Trend differences in lower stratospheric water vapour between Boulder and the zonal mean and their role in understanding fundamental observational discrepancies

Stefan Lossow[1], Dale F. Hurst[2], Karen H. Rosenlof[2], Gabriele P. Stiller[1], Thomas von Clarmann[1], Sabine Brinkop[3], Martin Dameris[3], Patrick Jöckel[3], Doug E. Kinnison[4], Johannes Plieninger[1], David A. Plummer[5], Felix Ploeger[6], William G. Read[7], Ellis E. Remsberg[8], James M. Russell[9], and Mengchu Tao[6]

[1]Karlsruhe Institute of Technology, Institute for Meteorology and Climate Research, Hermann-von-Helmholtz-Platz 1, 76344 Leopoldshafen, Germany.
[2]NOAA Earth System Research Laboratory, Global Monitoring Division, 325 Broadway, Boulder, CO 80305, USA.
[3]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, 82234 Oberpfaffenhofen-Wessling, Germany.
[4]University of Colorado, Atmospheric Chemistry Observations & Modeling Laboratory, P.O. Box 3000, Boulder, CO 80305-3000, USA.
[5]Environment and Climate Change Canada, Climate Research Branch, 550 Sherbrooke ouest, Montréal, Quebec H3A 1B9, Canada.
[6]Forschungszentrum Jülich, Institute for Energy and Climate Research: Stratosphere (IEK–7), Leo-Brandt-Straße, 52425 Jülich, Germany.
[7]Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA.
[8]NASA Langley Research Center, 21 Langley Boulevard, Hampton, VA 23681, USA.
[9]Hampton University, Center for Atmospheric Sciences, 23 Tyler Street Hampton, VA 23668, USA.

Correspondence: Stefan Lossow (stefan.lossow@kit.edu)

Version: Sunday 13th May, 2018 20:01 CET

Abstract. Trend estimates with different signs are reported in the literature for lower stratospheric water vapour considering the time period between the late 1980s and 2010. The NOAA (National Oceanic and Atmospheric Administration) frost point hygrometer (FPH) observations at Boulder (Colorado, 40.0°N, 105.2°W) indicate positive trends (about $0.1\,\mathrm{ppmv \cdot decade^{-1}}$ to $0.45\,\mathrm{ppmv \cdot decade^{-1}}$). Contrary, negative trends (approximately $-0.2\,\mathrm{ppmv \cdot decade^{-1}}$ to $-0.1\,\mathrm{ppmv \cdot decade^{-1}}$) are derived from a merged zonal mean satellite data set for a latitude band around the Boulder latitude. Overall, the trend differences between the two data sets range from about $0.3\,\mathrm{ppmv \cdot decade^{-1}}$ to $0.5\,\mathrm{ppmv \cdot decade^{-1}}$, depending on altitude. It has been proposed that a possible explanation for these discrepancies is a different temporal behaviour at Boulder and the zonal mean. In this work we investigate trend differences between Boulder and the zonal mean using primarily simulations from ECHAM/MESSy (European Centre for Medium-Range Weather Forecasts Hamburg/Modular Earth Submodel System) Atmospheric Chemistry (EMAC), WACCM (Whole Atmosphere Community Climate Model), CMAM (Canadian Middle Atmosphere Model) and CLaMS (Chemical Lagrangian Model of the Stratosphere). On shorter time scales we address this aspect also based on satellite observations from UARS/HALOE (Upper Atmosphere Research Satellite/Halogen Occultation Experi-

ment), Envisat/MIPAS (Environmental Satellite/Michelson Interferometer for Passive Atmospheric Sounding) and Aura/MLS (Microwave Limb Sounder). Overall, both the simulations and observations exhibit trend differences between Boulder and the zonal mean. The differences are dependent on altitude and the time period considered. The model simulations indicate only small trend differences between Boulder and the zonal mean for the time period between the late 1980s and 2010. These are clearly not sufficient to explain the discrepancies between the trend estimates derived from the FPH observations and the merged zonal mean satellite data set. Unless the simulations underrepresent variability or the trend differences originate from smaller spatial and temporal scales than resolved by the model simulations, trends at Boulder for this time period should be quite representative also for the zonal mean and even other latitude bands. Trend differences for a decade of data are larger and need to be kept in mind when comparing results for Boulder and the zonal mean on this time scale. Beyond that, we find that the trend estimates for the time period between the late 1980s and 2010 also significantly differ among the simulations. They are larger than those derived from the merged satellite data set and smaller than the trend estimates derived from the FPH observations.

1 Introduction

Water vapour in the stratosphere plays a fundamental role in the radiative budget and affects the ozone chemistry in this atmospheric layer. In the lower stratosphere water vapour is the most important greenhouse gas. As such, it is part of an important global warming feedback mechanism. A warmer climate increases lower stratospheric water vapour, leading to an even warmer climate. Dessler et al. (2013) estimated this feedback to be $0.3\,\mathrm{W}\cdot\mathrm{m}^{-2}$ for a temperature anomaly of 1 K at $500\,\mathrm{hPa}$. Besides that, water vapour is a fundamental component of polar stratospheric clouds. The heterogeneous chemistry on cloud particle surfaces is responsible for the severe ozone depletion in the lower stratosphere during winter and spring, especially in the Antarctic (Solomon, 1999). Water vapour is also the main source of hydrogen radicals (HO_x = OH, H, HO_2) in the stratosphere that contribute to ozone destruction through catalytic loss cycles (Brasseur and Solomon, 2005).

Thus, any change of stratospheric water vapour over a longer time scale has important implications (e.g. Dvortsov and Solomon, 2001; Forster and Shine, 2002; Stenke and Grewe, 2005; Solomon et al., 2010; Riese et al., 2012; Maycock et al., 2014; Gilford et al., 2016). In the past, the majority of studies related to longer term water vapour changes were based on observations by the balloon-borne NOAA frost point hygrometer at Boulder (a more detailed description of the measurement principle is provided in Sect. 2.2.4). These observations have been performed since 1980, typically once per month, providing the longest time series of water vapour in the lower stratosphere. Positive trends over Boulder were first reported by Oltmans and Hofmann (1995), then by Oltmans et al. (2000), Scherer et al. (2008) and finally Hurst et al. (2011). For the time period from 1980 to 2010, Hurst et al. (2011) showed an overall increase of $0.24\,\mathrm{ppmv}\cdot\mathrm{decade}^{-1}$ to $0.42\,\mathrm{ppmv}\cdot\mathrm{decade}^{-1}$ for the altitude range between $16\,\mathrm{km}$ and $26\,\mathrm{km}$ accompanied by significant variability on shorter time scales. 25% of the observed increase could be associated to changes of methane (Hurst et al., 2011). The oxidation of this trace gas is the most important in situ source of water vapour in the stratosphere. The other relevant source of water vapour in the stratosphere is transport from the troposphere, which mainly occurs through the cold tropical tropopause region. One major pathway is slow ascent (accompanied

by large horizontal motions, Holton and Gettelman, 2001) where the amount of water vapour entering the stratosphere is mainly controlled by the tropopause temperature (or better cold point temperature, Fueglistaler et al., 2009). Different changes of this temperature have been reported. Rosenlof and Reid (2008) reported an overall negative trend for the time period from 1980 to 2003, which would correspondingly result in a decrease of lower stratospheric water vapour. Recent work by Randel et al. (2017) indicates zero or slightly positive trends at the tropical tropopause for the time periods 1979 to 1997 and 1998 to 2014. The other pathway thought to be of importance is the convective lofting of ice particles (Moyer et al., 1996; Dessler et al., 2016; Avery et al., 2017). Once the particles reach the stratosphere, they evaporate and enhance the amount of stratospheric water vapour. This process is not dependent on the (cold point) temperature. Balloon-borne observations indicated no trend of the convective ice lofting into the stratosphere for the time period between 1991 and 2007 (Notholt et al., 2010). Based on all these results it is difficult to assess what process(es) caused the 30 year net increase of lower stratospheric water vapour observed by the FPH observations at Boulder (Hurst et al., 2011).

Satellite observations of stratospheric water vapour exist since 1978 (Gille and Russell, 1984), with some gaps. The instruments have limited life times and thus individual data sets do not allow a trend analysis on the same time scale as the FPH observations at Boulder. Recently, Hegglin et al. (2014) merged zonal mean data sets from seven satellite instruments. This merging was achieved with help of a CMAM simulation with specified dynamics (aka nudging) which acted as a transfer function. For each data set biases relative to the CMAM simulation were estimated. This assumes that the CMAM simulation provides a realistic representation of the water vapour variability (including trends) and that the satellite data sets do not have a drift in the bias estimation period. With this bias information the individual data sets were then adjusted relative to the Aura/MLS observations. Finally, the average over all bias-corrected data sets was used for the merged data set. This data set covers the time period between 1986 or 1988 (depending on latitude and altitude) and 2010, providing the opportunity to evaluate the trends observed by the FPH observations at Boulder and to address water vapour changes on a more global scale. The trends derived from the merged satellite data set for the zonal mean of the latitude around Boulder were negative below about $10\,\mathrm{hPa}$ and positive above. This behaviour could be essentially also observed at all other latitudes. Below $20\,\mathrm{hPa}$ the percentage changes up to 2010 were typically between -10% to -5%, which roughly corresponds to a trend between $-0.2\,\mathrm{ppmv}\cdot\mathrm{decade}^{-1}$ and $-0.1\,\mathrm{ppmv}\cdot\mathrm{decade}^{-1}$. Hegglin et al. (2014) attributed this trend to a reduced transport of water vapour into the stratosphere as a consequence of lower tropopause temperatures and a changed circulation in the stratosphere. During the same period as covered by the merged satellite data set, the FPH observations at Boulder still exhibit a clear increase of lower stratospheric water vapour (Hurst et al., 2011).

Figure 1 provides a summary of the trend discrepancies between the FPH observations and the merged satellite data. The trends derived from the merged satellite data set for the latitude band around Boulder (35°N – 45°N) are shown in green. Below (above) the dashed line the satellite trends are representative for the time period from 1988 to 2010 (1986 to 2010). The estimates are based on a digitisation of Fig. 5a in Hegglin et al. (2014). The extracted percentage trends were converted to volume mixing ratio trends using an average profile derived from all Aura/MLS observations in the latitude between 35°N and 45°N from August 2004 to December 2010. These data are chosen as all other satellite data sets are finally adjusted to the MLS data in the merging of Hegglin et al. (2014), as described above. Accordingly, the trends presented in Fig. 1 are approximations.

About the uncertainty of these trends we only know that they are at least statistically significant at the 2σ uncertainty level. Since the actual uncertainty level is unknown to us, we conservatively assume that the uncertainty is exactly at the 2σ level (which certainly overestimates the trend uncertainties). In red (blue) the trends derived from the FPH observations at Boulder are given for the time period from 1986 to 2010 (1988 to 2010). These were obtained by means of multi-linear regressions (see Eq. 2 later). Only small differences are observed between two time periods. The trend estimates do not change significantly if the vertical resolution of the FPH data is adjusted to that of the satellite observations. Likewise smoothing the FPH observations in time (with a 1 year running average), to reduce the scatter among individual observations, does not notably affect the trend estimates.

The lower panel of Fig. 1 shows an estimate of the differences in the trend estimates between the FPH observations and the merged zonal mean satellite data set. The differences vary with altitude ranging from about $0.3\,\mathrm{ppmv}\cdot\mathrm{decade}^{-1}$ to $0.5\,\mathrm{ppmv}\cdot\mathrm{decade}^{-1}$. Given the importance of water vapour in the lower stratosphere there is a dire need to reconcile these differences. Potential explanations could be the following or a combination of these:

(1) There might be problems with one of the data sets or even with both.

(2) The location of Boulder might be not representative for the zonal mean due to, e.g. local processes specific for the location (American monsoon, lee side of the Rocky Mountains, etc.)

(3) There might be unresolved differences between the measurement techniques, like due to the different spatial and temporal sampling and resolution.

In their discussion of the trend discrepancies between the FPH observations and the merged satellite data set Hegglin et al. (2014) opted for the second possible explanation, indicating that the temporal behaviour at Boulder is different than for the zonal mean of the latitude band around the Boulder latitude. Trends derived from the CMAM simulation at $100\,\mathrm{hPa}$ (considering the time period 1980 to 2010) indicated longitudinal differences at 40°N, but also at other latitudes. Subsampling the simulation to Boulder yielded better correlations with the FPH observations, in particular with respect to interannual variations. Yet, the trends derived from the FPH observations and the model simulation still disagreed, even in sign.

In this study we compare trend estimates for the Boulder location and the zonal mean for the latitude band around the Boulder latitude considering multiple time periods. For that we use several model simulations and observational data sets. The observations are meant to study this aspect on a decadal scale while the simulations will be used to analyse even longer time periods. This aims to understand how large the trend differences are general and how much they might contribute to the trend discrepancies shown in the lower panel of Fig. 1. In the next section the model simulations and observational data sets are briefly described while Sect. 3 outlines the analysis approach. The results of our analysis are presented in Sect. 4 and subsequently discussed in Sect. 5.

2 Data sets

In our analysis we primarily utilise model simulations. We consider results from EMAC, WACCM, CMAM and CLaMS. On the observational side we consider data from UARS/HALOE, Envisat/MIPAS and Aura/MLS. These data sets are analysed individually to avoid potential uncertainties and artefacts due to merging (e.g. Ball et al., 2017), providing results for the time periods 1992 – 2005, 2002 – 2012 and 2004 – 2016, respectively.

2.1 Model simulations

2.1.1 EMAC

The EMAC model is a numerical chemistry and climate simulation system that includes sub-models describing tropospheric and middle atmosphere processes and their interaction with oceans, land and human influences (Jöckel et al., 2010). It uses the second version of the Modular Earth Submodel System (MESSy2) to link multi-institutional computer codes. The core atmospheric model is the 5^{th} generation European Centre Hamburg general circulation model (ECHAM5, Roeckner et al., 2006). For the present study we applied EMAC (ECHAM5 version 5.3.02, MESSy version 2.50.5) in the T42L90MA-resolution, i.e. with a spherical truncation of T42 (corresponding to a quadratic Gaussian grid of approximately 2.8° by 2.8° in latitude and longitude) with 90 vertical hybrid pressure levels up to 0.01 hPa. The simulation was set up in accordance to the REF-C1SD (transient hindcast reference simulation with specified dynamics) scenario defined in the framework of the SPARC (Stratosphere-troposphere Processes And their Role in Climate) Chemistry-Climate Model Initiative (Eyring et al., 2013). Correspondingly, it considers nudging (by Newtonian relaxation) towards data from the Interim ECMWF (European Centre for Medium-Range Weather Forecasts) Reanalysis project (ERA-interim, Dee et al., 2011). Nudged parameters were the vorticity, divergence, the logarithm of the surface pressure, the temperature and the mean temperature (wave number zero in spectral space, Jöckel et al., 2016). Correspondingly, water vapour itself was not nudged and allowed to evolve freely. Depending on parameter the nudging time constant varied between 6 h and 48 h. The initial conditions (in 1979) were taken from a corresponding free-running simulation. In our analysis we use 10 hourly data, lasting until 2013.

2.1.2 WACCM

WACCM is an atmospheric component of the Community Earth System Model (CESM, Hurrell et al., 2013), a global climate model with interactive ocean, sea ice, land and atmosphere. WACCM itself extends from the Earth's surface into the thermosphere up to $5.1{\cdot}10^{-6}\,\mathrm{hPa}$ (about 140 km). The simulation used 88 vertical levels and its horizontal resolution amounts to 1.9° in latitude and 2.5° in longitude (Marsh et al., 2013). As EMAC, the WACCM simulation employed here was set up according to the REF-C1SD scenario. Meteorological fields from the MERRA (Modern Era Retrospective-Analysis for Research and Applications, Rienecker et al., 2011) reanalysis data set were nudged from the surface to 50 km. Above 60 km the model meteorological fields were fully interactive, with a linear transition in between. Here, temperature, zonal and meridional winds and surface pressure were used to drive the physical parameterisation that control boundary layer exchanges, advective and

convective transport and the hydrological cycle. The nudging time constant used in this study was 50 h. The initial conditions for the year 1979 were taken from a time-dependent REF-C1 simulation that started in 1955. Here we consider daily averaged data and 2014 is the last year of the simulation.

2.1.3 CMAM

The Canadian Middle Atmosphere Model is a well-established and comprehensive chemistry climate model (de Grandpré et al., 2000; Scinocca et al., 2008). The CMAM simulation we employ is the same that has been used for the merging of the satellite data sets (Hegglin et al., 2014). It covers the period from 1979 to 2010 and provides results from the Earth's surface up to 0.0007 hPa on 63 pressure levels. The horizontal resolution is 3.75° in latitude and longitude (T47). Horizontal winds and temperature data from ERA-interim were nudged up 1 hPa with a nudging time constant of 24 hours at all levels. The nudging was performed in spectral space and only spectral coefficients up to T21 were nudged (McLandress et al., 2013, 2014). For the initial conditions the same simulation setup was run up to 1979, but nudging ERA-40 reanalysis data (Uppala et al., 2005). In our analysis we employ 6 hourly data.

2.1.4 CLaMS

The CLaMS model is fundamentally different to the models presented so far, as it is a Lagrangian chemistry transport model (McKenna et al., 2002b, a). It is driven by horizontal winds, temperature and diabatic heating rates that are taken from a reanalysis data set. CLaMS uses a hybrid vertical coordinate system which considers isentropes above about 300 hPa. The calculation of water vapour volume mixing ratios is based on simplified dehydration scheme (Ploeger et al., 2013). Below about 500 hPa data from the driving reanalysis are used. Above, if saturation occurs along a trajectory the amount of water vapour in excess of the saturation ratio is frozen out and and partly sediments out, based on the fall speed of spherical ice particles of a mean size. Methane oxidation in the stratosphere is implemented using methane fields from the simulation and hydroxyl, oxygen and chlorine radicals from a model climatology. The simulation used in this work was driven by ERA-interim data. The results were interpolated on a regular pressure grid and use a horizontal resolution of 1° in latitude and longitude. We consider daily data (at 12 UTC) until 2010.

2.2 Observations

2.2.1 UARS/HALOE

HALOE was a solar occultation instrument deployed on UARS which was launched on 12 September 1991. Observations lasted until November 2005 shortly before the satellite was decommissioned. Based on the observation geometry 30 observations were performed per day. Those typically covered two distinct latitudes, one in the Northern and one in the Southern Hemisphere. Overall, latitudes between 80°S and 80°N were covered. HALOE measured in the infrared spectral region covering some specific bands between 2.5 μm and 11 μm. Water vapour information has been retrieved from a spectral band ranging from 6.54 μm to 6.67 μm, typically covering altitudes from the upper troposphere to the upper mesosphere. In this study we employ data

derived with retrieval version 19 (Kley et al., 2000). Occultations with anomalies regarding the trip angle (http://haloe.gats-inc.com/user_docs/events_terminate_below_150km.pdf) and the lockdown angle (http://haloe.gats-inc.com/user_docs/smoothed_lockdown_angles.pdf) were screened. Also, observations before March 1992 were discarded as they might be affected by aerosols from the Pinatubo volcanic eruption in June 1991.

2.2.2 Envisat/MIPAS

MIPAS was a high-resolution Fourier transform spectrometer flown on Envisat. The satellite was launched on 1 March 2002 and operated until 8 April 2012. The MIPAS instrument measured thermal emission in the infrared spectral region between 4.1 μm and 14.6 μm covering the entire latitude range (Fischer et al., 2008). Initially, the measurements used a spectral resolution of 0.025 cm^{-1} (unapodised). Due to an instrument failure in March 2004 the spectral resolution had to be reduced to 0.0625 cm^{-1}. Observations with the lower spectral resolution recommenced in January 2005. In accordance the MIPAS time period is split in two periods which are referred to as full (FR) and reduced (RR) resolution period. During the FR period more than 1000 scans were performed daily while during the RR period it were more than 1300 scans. Water vapour information is retrieved from 12 microwindows between 6.3 µm and 12.6 µm typically covering the upper troposphere to the middle mesosphere. Here we combine data from the retrieval version 20 for the FR period and version 220/221 for the RR period (Schieferdecker et al., 2015; Lossow et al., 2017), both generated with the research processor operated at IMK/IAA (Institut für Meteorologie und Klimaforschung (IMK) in Karlsruhe, Germany / Instituto de Astrofísica de Andalucía (IAA) in Granada, Spain). The overall time period ranges from July 2002 to April 2012. Before the analysis the data were screened considering the visibility flag and averaging kernel diagonal criterion (discard data with diagonal values < 0.03). The former flags data below the lowermost usable tangent altitude while the latter criterion concerns the measurement contribution to the retrieved data.

2.2.3 Aura/MLS

The Microwave Limb Sounder is an instrument aboard NASA's (National Aeronautics and Space Administration) Aura satellite. The satellite was launched on 15 July 2004 and uses a sun-synchronous orbit, as Envisat did. The MLS instrument measures microwave thermal emission at the limb of the Earth's atmosphere, covering the latitude range between 82°S and 82°N. An atmospheric scan takes about 25 s, resulting in more than 3400 observations per day (Waters et al., 2006). Water vapour information is derived from the strong emission line centred at 183 GHz, covering the altitude range from the upper troposphere to the upper mesosphere. In the analysis we used data from the latest retrieval version 4.2, considering the time period from August 2004 to December 2016. Prior any analysis the data were screened according to the criteria listed in the data quality document (Livesey et al., 2015).

2.2.4 NOAA frost point hygrometer

For the sake of completeness we provide here also a more detailed description of the NOAA frost point hygrometer. The FPH measurement principle is based on maintaining a thin, stable layer of frost on a chilled mirror as air flows past it at $5\,\mathrm{m}\cdot\mathrm{s}^{-1}$.

Stability in frost coverage is detected optically and maintained by rapidly adjusting the mirror temperature. When the frost coverage is stable, the ice and overlying water vapour are in equilibrium and the ice surface temperature (frost point temperature) is directly related to the partial pressure of water vapour in the air stream. At 50 hPa, a 0.5 ppmv (about 10%) change in the water vapour mixing ratio produces a 0.42 K change in the frost point temperature. The mirror temperature is measured by a thermistor calibrated to an accuracy better than 0.05 K. Hall et al. (2016) provide detailed descriptions of the instrument and its history, along with an assessment of its measurement uncertainties. The primary measurement uncertainty is related to instabilities in frost coverage that can produce frost point temperature errors as large as $\pm$0.5 K in the stratosphere. However, the instabilities are generally oscillatory in nature and therefore manifest as random errors, not systematic biases. Each thermistor is meticulously calibrated against a temperature probe certified by the National Institute of Standards and Technology (NIST) and, to ensure calibration stability over the long term (i.e. decades), a small archive of previously calibrated thermistors. Total FPH measurement uncertainties (95% confidence) in the stratosphere are estimated to be smaller than 0.3 ppmv (about 6%, Hall et al., 2016). The 30 year net increase ($\sim$1 ppmv, see Introduction) in stratospheric water vapour observed over Boulder translates to a 0.8 K rise in frost point temperatures that greatly exceeds the FPH measurement uncertainties.

3 Approach

3.1 Boulder time series

For the Boulder time series we consider simulated data and satellite observations that are spatially located within:

- a 1000 km radius around the Boulder FPH observation site.
- the latitude band between 35°N – 45°N.

In the analysis of the HALOE data set we use less strict criteria because of its sparseness relative to the other data sets. Instead of the radius criterion data in the wider longitude range between 130°W and 80°W are considered.

In temporal terms we consider for the Boulder time series two sets of data. Set #1 simply comprises all data in a given month. We will refer to these time series as full time series. Set #2 is adapted to the individual FPH observations at Boulder. From that we can also assess the role of the temporal sampling for the trend differences. For the simulations the data from the closest time step are used. For the observations all data obtained within $\pm$12 h of the FPH measurements are considered. These time series we will refer to as adapted time series.

All data obeying the spatial and temporal criteria are combined to monthly means. For the observations we consider only monthly means that are based on at least 5 measurements to avoid spurious results. As a result, a temporal adaption to the individual FPH observations is only meaningful for the MLS observations.

3.2 Zonal mean time series

For the zonal mean time series we consider monthly means of all data in the latitude range between 35°N to 45°N, resembling the merged satellite data set. Monthly zonal means derived from the satellite observations are discarded if they are not based on a minimum number of 20 measurements. If a monthly mean does not exist for the Boulder time series, e.g. because there were no FPH observations for the adapted time series or due to screening of the satellite data, this monthly mean is also not considered for the zonal mean results.

In addition, we also investigate how the trend estimates at Boulder compare to those for zonal means of other latitude bands. For that we consider the latitude bands 45°N – 55°N, 25°N – 35°N, Equator – 60°N and 60°S – 60°N. The first two bands are adjacent to the latitude band around the Boulder latitude. The last two bands cover a wider range of latitudes. This aims to investigate how representative trends at Boulder are on regional and more global scales.

3.3 De-seasonalisation

In our analysis we employ de-seasonalised data. This enhances the visibility of the long-term behaviour and has the positive side effect that the MIPAS observations from the FR and RR periods are homogenised. Between these periods typically a small bias in the absolute water vapour volume mixing ratios exists. The de-seasonalisation is achieved by means of regression, again motivated by the MIPAS data. This approach has the advantage of working for time series that cover a time period between 12 months and 24 months, which applies here to the MIPAS data for the FR period. The regression model contains an offset and a parametrisation for the semi-annual (SAO) and annual variation (AO) using orthogonal sine and cosine functions:

$$
\begin{aligned}
f_{\mathrm{d}}(t,\phi,z) = & C_{\mathrm{offset}}(\phi,z) + \\
& C_{\mathrm{SAO}_1}(\phi,z) \cdot \sin(2 \cdot \pi \cdot t / p_{\mathrm{SAO}}) + \\
& C_{\mathrm{SAO}_2}(\phi,z) \cdot \cos(2 \cdot \pi \cdot t / p_{\mathrm{SAO}}) + \\
& C_{\mathrm{AO}_1}(\phi,z) \cdot \sin(2 \cdot \pi \cdot t / p_{\mathrm{AO}}) + \\
& C_{\mathrm{AO}_2}(\phi,z) \cdot \cos(2 \cdot \pi \cdot t / p_{\mathrm{AO}}).
\end{aligned} \tag{1}
$$

In the equation $f_{\mathrm{d}}(t,\phi,z)$ denotes the fit of the regressed time series for a given time t (in years), latitude band ϕ and altitude z which is subsequently subtracted from the absolute time series to obtain the de-seasonalised time series. C are the regression coefficients of the individual model components and p_{SAO} and p_{AO} represent the time periods of the semi-annual (0.5 years) and annual variation (1 year), respectively. The regression coefficients are derived according to the method outlined by von Clarmann et al. (2010), using the standard errors of the monthly means (their inverse squared) as statistical weights. Autocorrelation effects and empirical errors (Stiller et al., 2012) are not considered in this regression.

For the de-seasonalisation of the simulations we consider data in the time period from 1985 to 2010. The start year is chosen because of obvious differences in the water vapour abundances among the simulations, related to differences in their initial conditions and spin up time (see Fig. 2 and Sect. 4.1). 2010 is the last year that is covered by all simulations. For the

observations it is not possible to use a consistent time period. Instead the entire time period covered by the individual data sets is used for the de-seasonalisation.

3.4 Trend estimates and trend differences

Like the de-seasonalisation of the time series, the estimation of the water vapour trends is based on regression. For this analysis the regression model is as follows:

$$\begin{aligned} f_{\mathrm{t}}(t,\phi,z) =& C_{\mathrm{offset}}(\phi,z) + C_{\mathrm{trend}}(\phi,z)\cdot t + \\ & C_{\mathrm{SAO_1}}(\phi,z)\cdot\sin(2\cdot\pi\cdot t/p_{\mathrm{SAO}}) + \\ & C_{\mathrm{SAO_2}}(\phi,z)\cdot\cos(2\cdot\pi\cdot t/p_{\mathrm{SAO}}) + \\ & C_{\mathrm{AO_1}}(\phi,z)\cdot\sin(2\cdot\pi\cdot t/p_{\mathrm{AO}}) + \\ & C_{\mathrm{AO_2}}(\phi,z)\cdot\cos(2\cdot\pi\cdot t/p_{\mathrm{AO}}) + \\ & C_{\mathrm{QBO_1}}(\phi,z)\cdot QBO_1(t) + \\ & C_{\mathrm{QBO_2}}(\phi,z)\cdot QBO_2(t). \end{aligned} \tag{2}$$

In comparison to the regression model used for the de-seasonalisation, it contains in addition a trend term C_{trend} and a parametrisation for the quasi-biennial oscillation (QBO). In our analysis we determine only a single trend for the entire time period. Trend changes within this period are correspondingly not analysed (see e.g. Hurst et al., 2011). Even though the regression is applied to de-seasonalised time series the SAO and AO terms are kept since the regression models for the de-seasonalisation and trend analysis differ. The QBO parametrisation is based on normalised winds at 50 hPa (QBO_1) and 30 hPa (QBO_2) observed over Singapore (1°N, 104°E), which are almost orthogonal. These data are provided by Freie Universität Berlin (webpage: http://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/qbo.dat). Unlike for the de-seasonalisation, in this regression we consider autocorrelation effects and empirical errors (Stiller et al., 2012), to obtain optimal estimates of the trends and their uncertainties.

To be consistent with our motivation shown in Fig. 1, we calculate the water vapour trends separately for the Boulder time series and the zonal mean time series and subsequently derive the trend differences. Correspondingly, the trend differences ($\Delta C_{\mathrm{trend}}$) and their uncertainties ($\varepsilon_{\mathrm{trend}}$) are given as:

$$\begin{aligned} \Delta C_{\mathrm{trend}}(\phi,z) &= C_{\mathrm{trend}}^{\mathrm{Boulder}}(z) - C_{\mathrm{trend}}^{\mathrm{zonal\ mean}}(\phi,z) \\ \varepsilon_{\mathrm{trend}}(\phi,z) &= \sqrt{\varepsilon_{\mathrm{trend}}^{\mathrm{Boulder}}(z)^2 + \varepsilon_{\mathrm{trend}}^{\mathrm{zonal\ mean}}(\phi,z)^2} \end{aligned} \tag{3}$$

Here $C_{\mathrm{trend}}^{\mathrm{Boulder}}$ represents the trends derived from the Boulder time series and $\varepsilon_{\mathrm{trend}}^{\mathrm{Boulder}}$ are the corresponding uncertainties. Likewise $C_{\mathrm{trend}}^{\mathrm{zonal\ mean}}$ and $\varepsilon_{\mathrm{trend}}^{\mathrm{zonal\ mean}}$denote the trends calculated from the zonal mean time series and their uncertainties.

4 Results

In this section we will first present the simulation results and subsequently the results derived from the observations. We focus on the altitude range between 100 hPa and 20 hPa that is typically covered by the FPH observations and in almost all cases completely in the stratosphere (Kunz et al., 2013).

4.1 Simulations

Figure 2 shows for the different model simulations the de-seasonalised Boulder time series (top panel) and the zonal mean time series around the Boulder latitude (latitude range 35°N – 45°N, middle panel) at 70 hPa. The differences between the two time series are shown in the lower panel as a complement. The time series adapted to the individual FPH observations (see Sect. 3.1 and 3.2) are marked with the suffix (A) in the figure legend. Overall, the Boulder and the zonal mean time series are visually rather similar, with the latter being more smooth. The difference time series show occasionally larger deviations (up to 0.6 ppmv in absolute terms), however any conspicuous behaviour or a trend appears to be absent. In general, the different simulations yield similar results for Boulder and the zonal mean. The most prominent exception is observed in the early 1980s. This relates to differences in the 1979 initial conditions and the spin up times among the simulations. Until 1985 the EMAC anomalies are significantly lower than for the other simulations. In the first years the largest anomalies are found in the CMAM results, which were probably caused by higher water vapour volume mixing ratios in the initial conditions based on the nudging of ERA-40 data (see Sect. 2.1.3). Presumably the best representation is provided by CLaMS which, as a Lagrangian model, does not need to deal with these aspects.

Figure 3 shows the trend estimates for the time series at Boulder (left column) and the zonal mean for the latitude band between 35°N and 45°N (middle column). The right column shows the corresponding difference according to Eq. 3. The different rows consider different time periods, i.e. 1985 – 2010 (top row), 1990 – 2010 (middle row) and 1995 – 2010 (bottom row). This is also indicated in the title of the centre panels. We have not included the time period from 1980 to 2010 here. The differences in the water vapour anomalies among the simulations in the early 1980s primarily affect the trends for Boulder and the zonal mean, yet the trend differences are comparable to those for the time period from 1985 to 2010. Trends and trend differences significant at the 2σ uncertainty levels are marked by triangles.

For the time period between 1985 and 2010, the EMAC results exhibit positive trend estimates at Boulder. They range between 0.04 $\mathrm{ppmv \cdot decade^{-1}}$ and 0.12 $\mathrm{ppmv \cdot decade^{-1}}$. The results derived from the other simulations indicate rather small trends at Boulder, typically within $\pm$0.05 $\mathrm{ppmv \cdot decade^{-1}}$. Small quantitative differences exist between the results derived from the full and the adapted time series. Those from the adapted time series are typically larger (up to about 0.04 $\mathrm{ppmv \cdot decade^{-1}}$). Overall, the spread among the trend estimates ranges from about 0.1 $\mathrm{ppmv \cdot decade^{-1}}$ at 100 hPa to 0.16 $\mathrm{ppmv \cdot decade^{-1}}$ at 20 hPa. The trend estimates derived from the zonal mean time series for the latitude band between 35°N and 45°N look very similar to those derived for Boulder. Correspondingly, the trend differences between Boulder and the zonal mean are very small. The differences never exceed 0.04 $\mathrm{ppmv \cdot decade^{-1}}$ in absolute terms. The largest differences are derived for EMAC and WACCM (based on the adapted time series) at 100 hPa. The trend differences are predominantly

positive below 70 hPa and mostly negative above 50 hPa. The exact altitude dependence differs in details among the different simulation results.

Both for Boulder and the zonal mean, the trend estimates for the time period from 1990 to 2010 are negative. There are differences among the individual simulations. The agreement is, however, better than for the time period between 1985 and 2010. The spread maximises at 100 hPa with about $0.12\,\mathrm{ppmv \cdot decade^{-1}}$ and is smallest above 40 hPa with about $0.06\,\mathrm{ppmv \cdot decade^{-1}}$. Differences among the results derived from the full and adapted time series are very small. The trend differences between Boulder and the zonal mean are of similar size as for the time period 1985 to 2010. A similar behaviour in terms of the altitude dependence is also visible.

The last time period we consider is from 1995 to 2010. At 100 hPa consistently positive trends are found, except for the adapted EMAC time series. Overall, the trend estimates vary between $-0.02\,\mathrm{ppmv \cdot decade^{-1}}$ and $0.14\,\mathrm{ppmv \cdot decade^{-1}}$. With increasing altitude the trend estimates typically decrease and above 45 hPa they are all negative. Higher up, the trends continue to get more negative, except in the CMAM results. At 20 hPa the trend estimates vary between $-0.24\,\mathrm{ppmv \cdot decade^{-1}}$ and $-0.08\,\mathrm{ppmv \cdot decade^{-1}}$ among the simulations, with significantly smaller differences between the results derived from the full and the adapted time series. The best agreement among the simulations is observed around 80 hPa where the spread is about $0.08\,\mathrm{ppmv \cdot decade^{-1}}$. The altitude dependence and the spread among the simulations is similar for the trend estimates derived from the zonal mean time series. Quantitatively there are larger differences between the Boulder and zonal mean trends, clearly surpassing those observed for the other time periods. Above 60 hPa the differences are still within $\pm 0.02\,\mathrm{ppmv \cdot decade^{-1}}$. Below this altitude the differences occasionally exceed $\pm 0.05\,\mathrm{ppmv \cdot decade^{-1}}$. The largest trend differences are derived from the adapted EMAC and CMAM time series.

To expand on the temporal development of the trend differences between Boulder and the zonal mean even more we derive these differences continuously for 11 year periods, as shown in Fig. 4. The results are assigned to the centre of the considered period, e.g. to 1995 for the time period between 1990 and 2000. The trend differences vary with time and altitude in size and sign. On this shorter time scale the differences are typically larger than observed for the longer time periods described in the last figure. There is also a more prominent distinction between the results derived from the full and the adapted time series. The latter yield larger differences on an absolute scale, but also some patterns are different.

For the full time series the trend differences are generally within $\pm 0.04\,\mathrm{ppmv \cdot decade^{-1}}$. Exceptions from this behaviour are primarily observed at the lowermost altitudes. In particular the EMAC results exhibit significantly larger differences, increasing to about $\pm 0.15\,\mathrm{ppmv \cdot decade^{-1}}$ at 100 hPa. The temporal development of the trend differences exhibits a number of common features among the simulations, even though quantitative differences are obvious. At the lowermost altitudes all simulations show negative trend differences from 1990 to about 1999. Afterwards positive trend differences are found. Higher up, i.e. above about 50 hPa, positive trend differences are visible from 1995 to about 2004.

The trend differences derived from the adapted time series are within $\pm 0.08\,\mathrm{ppmv \cdot decade^{-1}}$ above 60 hPa. Below, they increase again in absolute size, maximising at about $0.2\,\mathrm{ppmv \cdot decade^{-1}}$. The different simulations agree on some difference patterns, as observed for the results derived from the full time series. Most prominently, above 50 hPa the trend differences are typically positive from about 1990 to 2003 and negative afterwards. The bisection of trend differences at the lowermost

altitudes derived from the full time series is only visible in some simulations. Finally, it should be noted, that none of the trend differences shown in Fig. 4 are statistically significant at the 2σ uncertainty level.

To investigate the representativeness of the Boulder trends on a larger geographical scale Fig. 5 compares them to zonal mean trends for five latitude bands, namely 35°N – 45°N (row #1), 45°N – 55°N (row #2), 25°N – 35°N (row #3), Equator – 60°N (row #4) and 60°S – 60°N (row #5). The figure considers the time period between 1987 and 2010, approximately the time coverage of the merged zonal mean satellite data set. The results in the left column are the same for all rows and kept for the sake of convenience. The trends at Boulder are close to those obtained for the time periods 1985 – 2010 and 1990 – 2010 shown in Fig. 3. Note, that in Fig. 5 the x-axis is smaller, allowing a more detailed picture.

Overall, the trends at Boulder are within $\pm 0.07\,\mathrm{ppmv \cdot decade^{-1}}$. Clear differences among the simulations exist, while the differences between the results derived from the full and the adapted time series are typically smaller. The trend estimates derived from the full time series are again larger than those determined from the adapted time series, with few exceptions. The EMAC results indicate positive trends (up to almost $0.1\,\mathrm{ppmv \cdot decade^{-1}}$). Statistical significance is only visible at the highest altitudes. The trends derived from the CLaMS data are negative below $35\,\mathrm{hPa}$ and positive above, ranging from about $-0.05\,\mathrm{ppmv \cdot decade^{-1}}$ to $0.05\,\mathrm{ppmv \cdot decade^{-1}}$. They are relatively constant up to $60\,\mathrm{hPa}$ before they start to increase significantly. The WACCM and CMAM trends show a similar altitude dependence with maximum negative trends (around $-0.05\,\mathrm{ppmv \cdot decade^{-1}}$) in the altitude range between $50\,\mathrm{hPa}$ and $40\,\mathrm{hPa}$. For WACCM the trend estimates become positive below $80\,\mathrm{hPa}$ while those derived from the CMAM simulation are negative at all altitudes.

As observed in Fig. 3 the trends derived from the zonal mean time series for the latitude band between 35°N and 45°N are very similar to those for Boulder. Correspondingly, the trend differences are small, i.e. ranging from $-0.02\,\mathrm{ppmv \cdot decade^{-1}}$ to $0.04\,\mathrm{ppmv \cdot decade^{-1}}$. The differences are typically positive below $70\,\mathrm{hPa}$ and mostly negative above, affirming the picture observed for the time periods 1985 – 2010 and 1990 – 2010 in Fig. 3.

The trends derived from the zonal mean time series for the other latitude bands exhibit many common features with the results for the zonal mean between 35°N and 45°N. There are quantitative changes, but overall the trend estimates remain of the same order. Besides that, also the altitude dependence of the trends remains very similar and so do the relations among the different simulations. The trend differences between Boulder and the zonal mean for 45°N and 55°N remain within $\pm 0.04\,\mathrm{ppmv \cdot decade^{-1}}$. Again, the differences are typically positive below $70\,\mathrm{hPa}$ and predominantly negative at higher altitudes. The trend differences between Boulder and the zonal means for 25°N to 35°N and from the Equator to 60°N are quite similar, at least up to about $35\,\mathrm{hPa}$. In both cases the trend differences are within $\pm 0.03\,\mathrm{ppmv \cdot decade^{-1}}$. Typically the EMAC and CLaMS results are at the higher end of this interval while the WACCM and CMAM results are at the lower end. The largest trend differences compared to Boulder are observed for the zonal mean of the latitude band between 60°S and 60°N. These range from $-0.04\,\mathrm{ppmv \cdot decade^{-1}}$ to slightly more than $0.06\,\mathrm{ppmv \cdot decade^{-1}}$. There is clear separation between the CLaMS results and those from the other simulations. For the CLaMS simulation the trend differences are negative at $100\,\mathrm{hPa}$ (around $-0.015\,\mathrm{ppmv \cdot decade^{-1}}$). Around $90\,\mathrm{hPa}$ they turn positive and continue to increase within increasing altitude. At $20\,\mathrm{hPa}$ the differences amount to $0.05\,\mathrm{ppmv \cdot decade^{-1}}$ for the adapted time series and $0.06\,\mathrm{ppmv \cdot decade^{-1}}$ for the full time series, respectively. The other simulations indicate positive trend differences at $100\,\mathrm{hPa}$. Around $70\,\mathrm{hPa}$ the

differences become negative and peak in absolute size between 50 hPa and 40 hPa (between $-0.04\,\mathrm{ppmv\cdot decade^{-1}}$ and $-0.02\,\mathrm{ppmv\cdot decade^{-1}}$). Higher up, the trend differences get less negative again.

4.2 Observations

Figure 6 shows for the HALOE, MIPAS and MLS observations the de-seasonalised Boulder time series (top panel) and the zonal mean time series around the Boulder latitude at 70 hPa (middle panel). As in Fig. 2 the lower panel shows the differences between the two time series in addition. For MLS there is also a data set that is adapted to the FPH observations at Boulder (see Sect 3.1). Like the simulations the observations exhibit a rather similar picture for Boulder and the zonal mean. The difference time series indicate occasionally some larger deviations. For example in the second half of 2011 some substantial positive differences are observed, consistent in the MIPAS and MLS data. The largest differences typically occur in the MLS data set that is adapted to the FPH observations and for the HALOE data set, primarily due to its sparseness. In addition, there is a notable agreement between the MIPAS and MLS time series for Boulder and the zonal mean time series.

Figure 7 compares the trend estimates at Boulder with those derived from zonal mean time series for various latitude bands. The results for the different observational data sets consider different time periods as indicated in the figure legend. Thus, they are not comparable and will be addressed separately.

In the lower stratosphere the HALOE observations exhibit negative trends at Boulder for the time period between 1992 and 2005. This behaviour is primarily related to the significant drop in lower stratospheric water vapour in 2001 (Randel et al., 2006; Scherer et al., 2008; Brinkop et al., 2016). The relative dryness continued until 2005 (coinciding with the end of the HALOE observations), causing the 14 year HALOE trends to be negative. The largest trends are observed below 80 hPa with values around $-0.45\,\mathrm{ppmv\cdot decade^{-1}}$. Above, the trends get less negative with increasing altitude. At 20 hPa the trend amounts to about $-0.03\,\mathrm{ppmv\cdot decade^{-1}}$ (and is not statistically significant). The trends derived for the zonal mean between 35°N to 45°N have a similar altitude dependence, but their absolute sizes are smaller. Accordingly, the trend differences between Boulder and the zonal mean are negative. Above 80 hPa the differences are almost invariant with altitude. Here they amount to about $-0.05\,\mathrm{ppmv\cdot decade^{-1}}$. At lower altitudes the differences are larger maximising at 100 hPa with about $-0.1\,\mathrm{ppmv\cdot decade^{-1}}$. For the other latitude bands the zonal mean trends exhibit the same kind of altitude dependence as observed for the band from 35°N to 45°N. The most prominent variation concerns the exact altitude in which the negative trends exhibit their absolute maximum. For the latitude band between 45°N and 55°N this occurs close to 90 hPa. For the trends derived from the zonal mean from the Equator to 60°N and from 60°S to 60°N this maximum is observed around 70 hPa. The trend differences between Boulder and the zonal mean for the latitude band between 45°N and 55°N range from $-0.1\,\mathrm{ppmv\cdot decade^{-1}}$ to $0\,\mathrm{ppmv\cdot decade^{-1}}$ with the largest absolute values occurring below 75 hPa. For the latitude band between 25°N and 35°N the trend differences are close to zero, except below 75 hPa where they become significantly more negative. For the remaining two latitude bands the trend differences are quite similar. At 100 hPa the trend differences amount to $-0.15\,\mathrm{ppmv\cdot decade^{-1}}$. Towards 60 hPa the differences increase to around $0.05\,\mathrm{ppmv\cdot decade^{-1}}$. Between 60 hPa and 30 hPa the trend differences are rather constant. Higher up, they increase to more than $0.1\,\mathrm{ppmv\cdot decade^{-1}}$.

The MIPAS observations indicate positive trends at Boulder during the time period from 2002 to 2012. The trends decrease with increasing altitude from about 0.25 $\mathrm{ppmv \cdot decade^{-1}}$ at 100 hPa to 0.1 $\mathrm{ppmv \cdot decade^{-1}}$ at 20 hPa. For the zonal mean between 35°N and 45°N the trend estimates are also consistently positive. However, they show a slightly different altitude dependence than for Boulder. Below about 70 hPa the trends increase while higher up they decrease. In correspondence, the trend differences between the Boulder and zonal mean estimates are most pronounced below about 70 hPa, rising to 0.05 $\mathrm{ppmv \cdot decade^{-1}}$ at 100 hPa. Above 70 hPa the differences are close to zero. A very similar behaviour is observed for the trend differences between Boulder and the zonal mean considering the latitude band between 45°N and 55°N. The trend differences to the estimates for the latitude bands from 25°N to 35°N and the Equator to 60°N exhibit a pronounced altitude dependence. They decrease from more than 0.1 $\mathrm{ppmv \cdot decade^{-1}}$ at 100 hPa to -0.05 $\mathrm{ppmv \cdot decade^{-1}}$ at 20 hPa. The sign of the trend differences switches at about 60 hPa. The trend differences between Boulder and the zonal mean for 60°S to 60°N are positive at all altitudes. The smallest differences are close to zero and are observed between 45 hPa and 30 hPa. The largest difference is visible at 100 hPa with 0.15 $\mathrm{ppmv \cdot decade^{-1}}$.

The Boulder trends derived from the MLS observations from 2004 to 2016 are positive. They exhibit a pronounced altitude dependence. The trend estimates exhibit maxima at 70 hPa (0.4 $\mathrm{ppmv \cdot decade^{-1}}$) and 30 hPa (close to 0.3 $\mathrm{ppmv \cdot decade^{-1}}$). Minima are found at 100 hPa (0.2 $\mathrm{ppmv \cdot decade^{-1}}$), 45 hPa and 20 hPa (around 0.25 $\mathrm{ppmv \cdot decade^{-1}}$). The trends derived from the adapted time series are slightly larger than those calculated from the full time series. The trend differences between these two data sets are of a similar order as observed for the simulations addressed before. The MLS trends derived from the zonal mean time series for the different latitudes indicate a similar altitude dependence as that observed for Boulder. Overall, the trend differences between Boulder and the zonal means are generally within ± 0.05 $\mathrm{ppmv \cdot decade^{-1}}$. Prominent exceptions occur below 70 hPa for the differences to the zonal means from 25°N to 35°N, Equator to 60°N and 60°S to 60°N. Here, the differences can be as large as 0.15 $\mathrm{ppmv \cdot decade^{-1}}$. In addition, the trend differences between Boulder and the zonal mean from 60°S to 60°N are noticeably larger than for the other latitude bands, ranging from 0.05 $\mathrm{ppmv \cdot decade^{-1}}$ to 0.15 $\mathrm{ppmv \cdot decade^{-1}}$. Beyond that, the trend differences are consistently larger (by about 0.05 $\mathrm{ppmv \cdot decade^{-1}}$) for the adapted time series at altitudes around 40 hPa.

5 Discussion and conclusions

In this work we compared trend estimates for lower stratospheric water vapour between Boulder and zonal mean data around the Boulder latitude (35°N to 45°N) considering different time periods. For that we analysed multiple data sets, both from simulations and observations. The objective was to quantify how large these trend differences typically are and how much they could possibly help to explain the discrepancies in the trend estimates between the FPH observations at Boulder (Hurst et al., 2011) and a merged zonal mean satellite data set (Hegglin et al., 2014). For the time period from the late 1980s to 2010 the trend differences (FPH minus merged zonal mean satellite data set) range from 0.3 $\mathrm{ppmv \cdot decade^{-1}}$ to 0.5 $\mathrm{ppmv \cdot decade^{-1}}$, increasing with altitude.

Our analysis shows that there are differences in the trend estimates between Boulder and the zonal mean, both for the simulations and observations. These trend differences are dependent on altitude and the time period considered.

For the time period from the late 1980s to 2010 the simulations indicate trend differences between about $-0.02\,\mathrm{ppmv \cdot decade^{-1}}$ to $0.04\,\mathrm{ppmv \cdot decade^{-1}}$ (which are however not statistically significant different from zero). These are clearly smaller than the discrepancies in the trend estimates derived from the FPH observations and the merged satellite data set. The larger positive differences are observed close to 100 hPa. Here, the trend differences partly resolves the observational discrepancies. Above about 60 hPa the trend differences derived from the model simulations are however typically negative. This indicates that the trend estimates for the zonal mean data should be larger than at Boulder, which is contradictory to the observed trend differences between the FPH observations and the merged zonal mean satellite data set. Also, the simulations do not exhibit any pronounced deviations in the trend differences derived from time series using all data during a given month (which we referred to as full time series) or just using that closest in time to the actual FPH observations (which we referred to as adapted time series). This indicates that the temporal sampling has only a small influence on the trend differences on this time scale.

Given these model results, a different temporal behaviour between Boulder and the zonal mean is not a viable explanation for the discrepancies in the trend estimates derived from the local FPH observations and the merged zonal mean satellite data set presented by Hegglin et al. (2014). It still could be the case that the simulations underrepresent variability or that the trend differences originate from smaller spatial and temporal scales than are resolved by the model simulations (i.e. sub-grid processes). For the Boulder time series we used data in a 1000 km radius around the Boulder FPH observation site and within the latitude range from 35°N to 45°N. These criteria were primarily chosen for consistency with the analysis of the satellite observations whose exact measurement locations vary from orbit to orbit and day to day. In an additional analysis of the simulations, we considered for the Boulder time series only data from the closest grid point in space (EMAC: 40.5°N, 104.1°W, Δr=109 km; WACCM: 40.7°N, 105.0°W, Δr=80 km; CMAM: 39.0°N, 105.0°W, Δr=113 km; CLaMS: 40.0°N, 105.0°W, Δr=17 km). This analysis yields small quantitative changes (not shown here). Qualitatively, exactly the same conclusions can be drawn as from the standard analysis. The temporal resolutions of the analysed simulations vary (see Sect. 2.1). The CMAM simulation provides the best resolution in this analysis with 6 h. Accordingly the worst temporal mismatch to the actual FPH observations is 3 h. This gives an upper limit of temporal scales not covered in this analysis. Yet, arguably the different simulations yield similar results, as do the analysis of the full and the adapted time series.

For a single decade of data the trend differences between Boulder and the 35°N – 45°N zonal mean are typically larger than those discussed above for the entire time period from the late 1980s to 2010. The differences are typically within $\pm 0.10\,\mathrm{ppmv \cdot decade^{-1}}$, except close to 100 hPa where the differences can be occasionally as large as $\pm 0.2\,\mathrm{ppmv \cdot decade^{-1}}$. For the simulations, the trend differences derived from the adapted time series are typically larger than the trend differences obtained from the full time series on an absolute scale. A factor 2 is a common feature. In the MLS data, significant trend differences between the full and the adapted time series are observed around 40 hPa. These differences should to be kept in mind when comparing results for Boulder and the zonal mean on the shorter time scales.

In addition, we analysed trend differences between Boulder and the zonal means for a number of latitude bands. This aimed to investigate how representative the Boulder trends are for a more global scale. For the time period from the late 1980s to

2010 the simulations indicate trend differences within the interval from $-0.04\,\mathrm{ppmv \cdot decade^{-1}}$ to $0.06\,\mathrm{ppmv \cdot decade^{-1}}$. The largest differences occur when the Boulder trends are compared to those for the zonal mean of the latitude band between 60°S and 60°N. Based on these results, the Boulder trends should be quite representative (or a reasonable first guess) for the trends on more global scales during this time period. The caveats regarding missing variability or sub-grid processes in the simulations apply here as well. For shorter time periods, as covered by the individual satellite data sets, the representativeness gets smaller in general.

From our analysis it appears that a continued search for the reasons of the trend discrepancies between the FPH observations at Boulder and the merged satellite set is necessary (see list in the Introduction). In addition, even more differences become apparent. To start with, this considers the simulations themselves. The overall spread among the trend estimates derived from the different simulations can be almost as large as $0.2\,\mathrm{ppmv \cdot decade^{-1}}$. For the time period from the late 1980s to 2010 the spread varies between $0.06\,\mathrm{ppmv \cdot decade^{-1}}$ and $0.12\,\mathrm{ppmv \cdot decade^{-1}}$. To some degree this relativises the trend discrepancies between the FPH observations and the merged zonal mean satellite data set, if the spread among different simulations amounts to a considerable fraction of the discrepancies themselves. The reasons for the spread among the simulations are probably manifold, comprising general model characteristics (e.g. parameterisations, wave forcing, convection scheme, etc.), the choice of the nudged reanalysis data (and their quality over time, Fujiwara et al., 2017) or the exact details of the nudging (e.g. parameters, top height, relaxation time, etc.; see Sect. 2.1). Our analysis does not provide clear hints in a specific direction but leaves room for obvious followup activities.

Then, the trend estimates obtained from the simulations also differ from those derived from the FPH observations and the merged satellite data set (compare Fig. 1 and Fig. 5). Overall, they are closer to the trend estimates from the merged satellite data set, but consistently larger by about $0.05\,\mathrm{ppmv \cdot decade^{-1}}$ to $0.2\,\mathrm{ppmv \cdot decade^{-1}}$ depending on simulation and altitude. Compared to the FPH trend estimates the model results are consistently smaller by about $0.1\,\mathrm{ppmv \cdot decade^{-1}}$ to $0.45\,\mathrm{ppmv \cdot decade^{-1}}$. In many ways this situation is reminiscent to the results presented by Garcia et al. (2007) that indicated clear trend differences between the FPH observations, HALOE and a simulation from an older version of WACCM for the time period between 1992 and 2002.

A way forward is certainly to put more focus in understanding differences in time series of the water vapour anomalies instead of those in derived quantities. An example of this is shown in Fig. 8 which considers the difference between the de-seasonalised time series derived from the FPH observations at Boulder and the Boulder time series from the different simulations and satellite observations (i.e. FPH minus the other data sets) used in this work at $70\,\mathrm{hPa}$. For the simulations and the MLS data set the adapted Boulder time series are used while for the HALOE and MIPAS data sets the full time series are employed, as also indicated in the legend of the figure. Besides that, also the difference time series between the FPH observations and the zonal mean (35°N to 45°N) results exemplarily from the EMAC simulation as well as the ERBS/SAGE II (Earth Radiation Budget Satellite/Stratospheric Aerosol and Gas Experiment II) observations, derived with retrieval version 7.00 (Damadeo et al., 2013), are shown. These differences are marked with the suffix (zonal) in the figure legend. The ERBS/SAGE II data are considered here as they are also part of the merged satellite data set and actually define the start of the time series (Hegglin et al., 2014). The de-seasonalisation period of the FPH time series is always adapted to the time series it is compared to, i.e.

from 1985 to 2010 for the model simulations, from 1992 to 2005 for HALOE, from 2002 to 2012 for MIPAS, from 2004 to 2016 for MLS and from 1988 to 2005 for the SAGE II data set (see Sect. 3.3). For a clearer picture the differences are smoothed with a 1 year running mean. At least three valid data points during this period are required for a running mean to be considered further. The differences visible in the figure are also representative for other altitudes, even though some details are different. A number of aspects gain attention:

(1) The differences of the EMAC Boulder and zonal mean time series to the FPH observations are very similar. This is also true for the other simulations (not shown here). It highlights once more the main outcome of this study that the temporal behaviour at Boulder largely resembles that for the zonal mean around the Boulder latitude. More obvious deviations occur in the EMAC simulation between 1997 and 2000. This behaviour is also found in the WACCM simulation and to some degree in the CLaMS results, while the CMAM simulation show larger deviations around 1990. Also, in 2004 pronounced deviations are observed, consistently in all simulations (see also Fig. 2).

(2) Before 1986 the differences to FPH observations are predominantly negative (EMAC being the exception), while afterwards until 2011 they are mostly positive. As the trends in this work are derived by multi-linear regression with a single trend term, this behaviour is consistent with larger trend estimates for the FPH observations compared to the simulations for the time period from the late 1980s to 2010.

(3) While the SAGE II differences to the FPH observations mostly blend with the other data sets there is pronounced deviation between 1989 and 1991 (afterwards data are screened due to aerosol contamination by the Pinatubo eruption). During this time period the differences are more negative than for the model simulations. This behaviour is consistently observed below 30 hPa. Since this is close to the very beginning of the merged time series it has a pronounced effect on the trend estimates. It provides an explanation why the trend estimates derived from the merged satellite data set are smaller than those for the simulations considering the time period from the late 1980s to 2010. Overall, this might hint to a potential issue with the SAGE II data before the Pinatubo eruption. Alternatively, an issue might originate from the equal weighting of the pre- and post-Pinatubo SAGE II data in the merged satellite data set. More investigations are required to rule out any of these potential issues.

(4) The temporal development of the differences is quite consistent in qualitative terms for the various simulations and observational data sets. Features like the strong negative differences around 1993/1994, the subsequent increase until 2000, the relatively constant behaviour from 2001 to 2009 or the decrease starting in 2010 are visible for all simulations and satellite observations. Interestingly, we find a similar behaviour also in difference time series between frost point hygrometer observations at other stations and the simulations and satellite observations used in this work (not shown here). Explicitly, this applies to the NOAA FPH observations at Lauder (45°S, 169.7°E) and the CFH (cryogenic frost point hygrometer, Vömel et al., 2007) observations at San Jose (9.9°N, 84.0°W) and Lindenberg, (52.2°N, 14.1°E). In quantitative terms, the consistency of the differences is evidently is less good. The spread among the various data sets is on average 0.26 ppmv and is, thus, comparable to the differences between the FPH observations and the different simulations and satellite observations

themselves. In particular, between 1980 and 1985 there are huge deviations among the simulations in their differences to the FPH observations, relating to differences in the initial conditions and the spin up times among the simulations (except for CLaMS). After this period the average spread decreases to 0.21 ppmv.

In summary, understanding the differences shown in Fig. 8 and their temporal development, hopefully in combination with the merged satellite data set, should be a focal point of further research on lower stratospheric water vapour. This will inevitably yield better consistency in the trend estimates but also highlight the benefit of combining different data sources, as in situ observations, satellite measurements and modelling efforts.

Data availability.

Simulations:

- The data of the EMAC simulation described above will be made available in the Climate and Environmental Retrieval and Archive (CERA) database at the German Climate Computing Centre (DKRZ, website: http://cera-www.dkrz.de/WDCC/ui/Index.jsp). The corresponding digital object identifiers (DOI) will be published on the MESSy consortium website (http://www.messy-interface.org). Alternatively, the data can be obtained on request from Patrick Jöckel (patrick.joeckel@dlr.de).
- The WACCM data can be obtained on request from Doug Kinnison (dkin@ucar.edu).
- The CMAM simulation can be accessed from the following webpage: http://www.cccma.ec.gc.ca/data/cmam/CMAM/index.shtml.
- The CLaMS data can be obtained on request from Felix Ploeger (f.ploeger@fz-juelich.de).

Observations:

- The NOAA FPH data observed at Boulder can be downloaded from the FTP address ftp://ftp.cmdl.noaa.gov/data/ozwv/WaterVapor/Boulder_LEV or alternatively obtained on request from Dale Hurst (dale.hurst@noaa.gov).
- The HALOE data can be accessed on the following website: http://haloe.gats-inc.com/download/index.php.
- The MIPAS data are available on the following website: https://www.imk-asf.kit.edu/english/308.php.
- The MLS data can be downloaded from the following website: https://acdisc.gesdisc.eosdis.nasa.gov/data/Aura_MLS_Level2/ML2H2O.004/.
- The SAGE II data can be accessed from the following website: https://eosweb.larc.nasa.gov/project/sage2/sage2_table.

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

Acknowledgements. S. Lossow was funded by the DFG Research Unit "Stratospheric Change and its Role for Climate Prediction" (SHARP) under contract STI 210/9-2. The UTLS water vapour measurement record at Boulder, comprised of monthly balloon flights of the NOAA

FPH, continues in its 39$^{\text{th}}$ year thanks to funding from the NOAA Climate Program Office, the NASA Earth Science Division's Upper Atmospheric Composition Observations programme and the US Global Climate Observing System programme. We appreciate the HALOE Science Team and the many members of the HALOE project for producing and characterising the high quality HALOE data set. We would like to thank the European Space Agency (ESA) for making the MIPAS level-1b data set available. MLS data were obtained from the NASA Goddard Earth Sciences and Information Center. Work at the Jet Propulsion Laboratory, California Institute of Technology, was done under contract with the National Aeronautics and Space Administration. The EMAC simulation has been performed at the German Climate Computing Centre (DKRZ) through support from the Bundesministerium für Bildung und Forschung (BMBF). DKRZ and its scientific steering committee are gratefully acknowledged for providing the high performance computing (HPC) and data archiving resources for ESCiMo (Earth System Chemistry integrated Modelling) consortial project. We would like to thank the two anonymous reviewers for their valuable comments to improve the manuscript. We also appreciate comments from Ted Shepherd on an early version of this manuscript.

We acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of Karlsruhe Institute of Technology.

References

Avery, M. A., Davis, S. M., Rosenlof, K. H., Ye, H., and Dessler, A. E.: Large anomalies in lower stratospheric water vapour and ice during the 2015-2016 El Niño, Nature Geoscience, 10, 405 – 409, https://doi.org/10.1038/ngeo2961, 2017.

Ball, W. T., Alsing, J., Mortlock, D. J., Rozanov, E. V., Tummon, F., and Haigh, J. D.: Reconciling differences in stratospheric ozone composites, Atmospheric Chemistry & Physics, 17, 12 269 – 12 302, https://doi.org/10.5194/acp-17-12269-2017, 2017.

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

Brinkop, S., Dameris, M., Jöckel, P., Garny, H., Lossow, S., and Stiller, G.: The millennium water vapour drop in chemistry-climate model simulations, Atmospheric Chemistry & Physics, 16, 8125 – 8140, https://doi.org/10.5194/acp-16-8125-2016, 2016.

Damadeo, R. P., Zawodny, J. M., Thomason, L. W., and Iyer, N.: SAGE version 7.0 algorithm: application to SAGE II, Atmospheric Measurement Techniques, 6, 3539 – 3561, https://doi.org/10.5194/amt-6-3539-2013, 2013.

de Grandpré, J., Beagley, S. R., Fomichev, V. I., Griffioen, E., McConnell, J. C., Medvedev, A. S., and Shepherd, T. G.: Ozone climatology using interactive chemistry: Results from the Canadian Middle Atmosphere Model, Journal of Geophysical Research, 105, 26 475 – 26 492, https://doi.org/10.1029/2000JD900427, 2000.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553 – 597, https://doi.org/10.1002/qj.828, 2011.

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

Dessler, A. E., Ye, H., Wang, T., Schoeberl, M. R., Oman, L. D., Douglass, A. R., Butler, A. H., Rosenlof, K. H., Davis, S. M., and Portmann, R. W.: Transport of ice into the stratosphere and the humidification of the stratosphere over the 21st century, Geophysical Research Letters, 43, 2323 – 2329, https://doi.org/10.1002/2016GL067991, 2016.

Dvortsov, V. L. and Solomon, S.: Response of the stratospheric temperatures and ozone to past and future increases in stratospheric humidity, Journal of Geophysical Research, 106, 7505 – 7514, https://doi.org/10.1029/2000JD900637, 2001.

Eyring, V., Lamarque, J.-F., Hess, P., Arfeuille, F., Bowman, K., Chipperfield, M. P., Duncan, B., Fiore, A., Gettelman, A., Giorgetta, M. A., Granier, C., Hegglin, M. I., Kinnison, D., Kunze, M., Langematz, U., Luo, B., Martin, R., Matthes, K., Newman, P. A., Peter, T., Robock, A., Ryerson, T., Saiz-Lopez, A., Salawitch, R., Schultz, M., Shepherd, T. G., Shindell, D., Staehelin, J., Tegtmeier, S., Thomason, L., Tilmes, S., Vernier, J.-P., Waugh, D. W., and Young, P. J.: Overview of IGAC/SPARC Chemistry-Climate Model Initiative (CCMI) community simulations in support of upcoming ozone and climate assessments, SPARC Newsletter, 40, 48 – 66, 2013.

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

Forster, P. M. d. F. and Shine, K. P.: Assessing the climate impact of trends in stratospheric water vapor, Geophysical Research Letters, 29, 1086, https://doi.org/10.1029/2001GL013909, 2002.

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

Fujiwara, M., Wright, J. S., Manney, G. L., Gray, L. J., Anstey, J., Birner, T., Davis, S., Gerber, E. P., Harvey, V. L., Hegglin, M. I., Homeyer, C. R., Knox, J. A., Krüger, K., Lambert, A., Long, C. S., Martineau, P., Molod, A., Monge-Sanz, B. M., Santee, M. L., Tegtmeier, S., Chabrillat, S., Tan, D. G. H., Jackson, D. R., Polavarapu, S., Compo, G. P., Dragani, R., Ebisuzaki, W., Harada, Y., Kobayashi, C., McCarty, W., Onogi, K., Pawson, S., Simmons, A., Wargan, K., Whitaker, J. S., and Zou, C.-Z.: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems, Atmospheric Chemistry & Physics, 17, 1417 – 1452, https://doi.org/10.5194/acp-17-1417-2017, 2017.

Garcia, R. R., Marsh, D. R., Kinnison, D. E., Boville, B. A., and Sassi, F.: Simulation of secular trends in the middle atmosphere, 1950-2003, Journal of Geophysical Research, 112, D09 301, https://doi.org/10.1029/2006JD007485, 2007.

Gilford, D. M., Solomon, S., and Portmann, R. W.: Radiative impacts of the 2011 abrupt drops in water vapor and ozone in the tropical tropopause layer, Journal of Climate, 29, 595 – 612, https://doi.org/10.1175/JCLI-D-15-0167.1, 2016.

Gille, J. C. and Russell, J. M.: The Limb Infrared Monitor of the Stratosphere – Experiment description, performance, and results, Journal of Geophysical Research, 89, 5125 – 5140, https://doi.org/10.1029/JD089iD04p05125, 1984.

Hall, E. G., Jordan, A. F., Hurst, D. F., Oltmans, S. J., Vömel, H., Kühnreich, B., and Ebert, V.: Advancements, measurement uncertainties, and recent comparisons of the NOAA frost point hygrometer, Atmospheric Measurement Techniques, 9, 4295 – 4310, https://doi.org/10.5194/amt-9-4295-2016, 2016.

Hegglin, M. I., Plummer, D. A., Shepherd, T. G., Scinocca, J. F., Anderson, J., Froidevaux, L., Funke, B., Hurst, D., Rozanov, A., Urban, J., von Clarmann, T., Walker, K. A., Wang, H. J., Tegtmeier, S., and Weigel, K.: Vertical structure of stratospheric water vapour trends derived from merged satellite data, Nature Geoscience, 7, 768 – 776, https://doi.org/10.1038/ngeo2236, 2014.

Holton, J. R. and Gettelman, A.: Horizontal transport and the dehydration of the stratosphere, Geophysical Research Letters, 28, 2799 – 2802, https://doi.org/10.1029/2001GL013148, 2001.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S.: The Community Earth System Model: A framework for collaborative research, Bullentin of the American Meteorological Society, 94, 1339 – 1360, https://doi.org/10.1175/BAMS-D-12-00121.1, 2013.

Hurst, D. F., Oltmans, S. J., Vömel, H., Rosenlof, K. H., Davis, S. M., Ray, E. A., Hall, E. G., and Jordan, A. F.: Stratospheric water vapor trends over Boulder, Colorado: Analysis of the 30 year Boulder record, Journal of Geophysical Research, 116, D02 306, https://doi.org/10.1029/2010JD015065, 2011.

Jöckel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H., Baumgaertner, A., Gromov, S., and Kern, B.: Development cycle 2 of the Modular Earth Submodel System (MESSy2), Geoscientific Model Development, 3, 717 – 752, https://doi.org/10.5194/gmd-3-717-2010, 2010.

Jöckel, P., Tost, H., Pozzer, A., Kunze, M., Kirner, O., Brenninkmeijer, C. A. M., Brinkop, S., Cai, D. S., Dyroff, C., Eckstein, J., Frank, F., Garny, H., Gottschaldt, K.-D., Graf, P., Grewe, V., Kerkweg, A., Kern, B., Matthes, S., Mertens, M., Meul, S., Neumaier, M., Nützel, M., Oberländer-Hayn, S., Ruhnke, R., Runde, T., Sander, R., Scharffe, D., and Zahn, A.: Earth System Chemistry integrated Mod-

elling (ESCiMo) with the Modular Earth Submodel System (MESSy) version 2.51, Geoscientific Model Development, 9, 1153 – 1200, https://doi.org/10.5194/gmd-9-1153-2016, 2016.

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

Kunz, A., Müller, R., Homonnai, V., Jánosi, I. M., Hurst, D., Rap, A., Forster, P. M., Rohrer, F., Spelten, N., and Riese, M.: Extending water vapor trend observations over Boulder into the tropopause region: Trend uncertainties and resulting radiative forcing, Journal of Geophysical Research, 118, 11 269 – 11 284, https://doi.org/10.1002/jgrd.50831, 2013.

Livesey, N. J., Read, W. J., Wagner, P. A., Froidevaux, L., Lambert, A., Manney, G. L., Millan, L. F., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Fuller, R. A., Jarnot, R. F., Knosp, B. W., , and Martinez, E.: Aura/MLS version 4.2x level 2 data quality and description document, http://mls.jpl.nasa.gov/data/v4-2_data_quality_document.pdf, Last access: 18 August 2017, 2015.

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

Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N., and Polvani, L. M.: Climate change from 1850 to 2005 simulated in CESM1(WACCM), Journal of Climate, 26, 7372 – 7391, https://doi.org/10.1175/JCLI-D-12-00558.1, 2013.

Maycock, A. C., Joshi, M. M., Shine, K. P., Davis, S. M., and Rosenlof, K. H.: The potential impact of changes in lower stratospheric water vapour on stratospheric temperatures over the past 30 years, Quarterly Journal of the Royal Meteorological Society, 140, 2176 – 2185, https://doi.org/10.1002/qj.2287, 2014.

McKenna, D. S., Grooß, J.-U., Günther, G., Konopka, P., Müller, R., Carver, G., and Sasano, Y.: A new Chemical Lagrangian Model of the Stratosphere (CLaMS) 2. Formulation of chemistry scheme and initialization, Journal of Geophysical Research, 107, 4256, https://doi.org/10.1029/2000JD000113, 2002a.

McKenna, D. S., Konopka, P., Grooß, J.-U., Günther, G., Müller, R., Spang, R., Offermann, D., and Orsolini, Y.: A new Chemical Lagrangian Model of the Stratosphere (CLaMS) 1. Formulation of advection and mixing, Journal of Geophysical Research, 107, 4309, https://doi.org/10.1029/2000JD000114, 2002b.

McLandress, C., Scinocca, J. F., Shepherd, T. G., Reader, M. C., and Manney, G. L.: Dynamical control of the mesosphere by orographic and nonorographic gravity wave drag during the extended northern winters of 2006 and 2009, Journal of the Atmospheric Sciences, 70, 2152 – 2169, https://doi.org/10.1175/JAS-D-12-0297.1, 2013.

McLandress, C., Plummer, D. A., and Shepherd, T. G.: Technical Note: A simple procedure for removing temporal discontinuities in ERA-Interim upper stratospheric temperatures for use in nudged chemistry-climate model simulations, Atmospheric Chemistry & Physics, pp. 1547 – 1555, https://doi.org/10.5194/acp-14-1547-2014, 2014.

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

Notholt, J., Toon, G. C., Fueglistaler, S., Wennberg, P. O., Irion, F. W., McCarthy, M., Scharringhausen, M., Siek Rhee, T., Kleinböhl, A., and Velazco, V.: Trend in ice moistening the stratosphere - constraints from isotope data of water and methane, Atmospheric Chemistry & Physics, 10, 201 – 207, https://doi.org/10.5194/acp-10-201-2010, 2010.

Oltmans, S. J. and Hofmann, D. J.: Increase in lower-stratospheric water vapour at a mid-latitude Northern Hemisphere site from 1981 to 1994, Nature, 374, 146 – 149, https://doi.org/10.1038/374146a0, 1995.

Oltmans, S. J., Vömel, H., Hofmann, D. J., Rosenlof, K. H., and Kley, D.: The increase in stratospheric water vapor from balloonborne, frostpoint hygrometer measurements at Washington, D.C., and Boulder, Colorado, Geophysical Research Letters, 27, 3453 – 3456, https://doi.org/10.1029/2000GL012133, 2000.

Ploeger, F., Günther, G., Konopka, P., Fueglistaler, S., Müller, R., Hoppe, C., Kunz, A., Spang, R., Grooß, J.-U., and Riese, M.: Horizontal water vapor transport in the lower stratosphere from subtropics to high latitudes during boreal summer, Journal of Geophysical Research, 118, 8111 – 8127, https://doi.org/10.1002/jgrd.50636, 2013.

Randel, W. J., Wu, F., Vömel, H., Nedoluha, G. E., and Forster, P.: Decreases in stratospheric water vapor after 2001: Links to changes in the tropical tropopause and the Brewer-Dobson circulation, Journal of Geophysical Research, 111, D12 312, https://doi.org/10.1029/2005JD006744, 2006.

Randel, W. J., Polvani, L., Wu, F., Kinnison, D. E., Zou, C., and Mears, C.: Troposphere-stratosphere temperature trends derived from satellite data compared with ensemble simulations from WACCM, Journal of Geophysical Research, 122, 1 – 17, https://doi.org/10.1002/2017JD027158, 2017.

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

Riese, M., Ploeger, F., Rap, A., Vogel, B., Konopka, P., Dameris, M., and Forster, P.: Impact of uncertainties in atmospheric mixing on simulated UTLS composition and related radiative effects, Journal of Geophysical Research, 117, D16 305, https://doi.org/10.1029/2012JD017751, 2012.

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., Manzini, E., Schlese, U., and Schulzweida, U.: Sensitivity of Simulated Climate to Horizontal and Vertical Resolution in the ECHAM5 Atmosphere Model, Journal of Climate, 19, 3771, https://doi.org/10.1175/JCLI3824.1, 2006.

Rosenlof, K. H. and Reid, G. C.: Trends in the temperature and water vapor content of the tropical lower stratosphere: Sea surface connection, Journal of Geophysical Research, 113, D06 107, https://doi.org/10.1029/2007JD009109, 2008.

Scherer, M., Vömel, H., Fueglistaler, S., Oltmans, S. J., and Staehelin, J.: Trends and variability of midlatitude stratospheric water vapour deduced from the re-evaluated Boulder balloon series and HALOE, Atmospheric Chemistry & Physics, 8, 1391 – 1402, https://doi.org/10.5194/acp-8-1391-2008, 2008.

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

Scinocca, J. F., McFarlane, N. A., Lazare, M., Li, J., and Plummer, D.: Technical Note: The CCCma third generation AGCM and its extension into the middle atmosphere, Atmospheric Chemistry & Physics, 8, 7055 – 7074, 2008.

Solomon, S.: Stratospheric ozone depletion: A review of concepts and history, Reviews of Geophysics, 37, 275 – 316, https://doi.org/10.1029/1999RG900008, 1999.

Solomon, S., Rosenlof, K. H., Portmann, R. W., Daniel, J. S., Davis, S. M., Sanford, T. J., and Plattner, G.: Contributions of stratospheric water vapor to decadal changes in the rate of global warming, Science, 327, 1219 – 1223, https://doi.org/10.1126/science.1182488, 2010.

Stenke, A. and Grewe, V.: Simulation of stratospheric water vapor trends: impact on stratospheric ozone chemistry, Atmospheric Chemistry & Physics, 5, 1257 – 1272, 2005.

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

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quarterly Journal of the Royal Meteorological Society, 131, 2961 – 3012, https://doi.org/10.1256/qj.04.176, 2005.

Vömel, H., David, D. E., and Smith, K.: Accuracy of tropospheric and stratospheric water vapor measurements by the cryogenic frost point hygrometer: Instrumental details and observations, Journal of Geophysical Research, 112, D08 305, https://doi.org/10.1029/2006JD007224, 2007.

von Clarmann, T., Stiller, G., Grabowski, U., Eckert, E., and Orphal, J.: Technical Note: Trend estimation from irregularly sampled, correlated data, Atmospheric Chemistry & Physics, 10, 6737 – 6747, 2010.

Waters, J. W., Froidevaux, L., Harwood, R. S., Jarnot, R. F., Pickett, H. M., Read, W. G., Siegel, P. H., Cofield, R. E., Filipiak, M. J., Flower, D. A., Holden, J. R., Lau, G. K., Livesey, N. J., Manney, G. L., Pumphrey, H. C., Santee, M. L., Wu, D. L., Cuddy, D. T., Lay, R. R., Loo, M. S., Perun, V. S., Schwartz, M. J., Stek, P. C., Thurstans, R. P., Boyles, M. A., Chandra, K. M., Chavez, M. C., Chen, G.-S., Chudasama, B. V., Dodge, R., Fuller, R. A., Girard, M. A., Jiang, J. H., Jiang, Y., Knosp, B. W., Labelle, R. C., Lam, J. C., Lee, A. K., Miller, D., Oswald, J. E., Patel, N. C., Pukala, D. M., Quintero, O., Scaff, D. M., Vansnyder, W., Tope, M. C., Wagner, P. A., and Walch, M. J.: The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura Satellite, IEEE Transactions on Geoscience and Remote Sensing, 44, 1075 – 1092, https://doi.org/10.1109/TGRS.2006.873771, 2006.

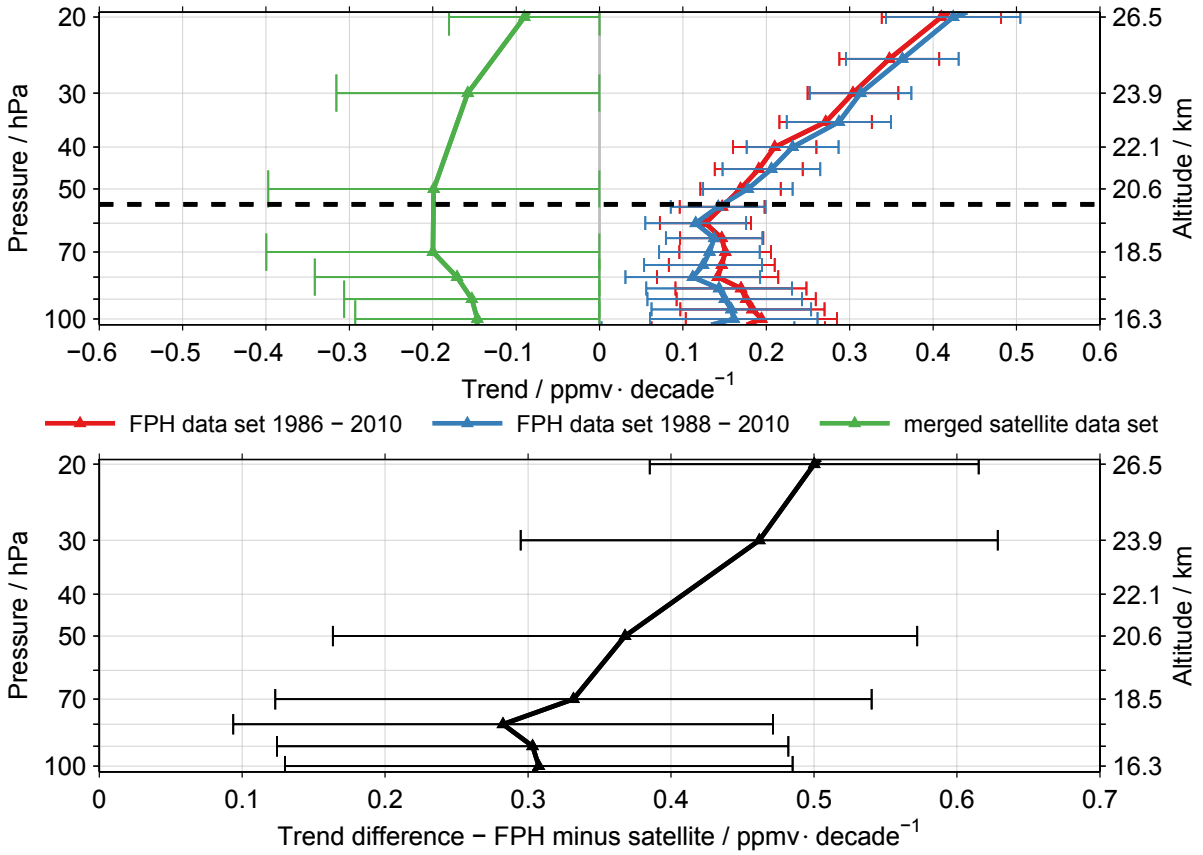

Figure 1. Upper panel: In green the approximated trend estimates derived from the merged satellite data set for the latitude band between 35°N and 45°N are shown. See text for more details. Below the dashed line these estimates consider the time period from 1988 to 2010, above they are for the time period from 1986 to 2010. In red and blue the corresponding trend estimates derived from the FPH observations at Boulder are shown. The error bars represent the 2σ uncertainty. The right axis provides an approximation of the altitude in geometrical terms. This information is derived from the MIPAS data. Lower panel: Difference between the trend estimates derived from the FPH observations and merged satellite data set.

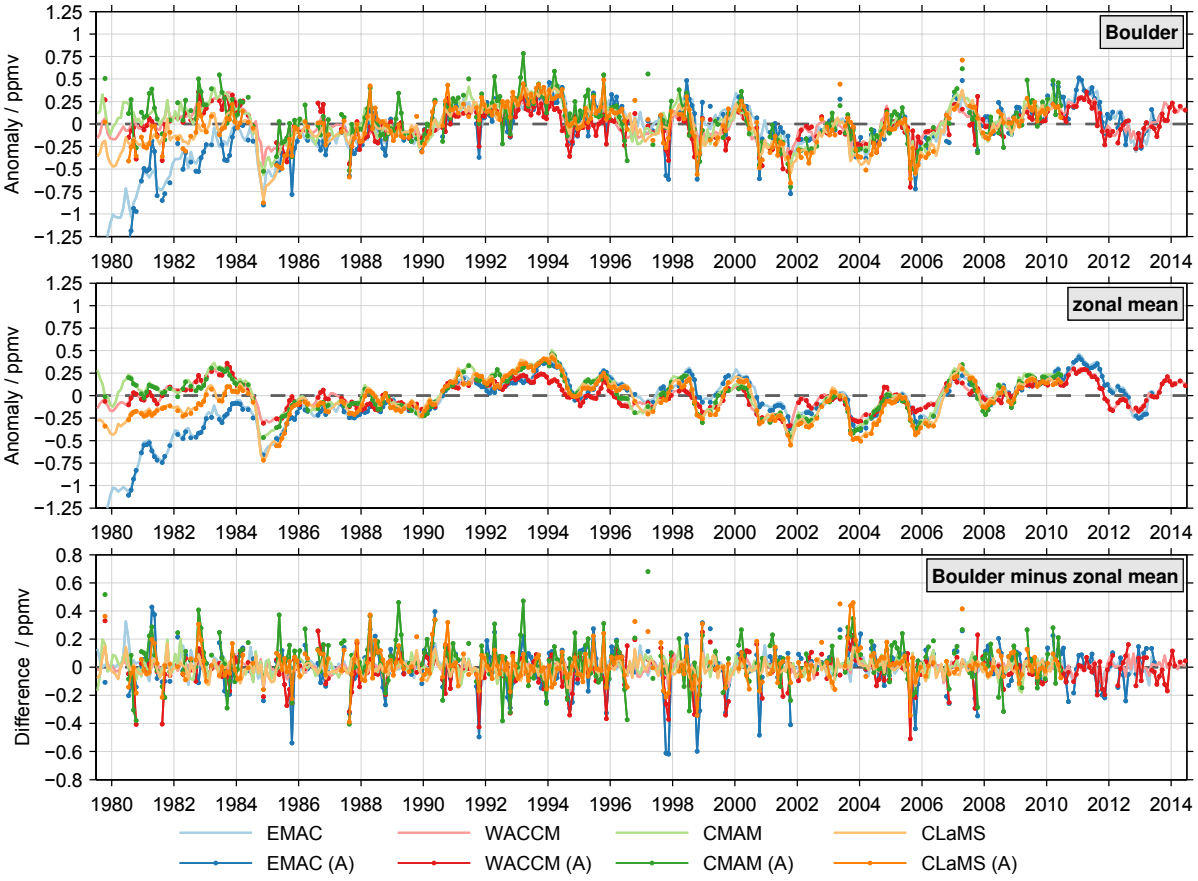

Figure 2. De-seasonalised time series for Boulder (top panel), the 35°N to 45°N zonal mean (middle panel) and its difference (low panel) for a number of model simulations considering the pressure level of 70 hPa. Results labelled with the suffix (A) are adapted to the actual FPH observations at Boulder, see text for more details. The time ticks consider the middle of the specified years.

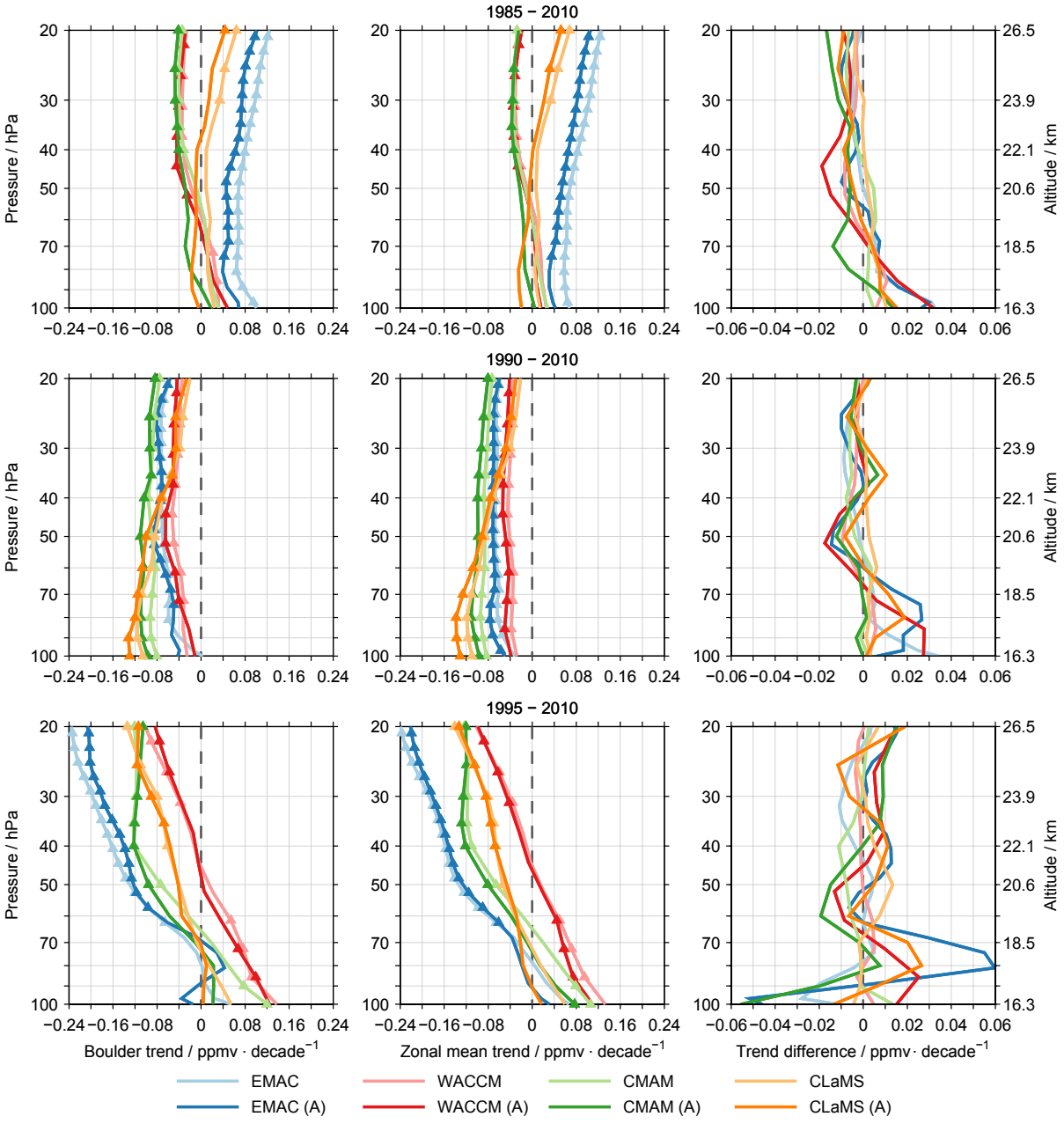

Figure 3. Trend estimates for the different model simulations for Boulder (left panels), the zonal mean for the latitude band between 35°N and 45°N (centre panels) and their corresponding differences (right panels). The different rows consider different time period as indicated in the title of the centre panels. Trends and trend differences significant at the 2σ uncertainty level are marked by triangles.

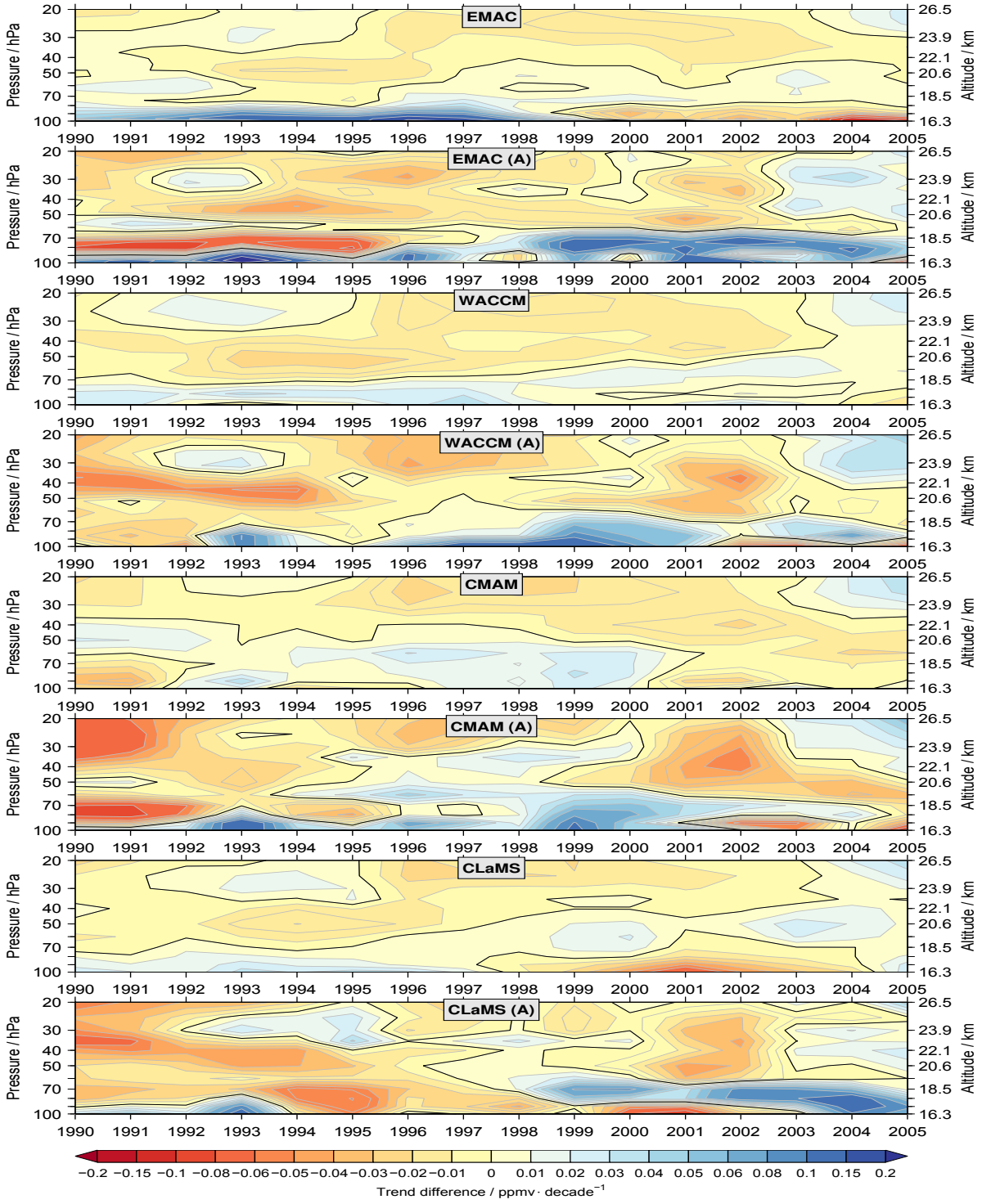

Figure 4. The temporal development of the trend differences between the Boulder and the zonal mean (35°N – 45°N) time series, based on 11 year time intervals. The results are given at the centre of the corresponding time intervals, i.e. in 1995 for the time period between 1990 and 2000. The black lines indicate zero trend differences.

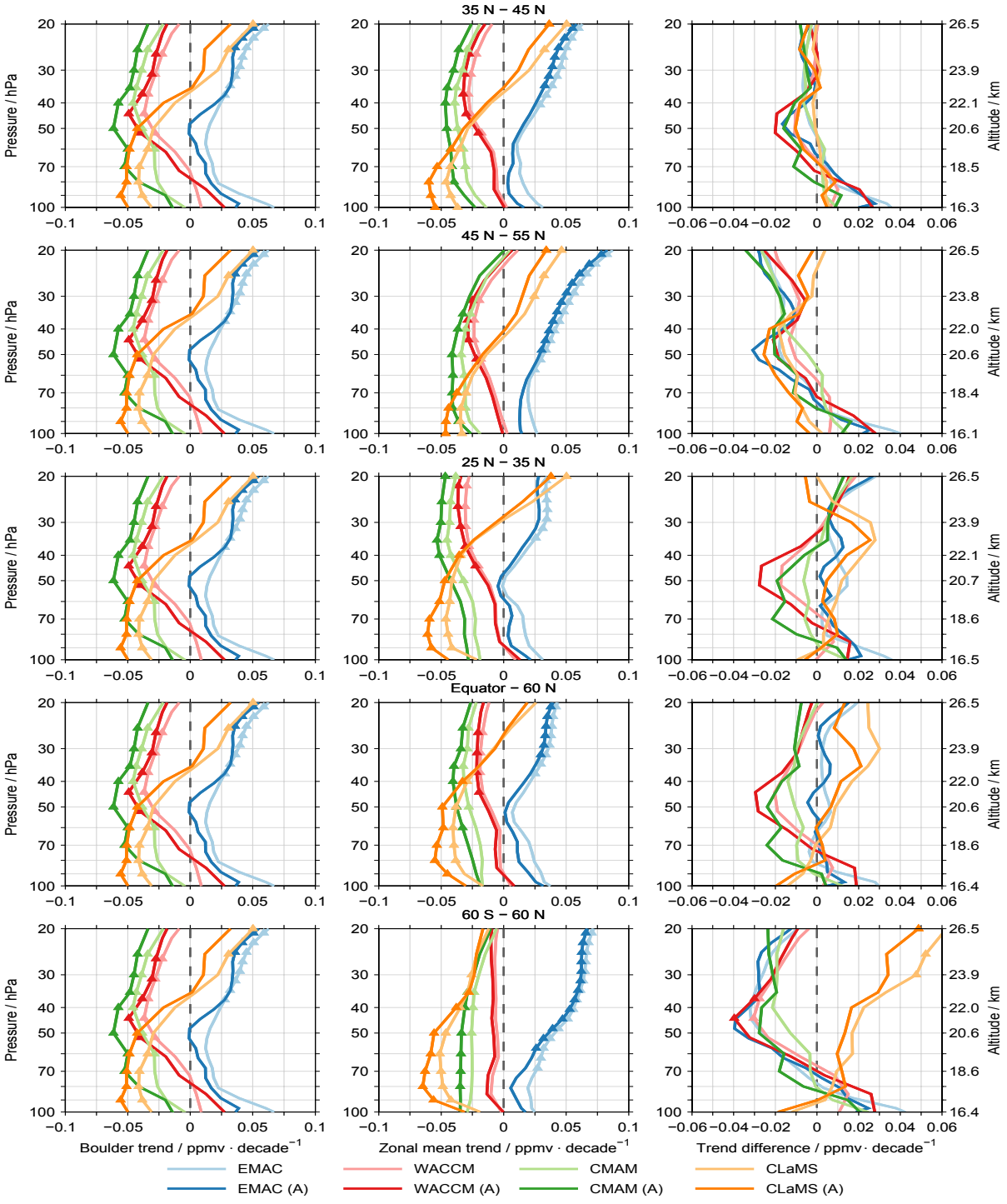

Figure 5. Comparison of the trend estimates at Boulder (left panels) and the zonal means for different latitude bands (centre panels) as indicted in the title. As in Fig. 3 the right panels show the difference between the two trends. The comparisons consider the time period between 1987 and 2010. The left panels are all the same and are repeated for convenience. Trends and trend differences significant at the 2σ uncertainty level are again marked by triangles.

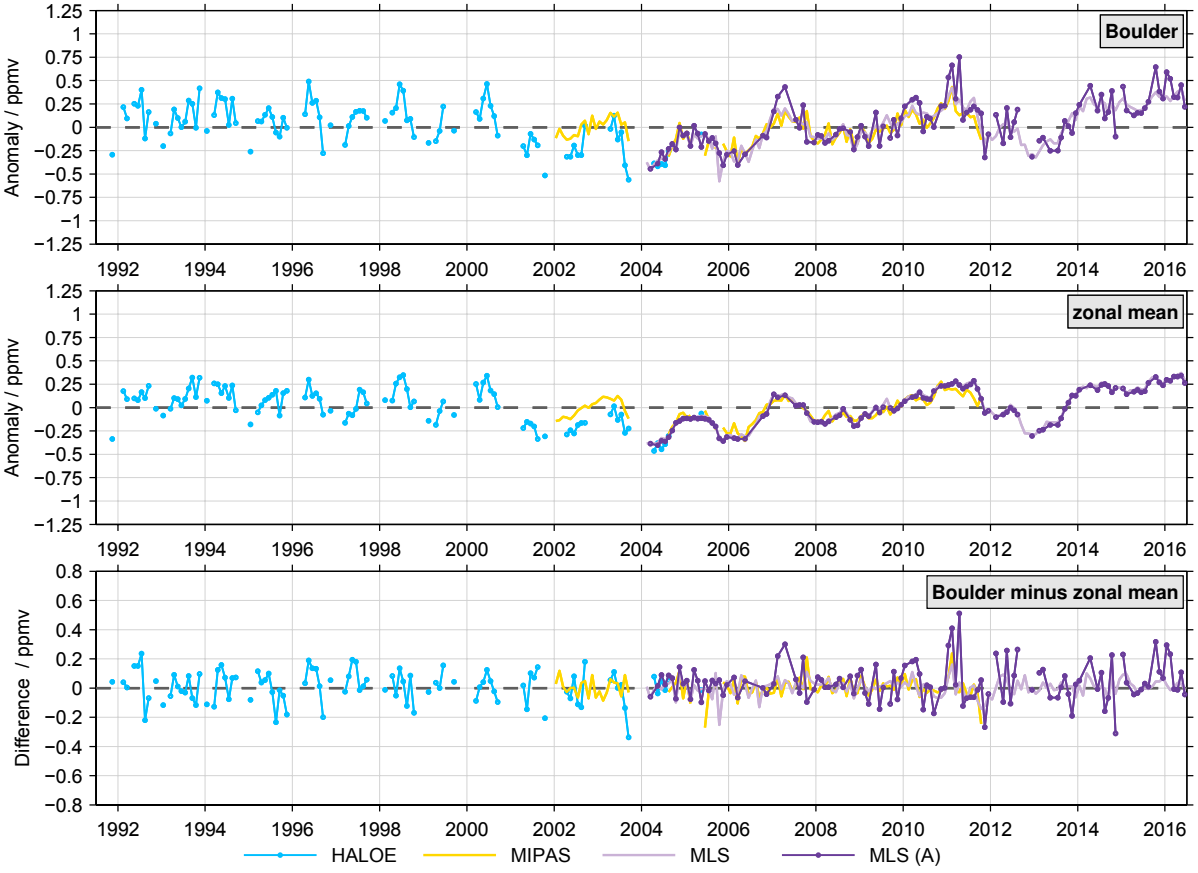

Figure 6. As Fig. 2 but here showing several observational results.

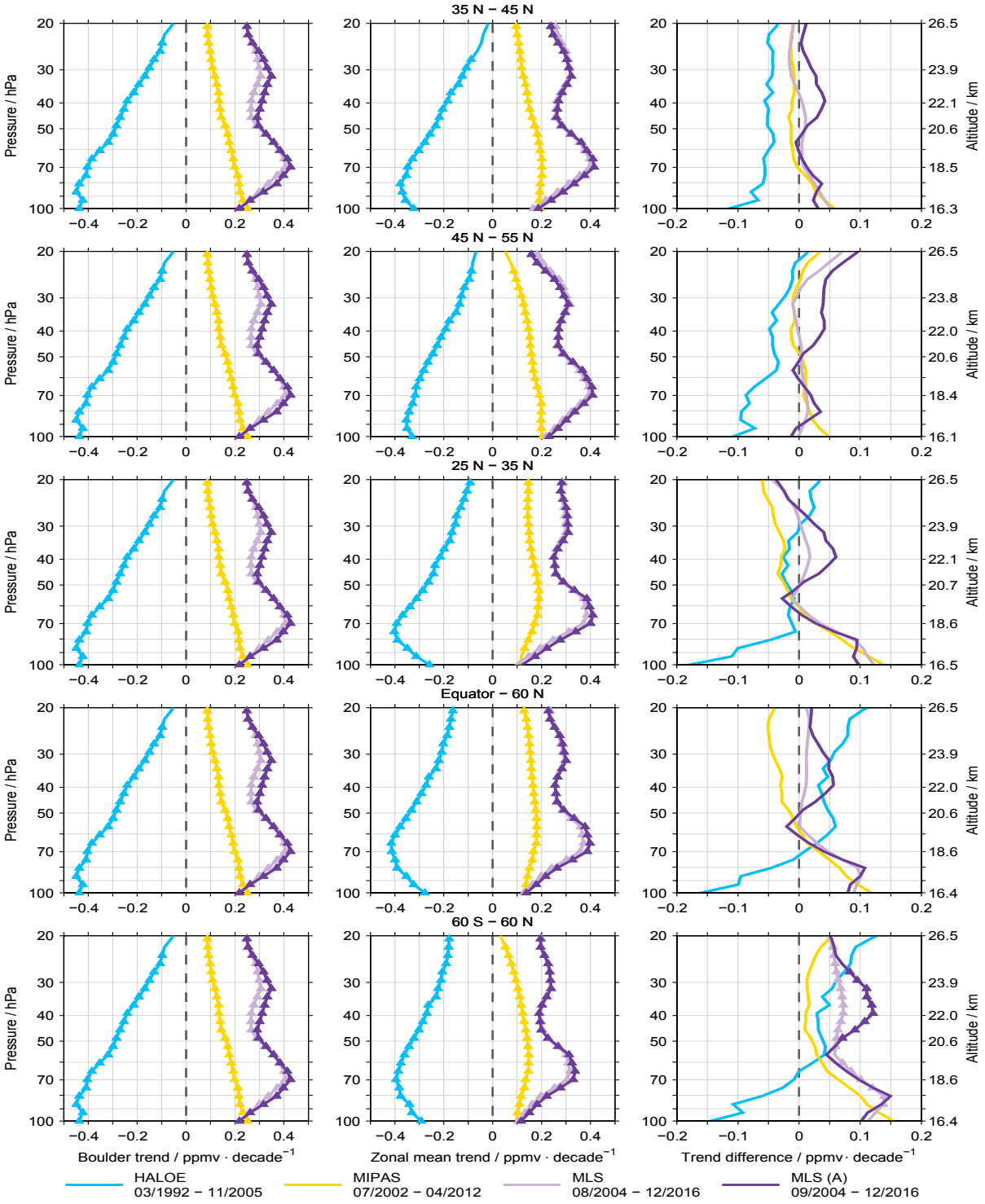

Figure 7. As Fig. 5 but here again for the observations.

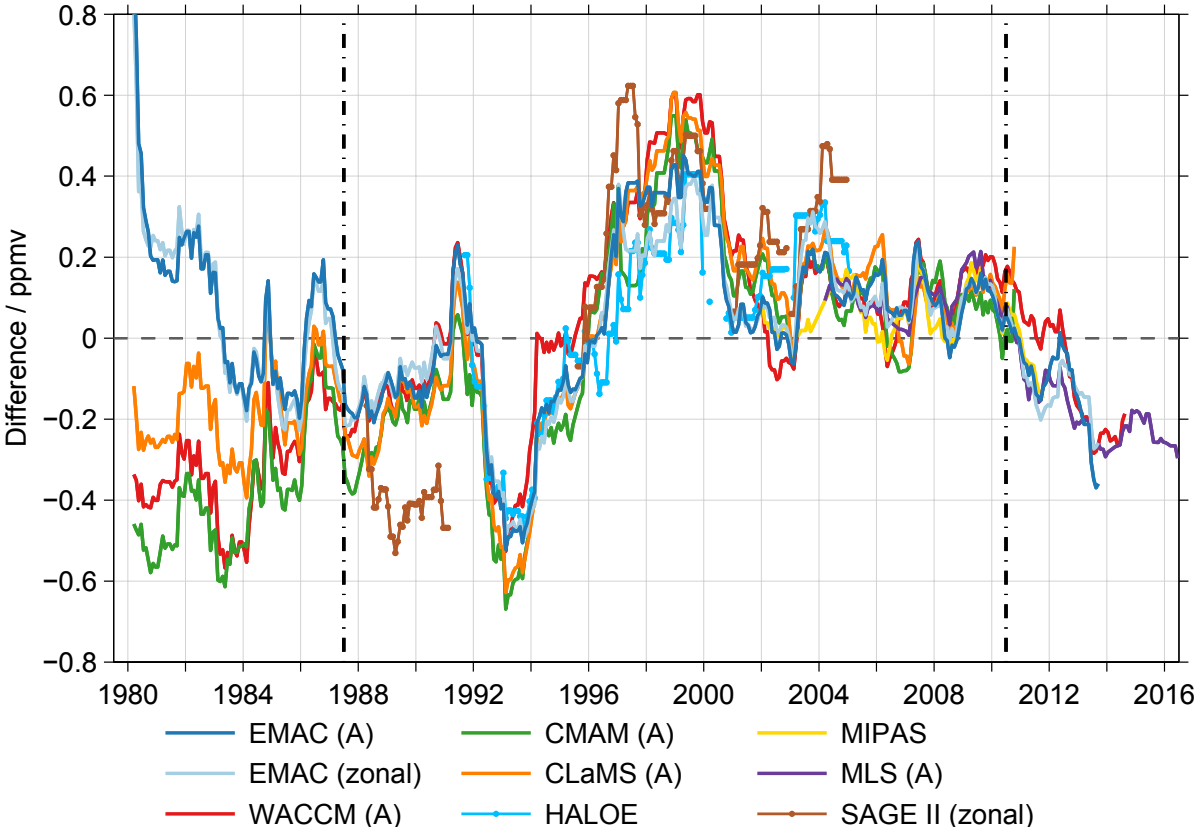

Figure 8. Differences between the de-seasonalised time series obtained from the FPH observations and the Boulder time series derived from the different simulations and observational results at 70 hPa. In addition, the differences between the FPH time series and the zonal mean (35°N to 45°N) time series for EMAC and SAGE II are shown. To provide a clearer picture, the differences are smoothed with a 1 year running average. At least three data points are required for a valid running average. The dashed-dotted line indicates the time period covered by the merged satellite data set at this altitude. The time ticks consider again the middle of the specified years.