Recent divergences in stratospheric water vapor measurements by frost point hygrometers and the Aura Microwave Limb Sounder

Dale F. Hurst[1,2], William G. Read[3], Holger Vömel[4], Henry B. Selkirk[5,6], Karen H. Rosenlof[7], Sean M. Davis[1,7], Emrys G. Hall[1,2], Allen F. Jordan[1,2], Samuel J. Oltmans[1,2]

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA
[2]Global Monitoring Division, NOAA Earth System Research Laboratory, Boulder, Colorado, USA
[3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California
[4]Earth Observing Laboratory, National Center for Atmospheric Research, Boulder, Colorado, USA
[5]Laboratory for Atmospheric Chemistry and Dynamics, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
[6]Goddard Earth Science Technology and Research, Universities Space Research Association, Columbia, Maryland, USA
[7]Chemical Sciences Division, NOAA Earth System Research Laboratory, Boulder, Colorado, USA

Correspondence to: Dale Hurst (Dale.Hurst@noaa.gov)

Abstract. Balloon-borne frost point hygrometers (FPs) and the Aura Microwave Limb Sounder (MLS) provide high-quality vertical profile measurements of water vapor in the upper troposphere and lower stratosphere (UTLS). A previous comparison of stratospheric water vapor measurements by FPs and MLS over three FP sites, Boulder, Colorado (40.0°N), Hilo, Hawaii (19.7°N) and Lauder, New Zealand (45.0°S), from August 2004 through December 2012, demonstrated agreement better than 1% between 68 and 26 hPa, but also exposed statistically significant biases of 2 to 10% at 83 and 100 hPa (Hurst et al., 2014). A simple linear regression analysis of the FPH-MLS differences revealed no significant long-term drifts between the two instruments. Here we extend the drift comparison to mid-2015 and add two FP sites, Lindenberg, Germany (52.2°N) and San José, Costa Rica (10.0°N) that employ FPs of different manufacture and calibration for their water vapor soundings. The extended comparison period reveals that stratospheric FP and MLS measurements over 4 of the 5 sites have diverged at rates of 0.03 to 0.07 ppmv yr^{-1} (0.6 to 1.5% yr^{-1}) from ~2010 to mid-2015. These rates are similar in magnitude to the 30-year (1980-2010) average growth rate of stratospheric water vapor (~1% yr^{-1}) measured by FPs over Boulder (Hurst et al., 2011). By mid-2015, the FP-MLS differences at some sites were large enough to exceed the combined accuracy estimates of the FP and MLS measurements.

1 Introduction

Water vapor in the Earth's atmosphere influences the radiation budget by strongly attenuating outgoing long-wave radiation. Though the lower troposphere holds the vast majority of atmospheric water vapor, abundance changes in the relatively dry upper troposphere and lower stratosphere (UTLS) can significantly impact global surface temperatures and climate (Forster and Shine, 2002; Solomon et al., 2010). Satellite-based remote sensors have greatly enhanced our ability to monitor UTLS water vapor on a near-global scale. However, because of the limited operational lifetimes of satellite sensors, an analysis of

trends over decadal or longer scales requires the merging of measurements by different instruments. Efforts to combine UTLS water vapor data sets from different satellites have demonstrated the need to reduce measurement biases between instruments before trend analyses are performed (Davis et al., 2016; Hegglin et al., 2014; Froidevaux et al., 2015). The necessity of adjusting data sets before they are merged adds an additional source of uncertainty to any determination of long-

5 term trends.

Balloon-borne frost point hygrometers (FPs) provide vertical profile measurements of water vapor at high resolution from the surface to the middle stratosphere (~28 km). Measurement programs with FPs typically focus on the UTLS for the purpose of long-term climate monitoring and/or studies of processes that influence humidity in the upper atmosphere (e.g.,

10 cloud microphysical processes that regulate dehydration). Though FP data sets are spatially and temporally sparse compared to those produced by satellite sensors, long-term records of UTLS water vapor, like the 36-year record over Boulder, Colorado, are invaluable for determining long-term trends (Oltmans and Hofmann, 1995; Oltmans et al., 2000; Rosenlof et al., 2001; Scherer et al., 2008; Hurst et al., 2011) and for validating satellite-based remote sensors like the Aura Microwave Limb Sounder (Vömel et al., 2007a; Hurst et al., 2014).

Nearly every day since August 2004 the Aura MLS has provided ~3500 near-global vertical profile measurements of water vapor measurements from the UT well into the mesosphere, and continues to do so today. Stratospheric water vapor measurements by the MLS and NOAA frost point hygrometers (FPHs) were recently compared to evaluate biases and temporal drifts between them during the period August 2004 through December 2012 (Hurst et al., 2014). Measurements

20 over three UTLS water vapor monitoring sites of the Global Monitoring Division of NOAA's Earth System Research Laboratory were compared: Boulder, Colorado; Hilo, Hawaii; and Lauder, New Zealand. Statistically significant FPH–MLS biases ranging from –0.10 (–2.2%) to –0.46 ppmv (–10.3%) were reported at 100 hPa over all three sites and at 83 hPa over Boulder and Hilo. Higher in the stratosphere, at the six MLS retrieval pressures from 68 to 26 hPa, the average FPH-MLS agreement was better than 0.04 ppmv (0.8%). FPH–MLS differences at each of the three sites were also analyzed for

25 temporal drifts using weighted linear regression fits to the full records. With a few minor exceptions the linear trends in FPH–MLS differences through the end of 2012 were not statistically different from zero (Hurst et al., 2014).

Here we present an updated comparison of stratospheric water vapor measurements by FPs and the MLS for the period August 2004 through June 2015. Data from two different types of FPs are used: the NOAA FPH (Mastenbrook and Oltmans,

30 1983; Hall et al., 2016) and the cryogenic frost point hygrometer (CFH) (Vömel et al., 2007b; Vömel et al., 2016). The balloon-borne measurements are compared to MLS profiles obtained during overpasses of Boulder, Hilo, Lauder and two additional FP sounding sites, Lindenberg, Germany and San José, Costa Rica (Table 1). Note that the Hilo and Lauder FP soundings were performed exclusively with the NOAA FPH, the Lindenberg and San José profiles are solely from the CFH, and the Boulder record combines soundings by both FP types. Though both FP types use the same measurement principle,

they are built from different parts, are independently calibrated and have subtle yet important differences in their software and frost control logic. Data from the two FP types are also independently processed and quality assured.

FP profiles at each site are independently compared to MLS version 3.3 (v3.3) and the latest v4.2 water vapor retrievals
5   using the same analysis methods. MLS v3.3 water vapor was retrieved until June 30, 2015, after which only v4.2 data are available. MLS v4.2 retrievals feature an improved cloud detection methodology, use more spectral channels and include an improved forward model for greater accuracy (Livesey et al., 2015). Unless otherwise noted, the values presented in the text and figures pertain to the comparison conducted with MLS v3.3 retrievals. Tables presenting results based on MLS v3.3 and v4.2 are so specified. We consider it essential to evaluate both MLS versions because many papers have been written using
10   v3.3 retrievals and many more will be published using v4.2 retrievals. All water vapor mixing ratios are reported as mole fractions (μmol mol^{-1} dry air) in units of parts per million by volume (ppmv).

2 Methods

For the purpose of comparison each FP profile was first convolved with the MLS averaging kernels to degrade its high vertical resolution to the ~3 km resolution of lower stratospheric MLS retrievals and place the FP mixing ratio "retrievals"
15   on the MLS pressure grid (Read et al., 2007; Lambert et al., 2007). FP profiles were independently convolved with the MLS v3.3 and v4.2 averaging kernels for 8 MLS retrieval pressure levels: 100, 83, 68, 56, 46, 38, 32 and 26 hPa. FP mixing ratios were not retrieved at pressures <26 hPa because the averaging kernels require data above the typical maximum altitude of high-quality FP measurements. Although convolved FP retrievals at pressures >100 hPa are feasible, the coincidence criteria applied to FP and MLS retrievals at pressure levels 100-26 hPa (see below) produced very noisy comparison results at >100
20   hPa, likely because of the much greater variability of water vapor at pressures >100 hPa, especially in the tropics. Applying more stringent coincidence criteria to improve the spatiotemporal matching of FP and MLS data below 100 hPa severely reduces the number of coincident profiles at each site and diminishes the value of the statistics generated by this type of comparison.

25   The same spatial criteria presented as "coincidence criteria set #1" in Hurst et al. (2014), within ±2° latitude and ±8° longitude, were employed to identify MLS profiles proximate to the five FP sounding sites. The spatially coincident MLS retrievals are plotted as time series along with the convolved FP "retrievals" at 68 hPa over each site (Figure 1). Note that towards the end of each record many of the FP mixing ratio retrievals reside near the lower limits of the MLS data envelope.

30   For this work a criterion of ±18 hours was used to identify temporally coincident MLS and FP profiles. This enabled MLS profiles to be compared with 94-100% of the FP soundings at each site. Employing the spatial and temporal criteria together, an average of 4-6 spatiotemporally coincident MLS overpass profiles were identified per FP sounding at each of the 5 sites

(Table 1). As in Hurst et al. (2014) the multiple MLS profiles coincident with each FP flight were distilled into a single "median" coincident profile composed of the median MLS mixing ratio at each pressure level. Our choice to use median rather than mean mixing ratios reduces the potential for any anomalous MLS retrievals to skew the values used for this comparison.

FP–MLS differences were calculated for each FP sounding by subtracting the MLS median coincident profile from the convolved FP profile. Statistical outliers were identified independently for each site and pressure level by evaluating the residuals of FP–MLS differences from smoothed time series of the differences. Points with absolute residuals exceeding twice the mean absolute residual were flagged as outliers and excluded from further study. Approximately 10% of the FP–
10   MLS differences were flagged as outliers.

For some sites the records of FP–MLS differences at 68 hPa visually exhibit time-dependent changes in trends (Figure 2). Many of the time series at other pressure levels over the sites (not shown) also show these same characteristics. Intuitively, full-record linear trend analyses of these time series of differences would greatly misrepresent the data. Instead, the time-
15   dependent changes in these records indicate they should be evaluated for a statistically significant "changepoint", the point where the mean of the time series first undergoes a structural pattern change. Such an analysis was performed on each time series of FP–MLS differences using the two-phase regression model described by Lund and Reeves (2002). The model considers every data point to be a potential undocumented changepoint and calculates an F-statistic for each. The F-statistic is a quantitative assessment of how much the sum of squared residuals is reduced when the time series is fit in two periods
20   (separated by the changepoint) instead of one period. The maximum in the time series of F-statistics, F_{max}, identifies the most probable changepoint in the time series.

The two-phase regression model was first applied to time series of smoothed FP–MLS differences at each site to look for conformity between the detected changepoints. Except for Hilo, nearly all of the changepoints identified for the 8 pressure
25   levels above each site were within ±1 yr of the mean changepoint for the site. This intra-site consistency facilitated the recognition of any non-conforming changepoints found when the model was applied to time series of unsmoothed FP–MLS differences. When an anomalous changepoint was detected in the unsmoothed differences, the time series of F-statistics was examined for a secondary maximum nearer in time to the consensus changepoint for that site. The value of the F-statistic at the secondary maximum was typically only slightly less than F_{max} so the more conforming changepoint of the secondary
30   maximum was used instead of the anomalous changepoint.

The dates and confidence levels of the changepoints for each time series of FP-MLS differences (except at Hilo) are presented in Table 2. For Hilo the analysis found no discernable maxima (F_{max}) in the time series of F-statistics, likely

because the record only began at the end of 2010, after the changepoints determined for most other sites. Visually the time series of differences at Hilo depict decreasing trends from the start of the record (Figure 2b).

The confidence level of each changepoint was calculated using the 90th, 95th and 99th percentiles of the F_{max} distribution as a

5 function of n (time series length) presented in Table 1 of Lund and Reeves (2002). Confidence levels for F-statistic values between the 90th and 99th percentiles, and for values below the 90th percentile were interpolated and extrapolated, respectively, using a quadratic fit to the n-dependent percentiles. Confidence levels for F-statistic values above the 99th percentile are reported as >99% (Table 2). Of the 32 changepoints identified for Lindenberg, Boulder, San José and Lauder, the confidence levels of 24 are ≥90% and all but 4 are >68%, substantiating the need to break each time series into two

10 separate intervals (periods 1 and 2) for trend analysis. The mean and standard deviation of the 8 changepoints for each site are also presented in Table 2. Dissimilarities between the mean changepoints for the four sites are likely due in part to the disparate lengths and data populations of the FP records prior to their changepoints.

Changepoints with high confidence levels were successfully identified in the time series of FP-MLS differences using

15 piecewise linear regression, so this same analysis method was also used to evaluate trends in the differences. Piecewise continuous linear regression fits (i.e., perfectly connected at the changepoint) were employed instead of non-continuous fits because there is no evidence of step jumps in FP-MLS differences at the changepoints. The absence of step jumps is confirmed by the lack of statistically significant (2σ) differences between 1-year averages of FP–MLS before and after the changepoints. The piecewise continuous linear fits included statistical weights (reciprocals of the squared uncertainties of the

20 FP–MLS differences) determined from the combined uncertainties (in quadrature) of the FP and MLS mixing ratios. Each MLS uncertainty was computed as the product of the standard error (σ/√n) of the median MLS mixing ratio and the Student-t value for 95% confidence. FP uncertainties were estimated (95% confidence) as 5% of the FP mixing ratios (see Section 5). Trends for periods 1 and 2 are presented with their uncertainties in Table 3 and Figure 3. Trend uncertainties were computed as the products of the fit slope uncertainties and the Student-t values for 95% confidence. Fits of the Hilo differences were

25 performed using weighted linear regression over the full record period (2010.95-2015.5). The resulting period 2 trends and their uncertainties are included in Table 3 and Figure 3.

For FP–MLS differences computed using MLS v4.2 retrievals the changepoints and confidence levels (Table 4) are very similar to those for v3.3 (Table 2). Mean changepoints for each of the four sites are different by ≤0.3 yr from those

30 calculated in the v3.3 analysis. Many of the trends determined from weighted, piecewise continuous linear regression fits to the FP–MLS v4.2 differences (Table 5) are also very similar to those for the v3.3 retrievals (Table 3).

3 Results for MLS v3.3

In the remainder of this work we report stratospheric averages of trends and changes in FP-MLS differences. These are computed as weighted averages over the 8 pressure levels above each site. Weights are the reciprocals of squared trend uncertainties (95% confidence), yielding uncertainties with 95% confidence. Unless otherwise noted, averages reported in

5   relative units (%) are based on the mean stratospheric water vapor mixing ratio of 4.4 ppmv from 100 to 26 hPa.

Almost all of the period 1 trends in FP–MLS differences over Lindenberg, Boulder and Lauder are positive, but for each of Lindenberg and Lauder these trends are statistically different from zero (95% confidence) at only 1 pressure level (Table 3). For Boulder, period 1 trends at 6 pressure levels are statistically significant, yielding a stratospheric average trend of $0.047 \pm$

10   0.011 ppmv yr^{-1} (Table 6). The period 1 stratosphere-averaged trend over Boulder translates to a mean change of 0.22 ± 0.05 ppmv ($5.0 \pm 1.2\%$) in FP-MLS differences over ~4.6 years (August 2004 to mid-2009). Stratosphere-averaged period 1 changes at Lindenberg and Lauder were smaller, 0.14 ± 0.11 and 0.06 ± 0.08 ppmv, respectively (Table 6). Period 1 trends at 7 of the 8 pressure levels above San José are negative, yielding a stratosphere-averaged change of -0.17 ± 0.06 ppmv ($-3.8 \pm 1.3\%$) in FP-MLS differences from 2005.5 to ~2009.9.

All but 3 of the 24 period 2 trends at Lindenberg, Boulder and Lauder are negative and statistically significant (Table 3, Figure 3). Stratosphere-averaged trends are -0.064 ± 0.016, -0.062 ± 0.009 and -0.052 ± 0.017 ppmv yr^{-1}, respectively (Table 6), demonstrating relatively consistent rates of change (-1.2 to -1.5% yr^{-1}) in the FP–MLS differences. These mean trends translate to stratosphere-averaged changes of -0.25 ppmv (-5.8%), -0.38 ppmv (-8.7%) and -0.25 ppmv (-5.7%) over the

20   period 2 lengths of roughly 4.0, 6.2 and 5.1 years, respectively. This is compelling evidence that FP–MLS differences at these three extra-tropical sites changed significantly during the 4-6 years prior to mid-2015.

All but one of the period 2 trends at Hilo are negative, but none are statistically significant due to the shorter FP measurement record. The stratosphere-averaged trend in FP-MLS differences at Hilo, -0.015 ± 0.019 ppmv yr^{-1}, also lacks

25   statistical significance (95% confidence). Period 2 trends at San José are split between positive and negative, with two of each being statistically different from zero (Table 3). The resulting stratosphere-averaged period 2 trend for San José is small and not statistically different from zero.

Changes in FP–MLS differences over the entire comparison period are calculated by summing the changes for periods 1 and

30   2 at each pressure level. For Lindenberg, Boulder and Lauder the stratosphere-averaged full record changes are -0.11 ± 0.13, -0.16 ± 0.08 and -0.19 ± 0.11 ppmv, respectively (Table 6). Uncertainties in the full-record changes were calculated from the combined (in quadrature) uncertainties of the period 1 and 2 changes at each of the 8 pressure levels, not from the

stratospheric averages in Table 6. Remarkably the stratosphere-averaged full record change of -0.12 ± 0.09 ppmv at San José is similar to those at the other sites despite the period 1 changes at San José being mostly negative.

4 Results for MLS v4.2

Trends in FP–MLS v4.2 differences (Table 5) are, for the most part, very similar to those determined for v3.3 (Table 3).

5 Period 2 trends calculated using v3.3 and v4.2 retrievals (Figure 4) demonstrate that the choice of MLS version makes little difference to the results. An exception is at Hilo where the switch from v3.3 to v4.2 strengthens the negative period 2 trends at 83 and 100 hPa, and intensifies the stratospheric-averaged trend from -0.015 ± 0.019 to -0.025 ± 0.019 ppmv yr^{-1}. Interestingly the choice of MLS retrieval version also makes a significant difference in the period 1 trends at San José, with v3.3 yielding a stronger stratosphere-averaged trend of -0.039 ± 0.013 ppmv yr^{-1} than v4.2 (-0.020 ± 0.012 ppmv yr^{-1}). The

10 choice of MLS version makes very little difference to the stratosphere-average period 2 trends at San José even though v4.2 reduces the number of pressure levels with significant trends from 4 to 2.

5 Discussion

The magnitudes of statistically significant stratosphere-averaged trends in FP–MLS differences (-0.6 to -1.5% yr^{-1}) from ~2010 to mid-2015 are similar in magnitude to the ~1% yr^{-1} average stratospheric water vapor increase reported from FP

15 measurements over Boulder during 1980-2010 (Hurst et al., 2011). Negative trends in FP–MLS differences imply that MLS measurements have biased high, FP measurements have biased low, or some combination of both has occurred over the last 4-6 years. Given these scenarios, an increasing trend in stratospheric water vapor would be exaggerated by MLS measurements that have biased high, and underestimated or undetected by FP measurements that have biased low. For a decreasing water vapor trend the effects of these temporally-changing biases would be reversed.

20

Here we assess the recent changes in FP-MLS differences in relation to the estimated accuracies of stratospheric water vapor measurements by the MLS and FPs. Accuracy estimates for MLS v3.3 and v4.2 retrievals are identical and range from 4 to 8% (0.18 to 0.32 ppmv) over the pressure levels of interest (Livesey et al., 2013; Livesey et al., 2015). Vömel et al. (2007a) assessed the stratospheric measurement uncertainties of the CFH and estimated the accuracy to be <10% (<0.5 ppmv), but a

25 recent reassessment lowers the uncertainty estimate (1σ) to <5% (Vomel et al., 2016). A recent evaluation of the NOAA FPH (Hall et al., 2016) demonstrates that the stratospheric measurement uncertainties (2σ) are <6% (<0.3 ppmv). Employing 3% and 5% as 1σ and 2σ accuracy estimates for the FPs, the combined (in quadrature) accuracy estimates of FP and MLS measurements of stratospheric water vapor at the 8 retrieval pressures range from 5.0 to 8.5% (0.23 to 0.34 ppmv) and 6.4 to 9.4% (0.29 to 0.40 ppmv), respectively. From here forward the combined accuracy estimates for FPs and MLS based on FP

30 measurement uncertainties of 3% and 5% are denoted ACC$_{FP3}$ and ACC$_{FP5}$, respectively.

Figure 5 displays the values of FP–MLS differences (v3.3) at the start of each record, at the changepoint and at the end of each record for the 8 pressure levels, as determined by the piecewise linear fits described above. By the end of the comparison period in mid-2015, 18 of the 40 differences exceeded the ACC_{FP3} and another 5 were within 0.05 ppmv of the

5   ACC_{FP3}. Endpoint differences surpassed the more conservative ACC_{FP5} estimates for 11 site/pressure level combinations and another 5 were within 0.05 ppmv of the ACC_{FP5}. Six of the endpoint differences exceeding the ACC_{FP5} were at 100 and 83 hPa, pressure levels for which FP-MLS biases of up to 10% have already been reported (Hurst et al., 2014).

By mid-2015 the FP–MLS differences at 7 pressure levels over Lindenberg exceeded the ACC_{FP5} (Figure 5). However, the

10   starting point differences for 4 of these 7 levels also exceeded or nearly exceeded the ACC_{FP5} (Figure 5), indicating that the large differences in mid-2015 resulted from the continuation of long-term biases rather than recent drifts. At the other 3 pressure levels over Lindenberg the endpoint differences exceeded ACC_{FP5} because of large decreases in FP-MLS differences during period 2. At Boulder, 6 and 4 endpoint differences exceeded or were within 0.05 ppmv of the ACC_{FP3} and ACC_{FP5}, respectively, with all but one (100 hPa) caused by strong negative period 2 trends. At Lauder and San José, 1 and 3

15   endpoint differences exceeded or were within 0.05 ppmv of the ACC_{FP5}, respectively, all of which resulted from strong declines. At Hilo the starting point and endpoint differences at 100 and 83 hPa exceeded or were within 0.05 ppmv of the ACC_{FP5}, consistent with the long-term biases already reported for these pressure levels (Hurst et al., 2014).

Very similar results were obtained when MLS v4.2 retrievals were employed (not shown). By mid-2015, 44% and 25% of

20   the FP–MLS differences (both MLS versions) exceeded the ACC_{FP3} and ACC_{FP5}, respectively. Likewise, 57% and 40% of the endpoint differences exceeded or were within 0.05 ppmv of the ACC_{FP3} and ACC_{FP5} estimates, respectively. If the recent divergences between FPs and MLS continue they will inevitably push FP-MLS differences at most pressure levels to exceed the combined accuracy estimates of the two instruments.

25   The causes of the recent divergences in stratospheric water vapor measurements by FPs and MLS at Lindenberg, Boulder and Lauder are currently unknown. It is intriguing that the period 2 trends at the three extratropical sites are similar to one another but disparate from those at tropical San José. The differences at Hilo have also drifted downward since late 2010, but the FPH record is too short to permit the detection of statistically significant trends. We deliberately compared MLS retrievals with 5 different records of in situ, balloon-borne measurements compiled using two independent FPs with different

30   manufacturers, calibration, frost control parameters and data processing. Our finding of similar divergences (not step changes) in FP and MLS measurements over the three extratropical sites suggests a positive drift in MLS retrievals over these locations, primarily because it is highly unlikely that the two different types of FPs are drifting at similar rates at the 3 sites. We plan to continue closely comparing MLS and FP measurements over these 5 sites to ascertain if they continue to diverge, settle into a stable bias, or start to re-converge.

6 Conclusions

The recent divergences in stratospheric water vapor measurements by FPs and MLS should prompt serious discussions about our future capabilities to monitor UTLS water vapor around the globe. Currently there is no comprehensive, long-term plan for a monitoring program that even approaches the 3500 near-global profiles per day by MLS (Müller et al., 2016). A third

5 generation Stratospheric Aerosol and Gas Experiment (SAGE III) spectrometer is ready to be deployed in late 2016 on the International Space Station where it will provide an average of 32 vertical profiles of UTLS water vapor each day. Ultimately, when Aura MLS fails, there will be an immediate 99% reduction in the spatiotemporal density of measurements because there is no plan to replace MLS with a satellite sensor of similar capabilities. For this reason Müller et al. (2016) have proposed the creation of a large network of FPs covering the globe and funded in a committed way that would make the

10 network sustainable for many decades. Towards this goal, a network of 20-30 globally-distributed FP sounding sites is in development as part of the Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN, Bodeker et al., 2016). However, even with a FP network of 100 sites performing weekly soundings the spatiotemporal density of UTLS water vapor measurements would be only 0.4% of what MLS is currently providing.

Acknowledgements

15 The NOAA frost point hygrometer network is supported in part by NOAA's Climate Program Office, the US Global Climate Observing System Program, and NASA's Upper Atmosphere Research Program. The FPH soundings used in this study were carefully conducted at Hilo by David Nardini and Darryl Kuniyuki, and at Lauder by Hamish Chisholm, Alan Thomas, Wills Dobson, and Richard Querel. Karen Rosenlof and Sean Davis's participation in this study was supported by NOAA resources targeted for water vapor research in the upper troposphere.

20 **References**

Bodeker, G. E., Bojinski, S., Cimini, D., Dirksen, R. J., Haeffelin, M., Hannigan, J. W., Hurst, D. F., Leblanc, T., Madonna, F., Maturilli, M., Mikalsen, A. C., Philipona, R., Reale, T., Seidel, D. J., Tan, D. G. H., Thorne, P. W., Vömel, H., and Wang, J.: Reference Upper-Air Observations for Climate: From Concept to Reality, B. Am. Meteorol. Soc., 97, 123–135, doi:10.1175/BAMS-D-14- 00072.1, 2016.

25 Davis, S. M., Rosenlof, K. H., Hassler, B., Hurst, D. F., Read, W. G., Vömel, H., Selkirk, H. B., and Fujiwara, M.: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: A long-term database for climate studies, Earth Syst. Sci. Data, Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-16, 2016.

Forster, P. M. de F. and Shine, K. P.: Assessing the climate impact of trends in stratospheric water vapor, Geophys. Res. Lett., 29, 1086, doi:10.1029/2001GL013909, 2002.

Froidevaux, L., Anderson, J., Wang, H. -J., Fuller, R. A., Schwartz, M. J., Santee, M. L., Livesey, N. J., Pumphrey, H. C., Bernath, P. F., Russell III, J. M., and McCormick, M. P.: Global OZone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS): methodology and sample results with a focus on HCl, H2O, and O3, Atmos. Chem. Phys., 15, 10471-10507, doi:10.5194/acp-15-10471-2015, 2015.

5   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 frostpoint hygrometer, Atmos. Meas. Tech. Discuss., Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-160, 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

10   stratospheric water vapor trends derived from merged satellite data, Nature Geoscience, 7, 1-9, doi:10.1038/NGEO2236, 2014.

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, J. Geophys. Res., 116, D02306, doi:10.1029/2010JD015065, 2011.

15   Hurst, D. F., Lambert, A., Read, W. G., Davis, S. M., Rosenlof, K. H., Hall, E. G., Jordan, A. F., and Oltmans, S. J.: Validation of Aura Microwave Limb Sounder stratospheric water vapor measurements by the NOAA frost point hygrometer, J. Geophys. Res. Atmos., 119, doi:10.1002/2013JD020757, 2014.

Lambert, A., et al.: Validation of the Aura Microwave Limb Sounder middle atmosphere water vapor and nitrous oxide measurements, J. Geophys. Res., 112, D24S36, doi:10.1029/2007JD008724, 2007.

20   Livesey, N. J., Read, W. G., Froidevaux, L., Lambert, A., Manney, G. L., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Cofield, R. E., Cuddy, D. T., Fuller, R. A., Jarnot, R. F., Jiang, J. H., Knosp, B. W., Stek, P. C., Wagner, P. A., and Wu, D. L.: Version 3.3 and 3.4 Level 2 data quality and description document, Tech. Rep. JPL D-33509, Jet Propulsion Laboratory, available at: http: //mls.jpl.nasa.gov/data/v3_data_quality_document.pdf, 2013..

Livesey, N. J., Read, W. G., Wagner, P. A., Froidevaux, L., Lambert, A., Manney, G. L., Millán Valle, L. F., Pumphrey, H.

25   C., Santee, M. L., Schwartz, M. J., Wang, S., Fuller, R. A., Jarnot, R. F., Knosp, B. W., and Martinez, E.: Version 4.2x Level 2 data quality and description document, Tech. Rep. JPL D-33509 Rev. A, Jet Propulsion Laboratory, available at: http: //mls.jpl.nasa.gov/data/v4-2_data_quality_document.pdf, 2015.

Lund, R. and Reeves, J.: Detection of undocumented changepoints: A revision of the two-phase regression model, J. Climate, 15, 2547-2554, 2002.

30   Mastenbrook, H. J. and Oltmans, S. J.: Stratospheric water vapor variability for Washington, D.C./Boulder, CO; 1964-82, *J. Atmos. Sci.*, *40*, 2157-2165, 1983.

Müller, R., Kunz, A., Hurst, D. F., Rolf, C., Krämer, M., and Riese, M.: The need for accurate long-term measurements of water vapor in the upper troposphere and lower stratosphere with global coverage. Earth's Future, 4, doi:10.1002/2015EF000321, 2016.

Oltmans, S. J. and Hofmann, D. J., Increase in lower-stratospheric water vapor at a midlatitude northern hemisphere site from 1981 to 1994, Nature, 374, 146-149, 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, Geophys. Res. Lett.,
5   27(21), 3453-3456, 2000.

Read, W. G., et al.: Aura Microwave Limb Sounder upper tropospheric and lower stratospheric H_2O and relative humidity with respect to ice validation, J. Geophys. Res., 112 (D24), D24S35, doi.10.1029/2007JD008752, 2007.

Rosenlof, K. H., et al.: Stratospheric water vapor increases over the past half-century, Geophys. Res. Lett., 28(7), 1195-1198, 2001.

10   Scherer, M., Vömel, H., Fueglistaler, S., Oltmans, S. J., and Staehelin, J.: Trends and variability of midlatitude stratospheric water vapor deduced from the re-evaluated Boulder balloon series and HALOE, Atmos. Chem. Phys., 8, 1391-1402, 2008.

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

Vömel, H., et al.: Validation of Aura MLS water vapor by balloon-borne Cryogenic Frostpoint Hygrometer measurements, *J.*
15   *Geophys. Res.*, *112*, D24S37, doi:10.1029/2007JD008698, 2007a.

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, J. Geophys. Res., 112, doi:10.1029/2006JD007224, 2007b.

Vömel, H., Naebert, T., Dirksen, R., and Sommer, M.: An update on the uncertainties of water vapor measurements using
20   Cryogenic Frostpoint Hygrometers, Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-44, 2016.

Table 1. Frost Point Hygrometer Site Information and Coincident MLS Profile Statistics

Site	Site Code	Altitude (masl)	Latitude (°N)	Longitude (°E)	Comparison Start Date[a]	FP Type	FP Profiles[b]	MLS Profiles v3.3[c]	MLS Profiles v4.2[d]
Lindenberg	LIN	112	52.21	14.12	2006.66	CFH	132	801	787
Boulder	BLD	1743	39.95	-105.20	2004.67	FPH/CFH	144/31	1000	990
Hilo	HIH	10	19.72	-155.05	2010.94	FPH	51	275	276
San José	SJC	1075	9.98	-84.13	2005.52	CFH	158	788	858
Lauder	LDR	370	-45.04	169.68	2004.65	FPH	117	463	458

[a] Decimal date of first FP profile after the 2004.59 start of MLS data reporting.

[b] Number of FP profiles with at least one coincident MLS profile.

[c] Number of MLS version 3.3 profiles coincident with the FP profiles.

5   [d] Number of MLS version 4.2 profiles coincident with the FP profiles.

Table 2. Changepoint Dates and Their Confidence Levels, MLS Version 3.3

	Lindenberg		Boulder		San José		Lauder	
Pressure (hPa)	Changepoint (year)	CL[a] (%)	Changepoint (year)	CL (%)	Changepoint (year)	CL (%)	Changepoint (year)	CL (%)
26	2011.4	83	2009.1	>99	2008.9	66	2010.4	>99
32	2011.1	>99	2009.0	>99	2010.8	45	2010.7	99
38	2011.3	>99	2009.5	>99	2009.3	48	2010.7	>99
46	2011.6	92	2009.5	>99	2008.9	71	2010.6	>99
56	2011.9	94	2009.3	>99	2010.1	>99	2010.6	>99
68	2011.4	99	2009.5	>99	2010.7	>99	2010.4	>99
83	2011.4	80	2009.3	>99	2010.6	>99	2010.7	93
100	2011.7	40	2009.1	>99	2009.1	73	2011.1	94
Mean[b]	2011.5		2009.3		2009.8		2010.6	
StdDev	0.3		0.2		0.9		0.2	

[a] Confidence levels for the listed changepoint dates, with >99 indicating a value between 99 and 100%.

10   [b] Mean and standard deviation of the changepoint dates for each site.

Table 3. Linear Regression Slopes of FP-MLS v3.3 Differences

MLS Pressure	Period	Lindenberg (ppmv yr^{-1})	Boulder (ppmv yr^{-1})	Hilo (ppmv yr^{-1})	San José (ppmv yr^{-1})	Lauder (ppmv yr^{-1})
26	1	-0.001 ± 0.089	**0.074 ± 0.040**		-0.054 ± 0.066	**0.066 ± 0.051**
32	1	0.026 ± 0.089	**0.089 ± 0.040**		-0.004 ± 0.038	0.032 ± 0.043
38	1	**0.073 ± 0.071**	**0.057 ± 0.032**		0.003 ± 0.048	0.027 ± 0.040
46	1	0.044 ± 0.061	**0.038 ± 0.031**		**-0.047 ± 0.046**	0.019 ± 0.037
56	1	0.014 ± 0.053	0.021 ± 0.030		**-0.033 ± 0.032**	0.020 ± 0.036
68	1	0.024 ± 0.060	0.014 ± 0.027		**-0.036 ± 0.029**	0.014 ± 0.035
83	1	0.026 ± 0.056	**0.039 ± 0.027**		**-0.065 ± 0.027**	-0.024 ± 0.030
100	1	0.022 ± 0.058	**0.089 ± 0.031**		**-0.053 ± 0.046**	-0.013 ± 0.028
26	2	**-0.058 ± 0.053**	**-0.054 ± 0.029**	-0.028 ± 0.060	-0.009 ± 0.036	**-0.055 ± 0.062**
32	2	**-0.079 ± 0.048**	**-0.065 ± 0.027**	-0.036 ± 0.055	**-0.053 ± 0.051**	-0.029 ± 0.059
38	2	**-0.077 ± 0.045**	**-0.070 ± 0.028**	-0.019 ± 0.055	**-0.033 ± 0.031**	**-0.050 ± 0.050**
46	2	**-0.057 ± 0.045**	**-0.065 ± 0.028**	0.001 ± 0.052	0.002 ± 0.026	**-0.058 ± 0.045**
56	2	**-0.080 ± 0.049**	**-0.054 ± 0.025**	-0.018 ± 0.053	0.007 ± 0.035	**-0.069 ± 0.045**
68	2	**-0.081 ± 0.038**	**-0.056 ± 0.024**	-0.030 ± 0.053	0.011 ± 0.042	**-0.051 ± 0.039**
83	2	**-0.054 ± 0.038**	**-0.065 ± 0.023**	-0.020 ± 0.049	**0.059 ± 0.036**	-0.036 ± 0.042
100	2	-0.027 ± 0.044	**-0.066 ± 0.024**	0.023 ± 0.052	**0.041 ± 0.031**	**-0.058 ± 0.047**

Slopes are presented with their 95% confidence intervals. Periods 1 and 2 refer to the intervals before and including the changepoint (Table 2) and immediately after the changepoint to June 30, 2015, respectively. Values in boldface type are statistically different from zero with 95% confidence.

Table 4. Changepoint Dates and Their Confidence Levels, MLS Version 4.2

Pressure (hPa)	Lindenberg Changepoint (year)	CL[a] (%)	Boulder Changepoint (year)	CL (%)	San José Changepoint (year)	CL (%)	Lauder Changepoint (year)	CL (%)
26	2011.4	96	2009.1	>99	2008.9	69	2010.6	>99
32	2011.4	99	2009.0	>99	2010.8	39	2010.7	>99
38	2011.3	>99	2010.0	>99	2010.8	76	2010.7	>99
46	2011.7	92	2009.6	>99	2009.5	86	2010.6	>99
56	2011.6	87	2009.3	>99	2010.1	96	2010.6	>99
68	2011.4	>99	2009.5	>99	2010.6	>99	2011.1	>99
83	2011.4	83	2010.2	>99	2010.9	>99	2010.7	97
100	2011.4	73	2009.1	>99	2009.1	78	2010.7	>99
Mean[b]	2011.4		2009.5		2010.1		2010.7	
StdDev	0.1		0.4		0.8		0.2	

[a] Confidence levels for the listed changepoint dates, with >99 indicating a value between 99 and 100%.

[b] Mean and standard deviation of the listed changepoint dates for each site.

Table 5. Linear Regression Slopes of FP-MLS v4.2 Differences

MLS Pressure	Period	Lindenberg (ppmv yr⁻¹)	Boulder (ppmv yr⁻¹)	Hilo (ppmv yr⁻¹)	San José (ppmv yr⁻¹)	Lauder (ppmv yr⁻¹)
26	1	0.007 ± 0.089	**0.077 ± 0.041**		-0.054 ± 0.065	**0.053 ± 0.047**
32	1	0.026 ± 0.082	**0.081 ± 0.040**		-0.015 ± 0.038	0.026 ± 0.043
38	1	**0.076 ± 0.071**	**0.052 ± 0.029**		0.009 ± 0.034	0.033 ± 0.040
46	1	0.046 ± 0.061	**0.038 ± 0.030**		-0.013 ± 0.039	0.011 ± 0.038
56	1	0.023 ± 0.059	0.022 ± 0.030		-0.026 ± 0.033	0.017 ± 0.035
68	1	0.032 ± 0.059	0.018 ± 0.027		-0.019 ± 0.031	0.005 ± 0.029
83	1	0.030 ± 0.058	**0.029 ± 0.028**		**-0.028 ± 0.025**	-0.017 ± 0.030
100	1	0.064 ± 0.066	**0.079 ± 0.030**		-0.037 ± 0.047	-0.023 ± 0.030
26	2	**-0.062 ± 0.053**	**-0.056 ± 0.030**	-0.015 ± 0.070	-0.011 ± 0.035	-0.044 ± 0.063
32	2	**-0.079 ± 0.051**	**-0.064 ± 0.027**	-0.032 ± 0.057	-0.029 ± 0.051	-0.024 ± 0.060
38	2	**-0.071 ± 0.045**	**-0.082 ± 0.031**	-0.014 ± 0.055	**-0.055 ± 0.047**	**-0.059 ± 0.050**
46	2	**-0.054 ± 0.046**	**-0.063 ± 0.029**	-0.005 ± 0.052	-0.007 ± 0.031	**-0.060 ± 0.045**
56	2	**-0.065 ± 0.044**	**-0.055 ± 0.026**	-0.030 ± 0.055	-0.002 ± 0.036	**-0.075 ± 0.043**
68	2	**-0.079 ± 0.039**	**-0.054 ± 0.024**	-0.038 ± 0.053	-0.013 ± 0.042	**-0.064 ± 0.047**
83	2	**-0.058 ± 0.038**	**-0.083 ± 0.028**	**-0.059 ± 0.052**	0.043 ± 0.041	-0.025 ± 0.045
100	2	-0.037 ± 0.039	**-0.077 ± 0.023**	-0.001 ± 0.050	0.022 ± 0.031	**-0.064 ± 0.041**

Slopes are presented with their 95% confidence intervals. Periods 1 and 2 refer to the intervals before and including the changepoint (Table 4) and immediately after the changepoint to June 30, 2015, respectively. Values in boldface type are
5   statistically different from zero with 95% confidence.

Table 6. Stratospheric Average Trends and Changes in FP–MLS Differences

Site	MLS Version	Period 1 Trend (ppmv yr⁻¹)	Period 1 Change (ppmv)	Period 2 Trend (ppmv yr⁻¹)	Period 2 Change (ppmv)	Full Record Change (ppmv)
Lindenberg	3.3	**0.029 ± 0.023**	**0.14 ± 0.11**	**-0.064 ± 0.016**	**-0.25 ± 0.06**	-0.11 ± 0.13
Lindenberg	4.2	**0.039 ± 0.023**	**0.19 ± 0.11**	**-0.062 ± 0.015**	**-0.25 ± 0.06**	-0.07 ± 0.13
Boulder	3.3	**0.047 ± 0.011**	**0.22 ± 0.05**	**-0.062 ± 0.009**	**-0.38 ± 0.06**	**-0.16 ± 0.08**
Boulder	4.2	**0.044 ± 0.011**	**0.22 ± 0.05**	**-0.066 ± 0.010**	**-0.40 ± 0.06**	**-0.18 ± 0.08**
Hilo	3.3			-0.015 ± 0.019	-0.07 ± 0.09	
Hilo	4.2			**-0.025 ± 0.019**	**-0.11 ± 0.09**	
San José	3.3	**-0.039 ± 0.013**	**-0.17 ± 0.06**	0.006 ± 0.012	0.04 ± 0.07	**-0.12 ± 0.09**
San José	4.2	**-0.020 ± 0.012**	**-0.10 ± 0.06**	-0.002 ± 0.013	-0.02 ± 0.07	**-0.11 ± 0.09**
Lauder	3.3	0.009 ± 0.013	0.06 ± 0.08	**-0.052 ± 0.017**	**-0.25 ± 0.08**	**-0.19 ± 0.11**
Lauder	4.2	0.006 ± 0.012	0.04 ± 0.07	**-0.054 ± 0.017**	**-0.26 ± 0.08**	**-0.21 ± 0.11**

Weighted averages of trends and changes in FP–MLS differences at all 8 pressure levels (100-26 hPa) over each site. Stratospheric averages are presented with their 95% confidence limits. All values were computed using the regression slopes
10   and their uncertainties in Tables 3 and 5. Values in boldface type are significantly different from zero with 95% confidence.
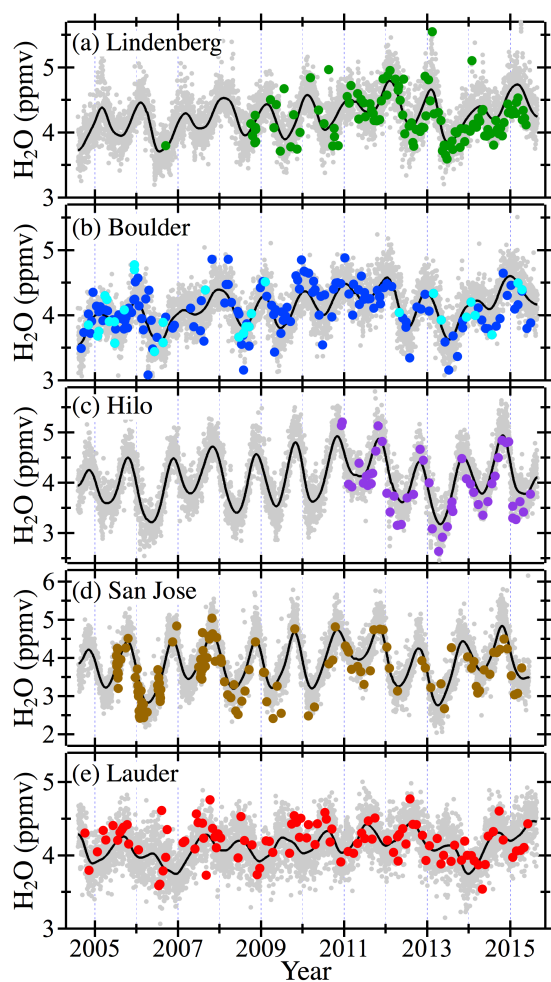
Figure 1: Daily average MLS version 3.3 overpass retrievals (gray markers, smoothed black curves) and in situ frost point hygrometer (FP) data at 68 hPa for individual soundings at each site (filled circles). Data from two types of FPs are shown: NOAA FPH at Boulder (dark blue), Hilo and Lauder, and CFH at Lindenberg, Boulder (cyan) and San Jose. Note the emerging biases between FP and MLS mixing ratios at all five sites towards the ends of their records.
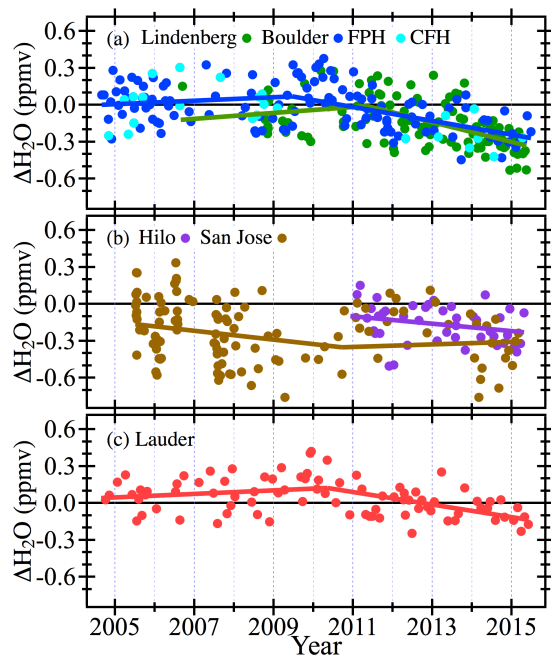
Figure 2: Differences between FP mixing ratios and spatiotemporally coincident MLS v3.3 water vapor retrievals at 68 hPa over the five FP sounding sites. In the top panel (a) dark blue and cyan markers for Boulder depict soundings made with the NOAA FPH and the CFH, respectively. Lines show the trends in FP-MLS differences in two distinct periods separated by a changepoint, except for Hilo where the shorter FPH records show no indications of statistically significant changepoints.
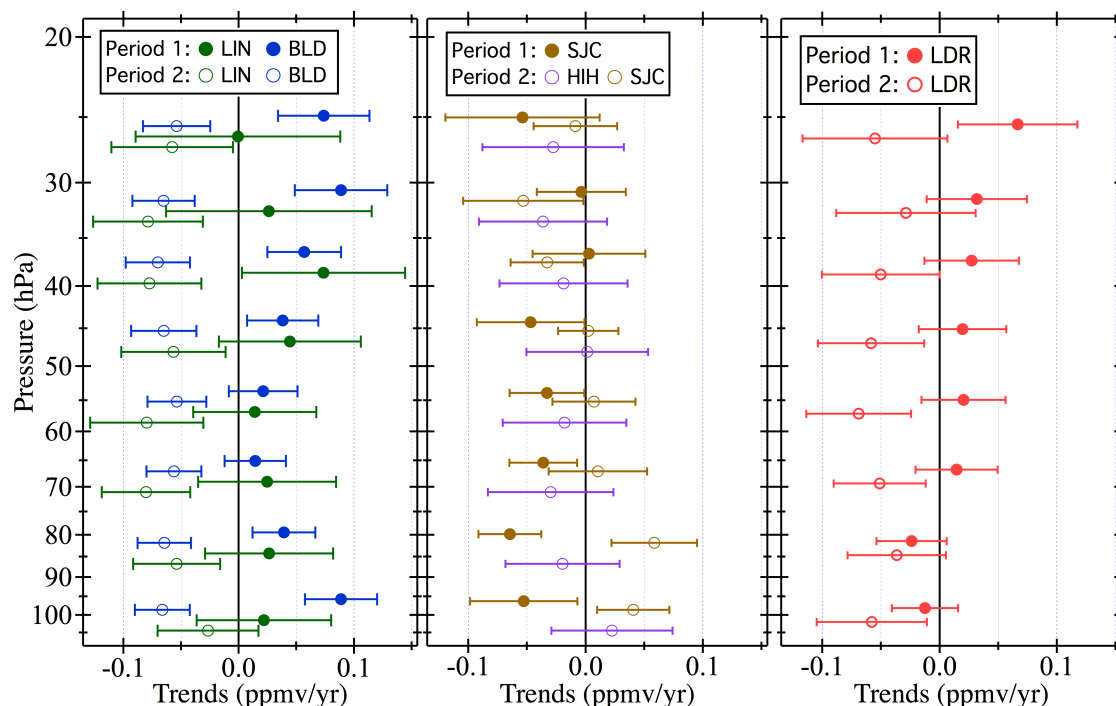
Figure 3. Trends in FP-MLS differences for the pre- and post-changepoint periods at 8 stratospheric pressure levels (100-26 hPa) over the five FP sounding sites. Markers for each pressure level are slightly offset in pressure for clarity. Horizontal error bars depict the 95% confidence intervals of the trends. Only period 2 trends are shown for Hilo because the shorter records show no indications of statistically significant changepoints.
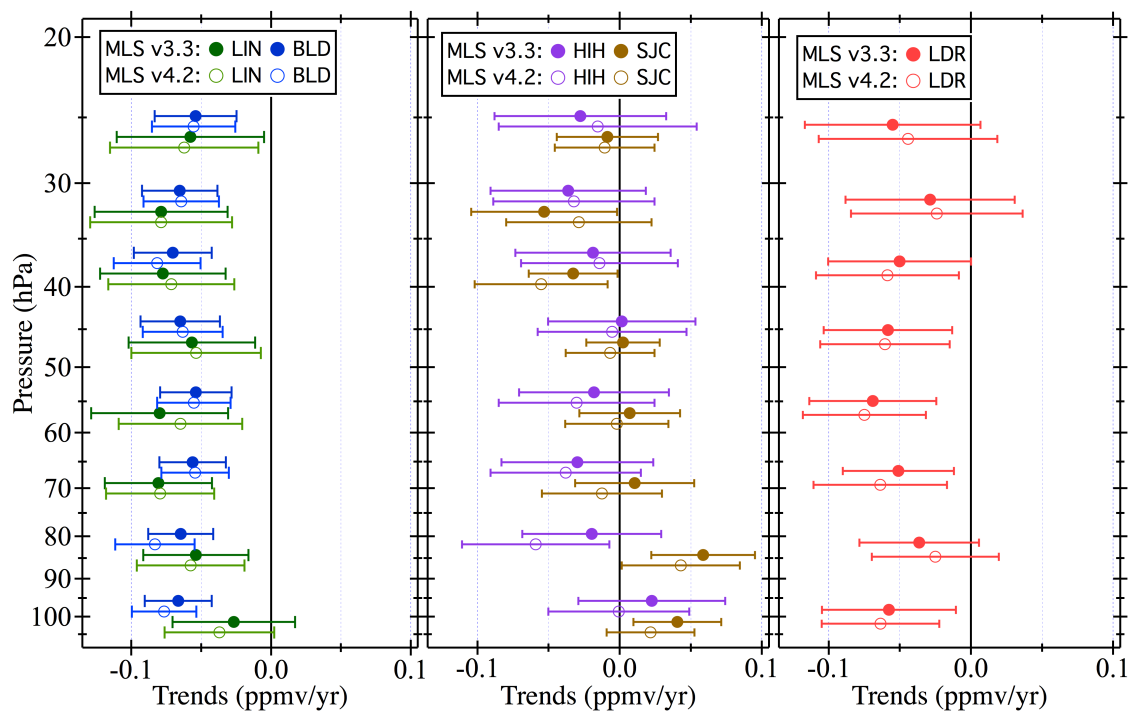
Figure 4. Period 2 trends in FP–MLS differences using MLS v3.3 (filled circles) and v4.2 (open circles) retrievals at 8 stratospheric pressure levels (100-26 hPa) over the five FP sounding sites. Markers for each pressure level are slightly offset in pressure for clarity. Horizontal error bars depict the 95% confidence intervals of the trends.
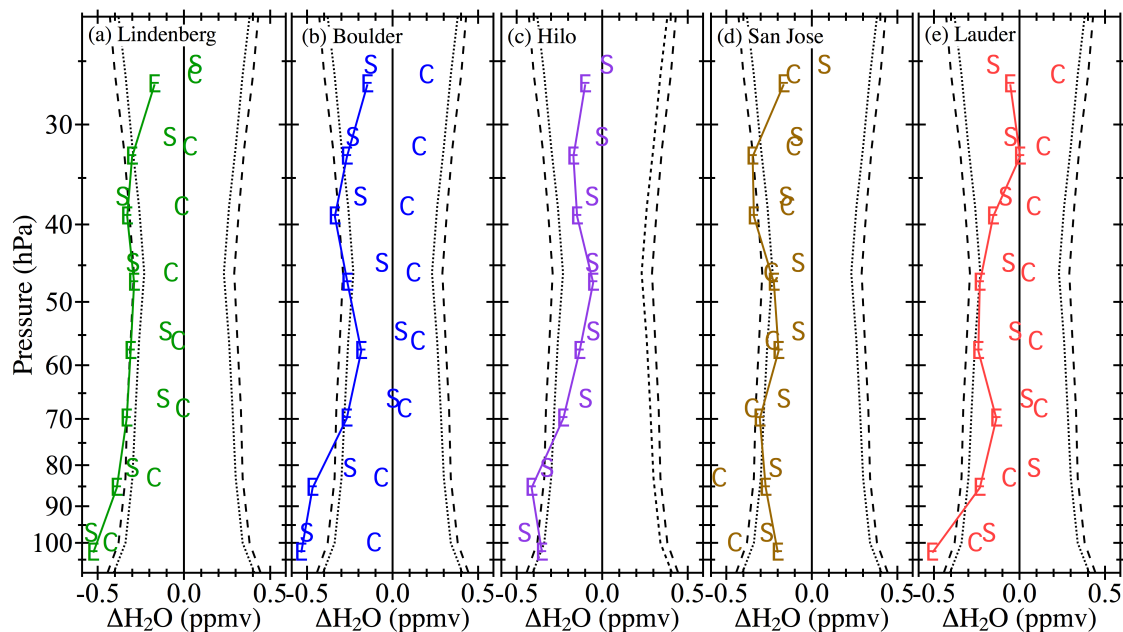
Figure 5. FP–MLS v3.3 differences at the starting points (S), changepoints (C) and ending points (E) of the time series as defined by piecewise continuous linear fits. Colored vertical curves join the ending points to serve as visual guides. Black vertical curves depict the combined accuracy estimates for FP and MLS measurements of stratospheric water vapor based on FP accuracy values of 3% (dashed) and 5% (dotted). Note that many of the ending point values (E) lie near or outside the combined accuracy estimates. For Hilo only the starting and ending point differencess are presented because no significant changepoints were detected in the shorter records.