Absolute accuracy of water vapor measurements from six operational radiosonde types launched during AWEX-G and implications for AIRS validation

Larry M. Miloshevich,[1] Holger Vömel,[2] David N. Whiteman,[3] Barry M. Lesht,[4] F. J. Schmidlin,[3] and Felicita Russo[5]

[1] A detailed assessment of radiosonde water vapor measurement accuracy throughout the tropospheric column is needed for assessing the impact of observational error on applications that use the radiosonde data as input, such as forecast modeling, radiative transfer calculations, remote sensor retrieval validation, climate trend studies, and development of climatologies and cloud and radiation parameterizations. Six operational radiosonde types were flown together in various combinations with a reference-quality hygrometer during the Atmospheric Infrared Sounder (AIRS) Water Vapor Experiment-Ground (AWEX-G), while simultaneous measurements were acquired from Raman lidar and microwave radiometers. This study determines the mean accuracy and variability of the radiosonde water vapor measurements relative to simultaneous measurements from the University of Colorado (CU) Cryogenic Frostpoint Hygrometer (CFH), a reference-quality standard of known absolute accuracy. The accuracy and performance characteristics of the following radiosonde types are evaluated: Vaisala RS80-H, RS90, and RS92; Sippican Mark IIa; Modem GL98; and the Meteolabor Snow White hygrometer. A validated correction for sensor time lag error is found to improve the accuracy and reduce the variability of upper tropospheric water vapor measurements from the Vaisala radiosondes. The AWEX data set is also used to derive and validate a new empirical correction that improves the mean calibration accuracy of Vaisala measurements by an amount that depends on the temperature, relative humidity, and sensor type. Fully corrected Vaisala radiosonde measurements are found to be suitably accurate for AIRS validation throughout the troposphere, whereas the other radiosonde types are suitably accurate under only a subset of tropospheric conditions. Although this study focuses on the accuracy of nighttime radiosonde measurements, comparison of Vaisala RS90 measurements to water vapor retrievals from a microwave radiometer reveals a 6–8% dry bias in daytime RS90 measurements that is caused by solar heating of the sensor. An AWEX-like data set of daytime measurements is highly desirable to complete the accuracy assessment, ideally from a tropical location where the full range of tropospheric temperatures can be sampled.

1. Introduction

[2] A great variety of atmospheric research is influenced, either directly or indirectly, by water vapor measurements from radiosondes. Radiosonde data from operational programs such as the U.S. National Weather Service (NWS) and similar programs worldwide are commonly assimilated into forecast models. Because of their high vertical resolution, radiosonde data are increasingly used to evaluate (validate) water vapor retrievals from ground-based and satellite remote sensors [e.g., *Soden et al.*, 2004; *Soden and Lanzante*, 1996]. Radiosonde water vapor measurements are also used in climate-related research, including studies of trends in upper troposphere (UT) water vapor, stratospheric dehydration and troposphere-stratosphere exchange processes, and initiation and maintenance of cirrus clouds, although it is questionable whether most radiosonde data are really accurate enough for these purposes. Since

[1]National Center for Atmospheric Research, Boulder, Colorado, USA.
[2]Cooperative Institute for Research in the Environmental Sciences, University of Colorado, Boulder, Colorado, USA.
[3]NASA/Goddard Space Flight Center, Greenbelt, Maryland, USA.
[4]Department of Energy/Argonne National Laboratory, Argonne, Illinois, USA.
[5]Department of Physics, University of Maryland-Baltimore County, Baltimore, Maryland, USA.

water vapor concentrations decrease by several orders of magnitude between the surface and the lower stratosphere (LS), it is particularly challenging to accurately measure relative humidity (RH) throughout the tropospheric column, especially with a low-cost operational radiosonde. It is important to realistically establish the accuracy of radiosonde water vapor measurements in order to evaluate the contribution of observational uncertainty to the uncertainty in forecast model results, remote sensor validations, radiative transfer calculations, cloud parameterizations, and other applications that use radiosonde data as input. An indication of radiosonde measurement accuracy is provided by radiosonde intercomparison experiments such as those conducted by the World Meteorological Organization [e.g., *Sapucci et al.*, 2005; *Yagi et al.*, 1996; *Schmidlin*, 1998; *Ivanov et al.*, 1991]; however, these studies compare operational radiosondes only to each other, which is not sufficient for estimating the absolute accuracy of the measurements. This study will quantify in detail the operational accuracy of radiosonde water vapor measurements throughout the troposphere, by comparing in situ radiosonde measurements to simultaneous measurements from a reference-quality research instrument of known absolute accuracy.

[3] This study was motivated by the need to establish the accuracy of radiance measurements and water vapor retrievals from the Atmospheric Infrared Sounder (AIRS) instrument onboard NASA's Aqua satellite. The general approach to AIRS water vapor validation involves comparing the AIRS-measured radiance spectrum to the spectrum calculated by a radiative transfer model that uses observed water vapor and temperature profiles as input [*Fetzer et al.*, 2003]. The goal for AIRS water vapor retrievals is 10% accuracy in 2 km layers throughout the troposphere, and thus the validation measurements must be *at least* this accurate. The AIRS Water Vapor Experiment-Ground (AWEX-G) was conducted in October–November 2003 at the Department of Energy Atmospheric Radiation Measurement (DOE/ARM) program's Southern Great Plains (SGP) research site. The primary purpose of AWEX-G was to evaluate the accuracy of several water vapor profiling instruments used in the AIRS validation effort [*Whiteman et al.*, 2006], including the ARM Raman lidar, the NASA Scanning Raman Lidar (SRL), and Vaisala RS90 radiosondes launched from the ARM SGP and Tropical Western Pacific (TWP) sites. During AWEX, six operational radiosonde types and a reference-quality water vapor instrument were launched in various combinations on the same balloon. The operational radiosondes included three Vaisala radiosonde types (RS80-H, RS90, and RS92); Sippican Mark IIa (formerly VIZ); Modem GL98; and the Meteolabor "Snow White" (SW) chilled mirror hygrometer. The radiosonde accuracy is evaluated relative to measurements by the University of Colorado (CU) Cryogenic Frostpoint Hygrometer (CFH), a fast-response, cryogenically cooled, chilled mirror hygrometer whose absolute measurement accuracy is quantified in section 2. Coincident water vapor measurements were also made by the SRL and by two ARM microwave radiometers. All balloon launches during AWEX were conducted at night, and therefore the influence of solar radiation on the radiosonde measurement accuracy cannot be assessed using the AWEX data set. However, in compensation, the absence of solar radiative effects permits

an assessment of the radiosonde calibration accuracy, which in turn permits the development of an empirical calibration correction for the Vaisala radiosonde types. A future AWEX-like experiment conducted during the daytime is needed for assessing solar radiative effects and possibly developing a means of correcting for them.

[4] The primary purpose of this study is to quantify the accuracy and its variability for the six operational radiosonde types launched during AWEX. A second purpose of this study is to evaluate the impact of physically based corrections for known sources of measurement error in Vaisala radiosonde measurements. The AWEX data set is also used to derive and validate a new empirical correction for inaccuracy in the Vaisala calibration, and the corrected measurements are found to be sufficiently accurate for AIRS validation throughout the troposphere.

[5] The accuracy and measurement characteristics of the CFH reference instrument are discussed in section 2, and the operational radiosondes are described in section 3. An overview of the qualitative performance of each radiosonde type is given in section 4, followed in section 5 by a detailed quantitative evaluation of radiosonde water vapor measurement accuracy and its variability, including the impact of time-lag and calibration corrections on the accuracy of Vaisala radiosonde measurements. The influence of solar radiation on the accuracy of Vaisala RS90 radiosonde measurements, and implications of this study for AIRS validation, are discussed in section 6. Evaluation of the NASA SRL measurement accuracy using the AWEX data set is given in a companion study by *Whiteman et al.* [2006].

2. CU CFH Reference Instrument

[6] The CFH is a new-and-improved version of the NOAA cryogenic hygrometer, which is the long-standing standard for balloon-borne stratospheric water vapor measurement [e.g., *Vömel et al.*, 1995]. Microprocessor control of the CFH electronics and optics allows accurate measurement of water vapor throughout the troposphere more reliably than its predecessor. The CFH water vapor measurement 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 (T_f). If the condensate is liquid water then the mirror temperature is equal to the dewpoint temperature of the air (T_d), although this paper will typically refer generically to the frostpoint temperature. The mirror temperature is measured by a tiny thermistor embedded in the surface of the mirror, and the accuracy of the thermistor calibration is <0.05°C throughout the temperature range +30 to −95°C. The uncertainty in the T_f measurement is dominated by the ability of the controller to maintain a constant condensate layer on the mirror, as the heating and cooling causes the mirror temperature to oscillate around the true frostpoint temperature. The total uncertainty in T_f from all sources is estimated to be ±0.3°C when 200 m (~40 s) averages are considered, as in this study. The uncertainty increases with decreasing averaging interval to an estimated ±0.5°C for 50 m (10 s) averages. A full description of the CFH design and measurement character-
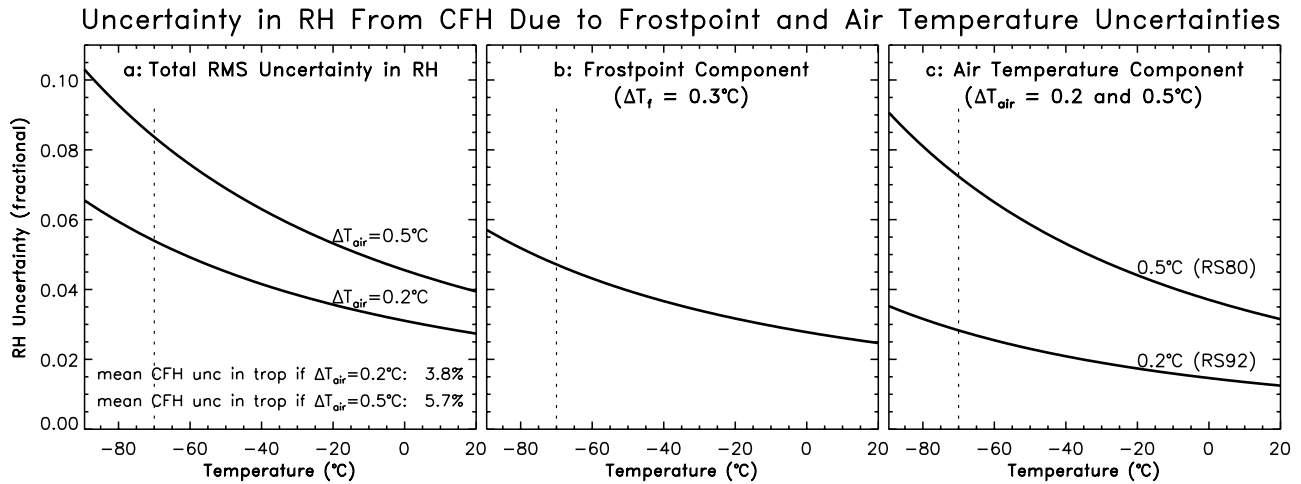
Figure 1. (a) Estimated fractional uncertainty in RH measurements from the CFH ($\Delta RH/RH$) as a function of temperature, which is given by the RMS sum of the independent uncertainties in (b) the CFH frostpoint measurement ($\Delta T_f = 0.3°C$) and (c) the radiosonde air temperature measurement ($\Delta T_{air} = 0.2$ or $0.5°C$). The mean RMS uncertainty in RH over the AWEX temperature range ($T > -70°C$, dashed line) is stated in Figure 1a for the two situations where the air temperature measurements are from either an RS92 or RS80 radiosonde.

istics is given by H. Vömel et al. (The University of Colorado Cryogenic Frostpoint Hygrometer (CFH), submitted to *Journal of Geophysical Research*, 2006).

[7] The RH with respect to liquid water is calculated from T_f and from the ambient air temperature (T) that is measured by an attached Vaisala RS80-H or RS92 radiosonde, according to $RH = e_i(T_f)/e_w(T) \times 100\%$, where e_i is the saturation vapor pressure (SVP) over ice as given by *Hyland and Wexler* [1983], and e_w is the SVP over liquid water as given by *Wexler* [1976]. If the condensate is liquid water, then $RH = e_w(T_d)/e_w(T) \times 100\%$. These particular SVP formulations are chosen for consistency with the Vaisala calibration procedure [*Miloshevich et al.*, 2001, hereinafter referred to as M01], and the formulas are summarized by *Miloshevich et al.* [2004, hereinafter referred to as M04]. A comparison of SVP formulations and discussion of the uncertainty in RH attributable to the choice of SVP formulation is given in Appendix A. Although the *Goff and Gratch* [1946] and *Goff* [1965] formulations for e_w are still in widespread use, they are inaccurate at low temperatures, as illustrated by the odd behavior at low temperatures in Figure A1c. It should be noted that some in the remote-sensing world define RH as the ratio of the water vapor mixing ratio (r) to the saturation mixing ratio (r_w); however, we adopt the convention in the radiosonde world that defines RH as the ratio of the vapor pressure to the saturation vapor pressure over liquid water.

[8] Radiosonde accuracy estimates reported in this paper are relative to the absolute accuracy of the CFH. In addition to the ±0.3°C uncertainty in T_f, additional uncertainty from the air temperature measurement is introduced when the frostpoint measurement is converted to RH for comparison to radiosonde measurements. In 9 of the 12 CFH soundings used in this study, the temperature measurement is from an attached RS92 radiosonde, whose uncertainty is given by Vaisala as ±0.2°C at the 2-sigma (95.5% confidence) level throughout the troposphere for nighttime measurements where solar radiation is not a factor [*Paukkunen et al.*,

2001], consistent with the detailed uncertainty analysis given by *Luers* [1997]. Three of the CFH soundings use the temperature measurement from an attached RS80-H radiosonde, whose nighttime uncertainty at the 2-sigma level, including time lag error which is not an issue for the smaller and faster RS92 temperature sensor, is estimated from Vaisala literature to be ±0.5°C, consistent with the uncertainty analysis of *Luers and Eskridge* [1995]. Figure 1 shows the fractional (percentage) uncertainty in calculated RH values ($\Delta RH/RH$), where the total uncertainty (Figure 1a) is the RMS sum of the contributions from the CFH frostpoint measurement (Figure 1b) and the radiosonde air temperature measurement (Figure 1c). The mean percentage uncertainty in the CFH RH measurements over the AWEX temperature range ($T > -70°C$) is about 4% when RS92 temperature measurements are used and about 6% when RS80-H temperature measurements are used.

[9] Two observations from the AWEX data set provide empirical estimates of the CFH measurement accuracy. The RH measured by the CFH during 6 instances when the balloon penetrated a liquid water (LW) cloud ranged from 99.9% RH to 103.0% RH, with a mean and standard deviation of 101.5±1.2% RH (Table 1). The RH measured by the Snow White (SW) chilled mirror hygrometer for these cases was nearly identical, indicating that the CFH and SW may contain a small 1−2% moist bias. Furthermore, the assumed ambient RH of 100% in LW clouds may in fact be slightly higher (i.e., closer to the mean CFH and SW measurements) if there is an updraft. Calculations from a detailed microphysical parcel model [e.g., *Heymsfield and Miloshevich*, 1993], when initialized with an updraft speed of 1 m s^{-1} and a droplet concentration of 100 cm^{-3} at +10°C, show a peak supersaturation in the lower part of the cloud of 1.0%, and an equilibrium supersaturation in the upper portion of the cloud of 0.2%. The observations in LW clouds support the conclusion that the CFH (and SW) mean absolute accuracy is 1−2%, at least under moist conditions in the lower to mid troposphere.

Table 1. Mean RH Measurements From CFH and SW in Layers That Contain Liquid Water (LW), as Identified by the Characteristic Increase in RH Below Cloud Base, Followed by Nearly Constant In-Cloud Measurements Near 100% RH, Followed by a Decrease in RH Above Cloud Top[a]

Date	T, °C	RH_{CFH}, %	RH_{SW}, %
29 Oct.	−12.8	102.5	103.0
6 Nov.	−2.1	103.0	100.8
6 Nov.	−7.9	101.5	101.6
10 Nov.	+4.8	100.5	100.7
10 Nov.	+5.7	101.4	100.1
11 Nov.	+13.6	99.9	100.5
Mean		101.5	101.1
Std dev		1.2	1.0

[a]Only the stable in-cloud measurements away from the cloud boundaries are averaged. The thickness of the LW layers was in the range 200–500 m, and the mean temperature in each layer is given. The mean and standard deviation of all 6 LW cases are shown for each instrument.

[10] A second empirical estimate of the CFH accuracy during AWEX is obtained by comparing the column-integrated precipitable water vapor (PWV) from the CFH with that retrieved from ARM microwave radiometer (MWR) measurements. The RMS accuracy of the MWR retrievals, including both instrumental and retrieval uncertainties, is typically about 0.4 mm in PWV, or a nominal accuracy of 2–3% for most conditions [*Liljegren et al.*, 2001, Appendix]. This accuracy will be poorer for very low PWV conditions and better for very high PWV conditions, and will vary with the season. A known moist bias of 3% in ARM MWR PWV retrievals after April 2002 (including the AWEX and AIRS timeframes) has been attributed to inaccuracy in the half-width of the 22 GHz water vapor line used in the Rosenkranz absorption model [*Liljegren et al.*, 2005]. To account for this bias, 3% has been subtracted from all ARM MWR PWV retrievals used in this study, which corresponds to using the smaller line half-width from the HITRAN compilation used by ARM prior to April 2002.

[11] The mean and standard deviation of the ratio PWV_{CFH}/PWV_{MWR} for the 10 AWEX cases where the CFH reached at least 10 km altitude is 1.03 ± 0.03 (Figure 2), suggesting a mean CFH moist bias of 3%, although 3% is within the nominal uncertainty in the MWR retrievals. The SW results were nearly identical, and further supports the conclusion that the absolute accuracy of the CFH (and SW) RH measurements is <3% in the lower troposphere where most of the PWV resides. Both the CFH measurements in LW clouds and the MWR PWV comparisons point to a very small moist bias in the CFH (and SW) RH measurements, but certainly these observations give confidence that the absolute accuracy of the CFH (and SW) RH measurements is at least as accurate as estimated from the frostpoint and air temperature uncertainties shown in Figure 1 (3% near the surface to 6% at the midlatitude tropopause). Since independent frostpoint measurements from both the CFH and SW indicate a similar 1–3% moist bias when converted to RH, the likely cause is a small bias in the air temperature measurements that are used in both cases to calculate the RH.

[12] Finally, an additional source of uncertainty in chilled mirror hygrometer measurements is ambiguity regarding the phase of condensate on the mirror. If the mirror and air temperatures are both above 0°C, then the condensate is surely liquid and the dewpoint temperature rather than the frostpoint temperature is measured. However, at mirror temperatures below 0°C but above the homogeneous nucleation temperature of about −35°C, the condensate may be either ice or supercooled liquid water. Detailed inspec-

Comparison of PWV Measurements from CFH and SW with MWR
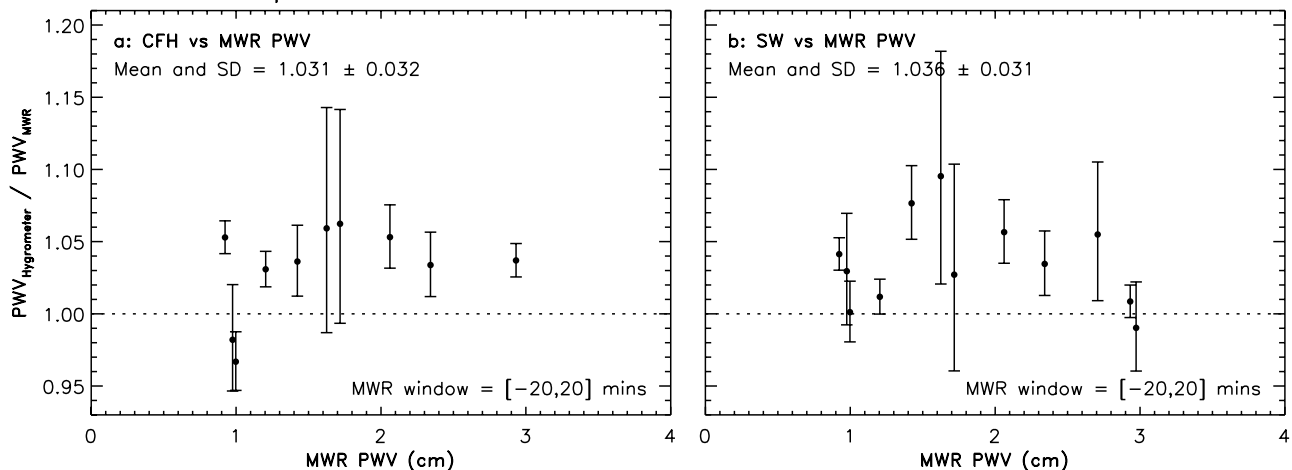

Figure 2. Ratio of the column-integrated precipitable water vapor (PWV) measured by either (a) the CFH or (b) the SW with the PWV measured simultaneously by an ARM microwave radiometer (MWR) for the 10 CFH and 12 SW profiles during AWEX that attained an altitude of at least 10 km. The MWR PWV value used to calculate the ratio is a 40 min average centered on the balloon launch time. Vertical bars represent the standard deviation of the ratio values calculated from the 75–100 individual MWR PWV values measured during the averaging window and thus represent the range of ratio values attributable to the combination of MWR measurement and retrieval uncertainty, plus actual change in the PWV during the averaging window. The annotation in each panel gives the mean of the ratio values for all profiles (dots) and the standard deviation of those values.

tion of many profiles measured by the NOAA hygrometer and the SW has shown that, although the temperature of the phase transition varies considerably, it is typically near $-20°C$. To address this ambiguity, the CFH is programmed to rapidly cool the mirror to $-40°C$ when the mirror temperature first reaches $-10°C$, ensuring that the frost-point temperature is measured thereafter, and leaving a "time stamp" in the data to identify the event. This forced freezing largely eliminates the phase ambiguity, as the condensate prior to the freezing event is almost certainly liquid in all cases.

3. Radiosonde RH Measurement Principles

[13] This section describes the RH measurement technology and known performance characteristics of the radiosonde types flown during AWEX. Manufacturer's accuracy specifications are not quoted in this paper, because these values often refer to only a limited set of conditions where the accuracy is best (e.g., static conditions at $+20°C$).

3.1. Vaisala RS80-H, RS90, and RS92

[14] Vaisala radiosondes use thin-film capacitance RH sensors, where a thin hydrophilic polymer layer on a glass substrate acts as the dielectric of a capacitor. The capacitance measured by the radiosonde is proportional to the number of water molecules captured at binding sites in the polymer structure, which in turn is proportional to the ambient water vapor concentration. The factory calibration procedure relates the measured capacitance to the RH with respect to liquid water at a standard temperature of $+20°C$. The RH is then adjusted for the measured ambient temperature on the basis of a "temperature-dependence" (TD) calibration model, which consists of RH- and T-dependent curve fits derived from the response of many sensors tested in the Vaisala calibration chamber. One source of "production variability" in the accuracy of an individual sensor is the deviation of the sensor's response from the mean response characteristics represented by the TD calibration model. The operational principles of Vaisala radiosonde RH sensors are described further by M01 and M04 and references therein.

[15] There are 4 types of Vaisala radiosondes: RS80-A, RS80-H, RS90, and RS92. All except the RS80-A were flown during AWEX. The two RS80 radiosonde types use the same temperature sensor, but the RH sensors use different polymers (A versus H). The H-type polymer is more stable against hysteresis than the A-type polymer, particularly at high RH, but the H-type polymer responds more slowly at low temperatures, and therefore the RS80-H has greater time lag error. The RS80-H, currently used at 2/3 of NWS radiosonde sites, is calibrated much more accurately at low temperatures than the RS80-A. The newer RS90 and the newest RS92 radiosondes both use the H-type polymer, but the polymer layer is smaller and thinner than the RS80-H, giving these sensors much faster time response at low temperatures. The RS90 and RS92 sensors and calibration procedures are essentially (but not quite) identical, and the nomenclature RS9x will be used in this paper to refer to either sensor when appropriate. The Vaisala calibration facility for RS9x sensors is more accurate than the RS80 calibration facility, and a detailed assessment of the RS9x calibration uncertainty is given by *Paukkunen et al.*

[2001]. The RS9x sensors also differ from the RS80 sensors in that they are actually dual sensors that are alternately heated while the other sensor makes the measurements, thereby eliminating susceptibility of the sensors to icing in clouds or in ice-supersaturated conditions.

[16] Several known sources of measurement error in Vaisala RH measurements, and corrections for some of them, are discussed by M04, M01, and *Wang et al.* [2002]. Although all Vaisala radiosondes are subject to the same general sources of RH measurement error, the magnitude of the error depends critically on the specific radiosonde type, so documentation of the radiosonde type and serial number by users is strongly recommended. The accuracy of the calibration is in part given by the absolute accuracy of the calibration references. Error in the calibration reference leads to bias error in the measurements that may vary with time as the calibration reference drifts, contributing to the "batch-dependent" bias error in RS80-H radiosondes reported by *Turner et al.* [2003]. The TD calibration model is also affected by "curve fit error," which is bias error that results from the inadequacy of curve fits in accurately representing the mean sensor response over the entire range of RH and T. As interest in upper tropospheric water vapor measurements has increased, shortcomings in the TD portion of the calibration at low temperatures became apparent as a dry bias in RS80 RH measurements, and led to efforts by Vaisala to improve the calibration at low temperatures. The RS80 TD calibration model was not changed by Vaisala, but TD corrections that can be applied to processed RS80 data were developed. The magnitude of the TD correction as a percentage of the measured RH for RS80-H radiosondes is 4% at $-40°C$, 13% at $-60°C$, and 32% at $-80°C$ [*Wang et al.*, 2002; M04]. The TD correction for RS80-A radiosondes is considerably larger: 10% at $-35°C$, 40% at $-50°C$, 80% at $-60°C$, 150% at $-70°C$, and 250% at $-80°C$ [M01; M04]. The RS9x calibration procedure is more accurate than the RS80-H calibration procedure [*Paukkunen et al.*, 2001], so no TD correction has been developed. However, RS90 radiosondes produced before 25 June 2001 used an early and inaccurate TD calibration model, but these early RS90 measurements can be adjusted to the current and more accurate TD model as described in Appendix B.

[17] In addition to calibration-related sources of measurement error, all sensors are subject to "time lag" (TL) error because of the finite response time to changes in the ambient humidity. The response time is characterized by the sensor time constant (the time required to respond to 63% of a step change in RH), which Vaisala measured for each sensor type over the temperature range $+20$ to $-60°C$. These time constant measurements were the basis of a time lag correction algorithm developed and validated by M04. Time lag error is not a bias error, but rather "smooths" the profile by an amount that depends on the temperature and the local humidity gradient, such that the time lag correction changes the shape of the RH profile in the upper troposphere (and in the middle troposphere for the slower RS80-H sensors). The greatest impact of time lag error is above and below cirrus layers and at the tropopause, where humidity gradients may be steep at low temperatures.

[18] Much has been written about the so-called "dry-bias" or "contamination" error in Vaisala RH measure-

ments, where nonwater molecules outgassed from styrofoam in the radiosonde packaging occupy binding sites in the sensor polymer and render these sites unavailable to water molecules [*Wang et al.*, 2002; M04]. Since June 2000, RS80 radiosondes are shipped with a sealed sensor cap that protects the polymer from exposure to contaminants, and the present study will show that the sensor cap largely solved the contamination problem. A statistical correction for contamination in RS80 radiosondes produced before June 2000 was developed by *Wang et al.* [2002] and modified by M04; however, M04 showed that considerable residual uncertainty remains in the data because of large variability in the contamination process between individual radiosondes and between calibration batches. Contamination is undoubtedly responsible for much of the dry bias in ARM RS80-H radiosondes reported by *Turner et al.* [2003] and *Soden et al.* [2004]. The RS90 does not use the sealed sensor cap, but the radiosonde packaging is cardboard rather than styrofoam, so the source of contaminants has been greatly reduced. The RS92 uses "regeneration," where the sensor is heated during the radiosonde launch preparations to drive off any contaminants and recover the original calibration accuracy [*Hirvensalo et al.*, 2002].

[19] Apart from the physically based corrections discussed above, other techniques have been developed to address bias error in Vaisala radiosonde measurements. Scaling of the measurements by a constant factor that matches the column-integrated PWV from the radiosonde with simultaneous PWV measurements from a microwave radiometer was found to improve the accuracy of lower-tropospheric radiosonde measurements [*Turner et al.*, 2003], as judged by decreased residuals and variability when radiosonde profiles are used as input to a radiative transfer model and results are then compared to downwelling longwave radiance measurements from the ARM Atmospheric Emitted Radiance Interferometer (AERI). A correction method based on assimilation of satellite radiance measurements [*Soden et al.*, 2004] was shown to substantially reduce the radiosonde dry bias in the upper troposphere relative to the ARM Raman lidar. However, both of these studies investigated older RS80-H radiosondes that were subject to contamination. The main disadvantages of these methods relative to correction methods based on principles of sensor behavior is that they rely on a second set of coincident measurements, and they are not necessarily effective throughout the troposphere.

[20] Finally, a diurnal bias of 3–4% was observed in ARM RS80-H measurements by *Turner et al.* [2003], which almost certainly results from solar heating of the RH sensor. Solar heating of the sensor causes the measured air that is in direct contact with the sensor polymer to be warmer than the ambient air, and therefore the measured RH of the warmer air is lower than the ambient RH, producing a consistent dry bias in daytime Vaisala (and probably other) radiosonde RH measurements. Since all soundings during AWEX occurred at night, solar radiation error is not a factor in the accuracy assessment given in this paper. The impact of solar radiation error on RS90 measurements used for AIRS validation is presented in section 6.

3.2. Meteolabor Snow White Hygrometer

[21] The Meteolabor Snow White (SW) hygrometer is an operational radiosonde in Switzerland where it is manufactured, and is often flown as a research instrument elsewhere. It operates on the same chilled mirror principle as the CFH, but the mirror is cooled electrically with a Peltier device rather than a cryogenic fluid like the CFH, and therefore its maximum cooling capacity and time response are less than the CFH. The maximum frostpoint depression attainable by a Peltier cooler under operational conditions is about 25°C, producing an inherent lower bound on the RH that is measurable by the SW. Comparisons between SW and the NOAA cryogenic hygrometer [*Vömel et al.*, 2003] revealed several limitations of the SW: (1) The lower RH detection limit, while dependent on temperature, is nominally about 6% RH; (2) extended dry layers below the RH detection limit can cause the SW to lose frost coverage on the mirror, which is sometimes regained above the dry layer but sometimes is not; and (3) mixed results are obtained in the UT, where sometimes the measurement accuracy is comparable to the NOAA hygrometer, and sometimes a bias (either dry or moist) is observed, which may be caused by instability in the controller at low temperatures. Apart from these limitations, the nominal SW measurement accuracy is comparable to the CFH accuracy that was estimated in Figure 1 and from the AWEX observations described in section 2, with two important exceptions.

[22] The sensor design and sampling inlet of the SW differs substantially from the CFH, and may lead to a new source of measurement error. Whereas the CFH sampling inlet is a metal tube with "straight through" air flow, the SW inlet is curved, composed of styrofoam, and the sensor casing restricts air flow. During transit through an ice cloud, ice particles can become trapped in the inlet and sensor casing, and then sublimate and artificially enhance the measured water vapor both within and above the cloud, as shown later.

[23] The SW has no mechanism analogous to the programmed force-freezing used by the CFH to address ambiguity in the phase of the condensate when the mirror temperature is below 0°C. For the AWEX data set, the phase transition from supercooled liquid water to frost was identified in the SW data by eye, from the sudden jump in RH as compared to the CFH and radiosonde measurements, and the RH is then calculated for the proper phase using the same SVP formulations as with the CFH. However, normally the phase of condensate on the mirror will be uncertain, and consequently there will be SW measurement error resulting from whatever assumption is made about the phase of the condensate. For example, if the data processing assumes that the frostpoint is measured whenever the mirror temperature (T_m) is below 0°C, then the percentage error in RH if in fact the dewpoint is really being measured is given by $(e_w(T_m) - e_i(T_m))/e_i(T_m) \times 100\%$. This "phase ambiguity error" is 10.3% if $T_m = -10°C$; 21.7% if $T_m = -20°C$; and 34.3% if $T_m = -30°C$. Although this source of error is eliminated in the analysis shown in this paper, under normal operation it can be a substantial source of SW measurement error in the lower to mid troposphere.

3.3. Modem GL98

[24] Modem radiosondes are produced in France and are currently used at French overseas sites. A U.S. radiosonde and data system manufacturer, Intermet, supplied Modem GL98 radiosondes and a Modem data system for use during

AWEX. Although Intermet and Modem share a historical relationship, Intermet has produced its own radiosonde that is currently undergoing qualifying tests for consideration by the NWS in its Radiosonde Replacement System, and the sensors and calibration procedure for the new Intermet radiosonde (iMet-1) are different from the Modem GL98. This study investigates only the Modem GL98. Modem RH sensors, like Vaisala RH sensors, measure the capacitance of a hydrophilic thin-film polymer layer, which is calibrated to measure RH. Corrections analogous to those developed for Vaisala radiosondes are not available for Modem radiosondes because there are no published studies on the calibration accuracy, sensor time response, or susceptibility to chemical contamination.

3.4. Sippican Mark IIa (Formerly VIZ)

[25] Relative humidity measurements from Sippican radiosondes, currently in operational use at about 1/3 of U.S. NWS sites, are based on an entirely different measurement technology: the carbon hygristor. The resistance of a carbon filament in the sensor varies with the ambient water vapor concentration and temperature, and the resistance is calibrated to measure RH. Manufacturer's specifications note that the measurement range of Sippican RH sensors is $5-100\%$ RH at temperatures above $-50°C$, thus these sensors are not capable of measuring all tropospheric conditions. Environmental chamber tests conducted by *Blackmore and Taubvurtzel* [1999] showed that the time response of VIZ carbon hygristors was extremely slow at $-40°C$, and the sensors showed little indication of functioning at all at $-60°C$. Similarly, *Ferrare et al.* [2004] found that Sippican water vapor measurements exhibited poor agreement with other water vapor measurements in the midlatitude UT.

4. Overview of Radiosonde Performance

[26] The general measurement characteristics and limitations of each radiosonde type are illustrated by the example AWEX soundings in Figure 3, relative to CFH (purple) as the reference standard. A detailed quantitative accuracy assessment is presented in section 5. The following observations from Figure 3 are supported by the AWEX data set as a whole, and are taken as general measurement characteristics of each sensor type.

[27] Only the Vaisala RS90 (red) and RS92 (green) show good qualitative agreement with CFH for all conditions in the midlatitude troposphere (Figures 3a–3c and 3e). The RS9x sensors are capable of reliable measurements after passage through a thick ice cloud (Figure 3b), because of their alternately heated dual-sensor design. The RS90 and RS92 are essentially equivalent sensors in terms of calibration accuracy and time response (Figures 3a, 3b, and 3d; A. Paukkunen, personal communication, 2004); however, only the RS92 performs "regeneration" during the launch procedure to drive off chemical contaminants. Three colaunches of RS90 and RS92 during AWEX produced profiles that were nearly identical, supporting the equivalence of the RS90 and RS92 measurements.

[28] The Vaisala RS80-H (black) shows generally good agreement with CFH (within $2-3\%$ RH) for temperatures above about $-35°C$ (Figure 3a, 3c, and 3e), but slow sensor time response and a less accurate calibration at lower temperatures prevents the RS80-H from accurately capturing vertical structure in the RH profile in the UT, particularly when the humidity gradient is steep (Figures 3c and 3e, below $-40°C$). Applying existing corrections for time lag and TD calibration errors (Figures 3c and 3e, dashed black) recovers the vertical structure that was "smoothed" by slow sensor response, but also reveals an apparent moist bias in the calibration at low temperatures. Unlike the dual heated RS9x sensors, the RS80-H is subject to "sensor icing" (i.e., ice deposition on the sensor in an ice-supersaturated environment), which can lead to suspiciously elevated RH measurements well into the stratosphere (Figure 3b).

[29] The Snow White measurements (yellow) are in good agreement with the CFH (Figures 3a, 3c, and 3e), except under very dry conditions (Figure 3a, at $3-9$ km), or within and above thick ice clouds (Figure 3b). Auxiliary temperature measurements from a bead thermistor were made during several SW flights at a position just upstream of the sensor in the inlet duct, which confirmed that heat from the mirror heater under conditions of ice supersaturation raises the temperature of the sampled air by $2-3°C$ above ambient, with the consequence that cloud particles are partially sublimated and RH measurements within the cloud are artificially elevated (Figure 3b, $6-11$ km). Erroneously high RH measurements above the cloud layer (Figure 3b, above tropopause) are thought to result from sublimation of ice particles trapped in the inlet duct and/or sensor casing; however, alternative explanations for anomalously high RH measurements by the SW above ice clouds in the UT include (1) sublimation of ice deposited onto the inlet duct or sensor casing in ice-supersaturated conditions, (2) outgassing of water vapor absorbed into the styrofoam of the radiosonde at lower levels, (3) insufficient remaining battery power to cool the mirror under dry conditions, or (4) instability in the controller circuitry at low temperatures as suggested by *Vömel et al.* [2003].

[30] The Modem GL98 measurements (dark blue) often show a moist bias compared to the other sensors (Figures 3a, 3d, and 3f). Time lag error in the UT is substantial (Figure 3d), and is similar in magnitude to the RS80-H time lag error (Figure 3c). The similar technology of the Modem and RS80-H sensors is also evident in the susceptibility of the Modem to sensor icing effects that persist into the stratosphere (Figure 3b). Very often the Modem measures 0% RH under dry conditions in the UT (Figures 3a and 3f), possibly indicating an inaccurate calibration at those temperatures, or possibly indicating susceptibility of the sensor to chemical contamination dry bias.

[31] The Sippican measurements (light blue; Figures 3a and 3f) exhibit slow sensor time response at low temperatures. The Sippican measurements in Figure 3f adequately reflect the vertical structure down to about $-25°C$, after which the response slows abruptly until the sensor becomes completely unresponsive at about $-45°C$. Analysis of Sippican soundings from the NWS Key West site in Florida during the NASA CRYSTAL-FACE experiment (L. Miloshevich, unpublished data, 2003) shows that the temperature at which Sippican sensors become unresponsive varies within the range -20 to $-50°C$, and therefore Sippican measurements cannot be considered reliable above the midtroposphere. The flat response in Figure 3a at
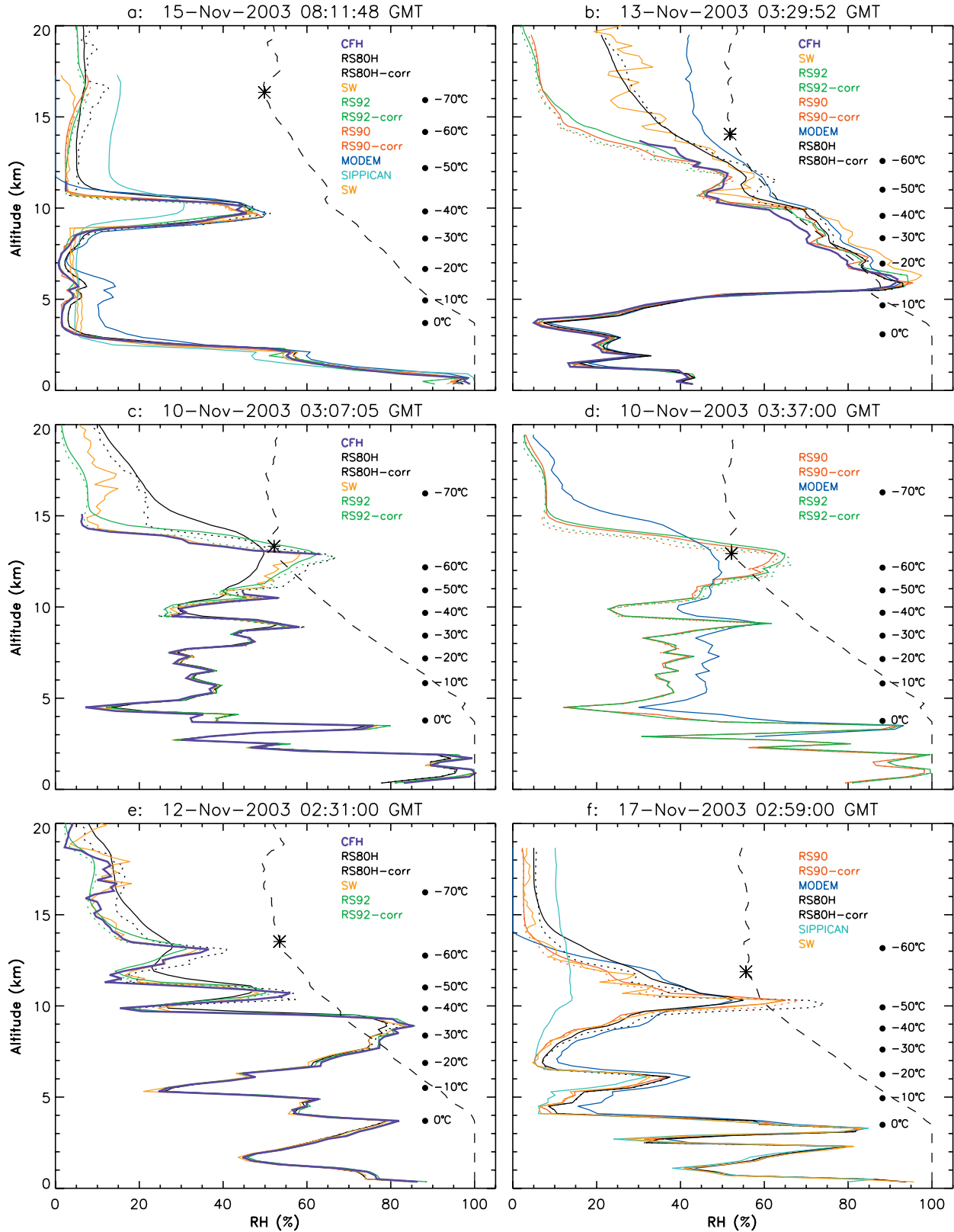
Figure 3. Example RH profiles from AWEX. (a and b) Profiles measured by sensors launched on separate balloons at approximately the same time. (c–f) Profiles measured by sensors on the same balloon. Dashed profiles are after applying a time lag correction to Vaisala measurements and, for RS80-H radiosondes, a TD calibration correction. Long dashes indicate ice saturation, and asterisks indicate the tropopause altitude.

3–9 km, similar in appearance to the SW response, indicates either an unreliable calibration for very dry conditions, or an inherent inability of the carbon hygristor to measure very dry conditions.

5. Radiosonde Accuracy Assessment

[32] Two categories of multiple-radiosonde launches occurred during AWEX: (1) radiosondes that were launched on the same balloon with CFH, such that the radiosonde measurements (nominally RS80-H, RS92, and SW) can be directly compared to the CFH measurements and (2) radiosondes that were launched on the same balloon with each other but not with CFH (nominally Modem, Sippican, RS80-H, and RS90), such that comparison to the CFH reference standard requires an indirect means. The analysis technique used in this study is described below, followed by an assessment of radiosonde accuracy from the direct comparisons to CFH, and then by an accuracy assessment for the remaining radiosonde types that is tied indirectly to the CFH using the RS90 as a proxy reference sensor and transfer standard. The impact of corrections applied to the Vaisala RH measurements is also assessed, and a new empirical calibration correction for Vaisala RH measurements is developed and validated.

5.1. Analysis Technique

[33] Several steps were taken in this study to maximize the accuracy of the radiosonde intercomparisons. Any two balloons, even if launched from the same location at the same time, will rise at different rates and sample different air, introducing differences between radiosonde measurements that are attributable to spatial and temporal variability in the water vapor field, in addition to differences attributable to sensor measurement characteristics, calibration accuracy, and production variability. Careful comparison of the RS92 measurements in Figures 3c and 3d (green), which were launched from the same location on different balloons separated in time by 30 min, shows substantial change in the atmosphere between 2 km altitude and the tropopause, such that comparison of the RS92 profile in Figure 3d with the CFH reference profile on a different balloon in Figure 3c would not yield a good assessment of the RS92 measurement capability. In this study, we only compare measurements acquired on the same balloon. Furthermore, we use the independent variable "time from launch" rather than the derived variable "altitude" for matching profiles that were measured by different sensors on the same balloon, thereby eliminating the occasionally large profile mismatch that results from differences in the accuracy of the pressure and temperature measurements used to derive altitude.

[34] Point-by-point comparisons between two profiles are not made for two reasons. First, the sampling interval used by different radiosondes varies from 1 s to 10 s (nominally 5 m to 50 m altitude intervals); and second, the different sensor types have different measurement characteristics as a result of fundamentally different measurement principles and degrees of postprocessing. The data systems for the Vaisala, Modem, and Sippican radiosondes provide filtered and otherwise postprocessed RH data that are smoother than the raw data, whereas the CFH and SW data are raw data characterized by oscillations around the true frostpoint

temperature. These factors are mitigated by first smoothing the CFH and SW data with an 11-point (13 s) boxcar average, which is comparable to the 11-point median filter applied during the standard processing of raw Vaisala RH data. Comparison profiles are then aligned using the variable "time," which is accomplished by adopting the Z(time) relationship from the designated reference sensor (nominally CFH), then redefining altitude Z for the other "test sensors" on the same balloon by interpolating their altitude variable to their time variable using the adopted Z(time) relationship. Finally, 200 m averages are computed for all profiles before conducting the comparisons.

[35] Most applications ultimately require an absolute water vapor quantity such as mixing ratio rather than RH, in which case it is more meaningful to express radiosonde measurement accuracy as a percentage of the measured water vapor rather than as a %RH value. Most of the sensor comparisons and accuracies reported in this paper are given as a percentage difference rather than a %RH difference, meaning in essence that we are investigating the absolute accuracy of radiosonde water vapor measurements rather than the RH measurement accuracy. As an example of the significance of this distinction, consider the specified 10% accuracy goal for validation of AIRS water vapor retrievals. Even an abysmal radiosonde accuracy of 10% RH meets this accuracy goal under conditions of water saturation, but under dry conditions of RH <10% the radiosonde accuracy must be <1% RH to meet the accuracy goal, which is quite a stringent requirement given that 1% RH is the (rather coarse) resolution at which some radiosonde measurements are reported. Thus radiosondes are inherently less accurate at measuring the absolute water vapor amount as the RH decreases.

5.2. Vaisala RS80-H, RS90, and RS92

[36] The Vaisala calibration procedure involves curve fits that are functions of RH and T, and therefore calibration-related uncertainty in Vaisala radiosonde measurements is also expected to depend on RH and T. The mean accuracy of the radiosonde measurements relative to the CFH is given by the mean of differences between simultaneous radiosonde and CFH measurements, and the variability in the radiosonde accuracy is taken to be the standard deviation of differences from the CFH (i.e., the RMS accuracy). The mean accuracy and its variability for RS80-H and RS92 radiosondes are shown in Figures 4 and 5 as successive corrections are applied, in three different ways. The top rows in Figures 4 and 5 show the radiosonde mean accuracy and variability in absolute terms (i.e., as a percentage of the CFH-measured RH, or $(RH_{sonde} - RH_{CFH})/RH_{CFH} \times 100\%)$. The middle rows in Figures 4 and 5 show the same mean accuracy and variability in units of RH (i.e., $RH_{sonde} - RH_{CFH}$). In each panel, the mean accuracy (top three curves) and the variability (bottom three curves) are shown as a function of RH for three temperature intervals that correspond roughly to the lower troposphere (LT), middle troposphere (MT), and upper troposphere (UT). The bottom rows in Figures 4 and 5 show the percentage (absolute) difference from CFH as a function of altitude, for both the individual profiles and the mean of all profiles. The left columns in Figures 4 and 5 show the mean accuracy and variability relative to the CFH for standard Vaisala data before any corrections are applied; the middle columns in

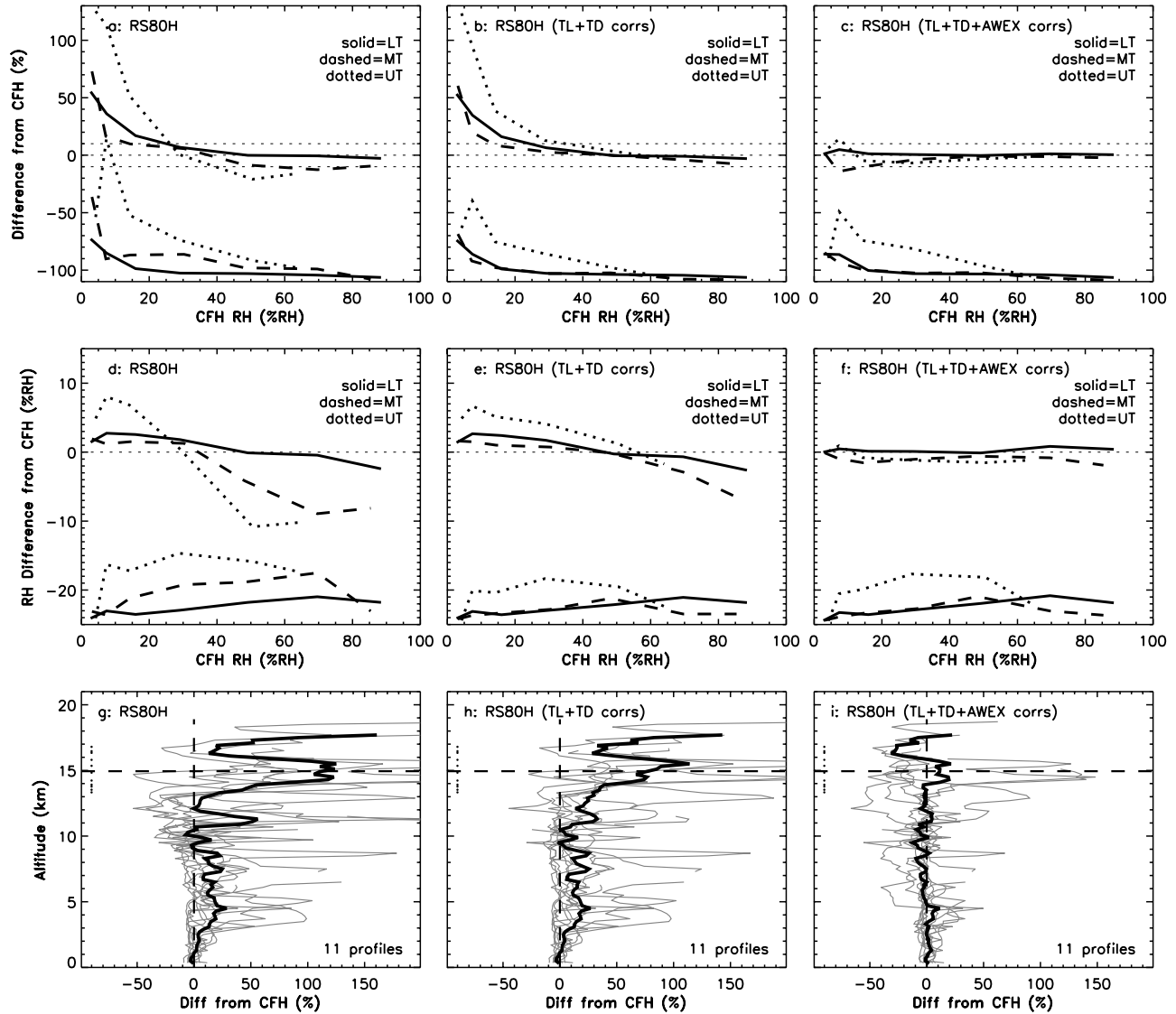## RS80-H Mean Accuracy and Variability Relative to CFH

Figure 4. (a–f) Mean and standard deviation of differences between Vaisala RS80-H and simultaneous CFH measurements as successive corrections are applied. Figures 4a–4c show absolute (percentage) differences (i.e., $(RH_{sonde} - RH_{CFH})/RH_{CFH} \times 100\%$), and Figures 4d–4f show RH differences ($RH_{sonde} - RH_{CFH}$). The top three curves in each panel show the mean difference from CFH as a function of RH, partitioned into three temperature intervals that correspond roughly to the lower troposphere (LT, T > $-20°$C, solid), the middle troposphere (MT, $-20 >$ T $> -50°$C, dashed), and the upper troposphere (UT, T $< -50°$C, dotted). The bottom three curves in each panel show the standard deviation of the differences from CFH, offset for clarity such that zero is at the bottom of the panel. The left column shows the difference from CFH for standard (uncorrected) RS80-H measurements, the middle column is after applying the time lag (TL) and temperature dependence (TD) corrections, and the right column is after also applying the empirical calibration correction that is derived from the AWEX data set as explained in section 5.2.1. Horizontal dashed lines in the top row indicate the AIRS ±10% goal for retrieval accuracy. The analysis uses 200 m averages from the surface to 3 km above the tropopause. (g–i) Altitude profiles of the percentage difference from CFH as successive corrections are applied. The thin curves are for the individual profiles, and the bold curves show the mean. The tropopause altitude for each profile is indicated by a dot at left, and the horizontal dashed line is the mean tropopause altitude.

Figures 4 and 5 are after correcting for sensor time lag (TL) error and for the RS80-H temperature-dependence (TD) calibration error; and the right columns in Figures 4 and 5 are after also applying an empirical calibration correction that is derived from the AWEX data set and will be discussed later in this section.

[37] Figures 4a or 4d show that standard RS80-H measurements exhibit a mean moist bias under dry conditions

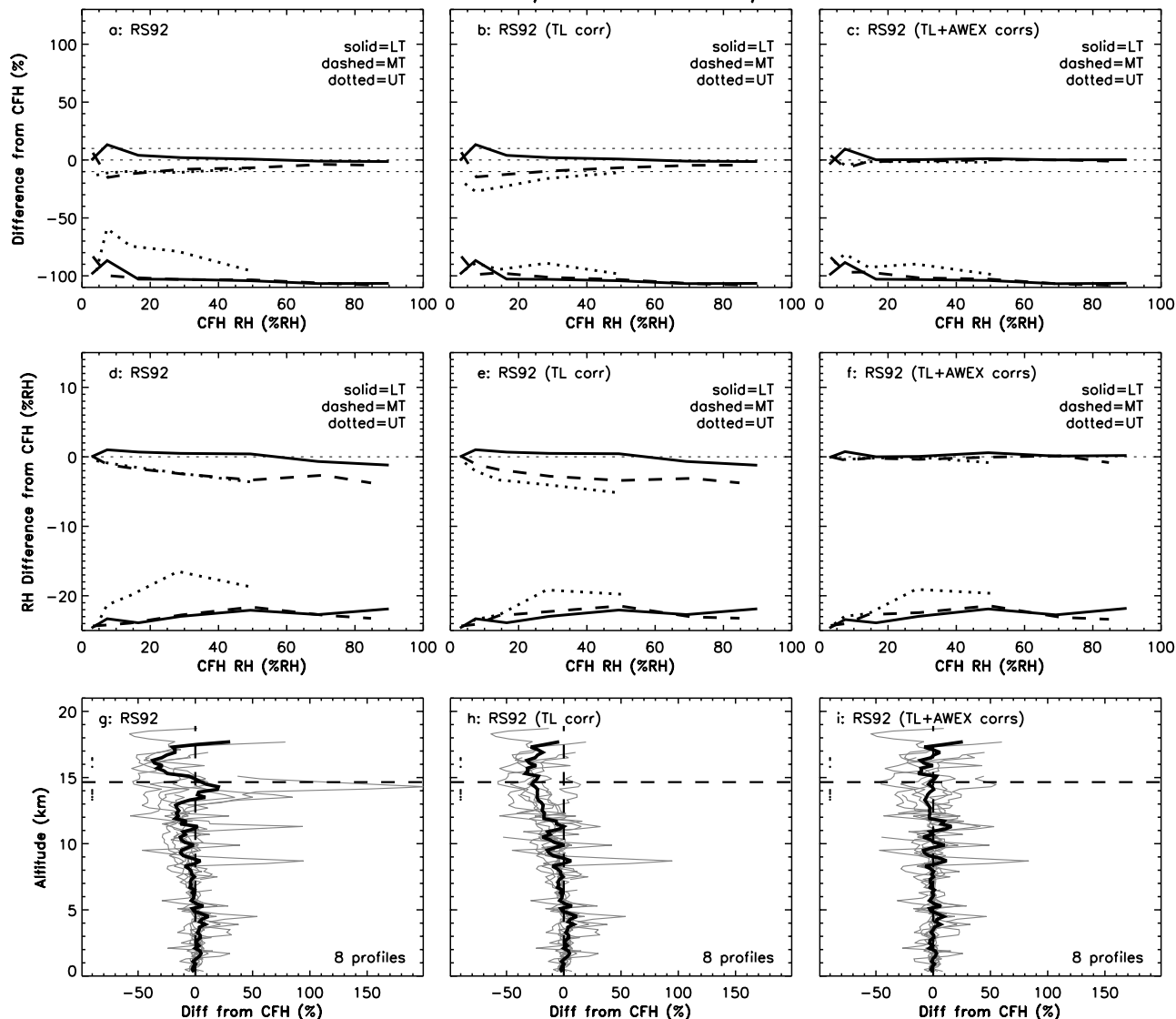## RS92 Mean Accuracy and Variability Relative to CFH

Figure 5. (a−f) Mean and standard deviation of differences between Vaisala RS92 and simultaneous CFH measurements as successive corrections are applied. Figures 5a–5c show absolute (percentage) differences (i.e., $(RH_{sonde} - RH_{CFH})/RH_{CFH} \times 100\%$), and Figures 5d–5f show RH differences ($RH_{sonde} - RH_{CFH}$). The top three curves in each panel show the mean difference from CFH as a function of RH, partitioned into three temperature intervals as defined in Figure 4. The bottom three curves in each panel show the standard deviation of the differences from CFH, offset for clarity such that zero is at the bottom of the panel. The left column shows the difference from CFH for standard (uncorrected) RS92 measurements, the middle column is after applying the time lag (TL) correction, and the right column is after also applying the empirical calibration correction that is derived from the AWEX data set as explained in section 5.2.1. Horizontal dashed lines in the top row indicate the AIRS ±10% goal for retrieval accuracy. The analysis uses 200 m averages from the surface to 3 km above the tropopause. (g−i) Altitude profiles of the percentage difference from CFH as successive corrections are applied. The thin curves are for the individual profiles, and the bold curves show the mean. The tropopause altitude for each profile is indicated by a dot at left, and the horizontal dashed line is the mean tropopause altitude.

and a mean dry bias under moist conditions. The dry bias when RH > 40% increases in magnitude with decreasing T to as much as 8% RH in the MT and 10% RH in the UT (Figure 4d), or, in absolute terms, a 10% dry bias in the MT and up to a 20% dry bias in the UT (Figure 4a). After application of the TD and TL corrections (Figures 4b or 4e),

both the mean accuracy and variability of the RS80-H measurements are substantially improved in the MT and UT, and much of the T-dependence of the measurement error disappears, which is mostly attributable to the TD correction. The TL correction, which recovers the vertical structure in the RH profile that was ''smoothed'' by slow

sensor response, is largely responsible for reducing the variability in the MT and UT.

[38] *Ferrare et al.* [2004] evaluated RS80-H and several other in situ and remote sensor water vapor measurements in the UT during the ARM-FIRE Water Vapor Experiment (AFWEX), conducted at the ARM SGP site in November 2000. The reference sensor for their accuracy assessment was the Lidar Atmospheric Sensing Experiment (LASE), an airborne DIAL lidar onboard the NASA DC-8. After correction for known systematic errors, the residual bias uncertainty in the LASE measurements during AFWEX was estimated to be 3–5%, and the absolute LASE accuracy including random error was estimated to be in the range 5–10%. Relative to LASE, the RS80-H exhibited a mean dry bias of ∼20% in the UT, which decreased to ∼5% (within the LASE uncertainty) after applying the TD and TL corrections. Almost all of the AFWEX research flights encountered very moist air in the UT (RH > 40%), and the 20% mean dry bias observed during AFWEX is in good agreement with the 15–20% mean dry bias relative to CFH observed in the UT during AWEX for uncorrected RS80-H measurements (Figure 4a, dotted, at RH > 40%). This agreement relative to two reference sensors of comparable absolute accuracy (LASE and CFH) gives confidence that the RS80-H measurement accuracy did not change substantially over the 2000 to 2003 timeframe, and also supports the general utility of the TD and TL corrections.

[39] It is important to note that the overall RS80-H mean accuracy in the UT during AWEX, when viewed as a function of altitude as was done by *Ferrare et al.* [2004], indicates a moist bias in the UT (Figure 4g) rather than a dry bias as observed during AFWEX. The reason for this apparent (but not real) inconsistency is that many of the AWEX profiles were quite dry in the UT, and Figure 4a shows that dry conditions contribute a large moist bias to the overall mean RS80-H accuracy in the UT during AWEX. Thus altitude-based observations in the UT of both a mean moist bias (AWEX) and a mean dry bias (AFWEX) are entirely consistent and explained by the RH dependence of the RS80-H measurement accuracy. More importantly, evaluating radiosonde accuracy in terms of altitude is less general and less illuminating than an accuracy evaluation in terms of RH and T, because the underlying RH dependence of the calibration accuracy is averaged over the specific profiles in the sample, and therefore accuracy conclusions given in terms of altitude are profile-dependent and are not generally applicable to other profiles having different characteristics. The RH and T dependences must be separated in order to assess radiosonde calibration accuracy in a profile-independent way.

[40] Standard RS92 measurements (Figures 5a or 5d) are considerably more accurate and less variable than standard RS80-H measurements, particularly for dry conditions (e.g., RH < 20%). The TL correction substantially reduces the RS92 variability in the UT (Figures 5b versus 5a, bottom dotted curves), but the RS92 mean accuracy in the UT is actually worse after applying the TL correction. By recovering a more realistic profile shape and thereby reducing the variability, the TL correction reveals underlying T-dependent calibration error that was hidden by variability in the time lag error. Reduced variability and the resulting RS92 mean dry bias of 10–25% in the UT is also apparent in the

individual altitude profiles (Figures 5h versus 5g, thin curves in the UT). An empirical correction for the residual mean bias error seen in the middle columns of Figures 4 and 5 is described in the next subsection, and application of this new correction leads to the right columns of Figures 4 and 5.

[41] Although the top rows (percentage accuracy) and the middle rows (RH accuracy) in Figures 4 and 5 contain the same information, their differences illustrate that modest errors of a few %RH correspond to quite large percentage errors under dry conditions. For example, at 10% RH the mean residual error after applying corrections is, depending on the temperature, 10–25% for RS92 (Figure 5b) and 15–70% for RS80-H (Figure 4b). The large spikes on the altitude plots in Figures 4 and 5 correspond to very dry conditions, where a small %RH difference corresponds to a large percentage difference. This error under dry conditions could be reduced by the manufacturer in two ways: (1) by minimizing curve fit error in the calibration, perhaps using a method similar to that used by M04 to reduce curve fit error from sensor time constant measurements and (2) by reporting RH measurements with a resolution that is better than the standard 1% RH.

5.2.1. AWEX Empirical Calibration Correction

[42] The mean accuracy curves in Figures 4 and 5 describe the accuracy of nighttime Vaisala RH measurements and its dependence on RH and T, and the standard deviation curves describe the sensor-to-sensor production variability. When expressed as a function of RH and T, the mean accuracy curves can be interpreted as describing the absolute accuracy of the Vaisala calibration (relative to CFH), if all non-calibration-related sources of bias error are removed. The primary noncalibration sources of error in Vaisala RH measurements are time lag error and solar radiation error (apart from the random production variability that should have no effect on the mean for a sufficiently large data set). All of the AWEX soundings occurred at night, and therefore after applying the time lag correction, the mean accuracy curves in the middle columns of Figures 4 and 5 describe the accuracy of the Vaisala calibration over the full range of RH and T conditions in the midlatitude troposphere. We can derive an empirical correction for calibration bias error from the mean accuracy curves by determining the RH- and T-dependent correction factor that, when applied to time lag-corrected measurements, yields corrected measurements that have zero mean bias relative to the CFH for all conditions of RH and T.

[43] The "AWEX empirical calibration correction" is essentially the reciprocal of the mean accuracy curves in Figures 4b and 5b, after algebraically recasting them in terms of the radiosonde-measured RH instead of the CFH-measured ("true") RH. Figure 6 shows the AWEX calibration correction for the RS80-H (Figures 6a and 6b) and the RS92 (Figures 6c and 6d), where the two panels for each sensor give the correction curves both before and after applying the "sensor-based" TL and TD corrections. The effect of the sensor-based corrections (Figures 6b versus 6a and 6d versus 6c) is to reveal the underlying regularity in the RH and T dependences of the calibration bias error, giving some confidence that only calibration-related bias error remains in the data. The RS80-H calibration correction after applying the TL and TD corrections (Figure 6b) depends almost exclusively on RH, which is encouraging
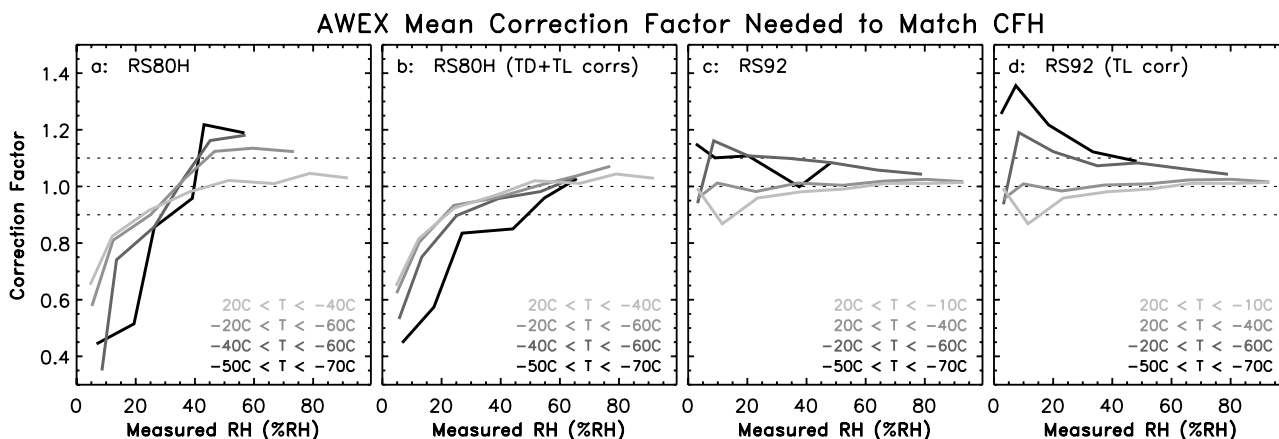
AWEX Mean Correction Factor Needed to Match CFH

Figure 6. Multiplicative correction factor curves for mean calibration error in (a and b) Vaisala RS80-H and (c and d) RS92 measurements, derived from the same 11 RS80-H profiles and 8 RS92 profiles assessed in Figures 4 and 5. Curves show the correction factor as a function of the measured RH for the indicated temperature intervals, as applies to both the standard (uncorrected) data (Figures 6a and 6c) and to the data after first correcting for the indicated measurement errors (Figures 6b and 6d). For the most accurate results this empirical correction should be applied after the other corrections (using Figures 6b or 6d), for reasons discussed in the text. Numerical values for the correction factors are given in Table 2.

since the TD correction was expected to reduce inaccuracy in the T-dependent portion of the calibration. The RS92 calibration correction after applying the TL correction (Figure 6d) depends on both RH and T, and has the appearance of addressing a curve fit error whose maximum magnitude occurs at 10% RH. The correction curves that apply to the standard data before correcting for the sensor-based error(s) (Figures 6a and 6c) are less regular in appearance primarily because they also contain (and correct for) the mean time lag error in this particular data set. Since time lag error is very profile-specific and is not a simple bias error that depends only on RH and T, correcting for only its mean value in the AWEX data set is much less accurate and less general than the recommended approach of first removing the time lag error and then applying the profile-independent empirical calibration correction (Figures 6b or 6d).

[44]   The empirical calibration correction from Figures 6b and 6d was applied to the AWEX RS80-H and RS92 measurements after applying the sensor-based corrections, resulting in the mean accuracy and variability relative to CFH that is shown in the right columns of Figures 4 and 5. In applying the correction, the curves from Figure 6 were linearly interpolated to the RH and T measurements, where the numerical values of the correction factors shown in Figure 6 are given in Table 2. The empirical calibration correction substantially improves the mean accuracy of the radiosonde data relative to CFH, which of course is by design. The mean RS92 accuracy (Figure 5c) is well within the AIRS ±10% absolute accuracy goal for all conditions in the midlatitude troposphere and lowermost stratosphere, and even the corrected RS80-H mean accuracy (Figure 4c) is essentially within the AIRS accuracy goal for all conditions. The greatest effect of the empirical calibration correction is to improve the absolute accuracy of the data under dry conditions, where the radiosonde calibration is the least accurate in terms of absolute water vapor amount. Note that the mean accuracy after applying the empirical calibration

correction is not identically zero for all RH and T conditions as might be expected, because the binning of the data in Figures 4 and 5 is different from that in Figure 6 used to derive the correction. The difference of the corrected mean accuracy from zero is attributed to two sources: (1) statis-

Table 2. Empirical Correction Factors for Mean Calibration Error in Vaisala RS80-H and RS92 Radiosonde RH Measurements Relative to the CFH, as Plotted in Figure 6[a]

T, °C	3%	10%	22.5%	37.5%	52.5%	67.5%	82.5%	94.5%
				RS80H Sensor				
−10	0.653	0.825	0.916	0.981	1.021	1.010	1.046	1.030
−40	0.578	0.810	0.900	1.027	1.124	1.135	1.123	...
−50	0.350	0.741	0.857	1.031	1.162	1.181	...	...
−60	0.444	0.515	0.856	0.958	1.218	1.189	...	...
				RS80H (TD + TL) Sensor				
−10	0.649	0.815	0.924	0.970	1.020	1.011	1.044	1.029
−40	0.623	0.803	0.933	0.961	1.002	1.036	1.071	...
−50	0.532	0.751	0.897	0.957	0.982	1.033	...	...
−60	0.448	0.574	0.835	0.850	0.959	1.027	...	...
				RS92 Sensor				
+5	0.994	0.868	0.959	0.981	0.991	1.011	1.011	1.014
−10	0.958	1.012	0.982	1.012	1.004	1.019	1.025	1.017
−40	0.941	1.161	1.108	1.099	1.084	1.058	1.043	...
−60	1.150	1.101	1.108	0.999	1.085	...	...	...
				RS92 (TL) Sensor				
+5	0.994	0.868	0.959	0.981	0.991	1.011	1.011	1.014
−10	0.955	1.009	0.984	1.005	1.009	1.023	1.025	1.016
−40	0.937	1.190	1.123	1.073	1.083	1.063	1.043	...
−60	1.256	1.356	1.217	1.122	1.089	...	...	...

[a]Correction factors are shown as a function of the radiosonde-measured RH for the indicated temperature, where the RH and T values are the bin centers. Correction factors are given for both the original radiosonde measurements and the measurements after correction for time lag (TL) error (and also TD error for RS80-H radiosondes). Much more accurate results are obtained if the TL and TD corrections are applied first, as discussed in the text.
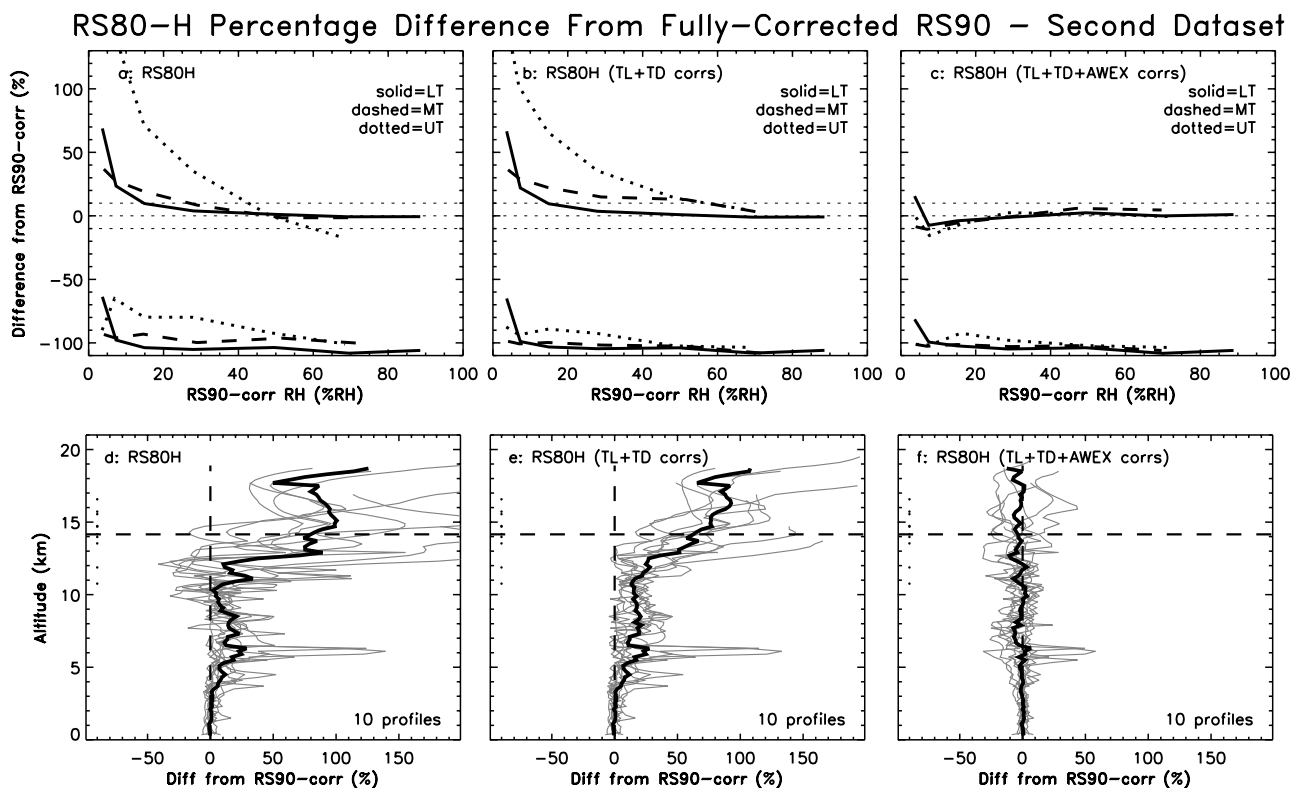
Figure 7. Percentage difference of RS80-H measurements from corrected RS90 measurements ("RS90-corr") as the reference standard, for a second (independent) AWEX data set of simultaneous RS80-H and RS90 measurements. The time lag and AWEX empirical calibration corrections were applied to the RS90 reference measurements, whose accuracy is tied to the CFH absolute reference standard via Figure 5c and the assumption that RS90 and RS92 RH sensors and calibrations are equivalent. (a–c) Mean and standard deviation of the percentage difference as a function of RH for three temperature intervals and (d–f) percentage difference as a function of altitude, as described in Figure 4. The left column shows the difference from RS90-corr for the standard (uncorrected) RS80-H data, the middle column is after applying the TL and TD corrections to the RS80-H data, and the right column is after also applying the AWEX empirical calibration correction to the RS80-H data.

tical uncertainty from the limited sample size that is binned in two dimensions to derive the correction and (2) the possibility of other small (noncalibration) bias errors of different dependence (e.g., unfiltered instances of RS80-H sensor icing in clouds). The residual bias error in the right columns of Figures 4 and 5, revealed by using different binning, might be termed "uncertainty in the method."

5.2.2. Validation of the AWEX Calibration Correction

[45] A second independent radiosonde/CFH data set is not available for validation of the empirical calibration correction; however, the generality of the correction can be assessed using a second AWEX data set of RS80-H and RS90 radiosondes that were flown on the same balloon, but were not flown with the CFH. As described earlier, RS90 and RS92 radiosondes use nearly identical RH sensors and calibration procedures, so we will apply the RS92 empirical calibration correction to the RS90 data. The fully corrected RS90 will act as a proxy reference sensor whose mean accuracy, given by the mean accuracy of the fully corrected RS92 relative to CFH (Figure 5c), is <2% under all conditions except near RH = 10%. Figure 7 shows the percentage difference between the RS80-H measurements and the fully corrected RS90 reference measurements as a

function of RH and T (top row) and as a function of altitude (bottom row). The final comparison to the RS90 reference measurements after applying all corrections to the RS80-H measurements (Figure 7c), is roughly equivalent to the mean accuracy of fully corrected RS80-H measurements relative to the CFH (Figure 4c). Within the aforementioned "accuracy of the method," the empirical calibration correction improves the mean accuracy of both the RS90 and RS80-H measurements in this second data set to a degree that is similar to its performance with the first (CFH) data set. This consistency gives confidence in the generality of the empirical calibration correction and the physical interpretation that it addresses inaccuracy in the Vaisala calibration. It is particularly noteworthy that the functional form of the RS80-H and RS9x empirical calibration corrections are entirely different, so the correction leads to RS80-H and RS90 measurements that converge on the same result from original measurements that are often quite different, especially in the UT.

[46] The validity of the sensor-based Vaisala corrections (TL and TD) has been established by previous studies, and is confirmed by this study. The new AWEX empirical calibration correction derived here, while validated with

Table 3. Mean Percentage Accuracy and its Variability (Standard Deviation) for Radiosonde Types Flown Together on the Same Balloon During AWEX, Given by the Percentage Difference From Simultaneous Measurements by the Designated Reference Sensor (Positive Accuracy Values Indicate Moister Than the Reference Sensor)[a]

T Range and Sensor	1–5% RH	5–10% RH	10–20% RH	20–40% RH	40–60% RH	60–80% RH	80–99% RH
Radiosonde Mean Percentage Accuracy and Variability Relative to CFH (Nighttime)							
Lower Trop (T > −20°C)							
RS80H	55.3 ± 37.1 (31)	36.1 ± 24.8 (14)	17.0 ± 11.3 (42)	6.9 ± 7.4 (94)	−0.1 ± 7.0 (65)	−0.6 ± 5.7 (35)	−2.8 ± 3.7 (33)
RS80H-corr	1.0 ± 24.0 (31)	4.9 ± 23.6 (14)	1.3 ± 9.7 (42)	0.6 ± 6.9 (94)	−0.3 ± 6.6 (65)	1.2 ± 5.9 (35)	0.5 ± 3.6 (33)
RS92	0.0 ± 11.5 (18)	13.2 ± 23.2 (13)	4.1 ± 7.1 (45)	2.1 ± 6.9 (74)	0.8 ± 5.8 (44)	−1.0 ± 3.3 (29)	−1.3 ± 3.5 (30)
RS92-corr	−3.8 ± 10.7 (18)	9.5 ± 21.5 (13)	0.2 ± 7.1 (45)	0.3 ± 6.7 (74)	1.1 ± 6.0 (44)	0.1 ± 3.2 (29)	0.2 ± 3.6 (30)
SW (no ice)	157.4 ± 97.5 (31)	23.5 ± 35.7 (17)	−1.0 ± 11.4 (47)	1.0 ± 6.4 (105)	−0.9 ± 2.8 (72)	−1.0 ± 2.4 (35)	−1.3 ± 1.7 (36)
Middle Trop (−50 < T < −20°C)							
RS80H	73.0 ± 73.7 (30)	16.5 ± 19.8 (27)	9.8 ± 23.1 (27)	5.7 ± 23.8 (38)	−8.4 ± 12.0 (34)	−12.6 ± 11.0 (17)	−9.5 ± 2.1 (10)
RS80H-corr	1.6 ± 25.1 (30)	−14.3 ± 16.6 (27)	−10.1 ± 10.4 (27)	−3.6 ± 7.4 (38)	−1.0 ± 8.0 (34)	−1.2 ± 2.9 (17)	−2.2 ± 1.6 (10)
RS92	6.3 ± 26.7 (27)	−15.0 ± 10.2 (25)	−12.0 ± 8.6 (19)	−8.0 ± 6.8 (35)	−6.7 ± 6.5 (28)	−3.8 ± 3.6 (20)	−4.4 ± 2.0 (15)
RS92-corr	3.9 ± 26.0 (27)	−7.1 ± 13.0 (25)	−1.4 ± 13.2 (19)	−1.0 ± 8.2 (35)	0.1 ± 7.1 (28)	0.1 ± 2.9 (20)	−1.0 ± 1.9 (15)
SW (no ice)	140.1 ± 121.2 (30)	4.7 ± 31.0 (27)	−5.0 ± 23.2 (27)	0.5 ± 6.5 (38)	−2.1 ± 4.7 (33)	−2.7 ± 4.4 (12)	−2.8 ± 1.2 (10)
Upper Trop (T < −50°C)							
RS80H	125.3 ± 56.2 (12)	113.0 ±124.1 (52)	51.9 ± 57.9 (71)	0.6 ± 36.3 (29)	−20.9 ± 18.4 (18)	−15.9 ± 12.0 (12)	
RS80H-corr	6.2 ± 20.5 (12)	14.3 ± 60.8 (52)	−4.7 ± 35.9 (71)	−6.8 ± 28.8 (29)	−3.0 ± 13.2 (18)	−1.7 ± 4.7 (12)	
RS92	−13.0 ± 14.4 (13)	−11.2 ± 50.8 (53)	−9.7 ± 35.6 (64)	−10.8 ± 31.1 (30)	−7.1 ± 13.8 (23)		
RS92-corr	1.9 ± 20.7 (13)	−4.4 ± 28.9 (53)	−1.5 ± 17.4 (64)	−1.4 ± 20.2 (30)	−1.9 ± 11.3 (23)		
SW (no ice)	39.0 ± 26.4 (12)	10.9 ± 33.2 (47)	−6.6 ± 29.7 (65)	−8.4 ± 35.7 (24)	−2.1 ± 11.5 (9)		
Radiosonde Mean Percentage Accuracy and Variability Relative to RS90-corr (Nighttime)							
Lower Trop (T > −20°C)							
RS80H	59.0 ± 37.8 (7)	25.6 ± 13.6 (16)	11.9 ± 7.7 (66)	5.1 ± 4.9 (59)	2.5 ± 7.0 (43)	−1.9 ± 1.9 (28)	−2.2 ± 4.0 (48)
RS80H-corr	8.7 ± 24.6 (7)	−5.9 ± 11.6 (16)	−4.5 ± 8.3 (66)	−1.1 ± 5.6 (59)	2.9 ± 7.1 (43)	−0.3 ± 2.0 (28)	1.0 ± 4.1 (48)
MODEM	217.0 ± 224.9 (55)	114.9 ± 44.7 (32)	52.3 ± 29.7 (74)	18.4 ± 12.3(133)	7.3 ± 9.9 (56)	4.8 ± 8.2 (35)	1.7 ± 4.8 (67)
SIPPICAN	276.5 ± 168.0 (43)	47.5 ± 96.5 (13)	22.0 ± 41.3 (34)	0.2 ± 17.2 (38)	−1.0 ± 7.7 (45)	−2.0 ± 4.2 (30)	1.2 ± 4.4 (24)
Middle Trop (−50 < T < −20°C)							
RS80H		25.0 ± 12.0 (13)	9.2 ± 14.7 (24)	1.2 ± 9.2 (76)	−5.2 ± 11.9 (18)		
RS80H-corr		−10.9 ± 8.3 (13)	−5.6 ± 9.6 (24)	−0.7 ± 7.0 (76)	5.4 ± 8.0 (18)		
MODEM	−11.5 ± 142.1 (47)	41.0 ± 67.5 (25)	27.5 ± 36.5 (32)	30.4 ± 26.9 (96)	−0.4 ± 9.9 (41)	−18.7 ± 14.2 (6)	
SIPPICAN	416.9 ± 173.4 (39)	12.9 ± 46.0 (13)	−1.5 ± 50.7 (20)	−10.5 ± 38.3 (41)	−21.7 ± 24.3 (18)		
Upper Trop (T < −50°C)							
RS80H	150.7 ± 15.6 (8)	114.0 ± 29.7 (27)	39.9 ± 28.8 (54)	23.7 ± 23.7 (81)	−8.8 ± 20.2 (17)	−21.2 ± 7.4 (17)	
RS80H-corr	−3.8 ± 9.6 (8)	−14.9 ± 9.4 (27)	−1.0 ± 19.9 (54)	3.7 ± 11.8 (81)	2.9 ± 8.4 (17)	−1.1 ± 6.6 (17)	
MODEM	−83.4 ± 51.5 (44)	−41.0 ± 98.9 (41)	−18.6 ± 80.8 (90)	−6.0 ± 39.0 (84)	−13.7 ± 8.5 (22)	−27.2 ± 4.3 (22)	
SIPPICAN	316.7 ± 128.3 (31)	202.2 ± 127.3 (47)	80.6 ± 55.8 (28)	−2.3 ± 44.9 (37)	−39.3 ± 19.9 (8)		

[a]Values in the top half of the table are relative to the CFH, whose absolute accuracy is estimated in section 2 and Figure 1. Sensor types with an appended "-corr" refer to measurements after the appropriate corrections (TL, TD, AWEX) have been applied. The SW values refer to the SW accuracy when measurements within and above thick ice clouds are removed, corresponding to Figure 8c. Values in the bottom half of the table are relative to the RS90 after applying the time lag and AWEX empirical calibration corrections ("RS90-corr"), where the RS90 accuracy and variability relative to CFH are taken as equivalent to the RS92 as given in the top half of the table ("RS92-corr") and in Figure 5c. The mean accuracy and variability are given in seven RH intervals for each of three temperature ranges. The number of samples for each comparison (200 m averages) is given in parentheses. The AWEX soundings reflect the radiosonde accuracy for nighttime measurements only; daytime soundings contain additional bias error caused by solar radiation, as discussed in section 6.

an independent AWEX data set, would nonetheless benefit from additional validation given the statistically small size of the data set used to derive the correction. Although the general conclusions and approximate magnitude of the correction probably would not change if the data set were larger, more data would reduce the uncertainty in the correction and would allow for more accurate determination of the production variability, including variability between calibration batches. The generality of the AWEX calibration correction would benefit from further study using a different CFH/radiosonde data set, ideally from a tropical location where the calibration accuracy at temperatures below the −70°C limit of the AWEX data could be investigated. Caution must be exercised in applying the AWEX correction to other data sets because Vaisala may change their calibration model without notice, as with the RS90 TD calibration model change on 25 June 2001 that is discussed in Appendix B. The RS80-H calibration is unlikely to change in the future, as is the RS90 calibration since the RS80-H and RS90 may soon be discontinued by Vaisala.

However, the RS92 TD calibration model was changed by Vaisala on 6 April 2004, so the AWEX calibration correction is not valid for RS92 radiosondes produced after that date (which can be determined from the serial number as described in the Appendix of M04). Finally, the accuracy curves derived in this section describe only the nighttime accuracy of Vaisala radiosondes (although the AWEX calibration correction should be valid day or night, since it addresses the fundamental calibration accuracy). The effect of solar radiation on daytime Vaisala RH measurements is discussed in section 6. Numerical values for the mean accuracy and variability of all radiosonde types investigated in this paper are summarized in section 5.5 (Table 3).

5.3. Snow White

[47] Direct comparisons between the Meteolabor Snow White (SW) chilled mirror hygrometer and the CFH are shown in Figure 8, as a function of RH and T in Figures 8a–8c, and as a function of altitude in Figures 8d–8f. The mean difference between the SW and CFH measurements is <2% in
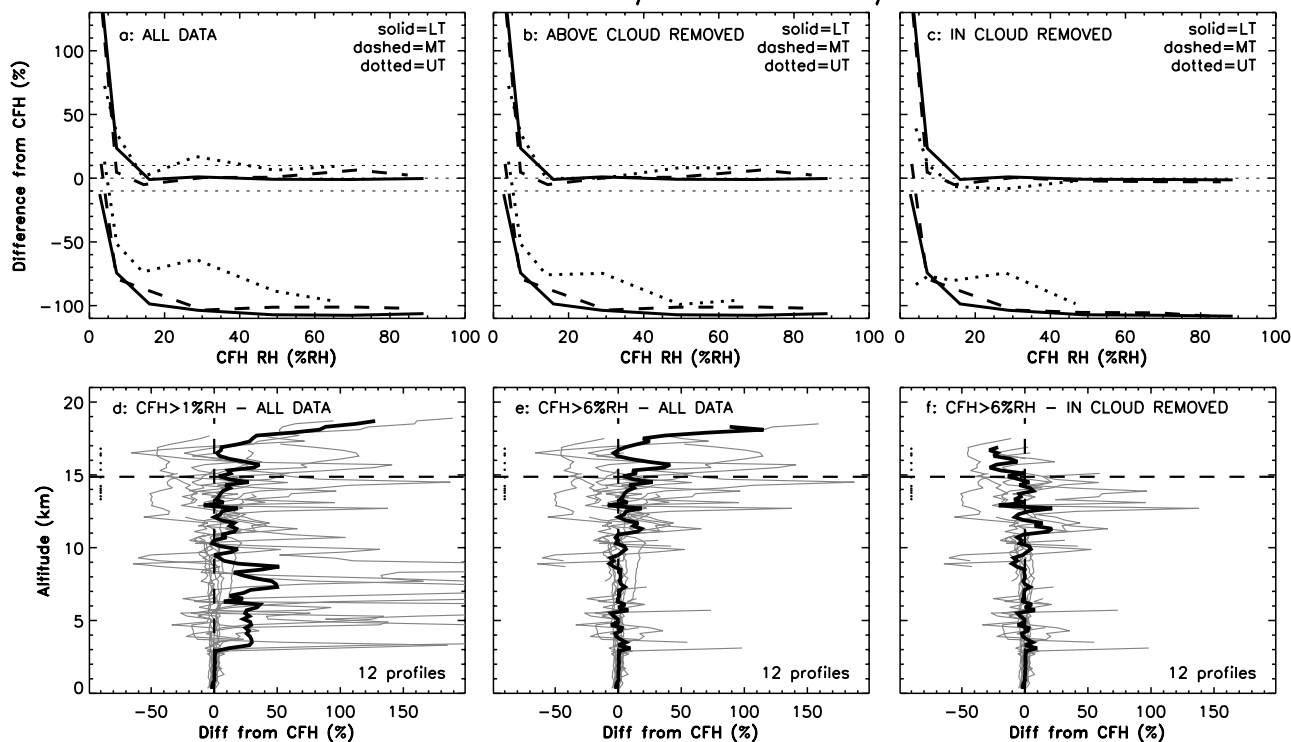
Figure 8. Percentage difference between simultaneous measurements from the Meteolabor Snow White hygrometer and the CFH. (a−c) Mean and standard deviation of the percentage difference as a function of RH for three temperature intervals and (d−f) individual profiles and mean percentage difference as a function of altitude, as described in Figure 4. All SW data above a minimum RH cutoff of 1% RH are considered in Figures 8a–8d. Data above thick ice clouds are removed for Figure 8b, and data both within and above thick ice clouds are removed for Figures 8c and 8f. The minimum RH cutoff is increased to 6% RH for Figures 8e and 8f.

the LT except under very dry conditions (Figure 8a), consistent with the finding in section 2 that the SW and CFH accuracies are comparable. However, the SW exhibits a moist bias under moist conditions of up to 8% in the MT and 5–15% in the UT. Inspection of the RH profiles shows that the SW usually exhibits an elevated RH relative to the CFH and RS92 when within and especially above thick ice clouds (e.g., Figure 3b), which is attributed to sublimation of ice particles trapped in the inlet duct and/or sensor casing by waste heat from the mirror heater. Much of the moist bias under somewhat dry conditions in the UT is eliminated when the SW measurements above thick ice clouds are removed from the analysis (Figure 8b), which leaves a moist bias of 5–8% in the MT and UT for conditions near ice saturation (i.e., within clouds). When the in-cloud SW measurements are also removed from the analysis (Figure 8c), both the mean accuracy and variability in the MT are improved and are comparable to the performance in the LT. However, removal of cloud-affected data reveals a residual dry bias of 5–10% in the UT, consistent with the SW performance in the UT observed during AFWEX by *Ferrare et al.* [2004], and consistent with the "mixed results" in the UT reported by *Vömel et al.* [2003], which might be caused by controller instability at low temperatures.

[48] The poor accuracy and large variability at all temperatures under very dry conditions seen in Figures 8a–8c results from the limitations of electrical cooling of the

mirror. The large spikes in some of the individual profiles in Figure 8d are the result of dry layers, as demonstrated by their disappearance and the consequent improvement in the mean SW accuracy throughout the profile if SW measurements are removed from the analysis whenever the RH measured by the CFH is <6% (Figure 8e), consistent with the SW performance under very dry conditions described by *Vömel et al.* [2003]. The remaining moist bias in the UT is eliminated when the SW measurements within and above thick ice clouds are removed (Figure 8f).

[49] The analysis in Figure 8 suggests that the SW is a reference-quality water vapor sensor of accuracy comparable to the CFH, but only under a limited set of conditions roughly described by T > −50°C, RH > 6%, and below thick ice clouds. It should be stated that the "potential accuracy" of the SW represented by Figure 8c is only attainable if cloud-affected data can be identified and removed, such as by comparing to measurements from a second sensor. Furthermore, without detailed quality control and data processing substantial SW measurement error can exist in the lower troposphere down to mirror temperatures of −20 to −30°C if the phase of the condensate on the mirror is not properly identified.

5.4. Modem and Sippican

[50] The mean accuracy and variability of RS80-H, RS9x, and SW radiosondes were evaluated in the preceding
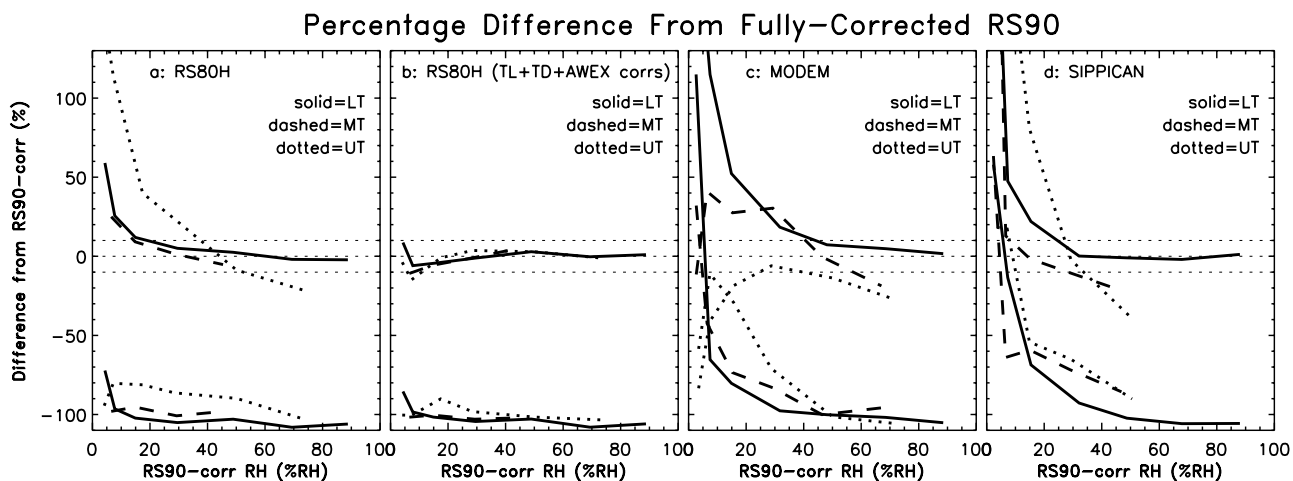
Figure 9. Mean percentage accuracy and variability (standard deviation) for three radiosonde types ((a and b) RS80-H, (c) Modem, and (d) Sippican) that were launched on the same balloon with RS90 but were not launched with the CFH. The reference sensor for these accuracy estimates is the RS90 after applying the time lag and AWEX empirical calibration corrections ("RS90-corr"). The mean accuracy of the corrected RS90 relative to the CFH, taken as equivalent to the RS92 mean accuracy, is given in Figure 5c. The RS80-H comparison is also shown after the TL, TD, and AWEX empirical calibration corrections were applied (Figure 9b). The data represent 8 RS80-H profiles, 15 Modem profiles, and 7 Sippican profiles. Curves are defined as in Figure 4.

sections by comparing to simultaneous measurements of known absolute accuracy from the CFH. A second set of multiple-radiosonde balloon launches that did not include the CFH was also acquired during AWEX, involving various combinations of RS80-H, RS90, Modem, and Sippican radiosondes. The mean accuracy and variability of the Modem and Sippican radiosondes are evaluated in this section in the same manner as in the previous sections, but using the RS90 rather than the CFH as the reference sensor. Specifically, comparisons are made to the fully corrected RS90 measurements (TL correction plus AWEX empirical calibration correction), where the mean accuracy of the corrected RS90 measurements relative to the CFH (Figure 5c) is 1–3%, except under very dry conditions. Note that this use of the RS90 as a "proxy reference and transfer standard" assumes that the RS90 and RS92 sensors and calibration procedures are essentially equivalent, which is supported by the successful application of the RS92 empirical calibration correction to the RS90 validation data in Figure 7, and also by the three dual RS90/RS92 colaunches conducted during AWEX.

[51] The mean percentage accuracy and variability of RS80-H, Modem, and Sippican measurements relative to the corrected RS90 reference measurements ("RS90-corr") are shown in Figure 9. The RS80-H comparisons involve approximately the same soundings that were used to validate the AWEX empirical calibration correction in Figures 7a and 7c, and are shown for comparison to the Modem and Sippican measurements. Figure 9a shows that in general the original RS80-H measurements are more accurate and less variable than the Modem or Sippican measurements, especially in the MT and UT, and in the LT when RH < 40%. However, the Sippican measurements are at least as accurate as the RS80-H measurements in the LT when RH > 30%. Only under these conditions are the Modem and Sippican measurements suitably accurate for use in AIRS validation.

The tabulated mean accuracy and variability values for each radiosonde type are presented in the next section.

[52] If the RS80-H corrections are applied, the RS80-H mean accuracy is <5% at all temperatures when RH > 15%, and is ~10% when RH < 15%, which is notably better than the Modem and Sippican accuracies. It is not justifiable to derive an empirical calibration correction for the Modem or Sippican measurements for three primary reasons: (1) Substantial time lag error in the UT and MT (see Figure 3) cannot be removed to reveal the underlying calibration error because the necessary sensor time constant information is not available; (2) the unresponsiveness of Sippican sensors in the UT and upper MT appears to be either severe time lag error or an inherent limitation of the carbon hygristor rather than a (correctable) calibration issue; and (3) insufficient information is available about the Sippican and Modem calibration procedures to conclude that mean differences are attributable to calibration accuracy alone, even if the time lag error could be removed.

[53] The mean accuracy of RS80-H (original and corrected), Modem, and Sippican measurements relative to RS90-corr is shown as a function of altitude in Figure 10, with a minimum RH cutoff of either 1% RH (top row) or 6% RH (bottom row). A minimum RH cutoff of 10% RH (not shown) is very similar to the 6% RH cutoff. The higher RH cutoff makes only a small difference for RS80-H radiosondes but a much larger difference for Modem and Sippican radiosondes, because the inaccuracy under very dry conditions is more severe for the Modem and Sippican radiosondes. As mentioned earlier, accuracy shown in terms of altitude as in Figure 10 is potentially misleading because the accuracy is fundamentally a function of RH and T, so accuracy results presented in terms of altitude are strongly dependent on the specific profiles measured, especially if dry layers are present. All 3 radiosonde types are very accurate near the surface, which is an important observation
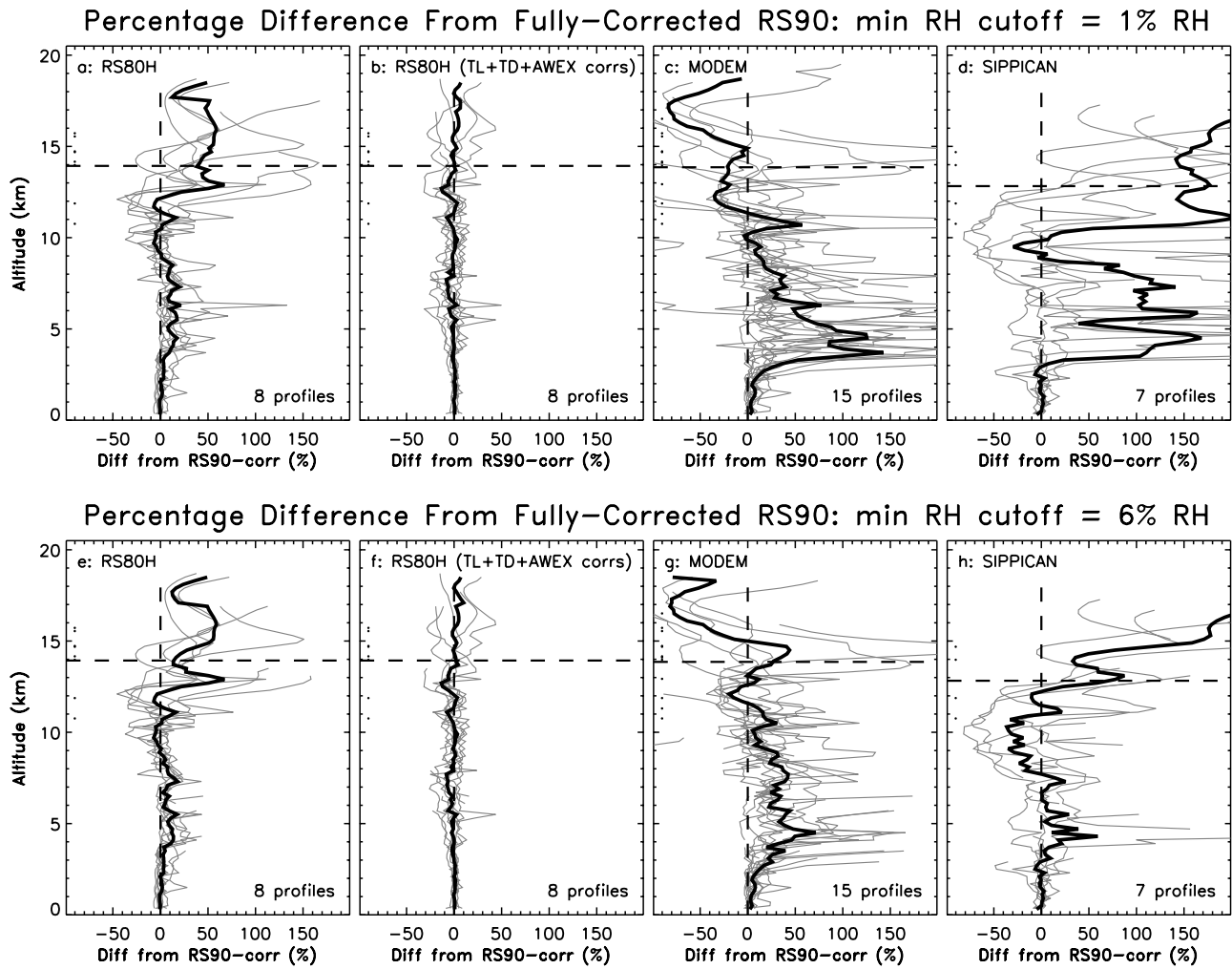
Figure 10. Mean percentage accuracy and variability as a function of altitude for three radiosonde types (RS80-H, Modem, and Sippican) that were launched on the same balloon as RS90 but were not launched with the CFH (same data as in Figure 9). The reference sensor for these comparisons is the RS90 after applying the time lag and AWEX empirical calibration corrections ("RS90-corr"). The mean percentage accuracy of the corrected RS90 relative to the CFH, taken as equivalent to the RS92 accuracy, is shown as a function of altitude in Figure 5i. The same data are plotted in the top and bottom rows, except that a different minimum RH cutoff of either (a–d) 1% RH or (e–h) 6% RH was used to show the impact of very dry conditions on the mean measurement accuracy. The RS80-H comparison is also shown after the TL, TD, and AWEX calibration corrections were applied (Figures 10b and 10f). Curves are defined as in Figure 4.

because it shows that evaluations of radiosonde accuracy based on comparison to surface reference sensors before launch are not at all representative of the overall radiosonde accuracy for soundings. Again it is seen that the RS80-H corrections are very effective, as the corrected measurement accuracy is generally <5% from the surface to the lower stratosphere (Figures 10b and 10f). The Modem measurements, even when very dry conditions are excluded (Figure 10g), are too moist by ~30–40% on average through much of the troposphere, with considerable variability between soundings. The Sippican measurements, when very dry conditions are excluded (Figure 10h), are comparable in accuracy to the RS80-H in the lowermost few km, but exhibit a moist bias of 10–30% in the MT, and are meaningless in the UT because of the lack of sensor

response at low temperatures, as was also found by *Ferrare et al.* [2004].

5.5. Summary of Radiosonde Measurement Accuracy

[54] The mean accuracy and variability of the radiosonde measurements evaluated in the previous sections are summarized as a function of RH and T in Table 3, which quantifies the comparisons shown graphically in Figures 4, 5, 8, and 9. The mean accuracy given in Table 3 is specifically the mean bias relative to the designated reference sensor, where the reference sensor for the radiosonde types listed in the upper half of the table (RS80-H, RS92, SW) is the CFH, and the reference sensor for the radiosonde types listed in the bottom half of the table (RS80-H, Modem, Sippican) is the corrected RS90 ("RS90-corr").

The mean accuracy listed in Table 3 is the most relevant parameter for statistical studies, such as climate trend analyses, or the statistical approach to validation of AIRS water vapor retrievals. However, the standard deviation, which is the 1-sigma uncertainty when any individual profile is considered, is the most relevant parameter for applications that depend on single radiosonde profiles, such as assessing the impact of observational error on a specific model forecast. The standard deviation values should be viewed only as a guideline and not as a robust description of manufacturing variability, mainly because the number of sensors tested is small (7–15).

[55] As discussed in section 2, the absolute percentage accuracy of the CFH measurements, against which the sensors in the top half of Table 3 are judged, is ∼3% in the LT and ∼6% in the UT. The mean percentage accuracy of the radiosonde types listed in the bottom half of Table 3 must also consider the accuracy of RS90-corr relative to CFH, taken as equivalent to the accuracy of RS92-corr relative to CFH, which is seen in the top half of Table 3 to be ∼1% in the mean for most conditions in the LT and MT (when RH > 10%), and ∼2% in the UT. The Vaisala RS92 (or RS90) is unquestionably the superior operational radiosonde type investigated in this study, in part because it is considerably more accurate and less variable than the other radiosonde types, and in part because it is capable of making reliable measurements within and above clouds.

[56] The accuracy of the radiosondes used in the NWS radiosonde network influences the accuracy of operational model forecasts, among many other applications. The radiosonde types used by the NWS were not chosen on the basis of the RH measurement accuracy alone, but also on the basis of the accuracy of the other thermodynamic and wind measurements, as well as economic and political considerations. As of this writing, 2/3 of NWS stations use Vaisala RS80-H radiosondes and 1/3 use Sippican radiosondes. For almost all conditions of RH and T, and considering both the mean accuracy and the variability, the RS80-H clearly provides the more accurate water vapor measurements, and this conclusion is amplified substantially if the RS80-H corrections are applied, one of which (TD) is very simple to implement. It is important for users of NWS radiosonde RH data to consider not only the accuracy and variability of the measurements from a given site, but also the impact of clouds (especially in the UT), which varies profile-to-profile, seasonally, and geographically. We conclude that NWS RH data are not suitable for rigorous climate-related research, in part because of the variable impact of clouds on RS80-H measurements, and in part because of the inability of Sippican radiosondes to measure RH reliably in the UT, upper MT, and under dry conditions at all temperatures.

6. Application to AIRS RS90 Validation Data

[57] One of the approaches to validating AIRS water vapor retrievals is to compare AIRS-measured spectral radiances with radiances calculated by a longwave radiative transfer model that uses radiosonde measurements coincident with Aqua satellite overpasses to specify the atmospheric state [Fetzer et al., 2003]. The ARM program, in support of AIRS validation, conducted a sounding campaign that targeted AIRS overpasses with pairs of RS90 soundings nominally separated in time by 45–60 min, such that during the overpass one radiosonde is in the UT while the other is in the LT. This approach allows the horizontal/temporal atmospheric variability within the AIRS footprint to be characterized, and a "Best Estimate" profile constructed from both soundings is used for AIRS validation [Tobin et al., 2006]. Six RS90 validation data sets, consisting of 3 time periods of dual soundings at both the ARM SGP and TWP sites, are evaluated in this section with the following two goals: to quantify diurnal bias in RS90 measurements caused by solar radiation and to quantify the impact of the time lag and AWEX empirical calibration corrections on the RS90 validation data. Although the corrected ARM RS90 validation data are available to any researcher involved in the AIRS validation effort, it is important to note that most validation investigations appearing in this special journal issue did not use the corrected data, as the empirical calibration correction was only recently developed and validated.

[58] The overall RS90 measurement accuracy in the LT can be evaluated by comparing the RS90 column-integrated precipitable water vapor (PWV) with simultaneous retrievals of PWV from an ARM microwave radiometer (MWR). Turner et al. [2003] developed a correction procedure whereby the radiosonde RH profile is scaled by a constant factor that matches the radiosonde PWV to the MWR PWV (i.e., scale factor $SF = PWV_{MWR}/PWV_{RS90}$). The intent of the scaling is to remove bias error from individual radiosonde soundings, using the MWR as a stable reference standard that has little or no diurnal variability. Turner et al. [2003] evaluated the MWR-scaling technique by using scaled and unscaled ARM RS80-H measurements as input to a radiative transfer model, then they compared the calculated downwelling longwave radiation to spectral measurements by the ARM Atmospheric Emitted Radiance Interferometer (AERI), and they found that the scaling substantially reduced the residuals and variability. The primary RS80-H bias error addressed by Turner et al. [2003] was the "chemical contamination error" that has since been addressed by Vaisala, but they also noted a dry bias in daytime RS80-H measurements of 3–4%, presumably caused by solar heating of the RH sensor. The MWR SF is used below to assess the impact of solar radiation on daytime RS90 measurements, and also to assess the impact of MWR scaling on the AIRS validation data. As discussed in section 2, all ARM MWR PWV measurements used in this study have been reduced by 3% to account for a bias error in the absorption model used in the MWR retrieval algorithm.

6.1. RS90 Daytime Dry Bias

[59] It was in one sense fortuitous that the AWEX soundings occurred at night, because otherwise the measurement error caused by solar radiation would have prevented isolation of the calibration-related uncertainties and subsequent development of the AWEX empirical calibration correction. Since the time lag and AWEX empirical calibration corrections are related to fundamental sensor principles, the AWEX correction should be applicable day or night, but solar heating of the RH sensor adds additional uncertainty to daytime measurements. As described earlier,

Table 4. Summary of RS90 Water Vapor Measurements and Comparison to Simultaneous Microwave Radiometer (MWR) Retrievals of Precipitable Water Vapor (PWV) for Six Sets of RS90 Soundings Conducted in Support of AIRS Validation at the ARM TWP and SGP Sites[a]

Data Set	Dates	RH_{surf},[b] %	PWV, cm	SF_Night	SF_Day	ΔSF	ΔSF_{corr} N/D
TWP-1	Sept. 2002 to May 2003	82.6 ± 5.3	5.24 ± 0.63	0.998 ± 0.018 (52)	1.058 ± 0.029 (52)	0.060	−0.004/−0.003
SGP-1	Oct. 2002 to May 2003	59.7 ± 20.5	1.63 ± 0.97	0.996 ± 0.024 (68)	1.057 ± 0.027 (73)	0.064	0.006/0.010
TWP-2	Sept. 2003 to March 2004	80.7 ± 6.4	4.94 ± 0.90	0.993 ± 0.028 (93)	1.055 ± 0.037 (77)	0.061	−0.002/0.000
SGP-2	Sept. 2003 to Feb. 2004	55.6 ± 21.8	1.42 ± 0.89	0.972 ± 0.033 (59)	1.056 ± 0.047 (61)	0.083	0.008/0.017
TWP-3	Apr. 2004 to Sept. 2004	82.0 ± 5.7	5.13 ± 0.71	0.974 ± 0.031 (67)	1.046 ± 0.045 (91)	0.070	−0.005/−0.003
SGP-3	Apr. 2004 to Aug. 2004	62.3 ± 19.0	2.77 ± 1.08	0.953 ± 0.031 (64)	1.037 ± 0.039 (72)	0.082	0.004/0.007

[a]Parameters shown are: mean and standard deviation of the RH below 1 km altitude (RH_{surf}); mean and standard deviation of the column-integrated PWV measured by the radiosondes for all soundings that reached at least 12 km altitude; mean and standard deviation of the MWR Scale Factor (SF), defined as the ratio of the MWR PWV to the radiosonde PWV, separated by day vs night, with number of qualifying soundings given in parentheses; difference between the mean SF for day (D) and night (N) soundings (ΔSF); and difference in the SF attributable to the radiosonde corrections (ΔSF_{corr}), separated by day vs night. The time lag and AWEX calibration corrections were applied to the RS90 data before calculating the scale factors. The MWR PWV is taken as the average PWV during a 40-min window centered on the radiosonde launchtime. Cases were eliminated if the MWR-measured liquid water path (LWP) exceeded 0.02 cm, because substantial liquid water may introduce a bias in the PWV retrieval.

[b]Mean RH in upper troposphere is near ice saturation at TWP; much less at SGP.

solar heating of the RH sensor causes the measured air in direct contact with the sensor polymer to be warmer than the ambient air, and therefore the RH of the measured air is lower than the ambient RH, resulting in a systematic dry bias in daytime RH measurements.

[60] The MWR scale factor (SF) was calculated for each of the RS90 validation soundings, and the mean and standard deviation of the SF partitioned by day versus night, plus other summary water vapor characteristics, are given for each of the six RS90 validation data sets in Table 4. The daytime RS90 measurements are on average 6−8% drier than the nighttime RS90 measurements relative to the MWR, which is double the 3−4% mean diurnal bias that *Turner et al.* [2003] observed in ARM RS80-H measurements. The greater susceptibility of the RS90 sensor to solar radiation probably results from the absence of a protective cap over the RS90 RH sensor, which on the RS80-H protects the sensor from solar radiation and precipitation (at the expense of decreased ventilation and sensor time response). The 6−8% daytime dry bias as determined from the PWV represents the solar radiative effect in the LT (where most of the PWV resides), and the daytime dry bias is probably greater in the UT. The data sets in Table 4 labeled TWP-1, SGP-1, and TWP-2 show near-perfect mean agreement with the MWR at night (<1% difference), suggesting that the RS90 calibration is on average quite accurate in the LT, in agreement with the CFH/RS92 comparisons. However, the nighttime mean SF for the data sets labeled SGP-2, SGP-3, and TWP-3 is 2−4% lower (moister) than the previous data sets, but with no change in the ~2−4% variability. Given that there were no changes to the MWR calibration or retrieval algorithm during this timeframe, a moistening of the RS90 calibration for conditions in the LT is indicated, which could be caused by either factory calibration changes or by drift in the Vaisala calibration reference, either of which would lead to a batch-dependent radiosonde accuracy. The implication that the RS90 calibration became less accurate with time rather than more accurate is troubling and difficult to accept, and so the meaning of this observation remains in question.

[61] The dependence of the solar radiation error on RH, T, and other factors such as cloud cover, altitude, and solar angle is not well known, but is unlikely to be well represented by a constant factor as assumed by the MWR-

scaling procedure. While MWR scaling may improve the accuracy of radiosonde measurements in terms of PWV, it may or may not improve the RH accuracy in various parts of the profile, especially parts that contribute little to the PWV, including dry layers and the middle and upper troposphere. Unlike the AIRS Best Estimate validation soundings produced to date, in this study the MWR SF was calculated after first applying the time lag and AWEX empirical calibration corrections, so the scaling addresses residual bias error, which includes solar radiation error and the random production variability of individual radiosondes. However, the impact of the time lag and empirical calibration corrections on the mean SF is small (final column in Table 4), because these corrections are relatively small in the LT.

6.2. Impact of Corrections on AIRS RS90 Data

[62] The magnitude of the time lag correction, AWEX empirical calibration correction, and MWR scaling for the RS90 AIRS validation profiles is shown in Figure 11 for the TWP-2 data set (top row) and the SGP-2 data set (bottom row). The results for the set-1 and set-3 data sets are similar to Figure 11 for the same site. However, Figure 11 shows that the magnitude of the corrections differs substantially between the TWP and SGP sites, which is attributed to the generally different characteristics of the RH and T profiles that typify each site (see Table 4).

[63] The time lag correction (Figures 11a and 11e) has no effect in the LT and MT, because RS90 sensors respond quickly at these temperatures. The effect of the time lag correction at TWP is to moisten the UT by ~5% and dry the LS by ~25% on average, thereby sharpening the troposphere-stratosphere transition layer. In contrast, the average profile at SGP in the tropopause region is dried by ~7%. Recalling that the time lag correction changes the shape of the profile and is sensitive to the local humidity gradient, any nonzero mean time lag effect for a data set indicates a climatological trend in the humidity gradient, such as the decrease in RH that consistently occurs above the tropopause. The moistening in the UT at TWP is caused by the consistent presence of a moist layer (and associated positive humidity gradient) near the tropopause, often producing a cirrus layer. However, the primary impact of the time lag correction in the UT is not necessarily its mean value, but
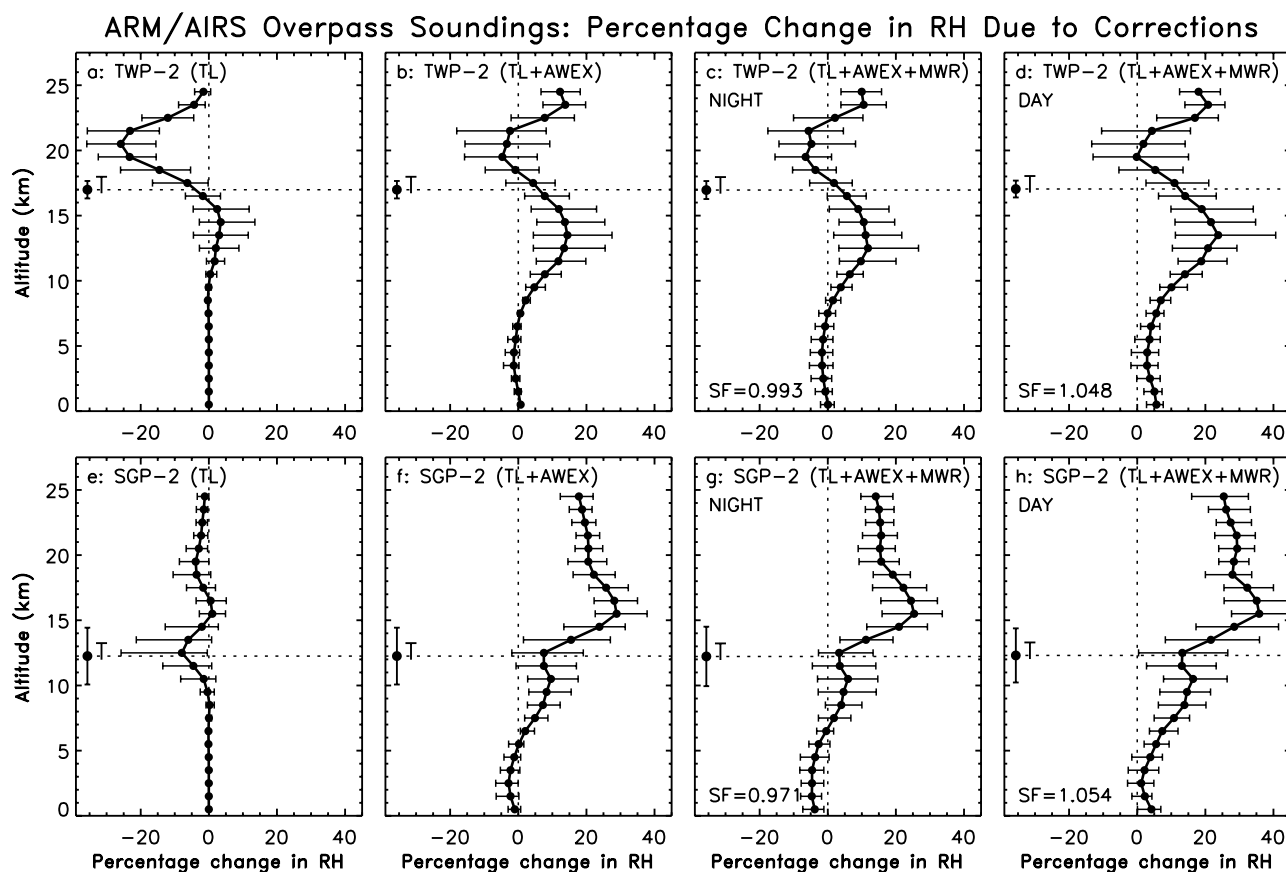
ARM/AIRS Overpass Soundings: Percentage Change in RH Due to Corrections

Figure 11. Mean (dots) and variability (bars) of the percentage change in RH resulting from successive corrections applied to the AIRS "phase 2" RS90 validation soundings from the (a–d) ARM TWP site and (e–h) SGP site, shown as a function of altitude in 1 km increments. The measure of variability is the 68th percentile above and below the mean, which is analogous to the standard deviation but more appropriate for an asymmetric distribution. Figures 11a and 11e show the magnitude of the time lag correction; Figures 11b and 11f also include the AWEX empirical calibration correction; the remaining panels also include MWR scaling, partitioned by nighttime (Figures 11c and 11g) versus daytime (Figures 11d and 11h). The MWR-scaled panels also show the mean scale factor (SF) for the data set. The horizontal dashed line indicates the mean tropopause altitude, and the vertical bar is its standard deviation. Additional characteristics of these data sets are given in Table 4.

rather its effect on individual profiles, which can vary substantially as indicated by the 68th percentile bars in Figure 11. By definition, 16% of the soundings received a time lag correction whose magnitude is beyond the ends of the bars, or >15% moistening in the UT at TWP. When the AWEX calibration correction is also applied (Figures 11b and 11f), then the AIRS validation profiles are moistened in the UT by an average of ~15% at TWP and ~10% at SGP, with 16% of soundings moistened by >25% at TWP and >17% at SGP. The AWEX calibration correction dries the profiles in the LT by 1–2% at TWP and 2–3% at SGP on average. The greater impact of the AWEX calibration correction at SGP, especially the greater range of variability in the LT and MT, results from the greater range of RH conditions that characterize the SGP environment compared to the TWP environment (see parameter "RH$_{surf}$" in Table 4). When the MWR scale factor is then applied to the data (Figures 11c, 11d, 11g, and 11h), its effect is to shift the mean correction profile left or right by an amount given by the mean scale factor.

[64] The diurnal bias in RS90 measurements, and the dependence of the RS90 accuracy on RH and T, are important considerations for AIRS validation, such as when AIRS-observed radiances are compared to calculated radiances that use the radiosonde profiles to specify the atmospheric state ("obs minus calc" analysis). Differences in obs-calc results at different locations, or diurnally at the same location, may be due in large part to the daytime RS90 dry bias and the dependence of the RS90 accuracy on the RH and T profiles typical of a given location, especially if the RS90 corrections are not applied.

7. Summary and Conclusions

[65] This study has yielded detailed estimates of radiosonde water vapor measurement accuracy for six operational radiosonde types launched during AWEX: Vaisala RS80-H, RS90, and RS92; Modem GL98; Sippican Mark IIa; and the Snow White chilled mirror hygrometer. The accuracy estimates (mean and variability) were derived by comparing

in situ radiosonde measurements with simultaneous measurements on the same balloon by a reference-quality sensor of known absolute accuracy, the CU CFH. This study also evaluated the impact of a correction for sensor time lag error on Vaisala radiosonde measurements, then the corrected data were used to derive and validate a new empirical correction for uncertainty in the Vaisala calibration. Although most radiosondes are calibrated in terms of RH rather than an absolute water vapor quantity such as mixing ratio, this study reports radiosonde accuracy in absolute terms as a percentage of the measured RH value, because it is the uncertainty in absolute water vapor amount that is the most relevant quantity for assessing observational error in most atmospheric applications. Radiosonde measurements are inherently less accurate in an absolute sense for dry conditions than for moist conditions (e.g., a 2% RH bias error is increasingly significant in an absolute sense as the RH decreases).

[66] A basic conclusion from this study is that there is no simple answer to the question "how accurate are water vapor measurements from a given radiosonde type?" Radiosonde accuracy varies substantially as a function of RH and T, and different radiosonde types have different strengths and weaknesses in different realms of RH and T space. Furthermore, the accuracy of RS80-H, Modem, and SW radiosondes may be substantially degraded by clouds, particularly thick ice clouds. Reliable use of data from these radiosonde types requires quality control to identify cloud influences. Only the Vaisala RS90 and RS92 radiosondes make reliable measurements within and above thick ice clouds, because of their alternately heated dual sensor design. Several other general characteristics of radiosonde performance were observed in this and previous studies: The SW is incapable of measuring very dry conditions (<6% RH), and substantial errors in the lower troposphere are possible if the proper phase of condensate on the mirror cannot be determined; the Sippican radiosonde cannot reliably measure dry conditions (RH < 20%), and the Sippican sensor becomes unresponsive at a temperature level that varies between -20 and $-50°C$; and the RS80-H and Modem radiosondes have substantial time lag error in the UT (although the RS80-H time lag error can be corrected).

[67] This study quantified the mean accuracy and variability of radiosonde water vapor measurements relative to the CFH as a function of RH and T (Figures 4, 5, 8, and 9 and Table 3), and then investigated the impact of the time lag and AWEX empirical calibration corrections on the accuracy and variability of Vaisala measurements. The absolute accuracy of the CFH reference sensor was shown from both observational and instrumental considerations to be about 3% in the LT and about 6% in the UT. Overall, the most accurate operational radiosonde tested is the Vaisala RS92 (and RS90), whose mean percentage accuracy relative to the CFH is <5% for most conditions in the LT, and <10% in the MT and UT. The corrections improve the RS92 mean accuracy relative to the CFH to <1% in the LT, <2% in the MT, and <3% in the UT, and the time lag correction substantially reduces the variability in the UT. Only the RS92 and RS90 are sufficiently accurate for AIRS validation throughout the troposphere, especially if the corrections are applied. The corrections also substantially improve the

RS80-H accuracy, such that corrected RS80-H data are marginally suitable for AIRS validation if the data are screened for the sensor-icing effect of clouds. The broad community would benefit from the operational application of the corrections to NWS RS80-H data. In contrast, the Sippican and Modem radiosondes are only reasonably accurate under relatively warm and moist conditions, and measurements from these radiosondes are generally not suitable for research purposes under cold or dry conditions.

[68] The quantitative accuracy assessment given in this paper applies only to nighttime radiosonde measurements, when solar radiation error is not an issue. The impact of solar radiation error on AIRS RS90 validation measurements was investigated by comparing to simultaneous retrievals of PWV from an ARM microwave radiometer, with the result that solar radiation produces a dry bias of 6–8% in RS90 measurements in the LT (probably more in the UT). Further investigation of the MWR-scaling technique as a means of correcting the solar radiation dry bias is warranted, particularly with regard to the dependence of solar radiation error on RH and T. Correction of the nonsolar component of RS90 measurement error (the time lag and AWEX empirical calibration corrections) leads to a moistening of the AIRS RS90 validation profiles in the UT by a mean of ∼15% at TWP and ∼10% at SGP, and ∼1–3% drying in the LT. Both corrections are sensitive to the individual profiles measured, which leads to considerable variability in the magnitude of the corrections between profiles, and the mean correction magnitudes give only a rough indication of the impact of the corrections on individual profiles.

Appendix A: Saturation Vapor Pressure Formulations

[69] The saturation vapor pressure (SVP) over liquid water (e_w) and over ice (e_i) are parameters needed to convert between absolute water vapor quantities such as mixing ratio and the water vapor quantity typically reported by radiosondes, RH with respect to liquid water: $RH \equiv RH_{water} = e/e_w(T) \times 100\%$, where e is the ambient water vapor pressure. Vaisala (and probably other) radiosonde calibrations are developed in part by measuring the sensor response (a capacitance) under controlled conditions below $0°C$ that are known to be saturated with respect to ice [M01], in which case $RH = e_i(T)/e_w(T) \times 100\%$. Therefore the value of RH that corresponds to ice saturation (or to any other conditions) is fundamentally tied to the choice of expressions for e_i and e_w used in the radiosonde calibration. Similarly, the conversion of chilled mirror hygrometer (e.g., CFH and SW) frostpoint and air temperature measurements to RH with respect to liquid water, $RH = e_i(T_f)/e_w(T) \times 100\%$, contains uncertainty related to the choice of SVP formulations, in addition to the instrumental uncertainty in the two temperature measurements.

[70] Most SVP formulations are derived from integration of the Clapyron equation. This study uses the *Wexler* [1976] formulation for e_w, and the *Hyland and Wexler* [1983] formulation for e_i, mostly for consistency with Vaisala's use of these expressions in their radiosonde calibrations, but also because *Elliott and Gaffen* [1991] recommend these formulations over the commonly used older expressions by

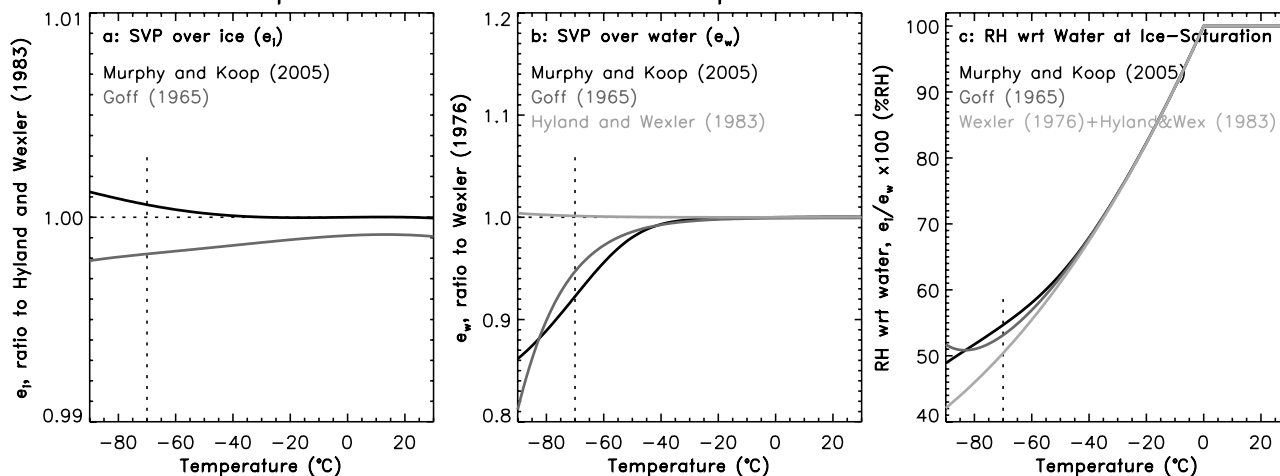## Comparison of Saturation Vapor Pressure Formulations

Figure A1. Comparison of formulations for the saturation vapor pressure (SVP) over ice (e_i) and over liquid water (e_w) as a function of temperature, given as a ratio to the SVP formulations used by Vaisala in their radiosonde calibrations. (a) e_i relative to *Hyland and Wexler* [1983]. (b) e_w relative to *Wexler* [1976]. (c) RH with respect to liquid water that corresponds to saturation over ice, $e_i/e_w \times 100\%$. The equations for these and other SVP formulations are summarized by *Murphy and Koop* [2005]. Dashed reference lines are shown at a ratio of 1.0, and at the $-70°$C low temperature limit of the AWEX data set.

Goff [1965] and *Goff and Gratch* [1946] because of the more recent and accurate measurement of certain fundamental constants. *Murphy and Koop* [2005] reviewed and assessed several SVP formulations and their consistency with experimental data, and they developed new SVP formulations. Figure A1 shows the ratio of several SVP formulations with the expressions used in this paper and by Vaisala. All common formulations for e_i, including several not shown in Figure A1a, agree with each other and with experimental data to better than 0.5% over the entire atmospheric temperature range, and therefore in a practical sense it is nearly irrelevant which e_i formulation is chosen. However, the *Wexler* [1976] formulation for e_w (and the nearly identical *Hyland and Wexler* [1983] expression) are considerably higher than most other formulations at low temperatures (Figure A1b), and are outside uncertainty limits that are estimated by *Murphy and Koop* [2005] on the basis of limited experimental data and plausibility arguments. Measurements of e_w over supercooled water are sparse and difficult to make, and are in fact impossible below the homogeneous nucleation temperature of about $-37°$C because even tiny droplets freeze spontaneously. By historical convention, radiosonde RH has always been reported with respect to liquid water, even at low temperatures where e_w used in calibrations has little physical meaning and is not experimentally verifiable. Figure A1c shows the range of RH values corresponding to conditions of ice saturation, $RH_i = e_i(T)/e_w(T) \times 100\%$, that result from different choices of SVP formulations (mainly e_w). The difference between the SVP formulations used by Vaisala and throughout this paper with the formulation developed by *Murphy and Koop* [2005] is <1.2% RH (<2% in absolute terms) at temperatures above $-50°$C, 4.2% RH (8.5%) at $-70°$C (the maximum effect for the AWEX data set), and 6.8% RH (16.2%) at a tropical tropopause temperature of $-90°$C. These values can be viewed as the amount by which Vaisala radiosonde measurements would be moist-

ened if the SVP formulations of *Murphy and Koop* [2005] were adopted. It is important to note that this source of uncertainty does not affect the Vaisala/CFH intercomparisons in this paper, because the same Vaisala formulations are used to convert the CFH measurements to units of RH. However, if AIRS measurements, or Raman lidar measurements, or frostpoint measurements, or models, or even other RH measurements assume a different e_w formulation when comparing to or using radiosonde measurements, then substantial uncertainty is introduced at low temperatures. Note that it is straightforward to convert RH measurements to values based on a different e_w formulation: $RH_{new} = RH_{old} \times (e_{w,old}/e_{w,new})$.

[71] The optimal solution to this SVP uncertainty is simple in principle, yet complicated in practice. Below 0°C, it is more physically meaningful to report RH with respect to ice, $RH_{ice} = e/e_i(T) \times 100\%$ (or, from frostpoint measurements, $RH_{ice} = e_i(T_f)/e_i(T) \times 100\%$), such that RH = 100% at ice saturation. Very little (<0.5%) uncertainty is related to the choice of e_i formulation, since they all agree closely with each other and with experimental data. The practical difficulty is overcoming the historical inertia of using a single RH parameter for temperatures both above and below 0°C. Simply put, changing the definition of RH in order to increase the consistency and accuracy of various water vapor measurements would likely wreak havoc on a wide range of modeling and other research that uses radiosonde data as input.

[72] We suggest that radiosonde manufacturers begin also reporting RH with respect to ice to eliminate uncertainty attributable to the choice of e_w formulation. As a practical interim solution, we suggest that radiosonde manufacturers adopt two changes: First, continue reporting RH with respect to liquid water as usual, but disclose to users in the data file header information the e_w formulation that was assumed in the calibration, thereby allowing users to convert to a different e_w formulation or to eliminate the

dependence on e_w altogether; and second, provide an additional RH parameter that is RH with respect to ice. This approach allows continuity with the current standard of radiosonde RH reporting while also allowing accurate conversion to other SVP formulations or to absolute water vapor quantities, and providing the RH_{ice} parameter gives the most accurate radiosonde measurements while raising awareness in the community that this detail and its ramifications are often buried deeply in models and other computer code.

Appendix B: TD Correction for RS90 Radiosondes Produced Before 25 June 2001

[73] The Vaisala radiosonde calibration procedure converts the fundamental capacitance measurements to RH using curve fits that depend on RH and T. Prior to 25 June 2001, Vaisala used a preliminary RS90 temperature dependence (TD) calibration model until a more accurate TD model was developed and implemented. The RS90 calibration change involved the coefficients on the radiosonde calibration tape, not changes to the ground station software, so the improved calibration accuracy was transparent but unknown to the user. Vaisala derived the following TD correction for RS90 radiosondes produced before 25 June 2001, which converts the original RS90 TD calibration to the current and more accurate RS90 TD calibration (A. Paukkunen, personal communication, 2004): $RH_{corrected} = RH_{measured} \times F(T)$, where $F(T) = a_0 + a_1 T + a_2 T^2 + a_3 T^3 + a_4 T^4$, with coefficients $a_0 = 0.9597$, $a_1 = 1.419e - 3$, $a_2 = 5.309e - 6$, $a_3 = -4.777e - 7$, and $a_4 = 7.427e - 9$. Data from RS90 radiosondes produced before 25 June 2001 contain a moist bias ($F(T) < 1.0$) in the temperature range $-48°C$ to $+30°C$ (minimum $F(T) = 0.94$ at $-20°C$), and a dry bias ($F(T) > 1.0$) outside this temperature range (i.e., in the UT), where the correction factor $F(T)$ is 1.08 at $-60°C$, 1.22 at $-70°C$, 1.43 at $-80°C$, and 1.7 at $-90°C$. The radiosonde calibration date can be determined from the serial number (see Appendix of M04).

References

Blackmore, W. H., and B. Taubvurtzel (1999), Environmental chamber tests of NWS radiosonde relative humidity sensors, paper presented at 15th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, Am. Meteorol. Soc., Dallas, Tex.

Elliott, W. P., and D. J. Gaffen (1991), On the utility of radiosonde humidity archives for climate studies, *Bull. Am. Meteorol. Soc.*, 72, 1507–1520.

Ferrare, R. A., et al. (2004), Characterization of upper-troposphere water vapor measurements during AFWEX using LASE, *J. Atmos. Oceanic Technol.*, 21, 1790–1808.

Fetzer, E., et al. (2003), AIRS/AMSU/HSB validation, *IEEE Trans. Geosci. Remote Sens.*, 41, 418–431.

Goff, J. A. (1965), Saturation pressure of water on the new Kelvin scale, in *Humidity and Moisture: Measurement and Control in Science and Industry*, vol. 3, edited by A. Wexler, p. 289, Reinhold, New York.

Goff, J. A., and S. Gratch (1946), Low-pressure properties of water from –160 to 212 F, in Transactions of the American society of heating and ventilating engineers, pp. 95–122, paper presented at 52nd Annual Meeting, Am. Soc. of Heat. and Ventilating Eng., New York.

Heymsfield, A. J., and L. M. Miloshevich (1993), Homogeneous ice nucleation and supercooled liquid water in orographic wave clouds, *J. Atmos. Sci.*, 50, 2335–2353.

Hirvensalo, J., J. Wahrn, and H. Jauhiainen (2002), New Vaisala RS92 GPS radiosonde offers high level of performance and GPS wind data availability, paper presented at presented at 12th Symposium on Meteorological Observations and Instrumentation, Am. Meteorol. Soc., Long Beach, Calif.

Hyland, R. W., and A. Wexler (1983), Formulations for the thermodynamic properties of the saturated phases of H_2O from 173.15 K to 473.15 K, *ASHRAE Trans.*, 2A, 500–519.

Ivanov, A., A. Kats, N. Kurnosenko, J. Nash, and N. Zaitseva (1991), WMO International Radiosonde Intercomparison—Phase III, *WMO Instrum. Obs. Methods Rep. 40*, 135 pp., World Meteorol. Organ., Geneva, Switzerland.

Liljegren, J. C., E. E. Clothiaux, G. G. Mace, S. Kato, and X. Dong (2001), A new retrieval for cloud liquid water path using a ground-based microwave radiometer and measurements of cloud temperature, *J. Geophys. Res.*, 106(D13), 14,485–14,500.

Liljegren, J. C., S. A. Boukabara, K. Cady-Pereira, and S. A. Clough (2005), The effect of the half-width 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.

Luers, J. K. (1997), Temperature error of the Vaisala RS90 radiosonde, *J. Atmos. Oceanic Technol.*, 14, 1520–1532.

Luers, J. K., and R. E. Eskridge (1995), Temperature corrections for the VIZ and Vaisala radiosondes, *J. Appl. Meteorol.*, 34, 1241–1253.

Miloshevich, L. M., H. Vömel, A. Paukkunen, A. J. Heymsfield, and S. J. Oltmans (2001), Characterization and correction of relative humidity measurements from Vaisala RS80-A radiosondes at cold temperatures, *J. Atmos. Oceanic Technol.*, 18, 135–156.

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.

Murphy, D. M., and T. Koop (2005), Review of the vapour pressures of ice and supercooled water for atmospheric applications, *Q. J. R. Meteorol. Soc.*, 131, 1539–1565.

Paukkunen, A., V. Antikainen, and H. Jauhiainen (2001), Accuracy and performance of the new Vaisala RS90 radiosonde in operational use, paper presented at 11th Symposium on Meteorological Observations and Instrumentation, Am. Meteorol. Soc., Albuquerque, N. M., 14–18 Jan.

Sapucci, L. F., L. A. T. Machado, R. B. Da Silveira, G. Fisch, and J. F. G. Monico (2005), Analysis of relative humidity sensors at the WMO radiosonde intercomparison experiment in Brazil, *J. Atmos. Oceanic Technol.*, 22, 664–678.

Schmidlin, F. J. (1998), Report of the WMO radiosonde relative humidity sensor intercomparison: Phase II, 8–26 Sept 1995, *WMO Instrum. Obs. Methods Rep., 70*, 21 pp., World Meteorol. Organ., Geneva, Switzerland.

Soden, B. J., and J. R. Lanzante (1996), An assessment of satellite and radiosonde climatologies of upper-tropospheric water vapor, *J. Clim.*, 9, 1235–1250.

Soden, B. J., D. D. Turner, B. M. Lesht, and L. M. Miloshevich (2004), An analysis of satellite, radiosonde, and lidar observations of upper tropospheric water vapor from the Atmospheric Radiation Measurement program, *J. Geophys. Res.*, 109, D04105, doi:10.1029/2003JD003828.

Tobin, D. C., H. E. Revercomb, R. O. Knuteson, W. F. Feltz, B. M. Lesht, T. Cress, L. L. Strow, S. E. Hannon, and E. J. Fetzer (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., 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, 117–132.

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, 1560–1567.

Wang, J., H. L. Cole, D. J. Carlson, E. R. Miller, K. Beierle, A. Paukkunen, and T. K. Laine (2002), Corrections of humidity measurement errors from the Vaisala RS80 radiosonde—Application to TOGA COARE data, *J. Atmos. Oceanic Technol.*, 19, 981–1002.

Wexler, A. (1976), Vapor pressure formulation for water in range 0 to 100°C: A revision, J, *Res. Natl. Bur. Stand. U.S., Sect. A*, 80, 775–785.

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.*, doi:10.1029/2005JD006429, in press.

Yagi, S., A. Mita, and N. Inoue (1996), WMO international radiosonde comparison: Phase IV—Tsukuba, Japan, 15 Feb.–12 Mar. 1993, final report, *WMO Instrum. Obs. Methods Rep., 59*, 130 pp., World Meteorol. Organ., Geneva, Switzerland.

————————————

B. M. Lesht, Department of Energy/Argonne National Laboratory, Argonne, IL 60439, USA.

L. M. Miloshevich, National Center for Atmospheric Research, Boulder, CO 80307, USA. (milo@ucar.edu)

F. Russo, Department of Physics, University of Maryland-Baltimore County, Baltimore, MD 21250, USA.

F. J. Schmidlin and D. N. Whiteman, NASA/Goddard Space Flight Center, Greenbelt, MD 20771, USA.

H. Vömel, Cooperative Institute for Research in the Environmental Sciences, University of Colorado, Boulder, CO 80309, USA.