In the upper troposphere, the strong underestimate of the simulated δD meridional gradient could be partly due to the poor representation of stratospheric δD, associated with mis-representation of the tropopause level, or of dynamical and chemical processes in the stratosphere. In MIPAS for example, we find that poleward of 45°S or 45°N, stratospheric δD accounts for more than 40% and 60% of the signal at 350 and 250 hPa respectively. However, within 25°S–25°N where much of the discrepancy in meridional gradients takes place, the model-data difference is completely due to discrepancies in tropospheric values. As a test, we replaced simulated δD by observed δD everywhere above the tropopause and applied the kernels: the
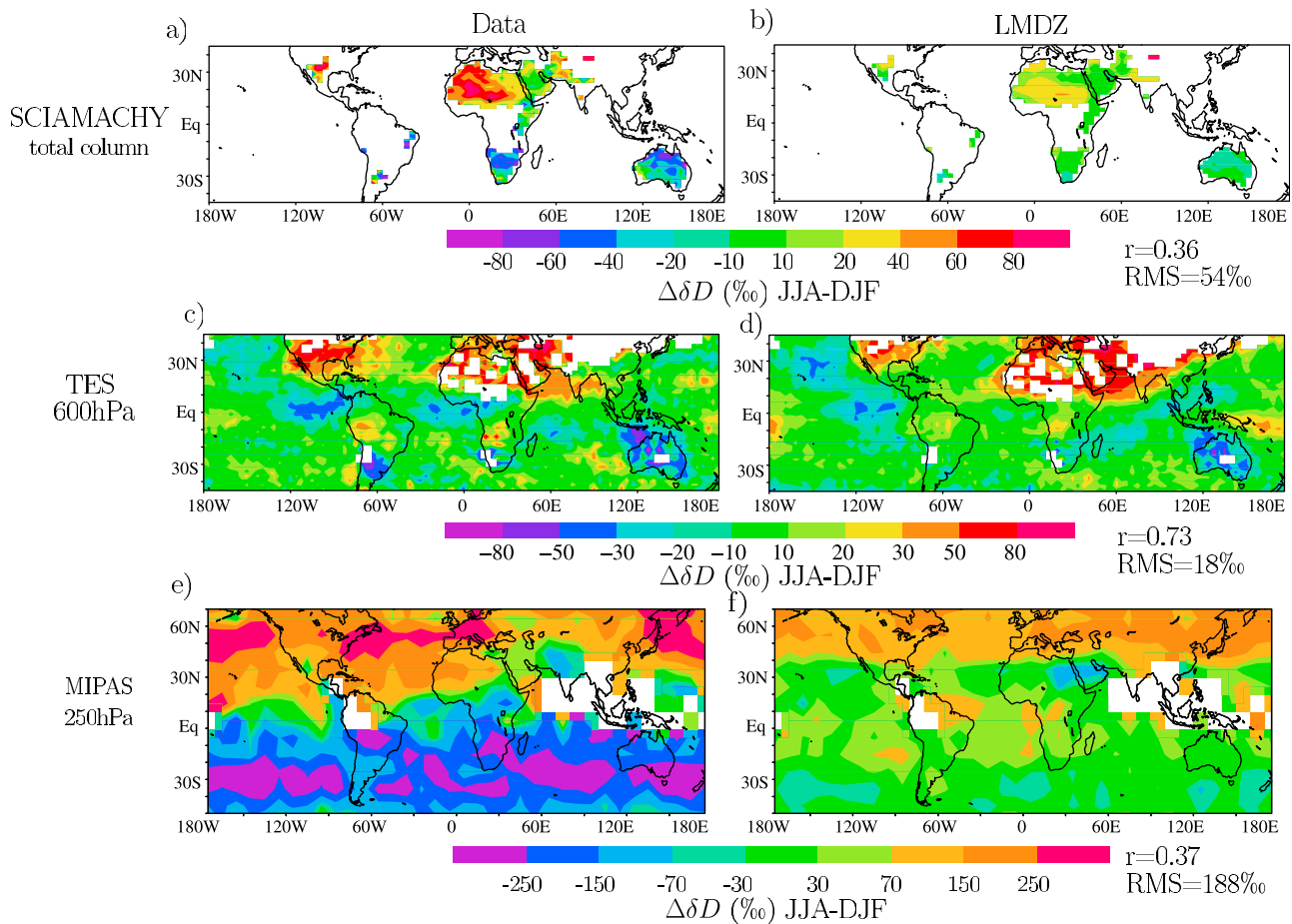
Figure 5. Same as Figure 4 but for seasonal variations (June–July August minus December–January–February).

impact on kernel-weighted δD values was smaller than 10‰ at both 350 and 250 hPa.

5.3. Seasonality

[53] Figures 3d, 3f, 3i, 3l, and 3o show the model-data differences for zonal mean seasonal δD variations. In the subtropics and mid latitudes of both hemispheres, LMDZ underestimates the observed seasonality. For example, at about 30°N, the δD seasonality is underestimated by about 20‰ (corresponding to a relative underestimation of −50%) compared to SCIAMACHY, by 80‰ (−75%) compared to MUSICA at Izaña both for total column and at 600 hPa, by 60‰ (−75%) compared to ACE at 350 hPa and by 80‰ (−50%) compared to MIPAS at 350 hPa. This underestimate is larger than all sources of model-data differences that we have previously described.

[54] Although the magnitude of the underestimate of the simulated δD seasonality varies depending on altitude and data set, the simulated bias in δD seasonality is robust in sign at all levels and compared to almost all data sets, with just two exceptions. The first exception is TES in the Northern Hemisphere, where model-data differences are very small. The second exception is at Wollongong at 600 and 350 hPa, where we have mentioned that the observed seasonality was shifted by two months. If looking at AMJ-OND variations, LMDZ underestimates the seasonality consistent with the

other data sets. Therefore, the consistency between almost all data sets and the large magnitudes of the model-data differences give us confidence that the underestimated δD seasonality is a robust and significant model bias.

[55] The most likely reason for this bias is investigated in detail in P2 and is shown to be overestimated vertical mixing, which preferentially increases δD in dry regions and during dry seasons.

5.4. Spatial Patterns

[56] To focus on spatial patterns, we show δD maps after subtracting the tropical mean δD of each data set and model output (Figures 4a–4d). In the lower troposphere, in the annual mean, SCIAMACHY and TES show δD maxima over deep convective regions over land, and lower δD over dry subtropical regions, such as the subsidence regions off the West coast of continents, and to a lesser extent over the Sahara. LMDZ captures these spatial patterns reasonably well, with model-data correlations of 0.63 for SCIA-MACHY and 0.94 for TES within 45°S–45°N. However, LMDZ underestimates the depletion over the driest regions. This model bias was also noticed in the GSM GCM [*Frankenberg et al.*, 2009], showing that LMDZ is not the only GCM exhibiting this problem. As for seasonality, a likely reason for this is excessive vertical mixing in dry regions (P2).
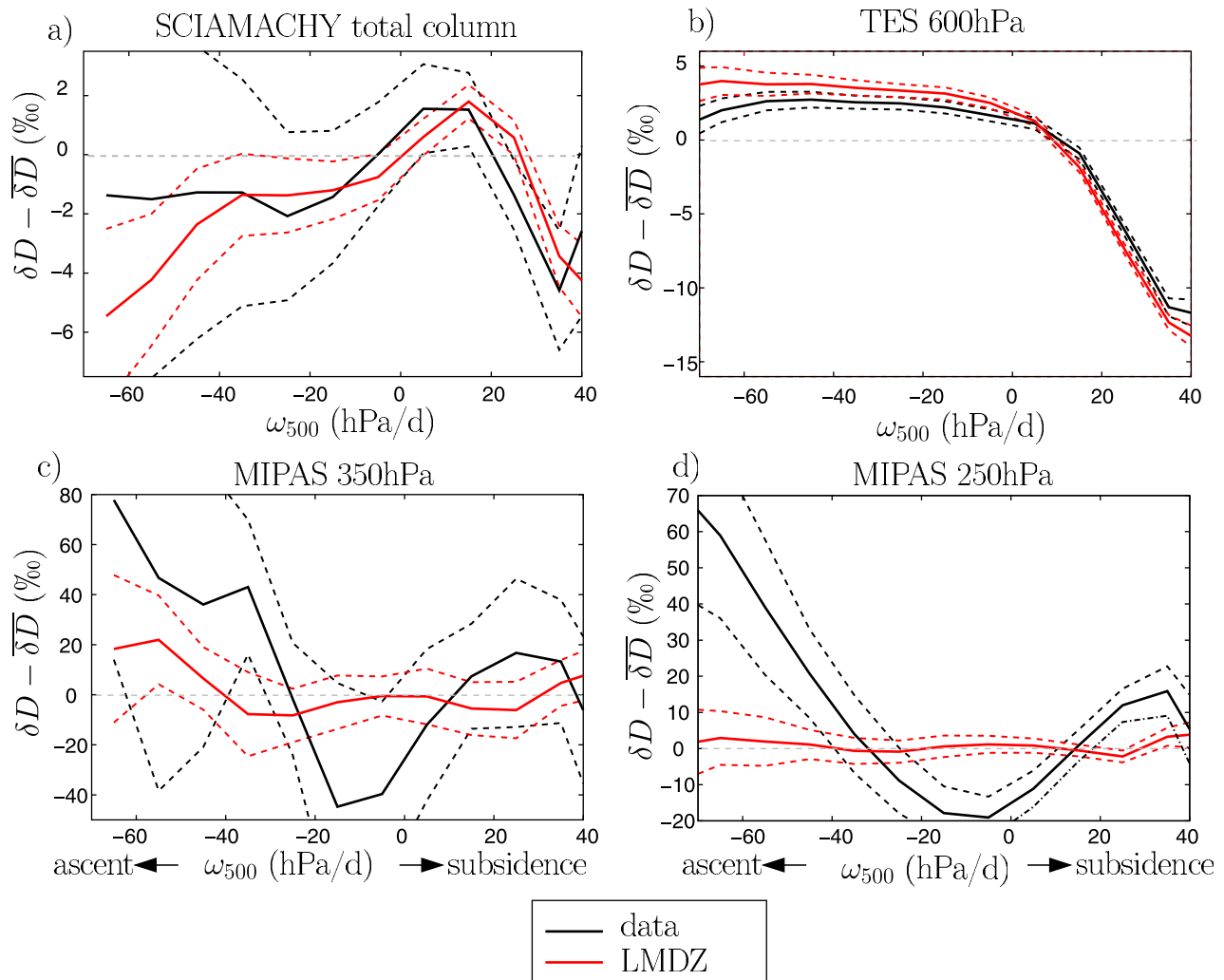
Figure 6. Composite of monthly mean δD as a function of monthly mean large-scale vertical velocity at 500 hPa (ω_{500}) in the tropics (30°S–30°N), at different levels, measured by satellite instruments (black) and simulated by LMDZ (red). We subtracted tropical mean δD to all data sets and model outputs to focus on spatiotemporal variations. (a) Total column δD compared to SCIAMACHY, (b) δD at 600 hPa compared to TES, (c) at 350 hPa compared to MIPAS and (d) 250 hPa compared to MIPAS. Negative values of ω_{500} indicate large-scale ascent (convective regions) while negative values indicate large-scale subsidence. The thick solid lines corresponds to the average, and the thin dashed lines correspond to the standard deviation divided by the square root of the number of samples, to give an envelope of confidence.

[57] In the lower troposphere, observed seasonal variations are much larger over land than over ocean and are largest over the driest land regions such as the Sahara (Figures 5a–5d). LMDZ reproduces this feature very well. LMDZ simulates a slightly lower δD during the wet season in the Warm Pool and South-East Asian region, which is not the case in TES. This model-data difference was already noticed in the CAM GCM [*Lee et al.*, 2009], though this problem is much less pronounced in LMDZ. Sensitivity tests performed both by *Lee et al.* [2009] and with LMDZ (P2) suggest that this problem is very sensitive to the fraction of precipitation arising from convection versus large-scale condensation.

[58] In the upper troposphere, only MIPAS has sufficient sampling to investigate spatial patterns, although the large discrepancies between ACE and MIPAS suggest that we

need to be very cautious when interpreting these results. MIPAS shows maximum δD over deep convective regions, and minima in the subtropics and high latitudes (Figure 4). There is a slight secondary δD maximum over the jet stream regions, for example in Northern midlatitude Atlantic. LMDZ does not capture the spatiotemporal variations of δD in the upper troposphere. Compared with the data, the simulated maximum δD in deep convective regions is too weak, consistent with the underestimated meridional gradient. The model has a maximum δD at about 30°N and 20°S, whereas the observed δD is a local minimum at these latitudes. In addition, LMDZ captures the seasonal cycle very poorly in all regions. Note that LMDZ would probably compare better to ACE, given the weaker meridional and seasonal variations in this data set.
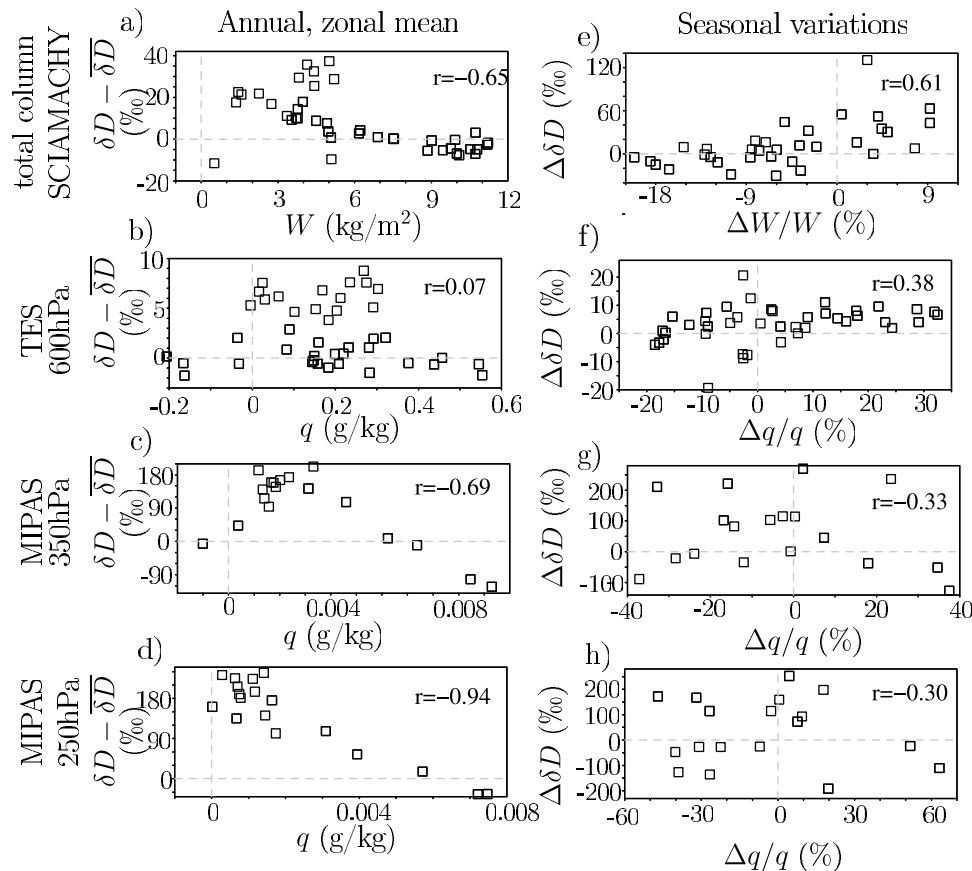
Figure 7. Relationship between biases in zonal mean δD and biases in zonal mean specific humidity. (a) Zonal, annual mean model-data difference in total column water δD as a function of zonal, annual mean model-data difference in precipitable water (W), compared with SCIAMACHY. We subtracted tropical average δD from both the data and the model to focus on meridional gradients. (b) Same as Figure 7a but for δD and specific humidity (q) at 600 hPa compared to TES. (c) Same as Figure 7b but at 350 hPa compared to MIPAS. (d) Same as Figure 7b but at 250 hPa compared to MIPAS. (e–h) Same as Figures 7a–7d but for zonal mean model-data difference of seasonal variations in total column δD as a function of zonal mean model-data difference of seasonal variations precipitable water. We normalized seasonal variations in precipitable water and show results in % to maximize the correlations. Correlation coefficients of the linear regressions are indicated. Note that for clarity in the notations, the Δ sign standing for bias (i.e. model-data difference) has been omitted on the labels. In Figures 7e–7h, the Δ sign stands for seasonal variations.

[59] To better visualize the δD contrast between wet and dry regions or seasons in the tropics, we classify the data and model outputs into dynamical regimes based on large-scale vertical velocity at 500 hPa (ω_{500}) as suggested by [*Bony et al.*, 2004] (Figure 6). We use ω_{500} from the ECMWF reanalysis for both the data and simulations, since simulations were nudged by the ECMWF winds. In the lower troposphere, for $\omega_{500} < 15$ hPa/day, observed δD decreases as the dynamical regime becomes more convective in SCIAMACHY (consistent with the amount effect) and remains almost constant in TES. For $\omega_{500} > 15$ hPa/day, observed δD strongly decreases as subsidence increases. This behavior is very well captured by LMDZ, although the ω_{500} threshold between the two regimes is slightly higher for LMDZ than for SCIAMACHY, explaining the overestimated δD in very dry regions. In the upper troposphere, observed δD increases as the dynamical regime becomes more convective for $\omega_{500} < 20$ hPa/day. LMDZ very poorly reproduces this behavior.

5.5. Link With Biases in Humidity

[60] The goal of this section is to investigate to what extent model-data differences in δD are linked to those in humidity, to assess the added value of δD measurements compared with humidity measurements alone. Traditionally, the isotopic distribution has been interpreted in terms of progressive depletion in deuterium as specific humidity (q) decreases, following Rayleigh distillation [e.g., *Dansgaard*, 1964; *Worden et al.*, 2007]. Spatially at the global scale [e. g., *Worden et al.*, 2007], or temporally in dry regions [e.g., *Galewsky et al.*, 2007; *Brown et al.*, 2008], δD has been shown to be strongly related to q. However, this does not necessarily mean that δD provides the same information as q. In this section, rather than looking at the relationship between q and δD, we look at the link between biases in q and biases in δD. Slight deviations from the Rayleigh-like behavior can arise from mixing between different air masses
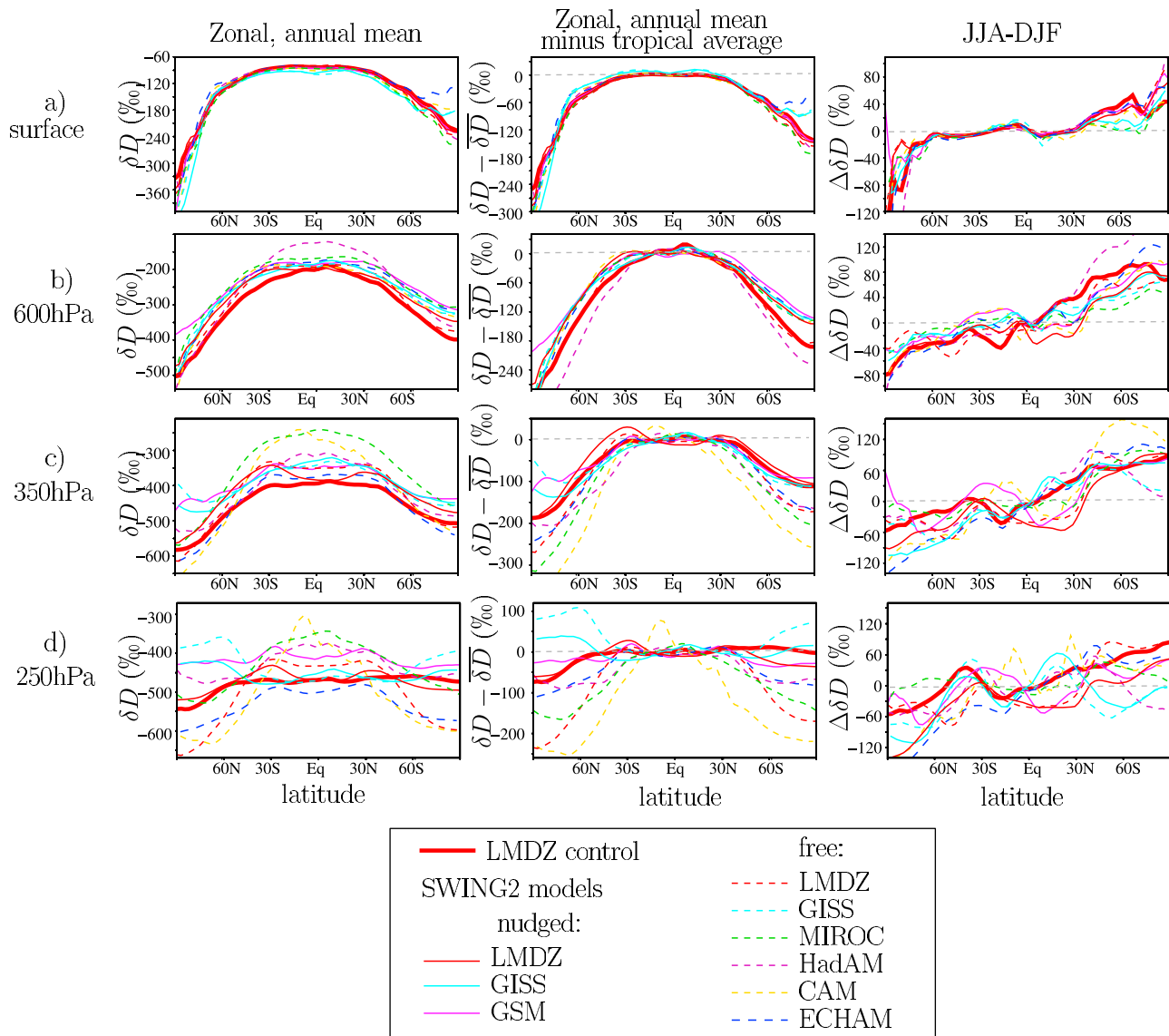
Figure 8. Similar to Figure 1 but for all SWING2 GCM simulations. The LMDZ control simulation (different from that given to SWING2), is also shown in thick red. Note that we focus here more on the spread between models than on the behavior of individual models.

[e.g., *Galewsky and Hurley*, 2010], from detrainment of condensate [e.g., *Moyer et al.*, 1996; *Dessler and Sherwood*, 2003] or from rainfall evaporation [*Worden et al.*, 2007; *Field et al.*, 2010]. If these processes are not appropriately simulated, then biases in δD differ from those in q.

[61] Figures 7a–7d show zonal, annual mean biases in δD as a function of zonal, annual mean biases in q. Correlations are either weak or negative. This means that in LMDZ, the underestimated meridional gradient in δD compared to all satellite data sets is not associated with an underestimated meridional gradient in q. Regions where δD is most overestimated compared to the global mean are not regions where q is most over-estimated. LMDZ overestimates q at all free tropospheric levels compared to all satellite data sets. This bias has been noticed in many GCMs for more than a decade [*Soden and Bretherton*, 1994; *Roca et al.*, 1997; *Allan et al.*, 2003; *Brogniez et al.*, 2005; *Pierce et al.*, 2006; *John and Soden*, 2007; *Chung et al.*, 2011]. This moist bias

is most pronounced in tropical and sub-tropical regions, whereas the high bias in δD is most pronounced in the subtropics and in midlatitudes. Therefore, the fact that LMDZ underestimates the meridional gradient in δD provides additional information on the model representation of physical processes compared to the information derived from q only.

[62] Figures 7e–7h show the relationship between biases in q seasonality and biases in δD seasonality. We use the relative seasonality in q (i.e. we normalize it by annual mean q) for both the data and for LMDZ. In this plot we would obtain an approximately straight line if δD and q were linked by pure Rayleigh distillation. In the lower troposphere, the correlation between bias in δD seasonality and bias in q seasonality is 0.61 (Figure 7b). This means that LMDZ underestimates the δD seasonality more so in regions where it also underestimates the relative seasonality in q. Therefore, the underestimated relative seasonality of q could
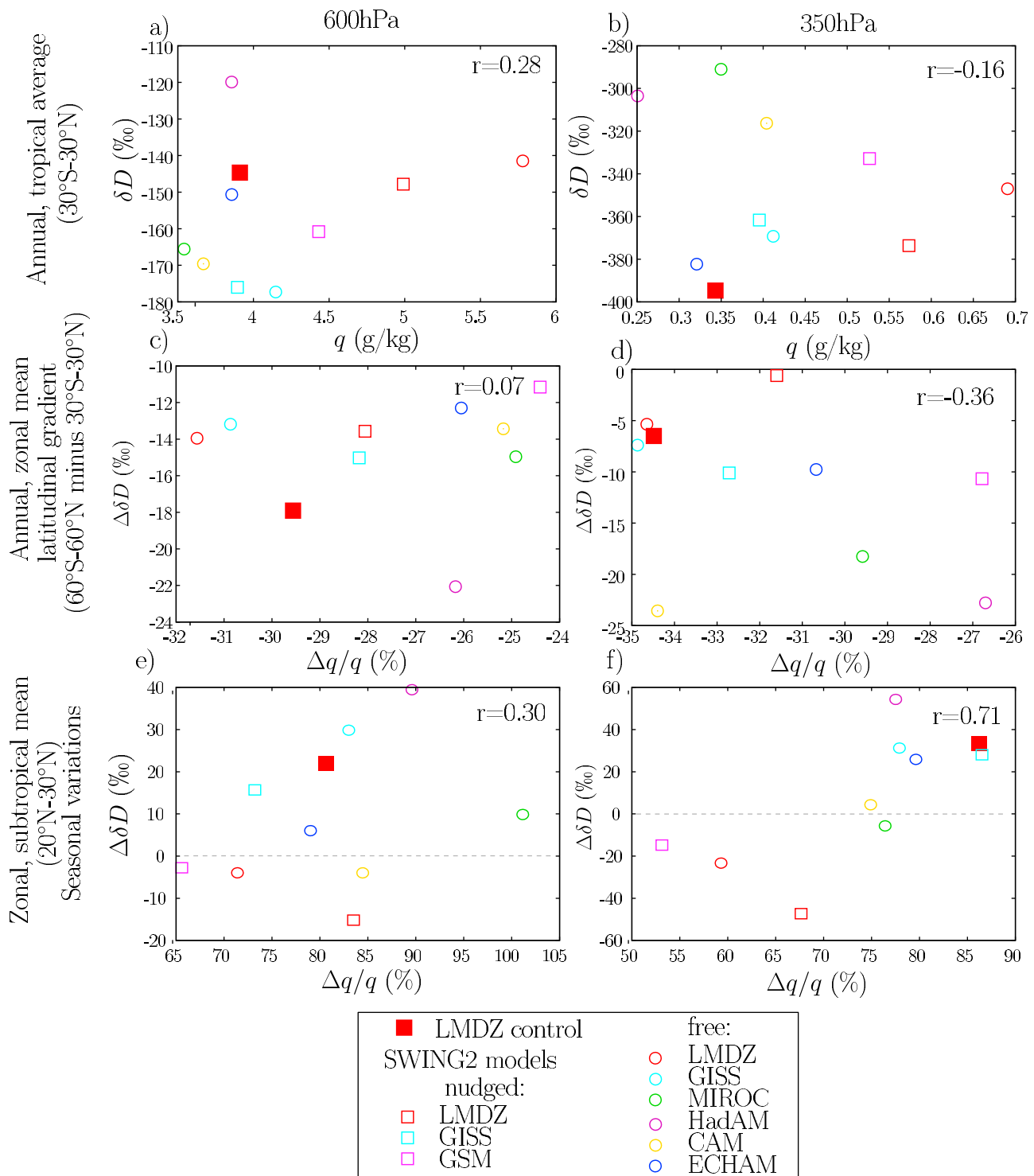
Figure 9. Link between δD features and humidity features among the different SWING2 models. The LMDZ control simulation is also shown in filled red square. (a) Annual mean δD as a function of annual mean specific humidity (q) in average over the tropics (30°S–30°N) at 600 hPa. (c) Zonal, annual mean meridional gradient in δD as a function of zonal, annual mean meridional gradient in q at 600 hPa. The meridional gradient is expressed as the difference between the average over 60°S–60°N and over 30°S–30°N. For q, the difference is normalized by the tropical average and expressed in %, to try to maximize the correlation between δD and q. (e) Subtropical average (20°N–30°N) of seasonal variations in δD as a function of subtropical average of seasonal variations in q. The seasonal difference in q is normalized by the annual mean and expressed in %, again to try to maximize the correlation between δD and q. (b, d, and f) Same as Figures 9a, 9c, and 9e but at 350 hPa.
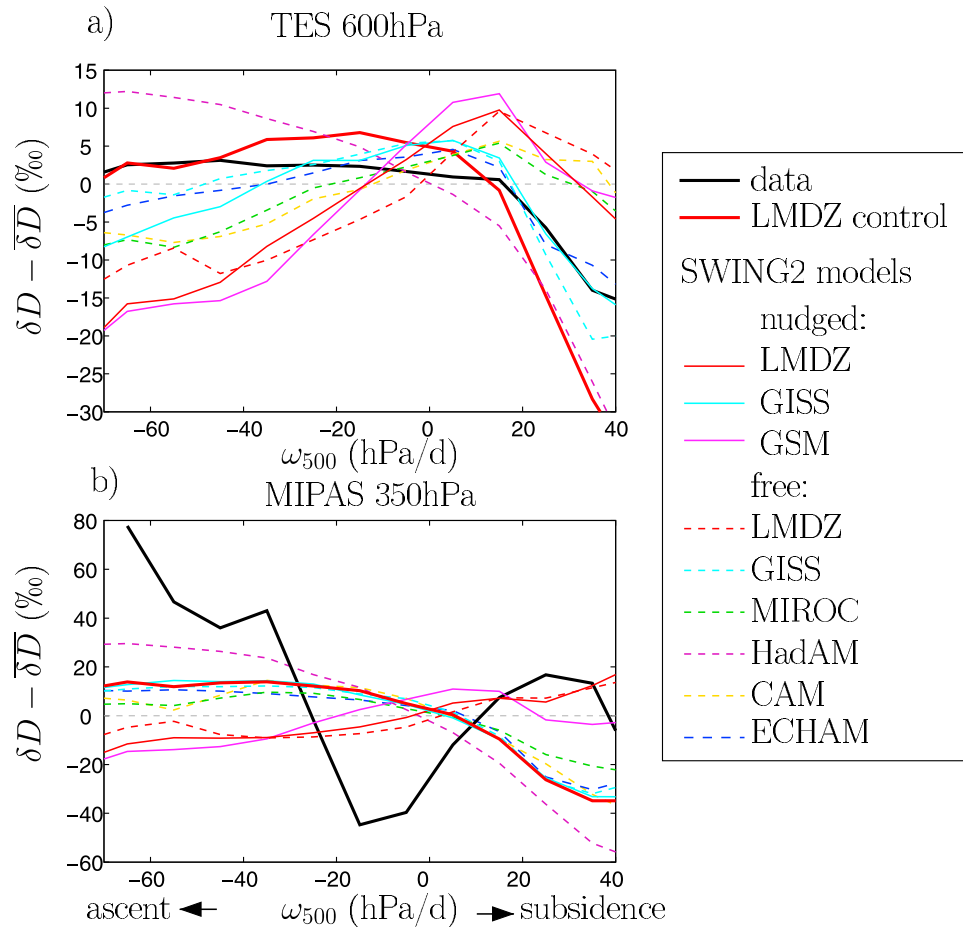
Figure 10. Composite of monthly mean δD as a function of monthly mean large-scale vertical velocity at 500 hPa (ω_{500}) in the tropics (30°S–30°N) at different levels, simulated by the different SWING2 models: (a) δD at 600 hPa and (b) at 350 hPa. We subtracted tropical mean δD to all model outputs to focus on spatiotemporal variations. The LMDZ control simulation is also shown in thick red. The δD measured by TES and MIPAS is also shown (black) as a crude comparison. No exact comparison can be done with the data in this Figure since none of these simulations have been collocated and kernel-weighted with the data. Negative values of ω_{500} indicate large-scale ascent (convective regions) while negative values indicate large-scale subsidence.

partly contribute to the underestimated δD seasonality. However, the correlation is weaker in the mid-troposphere and is negative in the upper troposphere. Therefore, the underestimated seasonality in δD cannot be explained by the underestimated relative seasonality in q in the mid and upper troposphere. P2 suggests that overestimated vertical mixing explain this underestimated seasonality.

[63] The spatial correlation between biases in q and biases in δD is weak for all satellite data sets for both annual mean and seasonal variations: for example, within 45°N–45°S, the correlation between annual mean biases in δD and in q is 0.01 for SCIAMACHY and 0.02 for TES, and the correlation between seasonal variation biases in δD and in q is 0.10 for SCIAMACHY and 0.21 for TES. Therefore, model-data differences in δD provide a different information on models shortcomings compared to humidity measurements only, showing the added value of isotopic measurements.

[64] To summarize this section, compared to satellite data sets, LMDZ successfully reproduces spatial patterns of δD in the lower troposphere, but underestimates meridional

gradients and contrasts between dynamical regimes in the upper troposphere. The most robust model data-difference is the underestimated δD seasonality in the subtropics and midlatitudes, which occurs at all levels in the vast majority of data sets.

6. Common Model-Data Differences Among Models

[65] Having established the robust features and biases in LMDZ using all data sets, we compare the spatiotemporal distribution of δD simulated by a set of 7 GCMs from the SWING2 project. The goal is threefold: (1) identify model-data differences that are robust across models; (2) assess the isotopic spread among models to see to what extent water isotopic measurements could help discriminate between models in their representation of processes controlling humidity; and (3) explore the link between the spread in δD and that in humidity, to evaluate the added value of isotopic measurements.

6.1. Meridional Gradient

[66] Figures 8a, 8d, 8g, and 8j show the meridional gradients simulated by the different SWING2 models. At the surface, all models give similar results within 20‰ between 60°S and 60°N. The spread in δD increases with altitude, reaching more than 200‰ at 250 hPa. The smaller spread at the surface may be explained by the fact that δD variations at the surface are partially damped by oceanic evaporation, whose δD is predicted by the same simple equation [*Craig and Gordon*, 1965] in all models.

[67] The spread in absolute δD values between models is not significantly related to that in q (Figures 9a and 9b). Although all models have a moist bias, models with the strongest moist bias are not those with the highest δD, showing again that isotopic composition provides additional information compared to q.

[68] Even after subtracting the tropical average (Figures 8b, 8e, 8h, and 8k), models show a very wide spread in meridional gradients. The difference between δD at the equator and at 60°S at 350 hPa varies from 60‰ in GISS to 220‰ in CAM. The models disagree on the sign of the gradient at 250 hPa. We showed in section 5.2 that LMDZ underestimated the meridional gradient compared to all remote-sensing data sets, and Figure 8 shows that it is quite typical of the set of models. Models with the strongest meridional gradients in δD are not the models with the strongest gradient in humidity (Figures 9c and 9d).

6.2. Seasonality

[69] The simulated JJA-DJF difference in SWING2 models agree quite well with each other at the surface, but again the spread increases with altitude (Figures 9c, 9f, 9i, and 9l). In the subtropics, at all free tropospheric levels, half of the models have higher δD values in summer, while the other half higher δD values in winter. We showed in section 5.2 that the control simulation with LMDZ underestimated the seasonality in the subtropics at all levels compared to almost all data sets by at least 50%. However, compared to SWING2 models, LMDZ has amongst the strongest seasonality. Therefore, all models are affected with this underestimated δD seasonality, and several even get the wrong sign. This problem may reveal a problem in the representation of humidity processes common to all GCMs. P2 suggests that all models overestimate vertical mixing.

[70] In the lower troposphere, models with the strongest seasonality in δD are not those with the strongest seasonality in q, again showing the added value of isotopic measurements to reveal GCM shortcomings. In the upper troposphere in contrast, models with the strongest seasonality in δD also have the strongest seasonality in q, suggesting that part of the isotopic behavior could be explained by that of q.

6.3. Spatial Patterns

[71] Figure 10 summarizes the spatial patterns in the tropics by classifying the model outputs based on ω_{500}. The data is also shown for reference but not for direct comparison, since model outputs were neither collocated nor kernel-weighted. Models show a very large spread in δD contrasts between convective and dry regions, both quantitatively and qualitatively. For example, at one extreme, at all levels, Had-AM simulates a strong increase in δD as ω_{500} decreases, with

maximum δD in convective regions. At the other extreme, at all levels, GSM simulates a strong decrease in δD as ω_{500} decreases from 20 hPa/day to below -70 hPa/day, with a pronounced δD minimum in convective regions. In some models like ECHAM, δD in convective regions is lower in the lower troposphere but higher in the upper troposphere than in subsidence regions. The behavior of LMDZ is within the model spread.

7. Conclusions and Perspectives

7.1. Conclusions

[72] We have evaluated the control version of the LMDZ GCM using a number of water vapor isotopic data sets from satellite, ground based remote-sensing and in situ techniques. This is the first time that so many data sets are being brought together and compared, that an isotopic GCM is so comprehensively evaluated for its water vapor, and using such a rigorous model-data comparison methodology. The different data sets show some consistent features: at all levels, the δD decreases with latitude, and at all levels, δD is higher during summer in the subtropics and in midlatitudes. There are however significant differences between data sets regarding absolute values of δD and the magnitude of meridional gradients and seasonality. We show that some of these differences can be explained by (1) different averaging kernels and a-priori values used in the retrievals of remote-sensing data sets, and (2) spatial and temporal sampling. A rigorous model-data comparison needs to take into account both these effects. We also show that systematic biases in the data are also likely the major source of δD difference between some data sets. The use of several data sets is therefore necessary to ensure the robustness of our conclusions. The lack of absolute calibration of remote-sensing data (e.g. due to the lack of aircraft validation campaigns for HDO) and possible discrepancies in the calibration of some in situ data sets further restricts our analysis to spatiotemporal variations.

[73] The simulated spatiotemporal distribution of δD shows strengths and weaknesses that are consistent across the different data sets used for the comparison. At the surface, in the lower and mid-troposphere, the simulation reproduces the observations reasonably well. In the upper troposphere, model-data differences are much larger, although the discrepancies between the MIPAS and ACE data sets prevent us from concluding about the magnitude of the model biases. The most consistent weakness of LMDZ is the underestimated seasonality at all levels in the subtropics and midlatitude of both hemispheres, compared to almost all data sets. Also, compared to all remote-sensing data sets, the model underestimates the meridional gradient of δD and the contrast in δD between convective and dry tropical regions. These biases in δD are not linked with those in humidity, which confirms that isotopic measurements provide additional information that can be used to expose model biases.

[74] Some of the problems that we have exposed in LMDZ are common in all SWING2 models. All models underestimate the seasonality in the subtropics at all levels, underestimate the meridional gradient compared to satellite data sets, and underestimate the δD contrast between convective and subsidence regions in the upper troposphere. However, the spread between models is very large, both quantitatively

and qualitatively. For example, models disagree on the sign of the δD contrast between convective and dry regions at all tropospheric levels. These differences between models in δD are not obviously related to those in specific humidity, suggesting that there is additional information provided by isotopic measurement compared to humidity measurements. In P2, these inter-model differences will be interpreted in terms of physical processes and will be used to expose the causes of biases in models.

7.2. Perspectives

[75] Unresolved differences between observed data sets (e. g. near-IR ground-based FTIR and SCIAMACHY or ACE and MIPAS) highlight the need for absolute calibration and improved measurement error characterization. While this is currently very difficult to achieve in practice, such a validation and calibration activity would enable GCM evaluation to be extended to include assessment of absolute values of δD, in addition to the spatiotemporal variability considered in this study. Our analysis shows that the upper troposphere is where GCMs disagree the most, so improving the calibration and frequency of measurements in this region would be particularly valuable.

Appendix A: Bias Correction in TES

[76] As discussed by *Worden et al.* [2006], TES data are known to be affected by a slight spectroscopic bias. Recent calibration studies against in situ measurements at Mauna Loa [*Worden et al.*, 2010] suggest that this bias may be corrected as follows:

$$x_{ci} = x_{di} - \sum_{k=1}^{n} (A_{xxik} \cdot 0.04)$$

where x_{ci} and x_{di} are the corrected HDO mixing ratio and the HDO mixing ratio provided in the data files, at level i, respectively; A_{xxik} is the averaging kernel for HDO and n is the number of TES levels.

Appendix B: Error Estimates for the Different Data Sets

[77] Errors estimates are either provided as part of the data sets, or calculated as a function of various retrieved quantities. When error estimates are given for H_2O and HDO independently and unless stated otherwise, we calculate the δD error based on the standard propagation of errors, as if the retrieval errors were independent:

$$E_{\delta D} = \frac{1000}{R_{SMOW}} \cdot \left(\frac{x_{err} \cdot q_{obs} + x_{obs} \cdot q_{err}}{q_{obs}^2} \right)$$

where $E_{\delta D}$ is the error on the δD retrieval, q_{obs} and x_{obs} are the observed H_2O and HDO mixing ratio at level respectively, and q_{err} and x_{err} are the errors on the observed H_2O and HDO mixing ratios respectively. The same applies for total column amounts instead of mixing ratio. In practice, retrieval errors are positively correlated, so this estimate represents an upper bound for the error in δD.

[78] When n samples are averaged, the random part of the error, when known, is divided by $\sqrt{n}$.

B1. SCIAMACHY

[79] The total error on observed δD, $E_{\delta D}$, is provided for each measurement as part of the data set. In zonal mean, annual average, the average error ranges from 50‰ in the tropics to 130‰ in mid latitudes (about 60°). If this error was random, the error on the zonal means would be reduced to about 2‰ in the tropics and up to 30‰ in mid latitudes. However, no detailed error budget is available for this data set so that the random part of the error is unknown. Systematic biases are not documented.

B2. TES Data

[80] In the TES data, the error on δD is smaller than that calculated from the errors on H_2O and HDO contents independently, due to the benefits of the joint retrieval [*Worden et al.*, 2006]. The TES retrievals provide data to calculate separately the random and non-random parts of the δD error. In the F04 version of the TES data, the total error on δD averaged over n samples at level j, $E_{\delta D}$, is calculated as:

$$E_{\delta D} = \frac{1000}{R_{SMOW}} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n} C_{non-rand,i} \cdot (R_i)^2 + \frac{1}{n^2}\sum_{i=1}^{n} C_{rand,i} \cdot (R_i)^2}$$

where $C_{non-rand,i}$ and $C_{rand,i}$ are the non-random and random parts of the H_2O/HDO error covariance respectively for retrieval i and where R_i is the retrieved H_2O/HDO ratio for retrieval i at level j. $C_{non-rand,i}$ and $C_{rand,i}$ are taken as the j-th diagonal terms of the corresponding error covariance matrices (named "HDO_H2ORatioMeasurementErrorCovariance" and "HDO_H2ORatioSystematicErrorCovariance" in the data files).

[81] We find that the precision is about 10–15‰ for individual measurements but is reduced down to 1–2‰ when calculating zonal averages.

B3. ACE

[82] $E_{\delta D}$ was calculated for each profile as following the standard propagation of errors as described earlier. As for SCIAMACHY, the random part of the error is unknown. In zonal, annual means, the errors range from about 20‰ in the tropics to about 50‰ in midlatitudes if the random contribution is zero, and are between 1 and 2‰ if the random contribution is 1.

B4. MIPAS

[83] As for TES, the error on δD is smaller than that calculated from the errors on H_2O and HDO contents independently, due to the benefits of the joint retrieval [*Steinwagner et al.*, 2007]. The different terms of the total error are given by *Steinwagner et al.* [2007]. Data files include random errors due to measurement noise only, which we consider in the following. The general formulation for the δD error for a given profile at level i, $E_{\delta D}$, is:

$$E_{\delta D} = \frac{1000}{R_{SMOW} \cdot q_{obs}^2}$$
$$\cdot \sqrt{x_{obs}^2 \cdot q_{err}^2 + q_{obs}^2 \cdot x_{err}^2 - 2 \cdot r_i \cdot x_{obs} \cdot x_{err} \cdot q_{err} \cdot q_{obs}}$$

where q_{obs} and x_{obs} are the observed H_2O and HDO mixing ratio at level i respectively, q_{err} and x_{err} are the errors on the observed H_2O and HDO mixing ratio respectively, and r_i is calculated as:

$$r_i = \frac{S_{HDO/H_2O}}{\sqrt{S_{HDO/HDO} \cdot S_{H_2O/H_2O}}}$$

where S_{HDO/H_2O}, $S_{HDO/HDO}$ and S_{H_2O/H_2O} are the i-th diagonal terms of the covariance matrices for HDO/H_2O, HDO/HDO and H_2O/H_2O respectively, provided for each profile. In practice, r_i was found to be of the order of 10^{-2}. We thus assumed $r_i = 0$ to avoid manipulating very voluminous data for little benefits.

[84] The measurement noise error for individual measurements are about 60‰ in the tropics and 150‰ in mid-latitudes. Once this measurement noise error is divided by the square root of the number of measurements, the error for zonal mean δD is between 1 and 3‰. Parameter errors (errors in the associated imperfectly known forward model parameters used in the retrieval) are of the order of 100‰ in the upper troposphere [*Steinwagner et al.*, 2007].

B5. Systematic Bias in MIPAS Related to Clouds

[85] In the absence of clouds, MIPAS observes from 70 km tangent altitude down to about 6 km tangent altitude with 3-km steps. In the presence of clouds, the spectra for the cloud-contaminated altitudes are discarded. However, propagation of systematic errors (e.g. due to undetected clouds) localized at lower altitudes (i.e. 6 km) may lead to systematic errors in δD profiles at higher altitudes [*Steinwagner et al.*, 2010].

[86] To estimate this effect, all scans which go down to the tangent altitude of 9 km were selected and test retrievals of these scans were performed omitting artificially the lowermost tangent height (i.e. 9 km). These test retrievals were then compared to the original results. This comparison shows that the retrievals from measurements which start at an altitude of 12 km instead of 9 km are biased high by 50 to 100‰.

[87] As an upper bound estimate of the impact of this systematic error on the observed δD distributions, we tried subtracting 100‰ from profiles that go down to 12 km or higher, 50‰ to profiles that go down to 9 km, and leaving the δD unchanged for profiles that go down to 6 km. At 200 ha, convective regions become slightly more depleted by about 15‰ in annual mean, and the δD seasonality in the subtropics becomes less pronounced by up to 30‰. This impact is much smaller than the model-data differences that we look at, and thus is unlikely to affect our conclusions.

B6. Ground-Based FTIR at MUSICA Sites

[88] The MUSICA profiles and column-integrated amounts are produced by an H_2O and δD optimal estimation approach. For H_2O, the ground-based NDACC FTIR systems are mainly sensitive up to about 10–15 km. The vertical resolution (full width half maximum of the kernels) is about 3 km in the lower troposphere, 6 km in the middle troposphere, and 10 km in the upper troposphere. The δD is mainly sensitive in the first 10 km above the surface and the

vertical resolution is 3 km in the lower troposphere and 10 km in the middle/upper troposphere.

[89] Measurement noise, uncertainties in the alignment of the instrument, detector non-linearities, uncertainties in the applied atmospheric temperature profiles, and uncertainties in the applied spectroscopic parameters are considered as the error sources. The propagation of these sources is estimated by a full treatment. Two retrievals are performed: a first with a correct parameter and a second with an erroneous parameter (e.g. 2 K increased lower tropospheric temperature, application of a 1% higher H_2O line strength parameter, etc.). The systematic and the random error are then given by the mean and the standard deviation of the difference between the two retrievals [*Schneider et al.*, 2010a]. For the ground-based MUSICA v101220_Ca.0 data used in this study, uncertainties in the applied atmospheric temperature profiles and the applied spectroscopic parameters are the leading random error sources, whereby uncertainties of the instrument's alignment are of secondary importance. The δD random error can reach 5‰ for column-integrated data and 15–25‰ for profiles. Systematic errors are dominated by uncertainties in the applied spectroscopic parameters. They can be 10‰ for column-integrated δD and 25–50‰ for δD profiles.

B7. Ground-Based FTIR at TCCON Sites

[90] The estimate measurement repeatability (1-σ) was calculated for each profile following the standard propagation of errors as described earlier. Annual mean measurement repeatability varies between 5‰ and 22‰ depending on sites. This is the random part of the error. The absolute calibration error cannot be readily estimated.

Appendix C: Applying Averaging Kernels to the Model Outputs

[91] The averaging kernel matrix defines the sensitivity of the retrieval at each level to the true state at each level. For a fair model-data comparison, it is necessary to take into account this sensitivity. This is done by applying to the model output the same averaging kernels as those calculated as part of the retrieval process.

C1. TES

[92] Let q and x be the volume mixing ratio in H_2O and HDO respectively. Subscripts denote the values simulated by LMDZ and interpolated on the TES retrieval grid (s), prescribed as the a-priori profile (p), or that would be measured by TES if TES were flying in LMDZ above an atmosphere similar to that predicted by the model (m). Then at level i, q_{mi} is calculated as [*Worden et al.*, 2006]:

$$\ln(q_{mi}) = \ln(q_{pi}) + \sum_{k=1}^{n} \left(A_{qqik} \cdot \left(\ln(q_{sk}) - \ln(q_{pk}) \right) \right) \quad (C1)$$

where A_{qq} is the averaging kernel for H_2O provided in the TES data set.

[93] The HDO mixing ratio is calculated similarly but involves cross terms due to the fact that H_2O and HDO are

jointly retrieved. The isotopic mixing ratio $R = x/q$ is thus calculated as:

$$\ln(R_{mi}) = \ln(R_{pi}) + \sum_{k=1}^{n} \left((A_{xxik} - A_{qxik}) \cdot (\ln(x_{sk}) - \ln(x_{pk})) \right.$$
$$\left. - (A_{qqik} - A_{xqik}) \cdot (\ln(q_{sk}) - \ln(q_{pk})) \right) \quad (C2)$$

where A_{xx} is the averaging kernel for H_2O and A_{xq} and A_{qx} are the cross kernels.

[94] Averaging kernels depend on surface temperature and atmospheric state, including the presence of clouds [*Lee et al.*, 2011]. However, since the averaging kernels for each individual profiles are computationally voluminous, and because the atmospheric conditions associated with a particular kernel in TES may be different in LMDZ despite the nudging, we did not attempt to use individual averaging kernels for each profile. Instead, we averaged averaging kernels for each month and each LMDZ grid box. The root mean square error between monthly mean model outputs transformed through individual kernels and model outputs transformed with monthly mean kernels is about 6‰ in average between 45°S–45°N. Therefore, using monthly mean kernels, and thus neglecting the day-to-day variability in the kernels, is a reasonable simplification.

[95] Another possible problem in comparisons is that differences in cloud properties as observed by TES and as simulated by LMDZ can also contribute to some uncertainty in the exact kernels to use. To quantify this effect, we made a test in which we applied to the model output the monthly mean kernels that were calculated after eliminating 30% of the cloudiest scenes in TES. The root mean square difference between monthly mean model outputs transformed by all-skies kernels and monthly mean model outputs transformed by clear-sky kernels is about 6‰ on average between 45°S–45°N. Therefore, the sources of model-data difference associated with kernel convolution is much smaller than the isotopic signals we look at in this paper. However, to examine smaller signals, for example compare isotopic signatures associated with clear sky and different cloud conditions, then a better account of cloud effects will be needed. This could be achieved, for instance, by taking advantage of the CALIPSO cloud data set [*Winker et al.*, 2007], which can allow collocating with TES, and whose observations can be emulated by the model [*Chepfer et al.*, 2008]. Incorporating this type of calculation in the present set of comparisons is beyond the scope of this paper.

C2.   MIPAS

[96] For MIPAS, the convolution is similar to equations (C1) and (C2) except that logarithms are not used and there are no cross terms. As for TES, we did not use individual averaging kernels for each profile. However, because averaged averaging kernels may lead to a broadening of the kernels, we used representative averaging kernels instead, as described below.

[97] The main source of kernel variability in MIPAS is the tropopause height. We binned individual profiles based on their tropopause height from 7 km to 17 km by bins of

0.5 km. The tropopause height was defined following the World Meteorological Organization as the lower bound of a layer in which the lapse rate is lower than 2 K/km, provided that this layer is at least 2 km thick. Then for each bin, we calculated an average profile of diagonal elements $\overline{A_{qqii}}$ and $\overline{A_{xxii}}$, and we selected the most "representative" kernel as the one minimizing $\sum_{i=1}^{n} \left(A_{qqii} - \overline{A_{qqii}} \right)^2 + \left(A_{xxii} - \overline{A_{xxii}} \right)^2$. This calculation was performed on 216 profiles collected during three days representative of different seasons. We checked on these 216 samples that using representative kernels rather than the individual kernels leads to an average error of about 50‰ only on the kernel-weighted δD at 200 hPa, which is within the measurement uncertainty.

[98] These representative averaging kernels were then applied to LMDZ output depending on the observed tropopause height for the time and location of each measurement. Due to systematically higher tropopause in LMDZ, we chose to select the appropriate representative averaging kernels depending on *observed*, and not *simulated*, tropopause height. This allows us to focus on the biases related to the isotopic composition rather than biases due to the tropopause height.

C3.   Ground-Based FTIR at MUSICA Sites

[99] For the ground-based FTIR data produced in the framework of MUSICA, the convolution is the same as for TES except that the cross terms involve the $H_2^{18}O$ and $H_2^{17}O$ mixing ratios as well. Since the $H_2^{17}O$ distribution simulated by a GCM has never been evaluated yet, we calculate the $H_2^{17}O$ mixing ratio based on the $H_2^{18}O$ mixing ratio, assuming an ^{17}O-excess of 20 permeg (consistent with orders of magnitudes given by *Landais et al.* [2010] and *Luz and Barkan* [2010]). The effect of this assumption on kernel-weighted δD profiles can be neglected.

C4.   Ground-Based FTIR at TCCON Sites

[100] For the ground-based FTIR at TCCON sites, the convolution transforms simulated profiles of specific humidity into total column water mass (Q) that would be observed by the instrument [*Rodgers and Connor*, 2003]:

$$Q = \sum_{k=1}^{n} \frac{\Delta P_k}{g} \cdot \left(A_k \cdot q_{sk} + (1 - A_k) \cdot q_{pk} \right)$$

where A_k is the column averaging kernel profile at level k, q_{sk} and q_{pk} are respectively the simulated and a priori specific humidity at level k, ΔP_k is the level thickness in the retrieval grid and g the gravity. After applying a similar equation for HDO, total column δD is finally calculated.

[101] A priori profiles are provided for every day. Once again, individual kernels are computationally voluminous. Since averaging kernels depend mainly on the solar zenith angle, we calculated a set of representative averaging kernels for different bins of solar zenith angles. To assess the error on the kernel-weighted δD resulting from using these representative averaging kernels rather than the individual kernels for each measurement, we tried using individual averaging kernels at the Lauder site over 2004–2007 (40% of all measurements). The difference between δD transformed by

individual kernels and with representative kernels is lower than 3‰. This confirms that using these representative averaging kernels is a good approximation.

Appendix D: Using LMDZ to Quantify Sources of Differences Between Data Sets

[102] Figure 1 shows large differences in δD between the different data sets. These differences can be explained by (1) spatiotemporal sampling, (2) instrument sensitivity, and (3) errors in each data sets. These three sources of differences are difficult to quantify directly since the data coverage is insufficient to quantify the effect of spatiotemporal sampling, and since vertical profiles through the troposphere are not available to explore the effect of instrument sensitivity. Therefore, we use LMDZ to quantify these three sources of differences between data sets.

[103] Hereafter, we consider averages over a given spatiotemporal domain. The average δD observed by instrument i, noted $\delta D_{i,obs,data}$, can be expressed as:

$$\delta D_{i,obs,data} = \delta D_{real,data} + \Delta \delta D_{i,colloc,data} + \Delta \delta D_{i,convol,data} + \Delta \delta D_{i,error,data} \tag{D1}$$

where $\delta D_{real,data}$ is the real average δD, which is independent of the instrument and will never be exactly known, $\Delta \delta D_{i,colloc,data}$ is the effect of spatiotemporal sampling, which can be taken into account in the model by collocation, $\Delta \delta D_{i,convol,data}$ is the effect of instrument sensitivity, which can be taken into account in the model by kernel convolution, and $\Delta \delta D_{i,error,data}$ are all the errors (e.g. spectroscopic) affecting the measurement.

[104] Similarly, in the model,

$$\delta D_{i,obs,model} = \delta D_{real,model} + \Delta \delta D_{i,colloc,model} + \Delta \delta D_{i,convol,model}$$

where

$$\Delta \delta D_{i,colloc,model} = \delta D_{i,colloc,model} - \delta D_{real,model}$$

and

$$\Delta \delta D_{i,convol,model} = \delta D_{i,obs,model} - \delta D_{i,colloc,model}$$

and where $\delta D_{real,model}$ is the average raw simulated δD, $\delta D_{i,obs,model}$ is the average simulated δD after both collocation and convolution and $\delta D_{i,colloc,model}$ is the average simulated δD after collocation only.

[105] We assume that LMDZ reproduces spatiotemporal δD patterns sufficiently well to predict correctly the effects of spatiotemporal sampling and instrument sensitivity:

$$\Delta \delta D_{i,colloc,data} = \Delta \delta D_{i,colloc,model} + \epsilon_{i,colloc}$$

and

$$\Delta \delta D_{i,convol,data} = \Delta \delta D_{i,convol,model} + \epsilon_{i,convol}$$

where $\epsilon_{i,colloc}$ and $\epsilon_{i,convol}$ are possible effects (hopefully small) of problems in the simulated δD patterns and of sub-daily and sub-grid sampling effects not resolved by our collocation. Their sum is noted ϵ_i.

[106] The difference between two data sets i and j can thus be decomposed as:

$$\delta D_{i,obs,data} - \delta D_{j,obs,data} = \left(\Delta \delta D_{i,colloc,model} - \Delta \delta D_{j,colloc,model} \right)$$
$$+ \left(\Delta \delta D_{i,convol,model} - \Delta \delta D_{j,convol,model} \right)$$
$$+ \left(\Delta \delta D_{i,error,data} - \Delta \delta D_{j,error,data} + \epsilon_i - \epsilon_j \right)$$

The first term on the right hand side is the effect of spatiotemporal sampling and the second is the effect of instrument sensitivity. These two terms are calculated from LMDZ outputs. The third term combines errors in each data set, possible problems in simulated δD patterns, calculated as a residual and sub-daily and sub-grid sampling effects.

[107] These terms are evaluated in Table 6. In the table headers and in section 4.2, for brevity we note $\Delta \delta D_{obs} = \delta D_{i,obs,data} - \delta D_{j,obs,data}$, $\Delta \delta D_{colloc} = \Delta \delta D_{i,colloc,model} - \Delta \delta D_{j,colloc,model}$, $\Delta \delta D_{convol} = \Delta \delta D_{i,convol,model} - \Delta \delta D_{j,convol,model}$ and $\Delta \delta D_{errors} = \Delta \delta D_{i,error,data} - \Delta \delta D_{j,error,data} + \epsilon_i - \epsilon_j$.

References

Aghedo, A. M., K. W. Bowman, D. T. Shindell, and G. Faluvegi (2011), The impact of orbital sampling, monthly averaging and vertical resolution on climate chemistry model evaluation with satellite observations, *Atmos. Chem. Phys.*, *11*, 6493–6514.

Allan, R., M. A. Ringer, and A. Slingo (2003), Evaluation of moisture in the Hadley Centre climate model using simulations of HIRS water-vapour channel radiances, *Q. J. R. Meteorol. Soc.*, 129(595), 3371–3389.

Angert, A., J.-E. Lee, and D. Yakir (2008), Seasonal variations in the isotopic composition of near-surface water vapour in the eastern Mediterranean, *Tellus, Ser. B*, 60(4), 674–684.

Bony, S., and J. Dufresne (2005), Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models, *Geophys. Res. Lett.*, 32, L20806, doi:10.1029/2005GL023851.

Bony, S., and K. A. Emanuel (2001), A parameterization of the cloudiness associated with cumulus convection: Evaluation using TOGA COARE data, *J. Atmos. Sci.*, 58, 3158–3183.

Bony, S., J.-L. Dufresne, H. Le Treut, J.-J. Morcrette, and C. Senior (2004), On dynamic and thermodynamic components of cloud changes, *Clim. Dyn.*, 22, 71–86.

Bony, S., et al. (2006), How well do we understand and evaluate climate change feedback processes?, *J. Clim.*, 19, 3445–3482.

Bony, S., C. Risi, and F. Vimeux (2008), Influence of convective processes on the isotopic composition ($\delta^{18}O$ and δD) of precipitation and water vapor in the tropics: 1. Radiative-convective equilibrium and TOGA-COARE simulations, *J. Geophys. Res.*, 113, D19305, doi:10.1029/2008JD009942.

Brogniez, C., R. Roca, and L. Picon (2005), Evaluation of the distribution of subtropical free tropospheric humidity in AMIP-2 simulations using METEOSAT water vapor channel data, *Geophys. Res. Lett.*, 32, L19708, doi:10.1029/2005GL024341.

Brown, D., J. Worden, and D. Noone (2008), Comparison of atmospheric hydrology over convective continental regions using water vapor isotope measurements from space, *J. Geophys. Res.*, 113, D15124, doi:10.1029/2007JD009676.

Chepfer, H., S. Bony, D. Winker, M. Chiriaco, J.-L. Dufresne, and G. Sèze (2008), Use of CALIPSO lidar observations to evaluate the cloudiness simulated by a climate model, *Geophys. Res. Lett.*, 35, L15704, doi:10.1029/2008GL034207.

Chung, E.-S., B. J. Soden, B.-J. Sohn, and J. Schmetz (2011), Model-simulated humidity bias in the upper troposphere and its relation to the large-scale circulation, *J. Geophys. Res.*, 116, D10110, doi:10.1029/2011JD015609.

Craig, H. (1961), Isotopic variations in meteoric waters, *Science*, 133, 1702–1703.

Craig, H., and L. I. Gordon (1965), Deuterium and oxygen-18 variations in the ocean and marine atmosphere, in *Stable Isotope in Oceanographic Studies and Paleotemperatures*, edited by E. Tongiogi, pp. 9–130, V. Lishi e Figli, Pisa, Italy.

Dansgaard, W. (1964), Stable isotopes in precipitation, *Tellus*, 16, 436–468.

Dessler, A. E., and S. C. Sherwood (2003), A model of HDO in the tropical tropopause layer, *Atmos. Chem. Phys.*, 3, 2173–2181.

Deutscher, N., et al. (2010), Total column CO_2 measurements at Darwin Australia—Site description and calibration against in situ aircraft, *Atmos. Meas. Technol.*, 3, 989–1021.

Dupuy, E., et al. (2009), Validation of ozone measurements from the Atmospheric Chemistry Experiment (ACE), *Atmos. Chem. Phys.*, 9, 287–343, doi:10.5194/acp-9-287-2009.

Ehhalt, D. H. (1974), Vertical profiles of HTO, HDO, and H_2O in the troposphere, *NCAR Tech. Note NCAR-TN-STR-100*, 138 pp., NCAR, Boulder, Colo.

Ehhalt, D. H., F. Rohrer, and A. Fried (2005), Vertical profiles of HDO/H_2O in the troposphere, *J. Geophys. Res.*, 110, D13301, doi:10.1029/2004JD005569.

Emanuel, K. A. (1991), A scheme for representing cumulus convection in large-scale models, *J. Atmos. Sci.*, 48, 2313–2329.

Emanuel, K. A., and M. Zivkovic-Rothman (1999), Development and evaluation of a convection scheme for use in climate models, *J. Atmos. Sci.*, 56, 1766–1782.

Field, R. D., D. B. A. Jones, and D. P. Brown (2010), The effects of post-condensation exchange on the isotopic composition of water in the atmosphere, *J. Geophys. Res.*, 115, D24305, doi:10.1029/2010JD014334.

Folkins, I., and R. Martin (2005), The vertical structure of tropical convection and its impact on the budgets of water vapor and ozone, *J. Atmos. Sci.*, 62, 1560–1573.

Frankenberg, C., et al. (2009), Dynamic processes governing lower-tropospheric HDO/H_2O ratios as observed from space and ground, *Science*, 325, 1374–1377.

Galewsky, J., and J. V. Hurley (2010), An advection-condensation model for subtropical water vapor isotopic ratios, *J. Geophys. Res.*, 115, D16116, doi:10.1029/2009JD013651.

Galewsky, J., M. Strong, and Z. D. Sharp (2007), Measurements of water vapor D/H ratios from Mauna Kea, Hawaii, and implications for subtropical humidity dynamics, *Geophys. Res. Lett.*, 34, L22808, doi:10.1029/2007GL031330.

Gates, W. L. (1992), AMIP: The Atmospheric Model Intercomparison Project, *Bull. Am. Meteorol. Soc.*, 73, 1962–1970.

Hoffmann, G., M. Werner, and M. Heimann (1998), Water isotope module of the ECHAM atmospheric general circulation model: A study on timescales from days to several years, *J. Geophys. Res.*, 103, 16,871–16,896.

Hourdin, F., and A. Armengaud (1999), The use of finite-volume methods for atmospheric advection of trace species. Part I: Test of various formulations in a General Circulation Model, *Mon. Weather Rev.*, 127, 822–837.

Hourdin, F., et al. (2006), The LMDZ4 general circulation model: Climate performance and sensitivity to parametrized physics with emphasis on tropical convection, *Clim. Dyn.*, 27, 787–813.

John, V. O., and B. J. Soden (2007), Temperature and humidity biases in global climate models and their impact on climate feedbacks, *Geophys. Res. Lett.*, 34, L18704, doi:10.1029/2007GL030429.

Johnson, L. R., Z. Sharp, J. Galewsky, M. Strong, P. Gupta, D. Baer, and D. Noone (2011), Hydrogen isotope measurements of water vapor and a correction for laser instrument measurement bias at low water vapor concentrations: Applications to measurements from Mauna Loa Observatory, Hawaii, *Rapid Commun. Mass Spectrom.*, 25, 608–616.

Jones, A., G. Qin, K. Strong, K. A. Walker, C. McLinden, M. Toohey, T. Kerzenmacher, P. Bernath, and C. Boone (2011), A global inventory of stratospheric NO_y from ACE-FTS, *J. Geophys. Res.*, 116, D17304, doi:10.1029/2010JD015465.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, 77, 437–470.

Kuang, Z., G. Toon, P. Wennberg, and Y. L. Yung (2003), Measured HDO/H_2O ratios across the tropical tropopause, *Geophys. Res. Lett.*, 30(7), 1372, doi:10.1029/2003GL017023.

Kurita, K., D. Noone, C. Risi, G. A. Schmidt, H. Yamada, and K. Yoneyama (2011), Intraseasonal isotopic variation associated with the Madden-Julian Oscillation, *J. Geophys. Res.*, 116, D24101, doi:10.1029/2010JD015209.

Kurylo, M., and R. Zander (2000), The NDSC—Its status after 10 years of operation, paper presented at XIX Quadrennial Ozone Symposium, Hokkaido Univ., Sapporo, Japan.

Landais, A., C. Risi, S. Bony, F. Vimeux, L. Descroix, S. Falourd, and A. Bouygues (2010), Combined measurements of $^{17}O_{excess}$ and d-excess in African monsoon precipitation: Implications for evaluating convective parameterizations, *Earth Planet. Sci. Lett.*, 298(1–2), 104–112.

Lee, J., J. Worden, D. Noone, K. Bowman, A. Elering, A. LeGrande, J.-L. Li, G. Schmidt, and H. Sodemann (2011), Relating tropical ocean clouds to moist processes using water vapor isotope measurements, *Atmos. Chem. Phys.*, 11, 741–752.

Lee, J.-E., I. Fung, D. DePaolo, and C. C. Fennig (2007), Analysis of the global distribution of water isotopes using the NCAR atmospheric general circulation model, *J. Geophys. Res.*, 112, D16306, doi:10.1029/2006JD007657.

Lee, J.-E., R. Pierrehumbert, A. Swann, and B. R. Lintner (2009), Sensitivity of stable water isotopic values to convective parameterization schemes, *Geophys. Res. Lett.*, 36, L23801, doi:10.1029/2009GL040880.

Lossow, S., et al. (2011), Comparison of HDO measurements from Envisat/MIPAS with observations by Odin/SMR and SCISAT/ACE-FTS, *Atmos. Meas. Tech.*, 4, 1855–1874.

Luz, B., and E. Barkan (2010), Variations of $^{17}O/^{16}O$ and $^{18}O/^{16}O$ in meteoric waters, *Geochim. Cosmochim. Acta*, 74(22), 6276–6286.

Marti, O., et al. (2005), The new IPSL climate system model: IPSL-CM4, technical report, IPSL, Paris.

Meehl, G. A., K. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. Taylor (2007), The WCRP CMIP3 multi-model dataset: A new era in climate change research, *Bull. Am. Meteorol. Soc.*, 7, 1383–1394.

Messerschmidt, J., R. Macatangay, J. Notholt, C. Petri, T. Warneke, and C. Weinzierl (2010), Side by side measurements of co2 by ground-based Fourier transform spectrometry (FTS), *Tellus, Ser. B*, 62, 749–758, doi:10.1111/j.1600–0889.2010.00491.x.

Moyer, E. J., F. W. Irion, Y. L. Yung, and M. R. Gunson (1996), ATMOS stratospheric deuterated water and implications for troposphere-stratosphere transport, *Geophys. Res. Lett.*, 23, 2385–2388.

Nassar, R., P. F. Bernath, C. D. Boone, A. Gettelman, S. D. McLeod, and C. P. Rinsland (2007), Variability in HDO/H_2O abundance ratio in the tropical tropopause layer, *J. Geophys. Res.*, 112, D21305, doi:10.1029/2007JD008417.

Noone, D. (2008), The influence of midlatitude and tropical overturning circulation on the isotopic composition of atmospheric water vapor and Antarctic precipitation, *J. Geophys. Res.*, 113, D04102, doi:10.1029/2007JD008892.

Noone, D., et al. (2011), Properties of air mass mixing and humidity in the subtropics from measurements of the D/H isotope ratio of water vapor at

the Mauna Loa Observatory, *J. Geophys. Res.*, *116*, D22113, doi:10.1029/2011JD015773.

Payne, V. H., D. Noone, A. Dudhia, C. Piccolo, and R. G. Grainger (2007), Global satellite measurements of HDO and implications for understanding the transport of water vapour into the stratosphere, *Q. J. R. Meteorol. Soc.*, *133*, 1459–1471.

Pierce, D. W., T. P. Barnett, E. J. Fetzer, and P. J. Gleckler (2006), Three-dimensional tropospheric water vapor in coupled climate models compared with observations from the AIRS satellite system, *Geophys. Res. Lett.*, *33*, L21701, doi:10.1029/2006GL027060.

Randall, D., et al. (2007), Climate models and their evaluation, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 591–662, Cambridge Univ. Press, Cambridge, U. K.

Risi, C., S. Bony, F. Vimeux, M. Chong, and L. Descroix (2010a), Evolution of the water stable isotopic composition of the rain sampled along Sahelian squall lines, *Q. J. R. Meteorol. Soc.*, *136*(S1), 227–242.

Risi, C., S. Bony, F. Vimeux, C. Frankenberg, and D. Noone (2010b), Understanding the Sahelian water budget through the isotopic composition of water vapor and precipitation, *J. Geophys. Res.*, *115*, D24110, doi:10.1029/2010JD014690.

Risi, C., S. Bony, F. Vimeux, and J. Jouzel (2010c), Water stable isotopes in the LMDZ4 General Circulation Model: Model evaluation for present day and past climates and applications to climatic interpretation of tropical isotopic records, *J. Geophys. Res.*, *115*, D12118, doi:10.1029/2009JD013255.

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

Roca, R., L. Picon, M. Desbois, H. Le Treut, and J.-J. Morcrette (1997), Direct comparison of Meteosat water vapor channel data and general circulation model results, *Geophys. Res. Lett.*, *24*(2), 147–150.

Rodgers, C. D., and B. J. Connor (2003), Intercomparison of remote sounding instruments, *J. Geophys. Res.*, *108*(D3), 4116, doi:10.1029/2002JD002299.

Sayres, D. S., et al. (2009), A new cavity based absorption instrument for detection of water isotopologues in the upper troposphere and lower stratosphere, *Rev. Sci. Instrum.*, *80*, 044102, doi:10.1063/1.3117349.

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

Schmidt, G., A. LeGrande, and G. Hoffmann (2007), Water isotope expressions of intrinsic and forced variability in a coupled ocean-atmosphere model, *J. Geophys. Res.*, *112*, D10103, doi:10.1029/2006JD007781.

Schneider, M., F. Hase, and T. Blumenstock (2006), Ground-based remote sensing of HDO/H_2O ratio profiles: Introduction and validation of an innovative retrieval approach, *Atmos. Chem. Phys.*, *6*, 4705–4722.

Schneider, M., G. Toon, J.-F. Blavier, F. Hase, and T. Leblanc (2010a), H_2O and δD profiles remotely-sensed from ground in different spectral infrared regions, *Atmos. Meas. Tech.*, *3*, 1599–1613.

Schneider, M., K. Yoshimura, F. Hase, and T. Blumenstock (2010b), The ground-based FTIR network's potential for investigating the atmospheric water cycle, *Atmos. Chem. Phys.*, *10*, 3427–3442.

Schneider, T., K. L. Smith, P. A. O'Gorman, and C. C. Walker (2006), A climatology of tropospheric zonal-mean water vapor fields and fluxes in isentropic coordinates, *J. Clim.*, *19*(22), 5918–5933.

Sherwood, S. C. (1996), Maintenance of the free tropospheric tropical water vapor distribution. Part II: Simulation of large-scale advection, *J. Clim.*, *11*, 2919–2934.

Soden, B. J., and C. Bretherton (1994), Evaluation of water vapor distribution in general circulation models using satellite observations, *J. Geophys. Res.*, *99*(D1), 1187–1210, doi:10.1029/93JD02912.

Soden, B. J., and I. M. Held (2006), An assessment of climate feedbacks in coupled ocean-atmosphere models, *J. Clim.*, *19*, 3354–3360, doi:10.1175/JCLI3799.1.

St. Clair, J. M., T. F. Hanisco, E. M. Weinstock, E. Moyer, D. Sayres, J. B. Smith, R. Lockwood, and J. G. Anderson (2008), A new photolysis laser induced fluorescence technique for the detection of HDO and H_2O in the lower stratosphere, *Rev. Sci. Instrum.*, *79*, 06401, doi:10.1063/1.2940221.

Steinwagner, J., M. Milz, T. von Clarmann, N. Glatthor, U. Grabowski, M. Höpfner, G. P. Stiller, and T. Röckmann (2007), HDO measurements with MIPAS, *Atmos. Chem. Phys.*, *7*, 2601–2615.

Steinwagner, J., S. Fueglistaler, G. Stiller, T. von Clarmann, M. Kiefer, P.-P. Borsboom, A. van Delden, and T. Röckmann (2010), Tropical dehydration processes constrained by the seasonality of stratospheric deuterated water, *Nat. Geosci.*, *3*, 262–266, doi:10.1038/NGEO822.

Tindall, J. C., P. Valdes, and L. C. Sime (2009), Stable water isotopes in HadCM3: Isotopic signature of El Niño–Southern Oscillation and the tropical amount effect, *J. Geophys. Res.*, *114*, D04111, doi:10.1029/2008JD010825.

Uemura, R., Y. Matsui, K. Yoshimura, H. Motoyama, and N. Yoshida (2008), Evidence of deuterium-excess in water vapour as an indicator of ocean surface conditions, *J. Geophys. Res.*, *113*, D19114, doi:10.1029/2008JD010209.

Uppala, S., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, *131*, 2961–3012.

Van Leer, B. (1977), Towards the ultimate conservative difference scheme: IV. A new approach to numerical convection, *J. Comput. Phys.*, *23*, 276–299.

Washenfelder, R., G. Toon, J.-F. Blavier, Z. Yang, N. Allen, P. Wennberg, S. Vay, D. Matross, and B. Daube (2006), Carbon dioxide column abundances at the Wisconsin Tall Tower site, *J. Geophys. Res.*, *111*, D22305, doi:10.1029/2006JD007154.

Webster, C. R., and A. J. Heymsfield (2003), Water isotope ratios D/H, $^{18}O/^{16}O$, $O/^{16}O$ in and out of clouds map dehydration pathways, *Science*, *302*, 1742–1746.

Werner, M., M. Heimann, and G. Hoffmann (2001), Isotopic composition and origin of polar precipitation in present and glacial climate simulations, *Tellus, Ser. B*, *53*, 53–71.

Winker, D., W. Hunt, and M. McGill (2007), Initial performance assessment of CALIOP, *Geophys. Res. Lett.*, *34*, L19803, doi:10.1029/2007GL030135.

Worden, J., et al. (2006), Tropospheric Emission Spectrometer observations of the tropospheric HDO/H_2O ratio: Estimation approach and characterization, *J. Geophys. Res.*, *111*, D16309, doi:10.1029/2005JD006606.

Worden, J., D. Noone, and K. Bowman (2007), Importance of rain evaporation and continental convection in the tropical water cycle, *Nature*, *445*, 528–532.

Worden, J., D. Noone, J. Galewsky, A. Bailey, K. Bowman, D. Brown, J. Hurley, S. Kulawik, J. Lee, and M. Strong (2010), Estimate of bias in Aura TES HDO/H_2O profiles from comparison of TES and in situ HDO/H_2O measurements at the Mauna Loa Observatory, *Atmos. Chem. Phys. Discuss.*, *10*, 25,355–25,388, doi:10.5194/acpd-10-25355-2010.

Wright, J. S., A. H. Sobel, and G. A. Schmidt (2009), The influence of condensate evaporation on water vapor and its stable isotopes in a GCM, *Geophys. Res. Lett.*, *36*, L12804, doi:10.1029/2009GL038091.

Wunch, D., et al. (2010), Calibration of the total carbon column observing network using aircraft profile data, *Atmos. Meas. Tech.*, *3*, 1351–1362, doi:10.5194/amt-3-1351-2010.

Wunch, D., G. C. Toon, J.-F. L. Blavier, R. A. Washenfelder, J. Notholt, B. J. Connor, D. W. T. Griffith, V. Sherlock, and P. O. Wennberg (2011), The Total Carbon Column Observing Network, *Philos. Trans. R. Soc. A*, *369*(1943), 2087–2112, doi:10.1098/rsta.2010.0240.

Yoshimura, K., M. Kanamitsu, D. Noone, and T. Oki (2008), Historical isotope simulation using reanalysis atmospheric data, *J. Geophys. Res.*, *113*, D19108, doi:10.1029/2008JD010074.

Zhang, C., B. Mapes, and B. J. Soden (2003), Bimodality in tropical water vapour, *Q. J. R. Meteorol. Soc.*, *129*(594), 2847–2866.

S. Barthlott, F. Hase, M. Kiefer, M. Schneider, and G. Stiller, IMK-ASF, Institute for Meteorology and Climate Research, Karlsruhe Institute of Technology, PO Box 3640, D-76021 Karlsruhe, Germany.

P. Bernath, Department of Chemistry, University of York, York YO10 5DD, UK.

S. Bony, LMD/IPSL, CNRS, Case Postale 99, 4 Place Jussieu, F-75252 Paris, France.

D. Brown, D. Noone, and C. Risi, Department of Atmospheric and Oceanic Sciences, University of Colorado at Boulder, Bldg. 318, 216 UCB, Main Campus, Boulder, CO 80309, USA. (camille.risi@colorado.edu)

N. Deutscher, J. Notholt, and T. Warneke, Institute of Environmental Physics, University of Bremen, Otto-Hahn-Allee 1, D-28334 Bremen, Germany.

C. Frankenberg, G. Toon, and J. Worden, Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA.

B. Funke, Instituto de Astrofísica de Andalucía, Apdo. 3004, E-18008 Granada, Spain.

O. García, Izaña Atmospheric Research Centre, Agencia Estatal de Meteorología, E-38001 Santa Cruz de Tenerife, Spain.

D. Griffith, Centre for Atmospheric Chemistry, University of Wollongong, Wollongong, NSW 2522, Australia.

J. Lee, Korea Polar Research Institute, Incheon 406-840, South Korea.

E. Mahieu, Institute of Astrophysics and Geophysics, University of Liège, B-4000 Liège, Belgium.

D. Sayres, School of Engineering and Applied Sciences, Harvard University, 12 Oxford St., Link Building, Cambridge, MA 02138, USA.

V. Sherlock and D. Smale, National Institute of Water and Atmospheric Research, PB 50061 Omakau, Lauder, Central Otago, New Zealand.

K. Strong and K. Walker, Department of Physics, University of Toronto, 60 St. George St., Toronto, ON M5S 1A7, Canada.

C. Sturm, Department of Geology and Geochemistry, Stockholm University, SE-10961 Stockholm, Sweden.

R. Uemura, Faculty of Science, University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa 903–0213, Japan.

P. O. Wennberg and D. Wunch, Division of Geological and Planetary Sciences and Engineering and Applied Science, California Institute of Technology, MS 150-21, 1200 E. California Blvd., Pasadena, CA 91125, USA.