Published in final edited form as:
J Geophys Res Atmos. 2014 February 27; 119(4): 1915–1935. doi:10.1002/2013JD020817.

Evaluation of UT/LS hygrometer accuracy by intercomparison during the NASA MACPEX mission

A. W. Rollins[1,2], **T. D. Thornberry**[1,2], **R. S. Gao**[1], **J. B. Smith**[3], **D. S. Sayres**[3], **M. R. Sargent**[3], **C. Schiller**[4,5], **M. Krämer**[4], **N. Spelten**[4], **D. F. Hurst**[2,6], **A. F. Jordan**[2,6], **E. G. Hall**[2,6], **H. Vömel**[7], **G. S. Diskin**[8], **J. R. Podolske**[9], **L. E. Christensen**[10], **K. H. Rosenlof**[1], **E. J. Jensen**[9], and **D. W. Fahey**[1,2]

[1]Chemical Sciences Division, NOAA Earth System Research Laboratory, Boulder, Colorado, USA

[2]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado, USA

[3]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, USA

[4]IEK-7, Forschungszentrum Jülich GmbH, Jülich, Germany

[6]Global Monitoring Division, NOAA Earth System Research Laboratory, Boulder, Colorado, USA

[7]GRUAN/Deutscher Wetterdienst, Lindenberg, Germany

[8]NASA Langley Research Center, Hampton, Virginia, USA

[9]NASA Ames Research Center, Moffett Field, California, USA

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

Abstract

Acquiring accurate measurements of water vapor at the low mixing ratios (< 10 ppm) encountered in the upper troposphere and lower stratosphere (UT/LS) has proven to be a significant analytical challenge evidenced by persistent disagreements between high-precision hygrometers. These disagreements have caused uncertainties in the description of the physical processes controlling dehydration of air in the tropical tropopause layer and entry of water into the stratosphere and have hindered validation of satellite water vapor retrievals. A 2011 airborne intercomparison of a large group of in situ hygrometers onboard the NASA WB-57F high-altitude research aircraft and balloons has provided an excellent opportunity to evaluate progress in the scientific community toward improved measurement agreement. In this work we intercompare the measurements from the Midlatitude Airborne Cirrus Properties Experiment (MACPEX) and discuss the quality of agreement. Differences between values reported by the instruments were reduced in comparison to some prior campaigns but were nonnegligible and on the order of 20% (0.8 ppm). Our analysis suggests that unrecognized errors in the quantification of instrumental background for some or all of the hygrometers are a likely cause. Until these errors are understood, differences at this level

Correspondence to: A. W. Rollins, andrew.rollins@noaa.gov.
[5]Deceased 3 March 2012

will continue to somewhat limit our understanding of cirrus microphysical processes and dehydration in the tropical tropopause layer.

1. Introduction

Water vapor (WV) in Earth's upper troposphere and lower stratosphere (UT/LS) exerts a significant control on the planet's climate through multiple mechanisms. First, transport of WV through the tropical tropopause layer (TTL) controls the concentration of stratospheric WV, which radiatively impacts surface temperatures [*Harries et al.*, 2008; *Solomon et al.*, 2010]. Second, deposition of WV onto particles in the TTL forms cirrus clouds. These clouds increase Earth's albedo relative to a clear sky and also serve to trap outgoing longwave radiation, with the net result being a cooling of Earth's surface and warming at the top of the atmosphere [*Lee et al.*, 2009]. Finally, WV influences stratospheric ozone chemistry. Both the gas phase reaction of water with electronically excited atomic oxygen and heterogeneous reactions on ice surfaces result in sinks for ozone. Thus, an increase in stratospheric WV has the potential to slow stratospheric ozone recovery [*Kirk-Davidoff et al.*, 1999; *Feck et al.*, 2008; *Vogel et al.*, 2011].

Measurements of UT/LS WV with high accuracy and long-term stability are crucial for establishing the magnitude and trends in the contributions of these various processes in the Earth system. Of particular importance in terms of absolute measurement accuracy is the ability to understand cirrus microphysics, which hinges on highly accurate and precise measurements of relative humidity with respect to ice (RH_i) both inside and outside of clouds [*Peter et al.*, 2006]. For example, RH_i accuracy on the order of 10% is necessary to distinguish between hexagonal and cubic ice phases [*Shilling et al.*, 2006] or to provide a satisfactory constraint on the dehydration efficiency of cirrus clouds. In situ measurements of WV and temperature are required to perform the necessary microphysical process studies on relatively small spatial and temporal scales that can provide and/or confirm the description of the processes that control dehydration in the TTL. Intercomparisons with these measurements also contribute significantly to confidence in the accuracy and stability of global data sets from satellite-borne sensors.

Given its relatively high mixing ratio (> 1 ppm (ppm = μmol/mol)) compared to most trace gases in the atmosphere, the challenge of measuring WV in the UT/LS is somewhat surprising. Coincident and near-coincident measurements in the TTL have often shown discrepancies greater than 10% and have at times exceeded 100% [*Albritton and Zander*, 1985; *Kley et al.*, 2000; *Jensen et al.*, 2005; *Read et al.*, 2007; *Vömel et al.*, 2007a; *Weinstock et al.*, 2009]. Discrepancies are not limited to in situ measurements, as various satellite instruments have well documented offsets relative to each other (e.g., Halogen Occultation Experiment/Microwave Limb Sounder/Stratospheric Aerosol and Gas Experiment II [*Kley et al.*, 2000; *Thomason et al.*, 2010]). During a series of NASA airborne in situ measurements in 2002–2006, extremely high (> 200%) RH_i values were observed in the TTL that could not be explained by conventional microphysical models [see *Peter et al.*, 2006]. These observations have raised questions about measurement accuracy and inspired careful investigations to understand WV measurement uncertainties.

In situ instruments that have contributed substantially to UT/LS WV measurement include balloon-borne frost point (FP) hygrometers (e.g., the cryogenic frost point hygrometer (CFH) [*Vömel et al.*, 2007a] and the National Oceanic and Atmospheric Administration (NOAA) frost point hygrometer (FPH) [*Hurst et al.*, 2011]); Lyman-α photofragmentation fluorescence hygrometers which include the Harvard Water Vapor instrument (HWV) [*Weinstock et al.*, 2009], the Fast In Situ Stratospheric Hygrometer (FISH) [*Zöger et al.*, 1999], and the Fluorescent Airborne Stratospheric Hygrometer (FLASH) [*Sitnikov et al.*, 2007]; and a number of tunable diode laser (TDL) based infrared absorption spectrometers such as the Jet Propulsion Laboratory Laser Hygrometer (JLH) [*May*, 1998], the Aircraft Laser Infrared Absorption Spectrometer (ALIAS) [*Webster et al.*, 1994], and the Integrated Cavity Output Spectrometer (ICOS) [*Sayres et al.*, 2009]. Most of the measurements have reported accuracy estimates of ±5–10%, which is inconsistent with the observed measurement differences that can exceed 50% at the sub-10 ppm level.

Water vapor measurements that have been particularly difficult to understand are those yielding RH_i values significantly greater than what can be explained given the current understanding of ice nucleation and ice/vapor partitioning in the atmosphere. These measurements have occurred both inside and outside of clouds. During the Cirrus Regional Study of Tropical Anvils and Cirrus Layers-Florida Area Cirrus Experiment (CRYSTAL-FACE) mission, JLH and HWV observed sustained RH_i of ~130% in dense cirrus where gas/particle equilibrium of water should be rapidly achieved, and RH_i near 100% is therefore expected [*Gao et al.*, 2004; *Jensen and Pfister*, 2005]. A separate problem has been the observations by at least JLH and HWV of clear-sky supersaturations that exceed the homogeneous nucleation threshold for ice (e.g., RH_i exceeding ~160% [*Koop et al.*, 2000]) during the Pre-Aura Validation Experiment (Pre-AVE) mission [*Jensen and Pfister*, 2005; *Jensen et al.*, 2005] and the Costa Rica Aura Validation Experiment (CR-AVE) mission [*Jensen et al.*, 2008]. These RH_i values exceeding ~160% have occurred exclusively at low temperatures ($T < 200$K) in the TTL. In contrast, an extensive data set of RH_i compiled by *Krämer et al.* [2009] and screened for measurement consistency between the FISH, FLASH, and Open path Jülich Stratospheric Tdl Experiment (OJSTER) (TDL) instruments exhibits very little anomalously high RH_i in the UT/LS.

Uncertainties in a single measurement point are due both to instrumental precision and accuracy limitations. Precision results in noise that may be random (e.g., photon counting) or systematic (e.g., frost point control algorithm), but should not lead to a systematic bias in the data, and uncertainties resulting from precision limitations can usually be significantly reduced by averaging a series of successive points. Accuracy-related uncertainties, however, are a result of intrinsic limitations in the knowledge of the response of the instrument to water vapor. Accuracy limitations may be due to fixed uncertainties in certain parameters (e.g., spectral line parameters) or due to uncertainties related to the instrument itself (e.g., potential biases in temperature or pressure measurements or uncertainties in the standards used to calibrate the instrument response). The data from the NASA Midlatitude Airborne Cirrus Properties Experiment (MACPEX) and elsewhere have demonstrated that precision is not the primary factor contributing to intercomparison discrepancies and limitations in the data for scientific work. Therefore, this work focuses on the accuracy limitations, which are

evaluated by comparing large numbers of data points such that noise is sufficiently averaged out.

A community reaction to the repeated WV measurement discrepancies was to conduct the extensive AquaVIT laboratory intercomparison at the Karlsruhe Institute of Technology Aerosol Interaction and Dynamics in the Atmosphere (AIDA) chamber in 2007. This experiment was unable to reproduce the large discrepancies between these measurements, and a conclusion was that these discrepancies occurred only when the instruments were operating in the UT/LS on airborne platforms (D. W. Fahey et al., The AquaVIT-1 intercomparison of atmospheric water vapor measurement techniques, submitted to *Atmospheric Measurement Techniques*, 2014). These observations have motivated the further intercomparisons described here between in situ instruments operating in the UT/LS.

In this work, WV measurements from the NASA MACPEX experiment are presented and discussed. MACPEX was a NASA WB-57F mission based in Houston, Texas, in March and April of 2011, providing nine flights with sufficient science-quality data adequate for intercomparison (Figure 1). The WB-57F payload included HWV, JLH, and ALIAS which had previously flown on the WB-57F; the FISH and NASA Langley/Ames Diode Laser Hygrometer (DLH) instruments that had flown in the UT/LS but not on the WB-57F; and one new instrument, the NOAA Chemical Ionization Mass Spectrometer (CIMS)-H_2O. In addition, six flights were coordinated with balloon soundings of either the CFH or FPH frost point hygrometers. Given the altitude range and payload of the WB-57F, MACPEX afforded an excellent opportunity to intercompare the UT/LS WV measurements and evaluate progress toward agreement. Here, we intercompare the measurements made during MACPEX and examine the scientific significance of the measurement discrepancies considering the recognized instrumental uncertainties and desired accuracy.

2. Instruments

The measurements discussed here were made with five different in situ hygrometers integrated on the NASA WB-57F as well as FP hygrometers deployed on meteorological balloons. These measurements represent four fundamentally different analytical techniques, and each instrument team independently calibrated their instrument using a variety of primary and secondary standards. Table 1 summarizes the stated accuracies of the instruments and briefly outlines the sources of the various uncertainties and potential measurement artifacts. In the remainder of this section, brief descriptions of the instrument principles of detection and calibration are provided.

2.1. HWV

HWV is a Lyman-α photofragmentation, fluorescence-based instrument [*Weinstock et al.*, 1994]. The HWV instrument has progressed through a number of in-flight configurations beginning in 1993. The HWV Lyman-α detection axis configuration used during MACPEX was first deployed in 2007 during the NASA TC^4 mission on the WB-57F aircraft out of San Jose, Costa Rica. A new aspect of the MACPEX HWV configuration is that a separate detection axis (Harvard Herriott Hygrometer (HHH)) was located upstream of the Lyman-α

detector. HHH data from MACPEX were only sparsely available and is discussed by Sargent et al. [2013].

HWV samples ambient air through an 8.9 × 10.2 cm forward facing duct extending forward of the wing from the nose of an aircraft wingpod. The sample duct is designed for very fast flow in order to render insignificant any evaporation of condensed-phase water due to compressional heating. Within the main inlet duct, gas is subsampled into a 5 × 5 cm secondary duct containing the fluorescence cell. This configuration allows for extremely high flow rates (~10^4 lpm) through the detection region, ensuring fast response and minimizing artifacts due to water outgassing from inlet surfaces. WV is detected by photofragmenting H_2O using a Lyman-α (121.6 nm) lamp and detecting fluorescence emission from the resulting OH* fragment. The HWV sensitivity is calibrated routinely on the ground at lower flow rates using a series of WV mixing ratios generated by dilutions of vapor-saturated flows over a range of pressures and temperatures. Optical background in the OH* emission band (315 ± 10 nm) is monitored periodically in situ by selectively blocking the Lyman-α flux from the lamp with a quartz window. Background fluorescence from illuminated components of the instrument has been evaluated to be insignificant (≤0.3 ppmv) in the laboratory with a flow of very dry air [*Hintsa et al.*, 1999]. Background artifact due to outgassing has been evaluated to be insignificant by changing the secondary duct flow rate during aircraft measurements. The instrument was calibrated before and directly after the MACPEX mission. During MACPEX, measurements of the output of the laboratory Harvard calibration system made with the NOAA MBW 373LX (MBW Calibration Ltd., Switzerland) reference frost point hygrometer yielded a correlation slope (saturator versus MBW) of 1.01 over the range of 2–40 ppm, further verifying the HWV laboratory calibration system.

2.2. FISH

FISH is a Lyman-α instrument that occupied the WB-57F fuselage pallet forward of CIMS-H_2O [*Zöger et al.*, 1999]. During MACPEX, FISH was operated in an enhanced total-water configuration (i.e., vapor plus amplified condensed phase), using a forward facing inlet that was heated in order to completely evaporate condensed water in particles prior to the detection of WV. The inlet sampling point was 35 cm from the fuselage, well outside the aircraft boundary layer. The inlet line was 2.5 m of 1 cm i.d. tubing (2 m of heated stainless steel tubing followed by 0.5 m of Teflon tubing) with a high flow rate (10–50 lpm, decreasing with increasing altitude). The FISH calibration constants were determined routinely between each flight. This was accomplished by sampling WV-in-air flows generated with a saturator and quantified simultaneously with an MBW hygrometer operated near 2 bar. During MACPEX the MBW instruments used for on-ground calibration by the FISH and CIMS-H_2O teams were compared and shown to be consistent to within 1%. The optical background of FISH is determined in situ in the same way as it is for HWV (blocking the UV lamp), and laboratory tests have verified the accuracy of this procedure.

2.3. DLH

DLH is an open-path TDL spectrometer that determines WV content by measuring absorption near 1.4 μm over a long path outside of the aircraft [*Diskin et al.*, 2002; *Podolske*

et al., 2003]. In the WB-57F configuration the DLH laser and detector were mounted in a wing hatch, and a turning mirror and retroreflective sheet were affixed to wing pods creating a folded path with a round trip optical path length of 20 m. During MACPEX the instrument suffered from a variable H_2O contamination internal to the laser/detector housing. Because the housing was maintained at an elevated pressure relative to the ambient, internal water could be spectroscopically discriminated, especially at high altitudes (low ambient pressures). Therefore, the contribution of internal WV contamination to the DLH measurement was removed during final data reduction, resulting in a measurement uncertainty of < 1 ppm. During the subsequent NASA Airborne Tropical Tropopause Experiment a number of additional measures were taken to address this issue. These included increasing the pressure differential between ambient air and the instrument enclosure, purging the enclosure, and directly measuring the WV inside the enclosure.

2.4. ALIAS

During MACPEX, ALIAS was mounted in the WB-57F right spear pod and used a heated, forward facing isokinetic inlet to vaporize condensed water before the sample was analyzed. ALIAS is a closed-path multichannel TDL spectrometer that employs a multipass analysis cell (36 m optical path length, 300 cm^3 volume). WV was measured using interlaced direct absorption and 2f detection [*Webster et al.*, 1994] at 2.65 μm. For data reduction, a Voigt line shape was used and the spectroscopic parameters as measured at Jet Propulsion Laboratory (JPL) were within 5% of the HITRAN2008 values [*Rothman et al.*, 2009].

ALIAS WV measurements were cross checked in the field after individual MACPEX flights as well as in the laboratory before the mission. In the field, ALIAS sampled synthetic air containing 50 ppm WV (3% uncertainty) after conditioning the instrument for several hours. In the laboratory, ALIAS sampled synthetic air containing WV concentrations from 3 to 50 ppm with flows that were continuous for several days. In the laboratory, water vapor flows were produced by a Thunder Scientific TS3900 humidity generator with a Vaisala frost point hygrometer (DMP501) sampling its output in parallel to ALIAS (2% combined uncertainty). ALIAS was significantly modified between MACPEX and its previous deployment in 2006 during CR-AVE. Increasing the wavelength from 1.88 μm to 2.65 μm allowed for a significant reduction in the cell volume from 16,700 cm^3 to 300 cm^3. The fore-optics were integrated into the measurement region (i.e., the gas that was sampled between the folding mirrors also flowed over the fore-optics). Additionally, a mesh with 100 μm openings was inserted into the flow on many flights to aid the sublimation of sampled ice particles. Subsequent comparison of mesh flights against nonmesh flights did not show obvious differences in outgassing properties, though further examination is required to raise confidence in this assessment.

2.5. CIMS-H_2O

The CIMS-H_2O instrument uses ion-molecule reactions in a small ionization volume to produce hydronium analyte ions from WV in the sample flow. The measurement technique and uncertainties are described in detail by *Thornberry et al.* [2013]. During MACPEX the instrument was mounted in the rear fuselage pallet of the WB-57F. The instrument sampled ambient air through an inlet line oriented orthogonally to the aircraft flight direction to reject

particulate water [*Perring et al.*, 2013]. The sampling point was located outside the aircraft boundary layer to sample air in the free stream. Inside the CIMS-H_2O, a portion of the sample flow passed through an ionization chamber where it was exposed to α-particle radiation. Air reacted with α particles from decay of ^{241}Am, producing mostly O_2^+ and N_2^+, which lead to a series of ion-molecule reactions resulting in the production of H_3O^+ ions from ambient WV. Ions were analyzed using a quadrupole mass spectrometer, and the measured count rate of the H_3O^+ analyte ions was used to quantify ambient WV.

The CIMS-H_2O used two independent calibration sources to sample known mixing ratios of WV into the instrument every 45 min during flight operation. The standards employed were (1) a mixture of WV in air near 5 ppm that was stored in an electro-polished stainless steel cylinder and (2) flows of gravimetrically prepared H_2 standards that were dynamically diluted in zero air with onboard mass flow controllers and then converted quantitatively to WV using a heated platinum (Pt) catalyst immediately prior to sampling [*Rollins et al.*, 2011]. The first standard was used to completely displace the ambient flow during a calibration procedure. The second standard was used both to completely displace the ambient flow with a WV/zero air mixture, as well as to add small quantities of WV to ambient air. The two systems are complementary because the H_2/Pt system produced a series of calibration points over a dynamic range of ~0.5–150 ppm, while the gravimetric WV standard provided a single point that was unlikely to be affected by the environmental conditions or by dilution system errors. Both calibration systems were routinely checked against a National Institute of Standards and Technology (NIST) traceable MBW 373LX reference hygrometer on the ground. Each system produced mixing ratios of WV known to better than ±10%, and the two systems were shown to be consistent with each other both on the aircraft and in the laboratory between flights.

2.6. Frost Point Hygrometers: CFH and NOAA FPH

Two different balloon-borne FP hygrometers were used in this study that are based on the same working principle but are of somewhat different designs. The instruments were not retrieved after each flight and were therefore each flown only once. Due to limitations in the available number of FP instruments and legal restrictions on balloon payloads, only one FP instrument was flown on each sounding.

The FPH is designed and built by the NOAA Earth System Research Laboratory [*Hurst et al.*, 2011]. The CFH is the design described by *Vömel et al.* [2007a]. Both instruments use a liquid cryogen to cool a mirror where frost forms from deposition of ambient WV. Intermittent heating of the cryogen-cooled mirror is used to maintain a stable layer of frost at a temperature where it is in equilibrium with the surrounding WV. The diffuse scattering of focused infrared light by frost on the mirror, as monitored by a temperature-controlled light-emitting diode and photodiode, provides a measure of frost coverage. This coverage signal is used as feedback to a proportional integral differential heater controller with adaptive gains to maintain the stable frost layer throughout each balloon flight. When the frost point temperature initially reaches −53°C during balloon ascent, the frost layer on the mirror is rapidly sublimated and regrown to ensure the formation of pure hexagonal ice. The mirror temperature is measured with a precision thermistor that is embedded in the mirror and

calibrated against a NIST-traceable temperature probe and a small archived set of calibrated thermistors. The WV partial pressure is calculated using the Goff-Gratch equation for hexagonal ice/vapor equilibrium [*Goff and Gratch*, 1946; *Murphy and Koop*, 2005] at the measured mirror temperature and is converted to a mixing ratio using coincident, calibrated pressure measurements by an accompanying radiosonde.

Vömel et al. [2007a] summarize uncertainties associated with the frost point measurements in detail. During MACPEX the NOAA FPH demonstrated somewhat higher precision than the CFH due to slightly different performance of the FPH frost control algorithm. The quality of frost control often varies from one flight to another and even during a flight, depending on how well the algorithm handles different atmospheric conditions. Over a wide range of stratospheric conditions the FPH and CFH generally control frost with similar ability and, therefore, have the same estimated measurement uncertainties (Table 1).

An important distinction between the FP instruments and the WB-57F instruments is that because there is a low rate for successful recovery of the balloon hygrometers, each flight is performed using a new instrument (produced to identical specifications). The FP measurements made during ascent are susceptible to outgassing of WV from the balloon and parachute, so here only measurements made during controlled descents are compared, even though data on ascent agree well with data on descent for most of the soundings. A valve in each balloon neck automatically opens at a preset pressure altitude to release helium before the balloon bursts, allowing the payload to descend at a controlled rate of ~5 m/s. This procedure is typically only used for FPH soundings: however, since both instruments were operated using NOAA Earth System Research Laboratory balloons during MACPEX, both FPH and CFH were used to acquire data under a controlled descent.

2.7. Other Data Sets

Other WV instruments that flew on the MACPEX WB-57F payload included JLH, as well as the JPL Unmanned Aircraft System Laser Hygrometer (ULH, an updated version of JLH), and the Closed-Path Laser Hygrometer (CLH) [*Davis et al.*, 2007]. JLH and ULH did not report data below 20 ppm during MACPEX, and, thus, these hygrometers are not considered in our analysis. CLH is designed to detect total water, and accurate measurements of low WV mixing ratios are not possible due to outgassing from the instrument inlet and detection cell surfaces as well as insufficient precision below 10 ppm. Therefore, CLH was also excluded from the present analysis.

2.8. Data Availability

Our intercomparison of the WB-57F measurements is restricted to the eight flights between 14 and 26 April when most instruments considered were typically producing science-quality data. Data from six FP sonde launches starting 13 April are also included in the intercomparisons. The data availability from the various instruments is summarized in Table 2. For intercomparison of the aircraft and balloon-based measurements, the only data used are from the WB-57F final descents, which were coordinated with the balloon launches in order to maximize temporal and spatial coincidence (see section 3.2). These interplatform intercomparisons are limited to the lower stratosphere (altitude > 15 km), where the WV

field is more spatially and temporally homogeneous than it is in the middle to lower troposphere. Because FISH measured total water, the FISH data are included only when the RH_i as calculated from the FISH data was less than 90%. An analogous RH_i criterion was used to screen data from ALIAS. Intercomparisons between the WB-57F hygrometers are restricted to WV mixing ratios below 150 ppm, the maximum mixing ratio for which CIMS-H_2O was calibrated.

3. H_2O Intercomparisons

Figure 1 is a map showing the WB-57F flight tracks from 13 April to 26 April, as well as a time series of altitude and WV data from a typical flight. MACPEX was primarily designed as a cirrus cloud study, and thus, in situ sampling of cirrus below the tropopause (~8–12 km) was typically the focus of the middle portion of the flights.

Data are intercompared here in two ways. The first is to use all available data that are coincident between a pair of two instruments. This approach is used exclusively for intercomparisons between the WB-57F hygrometers. While it would be desirable to intercompare measurements only when data were available from *all* instruments, this approach would result in elimination of more than 90% of the flight time, significantly reducing the quality of the statistics. The second approach used here is to intercompare only the data from the aircraft final descents from the maximum altitude leg immediately prior to landing. Due to the recognition that ascent data from some of the instruments are associated with larger uncertainties (e.g., larger background artifacts from water outgassing), the balloon launches were coordinated with the WB-57F final descents. The full range of data (< 150 ppm) are used to evaluate scaling (i.e., slope, sensitivity) differences between the instruments, and the data below 10 ppm are focused on to calculate offsets between the instruments. In the following sections we first focus on the intercomparison of the WB-57F instruments up to 150 ppm, then on the lower mixing ratios, and finally on the intercomparisons with the FP instruments.

3.1. WB-57F Measurements

Results of orthogonal linear regressions between all possible pairs of the WB-57F instruments (HWV, FISH, DLH, CIMS-H_2O, and ALIAS) using all available data < 150 ppm are shown in Table 3. Figure 2 shows a representative subset of the comparisons. For all instruments except ALIAS, 1 Hz WV values are compared. ALIAS reported data at a slightly lower data rate (approximately 0.58 Hz), and for these regressions HWV, CIMS-H_2O, FISH, and DLH data were averaged to 0.58 Hz.

Linear regressions were calculated for all instrument pairs for each flight for all data up to 150 ppm (Figure 2 insets). Table 3 shows the fit slopes from these regressions, both the range of correlation slopes for the individual flights, and the mean and standard deviation of the slopes. The regression slopes for the various hygrometer pairs confirm agreement within the fractional instrumental uncertainties (Table 1). Individual flight correlation coefficients (R^2) for all instrument pairs are greater than 0.99, indicating that linear regressions are an appropriate way to compare the data. The average slopes range from 0.976 to 1.070. This is evidence that the various calibration techniques used to determine the instrumental

sensitivities to WV were consistent with each other to within ±5% (a 9.4% range) and that some of the reported scaling uncertainties of ±6–10% (Table 1) may, if anything, be overestimated. The flight-to-flight variability in these slopes is taken as an indication of the flight-to-flight instrument measurement stability. This variability was relatively small (standard deviation < 6% for all comparisons) indicating that the instruments were stable to within the stated uncertainties relative to each other throughout these 2 weeks of science flights.

Figure 3 shows the percent differences between individual instruments and HWV using all available data below 150 ppm. For this figure, the HWV data set was used as the reference because HWV had the best data availability throughout MACPEX (see Table 2). Boxes and whiskers in Figure 3 show statistics for the percent differences (see Figure 3 caption) using logarithmically spaced WV bins. Statistics for all the data are shown in red and in black for only final descent data.

The FP measurements are also compared in Figure 3 using data from balloon and WB-57F descents as described in section 3.2. The goal in timing of the balloon flights was to match air masses. Figure 4 shows the UT/LS vertical profiles of ozone and temperature for the six balloon flights. The generally very good agreement indicates the high degree of comparability of the air masses separately sampled by the two platforms.

Figure 3 shows that at the low end of the mixing ratio measurements, larger fractional differences between instruments were typically observed as compared to the higher mixing ratios. These generally larger fractional differences between instruments at low mixing ratios imply that the most significant differences between instruments are being driven by background artifacts (offsets) as these are fractionally more important at low mixing ratios.

The low-mixing ratio differences are illustrated in Figure 5, where vertical profiles of WV measured by all instruments are shown on days when FP data were available. As was evident from the correlation analysis, profiles from the different instruments share much of the same vertical structure. However, the near-constant differences observed between instruments in the UT/LS region indicates that measurement differences are clearly dominated by systematic errors over random uncertainties. Differences in absolute value between the instruments varied somewhat from flight to flight. There were cases where the differences varied during an individual flight, indicating possible drifts. This variation is also evident from the scatter in the < 10 ppm comparisons in Figure 2.

To quantitatively distinguish offsets between instruments from errors in the calibration of instrumental sensitivity, the mean difference between each pair of instruments below 10 ppm is calculated after scaling one of the measurements by the slope difference found using all data below 150 ppm (e.g., CIMS/1.046-HWV, see Table 3). This was found to be a more reliable way to assess offsets between instruments as compared to using the *y* intercept of a linear regression, which is an extrapolation of the data and not necessarily well constrained for sparse data below 10 ppmv. To examine the variation of the differences at low mixing ratios, this analysis was performed between all pairs of instruments for each flight, using only data from the WB-57F final descent and using the scale difference found for that flight

between the two instruments. The results are summarized in Table 4 and shown graphically in Figure 6. For example, from Table 4 FISH has an average offset of 0.63 ppm relative to HWV, while HWV, CIMS, ALIAS, and DLH have average offsets of less than 0.1 ppm with respect to each other.

Using each of FISH, HWV, and CIMS-H_2O as a reference for the other measurements, Figure 6 shows that the instrument offsets covary to some degree, regardless of which instrument is used as the reference. For example, when FISH is used as the reference (Figure 6, x-FISH panel), most of the instruments follow a similar daily pattern with respect to their difference from FISH (down from 14 April to 20 April, then up to 23 April). Similar trends can be observed for various time segments using the other instruments as references. Table 4 and Figure 6 show that the variability in instrument offsets cannot be clearly attributed to a single instrument. The fact that the flight-to-flight variation was not limited to one of the instruments indicates that all instruments may have unrecognized variable background artifacts on the order of ~0.2 ppm (i.e., typical standard deviation of offsets in Table 4). During some of the flights a significant shift or drift in the difference between two instruments was observed over a relatively short time scale. Such differences can be clearly seen in the comparisons shown in Figure 2, for example, between HWV and CIMS-H_2O on 26 April (orange), or between FISH and CIMS-H_2O on 21 April (green), or between DLH and CIMS-H_2O on 25 April (pink) where drifts in the instrumental difference up to 1 ppm occur.

The most rapid changes in these differences typically occurred near the beginning of a flight or immediately following some other exposure to a high concentration of cloud water. This indicates that outgassing of water from instrument inlets or other surfaces occurred and that it was not adequately quantified on the necessary time scales and, hence, not accounted for in the reported WV values. Those instruments that intentionally ingest and sublimate cirrus particles (FISH and ALIAS) appear to be especially impacted by cirrus sampling. The CIMS-H_2O instrument which has a short stainless steel inlet did not show any indication of a sampling artifact.

As a typical example of this issue, Figure 7 shows two periods from the 21 April flight during which the WB-57F ascended to sample stratospheric air. The first portion of the time series shows data shortly after takeoff and climb out from the humid Houston boundary layer. Slow downward drifts relative to HWV are evident from ALIAS, FISH, and DLH on the order of 10–30% over the period of 1 h (CIMS-H_2O data were not available here). The second portion of the Figure 7 time series shows data ~3.5 h into the flight, immediately after a UT cirrus sampling period. The large differences (>100%) between the total water (FISH and ALIAS) and vapor measurements (CIMS-H_2O, HWV, and DLH) at the beginning of this period indicate cirrus along the flight track (gray shading). During ascent into the LS (8.45×10^4 s < UTC time < 8.5×10^4 s), large percentage differences between HWV and FISH (i.e., FISH more than 50% higher than HWV) are observed. In this case, RH_i calculated from both instruments is less than 20%, indicating that these differences cannot be due to real differences between WV and total water. During the high-altitude leg (i.e., UT time between 8.5×10^4 s and 8.75×10^4 s), a drift of up to 0.3 ppm occurs between HWV and CIMS-H_2O, while the difference between FISH and CIMS-H_2O changes by 0.7 ppm.

During this time ALIAS drifts down by 2 ppm (40%) and DLH up by 0.1 ppm relative to HWV. On descent, once WV mixing ratios become greater than ~50 ppm, the percentage difference between the instruments return to steady values of near 5–10%, which are dominated by uncertainties in instrument scale factors rather than background artifacts. In examples like this, the initial large differences between, for example, FISH and HWV is almost certainly due to cirrus artifacts. Therefore, given the framework of multiple measurements present here, these data can be easily eliminated from further scientific analysis. However, the longer drift between instruments observed after this initial period is still significant (~10–40%) and the causes are not obvious, making it problematic to reason from the multiple measurements alone which is likely the most accurate. These longer drifts can only be assessed with an independent, drift-free, in-flight reference standard such as was used with the CIMS-H_2O instrument.

The importance of long-term outgassing on the measurements is also illustrated in Figure 3. Systematic differences between ascent and descent data on the order of 5% are most likely a result of water steadily desorbing from instrumental surfaces and entering the detection region (or potentially other nonintentional and unswept optical paths for the TDL instruments). FISH and ALIAS values were typically lower than HWV on the aircraft descent. CIMS and DLH reported systematically higher WV values on the final descents relative to HWV, and the reason(s) for this is not clear.

3.2. Frost Point Hygrometers

WB-57F and balloon flights were coordinated to maximize spatial and temporal coincidence in the UT/LS during the final aircraft and balloon descents. Forecast winds were used to plan balloon launches to best match with the WB-57F location just before descent into Houston at aircraft maximum altitude. In practice, this resulted in coordination to within 50 km and 1 h of the balloon at 16.7 km. The similarity of the FP ascent and descent data typically indicated near-zero horizontal gradients in LS WV over such distances, making the interplatform comparison close to ideal. Comparisons of the WB-57F and balloon measurements of ozone, temperature, and WV demonstrated that sampling of very similar air masses was successful (Figures 4 and 5). To compare the WB-57F data to that of the balloon-borne FP hygrometers, the WB-57F final descent data were binned to the GPS altitude of the balloon (typically ~5 m bins), and mean WV values were calculated.

Comparisons of the matched FP and WB-57F data are shown in Figure 8 and summarized in Table 5. Figure 8 (top six panels) shows vertical profiles of the differences between each WB-57F instrument and the FP measurements using final descent data only (e.g., HWV-FP). Figure 8 (bottom six panels) also shows the WB-57F data plotted against the FP data above 15 km. Figures 4, 5, and 8 indicate that the coordination between the WB-57F and balloon platforms was not as good on 13 April as on the other flights, and therefore, this flight was not considered in the further analysis. The vertical profiles in Figure 8 also show the mean (orange dash) and standard deviation (gray box) of the differences above 15 km. The difference statistics are summarized by boxes (25th–75th percentile region) and whiskers (10th–90th percentiles). Scale factor differences are not considered here as these could not

be well constrained given the narrow range of mixing ratios available for comparison (~3.5–6 ppm).

The comparisons between the FP and WB-57F measurements were highly consistent from flight to flight. In Figure 5 it is clear qualitatively that the agreement for example between FP and FISH was very consistent from flight to flight regardless of whether the CFH or FPH was used. Figure 6 and a comparison of Tables 4 and 5 shows quantitatively that the flight-to-flight stability of the FP appeared to the best of all of the instruments (i.e., x-FP panel in Figure 6 shows the least flight-to-flight variability of all the panels in Figure 6). The fact that the six different FP hygrometers consisting of two different designs (CFH and FPH) agree so well with each other (by using the WB-57F instruments as a transfer standard; see Figure 6, top) is strong evidence that the average values measured by these two FP sondes are not significantly dependent on the subtle differences in their designs and frost control algorithms used.

4. Discussion

Our analysis of the data from MACPEX describes agreement between a suite of hygrometers that is better than has typically been observed in previous intercomparisons of UT/LS WV [*Kley et al.*, 2000; *Read et al.*, 2007; *Vömel et al.*, 2007a; *Jensen et al.*, 2008; *Weinstock et al.*, 2009]. As a point of reference for MACPEX, Figure 9 shows WV data from the 1 February 2006 flight during the CR-AVE experiment. Only data from the WB-57F final descent are shown. Of the NASA missions during which TTL air was sampled, CR-AVE encountered some of the driest conditions. This flight in particular was chosen for comparison because of the good data coverage and because as is shown in Figure 9 both the temperature and O_3 measurements indicated that the balloon and WB-57 sampled very similar air masses.

Figure 9 shows WV data from HWV, ICOS, JLH, CFH, and ALIAS (data downloaded from NASA online databases on 10 October 2012; data versions: HWV R4, ICOS R1, JLH R0, CFH R0, and ALIAS R2). To calculate fractional differences between the measurements, a mean profile was calculated from HWV, ICOS, JLH, ALIAS, and CFH binned by ambient pressure during WB-57F final descent and balloon ascent. From the balloon temperature and pressure, a corresponding profile of saturation mixing ratio was calculated, showing that during this coordinated intercomparison, supersaturated air was not reported by any instrument. A survey of the other comparisons from this mission revealed that while high supersaturations, including some exceeding the homogeneous freezing threshold, were reported by a number of the instruments at times during various flights, these never happened during the coordination with CFH [cf. *Jensen et al.*, 2005 cf. *Jensen et al.*, 2008].

A comparison of Figures 5, 8, and 9 shows that the MACPEX data agreement is an improvement over CR-AVE. Whereas in CR-AVE at an instrumental mean value of 5–8 ppm there is a somewhat constant difference of 25–30% (1–2 ppm) between CFH and, for example, HWV, during MACPEX this difference was closer to 10% (0.4 ppm). The magnitude of the MACPEX fractional differences between all of the instruments at the lowest mixing ratios observed (3–4 ppm) were generally within a 20% range, which is in

fact much closer to the differences observed in the AquaVIT laboratory intercomparison (D. W. Fahey et al., submitted manuscript, 2014). This improvement suggests that the causes of some of the in-flight-only problems that historically have resulted in large disagreements may have been mitigated, although the improvement cannot be clearly attributed to a specific set of changes in an instrument(s) or operational procedure(s). One caveat is that WV reached lower values in CR-AVE than in MACPEX (~2.2 versus 3.5 ppm). In MACPEX the observed spread in the lowest mixing ratios was ~0.8 ppm, and this was relatively consistent across the measurements below 10 ppm (Figure 5). Thus, if these offsets remained the same in the 1–2 ppm range, more significant fractional discrepancies (40–80%) would have been observed. We therefore argue that the instrument discrepancies in MACPEX, though improved, remain scientifically significant. Additionally, the temperatures encountered at the lowest mixing ratios during MACPEX were significantly above the ice saturation temperature. For example, at 100 hPa the lowest temperature encountered was near 200 K, which corresponds to a saturation mixing ratio greater than 12 ppm (i.e., more than twice any value seen in MACPEX at this altitude). These measurements, therefore, do not directly address the critical issue of very high supersaturations with respect to ice that were observed in the CR-AVE mission (>200% RH_i [e.g., *Jensen and Pfister*, 2005; *Jensen et al.*, 2008]).

4.1. Background Artifacts

We have shown that a likely cause of the scientifically most important differences between the reported WV mixing ratios are offsets between instruments which are attributed to unquantified background artifacts. Some of the background artifact issues appear to be due to instrument outgassing and/or cirrus artifacts that can have large amplitudes and undergo rapid changes in flight. Other slow drift or daily changes in background artifacts are significant but have less obvious sources. The issues associated with drift, which reduce the accuracy of the measurements, could be improved by implementing a technique to measure the instrument background including any outgassing in situ, at sufficiently frequent intervals. The spectroscopic techniques (i.e., Lyman-α, IR absorption) essentially measure the instrument background by tuning an excitation source off resonance, eliminating the signal due to any source of water. While this is advantageous because it allows for a very frequent and near-immediate metric of the background, these techniques cannot always quantify in real time the background artifact due to outgassing or trapped water in the instrument, nor can they be used to detect spectral interferences (e.g., other molecules in the gas phase or on instrument surfaces which may produce OH* from Lyman-α excitation). Open-path instruments such as DLH are generally not susceptible to contamination in the measurement optical path (i.e., outside the aircraft) but can be influenced by water in unintentional portions of the optical path inside the instrument housing, as was observed with DLH during MACPEX. Therefore, spectroscopic background detection techniques inherently allow the possibility of a single-sided (positive) artifact in the calculated WV mixing ratio.

The only method that we know of to quantify these background artifacts in real time is to sample known mixing ratios of water in the inlet at the same location from which ambient gas is sampled. This strategy is obviously not possible for open-path instruments or for instruments with extremely high flow rates. To our knowledge the CIMS-H_2O deployment

during MACPEX was the first for which the instrument background artifact could be quantified in situ using standard addition. Here, WV was quantified at the lowest mixing ratios with typical accuracies of ±10%. In the future, these uncertainties could be reduced to near ±5% by reducing the number of uncertainties incurred in determining the instrumental response and by linearizing the instrument sensitivity. The use of in situ standard addition may also help to evaluate background artifact for other instruments that use an inlet and moderate flow rates, e.g., FISH and ALIAS.

4.2. The Meaning of Agreement

Many of the reported mixing ratios from MACPEX using various instruments disagree enough that the differences are scientifically significant. However, the differences are often small enough compared to the combined uncertainties stated for each instrument that the measurements can be said to be statistically consistent with each other [e.g., *Immler et al.*, 2010] (in contrast to CR-AVE, when the differences frequently exceeded combined uncertainties). There is, therefore, a need for a quantitative evaluation of how well the measurements agree within stated uncertainties.

WV mixing ratios, c, with their associated uncertainties, are typically reported as $c \pm \delta c$, and often the definition of the δc uncertainty value is not explicitly stated by the investigators. Due to the fact that the uncertainties in many measurements are a result of combining a number of uncorrelated systematic and random errors, the probability distribution function (PDF) for many atmospheric measurements is frequently approximated as a normal distribution. In this case, if we assume that $\delta c = 2\sigma$ (σ^2 = variance), then we expect that the most likely true value of water is the value c and that there is ~95% confidence that the true value is in the range $c - \delta c$ to $c + \delta c$. A quantitatively quite different approach is if the PDF for the stated uncertainty is a boxcar function where all values in the range $c - \delta c$ to $c + \delta c$ are equally likely to be the true value. This point is illustrated in Figure 10 where average values measured from 16.5 to 16.7 km are shown from the 23 April descent. Error bars in this figure are not the standard deviation of the data points (which would be much smaller than the uncertainties shown) but rather are calculated from the instrumental accuracies listed in Table 1. For CIMS-H_2O, for example, the uncertainties are derived by propagation of various uncorrelated uncertainties accumulated in the delivery of calibration standards to the inlet and in the use of this calibration method to calculate ambient WV mixing ratios. Thus, the uncertainties are essentially normally distributed and symmetric. This is not the case with all instruments, some of which individually quantify uncertainties in the background artifact and sensitivity. For example, for HWV, uncertainty limits are nonsymmetric about the reported value due to WV in the zero air used in the laboratory that is known only to be below a certain value. The FISH uncertainties are ± (6% + 0.15 ppm) due to the independent uncertainties in the instrument sensitivity to WV and the instrumental background.

Due to the lack of precisely defined errors and historically poor agreement of UT/LS WV, “agreement” in this community has frequently been discussed to occur simply when the error bars of two instruments overlap to any degree, implying that boxcar uncertainties are assumed [e.g., *Weinstock et al.*, 2009; D. W. Fahey et al., submitted manuscript, 2014]. If

this is the case, most of the measurements of WV made during MACPEX could be said to agree. For example, we calculated for each flight the frequency with which the various instruments' reported uncertainties overlapped for the low WV mixing ratios observed on WB-57F final descent. Under this definition HWV and CIMS-H_2O always agree, HWV and FISH agree 96% of the time, and CIMS-H_2O and FISH agree 84% of the time. However, agreement in this case would necessarily mean that the true WV mixing ratio must fall in the common value region, which was nearly always at the low end ($c - \delta c$) of the uncertainties for CIMS-H_2O DLH and ALIAS and at the high end ($c + \delta c$) for FISH and FP uncertainties. This occurrence would of course be highly unlikely if the errors are truly normally distributed. This point is illustrated in Figure 10 (gray shaded region), which is the product of the FPH and CIMS-H_2O PDF functions. By integrating the product, the probability of the two instruments agreeing in this typical example is 5.7%.

5. Concluding Remarks

A successful and comprehensive airborne intercomparison of UT/LS WV measurements has been accomplished. The NASA WB-57F based in Houston, Texas was used to profile the UT/LS in the central U.S. and Gulf of Mexico. The aircraft was deployed with a payload that included the current versions of legacy instruments from the U.S. (i.e., HWV, ALIAS, JLH, and DLH), some of which have been associated with significant discrepancies in the past. The FISH instrument, which has played a major role in European UT/LS science, was on-board the same aircraft with those from the U.S. for the first time. The recently developed CIMS-H_2O instrument with redundant in situ calibration methods added a new and unique element to the payload. Soundings of both FPH and CFH balloon-borne frost point hygrometers were successfully coordinated with the WB-57F for high-quality intercomparisons in the lower stratosphere. A principal focus of the MACPEX campaign was intercomparison of these instruments below 10 ppm WV. The resulting data set has been used to evaluate instrument performance in this range where historically the most significant discrepancies have been observed.

The overall intercomparison between the WB-57F hygrometers considering data up to 150 ppm is generally quite good, with the systematic differences between instruments typically not exceeding 10% except at the lowest mixing ratios. The average systematic difference between various pairs of instruments (HWV, DLH, CIMS-H_2O, and FISH) in this range was < 7%, indicating that all techniques have intrinsic skill for measuring atmospheric WV and that the standards and procedures used for calibration of instrument-scale factors are consistent with each other to within their stated uncertainties.

In our assessment, instrument performance below ~8–10 ppm was distinctly different than it was at higher mixing ratios. Figure 3 shows that while at high mixing ratios the average percent deviations between instruments was generally within a 10% range, at the low mixing ratios the instruments systematically deviated from each other by 10–20%, which is at times in excess of the reported combined instrumental uncertainties. A likely cause of these differences between high and low mixing ratios appears to be variable background artifacts. These differences are quite similar to those observed in the AquaVIT laboratory intercomparison. Figures 3 and 7 show that a positive background artifact due to outgassing

(especially after cirrus sampling) most likely affects the total water FISH and ALIAS instruments to some degree. The CIMS-H_2O instrument routinely checks for background artifact in flight by completely displacing the ambient flow with a WV standard, and we have argued that this procedure is the only way that certain background artifacts can be accurately assessed.

The FPH and CFH instruments are used routinely to measure WV below 10 ppm throughout the annual cycle at tropical and midlatitudes. The long-term data set has been used for trend analysis [*Oltmans et al.*, 2000; *Scherer et al.*, 2008; *Hurst et al.*, 2011] and for satellite validation [*Read et al.*, 2007; *Vömel et al.*, 2007b; *Hurst et al.*, 2014]. The data collected with this family of hygrometers at Boulder, Colorado is the longest standing record of stratospheric WV in existence, making it an extremely valuable data set. The importance of this record makes it imperative to fully understand the measurement uncertainties. In addition, if the FP records are to be augmented with aircraft measurements, the cause of the intercomparison differences will need to be understood. During MACPEX the FP instruments consistently agreed quite well with the FISH instrument, but both reported values 0.38–0.89 ppm (~10–20%) drier than the other WB-57F hygrometers. The FP hygrometers are generally considered to be reliably accurate and less susceptible to long-term calibration drifts and shifts because WV mixing ratios are calculated from only thermistor temperature and ambient pressure using an equation relating temperature to vapor pressure over hexagonal ice. However, some historical instrumental changes and errors in the data reduction have required a few small systematic changes in the reported FP mixing ratios, demonstrating that they are not immune to errors [*Vömel et al.*, 2007a; *Scherer et al.*, 2008]. Although an extensive testing of chemical interferences from condensable atmospheric species other than H_2O has not been conducted, at least one study has demonstrated that nitric acid is unlikely to interfere under typical operation [*Thornberry et al.*, 2011]. We know of no physical explanation for why these instruments would have a systematic low bias, but given the importance of the FP data set the discrepancies observed here and elsewhere stress the importance of continued investigation of possible reasons for unrecognized artifacts.

Our evaluation of the reported measurement uncertainties from MACPEX demonstrates that unrecognized uncertainties with some or all of the hygrometers persist and therefore that measurements will likely continue to disagree with statistical significance. One possible explanation is that errors are less symmetrically distributed than is currently recognized by the investigators. Therefore, clear statements about the origins of the uncertainties are required if measurements are to be interpreted properly.

Measurements of UT/LS WV are important for at least two scientific objectives. The first is to produce accurate, high-precision measurements of RH_i to provide meaningful constraints on ice microphysical processes. In the UT/LS, mixing ratios can be as low as 1–2 ppm, requiring measurement uncertainty in WV better than ±10% or 0.1 ppm at the lowest mixing ratios [e.g., *Weinstock et al.*, 2009]. The second objective is to identify long-term trends in stratospheric WV and its associated radiative forcing for assessment of the importance of WV toward climate change. This is a significant challenge because the detection of trends near 1% per year have been reported, and thus, systematic drifts in an instrument must be

much lower than this for these observations to be useful. Due in part to the lack of a continuous record based on frequent use of a single instrument with a fixed configuration, trend detection has not been possible with aircraft hygrometers. The available comparisons between satellite measurements show consistent interannual variability, but significant offsets exist between the instruments [e.g., Fueglistaler et al., 2013].

The results from the present study show that routine agreement among measurements still has not been achieved that is sufficient for all of the objectives. With the introduction of formal laboratory and in situ intercomparison campaigns such as AquaVIT and MACPEX, respectively, the field is working toward agreement after many years of large discrepancies. As a new instrument, CIMS-H_2O has demonstrated the ability to reliably use calibration standards in-flight with mixing ratios as low as 1 ppm to an accuracy better than 10%. The CIMS-H_2O measurements agreed on average with those of the other instruments to better than 8% on all flight days (Table 3), and this agreement provides confidence in the success and promise of this system. Therefore, we believe that the use of in situ standard addition offers a valuable and unique constraint on the accuracy of UT/LS WV measurements and in the future may play a significant role in further reducing instrumental discrepancies and uncertainties by providing a drift-free in situ reference.

Acknowledgments

The authors thank NASA for MACPEX funding and the ground and aircrews of the NASA WB-57F for their support during the mission. We thank T. Peter for helpful discussions with regard to the manuscript. Support for NOAA and CIRES researchers was provided by the NOAA Climate Program Office, the NASA Upper Atmosphere Research Program, and the NASA Radiation Sciences Program.

References

Albritton, DL., Zander, RJ. Instrument Intercomparisons and Assessments. World Meteorol. Org.; Geneva: 1985. p. 963-968.

Davis SM, Hallar AG, Avallone LM. Measurement of total water with a tunable diode laser hygrometer: Inlet analysis, calibration procedure, and ice water content determination. J Atmos Oceanic Technol. 2007; 24:463–475.

Diskin GS, Podolske JR, Sachse GW, Slate TA. Open-path airborne tunable diode laser hygrometer. Proc SPIE. 2002; 4817:196–204.

Feck T, Grooß J-U, Riese M. Sensitivity of Arctic ozone loss to stratospheric H_2O. Geophys Res Lett. 2008; 35:L01803.doi: 10.1029/2007GL031334

Fueglistaler S, et al. The relation between atmospheric humidity and temperature trends for stratospheric water. J Geophys Res Atmos. 2013; 118:1052–1074. DOI: 10.1002/jgrd.50157

Gao RS, et al. Evidence that nitric acid increases relative humidity in low-temperature cirrus clouds. Science. 2004; 303(5657):516–520. DOI: 10.1126/science.1091255 [PubMed: 14739457]

Goff JA, Gratch S. Low-pressure properties of water from 160 to 212 F. Trans Am Soc Heating Ventilating Eng. 1946; 52:95–112.

Harries J, Carli B, Rizzi R, Serio C, Mlynczak M, Palchetti L, Maestri T, Brindley H, Masiello G. The far-infrared Earth. Rev Geophys. 2008; 46:RG4004.doi: 10.1029/2007RG000233

Hintsa EJ, Weinstock EM, Anderson JG, May RD. On the accuracy of in situ water vapor measurements in the troposphere and lower stratosphere with the Harvard Lyman-α hygrometer. J Geophys Res. 1999; 104:8183–8189.

Hurst DF, Oltmans SJ, Vömel H, Rosenlof KH, Davis SM, Ray EA, Hall EG, Jordan AF. Stratospheric water vapor trends over Boulder, Colorado: Analysis of the 30 year Boulder record. J Geophys Res. 2011; 116:D02306.doi: 10.1029/2010JD015065

Hurst DF, Lambert A, Read WG, Davis SM, Rosenlof KH, Hall EG, Jordan AF, Oltmans SJ. Validation of Aura Microwave Limb Sounder stratospheric water vapor measurements by the NOAA frost point hygrometer. J Geophys Res Atmos. 2014; 119doi: 10.1002/2013JD020757

Immler FJ, Dykema J, Gardiner T, Whiteman DN, Thorne PW, Vömel H. Reference quality upper-air measurements: Guidance for developing GRUAN data products. Atmos Meas Tech. 2010; 3:1217–1231. DOI: 10.5194/amt-3-1217-2010

Jensen E, Pfister L. Implications of persistent ice supersaturation in cold cirrus for stratospheric water vapor. Geophys Res Lett. 2005; 32:L01808.doi: 10.1029/2004GL021125

Jensen EJ, et al. Ice supersaturations exceeding 100% at the cold tropical tropopause: Implications for cirrus formation and dehydration. Atmos Chem Phys. 2005; 5(3):851–862. DOI: 10.5194/acp-5-851-2005

Jensen EJ, et al. Formation of large (≈ 100 μm) ice crystals near the tropical tropopause. Atmos Chem Phys. 2008; 8(6):1621–1633. DOI: 10.5194/acp-8-1621-2008

Kirk-Davidoff DB, Hintsa EJ, Anderson JG, Keith DW. The effect of climate change on ozone depletion through changes in stratospheric water vapour. Nature. 1999; 402:399–401. DOI: 10.1038/46521

Kley, D.Russell, JM., III, Phillips, C., editors. SPARC Assessment of Upper Tropospheric and Stratospheric Water Vapour. World Meteorol Org.; Geneva: 2000.

Koop T, Lou B, Tsias A, Peter T. Water activity as the determinant for homogeneous ice nucleation in aqueous solutions. Nature. 2000; 406:611–614. [PubMed: 10949298]

Krämer M, et al. Ice supersaturations and cirrus cloud crystal numbers. Atmos Chem Phys. 2009; 9(11):3505–3522. DOI: 10.5194/acp-9-3505-2009

Lee J, Yang P, Dessler AE, Gao B-C, Platnick S. Distribution and radiative forcing of tropical thin cirrus clouds. J Atmos Sci. 2009; 66(12):3721–3731. DOI: 10.1175/2009jas3183.1

May RD. Open-path, near-infrared tunable diode laser spectrometer for atmospheric measurements of H2O. J Geophys Res. 1998; 103:19,161–19,172. DOI: 10.1029/98JD01678

Murphy DM, Koop T. Review of the vapour pressures of ice and supercooled water for atmospheric applications. Q J R Meteorol Soc. 2005; 131:1539–1565.

Oltmans SJ, Vömel H, Hofmann DJ, Rosenlof KH, Kley D. The increase in stratospheric water vapor from balloon-borne, frost point hygrometer measurements at DC and Boulder, Colorado. Geophys Res Lett. 2000; 27:3453–3456.

Perring AE, Schwarz JP, Gao RS, Heymsfeld AJ, Schmitt CG, Schnaiter M, Fahey DW. Evaluation of a perpendicular inlet for airborne sampling of interstitial submicron black-carbon aerosol. Aerosol Sci Technol. 2013; 47(10):1066–1072.

Peter T, Marcolli C, Spichtinger P, Corti T, Baker MB, Koop T. When dry air is too humid. Science. 2006; 314(5804):1399–1402. [PubMed: 17138887]

Podolske JR, Sachse GW, Diskin GS. Calibration and data retrieval algorithms for the NASA Langley/Ames Diode Laser Hygrometer for the NASA Transport and Chemical Evolution Over the Pacific (TRACE-P) mission. J Geophys Res. 2003; 108(D208792)doi: 10.1029/2002JD003156

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

Rollins AW, Thornberry TD, Gao RS, Hall BD, Fahey DW. Catalytic oxidation of H_2 on platinum: A robust method for generating low mixing ratio H_2O standards. Atmos Meas Tech. 2011; 4(10): 2059–2064. DOI: 10.5194/amt-4-2059-2011

Rothman LS, et al. The HITRAN 2008 molecular spectroscopic database. J Quant Spectros Radiat Transfer. 2009; 110:533–572. DOI: 10.1016/j.jqsrt.2009.02.013

Sargent MR, Sayres DS, Smith JB, Witinski M, Allen NT, Demusz JN, Rivero M, Tuozzolo C, Anderson JG. A new direct absorption tunable diode laser spectrometer for high precision measurement of water vapor in the upper troposphere and lower stratosphere. Rev Sci Instrum. 2013; 84:074102.doi: 10.1063/1.4815828 [PubMed: 23902086]

Sayres DS, et al. A new cavity based absorption instrument for detection of water isotopologues in the upper troposphere and lower stratosphere. Rev Sci Instrum. 2009; 80(4):044,102–044,114.

Scherer M, Vömel H, Fueglistaler S, Oltmans SJ, Staehelin J. Trends and variability of midlatitude stratospheric water vapour deduced from the re-evaluated Boulder balloon series and HALOE. Atmos Chem Phys. 2008; 8(5):1391–1402. DOI: 10.5194/acp-8-1391-2008

Shilling JE, Tolbert MA, Toon OB, Jensen EJ, Murray BJ, Bertram AK. Measurements of the vapor pressure of cubic ice and their implications for atmospheric ice clouds. Geophys Res Lett. 2006; 33:L17801.doi: 10.1029/2006GL026671

Sitnikov NM, Yushkov VA, Afachine AA, Korshunov LI, Astakhov VI, Ulanovskii AE, Krämer M, Mangold A, Schiller C, Ravegnani F. The FLASH instrument for water vapor measurements on board the high-altitude airplane. Instrum Exp Tech. 2007; 50(1):113–131. DOI: 10.1134/S0020441207010174

Solomon S, Rosenlof KH, Portmann RW, Daniel JS, Davis SM, Sanford TJ, Plattner G-K. Contributions of stratospheric water vapor to decadal changes in the rate of global warming. Science. 2010; 327(5970):1219–1223. DOI: 10.1126/science.1182488 [PubMed: 20110466]

Thomason LW, Moore JR, Pitts MC, Zawodny JM, Chiou EW. An evaluation of the SAGE III version 4 aerosol extinction coefficient and water vapor data products. Atmos Chem Phys. 2010; 10(5): 2159–2173. DOI: 10.5194/acp-10-2159-2010

Thornberry T, Gierczak T, Gao RS, Vömel H, Watts LA, Burkholder JB, Fahey DW. Laboratory evaluation of the effect of nitric acid uptake on frost point hygrometer performance. Atmos Meas Tech. 2011; 4(2):289–296. DOI: 10.5194/amt-4-289-2011

Thornberry TD, Rollins AW, Gao RS, Watts LA, Ciciora SJ, McLaughlin RJ, Voigt C, Hall B, Fahey DW. Measurement of low-ppm mixing ratios of water vapor in the upper troposphere and lower stratosphere using chemical ionization mass spectrometry. Atmos Meas Tech. 2013; 6(6):1461–1475. DOI: 10.5194/amt-6-1461-2013

Vogel B, Feck T, Grooß JU. Impact of stratospheric water vapor enhancements caused by CH_4 and H_2O increase on polar ozone loss. J Geophys Res. 2011; 116:D05301.doi: 10.1029/2010JD014234

Vömel H, David DE, Smith K. Accuracy of tropospheric and stratospheric water vapor measurements by the cryogenic frost point hygrometer: Instrumental details and observations. J Geophys Res. 2007a; 112:D08305.doi: 10.1029/2006JD007224

Vömel H, et al. Validation of Aura/MLS water vapor by balloon borne cryogenic frostpoint hygrometer measurements. J Geophys Res. 2007b; 112:D24S37.doi: 10.1029/2007JD008698

Webster CR, May RD, Trimble CA, Chave RG, Kendall J. Aircraft (ER-2) laser infrared absorption spectrometer (ALIAS) for in-situ stratospheric measurements of HCI, N_2O, CH_4, NO_2, and HNO_3. Appl Opt. 1994; 33(3):454–472. [PubMed: 20862038]

Weinstock EM, Hintsa EJ, Dessler AE, Oliver JF, Hazen NL, Demusz JN, Allen NT, Lapson LB, Anderson JG. New fast response photofragment fluorescence hygrometer for use on the NASA ER-2 and the Perseus remotely piloted aircraft. Rev Sci Instrum. 1994; 65(11):3544–3554.

Weinstock EM, et al. Validation of the Harvard Lyman-α in situ water vapor instrument: Implications for the mechanisms that control stratospheric water vapor. J Geophys Res. 2009; 114:D23301.doi: 10.1029/2009JD012427

Zöger M, et al. Fast in situ stratospheric hygrometers: A new family of balloon-borne and airborne Lyman α photofragment fluorescence hygrometers. J Geophys Res. 1999; 104(D1):1807–1816. DOI: 10.1029/1998JD100025

Key Points

- Agreement among in situ measurements of UT/LS water vapor is improved
- Scientifically significant disagreements between instruments still exist

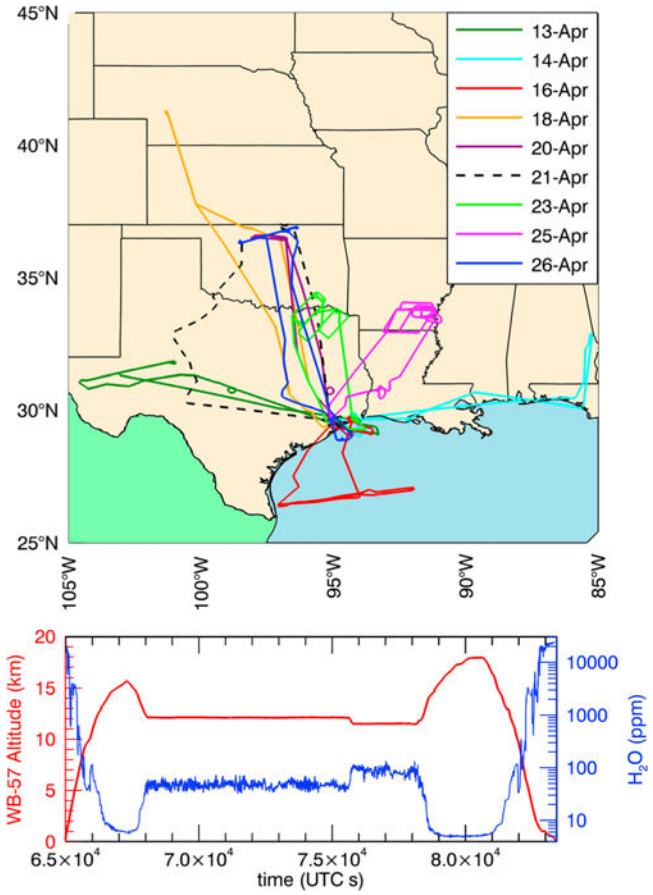

Figure 1.

(top) Map of MACPEX flight tracks for all flights included in the comparisons. (bottom) Time series of altitude and WV (DLH) mixing ratios during a typical MACPEX WB-57F flight on 23 April.

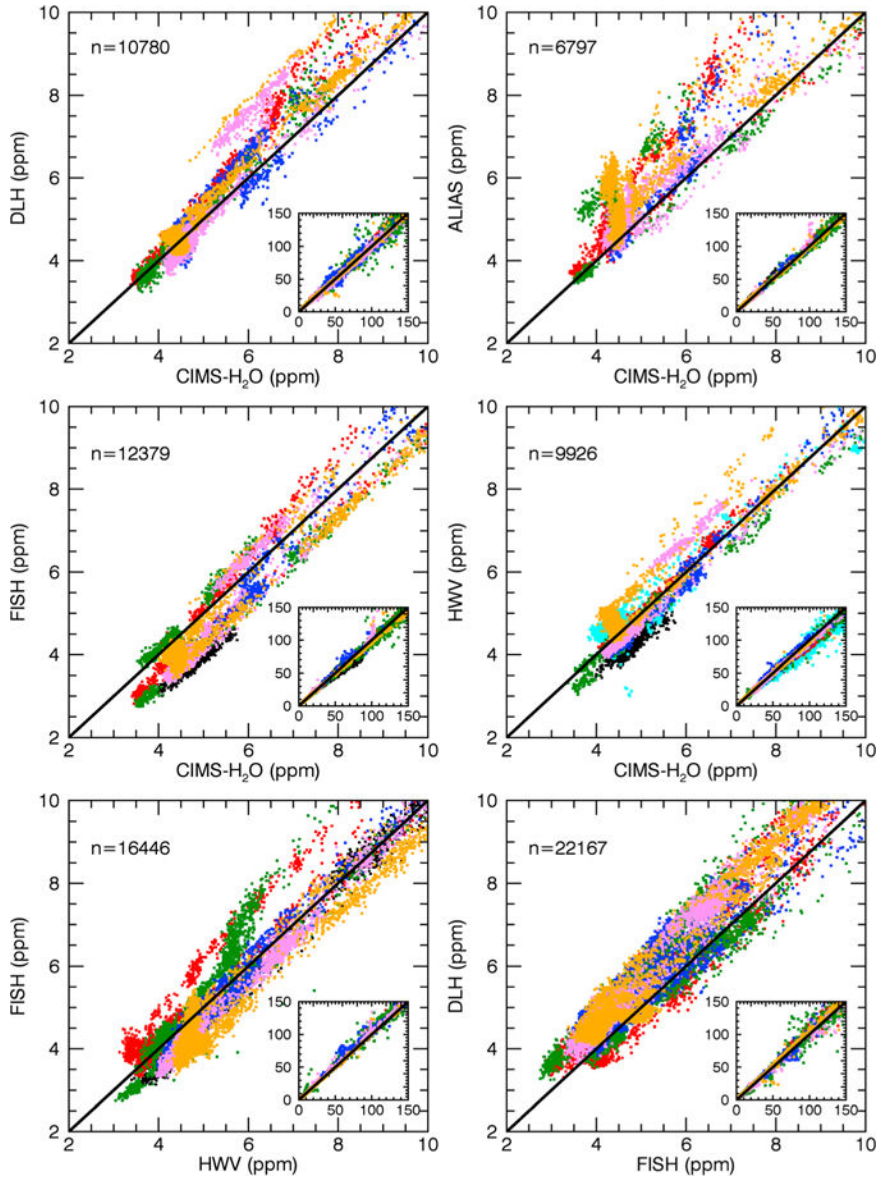

Figure 2.

Example correlation plots of DLH, ALIAS, FISH, HWV, and CIMS-H_2O. Main panels show the data in the 2–10 ppm range, and insets show all data up to 150 ppm. Different flight days from 14 April to 26 April are indicated using different color markers: 14 (black squares), 18 (cyan squares), 20 (red squares), 21 (green squares), 23 (blue squares), 25 (pink squares), and 26 (orange squares) April. The solid lines are the 1-1 lines. Regression statistics are given in Tables 3 and 4. Numbers indicated in each panel (e.g., n = 10,780) indicate the number of data points shown below 10 ppm.

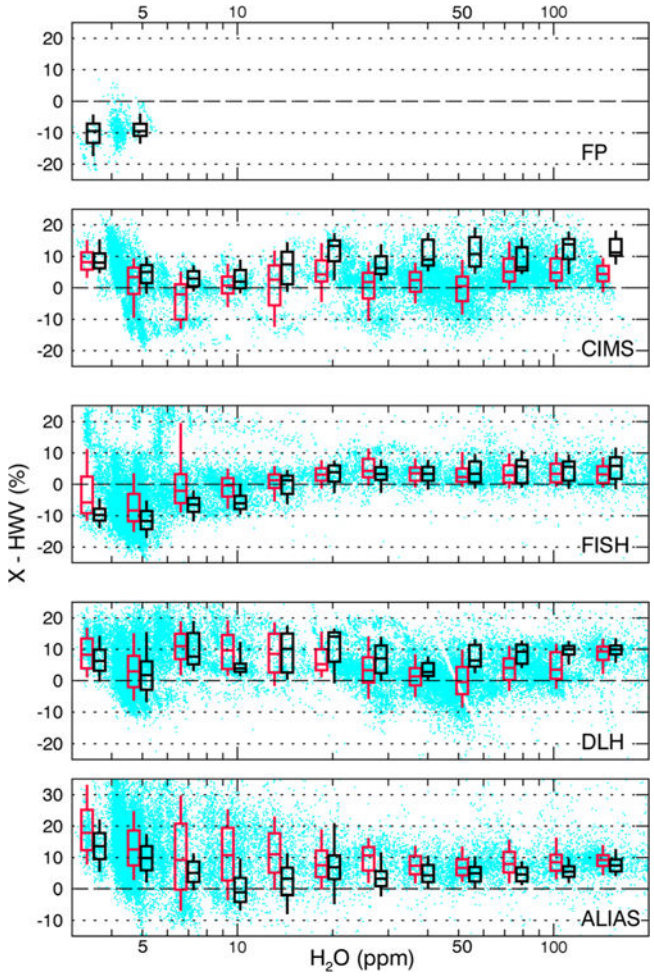

Figure 3.

Percent differences between HWV and FP, CIMS, FISH, DLH, and ALIAS as a function of WV mixing ratio (HWV data used for abscissa). Dots are all reported data in original time resolution from the flights in Table 2. Data are reported at 1 Hz for DLH, HWV, CIMS-H_2O, and FISH and at ~0.58 Hz for ALIAS. Boxes show median (horizontal line) and interquartile range. Whiskers show the 10–90 percentile range. Red boxes show statistics for all data, while black boxes show only descent data. Note that although different ranges are shown on the y axes, each panel shows a total range of 50%. FP data only shown for balloon descent above 15 km.

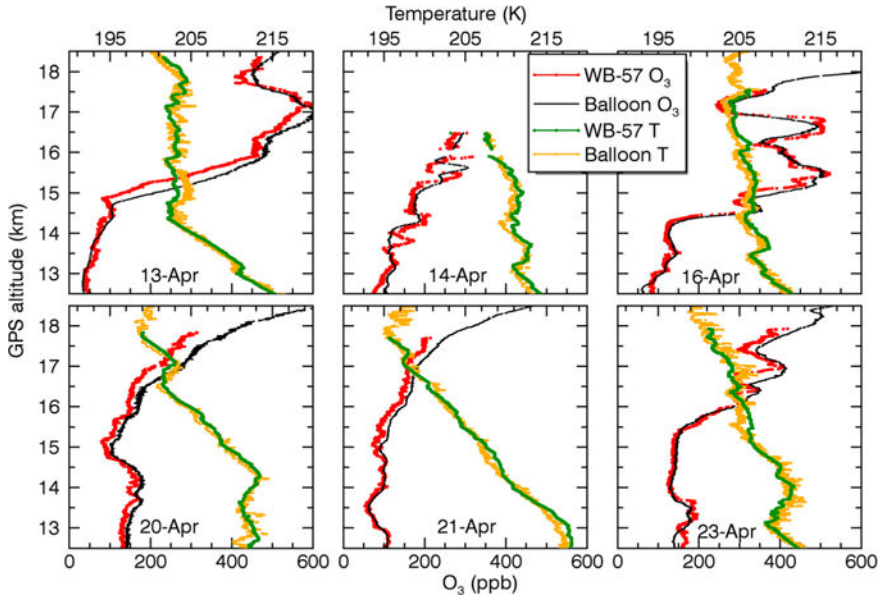

Figure 4.

Vertical profiles of temperature and ozone as measured by instruments on the WB-57F and balloon.

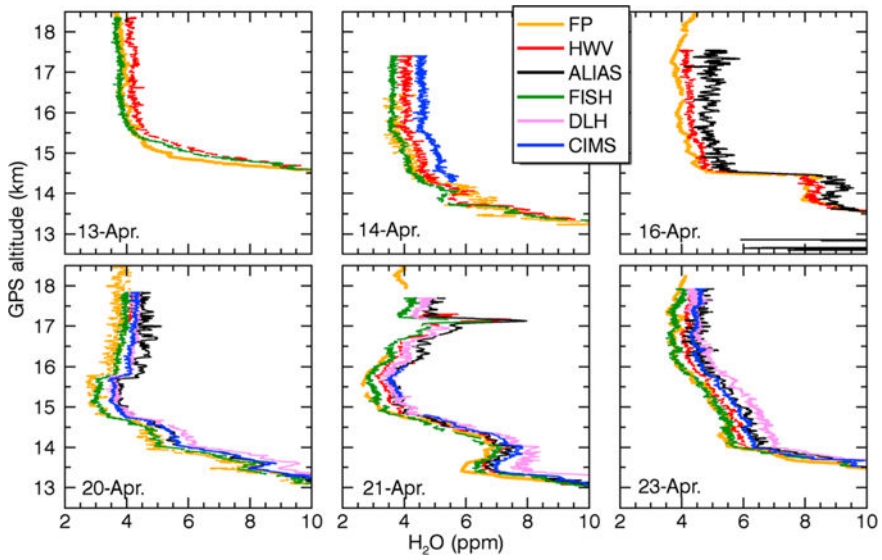

Figure 5.

Vertical profiles of WV during 13–23 April from days when FP measurements were available. Data shown are at original time resolution and are only from the final WB-57 descents. Frost point profiles are from CFH (14 April and 20 April) and FPH (13, 16, 21, and 23 April). FP data shown are only from the balloon descent.

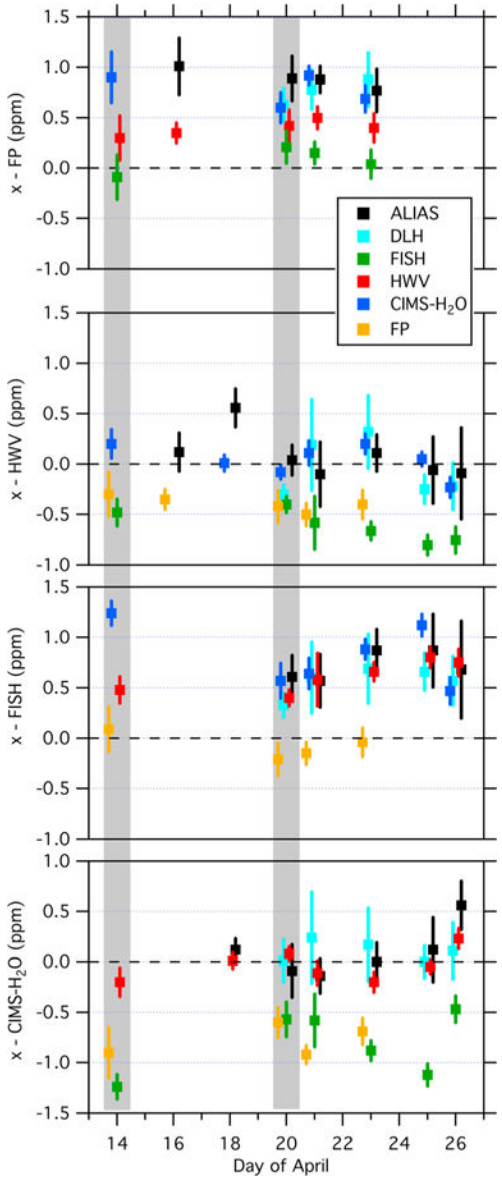

Figure 6.

Daily offsets between instruments using (from top) FP as a reference, HWV as a reference, FISH as a reference, and CIMS-H_2O as a reference. Data are from the final descents only. Gray regions indicate CFH flights. Other FP days are FPH. For each point, the marker and whiskers shows mean ± 1σ of the differences of the 1 s data.

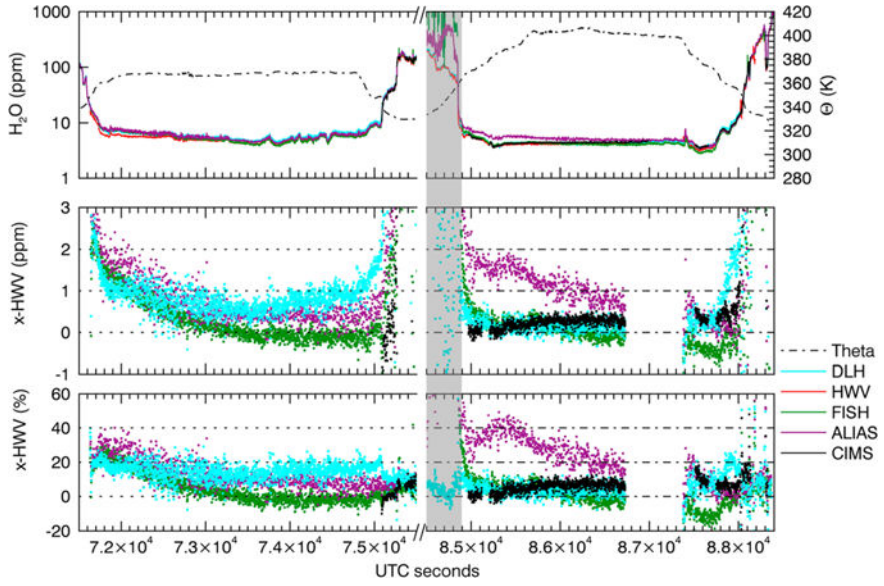

Figure 7.

Sample time series from the flight on 21 April during the two high-altitude legs. (top) Measured WV mixing ratios plotted on left axis and potential temperature (theta) on the right axis. (middle) Absolute differences from HWV (e.g., FISH – HWV) for FISH, ALIAS, CIMS-H_2O, and DLH with same colors as the top panel. (bottom) Data in middle panel plotted as percent differences from HWV (e.g., (FISH HWV)/HWV × 100%). Missing data from 86,700 to 87,400 s was due to a failure of the HWV data acquisition hardware. Gray region indicates cloud detected.

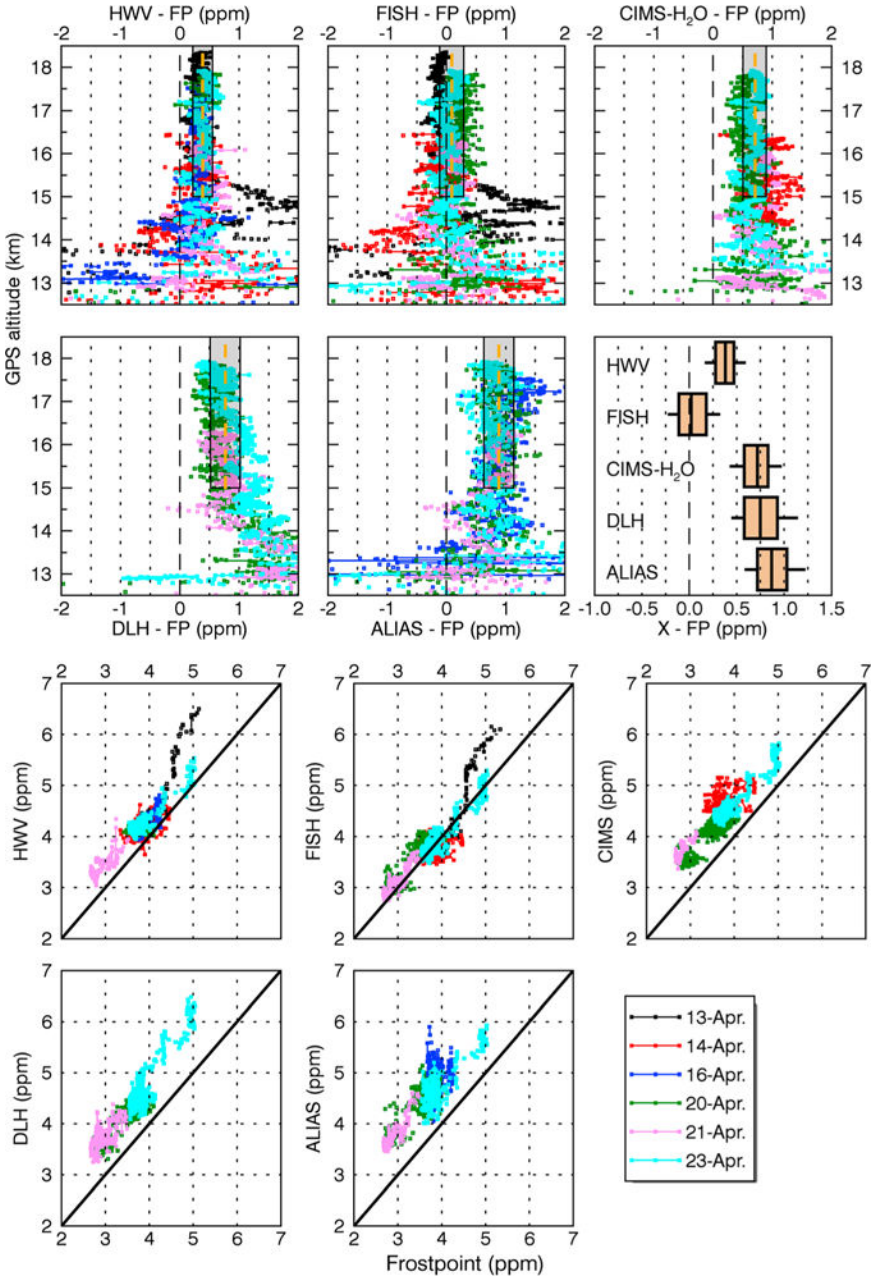

Figure 8.

Comparison between the frost point balloon measurements and FISH, CIMS-H_2O, HWV, ALIAS, and DLH. Data are for all FP flights in Table 5. Data are restricted to altitude above 15km. (top six panels) Difference between WB-57F and FP data. Orange dash shows average and gray box standard deviation. Summary panel (box and whiskers) shows 25–75 percentile (boxes) and 10–90 percentile (whiskers) for all measurements excluding 13 April data. Marker colors indicate flight day. (bottom six panels) Scatter plots comparing matched FP and WB-57F data. Solid lines indicate 1:1 line.

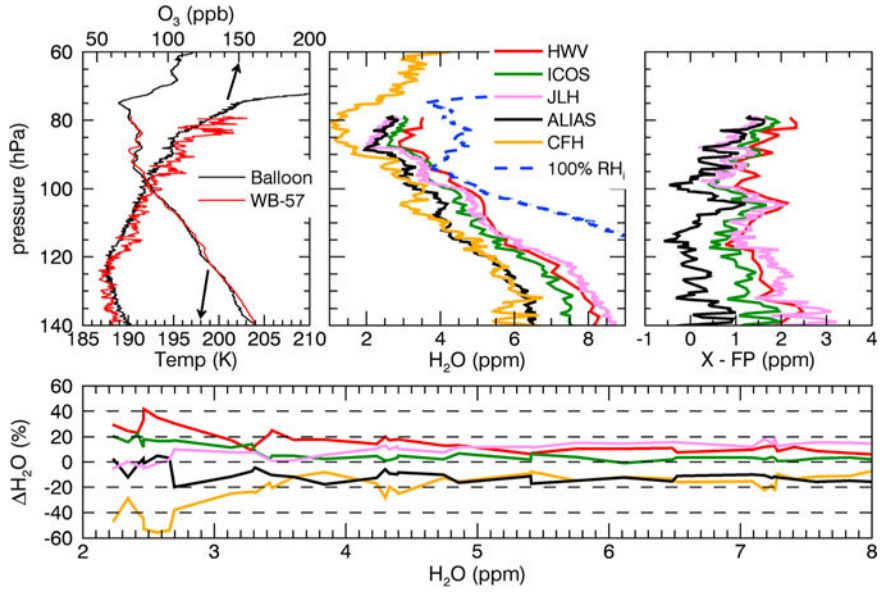

Figure 9.

Data from the 1 February 2006 flight during CR-AVE. (top left) Profiles of temperature (plotted on bottom axis) and O_3 (plotted on top axis) measured from the WB-57F (red) and balloon (black) show good correspondence between air masses sampled from these different platforms. (top middle) WV mixing ratios measured by the balloon-borne CFH instrument and four instruments on the WB-57F (HWV, ICOS, JLH, and ALIAS). Saturation mixing ratios were calculated using balloon temperature and pressure. (top right) Profiles showing differences between each of the WB-57F instruments and the FP balloon instrument. (bottom) Deviations from a mean profile (mean profile = equally weighted HWV, ICOS, JLH, ALIAS, and CFH) for HWV (red), ICOS (green), JLH (violet), ALIAS (black), and CFH (gold).

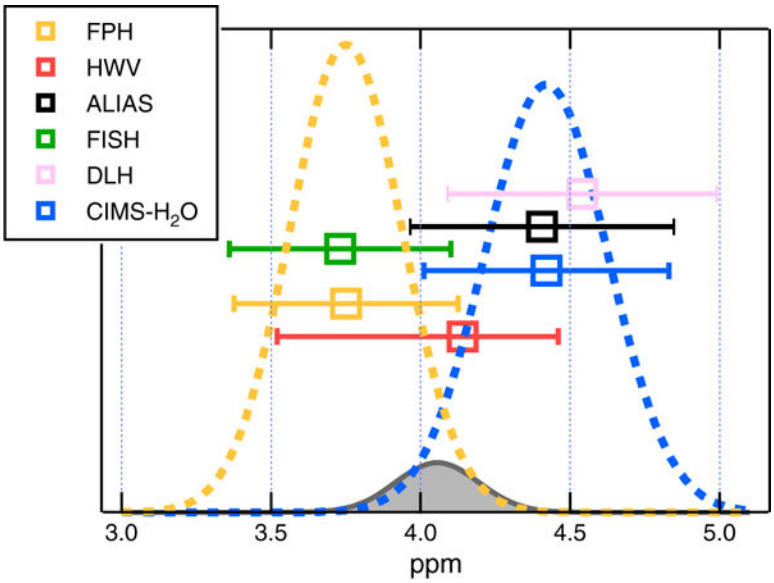

Figure 10.

Measurement spread from 23 April final descents from 16.7 to 16.5 km. Box and whiskers show measurement mean with stated uncertainty range. For this case all instrumental error estimates overlap near 4.1 ppmv. For FPH and CIMS the dashed lines show normal distributions about the mean values illustrating the PDF of WV values implied by the stated uncertainties. The product of the PDFs (gray) is the probability of agreement between FPH and CIMS-H_2O, which in this case is 5.7%. We note that the asymmetric uncertainties for HWV could not be approximated by a normal distribution as is done here for FPH and CIMS.

Table 1

Summary of Uncertainties for WV Instruments Used in the Intercomparison[a]

Instrument/Primary Measurement	Mixing Ratio Accuracy[b]	Sources of Uncertainties
Chemical Ionization Mass Spectrometer (CIMS-H_2O)[c]/mixing ratio	≈ ±10%	*Systematic errors:* Associated with standard addition of WV calibration mixing ratios and the application of calibration data to calculate the ambient mixing ratios < ±2% error from H_2 standards < ±3% error in the dynamic dilution of H_2 standards <−1% bias from converting H_2 to H_2O < ±4% error from interpolating between successive in-flight calibrations *Background artifact:* < + 0.1ppm artifact from outgassing from inlet materials as measured in flight with zero air addition < ± 0.1ppm artifact from uncertainty in the WV content of zero air *Interferences:* H_3O+ analyte ion production efficiency is affected by ambient CO_2. Calibrations performed in zero air containing 385 ppm CO_2 reduce this effect to < ± 0.5%
Harvard Water Vapor Lyman-α (HWV)[d]/mixing ratio	+7.5% −(7.5%+ 0.3 ppm)	*Systematic errors:* Laboratory measurements of calibration constant ±7.5% due to uncertainties from producing WV calibration standards and a drift in the empirically determined calibration constants over the course of the mission *Background artifact:* Outgassing from inlet materials has been evaluated to be insignificant by varying flow speed through duct during aircraft in-flight measurements < + 0.3 ppm from using the driest available zero air for calibrations *Interferences:* No known interferences
Fast In Situ Stratospheric Hygrometer Lyman-α (FISH)[e]/mixing ratio	±(6% + 0.15ppm)	*Systematic errors:* Laboratory measurements of calibration constants ±4% due to uncertainty of reference hygrometer used to quantify calibration mixing ratios. Additional 2% due to uncertainties of the stability of calibration factors throughout campaigns *Background artifact:* Outgassing from inlet materials assumed to be zero *Interferences:* No known interferences
Diode Laser Hygrometer (DLH)[f]/mixing ratio	±10%	*Systematic errors*: Uncertainty in laser modulation depth relative to absorption linewidth; absorption line spectroscopic parameters, ambient pressure (±03 hPa), and temperature (±03 K) *Background artifact:* Persistent moisture in internal portion of optical path; artifact largely removed spectroscopically. Harmonic baseline due to nonlinearity in power modulation; artifact largely removed via periodic spectral scans *Interferences:* No known interferences
Aircraft Laser Infrared Absorption Spectrometer (ALIAS)[g]/mixing ratio	±10%	*Systematic errors*: Spectroscopic parameters and ambient pressure and temperature *Background artifact:* Outgassing of inlet and detection cell surfaces *Interferences:* No known interferences
Cryogenic Frost Point Hygrometer (CFH) and NOAA Frost Point Hygrometer (FPH)[h]/frost point temperature	±7–10%	*Systematic errors*: Random and systematic uncertainties in determination of frost temperature, totaling in 0.51°C uncertainty. Uncertainty in total pressure of 0.5 hPa. For stratospheric frost points and pressures, these total to 7–10% uncertainty in mixing ratio *Background artifact*: Assumed to be zero *Interferences:* No known interferences

[a] Reported instrument accuracy estimates and known potential sources of error in the WV measurements in three categories: (1) intrinsic sources of systematic errors, (2) potential background artifacts, and (3) potential chemical interferences. Details of the instruments and investigation of their uncertainties are described in the footnoted reference(s). Uncertainties stated as "<" indicate the maximum absolute value of the error.

[b] As provided by the investigator in the indicated reference(s).

[c] *Thornberry et al.* [2013].

[d] *Weinstock et al.* [2009].

[e] *Zöger et al.* [1999].

[f] *Diskin et al.* [2002]; *Podolske et al.* [2003].

[g] *Webster et al.* [1994].

[h] *Vomel et al.* [2007].

Table 2

Summary of Available Data During MACPEX[a]

Date	HWV	CIMS-H_2O	FISH	DLH	ALIAS	Balloon
13 Apr	X		X		X	FPH
14 Apr	X	X	X		X	CFH
16 Apr	X				X	FPH
18 Apr	X	X			X	
20 Apr	X	X	X	X	X	CFH
21 Apr	X	X	X	X	X	FPH
23 Apr	X	X	X	X	X	FPH
25 Apr	X	X	X	X	X	
26 Apr	X	X	X	X	X	

[a]Available data are indicated by X. Under the balloon column, CFH and FPH indicate which device was used.

Table 3

Summary of Comparisons Between In Situ Instruments[a]

	HWV	CIMS	FISH	DLH
	Total number of points for comparison (< 150ppm)			
CIMS	23216			
FISH	27254	15810		
DLH	31701	27822	29931	
ALIAS	23249	12290	23137	21321
	Range of correlation slopes for individual flights (min, max)			
CIMS	1.014,1.079			
FISH	1.008,1.077	0.961,1.083		
DLH	1.015,1.109	0.975, 1.032	1.006, 1.058	
ALIAS	1.037, 1.098	0.972, 1.060	0.975, 1.085	0.946, 1.009
	Average ± standard deviation of correlation slopes for individual flights (mean ± std)			
CIMS	1.046 ±0.023			
FISH	1.048 ±0.028	1.024± 0.051		
DLH	1.057 ±0.036	1.002 ± 0.022	1.031 ± 0.020	
ALIAS	1.070 ±0.023	1.013 ± 0.033	1.012 ± 0.035	0.976±0.026

[a]Data in the range 0–150 ppm from all flights. Column headings are the abscissa, and row headings are the ordinate (e.g., first comparison is CIMS versus HWV). Data points are 1 s measurements except for ALIAS (~1.7 s).

Table 4

Summary of Offsets Calculated for Instrument Pairs[a]

	HWV	CIMS	FISH	DLH
Range of offsets for individual flights (min, max in ppm)				
CIMS	−0.23, 0.20			
FISH	−0.80, −0.40	−1.12, −0.47		
DLH	−0.30, 0.32	0.00, 0.24	0.33, 0.69	
ALIAS	−0.10, 0.24	−0.14, 0.56	0.57, 0.87	0.00, 0.32
Average ± standard deviation of offsets for individual flights (ppm)				
CIMS	0.00 ±0.16			
FISH	−0.63 ±0.16	−0.77 ±0.29		
DLH	−0.03 ±0.27	0.11 ±0.10	0.55 ±0.15	
ALIAS	0.04±0.12	0.12 ±0.27	0.72 ±0.13	0.15 ±0.12

[a]Comparison uses data less than 10 ppm from all flights. The sign of the offsets are determined as row headings minus column headings (e.g., first comparison is CIMS-HWV).

Table 5

Summary of Comparisons With the FP Measurements[a]

	HWV	FISH	CIMS-H_2O	DLH	ALIAS
14 Apr	0.30 ±0.23	−0.09 ±0.23	0.90±0.26		
16 Apr	0.35 ±0.11				1.01 ±0.29
20 Apr	0.42 ±0.17	0.21 ±0.17	0.60±0.16	0.63 ± 0.17	0.89 ±0.23
21 Apr	0.50 ±0.17	0.15 ±0.12	0.92 ±0.10	0.78 ±0.20	0.88 ±0.14
23 Apr	0.40±0.14	0.04±0.15	0.69±0.14	0.88 ±0.27	0.77 ±0.22
Mean difference (all data)	0.38±0.16	0.04±0.21	0.71 ±0.20	0.77 ± 0.27	0.89 ±0.24

[a]Data only from final descents above 15 km as shown in Figure 8. Values are reported, for example, as HWV minus FP values in ppm.