

Analysis of Raman lidar and radiosonde measurements from the AWEX-G field campaign and its relation to Aqua validation

D. N. Whiteman,¹ F. Russo,² B. Demoz,¹ L. M. Miloshevich,³ I. Veselovskii,⁴ S. Hannon,² Z. Wang,⁵ H. Vömel,⁶ F. Schmidlin,¹ B. Lesht,⁷ P. J. Moore,⁸ A. S. Beebe,^{9,10} A. Gambacorta,² and C. Barnet¹¹

Received 27 June 2005; revised 14 September 2005; accepted 8 November 2005; published 12 April 2006.

[1] Early work within the Aqua validation activity revealed there to be large differences in water vapor measurement accuracy among the various technologies in use for providing validation data. The validation measurements were made at globally distributed sites making it difficult to isolate the sources of the apparent measurement differences among the various sensors, which included both Raman lidar and radiosonde. Because of this, the AIRS Water Vapor Experiment–Ground (AWEX-G) was held in October–November 2003 with the goal of bringing validation technologies to a common site for intercomparison and resolving the measurement discrepancies. Using the University of Colorado Cryogenic Frostpoint Hygrometer (CFH) as the water vapor reference, the AWEX-G field campaign permitted correction techniques to be validated for Raman lidar, Vaisala RS80-H and RS90/92 that significantly improve the absolute accuracy of water vapor measurements from these systems particularly in the upper troposphere. Mean comparisons of radiosondes and lidar are performed demonstrating agreement between corrected sensors and the CFH to generally within 5% thereby providing data of sufficient accuracy for Aqua validation purposes. Examples of the use of the correction techniques in radiance and retrieval comparisons are provided and discussed.

Citation: Whiteman, D. N., et al. (2006), Analysis of Raman lidar and radiosonde measurements from the AWEX-G field campaign and its relation to Aqua validation, *J. Geophys. Res.*, *111*, D09S09, doi:10.1029/2005JD006429.

1. Introduction and Background

[2] The Aqua satellite validation activity funded by NASA includes the use of different water vapor profiling radiosondes and Raman lidar systems for acquisition of measurements during Aqua overpasses. Numerous special measurement campaigns have been staged from various geographic locations in order to acquire data of the highest quality for calibration and validation of the satellite mea-

surements and retrievals. It is fundamentally important that these special data sets possess higher absolute accuracy than required of the satellite data products for this validation technique to work. Early comparisons of many validation measurements with the Atmospheric Infrared Sounder (AIRS), through the use of the AIRS fast forward radiative transfer model, SARTA [Strow *et al.*, 2003], revealed apparent large calibration differences among the various water vapor profiling technologies being used. The differences were largest in the upper troposphere (UT) where differences between AIRS radiances and calculations of AIRS radiance using SARTA, when translated to UT relative humidity (RH), implied differences in the calibration of the water vapor measurement systems that exceeded 25% in some cases. This is to be contrasted with the Aqua retrieval accuracy goal, where a retrieval involves a minimization of differences between observed and calculated radiances, of 10% in 2-km layers. The apparent inadequacy of many of the validation measurement systems to provide data of sufficient quality to validate retrievals at this accuracy level created questions both about the validation sensor technologies and how to improve the quality of water vapor measurements used for Aqua validation. For this reason, a dedicated field program called the AIRS Water Vapor Experiment–Ground (AWEX-G) was held in October–November 2003 with the goal of resolving the measurement

¹NASA/Goddard Space Flight Center, Greenbelt, Maryland, USA.

²Department of Physics, University of Maryland Baltimore County, Baltimore, Maryland, USA.

³National Center for Atmospheric Research, Boulder, Colorado, USA.

⁴Goddard Earth Science and Technology Center, University of Maryland Baltimore County, Baltimore, Maryland, USA.

⁵Department of Atmospheric Sciences, University of Wyoming, Laramie, Wyoming, USA.

⁶Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA.

⁷U.S. Department of Energy/Argonne National Laboratory, Chicago, Illinois, USA.

⁸LJT and Associates, Wallops Island, Virginia, USA.

⁹SGT, Inc., Greenbelt, Maryland, USA.

¹⁰Now at Department of Mathematics, University of Maryland Eastern Shore, Princess Anne, Maryland, USA.

¹¹NOAA, Camp Springs, Maryland, USA.

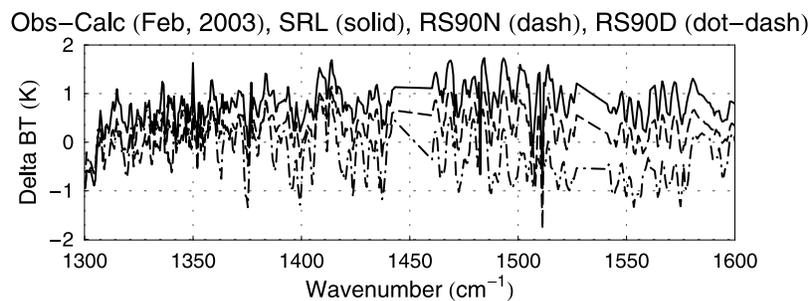


Figure 1. Mean comparisons of radiances observed by AIRS and calculations of radiances using version 3 of the AIRS fast model, SARTA, to demonstrate the state of the AIRS fast model validation effort as of February 2003. Three sets of water vapor input data acquired in coordination with Aqua overpasses are studied: (1) SRL water vapor measurements from fall 2002 at GSFC, (2) nighttime RS-90 radiosondes launched at the SGP site, and (3) daytime RS-90 radiosondes launched at the SGP site.

differences observed among Vaisala radiosonde and Raman lidar and to develop new analytical tools to improve the absolute accuracy of those measurement systems.

[3] This paper provides the motivation for the AWEX-G field campaign, discusses the field activities and the major results of the field activity and then puts those results in the context of Aqua validation. It is organized as follows. Early AIRS radiance validation comparisons are presented to illustrate some of the first discrepancies that were uncovered in the validation activity and that helped to motivate AWEX-G. The AWEX-G field campaign is then described and the major results summarized. This paper will focus mostly on the Raman lidar measurements and results, which included corrections to Raman lidar water vapor measurements that account for the temperature dependence of Raman scattering. A companion paper [Miloshevich *et al.*, 2006] provides the details of the radiosonde intercomparisons and radiosonde accuracy assessment that occurred during AWEX-G and correction techniques for Vaisala RS80-H and RS90/92 measurements that were derived from AWEX-G measurements. The radiosonde and lidar sensors are compared here both in terms of profiles and layer mean upper tropospheric precipitable water. These results are compared with the corresponding results from a similar lidar/radiosonde intercomparison experiment that was held in 2000. The effect of the new Raman lidar and radiosonde corrections on Aqua validation activities is then demonstrated using examples of both radiance and retrieval comparisons.

2. Early Discrepancies Between Raman Lidar and Radiosonde Measurements and AIRS

[4] NASA has funded special launches of Vaisala RS90 radiosondes to coincide with Aqua satellite overpasses of the Department of Energy (DOE) Atmospheric Radiation Measurements (ARM) facilities on the North Slope of Alaska (NSA), the Southern Great Plains (SGP) and the Tropical Western Pacific (TWP). A total of 90 overpasses were targeted in each of three sets of these special launches with the goal of providing validation data in a variety of seasons and from different geographic locations [Tobin *et al.*, 2006]. For each targeted satellite overpass, the goal was to launch two radiosondes separated in time so that one was in the upper troposphere and one in the lower troposphere at

the actual time of overpass. Using these sonde measurements and other ancillary data, a best estimate (BE) product is generated that performs a temporal and spatial interpolation over an Aqua retrieval region [Tobin *et al.*, 2006]. The BE database is one of the main sources of validation data for the Aqua validation activity.

[5] Other validation measurement campaigns were held including one at NASA/Goddard Space Flight Center (GSFC) in the fall of 2002 involving water vapor, temperature and pressure measurements coordinated with 26 nighttime overpasses of the Aqua satellite under both clear and partially cloudy conditions. For these measurements, the Scanning Raman Lidar (SRL) was used for the water vapor profiles, Sippican radiosonde for temperature and pressure profiles and SuomiNet Global Positioning System (GPS) [Ware *et al.*, 2000] for the total precipitable water. The Raman lidar profiles were calibrated so that the integrated water vapor amount agreed with the total precipitable water measured by GPS [Whiteman *et al.*, 2006a].

[6] The BE product generated from the first set of the special Aqua radiosonde launches and the GSFC validation data were among those validation data sources studied in early 2003 using version 3 of SARTA. Figure 1 presents comparisons of the mean brightness temperature differences between AIRS measurements (denoted “Obs”) and the output of version 3 of SARTA (denoted “Calc”), where the water vapor input to SARTA was one of the validation data sets, for three different sets of AIRS observations. The first comparison, shown with a solid line, is the mean difference (in degrees K of equivalent brightness temperature) between 82 AIRS fields of view (FOV) from 15 different overpasses of NASA/GSFC between September and November 2002 and the SARTA-calculated brightness temperature on the basis of the corresponding GSFC SRL water vapor and Sippican radiosonde temperature measurements. The second, shown with a dashed line, is the mean difference between approximately 410 AIRS FOVs from approximately 75 different nighttime overpasses of the ARM SGP site in late 2002 and the calculations based on the BE product using Vaisala RS90 data as water vapor input. The third set of measurements, shown in a dash-dot style, is similar to the second set except that it focusses on daytime overpasses of SGP during the same time period. Manual cloud-clearing has been performed on these ensembles.

[7] In general the water vapor line strengths increase moving toward higher wavenumbers in the water band shown in Figure 1. Therefore progressing from lower to higher wavenumbers, the equivalent brightness temperature, B_T , for spectral locations that correspond to the centers of water vapor absorption lines will increasingly be influenced by the amount of water vapor present in the upper troposphere and lower stratosphere. Between the lines, where absorption is lower, the brightness temperatures reflect water vapor concentrations lower in the atmosphere. Thus, given perfect knowledge of trace gas and continuum absorption spectra and accurate radiative transfer calculations based on them, differences in B_T in Figure 1, either on the line centers or in between the lines, can be taken to represent differences in the amount of water vapor inducing the radiances measured by AIRS (Obs) and the water vapor measured by the validation sensor (Calc).

[8] Several points can be made from Figure 1. The first is that approximately a 25% range in apparent water vapor calibration is implied by the $\sim 2\text{K}$ uncertainty observed through much of the water band using the approximate rule of thumb that a 1K difference in brightness temperature in the upper troposphere corresponds to a relative difference in UT water vapor amount of approximately 12% [Soden *et al.*, 2000] (12% absolute, not to be confused with a difference of 12% RH). It should be mentioned here that Obs-Calc comparisons based on other Aqua validation data sets (not shown), some acquired using Vaisala RS90s as well, implied even larger uncertainties in apparent water vapor calibration. The correlation of the high-frequency structure in all the measurements seen in Figure 1 implied, considering that the radiosonde and lidar measurements were acquired from different locations and at different times, that there likely were errors in the absorption cross sections of the water vapor lines used in SARTA. However, given the 2K range of brightness temperature differences displayed in Figure 1, it was not clear what should be used as validation data to help isolate the spectroscopy errors and to judge the overall accuracy of SARTA calculations. The questions of what data source to use for AIRS validation seemed largest in the upper troposphere where cold temperatures make reliable radiosonde measurements of water vapor more difficult and since the disagreements of the measurements shown in Figure 1 were largest in the higher-wavenumber portion of the water band, which corresponds roughly to the upper troposphere.

[9] These large apparent discrepancies in water vapor measurements were contrasted, however, with previous field mission experience that indicated that it was possible for several water vapor sensors to be used in a coordinated fashion over a period of several weeks from the same location and achieve mean upper tropospheric water vapor calibration in agreement at the $\pm 5\%$ level. This was the result of the ARM–First International Satellite Cloud Climatology Project (ISCCP) Regional Experiment (FIRE) Water Vapor Experiment (AFWEX), which was held in the fall of 2000 at the DOE ARM SGP Climate Research Facility (CRF) site in northern Oklahoma [Ferrare *et al.*, 2004].

[10] The sensors that agreed at the 5% level from AFWEX were the airborne NASA LaRC Lidar Atmospheric Sensing Experiment (LASE) water vapor lidar and Diode Laser Hygrometer (DLH), the Vaisala RS80-H radiosonde

(after application of corrections for calibration bias and sensor time lag), the DOE CRF (Climate and Radiation Facility) Raman Lidar (CARL) and the SRL. There were several reasons, however, that AFWEX results could not be applied to help resolve the measurement discrepancies illustrated in Figure 1. The SRL hardware configuration changed between the time of AFWEX and the fall of 2002 when the Aqua validation measurements shown in Figure 1 were performed. Therefore there was uncertainty about the current calibration of the SRL upper tropospheric water vapor measurements. Also, the Vaisala RS90 radiosonde was not tested during AFWEX. The RS90 calibration accuracy, time response, and susceptibility to sensor icing in clouds is improved substantially over those of the RS80-H and the RS90 has had changes in its calibration model in the time between AFWEX and the present [Miloshevich *et al.*, 2006]. The CARL lidar could also not be used since its performance was found to have degraded significantly since the time of AFWEX calling into question any extrapolations based on CARL measurements. The AWEX-G field campaign thus was held with the goal of providing AFWEX-quality accuracy assessments of water vapor measurement technologies in the Aqua era.

3. AIRS Water Vapor Experiment–Ground

[11] Drawing on the success of the AFWEX field campaign, the AIRS Water Vapor Experiment–Ground (AWEX-G) was held at the ARM SGP site between 27 October and 16 November 2003 in order to accomplish the following:

[12] 1. Bring a majority of the water vapor measurement technologies in use for Aqua and AIRS validation together at the ARM SGP site for intercomparison. The technologies that were used during AWEX-G included radiosondes (of various technologies including frostpoint hygrometer), Raman lidars, GPS (Global Positioning System) and microwave radiometer (MWR).

[13] 2. Operate these instruments over a 3-week period of time focussing on nighttime, clear weather conditions including many instances of multiple radiosondes launched on the same balloon. The nighttime period was chosen so as to permit Raman lidar measurements to extend into the upper troposphere and also to avoid the issues relating to daytime heating of radiosondes that can create a dry bias in the measurements [Miloshevich *et al.*, 2006].

[14] 3. Characterize the measurement differences of the water vapor technologies and use this information to better understand the existing differences in the Aqua validation activity and to develop schemes for improving the accuracies of those measurements.

[15] The “G” designation in the name AWEX-G indicated the possibility that ground-based measurements alone might not resolve the measurement discrepancies that motivated AWEX-G and that it might be necessary in the future to hold another experiment involving airborne instrumentation. The fall period was chosen for AWEX-G since this is the season that offers the highest probability of clear skies because of the frequency of strong frontal passages. It also was the season during which AFWEX occurred.

[16] During AWEX-G, 56 balloons carrying 112 radiosonde packages were launched. The radiosonde technolo-

gies that were tested were Vaisala RS80-H, RS90, RS92, University of Colorado–Cryogenic Frost point Hygrometer CFH) [Vömel *et al.*, 2003], Meteorolabor SnowWhite hygrometer [Wang *et al.*, 2003], Modem GL98 (manufactured in France) and Sippican Mark Ila (formerly VIZ). In addition, more than 40 hours of Scanning Raman Lidar measurements of water vapor were acquired in coordination with the radiosonde launches. The CARL Raman Lidar [Turner and Goldsmith, 1999] ran continuously through the experiment as did the CRF Microwave Radiometer (MWR) and a SuomiNet GPS system that was deployed along with the SRL and was the source of calibration of the SRL profiles from the fall 2002 measurements used in Figure 1.

[17] The primary goal of AWEX-G was to intercompare water vapor technologies and not to provide a statistically meaningful set of validation measurements during Aqua overpasses. Nonetheless, AWEX-G operations were coordinated with nighttime Aqua overpasses as much as possible. A period of heightened solar activity caused a shutdown of the AIRS sensors between 29 October and 2 November, however. It was not until 19 November that AIRS was recalibrated and declared operational. Therefore there was only one useful Aqua overpass during the AWEX-G experiment.

[18] In the following subsections concerning AWEX-G, the Scanning Raman Lidar and the AWEX-G reference instrument, the CFH, are briefly described. Then there is a comparison of GPS and MWR total precipitable water measurements. The corrections for both Raman lidar and Vaisala radiosonde that were derived from the AWEX-G field campaign are then summarized followed by comparisons of corrected lidar profiles with fully corrected Vaisala RS90/92 radiosondes. Mean percentage differences in upper tropospheric water vapor measurements are studied for Vaisala radiosondes, SnowWhite and SRL using three techniques of comparison.

3.1. NASA/GSFC Scanning Raman Lidar (SRL)

[19] The SRL is a mobile Raman lidar system designed to measure water vapor, aerosols, clouds and other quantities. It was first deployed in the field in 1991 for the Spectral Radiance Experiment (SPECTRE) [Ellingson and Wiscombe, 1996] in Coffeyville, Kansas, sponsored by DOE and NASA. It has received numerous upgrades since that time and now consists of a Nd:YAG laser operating at the tripled frequency of 354.7 nm, a 0.76 m horizontally mounted telescope that is coaligned with a single axis scanner that permits horizon to horizon measurements. The 0.76 m telescope is used for the high-altitude measurements, while a second, coaligned 0.25 m telescope mounted inside of the 0.76 m telescope measures aerosol depolarization and the low-altitude Raman signals. Measurements of water vapor and other quantities are performed during the day and night using the narrow field of view, narrow spectral band technique [Whiteman *et al.*, 2006a]. The SRL was in essentially the same experimental configuration as during the International H₂O Project_2002 (IHOP) that occurred in May–June 2002. Many more details of the instrumental configuration of the SRL and the analysis techniques used to process the data can therefore be found in references for the IHOP field experiment [Whiteman *et al.*, 2006a, 2006b].

3.2. AWEX-G Water Vapor Reference: CFH

[20] The CFH is an improved version of the NOAA cryogenic frostpoint hygrometer [Vömel *et al.*, 1995; Miloshevich *et al.*, 2006], which has long been a standard for balloon-borne stratospheric water vapor measurement. Its operation is based on the chilled mirror principle, where a small mirror is electrically heated or cryogenically cooled to maintain a constant thin layer of frost that is optically detected, in which case the frost layer is in equilibrium with the environment and the mirror temperature is equal to the frostpoint temperature of the air. The mirror temperature is measured by a tiny thermistor embedded in the surface of the mirror. During AWEX-G, it was always launched with Vaisala RS80-H and Meteorolabor SnowWhite packages attached. Most of the launches included a Vaisala RS92 sensor as well. An analysis of the measurement errors of the CFH [Miloshevich *et al.*, 2006] indicates that the mean percentage uncertainty in the CFH RH measurements over the AWEX-G temperature range is approximately 4% when RS92 temperature measurements were used to convert the frostpoint measurements to RH and approximately 6% when RS80-H temperature measurements were used. This reduction in absolute error when using the RS92 temperatures is due to the smaller mean temperature uncertainty in the RS92 radiosonde versus the RS80-H (~0.2 versus ~0.5 K).

3.3. SuomiNet GPS PWV Comparison With ARM MWR

[21] The technique that has been adopted for calibrating both Vaisala radiosonde [Turner *et al.*, 2003] and Raman lidar [Turner *et al.*, 2002] within the U.S. DOE ARM program is to constrain the vertical profile of water vapor so that it possesses the same total precipitable water vapor (PWV) as that measured by collocated MWR. Research done within the DOE ARM program indicates that carefully calibrated and analyzed microwave radiometer data possess an accuracy of approximately 2–3% or 0.4 mm, whichever is larger [Liljegren *et al.*, 2005]. Therefore, under dry conditions the absolute accuracy of the MWR PWV measurement can rise above the 2–3% figure stated but for measurements with PWV exceeding approximately 13 mm the accuracy specification should hold. The high accuracy of the MWR under most atmospheric conditions encountered at a location like the ARM SGP site makes it an excellent calibration standard for atmospheric research.

[22] A SuomiNet GPS system measuring total precipitable water was used in a similar fashion for total column water calibration of the SRL during the fall 2002 validation measurements used in Figure 1. One of the AWEX-G research objectives, therefore, was to compare the measurements of the GPS system with that of the ARM MWR and investigate if biases in the GPS measurements might help to explain the differences observed in Figure 1 between SARTA calculations based on SRL water vapor profiles and AIRS observations. The GPS measures over a much larger volume than the MWR so that individual comparisons can show considerable disagreement under conditions of atmospheric inhomogeneity. Line of site comparisons of the two instruments have been performed to address these differences in measurement techniques and have shown excellent agreement [Braun *et al.*, 2003]. During AWEX-G,

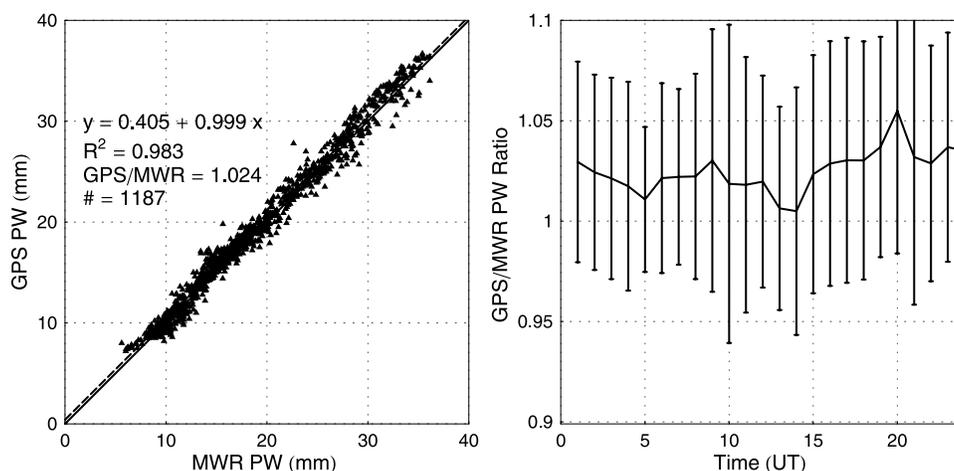


Figure 2. (left) An analysis of the precipitable water measurements from the SuomiNet GPS and the U.S. DOE microwave radiometer at the Southern Great Plains Climate Research Facility. There is a mean offset of $\sim 2.4\%$, which is within the absolute uncertainty of the MWR. The best fit regression is shown with a dashed line. The solid line indicates perfect 1-1 correlation. (right) Same data as plotted in Figure 2 (left) except this time divided by time of day. The data are grouped in bins of 1 hour each. The error bars indicate the standard deviation of the comparison within each 1 hour interval. Sunrise and sunset were at approximately 1200 and 2400 UT.

an approximately month-long comparison of 30-min average GPS and MWR vertical precipitable water measurements was performed in order to minimize the effects of short-term spatial inhomogeneities. The results of that comparison are shown in Figure 2.

[23] Figure 2 (left) shows that the PWV from GPS was on average 2.4% higher than MWR because of an offset of ~ 0.4 mm in the best fit regression. However, the slope of the regression between GPS and MWR was essentially unity. This overall agreement of the two sensors is within the uncertainty of the MWR, supporting the use of the GPS system as an independent source for calibration. This comparison was done after reducing the MWR PWV by 3% to account for recent updates in the ARM MWR PWV processing [Turner *et al.*, 2004]. Figure 2 (right) presents the comparison of MWR and GPS PWV throughout the 24-hour period. There is approximately a 1% increase in the moist bias during the daytime period although more data would be required to determine if this is statistically significant.

[24] This analysis indicates that the GPS calibration agrees well with the ARM MWR and offers similar absolute accuracy as a source for Raman water vapor lidar calibration provided sufficient statistics are accumulated to reduce the effects of spatial inhomogeneities in the atmosphere. Therefore the disagreements observed in Figure 1 between the calculations based on the Raman lidar water vapor profile and AIRS observations are unlikely to be due to the use of GPS as the water vapor calibration reference. The corrections to Raman lidar measurements that were derived from the AWEX-G experiment and that do offer an explanation for most of this disagreement will now be discussed along with the AWEX-derived corrections to Vaisala radiosonde data.

3.4. Corrections to Raman Lidar and Radiosondes Derived From AWEX-G

[25] One of the achievements of the AWEX-G effort was to develop and validate correction techniques for Raman

Lidar and Vaisala RS80-H and RS90/92 radiosondes. The Raman lidar corrections were developed independently of the CFH and accounting for the temperature dependence of Raman scattering from water vapor and have been discussed in detail recently in the context of the International H₂O Project (IHOP) [Whiteman *et al.*, 2006a, 2006b]. Therefore the details of the correction technique will only be summarized here. The corrections developed for the Vaisala RS80-H and RS90/92 radiosonde are based on comparisons of Vaisala radiosondes and the CFH when launched on the same balloon. The details of those corrections are described in a companion paper [Miloshevich *et al.*, 2006] in this volume and will be described only briefly as well.

3.4.1. Corrections for Raman Lidar Water Vapor Measurements

[26] Two significant corrections for SRL water vapor measurements were developed and tested during the AWEX-G field campaign. Lidar measurements at short range are usually influenced by what is known as the “overlap function.” The overlap function accounts for the fact that for ranges close to the lidar system, there is a nonlinear relationship between the received power in a Raman lidar and the number density of the scatterers being probed. This nonlinearity is due to both geometrical and optical effects and is difficult to compute from first principles with high accuracy. Therefore an empirical correction for residual effects of the lidar overlap function was made on the basis of comparisons of water vapor mixing ratio measured by SRL and Vaisala RS90 radiosondes. This technique has been previously described [Whiteman *et al.*, 2006a] and is similar to what has been applied to the CARL lidar system [Ferrare *et al.*, 2004]. It resulted in a single, mean overlap correction vector that was applied to all SRL water vapor profile measurements during AWEX-G. The net result of this overlap correction for the AWEX-G experiment was to reduce the SRL calibration factor by approximately 5%.

[27] In addition, a correction for the temperature dependence of the Raman lidar measurement of water vapor mixing ratio was also performed. This temperature dependence is primarily due to the temperature dependence of Raman scattering by water vapor [Whiteman, 2003a]. The use of narrow spectral band detection for Raman water vapor measurements, as is done in the SRL, significantly improves the signal-to-noise of upper tropospheric water vapor measurements thereby improving the precision of measurements of the type required for Aqua validation. However, because of the width of the Raman water vapor spectrum and its temperature dependence, it is possible that the effective cross section of water vapor can change significantly over the range of temperatures present in the troposphere. The temperature correction technique used here is based on a recently developed form of the lidar equation [Whiteman, 2003a] and its use in the calculation of water vapor mixing ratio [Whiteman, 2003b]. The equation for mixing ratio involves the ratio of two factors, $F_N(T)$ and $F_H(T)$, that account for the temperature dependence of Raman N_2 (N) and H_2O (H) scattering. The entire temperature dependence of the water vapor mixing ratio measurement by Raman lidar is contained in the ratio $F_N(T)/F_H(T)$, examples of which have recently been published [Whiteman et al., 2006a]. All of the SRL measurements made during AWEX-G were acquired during the times of radiosonde launches permitting this ratio to be calculated for each of the SRL measurement periods during AWEX-G by use of the radiosonde measured temperature. In general, the effect of the temperature correction during AWEX-G was to reduce the magnitude of the water vapor mixing ratio measurements increasingly with altitude. This effect reaches a maximum of approximately 8% at 14 km. The standard deviation of the temperature correction at any altitude was less than 1% with the largest standard deviation existing in the boundary layer. It should be pointed out that for the configuration of the SRL, $F_N(T)/F_H(T)$ can be made nearly temperature-independent by use of a water vapor center wavelength of 407.45 nm [Whiteman et al., 2006a]. Since the time of the AWEX-G measurements, the SRL water vapor filter has been tilt-tuned to this shorter wavelength so that future SRL water vapor measurement should be essentially temperature-independent.

3.4.2. Corrections for Vaisala Radiosondes

[28] During AWEX-G, most of the balloons launched carried multiple sonde packages permitting simultaneous measurements to be acquired by Vaisala RS80-H, RS-92 and the reference CFH sensor while traveling through the same atmosphere. This data set was analyzed by Miloshevich et al. [2006] by first applying previously known time lag and, in the case of the RS80-H, temperature-dependence corrections for Vaisala radiosondes [Miloshevich et al., 2004] in order to characterize residual mean differences in the measurements of the Vaisala RS80-H and RS92 with respect to the CFH reference sensor. These differences were then characterized as a function of relative humidity and temperature, in terms of which the Vaisala calibration functions are defined, in order to derive what are referred to as the “AWEX-G empirical calibration corrections.” The AWEX-G corrections are therefore applied to Vaisala RS80-H and RS92 as a function of relative humidity and temperature and minimize the mean differences between the Vaisala radiosonde mea-

surements and those of the CFH. After applying these corrections, the mean comparison of both RS80-H and RS92 and CFH agree within $\pm 10\%$ (at worst and usually much better) up to the tropopause and in the case of the RS92, into the lower stratosphere making both radiosonde sensors suitable for Aqua validation work if the corrections are applied. Other tests performed during AWEX-G indicated that the Vaisala RS90 and RS92 are virtually identical in their water vapor measurement performance permitting the correction function derived on the basis of simultaneous measurements of RS92 and CFH to be applied to RS90 measurements as well. See Miloshevich et al. [2006] for more details.

3.5. Water Vapor Profile Comparisons

[29] Water vapor measurements acquired by SRL and six radiosonde sensors (Vaisala RS80-H, Vaisala RS90, Vaisala RS92, Sippican Inc. Mark II-a, Modem GL-98 and Snow White chilled mirror hygrometer) were compared with the reference CFH profiles. The profile comparisons of the different radiosondes with CFH are fully described in a companion paper [Miloshevich et al., 2006], so the assessment of lidar profile accuracy will be focussed upon in this section. The two results of the companion study that are pertinent to the analysis done here are the following:

[30] 1. AWEX-G testing indicated that the Vaisala RS90 and 92 sensors have essentially identical water vapor measurement performance because of the use of nearly identical capacitance RH sensors in the two instruments. Because of this, these sensors will be referred to as RS9X in this analysis.

[31] 2. Overall, the most accurate operational radiosonde tested was the Vaisala RS9X, whose mean percentage accuracy relative to CFH, after all corrections, was $<1\%$ in the lower troposphere (LT), $<2\%$ in the midtroposphere (MT) and $<3\%$ in the upper troposphere (UT).

3.5.1. Data Selection

[32] In the comparison of Scanning Raman Lidar profiles with the balloon-borne sensors during the AWEX-G field campaign, certain data rejection criteria were used. Many of the AWEX-G measurement periods had at least some periods of cloudiness. In general, in comparisons of lidar and radiosonde, if the lidar senses a cloud at a certain altitude there is no guarantee that the radiosonde, which has drifted downwind, is also sensing a cloud at the same altitude. Therefore a cloud mask product was created from the SRL aerosol scattering ratio data and used to screen out cloudy comparisons during AWEX-G. Comparisons between SRL and sonde were not performed for altitudes at which the SRL aerosol scattering ratio indicated the presence of a cloud. Furthermore, if the random error in the SRL water vapor mixing exceeded 100%, these SRL data were rejected. Certain points were rejected from the SnowWhite and CFH data records because of icing or controller instability and from the RS9X because of anomalously large disagreements with SRL that indicated different air masses were likely being sampled. SnowWhite data below 6% RH were rejected from the analysis because of the inability of the Peltier cooler used in the SnowWhite to consistently maintain a frost layer under very dry conditions [Miloshevich et al., 2006].

3.5.2. SRL Profile Comparisons With Fully Corrected Vaisala RS9X

[33] The profile comparison of SRL with the reference CFH showed agreement generally within 10% up to an

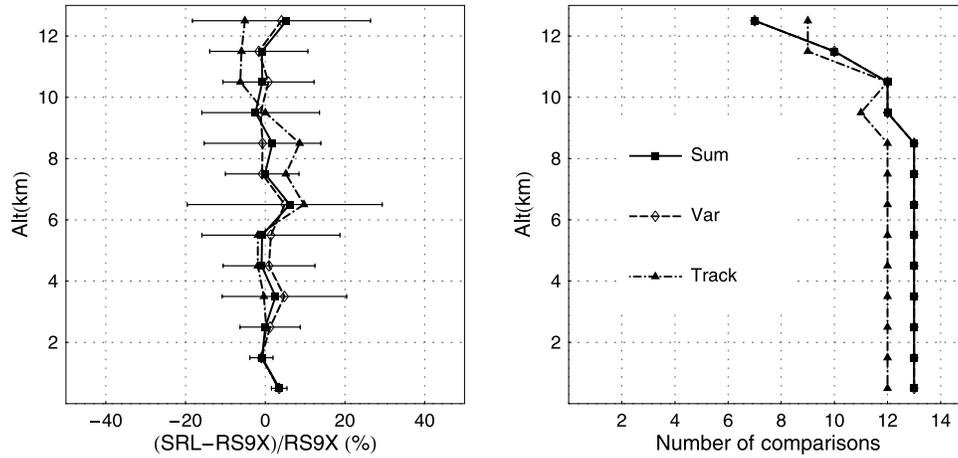


Figure 3. (left) Mean profile comparisons of SRL and the RS9X using the three methods of analyzing the SRL profiles described in the text. Layer averages of 1 km have been used and the error bars indicate standard deviation within each 1-km interval. (right) Number of comparisons at each altitude.

altitude of 9 km, but poor statistics above this altitude due to the presence of clouds prevented a robust comparison of SRL and CFH above this altitude. However, one of the conclusions of the AWEX-G radiosonde analysis [Miloshevich *et al.*, 2006] was that the mean, fully corrected RS9X agreed with the CFH to better than 3% at all altitudes. The fully corrected RS9X sensor, therefore, was used as a transfer standard of the CFH calibration to assess the accuracy of the SRL measurements since there were significantly more comparisons of SRL and RS9X than with CFH.

[34] The mean percentage differences between the SRL and RS9X using three methods of analyzing the SRL data are therefore shown in Figure 3 using the data rejection criteria described earlier. For these comparisons, the beginning of the SRL averaging window coincided with the radiosonde launch time. The three methods of analyzing the SRL data, referred to as “Sum,” “Var,” and “Track,” are described more fully in Appendix A. Briefly, the “Sum” method uses the same temporal averaging period independent of altitude, “Var” uses a variable temporal average as a function of altitude and “Track” implements a technique that attempts to compensate for the fact that the radiosonde drifts with the wind. For those comparisons, the definition of the layer mean percentage difference is $Mean[(S_i - V_i)/V_i]$ where S indicates the SRL measurement and V indicates the Vaisala RS9X measurement and where each i^{th} layer is 1 km thick. Table 1 provides the mean comparison of the sensors based on a regression analysis where the percentage difference between SRL and RS9X is calculated from

ordered pairs formed at 1-km resolution, the same as used in the profile analysis shown in Figure 3.

[35] Referring to Figure 3, the three methods of analyzing the SRL data with respect to RS9X show agreement to better than 10% at all altitudes below 13 km. Below 6 km, all methods agree to better than 5–7% with the RS9X measurements. In this altitude range, the track sonde technique yields a near 0% bias and the variable and straight sum techniques show the largest differences of 5–7% between 3 and 4 km. Above 6 km, the variability in the track sonde comparisons increases so that the track sonde analysis in general yields the largest differences with RS9X. The straight sum and variable sum comparisons perform similarly throughout the profiles. The regression analysis shows nearly unity slope for all three methods of reducing the lidar data and mean percentage difference of less than 3% for all techniques.

3.6. Mean AWEX-G Upper Tropospheric Accuracy Assessment

[36] Because of the importance of upper tropospheric (UT) water vapor for Aqua validation and the difficulty in accurately measuring it, the UT measurement accuracy of the Vaisala RS80-H, RS9X, SnowWhite and SRL were studied. Only those radiosonde measurements made on the same balloon were used in the accuracy assessment with respect to CFH. The SRL accuracy was assessed with respect to fully corrected RS9X because of the larger number of comparisons available.

Table 1. Full Profile Regression Comparison of SRL Water Vapor Mixing Ratio Profiles and RS9X During AWEX-G^a

SRL Versus RS9X, Technique	Regression Statistics						$\bar{\Delta}(\%)_{AWEX} \cdot \left(\frac{MR_{S_i} - MR_{V_i}}{MR_{V_i}} \right)$
	N_{profs}	N_{pts}	m	b, g/kg	R ²	$\bar{\Delta}$, g/kg	
SRL (Sum)	13	160	1.01	-0.013	0.998	0.004	$1.3 \pm 23.9(6.6)$
SRL (Var)	13	160	1.01	-0.007	0.998	0.009	$1.7 \pm 14.8(4.1)$
SRL (Track)	12	155	1.01	0.014	0.998	0.004	$2.4 \pm 15.2(4.4)$

^aThe columns, in order, are the number of profiles compared, the number of ordered pairs of points used in the regressions, the slope and intercept of the best fit line, the correlation coefficient of the regression and the mean difference between the best fit line and the SRL data. The last column gives mean percentage difference, defined as $100 \cdot Mean[(SRL-RS9X)/RS9X]$, for the entire profile. In the formula given in the table, S refers to the SRL, and V refers to the Vaisala RS9X sonde. The standard deviation and, in parentheses, the standard error of the mean percentage differences are also given.

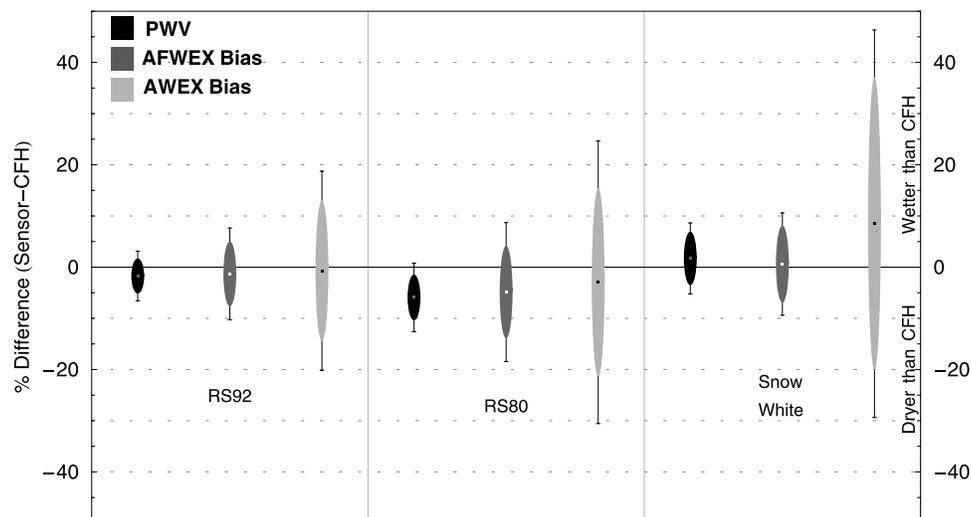


Figure 4. A comparison of the upper tropospheric water vapor measurements of various sensors versus the reference CFH sensor. Three methods of quantifying percentage difference are displayed and explained in the text.

[37] The accuracy assessment was performed in three different ways: The first was based on the total precipitable water measured by the various sensors from 7 km to the tropopause (typically between 14 and 15 km during AWEX-G). This technique gives greater weight to the lower altitudes where most of the precipitable water is located. The second was based on a linear regression of ordered pairs of values from the two sensors being tested and uses a similar approach as in the AFWEX analysis [Ferrare *et al.*, 2004], where the mean percentage differences were determined using the formula $\overline{X_i - REF_i} / \overline{REF_i}$, where $\overline{REF_i}$ refers to the mean of the reference sensor measurements in the entire 7 km to tropopause layer. This technique for calculating mean percentage difference tends to suppress the influence of dry layers. The third technique was based on the same linear regression results but used a different formula for calculating mean percentage difference: $(X_i - REF_i) / REF_i$, so that in this method the percentage difference is determined strictly layer by layer. This technique equally weights dry and wet layers in the computation of the mean. For these regressions of same-balloon radiosonde data, ordered pairs of points were formed using approximately 20 m vertical resolution where all qualifying points within the altitude range of 7 km to the tropopause were used. For the regressions of SRL and RS9X data, both 200-m and 1-km resolution data were studied. The regressions presented here use 1-km resolution; the 200-m resolution regressions gave similar results but with higher variability.

3.6.1. Comparisons With Respect to CFH

[38] The mean percentage differences of the Vaisala RS80-H, Vaisala RS92 and Meteolabor SnowWhite water vapor mixing ratio measurements with respect to CFH are shown in Figure 4. Water vapor mixing ratio was calculated using the respective temperature and pressure information from each individual sensor. For the CFH, which was always launched with an RS80-H attached and usually also with an RS92, the RS92 temperature and pressure were used when available because of their higher accuracy. For the SnowWhite, the RS80-H temperatures were used since these are the temperatures that are available in its nominal configuration when launched with the CFH. Ordered points were formed at a vertical resolution of approximately every 20 m. The Vaisala radiosondes have received the full corrections described above including MWR scaling. The SnowWhite profiles have only been microwave scaled. The results are summarized in Table 2.

[39] Figure 4 and Table 2 indicate that the various methods of analysis show approximately a 1–2% dry bias of RS92, a 3–6% dry bias of RS80-H and a 2–8% wet bias of SnowWhite with respect to CFH. Nearly all of these comparisons, however, are within one standard error of the reference CFH value. Only the comparison of RS80-H and CFH based on precipitable water shows a statistically significant, small dry bias. As would be expected, the variability in the three mean percentage differences ($\overline{\Delta}(\%)$) presented (PWV, AFWEX, AWEX-G) increases consistently from left to right. The PWV method of calcu-

Table 2. Upper Troposphere Bias Comparison of Various Sensors and CFH During AWEX-G^a

UT (7 km-trop) Versus CFH, Sensor	Regression Statistics						$\overline{\Delta}(\%)_{PWV}$, $\frac{PWV_s - PWV_c}{PWV_c}$	$\overline{\Delta}(\%)_{AFWEX}$, $\frac{MR_{S_i} - MR_{C_i}}{MR_{C_i}}$	$\overline{\Delta}(\%)_{AWEX}$, $\left(\frac{MR_{S_i} - MR_{C_i}}{MR_{C_i}}\right)$
	N_{profs}	N_{pts}	m	b , g/kg	R^2	$\overline{\Delta}$, g/kg			
RS-92	8	2370	0.988	-0.0002	0.997	-0.0024	$-1.73 \pm 4.86(1.7)$	$-1.31 \pm 8.96(3.2)$	$-0.71 \pm 19.5(6.9)$
RS80-H	9	2940	0.976	-0.0040	0.994	0.0080	$-5.92 \pm 6.71(2.2)$	$-4.86 \pm 13.6(4.5)$	$-2.95 \pm 27.6(9.2)$
SnowWhite	7	2346	0.978	0.0055	0.994	0.0012	$1.71 \pm 6.93(2.6)$	$0.60 \pm 9.99(3.8)$	$8.48 \pm 37.9(14.3)$

^aThe columns have similar meaning as in Table 1. The percentage difference is calculated three ways as explained in the text. In the formulas provided, S refers to the sensor under study, and C refers to the reference CFH instrument.

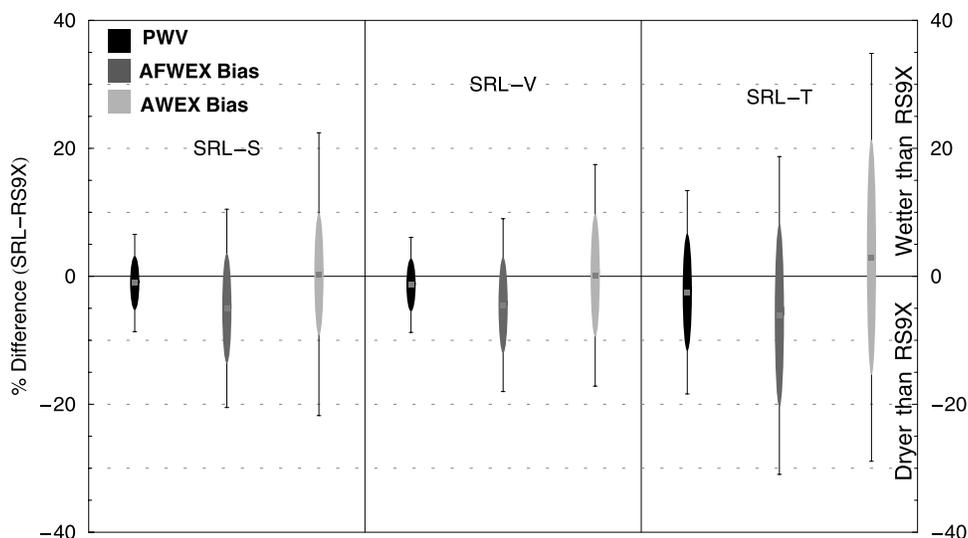


Figure 5. A comparison of mean upper tropospheric measurements of SRL (using the three different methods of averaging the SRL data described in the text) and RS-9X after fully correcting both sensors. In general, agreement between SRL and RS9X is better than 5% using any of the techniques. The error bars represent the standard deviation, while the shaded region represents twice the standard error.

lating percentage difference generates one value for each profile, thus increasing the statistical robustness. The AFWEX method calculates a mean percentage difference but uses the same mean value from 7 km to the tropopause to normalize each of the differences used in the mean while the AWEX-G method normalizes each of these differences by the CFH value used in each difference. Thus, from left to right the results shown in Table 2 of the different methods of calculating percentage difference are increasingly more subject to the influence of small absolute differences (and corresponding large percentage differences) in dry layers.

3.6.2. Comparisons of SRL With Respect to Fully Corrected RS90/92

[40] Miloshevich et al. [2006] showed that in the mean, the fully corrected RS9X could be used as a transfer of the CFH absolute calibration with mean uncertainties less than 3% in the upper troposphere. Because of the much larger number of comparisons of SRL and RS9X than with CFH, therefore, the analysis of SRL upper tropospheric measurements was done with respect to RS9X. For the regression analyses shown here, the ordered pairs were formed using 1-km averages to decrease the influence of small differences in dry layers that can result in large percentage differences. The results of the PWV and bias comparisons are presented in Figure 5. The corresponding tabular information is provided in Table 3.

[41] Figure 5 presents the upper tropospheric precipitable water comparison of the fully corrected SRL and RS9X data using the three methods of analyzing the SRL measure-

ments. All methods of analysis show agreement to better than $\pm 5.4\%$ with the standard error indicating no significant difference between the mean UT measurements of the SRL and RS9X. There is, however, a distinct tendency for the track-sonde method to produce higher variability than the other methods of reducing the SRL data. The slope of the best fit line using all three methods of processing the SRL data is less than unity, whereas the full profile regression results shown in Table 1 indicated a nearly unity slope. A study of the mean water vapor differences versus water vapor amount and altitude (not shown) revealed essentially no statistically significant differences from perfect agreement between the two sensors. This implies that a larger body of comparisons would be required to determine if the less than unity slopes in Table 1 are statistically significant.

[42] Of the three techniques, the variable sum technique yields a slope that is closest to unity, an intercept that is closest to zero, shows the lowest mean variability, and also provides the best overall agreement with the fully corrected RS9X. It is for these reasons that the variable sum technique of analyzing the SRL data will be used to demonstrate the effect of the corrections to the lidar data later in the paper. It also was the technique used to analyze the SRL data submitted to the Aqua validation archive.

4. Relationship of AWEX-G and AFWEX Results

[43] Both AWEX-G and AFWEX resulted in a core group of sensors that agree within $\pm 5\%$ on the basis of several

Table 3. Upper Troposphere Bias Comparison of the Three Methods of Analyzing SRL Data and RS-9X During AWEX-G^a

SRL(UT) Versus RS9X	Regression Statistics						$\bar{\Delta}(\%)_{PWV}, \frac{PWV_S - PWV_V}{PWV_V}$	$\bar{\Delta}(\%)_{AFWEX}, \frac{MR_{S_i} - MR_{V_i}}{MR_{V_i}}$	$\bar{\Delta}(\%)_{AWEX}, \left(\frac{MR_{S_i} - MR_{V_i}}{MR_{V_i}} \right)$
	N_{profs}	N_{pts}	m	$b, g/kg$	R^2	$\bar{\Delta}, g/kg$			
SRL (Sum)	13	69	0.89	0.008	0.97	-0.004	$-1.0 \pm 7.6(2.1)$	$-4.3 \pm 15.1(4.2)$	$1.1 \pm 17.2(4.8)$
SRL (Var)	13	69	0.92	0.005	0.97	-0.003	$-1.3 \pm 7.4(2.1)$	$-3.8 \pm 13.1(3.6)$	$0.9 \pm 17.2(4.8)$
SRL (Track)	12	71	0.89	0.009	0.93	-0.005	$-2.5 \pm 15.9(4.6)$	$-5.4 \pm 24.5(7.1)$	$3.8 \pm 32.1(9.3)$

^aSee Tables 1 and 2 for the definitions of the quantities.

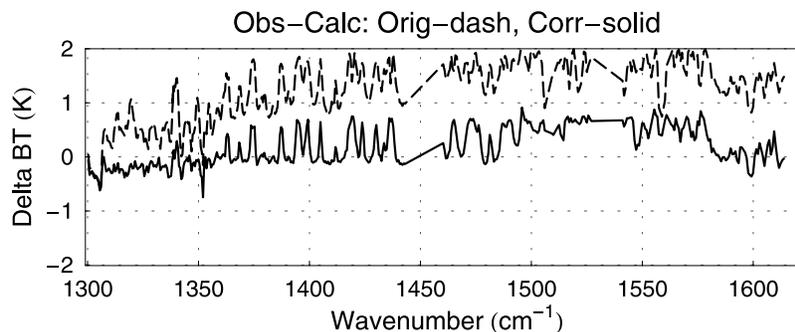


Figure 6. An Obs-Calc comparison of the original SRL water vapor profiles from September to November 2002 and those released in late 2004, which included the corrections described here. Both comparisons use the January 2004 version of the AIRS fast model. The calculations based on the updated SRL profiles agree in the mean with the AIRS observations within $\pm 0.5K$.

analysis methods. The AFWEX field campaign used the LASE airborne DIAL instrument as its reference and AWEX-G used the CFH. These two instruments have not been directly compared but an analysis of the results of the two field campaigns can give some indication of the relative performance of the LASE and CFH and thus of the conclusions of AWEX-G and AFWEX. There were two water vapor profilers common to the two field campaigns that operated nominally and for which there are sufficient statistics to make useful conclusions about their operations. Those sensors were the Vaisala RS80-H and the SRL.

[44] As detailed by *Miloshevich et al.* [2006], apparent differences in the performance of the RS80-H sensor prior to applying corrections in the two experiments are attributed to the generally moister profiles that occurred in AFWEX versus AWEX-G. Therefore, after considering the differences in the measurement conditions found in AFWEX and AWEX-G, the RS80-H performance, relative to the respective reference systems, was similar in the two experiments.

[45] Considering the SRL results, the mean UT bias of the SRL with respect to LASE in AFWEX (based on a regression of points between 7 km and the tropopause using what we have referred to as the AFWEX technique here) was approximately +4% [*Ferrare et al.*, 2004]. We have reanalyzed the SRL measurements from AFWEX using the correction techniques developed here but otherwise using the same methods employed in AFWEX and found that the mean upper tropospheric bias of the SRL decreased by 3%. This brings the mean calibrations of LASE and SRL to within 1% of each other during AFWEX. Table 3 shows that using the variable sum technique, the SRL agreed within 1–4% of the fully corrected RS9X, which agreed with CFH to better than 3%. Therefore the implication of these relative comparisons is that the AFWEX water vapor reference sensor (LASE) and the AWEX-G water vapor reference sensor (CFH, which in the work by *Miloshevich et al.* [2006] was calculated to have absolute accuracy of 4% depending on RH) are equivalent to within approximately 5%.

5. Effect of the Raman Lidar and Radiosonde Corrections on AIRS Radiance Comparisons

[46] Both Raman lidar and radiosonde correction techniques have been described above. In this section, the

influence of those corrections on AIRS radiance comparisons is illustrated.

5.1. Raman Lidar Overlap and Temperature Dependence Corrections

[47] The radiance comparison shown in Figure 1 has been reanalyzed using the Raman lidar correction techniques described here and also using version 4 of SARTA. Figure 6 shows the Obs-Calc comparison of AIRS observations and SARTA calculations based on (1) the SRL data as released in early 2003 without the corrections described here (dashed) and (2) the SRL water vapor measurements after applying the corrections developed and verified in the AWEX-G field experiment (solid). Using version 4 of SARTA, the calculations based on the original SRL data show a significant positive bias with respect to the AIRS observations. The calculation based on the revised SRL data, however, agrees in the mean to within $\pm 0.5K$ with the AIRS observations throughout the water band although there is an indication of dry bias to the calculation of $\sim 0.5K$, or approximately 5% in RH, in the high-wavenumber region of the water band. It is worth pointing out that the total precipitable water standard used for the calibration of the SRL measurements during this fall 2002 field campaign did not change from the original and corrected data shown in Figure 6. Therefore the total precipitable water is the same for both versions of the SRL profiles. The large differences shown in Figure 6 can all be attributed to differences in how the precipitable water is distributed in the SRL profiles used in the SARTA calculations.

5.2. Vaisala RS-90 Corrections

[48] At the current time, the channel transmittances in the AIRS fast forward model, SARTA, have been tuned to minimize the differences between fast model calculations and the BE product based on the first phase of RS90 launches that occurred at the Tropical Western Pacific (TWP) ARM site [*Tobin et al.*, 2006]. This tuning was done prior to the availability of the AWEX-G corrections. Figure 7 presents a comparison of 68 AIRS FOVs with calculations using the version 4 of SARTA for an ensemble of 13 semicircular overpasses of TWP from the second phase of the special Aqua launches that occurred between September 2003 and March 2004. The dashed curve

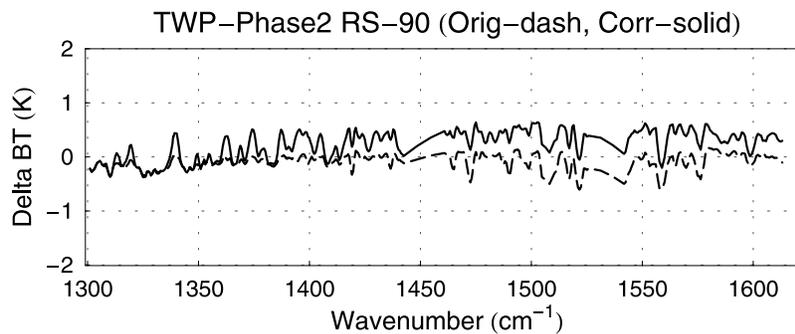


Figure 7. A comparison of AIRS observations and calculations using the AIRS fast model for 13 overpasses of the ARM TWP site. Two sets of water vapor input information have been used: the original RS-90 radiosondes prior to application of the time lag and AWEX corrections (dashed) and after applying those corrections (solid). The mean effect of the corrections is to increase upper tropospheric RH by approximately 5% in this example. However, for individual sondes the magnitude of these corrections can exceed 40% in parts of the upper troposphere [Miloshevich *et al.*, 2006].

presents Obs-Calc using the microwave-scaled Vaisala RS90 radiosondes without any other corrections as input to SARTA. The solid curve is the Obs-Calc for the same RS90 radiosondes but after applying the time lag correction [Miloshevich *et al.*, 2004], the AWEX-G correction and microwave scaling. In the range of 1400–1500 cm^{-1} , the difference between the two curves is generally 0.4–0.5K, with the AWEX-G correction being responsible for most of this effect. Using the rule of thumb that 1 K corresponds to a relative change of 12% in the amount of water vapor in the upper troposphere [Soden *et al.*, 2000], the net effect of the corrections on this ensemble of radiosondes equates to approximately a 5–6% increase in UT water vapor mixing ratio. Although the absolute uncertainty of the corrected RS90 data is not better than 5%, the implication of the comparison shown in Figure 7 is that there could be a dry bias in the spectroscopy of SARTA in the upper troposphere of approximately 5%. This could reflect the fact that SARTA was tuned with respect to radiosonde data prior to applying the corrections developed here. Although the comparison with respect to the lidar profiles shown in Figure 6 shows slightly better overall agreement with the AIRS observations than Figure 7 on the basis of RS90 radiosondes from TWP, both comparisons tend to indicate a dry bias to SARTA spectroscopy of approximately 5% in absolute water amount in the high-wavenumber region of the water band. Although the absolute accuracy of the CFH sensor used to generate the AWEX-G calibration correction for the RS90 is also approximately 5%, the implication of both of these comparisons based on independently calibrated validation data is that there is a small dry bias in the current version of SARTA.

[49] It is important to point out that Figure 7 represents the mean comparison of an ensemble of measurements. This is quite useful for radiance validation experiments where large sets of statistics are desired. However, the magnitude of the corrections for individual radiosondes can be considerably larger than the ~ 5 –6% noted above. For example, when considering the individual radiosondes that were launched at TWP during phase 2 of the special Aqua radiosonde launches, the changes in upper tropospheric water vapor values due to the corrections developed from AWEX-G exceed 40% in some cases [Miloshevich *et al.*,

2006]. Therefore, if studies are to be done on single overpass cases or with small data sets it is particularly important to work with fully corrected radiosonde data.

6. Comparison of GPS and Corrected SRL Measurements From Fall 2002 to Aqua Version 4 Retrievals

[50] Using corrected SRL data, the 26 overpass measurements acquired at GSFC in the fall of 2002 were divided into clear and high cloud cases and compared with the latest version 4 Aqua retrievals after degrading the validation measurements to the vertical resolution of the retrievals. Only those Aqua retrievals, which combine the data from AIRS, AMSU and HSB, for which all quality assurance (QA) flags were zero were used. The comparison is shown on the left in Figure 8. Although the number of cases is not large, the profile comparisons show a distinct moist bias to the retrievals in the lower and mid troposphere reaching mean values of 20–30%. There is also some indication that the presence of cirrus clouds may bias the retrievals drier both in the lower and upper troposphere. A more extended comparison (August 2002 to April 2004) of precipitable water vapor measurements from SuomiNet GPS at GSFC (which was the calibration source for the Raman lidar profiles used in the comparison on the left) and Aqua retrieval is shown on the right. The best fit regression line has a slope of ~ 0.82 and the mean ratio of Aqua and GPS PWV is 1.21. Both measures indicates a similar moist bias to the retrievals of approximately 20% compared with the SuomiNet GPS, which agreed with the DOE MWR PWV within 3% during AWEX-G.

7. Summary and Conclusions

[51] The early AIRS comparisons using Aqua validation data demonstrated a large uncertainty in the water vapor measurements that were being acquired under the Aqua validation activity. Discrepancies between AIRS observations and SARTA calculations using Scanning Raman lidar and Vaisala RS90 radiosonde data indicated an apparent uncertainty in upper tropospheric water vapor calibrations of at least 25%. Other validation data showed even larger

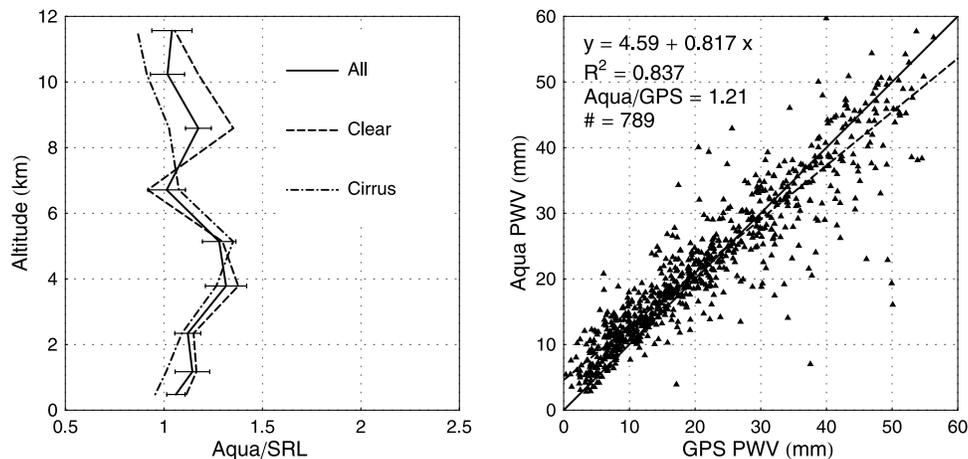


Figure 8. (left) Comparison of water vapor profiles from SRL and Aqua retrievals at NASA/GSFC in fall 2002. There is a persistent moist bias to the retrievals in the midtroposphere. (right) Comparison of precipitable water vapor from SuomiNet GPS and Aqua retrievals over the period of August 2002 to March 2004. A similar wet bias in the retrievals is observed.

discrepancies. This was a motivator for the AIRS Water Vapor Experiment–Ground (AWEX-G) that took place between 27 October and 19 November 2003. The objective of AWEX-G was to bring together in one place various water vapor technologies in use for Aqua validation and operate them over an extended period in order to resolve the discrepancies observed.

[52] During AWEX-G various radiosondes (Vaisala RS80-H, RS90, RS-92, Modem, Sippican, SnowWhite) and Scanning Raman Lidar were operated along with the reference Cryogenic Frostpoint Hygrometer (CFH). The total precipitable water measurements of GPS were also studied with respect to MWR. The month-long comparison of GPS and MWR precipitable water measurements during AWEX-G indicated mean agreement of better than 3%. The conclusion was that GPS is an accurate precipitable water calibration source that can be used to constrain profile measurements of water vapor mixing ratio by Raman lidar, provided that ensembles of cases are used to reduce variability that can be introduced by the large-volume average used by GPS.

[53] Corrections for the effects of the lidar overlap function and the temperature dependence of Raman scattering on Raman water vapor measurements were also validated using the AWEX-G measurements. The combined effect of the overlap and temperature-dependence corrections on the SRL water vapor measurements was to reduce the water vapor mixing ratio in the upper troposphere by 10–15% at the highest altitudes. The temperature dependence correction is a sensitive function of the exact transmission characteristics of the lidar system. In the case of the SRL, the temperature dependence of the water vapor mixing ratio measurements can be essentially eliminated with careful selection of the bandpass characteristics of the water vapor interference filter.

[54] Atmospheric variability was found to be potentially a significant source of error in this study. For this reason radiosonde comparisons with CFH were limited to sensors flown on the same balloon as the CFH. Comparisons of sensors with CFH launched simultaneously but on different

balloons consistently showed significantly higher variability. The contribution of atmospheric variability to the comparison of lidar and radiosonde was noticeable and required some manual rejection of data. On the basis of our experience in AWEX-G, it is suggested that, if possible, accuracy assessments of radiosondes be done with respect to sensors on the same balloon and that accuracy assessments of lidar be done with respect to other lidar systems so as to minimize the effects of sampling different atmospheres. In order to avoid the need for manual rejection of data, larger ensembles of lidar/radiosonde comparisons than acquired in AWEX-G are encouraged in future experiments.

[55] The AWEX-G radiosonde intercomparison activity is detailed in a companion paper [Miloshevich *et al.*, 2006] in this same issue. Among the conclusions of that effort were that the Vaisala RS90 and RS92 have essentially identical measurement performance and thus could be treated as identical sensors. A significant result of that detailed accuracy assessment was the development of new empirical corrections for Vaisala RS80-H and RS9X that address errors in the Vaisala calibration model and bring these Vaisala sensors into excellent mean agreement with the CFH. For RH >10%, the RS80-H agrees in the mean with CFH to within $\pm 5\%$ and the RS-9X to within $\pm 2\%$. For RH <10%, the mean agreement between RS80-H and CFH is within $\pm 10\%$, while that for RS-9X is within $\pm 5\%$. It was judged that only the Vaisala RS9X is sufficiently accurate for Aqua water vapor validation throughout the troposphere (given the 10% absolute accuracy goal for Aqua retrievals), especially if the corrections are applied.

[56] Using the correction techniques for both radiosonde and Raman lidar, the mean upper tropospheric water vapor accuracies of the Vaisala RS-80H, Vaisala RS92 and Meteorolabor SnowWhite were assessed using three techniques versus the reference CFH. Full profile and upper tropospheric accuracy comparisons were also made between the SRL and fully corrected RS9X. For the purposes of this study, the upper troposphere was defined as extending from 7 km to the tropopause; the same as in the AFWEX study [Ferrare *et al.*, 2004]. Three techniques were used to

compare upper tropospheric water vapor measurement performance. The first of these was in terms of integrated precipitable water between 7 km and the tropopause, the second and third were both in terms of biases determined from the linear regression of ordered pairs of points within the same altitude range, where two different formulas were used to calculate the biases. Using data selection criteria, mean agreement within 5% was generally achieved between the operational radiosondes and the CFH using any of the methods of comparison.

[57] Of the three methods of characterizing sensor UT accuracy, the one based on precipitable water is likely to be more important for assessing the influence of measurement error on outgoing longwave radiation (OLR) calculations since it has been shown [Ferrare *et al.*, 2004] that errors of 5% in the integrated PWV above 7 km introduce errors in OLR calculations of $\sim 0.5 \text{ W/m}^2$. It should be pointed out, however, that the PWV above 7 km is typically dominated by the water vapor found within a region 1–2 km above this altitude. So this technique of comparing sensors will be less sensitive to measurement differences at high altitudes where water vapor concentrations generally are lower. The regression analyses can therefore provide more information about the relative performance of sensors in the upper portions of the upper troposphere than can the PWV approach and could therefore be more appropriate indications of a sensor's ability to provide useful measurements in cirrus cloud formation studies, for example. Only if regression results reflect relative performance of sensors that is independent of altitude may they be taken as a substitute for an accuracy assessment done directly in terms of precipitable water. Accuracy assessments based both on total precipitable and regression biases provide a more complete characterization of a sensor's relative performance than either approach separately.

[58] *Miloshevich et al.* [2006] showed that the fully corrected RS9X agrees in the mean with CFH to generally better than 3% throughout the troposphere (better in the lower and middle troposphere). Therefore, because of a small number of direct SRL/CFH comparisons in the UT, the RS9X was used as an accurate transfer standard of the water vapor calibration of the CFH reference sensor in order to characterize the water vapor measurement accuracy of the SRL. There were three techniques of analyzing the Raman lidar data used in order to address the inherent spatiotemporal measurement differences between lidar and radiosonde. These techniques are referred to as "Var," "Sum," and "Track." The 1-km average profile comparison of fully corrected SRL and fully corrected RS9X showed mean agreement between the two sensors of generally better than 5% up to an altitude of 12 km using any of the three techniques. The "Track" method showed larger variability than the "Sum" and "Var" methods resulting in agreement of better than 10% to 12 km. The conclusion regarding the three methods of analyzing the Raman lidar data is that on the basis of this ensemble of measurements, there is a small advantage to the use of the variable sum technique over the straight sum technique, but that the track-sonde approach of analyzing the data introduced larger variability in the results in the mid-to-upper troposphere than either of the other techniques. It is apparent, at least for this set of measurements, that the assumption of the track-sonde approach, that

the atmosphere translates uniformly and unchanging at each altitude, was violated such that this technique of analyzing the data actually degraded the comparisons, at least in the middle to upper troposphere.

[59] Mean percentage differences of both standard deviation and standard error are presented in the regression results provided here. The use of standard error assumes that the relative difference between two sensors being compared is independent of the ordered pairs used in the regression even though these pairs may represent very different measurement conditions (RH, T, mixing ratio, altitude) that can influence measurement performance. On the basis of the results here, this assumption is more true for the same-balloon radiosonde comparisons than the lidar/radiosonde comparisons. It is not clear if the tendency for the upper tropospheric lidar/radiosonde regressions to yield slopes less than unity reflects a measurement problem in the lidar or the radiosonde or if some element of atmospheric variability not removed through data selection might be the cause. Despite these small disagreements between SRL and RS9X, mean agreement in the UT comparisons was achieved generally at the 5% level using all methods of comparison of the two instruments.

[60] On the basis of this analysis, a core group of sensors was found to agree at approximately the $\pm 5\%$ level after applying all corrections that were developed from AWEX-G. These sensors were the Vaisala RS80-H, RS90, RS92, SnowWhite, SRL and the CFH. In the case of the RS80-H, cloud contamination must be eliminated to achieve this accuracy. In the case of the SnowWhite, cloud contamination must be eliminated and the RH comparisons must further be restricted to values above 6% RH. The comparison of fully corrected SRL and RS9X indicated agreement generally to better than 5% both in the full profile comparisons and in the upper tropospheric water vapor comparisons. A study of the common results of AFWEX and AWEX-G experiments implies that the reference sensors from AFWEX (LASE) and the CFH reference sensor from AWEX-G should agree within approximately $\pm 5\%$. The effect of the corrections was demonstrated both in radiance and retrieval comparisons using SRL data from NASA/GSFC and Vaisala RS90 data from the ARM TWP site. The radiance comparison indicates generally good agreement with the AIRS fast forward model, SARTA, although there is an indication of a small dry bias in the SARTA calculations of approximately 5% in the high-wavenumber region of the water band. Comparison of Aqua retrievals and corrected Raman lidar data indicate a moist bias to the retrieval of approximately 20% and that the presence of cirrus clouds may influence these results.

[61] The main research goal of the AWEX-G field campaign was to study and resolve apparent measurement differences among Vaisala radiosondes and Raman lidars in use in the Aqua validation effort. The results presented here indicate that measurement errors of significant magnitude were found in both the Vaisala radiosondes and Scanning Raman Lidar and that the correction techniques developed for these instruments bring them into agreement within approximately 5% both with each other and with the reference sensor, the Cryogenic Frostpoint Hygrometer, which itself was calculated to possess an absolute accuracy of approximately 4%. Given that the goal for Aqua retrieval

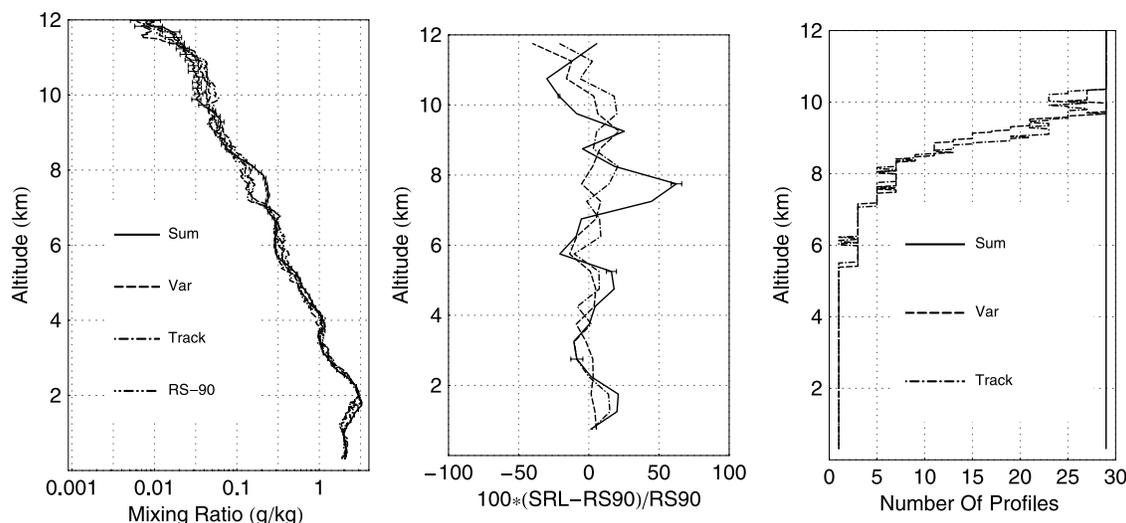


Figure A1. An illustration of the three profile averaging schemes that were investigated for the processing of the Raman lidar water vapor mixing ratio data. “Sum” refers to using the same temporal averaging as a function of altitude, in this case 29 min. “Var” refers to using a variable temporal average as a function of altitude in order to maintain the random error below a fixed value here chosen to be 10%. An upper limit of 29 profiles is also selected in this case. “Track” refers to a technique that attempts to address the fact that the radiosonde moves downwind with altitude whereas the lidar can measure purely vertically. A variable temporal smoothing is also employed in this technique.

absolute accuracy is 10%, the corrected profiles provided by Vaisala RS80-H, RS9X and Scanning Raman Lidar are found to be suitably accurate for Aqua validation efforts.

Appendix A: Efforts to Address the Spatiotemporal Mismatch Between Lidar and Radiosonde Data

[62] A fundamental difference in the water vapor measurement techniques of Raman lidar and radiosonde is that the lidar measures along a straight line and radiosondes drift with the wind. Furthermore, mixing ratio measurements by Raman lidar require varying amounts of averaging time to provide similar signal-to-noise characteristics as a function of altitude. For example, under nighttime conditions such as those studied during the AWEX-G field campaign, Figure A1 shows a typical example of the number of 1-min profiles that must be summed together to achieve no more than 10% random error as a function of altitude. It varies from 1 profile below 5 km to 29 at 10 km. Above 10 km, the random error in the summed profile exceeds 10%. By contrast, the radiosonde makes all of its measurements nearly instantaneously but requires approximately 30 min to ascend to the upper troposphere. Because of the fundamental differences in water vapor measurement resolution of Raman lidar and radiosonde, three techniques of averaging the Raman lidar data were explored in the comparisons with radiosonde data.

[63] The first technique of averaging the Raman lidar data consisted of using a straight sum of a fixed number of profiles independent of altitude. This is the technique that has been used in similar intercomparison studies such as AFWEX [Ferrare *et al.*, 2004] and is referred to here as “Sum.” The second technique of averaging used a variable number of profiles in the summation as a function of altitude so as to maintain the random error at each altitude

below some threshold value, here set to be 10%. This technique has been described recently [Whiteman *et al.*, 2006b] and is referred to as “Var.” The third technique attempted to compensate for the fact that the radiosonde drifts with the wind. The height of the radiosonde as a function of time was used to shift, the time/height series of lidar mixing ratio. This shifting was done so that the data used at a certain altitude were those which were measured by the lidar at the time corresponding to when the radiosonde passed through that same altitude. This technique of attempting to track the sonde is referred to as “Track.” It essentially assumes a homogeneous atmosphere uniformly translating horizontally as a function of altitude over the averaging period. Any change in wind speed or direction at a given altitude is not accounted for in this scheme. After performing this translation of the lidar data, the same variable averaging technique used in the “Var” method was then used. Figure A1 illustrates the use of these different techniques in a comparison with a Vaisala RS90 radiosonde on the night of 19 November at the end of the mission.

[64] Figure A1 (left) presents the comparison of the individual profiles, while Figure A1 (middle) shows the percentage difference of the three methods of processing the SRL data with respect to RS90. On the right is shown the number of 1-min profiles used as a function of altitude for the three lidar averaging techniques. There is good general agreement among all 4 profiles plotted. However, in Figure A1 (middle) one can see that between the altitudes of 7–8 km the straight summation technique, which used a 29-min average at all altitudes, diverges from the other measurements by more than 50%. Both the variable sum and track-sonde techniques used between 3 and 7 profiles in this altitude range implying that there was significant atmospheric variation during the period sampled by the

“Sum” technique that was not sampled using the other techniques. Between 1 and 2 km, both the “Sum” and “Track” techniques diverge from the radiosonde and the “Var” results due presumably to atmospheric variation captured by those techniques that was not sampled by the “Var” technique. The comparisons detailed in Table 1 indicate that, of the three techniques, the “Var” technique yielded a slope that is closest to unity, an intercept that is closest to zero, showed the lowest mean variability, and also provided the best overall agreement with the fully corrected RS9X. For these reasons the variable sum technique was used to analyze the SRL data submitted to the Aqua validation data archive.

[65] **Acknowledgments.** The support of the Research Division of NASA’s Office of Earth Science, managed by Jack Kaye, and the EOS Validation Activity managed by David Starr are gratefully acknowledged.

References

- Braun, J., C. Rocken, and J. Liljegren (2003), Comparisons of line-of-sight water vapor observations using the global positioning system and a pointing microwave radiometer, *J. Atmos. Oceanic Technol.*, *20*(5), 606–612.
- Ellingson, R. G., and W. J. Wiscombe (1996), The spectral radiance experiment (SPECTRE): Project description and sample results, *Bull. Am. Meteorol. Soc.*, *77*(9), 1967–1985.
- Ferrare, R. A., et al. (2004), Characterization of upper-troposphere water vapor measurements during AFWEX using LASE, *J. Atmos. Oceanic Technol.*, *21*, 1790–1808.
- Liljegren, J. C., S. A. Boukabara, K. Cady-Pereira, and S. A. Clough (2005), The effect of the halfwidth of the 22-GHz water vapor line on retrievals of temperature and water vapor profiles with a twelve-channel microwave radiometer, *IEEE Trans. Geosci. Remote Sens.*, *43*, 1102–1108.
- Miloshevich, L. M., A. Paukkunen, H. Vömel, and S. J. Oltmans (2004), Development and validation of a time-lag correction for Vaisala radiosonde humidity measurements, *J. Atmos. Oceanic Technol.*, *21*, 1305–1327.
- Miloshevich, L. M., H. Vömel, D. N. Whiteman, B. M. Lesht, F. J. Schmidlin, and F. Russo (2006), Absolute accuracy of water vapor measurements from six operational radiosonde types launched during AWEX-G and implications for AIRS validation, *J. Geophys. Res.*, *111*, D09S10, doi:10.1029/2005JD006083.
- Soden, B., et al. (2000), An intercomparison of radiation codes for retrieving upper-tropospheric humidity in the 6.3- μ m band: A report from the first GVAP workshop, *Bull. Am. Meteorol. Soc.*, *81*(4), 797–808.
- Strow, L. L., S. E. Hannon, S. De Souza-Machado, H. E. Motteler, and D. Tobin (2003), An overview of the AIRS radiative transfer model, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 303–313.
- Tobin, D. C., H. E. Revercomb, R. O. Knuteson, B. Lesht, L. L. Strow, S. E. Hannon, E. J. Fetzer, W. F. Feltz, L. Moy, and T. Cress (2006), Atmospheric Radiation Measurement site atmospheric state best estimates for Atmospheric Infrared Sounder temperature and water vapor retrieval validation, *J. Geophys. Res.*, *111*, D09S14, doi:10.1029/2005JD006103.
- Turner, D. D., and J. Goldsmith (1999), Twenty-four-hour Raman lidar water vapor measurements during the Atmospheric Radiation Measurement program’s 1996 and 1997 water vapor intensive observation periods, *J. Atmos. Oceanic Technol.*, *16*(8), 1062–1076.
- Turner, D. D., R. A. Ferrare, L. A. H. Brasseur, and W. F. Feltz (2002), Automated retrievals of water vapor and aerosol profiles from an operational Raman lidar, *J. Atmos. Oceanic Technol.*, *19*(1), 37–50.
- Turner, D. D., B. M. Lesht, S. A. Clough, J. C. Liljegren, H. E. Revercomb, and D. C. Tobin (2003), Dry bias and variability in Vaisala RS80-H radiosondes: The ARM experience, *J. Atmos. Oceanic Technol.*, *20*(1), 117–132.
- Turner, D. D., K. L. Gaustad, S. A. Clough, and E. J. Mlawer (2004), Improved PWV and LWP retrievals from the microwave radiometer for ARM, paper presented at Fourteenth ARM Science Team Meeting, U.S. Dep. of Energy, Albuquerque, N.M., 22–26 March.
- Vömel, H., S. J. Oltmans, D. J. Hofmann, T. Deshler, and J. M. Rosen (1995), The evolution of the dehydration in the Antarctic stratospheric vortex, *J. Geophys. Res.*, *100*, 13,919–13,926.
- Vömel, H., M. Fujiwara, M. Shiotani, F. Hasebe, S. J. Oltmans, and J. E. Barnes (2003), The behavior of the Snow White chilled-mirror hygrometer in extremely dry conditions, *J. Atmos. Oceanic Technol.*, *20*(11), 1560–1567.
- Wang, J. H., D. J. Carlson, D. B. Parsons, T. F. Hock, D. Lauritsen, H. L. Cole, K. Beierle, and E. Chamberlain (2003), Performance of operational radiosonde humidity sensors in direct comparison with a chilled mirror dew-point hygrometer and its climate implication, *Geophys. Res. Lett.*, *30*(16), 1860, doi:10.1029/2003GL016985.
- Ware, R. H., D. W. Fulker, S. A. Stein, D. N. Anderson, S. K. Avery, R. D. Clark, K. K. Droegemeier, J. P. Kuettner, J. B. Minster, and S. Soroshian (2000), SuomiNet: A real-time national GPS network for atmospheric research and education, *Bull. Am. Meteorol. Soc.*, *81*(4), 677–694.
- Whiteman, D. N. (2003a), Examination of the traditional Raman lidar technique. I. Evaluating the temperature-dependent lidar equations, *Appl. Opt.*, *42*(15), 2571–2592.
- Whiteman, D. N. (2003b), Examination of the traditional Raman lidar technique. II. Evaluating the ratios for water vapor and aerosols, *Appl. Opt.*, *42*(15), 2593–2608.
- Whiteman, D. N., et al. (2006a), Raman water vapor lidar measurements during the international H₂O project. I. Instrumentation and analysis techniques, *J. Atmos. Oceanic Technol.*, *23*, 157–169.
- Whiteman, D. N., et al. (2006b), Raman water vapor lidar measurements during the international H₂O project. II. Case studies, *J. Atmos. Oceanic Technol.*, *23*, 170–183.
- C. Barnett, NOAA, Camp Springs, MD 20746, USA.
- A. S. Beebe, Department of Mathematics, University of Maryland Eastern Shore, 1 Backbone Road, Princess Anne, MD 21853, USA.
- B. Demoz, F. Schmidlin, and D. N. Whiteman, NASA/Goddard Space Flight Center, Code 613.1, Greenbelt, MD 20771, USA. (david.n.whiteman@nasa.gov)
- A. Gombacorta, S. Hannon, and F. Russo, Department of Physics, University of Maryland Baltimore County, Baltimore, MD 21250, USA.
- B. Lesht, U.S. Department of Energy/Argonne National Laboratory, Chicago, IL 60439, USA.
- L. M. Miloshevich, National Center for Atmospheric Research, Boulder, CO 80307-3000, USA.
- P. J. Moore, LJT and Associates, Inc., 9881 Broken Land Parkway, Suite 400, Columbia, MD 21046, USA.
- I. Veselovskii, Goddard Earth Sciences and Technology Center, University of Maryland Baltimore County, Baltimore, MD 21228, USA.
- H. Vömel, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA.
- Z. Wang, Department of Atmospheric Sciences, University of Wyoming, Laramie, WY 82071, USA.