Water Vapor Measurements by Howard University Raman Lidar during the WAVES 2006 Campaign

M. ADAM,*,+ B. B. DEMOZ,* D. N. WHITEMAN,# D. D. VENABLE,* E. JOSEPH,*
A. GAMBACORTA,@ J. WEI,@ M. W. SHEPHARD,& L. M. MILOSHEVICH,** C. D. BARNET,++
R. L. HERMAN,## J. FITZGIBBON,@@ AND R. CONNELL*

* *Howard University, Washington, D.C.*
\# *NASA Goddard Space Flight Center, Greenbelt, Maryland*
@ *Perot System Government Services, and NOAA/NESDIS, Camp Springs, Maryland*
& *Atmospheric and Environmental Research, Inc., Lexington, Massachusetts*
** *National Center for Atmospheric Research, Boulder, Colorado*
++ *NOAA/NESDIS, Camp Springs, Maryland*
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California
@@ *National Weather Service, Sterling, Virginia*

(Manuscript received 6 May 2009, in final form 26 August 2009)

ABSTRACT

Water vapor mixing ratio retrieval using the Howard University Raman lidar is presented with emphasis on three aspects: (i) comparison of the lidar with collocated radiosondes and Raman lidar, (ii) investigation of the relationship between atmospheric state variables and the relative performance of the lidar and sonde (in particular, their poor agreement), and (iii) comparison with satellite-based measurements. The measurements were acquired during the Water Vapor Validation Experiment Sondes/Satellites 2006 campaign. Ensemble averaging of water vapor mixing ratio data from 10 nighttime comparisons with Vaisala RS92 radiosondes shows, on average, an agreement within ±10%, up to ~8 km. A similar analysis of lidar-to-lidar data of over 700 profiles revealed an agreement to within 20% over the first 7 km (10% below 4 km). A grid analysis, defined in the temperature–relative humidity space, was developed to characterize the lidar–radiosonde agreement and quantitatively localizes regions of strong and weak correlations as a function of altitude, temperature, or relative humidity. Three main regions of weak correlation emerge: (i) regions of low relative humidity and low temperature, (ii) regions of moderate relative humidity at low temperatures, and (iii) regions of low relative humidity at moderate temperatures. Comparison of Atmospheric Infrared Sounder and Tropospheric Emission Sounder satellite retrievals of moisture with those of Howard University Raman lidar showed a general agreement in the trend, but the satellites miss details in atmospheric structure because of their low resolution. A relative difference of about ±20% is usually found between lidar and satellite measurements for the coincidences available.

1. Introduction

Water vapor is an important constituent of the atmosphere. The vertical distribution of moisture is important in determining atmospheric stability. Water vapor is also the most radiatively active atmospheric trace gas in the infrared (Ramanathan 1988) and thus could produce strong forcing from feedback associated with anthropogenically driven climate change (Cess et al. 1990). In addition, there is significant variability in the distribution of water vapor on temporal and spatial scales smaller than is currently being measured by the standard techniques [radiosonde (RS) and satellite]. A better undertanding of both the global climatology and the small-scale variability of water vapor is required to improve simulation of current and future climates by global circulation models (Moncrieff et al. 1997; Ingram 2002). Without significantly improved water vapor data, we will continue to be limited in our understanding of important moist processes

+ Current affiliation: European Commission, JRC, Ispra, Italy.

Corresponding author address: Mariana Adam, Howard University, 2355 6th St. NW, Washington, DC 20059.
E-mail: madam@howard.edu

DOI: 10.1175/2009JTECHA1331.1

© 2010 American Meteorological Society

within the atmosphere (e.g., Korolev and Mazin 2003; Peter et al. 2006; Demoz et al. 2006).

Routine measurements of atmospheric water vapor have serious limitations. Upper-air radiosondes are generally launched only twice daily. The quality of the routine global upper-air radiosonde measurement of water vapor is inadequate for many purposes such as radiation modeling and climate studies (WMO 2008). In the United States, the data provided by the standard National Weather Service (NWS) radiosonde sensors can perform poorly in cold dry regions or when the package becomes wet either in clouds or during precipitation (Wade 1994; Blackwell and McGuirk 1996; Miloshevich et al. 2006). Although current satellite remote sensing holds promise for providing high-quality global water vapor observations, it is limited by its vertical and spatial resolution. Even high-quality data from recent satellites (e.g., *Aqua* and *Aura*) do not observe the fine vertical structure of the water vapor. Raman lidars, although generally limited to a single location, provide high temporal and spatially resolved water vapor mixing ratio profiles and are capable of continuous measurements over hours or days (Turner and Goldsmith 1999). This paper discusses the temporal and spatial retrievals of the water vapor mixing ratio (WVMR) from the Howard University Raman lidar (HURL) and its comparisons with satellite, radiosonde, and a Raman lidar. It focuses on three main aspects: (i) the performance of the relatively new HURL system through comparisons with collocated Vaisala RS92 radiosondes and a National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) Raman lidar; (ii) a detailed analysis of the WVMR as measured by lidar and the standard NWS upper-air sounding package, the Sippican Mark IIA radiosonde; and (iii) comparisons between HURL and satellite (*Aura* and *Aqua*) retrievals of WVMR. The data used in this paper were collected during the Water Vapor Validation Experiment Sondes/Satellites (WAVES) experiment that was held at the Howard University Beltsville Research Campus between 7 July and 12 August 2006. The Howard University Beltsville Research Campus is located in Beltsville, Maryland (39°N, 76.9°W, around 18 km northeast of Washington, D.C.). The objective of WAVES 2006 was (i) to provide high-quality measurements of water vapor and ozone profiles for comparison with *Aura* satellite retrievals, (ii) to assess radiosondes performance, and (iii) to study upper-tropospheric water vapor measurements by Raman lidar systems. The paper is structured as follows: In section 2, a short description of the 2006 WAVES campaign and HURL system is provided. Section 3 describes the lidar WVMR calculations. Section 4 presents the data comparison results and discussion. Concluding remarks are presented in section 5.

2. WAVES campaign and the Howard University Raman lidar

a. WAVES campaign

The WAVES 2006 field campaign took place at the Howard University Research Beltsville Campus from 7 July to 10 August. Groups from 13 academic and government institutions participated. The key government collaborators included groups from National Ocean and Atmospheric Administration (NOAA)/NWS and NASA GSFC. The field campaign was intended to provide high-quality measurements of water vapor and ozone for validation of *Aura* satellite retrievals, to assess the accuracy of upper-tropospheric water vapor measurements using radiosoundings, and to observe mesoscale processes that influence local air quality and regional water vapor variability. The operations include intensive observations by multiple radiosonde–ozonesonde sensors and several lidar systems during overpasses of *Aura* and other A-Train satellites. In addition to the lidar–radiosondes operations, continuous meteorological measurements were recorded using a suite of sensors: a 31-m instrumented tower [for temperature (T), pressure, relative humidity (RH), flux, and wind]; various broadband and spectral radiometers; a two-channel microwave radiometer; a Doppler C-band radar; various aerosol chemical parameters; a 915-MHz wind profiler operated by the Maryland Department of Environment (MDE); a sun photometer [operated by the U.S. Agricultural Services and a part of Aerosol Robotic Network (AERONET) network]; and a Suominet GPS. More information on WAVES and associated experiments can be found in Whiteman et al. (2006d) and on the WAVES Web site (available online at http://www.ecotronics.com/lidar-misc/waves_06/waves06.htm).

Satellite validation is often preferred in homogeneous locations: over water or at locations where pollution is low and surface albedo has been well characterized (e.g., Tobin et al. 2006). The Beltsville facility is located in a high population area and major pollution corridor in the eastern United States. It belongs to a somewhat urban region, where a wide range of meteorological conditions occur throughout the year. It provides an environment very different than Atmospheric Radiation Measurement Program (ARM) sites. Also, great opportunities for interagency and university collaboration exists (e.g., HU–MDE–NOAA–NASA). The atmospheric measurement program at Beltsville has developed expertise in upper-air ozone sounding over the past several years through participation in the Intercontinental Chemical Transport Experiment (INTEX) Ozonesonde Network Study (IONS; Thompson et al. 2007) and the State of Maryland Department of the Environment summer air quality monitoring campaigns.

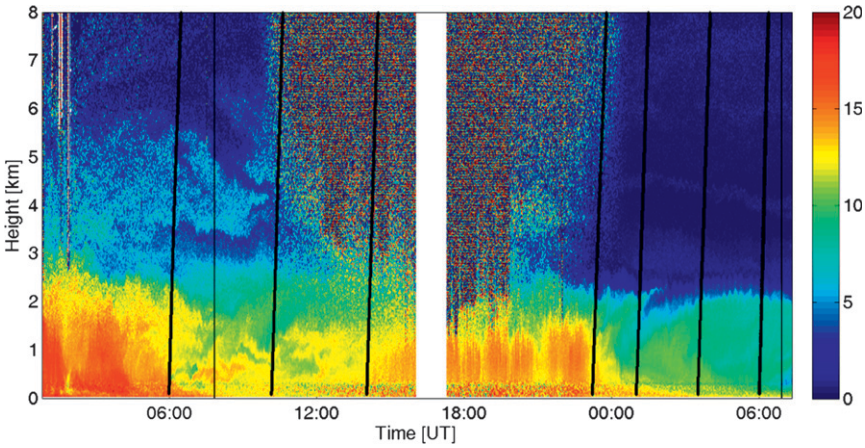

FIG. 1. Temporal series of water vapor mixing ratio from 4 and 5 Aug 2006. The thick black lines represent the trajectory of the radiosondes launched during lidar measurements. The thin black vertical lines represent the AIRS overpasses (0751 UTC 4 Aug and 0656 UTC 5 Aug).

b. HURL system

The HURL system operated over a 14-day period between 7 July and 12 August 2006 as part of WAVES 2006. Figure 1 shows an example of a time series of WVMR (g kg^{-1}) profile data covering around 30.5 h: starting at 0050 UTC 4 August and ending at 0721 UTC 5 August 2006. Temporally, convective clouds were present at the top of the planetary boundary layer (PBL; better seen in aerosol backscatter ratio), specifically over the first 3 h and during daytime operation, ~1200 and 1930 UTC 4 August. The data gap at 1600 UTC 4 August is due to HURL interruption resulting from heavy rainfall. The temporal resolution of the data is 1 min, whereas the vertical resolution is 30 m. No smoothing was applied to the data. The data represent the passage of a cold front over the site, which cleared the PBL moisture after about 0000 UTC 5 August 2006. Note the highly variable WVMR structure in the boundary layer revealed by the Raman lidar data. A rigorous study of this case supported with modeling is in progress and will be reported elsewhere. It is shown here (i) as an illustration of the HURL capability and (ii) to add temporal context to comparisons of HURL and radiosonde water vapor mixing ratio profile comparison from this case discussed later. In summary, a total of 133 HURL operational hours of data were collected, of which 84 h were during nighttime and 49 h were during daytime. Several types of radiosonde packages, collocated Raman lidars, and satellite datasets were also operated. A comparison of the HURL data with these datasets is discussed in section 4. The Howard University Raman lidar system was developed to provide both daytime and nighttime measurements of lower- and middle-tropospheric water vapor mixing ratios and aerosol scattering profiling with high temporal and spatial resolution. HURL utilizes an Nd:YAG laser that operates at the third-harmonic wavelength (354.7 nm). HURL is a narrow-field-of-view, coaxial, three-channel, fiber-optic coupled system that uses narrow bandpass (0.25 nm) filters to measure backscattered radiation at 354.7 nm and Raman scattered radiation from N_2 molecules at 386.7 nm and from water vapor molecules at 407.5 nm. HURL utilizes Licel Transient Recorders (available online at http://www.licel.com/index.html) for data acquisition that simultaneously obtain both analog (AD) and photon counting (PC) signals thereby expanding the dynamic range of the detection system. The system includes an exit/receiving window that allows operations in inclement weather. A detailed description of the system is given elsewhere (Venable et al. 2005). HURL was designed jointly by Howard University and NASA GSFC and therefore shares some technologies with that of the NASA GSFC scanning Raman lidar (SRL; Whiteman et al. 2006a). The main products of HURL are profiles of WVMR and aerosol backscatter coefficient at high temporal and spatial resolution (typically 1 min and 7.5 m, respectively). Although HURL is not designed to operate continuously, it is capable of operating for extended periods of time if needed. As mentioned earlier, the high temporal and spatial resolution is necessary in the study of the atmospheric dynamics, especially in the PBL, where the dynamics are more variable. The Raman lidar products also constitute a source of validation for the satellite water vapor mixing ratio retrieval. Here, we provide initial comparison examples with retrievals provided by Atmospheric Infrared Sounder (AIRS) on the *Aqua* satellite and Tropospheric Emission Sounder (TES) on the *Aura* satellite.

3. HURL water vapor mixing ratio calculation

Many of the steps that are required to convert raw lidar returns (photon counts) to a usable quantity such as the water vapor mixing ratio are abundantly described in the literature and will not be repeated here. However, there are some improvements in the instrument, and specific details for HURL are new; thus, they are briefly summarized in this section. These are mainly the gluing of the photon counting and analog signals, temperature sensitivity, overlap correction, and the calibration factor.

a. Gluing of signals

Two types of signals are received by HURL: PC and AD. This choice allows for analog data usage in the strong-signal (lower altitude) regions and the PC data in the weak-signal (high altitude) regions. These signals, from different altitude ranges, must be reconciled and "glued" to form a single profile. Before gluing the signals, the raw photon counts received, converted into count rates, are corrected for system dead time (pulse pileup), using a nonparalyzable assumption (for details, see Whiteman et al. 2006a). The resolving time (5 ns) was determined using the "zero offset method" (corresponding to a zero intercept) for all three channels. This value has been verified for numerous data profiles and is consistent with the pulse width of 4–5 ns of the Hamamatsu 1924 photomultiplier tubes and the high-speed (250 MHz) Licel transient recorders used in HURL (Venable et al. 2005). Background-subtracted photon count rates and analog data are selected in a region where both are considered to be performing in a reasonably linear fashion, and a regression is performed. The region is selected setting minimum and maximum thresholds for PC signal as 1 and 10 MHz for the aerosol and nitrogen channels and 0.5 and 2.5 MHz for the water vapor channel, respectively. The regression determines the gain coefficient that is then used to convert the AD scale to a "virtual" photon count rate scale (Whiteman et al. 2006a). The gluing procedure consists of two steps. First, the gluing coefficients are determined for individual profiles through regression (at least 25 points are used in regression). Second, the mean gluing coefficients are determined for each of the aerosol, nitrogen, and water vapor profiles and used for final gluing (for further details, see Adam et al. 2007). The use of the mean gluing coefficients is beneficial in regions where individual profiles may not be reliably determined (e.g., in regions where the profiles are affected by clouds). In this study, only the nighttime measurements were used to determine the gluing coefficients. For the daytime cases, gluing coefficients from the closest night were used following the previously suggested practice of calculating gluing coefficients from nighttime data and applying to daytime data (Whiteman et al. 2006a; Newsom et al. 2009).

b. Temperature sensitivity correction

When an interference filter with full width at half maximum (FWHM) less than a few nanometers (narrow filter) is used for signal detection in lidar systems to minimize daylight background, the total differential backscatter cross section is temperature sensitive (for detailed discussion, see Measures 1984; Sherlock et al. 1999; Whiteman 2003; Adam et al. 2007; Adam 2009): thus, the need to define the total system efficiency as a product of the interference filter transmission efficiency $\xi(\lambda_{X,i})$ and $\kappa(\lambda_X)$ that represents all the other system efficiencies (telescope reflectivity, transmission through the conditioning optics, quantum efficiency of the detector, other filters transmission, etc.; for details, see Adam 2009). The transmission efficiency of the interference filter is wavelength dependent (given by manufacturer), whereas $\kappa(\lambda_X)$ is considered wavelength independent within the range of the interference filter bandpass $\Delta\lambda_X$ (less than 2 nm). Subscript X stands for either laser-emitting wavelength (L) or Raman-shifted wavelength (H for water vapor and N for nitrogen), whereas i denotes a wavelength within the filter bandpass associated with the Raman-shifted wavelength λ_X. The temperature-dependent lidar equation for the Raman channel can then be written as (Whiteman 2003; Adam and Venable 2007; Adam 2009)

$$P(\lambda_X, r) = P_0 \frac{c\tau}{2} \frac{[O_X(r) A \kappa(\lambda_X) N_X(r) F_X(T)] \left[\frac{d\sigma_t(\lambda_X, \pi)}{d\Omega}\right] \xi(\lambda_X)}{r^2} \exp\left\{-\int_0^r [\alpha(\lambda_L, r') + \alpha(\lambda_X, r')]\, dr'\right\}, \quad (1)$$

where $P(\lambda_X)$ is the backscatter power (in W), P_0 is the outgoing laser power (in W), c is the speed of light (in m s^{-1}), τ is the pulse duration (in s), $O_X(r)$ is the overlap function (dimensionless), A/r^2 is the solid angle (in sr) defined by the telescope area A (in m^2) and the distance r (in m) from the lidar to backscatter, N_X is the number of molecules of the X species, α is the extinction coefficient at laser wavelength λ_L and Raman shift wavelength λ_X, and O_X is the overlap function at wavelength λ_X. The exponential term represents the round-trip

transmission from the lidar to the backscatter (molecule or particle). The total extinction coefficient α is the sum of molecular and aerosol components (m^{-1}). The molecular backscatter coefficient is the product of the total number of molecules N_X (calculated from radiosonde pressure and temperature measurements) and the total molecular differential backscatter cross section $d\sigma_t(\lambda_L, \pi)/d\Omega$ (in $m^2\ sr^{-1}$). The latter is the sum of all individual cross sections over Q, S, and O branches (Raman vibrational–rotational lines). The temperature dependence factor $F_X(T)$ at wavelength λ_X is defined in the integral form (Whiteman 2003), whereas, in practice, the computations are done using the sum (Adam and Venable 2007; Adam 2009)

$$F_X(T) = \frac{\sum_i \frac{d\sigma(\lambda_{X,i}, T, \pi)}{d\Omega}\xi(\lambda_{X,i})}{\frac{d\sigma_t(\lambda_X, \pi)}{d\Omega}\xi(\lambda_X)}, \tag{2}$$

where i goes over all spectral lines within interference filter bandpass and $d\sigma(\lambda_{X,i}, T, \pi)/d\Omega$ and $\xi(\lambda_{X,i})$ are the individual backscatter cross sections and interference filter transmission efficiency (wavelength dependent) within the filter bandpass $\Delta\lambda_X$. The temperature data are provided from radiosounding coincident with the lidar measurements.

Recall that the WVMR is defined as the ratio between the mass of water vapor and the mass of dry air. The expression of the WVMR, after overlap ratio correction is applied, simply becomes (Adam and Venable 2007; Adam 2009)

$$\mathrm{WVMR} = C\kappa(\lambda_N, \lambda_H)\left[\frac{P(\lambda_H, r)}{P(\lambda_N, r)}\right] \times \left[\frac{F_N(T)\frac{d\sigma_t(\lambda_N, \pi)}{d\Omega}\xi(\lambda_N)}{F_H(T)\frac{d\sigma_t(\lambda_H, \pi)}{d\Omega}\xi(\lambda_H)}\right]\Delta\tau(\lambda_N, \lambda_H, r), \tag{3}$$

where $C \cong 0.485$ (Whiteman et al. 2006a), $\Delta\tau(\lambda_N, \lambda_H, r)$ is the differential transmission, and $P(\lambda_N)$ and $P(\lambda_H)$ are the backscatter signals for nitrogen and water vapor [Eq. (1)]. The system calibration factor $\kappa(\lambda_N, \lambda_H)$ is the ratio of $\kappa(\lambda_N)$ to $\kappa(\lambda_H)$. The computation of $F_X(T)$ [Eq. (2)], including the calculation of the individual and total backscatter cross sections, is described in details elsewhere (Adam 2009). Following the overlap correction (next subsection), the ratio for the overlaps $O_N(r)/O_H(r)$ is considered unity in Eq. (3) and is omitted. The differential transmission was computed using radiosonde pressure and temperature to account for the molecular attenuation and a constant aerosol extinction coefficient in the PBL, derived from AERONET (Holben et al. 1998), to account for aerosol attenuation (Adam et al. 2007) of all the signals.

c. Overlap correction

In Eq. (3), the ratio of the nitrogen overlap function to the water vapor overlap function might not be unity in the region of incomplete overlap. To extract useful information from that region, a correction has to be applied. In the present study, a correction function was determined using the ratios of lidar- and radiosonde-(Vaisala RS92) measured WVMR profiles (Turner and Goldsmith 1999; Ferrare et al. 2004; Whiteman et al. 2006a,c). A set of profiles (14 in this case) over the measurement period were used to derive the correction. The optimal analytical fit for the mean ratio was determined by considering two third-order polynomial fits over two regions (altitude ranges: 45–345 and 345–1600 m) as a function of $\exp(-h)$, where h is height, in kilometers. Above 1600 m, the comparison between sonde and lidar does not show consistent differences; therefore, the overlap correction function is considered to be unity above 1600 m. A difference of up to -15% (at ~350 m) between lidar and radiosonde values was found, and the lidar WVMR profiles were corrected accordingly. More details on the experimental and the analytical fit are presented in the appendix. An additional improvement in the overlap correction and water vapor mixing ratio data quality was achieved by minimizing reflections in the laboratory of the elastic signal, which we speculate contaminated the nitrogen and water vapor channels. This new configuration of the system, which includes a baffle, which shielded the transmitted laser beam from the telescope to the exit–receiving window, eliminated the observed strong increase in the elastic backscatter signal and the deviation from unity of the lidar–sonde WVMR ratio above 400 m.

d. Raman lidar calibration factor

Further, routine calibration of lidar measured water vapor mixing ratio data by tracking the calibration factor, $\kappa(\lambda_N, \lambda_H)$ in Eq. (3), is performed by comparison of the nighttime lidar integrated precipitable water (IPW) measurements with IPW measured using a collocated microwave radiometer (MWR), as is routinely done in Raman lidar calibration (see, e.g., Turner and Goldsmith 1999; Turner et al. 2002; Ferrare et al. 2004; Whiteman et al. 2006a; Adam and Venable 2007). The MWR IPW is assumed to be reliable, with its uncertainty in the range of 0.2–0.4 mm, depending on absolute value of

IPW (Cimini 2003). The absolute IPW measurements observed during WAVES 2006 campaign varied between 10 and 60 mm, with most of the values being above 20 mm. Consequently, expected errors for the MWR IPW are less than 2%. The MWR-to-lidar IPW ratio is required for the lidar calibration factor for water vapor mixing ratio. For each period analyzed, as a first step, the mean and the standard deviation (STD) of the lidar calibration factor are computed. If the ratio (STD/mean) is >0.01, the outliers (individual calibration factors outside ±1 STD) are excluded. With the remaining set of individual calibration factors, a new mean and STD are computed. The process continues until the condition (STD/mean) <0.01 is achieved (often 1–3 steps). Note that for the 10 days of calibration calculation, this procedure eliminates a number of outliers between 0% (2 cases) and 38% (one case; here, the remaining number of individual calibration is 328). The number of remaining individual calibrations that gives the final mean varies between 91 and 428, which we believe is enough for statistics. The mean calibration factor is then used for the calibration of all the profiles during that period. As in the case of gluing coefficients, this assures that a calibration factor is obtained, even in regions where it cannot be reliably calculated because of cloud or other reasons. For the entire WAVES 2006 campaign, the mean calibration factor is $\sim$501.97 g g^{-1}, whereas the STD is $\sim$15.68 g g^{-1}. The corresponding relative error [100(STD/mean)] is found to be 3.12%. Variation of the calibration factor has been reported for a number of Raman lidar systems; for example, in the case of absolute Raman calibration, Sherlock et al. (1999) reported a 10%–12% uncertainty. For radiosounding-based calibration (using RS80 sondes), Ferrare et al. (1995) reported a 1% change, but when an Atmospheric Instrumentation Research (AIR) radiosonde was used the calibration changed by 5% over two years. Using a Meisei RS2–91 radiosonde, Sakai et al. (2007) reported an 11% change over 18 months. Whiteman et al. (2006a) report a 6% change when calibration was performed with GPS IPW. Variation of the calibration factor (3%), when performed with respect to a MWR IPW is reported by Turner and Goldsmith (1999), who combine two intensive campaigns over 1996 and 1997. The comparisons were made over 10-min averages and finally collected into 30-min bins. Ferrare et al. (2004) compared seven airborne-based lidar IPW with ground-based MWR and found a 3% change (a variable smoothing was applied to lidar data to constrain the random error within 2%–5%). The HURL calibration constant variations reported here fall in the low range of variation in the calibration constants reported, an indication of the stability of the system over time. Because system performances are expected to change over time (degradation of components, laser power, etc.), the variation will be tracked over the HURL lifetime.

4. Results and discussion

a. System performance: HURL–Vaisala radiosonde (RS92) comparisons

Ten comparisons with Vaisala RS92 radiosondes were available for nighttime operations. However, prior to comparing lidar–sonde data values, the RS92 RH measurements were corrected for known measurement errors using an algorithm similar to the time-lag and empirical bias correction described by Miloshevich et al. (2006). The RS92 data were first corrected for time-lag error (slow response of the RH sensor at low temperatures) based on laboratory measurements of the sensor time constant as a function of temperature (Miloshevich et al. 2009). Then, an empirical correction for mean calibration bias was applied, which was derived as a function of RH and altitude from dual RS92–cryogenic frost-point hygrometer (CFH) soundings conducted during several experiments (including WAVES). These corrections resulted in a mean accuracy of about ±[(4% + 0.5%)/RH] for all RH conditions throughout the troposphere and a standard deviation of RS92–CFH differences of about 5% (Miloshevich et al. 2009). The RS92 calibration bias below the 700-mb level was derived in part from comparisons of RS92 to collocated MWR retrievals of IPW using the latest MWR physical retrieval algorithm (Turner et al. 2007). Daytime RS92 measurements are also affected by a solar radiation error, which is often a dry bias caused by solar heating of the RH sensor (Vömel et al. 2007). A daytime RS92 correction for RH and height dependence of solar radiation error is derived from dual RS92–CFH soundings, and dependencies of the error on the solar altitude angle is derived from the day–night difference between the RS92 and MWR measurements, with results similar to those of Cady-Pereira et al. (2008).

In forming the HURL–RS data pairs for comparison, the lidar profiles are selected according to the radiosonde trajectory (a process referred to as RS tracking) and are shown by black lines on Fig. 1 (note that the dashed line indicate time of *Aqua* satellite overpass). The assumption in forming the RS tracking is that, at each moment (time stamp), the atmosphere is horizontally homogeneous to account for possible horizontal shifts in the RS trajectory. A version of this technique was described by Whiteman et al. (2006c) and used during the AIRS Water Vapor Experiment (AWEX) campaign, where the authors applied a variable smoothing

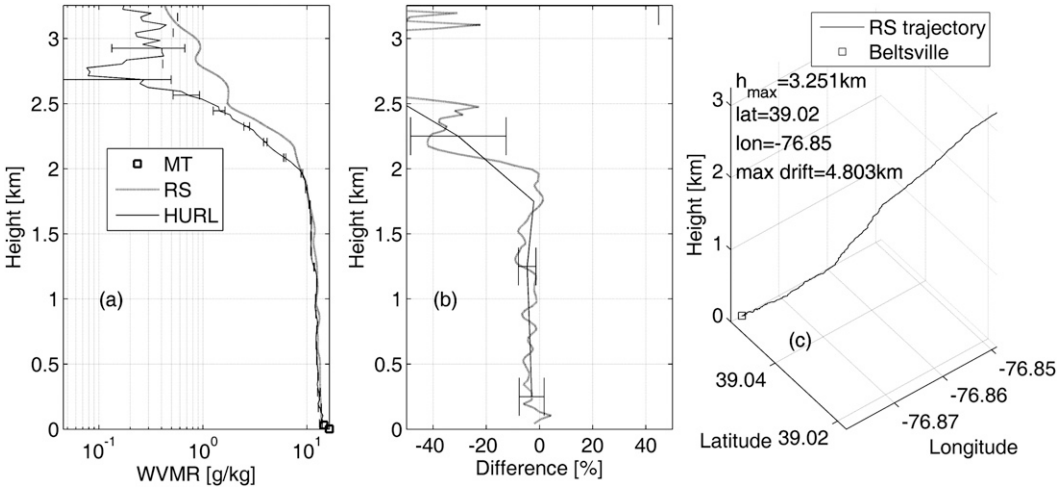

FIG. 2. (a) Water vapor mixing ratio as measured by HURL and RS92 at 2313 UTC 4 Aug 2006. (b) The relative difference between RS and HURL [100(HURL/RS) − 1] (%). The solid curves and error bars represent the mean and STD, respectively, over 500-m blocks. (c) RS trajectory.

to the lidar data. In this case, a moving average was performed over 5 temporal profiles. An additional moving average over 31 min was applied for altitudes higher than 5 km. A variable vertical smoothing was performed as follows: for altitude ranges of 1–2, 2–4, 4–6, 6–8, and ≥8 km, a moving average over 3, 5, 7, 9, and 11 bins, respectively, was performed (each bin represents 30 m). In cases of temporally homogenous datasets, the RS tracking method gives similar results to the nontracking averages (averaging time starts at RS launch time). In cases where there is significant atmospheric temporal variability over the course of the RS flight, large difference is found between the two methods. Consequently, the RS tracking method is used for the lidar comparison with RS. Our current methodology to track RS is based on two steps. First, mean values of the RS data are calculated for each layer corresponding to a lidar altitude bin (30 m). Next, the lidar data are linearly interpolated to match the RS time stamps corresponding to the RS means calculated in the first step. Two profile comparisons using data from Fig. 1 (the fourth and the last RS trajectories with launch times at 2313 UTC 4 August and 0601 UTC 5 August) are discussed to demonstrate HURL–RS92 comparisons (Figs. 2–3). The water vapor mixing ratio from two sensors on the meteorological tower (MT), at 1.5 and 31.8 m, are also shown as squares in the figures.

In the first example, the lidar retrieval in Fig. 2 is limited to about 3 km because of low signal-to-noise ratio (SNR) above this altitude resulting from daytime solar noise. In the first 2 km, an agreement within −5% is found degrading at higher altitudes mainly because of low lidar SNR (see also larger error bars in this region). The solid black curve and the error bars represent the mean relative difference and STD over 500-m blocks, whereas the dotted curve represents the mean relative difference at 30-m resolution (Fig. 2b). The maximum drift reached by the radiosonde (4.8 km) occurred at an altitude of 3.25 km, the highest altitude considered for comparison in Fig. 2c. In the second example (Fig. 3), on average, the sonde profile is 10% moister than the lidar profile. Larger differences are found in the region between 2 and 5 km as well as above 10 km, when the sonde drifted farther away (∼13 km). Note that, between 2 and 5 km, RH is ≤3%, which translates into an RS uncertainty of 21%.

The profile comparisons shown here demonstrate the range of variability that can occur when comparing lidar–sonde profiles, even within this relatively short time difference. This variability in lidar–sonde differences can be due to a combination of factors: sensor performance and/or atmospheric variability. One way of minimizing these effects of variability is to perform an ensemble average of a large number of profiles. The ensemble average is the mean over an ensemble of measurements (e.g., 10 profiles of WVMR) taken over the same altitude range but at different times. Taking into account that, for one single profile, the relative difference between two measurements (x and y) taken by two sensors is $100[(x/y) - 1]$ (%), the ensemble relative difference simply becomes $100(\langle x/y \rangle - 1)$ (%), where $\langle\rangle$ signifies the mean.

In Fig. 4, an ensemble average over 10 events is plotted. On average, a relative difference of less than ±20% below 7 km is found, increasing up to ±40% above this altitude. Moreover, the ensemble average revealed a moist bias for RS92 (or a dry bias for HURL) of about 4% over the first 2 km. Note that, in Fig. 4, the

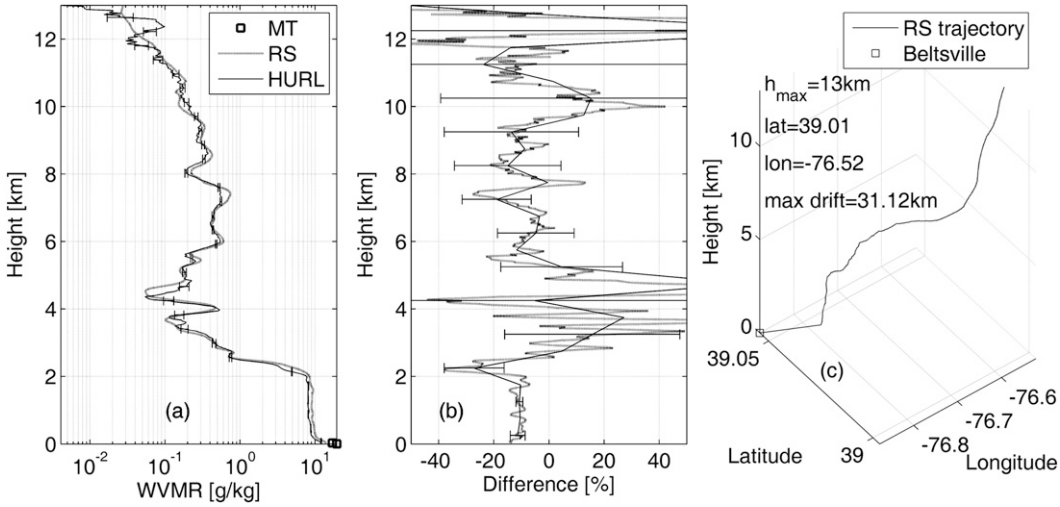

FIG. 3. As in Fig. 2, but for 0601 UTC 5 Aug 2006.

spatial resolution is 30 m (no vertical smoothing applied). The error bars represent STD over the number of profiles averaged, shown in Fig. 4b. To compare these results with other reported performances of Raman lidars, a different averaging was employed. First, relative differences are presented as averages over blocks of hundreds of meters (Whiteman et al. 2006b,c; Ferrare et al. 2004; Behrendt et al. 2007). In addition, different procedures are followed when computing the error bar of the mean profile, which is not always explicitly mentioned in the published literature. Here, we show two methods to calculate STD of the mean profile. In the first method, the mean of the relative difference is computed for each profile and for each block; then, the mean relative differences are computed (similar to Whiteman et al. 2006b). In the second method, for each block, STD is computed taking into consideration all the measurement points available (the population) from all available

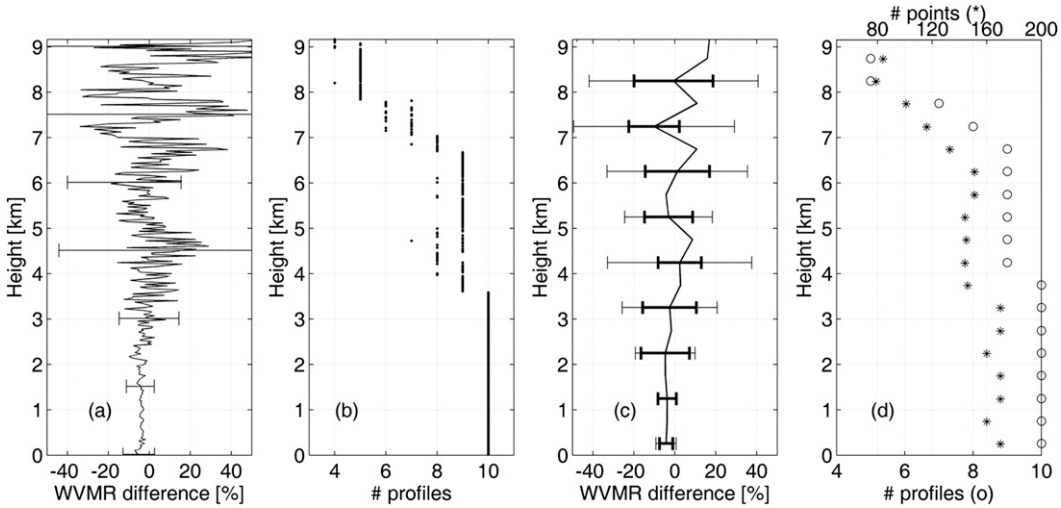

FIG. 4. Water vapor mixing ratio ensemble average for the nighttime coincidences. (a) Relative difference with respect to RS92. The error bars represent the STD over the number of profiles. (b) Number of profiles available for ensemble average. The altitude resolution for (a) and (b) is 30 m. (c) The thick error bars represent the STD over the number of profiles, whereas the thin error bars represent the STD over all measurements (points) available in a block (see text for explanation). (d) Number of profiles (open circles) and number of points (asterisks) available for ensemble average for each 500-m block. The relative difference is defined as $100(\langle \mathrm{HURL/RS} \rangle - 1)$ (%), where $\langle\rangle$ represents the mean.

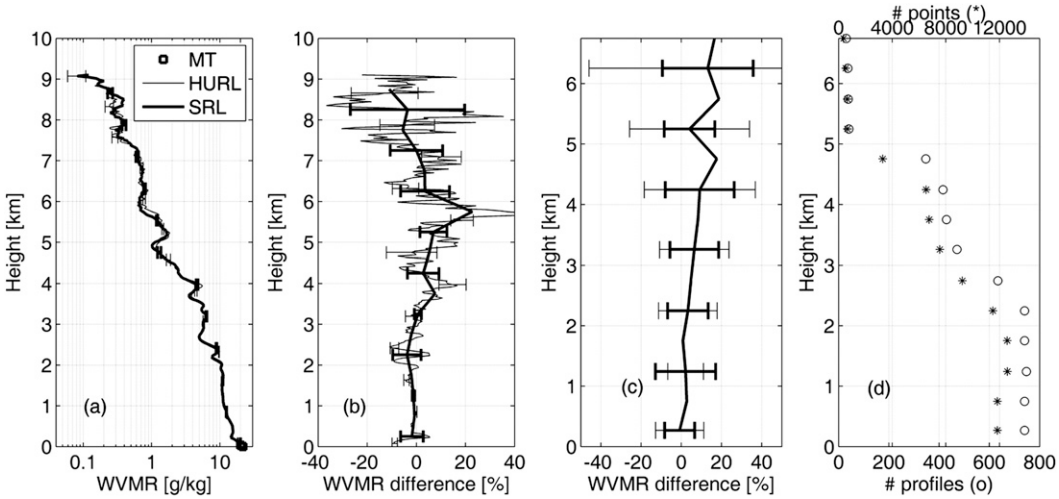

FIG. 5. (a) Water vapor mixing ratio as measured by HURL and SRL at 0713 UTC 3 Aug 2006. (b) The relative difference between SRL and HURL. The solid curves and error bars respectively represent the mean and STD over 500-m blocks. (c) Ensemble relative difference between HURL and SRL. The thin and thick errors bars are computed according to the two methods (see text). (d) Number of lidar profiles (open circles) and number of points (asterisks) available for each 500-m block. The relative difference is defined as 100[(HURL/RS) − 1] (%) for (a) and as 100(⟨HURL/RS⟩ − 1) (%) for (c), where the angle brackets ⟨⟩ represent the mean.

profiles (similar to Ferrare et al. 2004). Figure 4c shows the relative difference and the STD determined by the two methods using 500-m blocks in altitude. We can infer that the thick error bars (the first method) can be associated with the atmospheric vertical variation and the thin error bars (the second method) can be associated with both vertical and temporal atmospheric heterogeneity. The mean HURL–RS92 water vapor mixing ratio relative difference is less than ±10% over the entire region, except the uppermost block. This represents a very good result as compared with similar comparisons reported.

b. *System performance: HURL–SRL intercomparisons*

HURL and the NASA GSFC SRL (Whiteman et al. 2006a) were operated side by side. The SRL operates using the third harmonic of a Nd:YAG laser (354.7 nm). The receiving system measures returns at 354.7, 386.7, and 407.5 nm. It operates at a frequency of 30 Hz and uses PC and AD for data acquisition. SRL measures water vapor mixing ratio, aerosol backscatter and extinction coefficients, depolarization, and liquid water content using two telescopes: 25 (high channel) and 76 cm (low channel). The SRL is a well-established instrument with a long history of making well-calibrated measurements of tropospheric water vapor (see Whiteman et al. 2006a,b,c; and references therein). The SRL water vapor mixing ratio for the high channel was calibrated using the same MWR used for HURL. For the day analyzed (3 August 2006), there are almost 13 h (766 profiles) of coincidental measurements. Figures 5a,b show an example of an individual profile comparison. A temporal and vertical smoothing was applied for both lidar data as follows: First, a temporal moving average over 11 min is applied at all altitudes. Next, an additional temporal moving average over 11 profiles is applied for altitudes above 5 km. Finally, the same vertical smoothing as in the HURL–RS92 case is applied. The thick curve and error bars (Fig. 5b) represent the mean and STD over 500-m blocks. On average, a ±10% agreement was found between the lidars, except in the altitude range between 5.5 and 6 km. Note that the SRL has a better SNR at high altitudes, because it operates using a more powerful laser. Ensemble average is also performed over a maximum of 766 profiles (in the lower troposphere), with no vertical smoothing applied (Fig. 5c). The error bars corresponding to the relative differences are computed using the two methods described in the previous subsection. Several additional criteria were followed in computing the mean profiles and their relative difference. First, the lidar profiles are restricted to a region where the noise did not overwhelm the signal, choosing only the region where the WVMR is always positive. Second, only points with relative error (100 STD/mean) smaller than 30% are taken into consideration. Consequently, out of the maximum available number of profiles in the lower troposphere (766), only 30 profiles survived this restricted conditions at high altitudes (circles in Fig. 5d). A similar decrease occurs for the

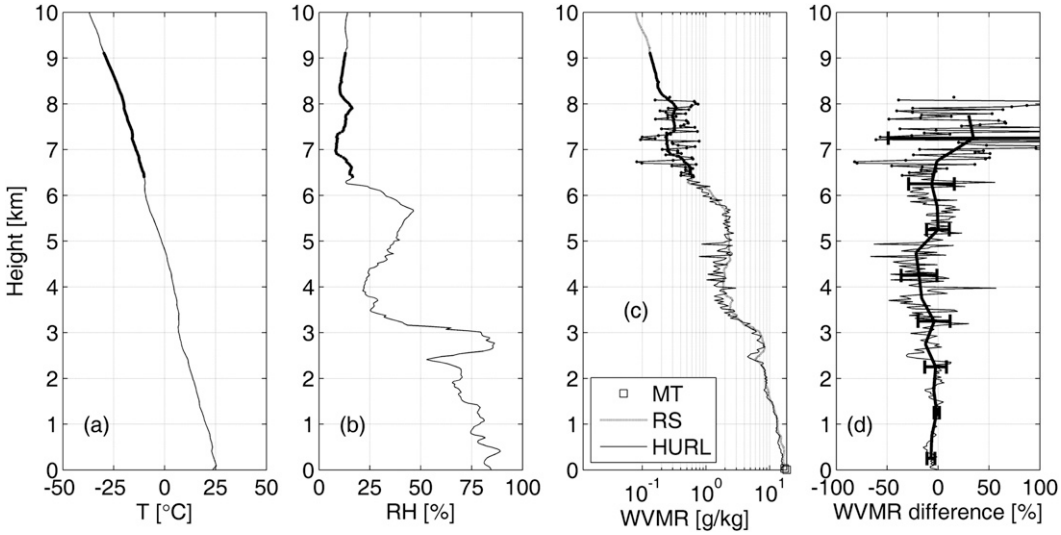

FIG. 6. Water vapor mixing ratio as measured by HURL and NWS Sippican Mark IIA at 0336 UTC 27 Jul 2006: (a) Temperature and (b) relative humidity during RS flight; (c) WVMR; and (d) relative difference with respect to RS. The thick curve represents the mean profiles over 500-m blocks. The relative difference is defined as 100[(HURL/RS) − 1] (%).

number of points when the data are averaged over 500-m blocks (asterisks in Fig. 5d). In general, a mean ensemble relative difference of less than 10% below 5 km was found, where the difference was below 20% over all altitudes. In addition, HURL WVMR profiles were slightly moister than that of SRL above 2 km, which at present is not explained. Analysis of additional data from the WAVES 2006 campaign is needed to further investigate this behavior.

c. *Grid method*

Numerous comparisons of lidar- and radiosonde-derived water vapor mixing ratios are reported in the literature. Most of these comparison studies are presented either as profile-by-profile ensemble averages or as IPW correlations. The effects of temperature and relative humidity on instrument performance are often studied separately. To visualize the instrument error characteristics in a unified form, a grid method is developed. This method is used to investigate the meteorological conditions (in terms of temperature and relative humidity) under which poor agreement occurs between the lidar and radiosonde profiles. In the present paper, the method is applied to HURL and NWS Mark IIA sondes. This study is part of a NOAA Center for Atmospheric Sciences (NCAS) and NWS long-term collaboration as part of the NWS radiosonde testing and replacement program. The goal is to validate the new sensors using HURL and other supporting observations that are performed at Beltsville. Two examples, typical of the data during this campaign, are used (Figs. 6, 7) to demonstrate the method. The temporal smoothing applied to HURL data is the same as done previously in the case of HURL–RS92 comparisons. However, no vertical smoothing is applied, keeping a 30-m spatial resolution over the entire altitude range of the analysis. The first example, from 0336 UTC 27 July 2006, is representative of the good comparisons with relative differences below ±20% up to 6 km and much smaller below 2 km (Figs. 6a–d). The second example, taken from 0330 UTC 5 August 2006, is a poor case of comparison where there is relatively large moist bias between HURL and the Mark IIA sonde data (Figs. 7a–d). The thick curves in the plots of Figs. 6d and 7d represent the mean and STD of the relative difference computed for 500-m blocks.

As is evident from Fig. 1, the lidar revealed a dry region between 2 and 6 km, where the mixing ratio decreases above the boundary layer. In such dry regions, the accuracy of the Mark IIA–derived relative humidity shows substantial errors, and these errors were observed in several profiles. This limitation of the Mark IIA may be a result of the errors in calibration, sensor hysteresis, and sensor response time (Blackmore and Taubvurtzel 1999). According to Blackmore and Taubvurtzel (1999), at low temperatures, the calibration (lock-in resistance) increases, whereas the time response slows. Low temperatures and sensor hysteresis cause errors up to 10% in RH. At transition from high to low RH, the sensor hysteresis can also induce errors within 10% in RH (drier RH). However, the present study reveals much

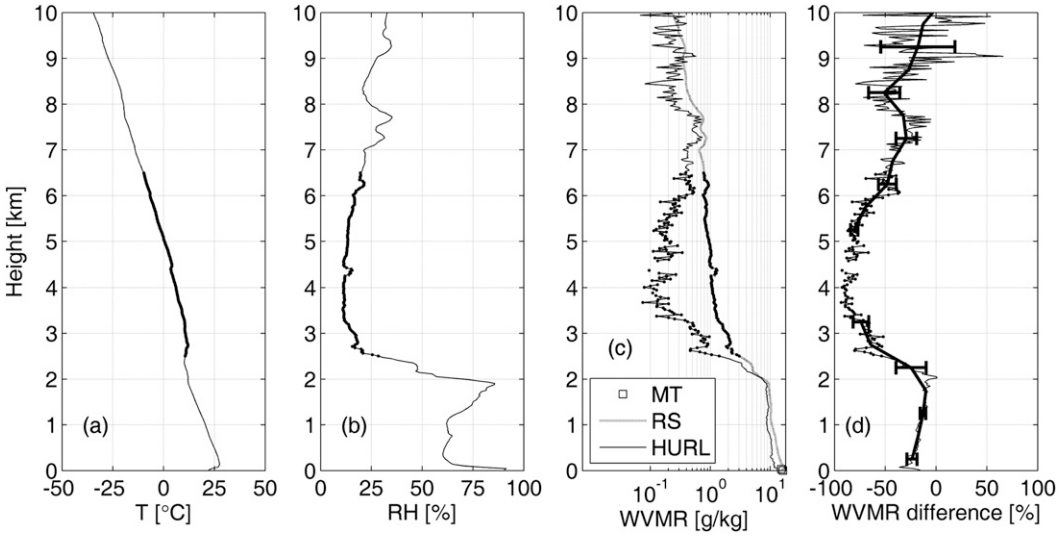

FIG. 7. As in Fig. 6, but at 0330 UTC 5 Aug 2006.

larger mixing ratio differences at the transition between high to low RH. These findings are consistent with those reported by Ferrare et al. (2004) and Sakai et al. (2007). Da Silveira et al. (2003) reported substantial large disagreements between Mark IIA and other sensors in their study of GPS–sonde intercomparison, whereas Wang et al. (2003) reported time-lag errors and failure of the sonde to respond to humidity changes in the upper and middle troposphere. Miloshevich et al. (2006) report slow time response at low temperatures and a moist bias in middle troposphere of 10%–30% as compared with RS80-H. These authors consider the measurements to be suspect between −20° and −50°C and all temperatures when operated under dry conditions. Ensemble relative difference plots in Fig. 8 clearly show these findings; a large moist bias for sonde is revealed between ~2.5 and 5.5 km.

Because the atmosphere can have several cold–dry regions as well as fast RH transition areas that result in large differences, a grid analysis in the dual-variable T–RH space was developed. This allows for quantification and easy visualization of instruments (dis)agreement, here expressed as the WVMR root-mean-square error (RMSE). Note that the temperature and relative humidity are provided by the radiosondes and the focus here is on large discrepancies (large RMSE) between the two instruments. The methodology is as follows: For each increment of ΔT = 10°C and ΔRH = 10%, we compute RMSE as

$$\mathrm{RMSE} = 100\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathrm{WVMR}_{\mathrm{HURL},i}}{\mathrm{WVMR}_{\mathrm{RS},i}} - 1\right)^2}\ (\%), \qquad (4)$$

where n represents the number of WVMR data points within each ΔT and ΔRH space. The RMSE [Eq. (4)] is then contoured in T–RH space as shown in Figs. 9 and 10. The first case shows that the largest RMSE occurs for $\Delta T = [-30, -20]$ (°C) and $\Delta\mathrm{RH} = [10, 20]$ (%):

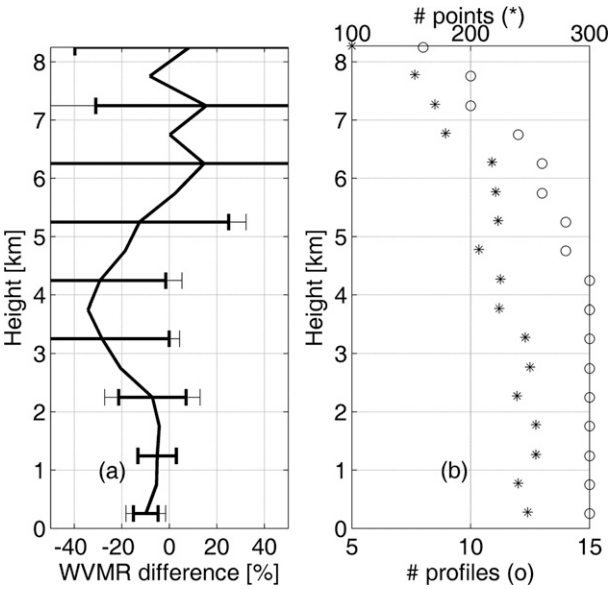

FIG. 8. (a) Ensemble relative difference with respect to Mark IIA. The thick error bars represent the STD over the number of profiles, whereas the thin error bars represent the STD over all measurements (points) available in a block. See text for explanation. (b) Number of lidar profiles (open circles) and number of points (asterisks) available for each 500-m block. The relative difference is defined as 100(⟨HURL/RS⟩ − 1) (%).

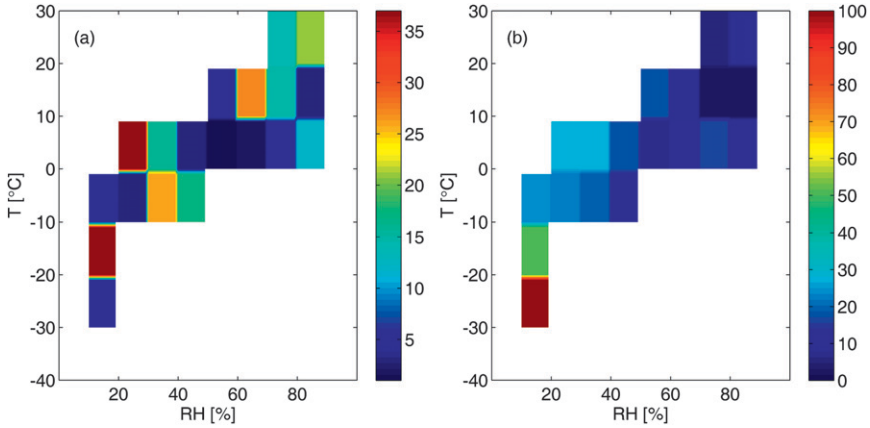

FIG. 9. HURL–Mark IIA RS grid comparison for 0336 UTC 27 Jul 2006. (a) Number of events for each grid box, characterized by $\Delta T = 10°C$ and $\Delta RH = 10\%$. (b) RMSE (%) for each grid box. The largest RMSE occurs over the region characterized by $T = [-30, -20]$ (°C) and RH = [10, 20] (%).

that is, in cold and dry regions. The second case shows the largest RMSE in the region characterized by $\Delta T = [0, 10]$ and $\Delta RH = [10, 20]$. Relative large RMSEs (>50%) occur also in boundary layer regions (relatively warm and moist) characterized by $\Delta T = [10, 20]$ and $\Delta RH = [10, 30]$. Note again that this mapping of RMSE into the T–RH space can be easily reversed, and the T–altitude and RH–altitude pairs can be extracted. A plot of the data points for which $RMSE_{threshold} = 50\%$ are shown in Figs. 6 and 7 by dots on the curves (shown as heavy lines). Thus, the box with the largest RMSE in Fig. 9 corresponds to the thick part of the curve shown in Fig. 6 (~6.5–9 km). Similarly, for Fig. 10, the regions with RMSE > 50% correspond to the thick curves in Fig. 7 (~2.5–6.5 km). In summary, the ensemble average of the RMSE (Fig. 11) reveals the largest HURL–Mark IIA discrepancies over cold and dry regions, characterized by $\Delta T = [-30, -20]$ and $\Delta RH = [10, 20]$, where RMSE reaches ~93%. Relatively large values (>50%) also occur elsewhere where RH < 60%, whereas T varies. Note the box where $\Delta T = [10, 20]$, which suggests that the region is somewhere within the boundary layer. Note also that the differences in cold and dry regions should not be attributed to Mark IIA sonde (inadequate RH sensor response at low T) alone; lidars also are affected by the low SNR and low quantity of water vapor molecules. In other conditions, the difference is primarily attributed to the inadequacies in Mark IIA RS sensor response in high gradient moisture regions (as from high to low RH). More detailed analyses can and should be performed by choosing a higher-resolution grid defined by smaller ΔT and ΔRH intervals than used here, provided that a

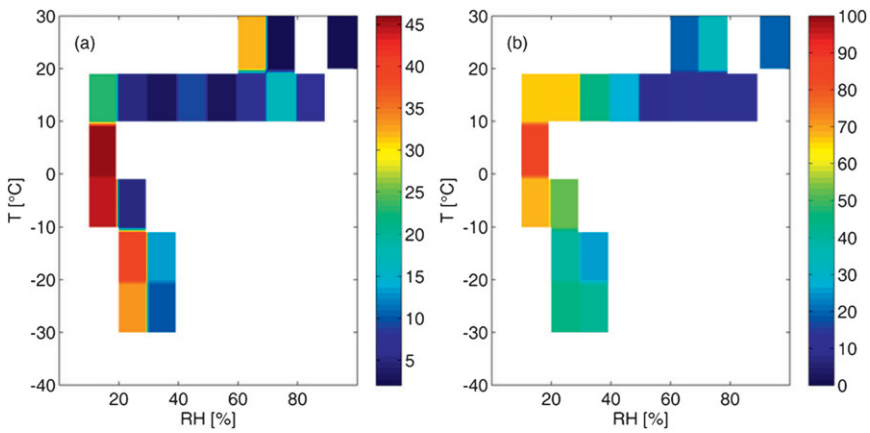

FIG. 10. As in Fig. 9, but for 0330 UTC 5 Aug 2006. The largest RMSE occurs over the region characterized by $T = [0, 10]$ (°C) and RH = [10, 20] (%).

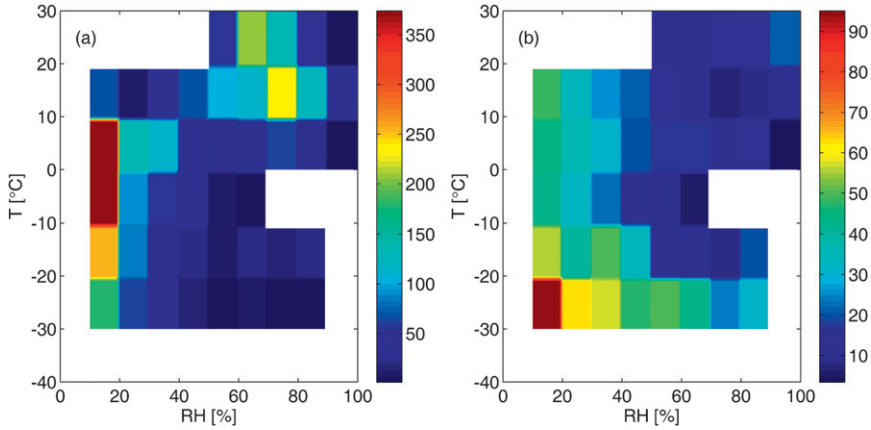

FIG. 11. HURL–Mark IIA RS grid comparison for the ensemble average over all 15 comparisons. (a) Number of events for each grid box, characterized by $\Delta T = 10°C$ and $\Delta RH = 10\%$. (b) RMSE (%) for each grid box. The largest RMSE occurs over the region with $T = [-30, -20]$ (°C) and RH = [10, 20] (%; RMSE = 95%). Relatively large RMSE can be observed over regions with $T = [-30, -20]$ and RH = [20, 30]; (RMSE = 62.4%) and other regions with $T = [-30, -20]$ and RH = [20, 70] or $T = [-20, 20]$ and RH = [10, 20].

statistically significant number of data points exist for each grid box.

d. WVMR: HURL–satellites

As mentioned earlier, one of the main objectives of the WAVES campaign was to provide ground-based measurements for the validation of the *Aura* sensors. However, because the *Aura* and *Aqua* overpasses occur only 15 min apart, both are studied. The WVMR profiles derived from lidar are compared with those from TES on *Aura* and AIRS on *Aqua*. AIRS and TES are infrared spectrometers. AIRS is a scanning instrument, whereas TES is a nadir-viewing instrument. In brief, AIRS was launched to provide temperature and water vapor profiles or spectral radiances for assimilation in numerical weather prediction and to provide an improved understanding of the atmospheric branch of the hydrological cycle and climate processes (Aumann et al. 2003; Fetzer 2006; and references therein). An ultimate goal of the AIRS validation effort is to achieve WVMR RMSE uncertainties of 10% over 2-km layers in the troposphere (Fetzer et al. 2003; Tobin et al. 2006). For TES, the main objective is to measure the global profiles of tropospheric ozone and its precursors, among which water vapor is particularly important (Shephard et al. 2008, and references therein). The criteria used for selection of profiles in this study are such that the *Aura* satellite ground track lies within 50 km of the Beltsville site. Note also that the daytime comparisons were restricted to below 5-km altitude, because of either low SNR in the lidar signals or the presence of convective clouds in the boundary layer, whereas at night the altitude range for comparison extended on average to 10 km.

AIRS version 5 tropospheric moisture retrieval resolution as determined by the FWHM of the averaging kernels ranges between 2.7 km near the surface and 4.3 km near the tropopause (Maddy and Barnet 2008), which is similar to TES performance (Shephard et al. 2008). In addition, AIRS moisture retrieval degrees of freedom for nominal midlatitude cases is nearly 4.0, which is also very close to the TES moisture retrieval reported in Shephard et al. (2008). We therefore would expect similar performance in the AIRS and TES water retrievals if we accounted for the a priori dependence of the AIRS retrievals using averaging kernels (Maddy and Barnet 2008). For consistency with previous AIRS water vapor validation efforts (Whiteman et al. 2006c), we have chosen to compare the AIRS retrievals using traditional simple layer techniques (i.e., without the use of averaging kernels). In this study, AIRS level 2 products are used where temperature and water vapor profiles are reported on 100 vertical grid layers. The calculation of mean mixing ratio within a layer takes into account the conservation of the number of molecules within each layer and is given by the ratio of the water vapor to dry air column densities.

The TES temperature and water vapor volume mixing ratio are reported on a standard 67 pressure level grid. The TES footprint, at nadir, is 8×5 km. TES utilizes an optimal estimation retrieval approach to simultaneously minimize the difference between observed and model spectral radiances to estimate atmospheric profiles (Bowman et al. 2006). With each TES retrieved profile,

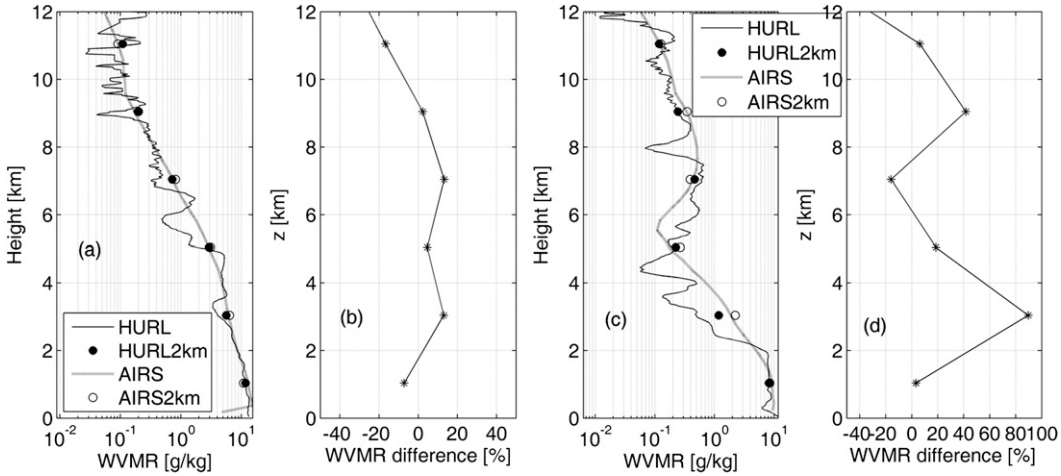

FIG. 12. HURL–AIRS water vapor mixing ratio for (a) 4 (*Aqua* overpass 0751 UTC) and (c) 5 Aug 2006 (*Aqua* overpass 0656 UTC). The HURL profiles are shown at both original resolution and averaged over 2-km layers. (b),(d) Relative difference with respect to HURL within 2-km layers for the same times as in (a),(b). The relative difference is defined as 100[(AIRS/HURL) − 1] (%).

the corresponding averaging kernel, which describes the sensitivity of the retrieval, is provided. The FWHM of the rows of the averaging kernels provide the vertical resolution of the retrieval. To help provide some additional insight, the sum of the rows of the averaging kernels can be thought of as the fraction of information in the retrieval that comes from the measurement rather than the a priori. Note that the sensitivity of TES varies profile to profile depending on atmospheric state (concentration of the species of interest, temperature, etc.), clouds, and constraints used in the retrieval. Because the goal of these comparisons is to validate the satellite measurements, the TES averaging kernels and a priori profile were applied to the sondes and/or lidar profiles to account for the a priori bias, sensitivity, and vertical resolution of the TES retrievals (Shephard et al. 2008). The TES standard procedure maps the in situ measurements (radiosonde or lidar) to the TES reported grid levels and then applies the TES averaging kernels A_{xx} and the a priori profile X_a to the mapped in situ profile $X_{\text{insitu}}^{\text{mapped}}$

$$X_{\text{insitu}}^{\text{est}} = X_a + A_{xx}(X_{\text{insitu}}^{\text{mapped}} - X_a), \tag{5}$$

Using this method, we obtain a lidar or radiosonde profile that represents what TES would ''see'' for the same atmospheric state (thus yielding a profile that accounts for TES sensitivity and vertical resolution). The TES vertical resolution, computed at FWHM of the averaging kernels (Shephard et al. 2008), is ~2 km from the surface up to close to 4 km, ~3 km between 4 km and close to 8 km, and 3.5–4 km between 8 and 11 km.

During the WAVES 2006 campaign, 13 coincidences were found for AIRS–HURL water vapor mixing ratio comparison. Two examples of such a comparison for the nighttime cases (4 August, overpass time 0751 UTC, and 5 August, overpass time 0656 UTC) are shown in Fig. 12. The averaging was performed over 2-km layers to have a direct comparison with previous studies by the AIRS community (e.g., Tobin et al. 2006). The average over the 2-km layer was performed as follows: Within each 2-km layer, the mean value was determined as the integral of all available points in the layer divided by the thickness of the layer (2 km). The common practice in the AIRS community of weighting the statistics by the water vapor layer amounts was not applied here; thus, the relative differences between the lidar and AIRS are more comparable to the larger nonweighted results reported by Tobin et al. (2006). The *Aura* ground track was 48.5 km away from Beltsville on 4 August and 30.96 km away on 5 August. In addition, the HURL-derived time–height evolution of mixing ratio (Fig. 1) during the overpass time shows a strong temporal and vertical variability over the region, including the presence of dry and moist layers above (e.g., ~3 km, 5–6 km). Note that averaging the HURL data in 2-km layers to compare with the AIRS resolution removes a lot of the atmospheric variability and the small-scale structures in the profile. The HURL–AIRS comparisons show relative differences below ±20% for the 4 August case. For the August 5 case, large differences occur around 3 km, most likely associated with the frontal surface variability. As can be seen in the original HURL profile (and also in Fig. 1), the lidar reveals a quick reduction in the

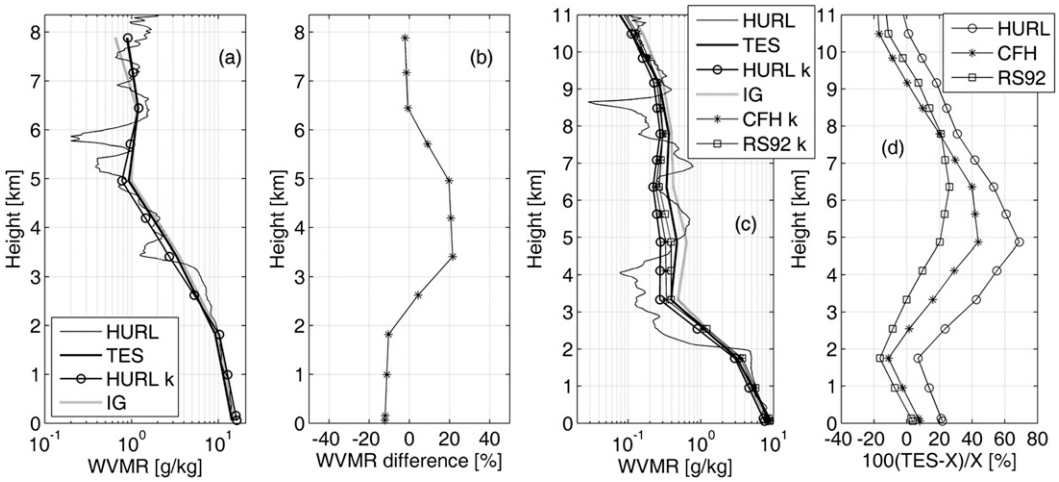

FIG. 13. (a) HURL–TES water vapor mixing ratio for 11 Jul 2006, *Aura* overpass 0715 UTC. (b) Relative difference with respect to HURL after applying a priori profile and kernels. (c) HURL–TES–CFH–RS92 water vapor mixing ratio for 12 Aug 2006, *Aura* overpass 0716 UTC. (d) Relative difference with respect to HURL, CFH, and RS92 after applying a priori profile and kernels, where X stands for HURL, CFH, or RS92. The relative difference is defined as $100[(\mathrm{TES}/X) - 1]$ (%), where X is (b) HURL and (d) HURL, CHF, or RS92. HURL k, CFH k, and RS92 k are computed with Eq. (5).

water vapor mixing ratio results above the elevated moist layer lifted because of the cold frontal surface (above 2 km). Thus, the layer-averaged value in the 2–4-km altitude region is much smaller than the retrieved value by AIRS, which has a much wider footprint as discussed previously. On average, AIRS captures the general shape of the lidar profiles but fails to catch the finer atmospheric because of its inherent lower spatial resolution. Note also that the small-scale structures reported by HURL during this time are also present in the RS data at 0600 UTC. Moreover, the values recorded at the meteorological tower (at 1.5 and 31.8 m) agree with lidar and RS data close to the ground. In the past, a relative difference up to 30% between AIRS and SRL is reported by Whiteman et al. (2006c) for ensemble average and over a variable vertical resolution (1-km layers below 4 km and 2-km layers above). Ensemble average over all the WAVES 2006 six nighttime cases (not shown here) resulted in a large relative difference (±70%) centered at about 3 km and mostly because of contributions from the cases of 0656 UTC 5 August 2006 and 0702 UTC 12 August 2006. Outside this midtroposphere region, the relative errors are generally within ±20%. These results are similar to the relative difference of 30% between AIRS and SRL reported by Whiteman et al. (2006c) and the 20% (nonwater vapor layer weighted statistics) bias reported by Tobin et al. (2006) when comparing AIRS with radiosondes below 400 mb (~7.5 km). The HURL–AIRS bias increases to −10% around the 200-mb (~12 km) altitude region, and the calculations were performed using 2-km block averages and 1500 pairs of data. However, a more robust set of comparisons is needed to get a statistically significant result. In the future, we plan to complete the analysis over the whole three years of the experiment. Note that the mean values in the layers reported by Tobin and Whiteman are computed slightly differently than our approach; thus, a direct comparison between various studies is always questionable.

The next examples show water vapor mixing ratio profile comparison from HURL and TES. A study by Shephard et al. (2008) describes the TES–radiosondes comparisons at Beltsville. As in the previous study, version V003 of the TES data was used in these comparisons. There were seven TES overpasses that occurred during HURL lidar operations in WAVES 2006. After excluding the cloudy sky cases, only two cases for each daytime and nighttime overpass were available for comparison. For the daytime comparisons, the altitude ranges were 4 and 5 km, respectively; for nighttime comparisons, the altitude ranges were 8 and 10 km, respectively. The HURL–TES nighttime comparisons (0715 UTC 11 July 2006 and 0716 UTC 12 August 2006) are plotted in Fig. 13. TES products are converted from volume mixing ratio to mass mixing ratio. In the first example, the TES ground track was 31.1 km away from HURL. In addition, thin cirrus clouds were present around 8–9 km at the time of the overpass. In the second example, the TES ground track was 0.39 km away from the site. In the figure, both the lidar profile at its standard high vertical resolution (showing the fine water vapor profile

structure) and the smoothed profile (to match the TES sensitivity) are shown. As in the case with the AIRS comparisons, the fine vertical structure caught by HURL is not seen in TES retrievals, which is expected because of satellite vertical resolution. In the first case, the TES retrieval shows smaller values (~−10% difference) in the first 2 km compared to HURL; above 3 km, the values are larger by up to 20%. In the second example, the relative difference is larger, reaching 40%–68% over a 3–7-km range. Note that the HURL–CFH comparison (not shown here) revealed large differences above 3 km, with a systematic bias increasing with height (lidar moister). On the other hand, HURL comparisons with RS92 shows better match, with lidar being drier. In Figs. 13c,d, we have added comparisons of TES with CFH and RS92. As observed, for this particular case, TES shows a better agreement with RS92. Note that both radiosondes were launched at 0601 UTC. Further investigation is needed to compare and interpret the large differences between various sensors (HURL, SRL, RS92, CFH, AIRS, and TES). Also note that this second case was presented by Shephard et al. (2008) through the comparison with CFH. Their results for this case show the maximum relative difference was found to be ~30% over the same region (specifically around 500 mb). The authors also report ensemble average of TES and radiosonde data. The results show a mean relative difference of about 5% in the lower troposphere and about 20% in the upper troposphere (note that these statistics were not weighted by the water vapor layer amount). The comparisons were made for 21 TES–RS coincidences within 150 km and within 1.5 h of sonde launch. Work is in progress to validate TES retrievals using both ground-based and airborne lidar systems and radiosonde data for which TES overpasses are within less than 50 km from the ground site and within less than 1 h.

The entire dataset over three years of WAVES experiments (2006–08) will provide a robust and statistically significant set of HURL data available for satellites comparisons. Also, because the standard comparison methods used for the AIRS and TES are different, the differences in the performance of the AIRS and TES retrieval algorithms are beyond the scope of this paper.

5. Conclusions

One of the HURL goals during the WAVES 2006 campaign was to test its performance. In the present study, HURL performance is compared to collocated Vaisala radiosonde (RS92); standard NWS Mark IIA radiosonde packages; satellite measurements from AIRS on *Aqua* and TES on *Aura* satellites; and a more established Raman lidar, the scanning Raman lidar (SRL) from NASA GSFC. On average, a relative difference between HURL and RS92 below ±10% is obtained for altitudes up to 8 km. The relative difference with respect to SRL is on average less than 20% over ~7 km and less than 10% below 4 km. Within the Howard University–NOAA collaboration, one goal was to test the new NWS sensors and validate them with respect to HURL measurements. Within these analyses, a grid method was developed to reveal regions with strong or weak agreement (quantified by RMSE) and characterize them in terms of T and RH. Ensemble averages over 15 cases showed two main regions where large discrepancies occur. Although one occurs at cold T and low RH (where usually RS RH sensors do not respond properly, whereas lidar has a poor SNR) the other occurs at either milder T (−20 to +10) (in °C) and low RH (10–20) (in %) or low temperatures (−30 to −20) (°C) and larger RH (20–70) (%). The typical situation for the latter case is when a strong gradient occurs in RH (usually above the PBL), where we speculate that the RS RH sensor does not respond accurately at this change. Further investigations as well as laboratory tests are required to confirm our suppositions.

HURL compared relatively well with satellite retrievals from two satellite sensors (AIRS on *Aqua* and TES on *Aura*). The main water vapor mixing ratio trend is captured in the satellite data, but the details of the atmospheric layers, shown by the lidar, were not caught because of their low vertical resolution. In general, discrepancies on the order of ~±20% are found between lidar and both AIRS and TES. As mentioned, AIRS and TES retrievals follow different approaches; thus, a comparison of performance of the AIRS and TES retrieval algorithms is beyond the scope of this paper. HURL–AIRS comparisons were made in 2-km atmospheric layers, whereas TES comparisons were performed on satellite levels and from TES prospective (i.e., applying averaging kernels and a priori profiles). Although generally in agreement in the moisture trends and within ±20% relative difference, in specific cases (Figs. 12d, 13d) maximum relative differences of ~90% and ~70% were found for AIRS and TES, respectively. Further investigation into obtaining robust HURL–AIRS statistics is needed and is underway. Although RS provides a high spatial resolution over a range up to 20–30 km as support for satellite validation, a Raman lidar can provide high-resolution profiles over the lower and middle troposphere (~>10 km, depending on the lidar SNR). The satellite and lidar observations are complementary in trying to monitor the atmospheric water vapor, because the satellites provide global spatial coverage, and the lidar provides high vertical and spatial observations at a single location. However, care has to

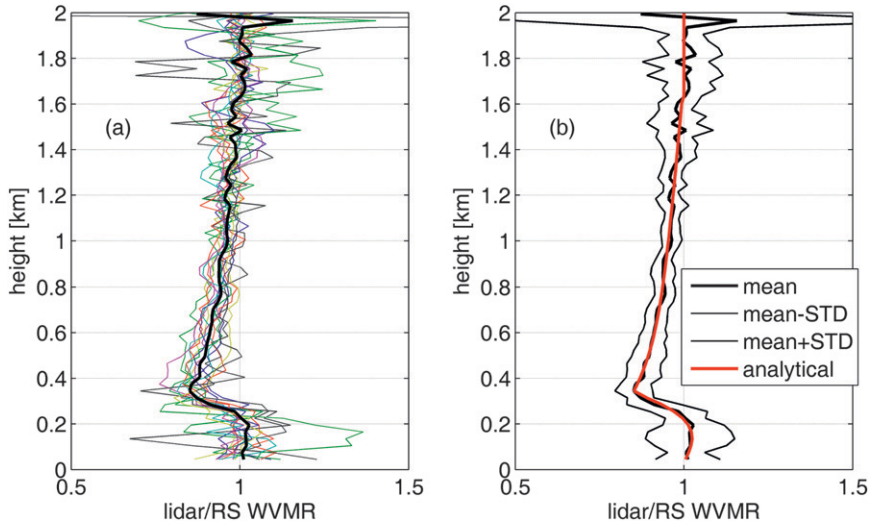

FIG. A1. The overlap correction: (a) individual and the mean corrections and (b) the experimental mean and its STD and the analytical fit.

be taken when performing comparisons between the measurements because of their different resolutions.

Acknowledgments. This research was partially supported by the NOAA Educational Partnership Program Cooperative Agreement and the NASA Aura Validation Program.

APPENDIX

Overlap Correction

The overlap correction was determined using the ratio of WVMR as calculated from the lidar and RS. To derive the correction, 14 profiles over the measurement period were used (Fig. A1a). The optimal analytical fit for the mean ratio was determined by considering two third-order polynomial fits over two regions (altitude ranges of 45–345 and 345–1600 m) as a function of $\exp(-h)$, where h is the height (in km). The ratio is considered unity above 1600 m, where, on average, the experimental ratio is unity. The two third-order polynomial fits are

$$O_1(h) = 19.50[\exp(-h)]^3 - 55.06[\exp(-h)]^2 + 51.52[\exp(-h)]^1 - 14.96,$$
$$0.045\,\text{km} < h < 0.345\,\text{km}, \quad \text{and}$$

$$O_2(h) = -0.94[\exp(-h)]^3 + 1.08[\exp(-h)]^2 - 0.62[\exp(-h)]^1 + 1.08,$$
$$0.345\,\text{km} < h < 1.6\,\text{km}.$$

The relative difference between the mean experimental overlap and the analytical fit (merging the two fits) is below ±3%, except in the region around 2 km, where it is almost ±15% (Fig. A1b).

As mentioned, the late adding (during 2007) of a baffle (carbon paper) changed overlap ratio $O_N(r)/O_H(r)$ such that the complete overlap starts at ~400 m. The baffle reduced the reflections and other secondary effects, which artificially increased the individual signals (especially in the elastic channel). Figure A2 shows the setup of the baffle.

FIG. A2. Lidar setup, including the telescope, the periscope, and the baffle. On the top of the baffle, there is the exiting–receiving window.

REFERENCES

Adam, M., 2009: Notes on temperature-dependent lidar equations. *J. Atmos. Oceanic Technol.,* **26,** 1021–1039.

——, and D. D. Venable, 2007: Systematic distortions in water vapor mixing ratio and aerosol scattering ratio from a Raman lidar. *Lidar Technologies, Techniques, and Measurements for Atmospheric Remote Sensing III,* U. N. Singh and G. Pappalardo, Eds., International Society for Optical Engineering (SPIE Proceedings, Vol. 6750), doi:10.1117/12.738205.

——, ——, R. Connell, E. Joseph, D. N. Whiteman, and B. B. Demoz, 2007: Performance of the Howard University Raman lidar during 2006 WAVES campaign. *J. Optoelectron. Adv. Mater.,* **9,** 3522–3528.

Aumann, H. H., and Coauthors, 2003: AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems. *IEEE Trans. Geosci. Remote Sens.,* **41,** 253–264.

Behrendt, A., and Coauthors, 2007: Intercomparison of water vapor data measured with lidar during IHOP_2002. Part I: Airborne to ground-based lidar systems and comparisons with chilled-mirror hygrometer radiosondes. *J. Atmos. Oceanic Technol.,* **24,** 3–21.

Blackmore, W., and B. Taubvurtzel, 1999: Environmental chamber tests of NWS radiosonde relative humidity sensors. Preprints, *15th Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology,* Dallas, TX, Amer. Meteor. Soc., 259–262.

Blackwell, K. G., and J. P. McGuirk, 1996: Tropical upper-tropospheric dry regions from TOVS and rawinsondes. *J. Appl. Meteor.,* **25,** 464–481.

Bowman, K. W., and Coauthors, 2006: Tropospheric emission spectrometer: Retrieval method and error analysis. *IEEE Trans. Geosci. Remote Sens.,* **44,** 1297–1307, doi:10.1109/TGRS.2006.871234.

Cady-Pereira, K. E., M. W. Shephard, D. D. Turner, E. J. Mlawer, S. A. Clough, and T. J. Wagner, 2008: Improved daytime column-integrated precipitable water vapor from Vaisala radiosonde humidity sensors. *J. Atmos. Oceanic Technol.,* **25,** 873–883.

Cess, R. D., and Coauthors, 1990: Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. *J. Geophys. Res.,* **95** (D10), 16 601–16 615.

Cimini, D., 2003: Accuracy of ground-based microwave radiometer and balloon-borne measurements during the WVIOP2000 field experiment. *IEEE Trans. Geosci. Remote Sens.,* **41,** 2605–2615.

Da Silveira, R. B., G. Fisch, L. A. T. Machado, A. M. Dall'Antonia Jr., L. F. Sapucci, D. Fernandes, and J. Nash, 2003: Executive summary of the WMO intercomparison of GPS radiosondes. WMO Instruments and Observing Methods Rep. 76, WMO/TD 1153, 17 pp. [Available online at http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-76-GPS-RSO/Intercomp-RSO-Brazil2001-ExecSummary.pdf.]

Demoz, B., and Coauthors, 2006: The dryline on 22 May 2002 during IHOP_2002: Convective-scale measurements at the profiling site. *Mon. Wea. Rev.,* **134,** 294–310.

Ferrare, R. A., S. H. Melfi, D. N. Whiteman, K. D. Evens, F. J. Schmidlin, and D. O'C. Starr, 1995: A comparison of water vapor measurements made by Raman lidar and radiosondes. *J. Atmos. Oceanic Technol.,* **12,** 1177–1195.

——, and Coauthors, 2004: Characterization of upper-troposphere water vapor measurements during AFWEX using LASE. *J. Atmos. Oceanic Technol.,* **21,** 1790–1808.

Fetzer, E. J., 2006: Preface to special section: Validation of Atmospheric Infrared Sounder observations. *J. Geophys. Res.,* **111,** D09S01, doi:10.1029/2005JD007020.

——, and Coauthors, 2003: AIRS/AMSU/HSB validation. *IEEE Trans. Geosci. Remote Sens.,* **41,** 418–431.

Holben, B. N., and Coauthors, 1998: AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sens. Environ.,* **66,** 1–16.

Ingram, W. J., 2002: On the robustness of the water vapor feedback: GCM vertical resolution and formulation. *J. Climate,* **15,** 917–921.

Korolev, A. V., and I. P. Mazin, 2003: Supersaturation of water vapor in clouds. *J. Atmos. Sci.,* **60,** 2957–2974.

Maddy, E. S., and C. D. Barnet, 2008: Vertical resolution estimates in version 5 of AIRS operational retrievals. *IEEE Trans. Geosci. Remote Sens.,* **46,** 2375–2384, doi:10.1109/TGRS.2008.917498.

Measures, R. M., 1984: *Laser Remote Sensing: Fundamentals and Applications*. John Wiley & Sons, 510 pp.

Miloshevich, L. M., H. Vömel, D. Whiteman, B. Lesht, F. J. Schmidlin, and F. Russo, 2006: Absolute accuracy of water vapor measurements from six operational radiosondes types launched during AWEX-G and implications for AIRS validation. *J. Geophys. Res.,* **111,** D09S10, doi:10.1029/2005JD006083.

——, ——, D. N. Whiteman, and T. Leblanc, 2009: Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements. *J. Geophys. Res.,* **114,** D11305, doi:10.1029/2008JD011565.

Moncrieff, M. W., S. K. Krueger, D. Gregory, J.-L. Redelsperger, and W.-K. Tao, 1997: GEWEX Cloud System Study (GCSS) Working Group 4: Precipitating convective cloud systems. *Bull. Amer. Meteor. Soc.,* **78,** 831–845.

Newsom, R. K., D. D. Turner, B. Mielke, M. Clayton, R. Ferrare, and C. Sivaraman, 2009: Simultaneous analog and photon counting detection for Raman lidar. *Appl. Opt.,* **48,** 3903–3914.

Peter, T., C. Marcolli, P. Spichtinger, T. Corty, M. B. Baker, and T. Koop, 2006: When the dry air is too humid. *Science,* **314,** 1399–1402.

Ramanathan, V., 1988: The greenhouse theory of climate change: A test by an inadvertent global experiment. *Science,* **240,** 293–299.

Sakai, T., T. Nagai, M. Nakazato, T. Matsumura, N. Orikasa, and Y. Shoji, 2007: Comparisons of Raman lidar measurement of tropospheric water vapor profiles with radiosondes, hygrometers, on the meteorological observation tower, and GPS at Tsukuba, Japan. *J. Atmos. Oceanic Technol.,* **24,** 1407–1423.

Shephard, M. W., and Coauthors, 2008: Comparison of Tropospheric Emission Spectrometer water vapor retrievals with in situ measurements. *J. Geophys. Res.,* **113,** D15S24, doi:10.1029/2007JD008822.

Sherlock, V., A. Hauchecome, and J. Lenoble, 1999: Methodology for the independent calibration of Raman backscatter water-vapor lidar systems. *Appl. Opt.,* **36,** 5816–5837.

Thompson, A. M., and Coauthors, 2007: Intercontinental Chemical Transport Experiment Ozonesonde Network Study (IONS) 2004: 2. Tropospheric ozone budgets and variability over northeastern North America. *J. Geophys. Res.,* **112,** D12S13, doi:10.1029/2006JD007670.

Tobin, D. C., and Coauthors, 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. E. M. 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,** 1062–1076.

——, R. A. Ferrare, L. A. Heilman Brasseur, W. F. Feltz, and T. P. Tooman, 2002: Automated retrievals of water vapor and aerosol profiles from an operational Raman lidar. *J. Atmos. Oceanic Technol.,* **19,** 37–50.

——, S. A. Clough, J. C. Liljegren, E. E. Clothiaux, K. E. Cady-Pereira, and K. L. Gaustad, 2007: Retrieving liquid water path and precipitable water vapor from the Atmospheric Radiation Measurement (ARM) microwave radiometers. *IEEE Trans. Geosci. Remote Sens.,* **45,** 3680–3690.

Venable, D. D., E. Joseph, D. N. Whiteman, B. Demoz, R. Connell, and S. Walford, 2005: Development of the Howard University Raman Lidar. Preprints, *Second Symp. on Lidar Atmospheric Applications,* San Diego, CA, Amer. Meteor. Soc., P1.1. [Available online at http://ams.confex.com/ams/pdfpapers/85650.pdf.]

Vömel, H., D. E. David, and K. Smith, 2007: Accuracy of tropospheric and stratospheric water vapor measurements by a cryogenic frost point hygrometer: Instrumental details and observations. *J. Geophys. Res.,* **112,** D083305, doi:10.1029/2006JD007224.

Wade, C. G., 1994: An evaluation of problems affecting the measurement of low relative humidity on the United States radiosonde. *J. Atmos. Oceanic Technol.,* **11,** 687–700.

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,** 1860, doi:10.1029/2003GL016985.

Whiteman, D. N., 2003: Examination of the traditional Raman lidar technique. I. Evaluating the temperature-dependent lidar equations. *Appl. Opt.,* **42,** 2571–2592.

——, and Coauthors, 2006a: Raman lidar measurements during the International H_2O Project. Part I: Instrumentation and analysis techniques. *J. Atmos. Oceanic Technol.,* **23,** 157–169.

——, and Coauthors, 2006b: Measurements during the International H_2O Project. Part II: Case studies. *J. Atmos. Oceanic Technol.,* **23,** 170–183.

——, and Coauthors, 2006c: 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.

——, and Coauthors, 2006d: Water Vapor Validation Experiment—Satellite/sondes—Overview and preliminary results. Preprints, *Third Symp. on LIDAR Atmospheric Applications,* San Antonio, TX, Amer. Meteor. Soc., 4.4.

WMO, 2008: GCOS Reference Upper Air Network (GRUAN): Report of the GCOS Reference Upper-Air Network Implementation Meeting, Lindenberg, Germany, 26-28 February 2008. WMO Tech. Doc. WMO/TD 1435, GCOS-121, 49 pp. [Available online at http://www.wmo.int/pages/prog/gcos/Publications/gcos-121.pdf.]