

**Accuracy Assessment and Correction of Vaisala RS92 Radiosonde
Water Vapor Measurements**

by

Larry M. Miloshevich
National Center for Atmospheric Research*
Boulder, Colorado

Holger Vömel
University of Colorado (CIRES)
Boulder, Colorado

David N. Whiteman
NASA/GSFC
Greenbelt, Maryland

Thierry Leblanc
NASA/JPL
Wrightwood, California

*The National Center for Atmospheric Research (NCAR) is sponsored by the National Science Foundation.

Abstract

Relative humidity (RH) measurements from Vaisala RS92 radiosondes are widely used in both research and operational applications, although the measurement accuracy is not well characterized as a function of its known dependences on height, RH, and time of day (or solar altitude angle). This study characterizes RS92 mean bias error as a function of its dependences by comparing simultaneous measurements from RS92 radiosondes and from three reference instruments of known accuracy. The cryogenic frostpoint hygrometer (CFH) gives the RS92 accuracy above the 700 mb level; the ARM microwave radiometer gives the RS92 accuracy in the lower troposphere; and the ARM SurTHref system gives the RS92 accuracy at the surface using 6 RH probes with NIST-traceable calibrations. These RS92 assessments are combined using the principle of Consensus Referencing to yield a detailed estimate of RS92 accuracy from the surface to the lowermost stratosphere. An empirical bias correction is derived to remove the mean bias error, yielding corrected RS92 measurements whose mean accuracy is estimated to be $\pm 3\%$ of the measured RH value for nighttime soundings and $\pm 4\%$ for daytime soundings, plus an RH offset uncertainty of $\pm 0.5\%RH$ that is significant for dry conditions. The accuracy of individual RS92 soundings is further characterized by the $1-\sigma$ “production variability,” estimated to be $\pm 1.5\%$ of the measured RH value. The daytime bias correction should not be applied to cloudy daytime soundings, because clouds affect the solar radiation error in a complicated and uncharacterized way.

1. Introduction

Atmospheric water vapor measurements are used in a wide variety of both operational and research applications, including as input to forecast models and radiative transfer calculations, for validation of ground-based and satellite remote sensor retrievals, and for development of water vapor and cloud parameterizations, among other applications. The high vertical resolution of radiosonde measurements is well suited to these measurement needs, if not for their substantial inaccuracy under certain atmospheric conditions, especially in the upper troposphere (UT) and lower stratosphere (LS). As with all measurements, their scientific value is tied to estimates of uncertainty in the measurements, which must be known if uncertainty in subsequent results is to be estimated. Unfortunately, the accuracy of radiosonde relative humidity (RH) measurements differs between measurement technologies, between radiosonde manufacturers and models, and even with time for a given model due to hardware, manufacturing, or calibration changes. Furthermore, the accuracy of radiosonde RH measurements has been shown to vary with height, RH, and time of day (or solar altitude angle). The aim of this paper is to characterize the accuracy of Vaisala RS92 radiosonde water vapor measurements as a function of its dependences, and then develop and evaluate an empirical correction that removes the mean bias error.

Several methods of characterizing and improving the accuracy of RS92 RH measurements have been developed, although these methods generally address only a subset of the main sources of measurement error. These sources include calibration error that reflects the accuracy of the Vaisala calibration model and calibration references, solar radiation error (SRE) caused by solar heating of the RH sensor, and time-lag error caused by slow sensor response at low temperatures. *Vömel et al.* [2007] used dual soundings of RS92 and the cryogenic frostpoint hygrometer (CFH) to show that daytime RS92 measurements suffer from SRE that increases with height from 9% near the surface to 50% at the tropical tropopause for high solar altitude angles ($\alpha > 60^\circ$), and they derived a height (pressure) dependent correction that removes this mean bias. *Miloshevich et al.* [2006] (hereafter “M06”) compared ARM microwave radiometer (MWR) and RS90 radiosonde measurements and found that the SRE is 6-8% in terms of precipitable water vapor (PW), where this bias represents an average over all solar altitude

angles and RH conditions in the lower troposphere (LT). M06 also used dual RS92/CFH soundings to characterize the mean bias error for nighttime RS92 soundings (i.e., zero SRE), and derived a correction that is dependent on height (temperature) and on RH. *Cady-Pereira et al.* [2008] determined the dependence of SRE on solar altitude angle by comparing ARM MWR PW measurements to column-integrated RS90 and RS92 PW measurements, and derived a correction that removes the mean bias in the LT as a function of the solar altitude angle. *Miloshevich et al.* [2004] (hereafter “M04”) used laboratory measurements of the sensor time-constant as a function of temperature to derive a correction for sensor time-lag error caused by slow sensor response at low temperatures, and *Vömel et al.* [2007] showed that addressing time-lag error markedly improves the agreement between RS92 and CFH in the UT.

This study will evaluate the accuracy of RS92 water vapor measurements using the principle of Consensus Referencing [*Fitzgibbon, 2008; Facundo and Fitzgibbon, 2007*], whereby RS92 measurements are compared to measurements from different reference-quality instruments for the conditions under which each instrument performs well and its accuracy is known. Simultaneous measurements from the RS92 and 3 reference instruments --- CFH, MWR, and calibrated RH probes --- are compared in the next section. The comparisons are synthesized in section 3 to produce an estimate of the RS92 accuracy as a function of height, RH, and solar altitude angle. An empirical correction that removes the RS92 mean bias relative to the reference instruments is derived in section 4, and the accuracy of the corrected data are evaluated.

2. Instrumentation and Data

a. RS92

Vaisala radiosondes use thin-film capacitance RH sensors, where a 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 sensor calibration relates the measured capacitance to the RH with respect to liquid water at +25°C, and then compensates for temperature using a sensor temperature-dependence model.

The calibration is optionally adjusted during the ground check (GC) procedure prior to launch, where the RH sensor is sealed in a container of desiccant assumed to be at 0.0% RH, and the radiosonde measurement under these conditions is used as an RH offset correction in the calibration. Unlike earlier RS80 radiosondes, RS90 and RS92 radiosondes use dual RH sensors that are alternately heated while the other sensor measures the ambient RH, which eliminates the problem of sensor icing in liquid water clouds or ice-supersaturated conditions. The operational principles of Vaisala radiosonde RH sensors are described further by M06 and references therein. The following sources of RS92 measurement error are considered in this study:

- *Mean calibration bias* reflects the absolute accuracy of the Vaisala calibration references and their variation with time, which *Paukkunen et al.* [2001] estimates as 0.6-2% RH over the range 0-90% RH. Calibration bias also arises from inaccuracy in the Vaisala calibration model, including curve-fit error that is inherently a function of RH and T.
- *Production variability* is the random sensor-to-sensor variability relative to the mean calibration accuracy that reflects such things as manufacturing variability and inhomogeneous conditions within the calibration chamber. A complete accuracy specification requires not only an estimate of the mean (bias) uncertainty but also a measure of the variability, such as the standard deviation of differences from the mean for a batch of sensors.
- *Time-lag error* arises from slow sensor response to changing RH conditions at low temperatures, which has the effect of “smoothing” the RH profile in the UT and LS. This study uses the time-lag correction described by M04, which is a numerical inversion algorithm that recovers the “true” shape of the RH profile from the measured RH and T profiles based on measurements of the sensor time-constant and its dependence on temperature.
- *Solar radiation error* is a dry bias in daytime measurements caused by solar heating of the RH sensor [*Vömel et al.*, 2007]. The magnitude of the SRE depends on the incident solar flux and is therefore a function of numerous factors including the solar altitude angle (α), height (or pressure), the angle between the sun and the sensor normal, the cloud optical depth along a sun-sensor line, and the transmissivity of the airmass. The net heating of the RH sensor is also

affected by the ventilation rate (ascent rate), and by the thermal and radiative characteristics of the sensor. Although the SRE could be reduced by increasing the reflectivity of the sensor or adding some type of sensor shield, the underlying cause of the error is that the proper temperature is not used in the data processing. The ambient air temperature that is accurately measured by the temperature sensor is used in the RH calibration equation, but the appropriate temperature is actually the (unknown) temperature of the sensor polymer itself. Measuring the temperature of the RH sensor would also address other minor measurement errors that occur when the temperature gradient changes abruptly (e.g., at the tropopause or at temperature inversions), and in the first 100 m or so of flight. These errors result from the different thermal time constants of the temperature and (relatively massive) RH sensors, because the measured air temperature doesn't accurately represent the temperature of the RH sensor.

- *Ground check related uncertainty* arises from both “operator error” (improper procedures) and from the unrealistic assumption that the desiccant can truly create an environment of 0.0% RH. The mean GC correction for the 2006-2007 ARM RS92 dataset used in this study is -0.5% RH; however, it is easy to introduce error greater than this if the desiccant is not fresh or the chamber lid is not tightly closed, in which case the environment in the chamber is >0.0% RH and the GC correction introduces bias error. In addition, even fresh desiccant is unlikely to produce an environment of precisely 0.0% RH as assumed, leading to an RH offset error equal in magnitude to the true RH in the desiccant chamber. The best of a selection of desiccants tested by the Australian Bureau of Meteorology achieved a minimum of only 0.5% RH [Gorman, 2002].

- *Roundoff error* is the unnecessary loss of information when the RH measurements reported in the standard RS92 processed (“EDT”) data files are rounded to the nearest integer. Rounding introduces uncertainty in a given data point of $\pm 0.5\%$ RH, which is substantial for dry conditions (e.g., $\pm 10\%$ uncertainty at 5% RH). This 1% RH resolution also limits the ability of the time-lag correction to recover vertical structure in the profile, because that structure is degraded by the rounding. The RS92 data used in this study are instead from the Vaisala FLEDT (floating-point EDT) data files, which report the RH values with a resolution of 0.01% RH (FLEDT files are only available from the Vaisala DigiCora-III data system, version 3.12 and newer). The use of

high RH resolution RS92 data and other recommendations for maximizing the accuracy of RS92 soundings are given in the Appendix.

b. CFH

The CFH measures water vapor based on the chilled-mirror principle, where a small mirror is electrically heated against a cryogenic cold sink to maintain a constant layer of condensate that is optically detected. When the condensate layer is in equilibrium with the environment the mirror temperature is equal to the dewpoint temperature (T_d) if the condensate phase is liquid, or the frostpoint temperature (T_f) if the condensate is ice. To remove ambiguity as to the phase of water on the mirror, the CFH forces the liquid layer to freeze by briefly cooling the mirror to -40°C when T_d first reaches -15°C , thereby guaranteeing that T_f is measured thereafter. The mirror temperature is converted to RH with respect to liquid water using the air temperature measured by an attached RS80 radiosonde, and using the *Wexler* [1976] saturation vapor pressure formulation that is implicit in Vaisala's calibration procedure (see discussion in M06, Appendix A). A full description of the CFH design and measurement characteristics is given by *Vömel et al.* [2007a]. They estimate the uncertainty in CFH RH measurements to be 4% of the measured RH value near the surface, increasing with height to 9% at the tropopause. Daytime CFH measurements are not affected by SRE because the mirror assembly is protected from solar heating inside a well-ventilated sampling tube of low thermal mass, where the sampled air transits the inlet in <0.1 s.

It is critical to distinguish between percentage differences and RH differences in this paper, since differences will generally be given as a percentage of the measured value (%), but the unit of measurement for RH is also called %. The above uncertainty in CFH RH measurements is 4-9% of the measured RH value, not 4-9% RH. Consider a difference between CFH and RS92 measurements of 1% RH: this is a difference of 2% if RH=50%, and a difference of $>50\%$ for conditions drier than 2% RH. Most applications are concerned with percentage differences rather than RH differences, because most applications measure or convert to absolute measures of water vapor concentration like mixing ratio. Care is taken in this paper to clearly and consistently distinguish % from %RH.

This study uses simultaneous measurements from CFH and RS92 instruments launched on the same balloon during 4 NASA satellite and Raman lidar water vapor validation experiments. Two experiments called WAVES (Water Vapor Validation – Satellite/Sondes) were conducted at the Howard University (HU) Research Campus in Beltsville, Maryland in July and August of 2006 and 2007 [Whiteman *et al.*, 2008]. Two experiments called MOHAVE (Measurements of Humidity in the Atmosphere and Validation Experiments) were conducted at the JPL Table Mountain Facility in Wrightwood, California in October of 2006 and 2007 [Leblanc *et al.*, 2008]. The dataset consists of 33 nighttime dual soundings and 12 daytime dual soundings, although only 7 of the daytime soundings that are cloud free are used in this analysis. Nine of the CFH soundings from the first MOHAVE experiment had two RS92 radiosondes on each balloon, to investigate RS92 sensor-to-sensor variability. The standard deviation of the percentage difference between the RS92 pairs is about $\pm 1.5\%$ of the measurement for conditions above 10% RH, and $\pm 3\%$ for conditions below 10% RH, constant with height throughout the troposphere. These values are an estimate of the 1- σ random production variability for RS92 sensors, and they indicate good consistency between sensors.

Figure 1 shows example CFH and RS92 profiles from the summer 2006 WAVES experiment that illustrate the nature of the time-lag correction and the fundamental difference in the accuracy of daytime and nighttime RS92 measurements. The RS92 measurements above the -45°C level in Fig. 1a have been “smoothed” by time-lag error (black), yet the information about changes in the humidity gradient is present in the data and is recovered by the time-lag correction (red), as judged by comparing to the vertical structure and steepness of the troposphere-stratosphere transition that is measured by the CFH (purple). Removing the time-lag error reveals an RS92 bias relative to CFH that varies with height (or pressure) and with RH. The RS92 bias is much greater for daytime soundings than nighttime soundings due to the SRE, and the daytime and nighttime RS92 measurements will be characterized separately. Note that the time-lag effect in the UT is much smaller for the daytime example than the nighttime example. This is an artifact of the choice of soundings, where the temperature in the UT is much warmer for the daytime example as a result of the lower tropopause height, and the RH sensor responds more quickly at the warmer temperatures.

When RS92 and CFH measurements on the same balloon are compared quantitatively, we align the profiles based on time rather than altitude, because the clocks in the two radiosondes are more accurate and less variable than are the altitudes derived from separate pressure and temperature sensors. The percentage difference between the RS92 and CFH measurements for all nighttime dual soundings is shown in Figs. 2a-c as a function of altitude, and in Figs. 2d-f as a function of height (pressure rather than altitude) and partitioned into 7 RH intervals. The nighttime RS92 bias relative to CFH is shown for three circumstances: using the original RS92 FLEDT measurements (a and d); after applying the time-lag correction (b and e); and including only the data after the CFH forced freezing event that typically occurs at 2.5 to 4.5 km altitude (c and f). For the RH intervals in panels d-f, each dot in an upper curve is the mean percentage difference from CFH for all data in a P-RH bin, and the corresponding point in the lower curve is the variability in that bin (standard deviation of differences), offset to zero at the bottom of the panel for clarity.

The percentage difference between the CFH and original RS92 measurements contains much greater sensor-to-sensor variability in the UT/LS region than at lower levels (Fig. 2a, and lower curves in Fig. 2d below 300 mb). The time-lag correction recovers vertical structure in the RS92 profiles and reduces the variability in the UT/LS to about the same as at lower levels (Figs. 2b and 2e). Removing time-lag error also reveals that the RS92 mean bias in the UT/LS varies smoothly with altitude and has a clear RH-dependence. The very large mean moist bias above ~20 km occurs because the RS92 is insensitive to water vapor in the stratosphere and typically measures a nearly constant value of 1-2% RH, whereas the CFH is far more accurate under these conditions and typically measures RH values that decrease with height to $\ll 1\%$ RH at the top of the profile.

The RS92 mean bias relative to CFH varies smoothly with height from a moist bias of 9% at 4.5 km to a dry bias of 20% in the UT/LS region (Fig. 2b), but in the LT the mean bias changes slope at 4.5 km and then changes abruptly at 2.5 km. Altitude profiles of the RS92 bias can be misleading because the bias varies with RH and not just height, and the dataset contains a mixture of MOHAVE soundings that are generally drier and begin at a surface altitude of 2.3 km, and WAVES soundings that are generally moister and begin at a surface altitude of 0.1 km.

However, the change in character of the mean bias in the 2.5-4.5 km altitude range in Fig. 2b is also seen to occur for the separate RH intervals (Fig. 2e at about 700 mb), indicating that the abrupt change at 2.5 km is not an artifact of the mix of stations or the RH-dependence of the RS92 bias. The abrupt change also coincides with the CFH forced freezing event during WAVES. When the data prior to the forced freezing event are eliminated (Figs. 2c and 2f), then the mean bias above the forced freezing event varies smoothly with height and with RH, suggesting that this anomaly might be attributable to CFH behavior. Comparison of an ensemble of profiles measured by the CFH and by two Raman lidars during WAVES showed that the CFH measured moister in the mean than the lidars by about 7% up to the altitude of the forced freezing event, but above this level the CFH and Raman lidar agreed well. The CFH also measured 4-5% moister in the mean than the HU MWR in terms of PW, where the PW is dominated by water vapor in the lowest few km. These observations are not conclusive by themselves, but analysis of data from additional sensors presented later in this paper will show that the CFH measurements for some soundings during WAVES have an unexplained moist bias of about 5% in the LT that is related to the presence of liquid condensate on the mirror (i.e., prior to the forced freezing event at $T_d = -15^\circ\text{C}$). Therefore, this RS92 accuracy assessment will use the CFH measurements only above the forced freezing event, as shown in Figs. 2c and 2f. This CFH mean moist bias in the LT during WAVES is a conclusion of this study that will be further examined at the end of this section after all the data have been presented.

When the CFH measurements prior to the forced freezing event are excluded (Fig. 2f), a pattern emerges that may reflect underlying curve fit error in the Vaisala calibration. The curves exhibit an inflection point at about 400 mb, and all curves in the center portion of the RH range (4-36% RH) cross at the same point. The RH-dependence of the bias varies smoothly but differently on both sides of the inflection point: on the right at 700 mb the RS92 has a moist bias that decreases rapidly with increasing RH up to about 20% RH and then varies little with RH, while in the UT the RS92 has a dry bias of about 20% for conditions below 20% RH that decreases with increasing RH for moister conditions. A single curve represents the broad interval 48-99% RH, because subintervals all lie on top of this same curve. The standard deviation of the RS92 difference from CFH is generally 5% or less, except for the driest conditions where even a tiny RH offset error produces a large percentage difference.

The RS92 mean bias for nighttime measurements in Fig. 2f represents the RS92 mean calibration accuracy relative to CFH (above the 700 mb level), whereas the RS92 mean bias for the daytime dual soundings (Fig. 3) represents the sum of calibration bias plus solar radiation error for clear (“full-sun”) conditions and high solar altitude angles (62-70°). The dependence of the RS92 bias on solar altitude angle will be explored in the next section. The daytime dataset is much smaller than the nighttime dataset, and therefore contains greater statistical uncertainty. The regions of individual curves in Fig. 3c that depart substantially from the mean are dry layers, and illustrate the need to consider the RH-dependence of the RS92 bias and not just the height dependence. The RS92 daytime bias in Fig. 3f shows characteristics similar to the nighttime pattern, but modified by the SRE. As with the nighttime pattern, the daytime bias curves intersect, but at a pressure of 300 mb rather than 400 mb, and the curves for super-dry conditions (RH<3%) are offset from the rest of the RH range. One expects the SRE component of the bias to have a pressure dependence because the net sensor heating will increase with height as the cooling from ventilation decreases with decreasing pressure. This additional pressure-dependence from SRE modifies the pressure-dependence of the nighttime (calibration) bias, apparently shifting the crossover point to 300 mb and removing the upward curvature in the UT. The RS92 daytime dry bias in Fig. 3f increases with decreasing pressure from 8% at 700 mb for moist conditions (black curve) to 40-50% in the UT, and the standard deviation of the difference from CFH is ~5% in the troposphere, similar to the nighttime measurements.

c. ARM MWR

The performance of the RS92 in the lower troposphere is evaluated by comparing PW measurements from an ARM MWR with the column-integrated PW from colocated RS92 radiosondes at the ARM Southern Great Plains (SGP) site for the years 2006-2007 (except the period 4 August to 18 November 2007 when MWR data were unavailable). The Radiometrics 2-channel MWR (23.8 and 31.4 GHz) measures downwelling radiation that is converted to sky brightness temperature, from which the PW is retrieved. Historically, ARM has used a statistical PW retrieval algorithm called MWRLOS. Recently, *Turner et al.* [2007] developed and implemented at ARM a substantially more accurate MWR retrieval algorithm called MWRRET, which accounts for systematic error in the measurements and in the spectroscopy of the forward

radiative transfer model by deriving and applying small brightness temperature offsets that result in retrieved liquid water path (LWP) values that are near zero for clear-sky conditions. A further improvement in the retrieval accuracy is achieved at times of radiosonde launches using a physical-iterative approach (“PHYS”) that incorporates the additional information about the atmospheric temperature and water vapor structure. This study primarily uses the PHYS PW retrieval from the ARM MWRRET datastream, whose mean accuracy based on variability in the brightness temperature offsets at the SGP site in recent years (Table 2 in *Turner et al.* [2007]), is estimated to be about 0.25 K in terms of brightness temperature, which translates to uncertainty in PW of about 0.02 cm, or 2% if PW=1 cm, and <1% if PW>2 cm.

The “old” MWRLOS retrieval is widely used, and it is compared to the MWRRET PHYS retrieval in Fig. 4a. The MWRLOS retrieval is drier than PHYS by 1.2% when PW>2 cm, by 3% when PW=1 cm, and by 5% when PW=0.5 cm. The difference between retrieval algorithms will exhibit a seasonal cycle that can be judged from the time series of the PW for the ARM dataset (Fig. 4b). It would be reasonable to use the relationship in Fig. 4a to adjust ARM MWRLOS data to be equivalent in the mean to the PHYS retrieval for data from the SGP site since about 2004, but the relationship is not general due to time variability of the brightness temperature offsets in the PHYS algorithm and also due to the dependence of the algorithm on the mean atmospheric profile at a given site.

MWR/RS92 comparisons complement the CFH/RS92 comparisons in that the PW is dominated by water vapor in the LT, where the CFH measurements are not used. Calculations of the differential contribution of each altitude level in a sounding to the total PW show for the 2-year ARM RS92 dataset that 50% of the PW lies below 1.50 ± 0.43 km above ground level (AGL); 75% of the PW lies below 2.75 ± 0.69 km AGL; and 90% of the PW lies below 4.18 ± 0.93 km AGL. Thus, the PW is dominated by water vapor in the lowest 1.5 km, and the PW is influenced little by water vapor above 3-5 km, depending on the vertical distribution of RH. In contrast to the direct comparison of RH between RS92 and CFH, comparisons of PW between RS92 and MWR represent an average over the RH conditions in the LT, weighted toward lower altitudes and moister portions of the profile because these dominate the PW. It is noted for instruments with measurement issues near the surface (e.g., radiosondes before they equilibrate

by ventilation, and Raman lidar due to the overlap issue) that at SGP the lowest 100, 200, and 500 m AGL contribute $4.4\pm 1.4\%$, $8.7\pm 2.6\%$, and $20.5\pm 5.8\%$ of the PW respectively.

An additional consideration when comparing radiosondes and remote sensors is that the instruments measure different air. The impact of this sample volume mismatch depends on the degree of horizontal atmospheric variability, and in general the mismatch increases with height for vertically-pointing remote sensors. The MWRLOS retrievals (referred to as “ORIG” in the ARM MWRRET files) are not used in this study for assessing RS92 accuracy, but they can provide information about the influence of atmospheric variability on MWR/RS92 comparisons because they are calculated continuously as a function of time (every 20-30 s) rather than only at radiosonde launch times like the PHYS retrievals. The MWRLOS PW is taken to be the average PW in a 40 min window centered on the radiosonde launch time, where the PW varies with time due to both random error (noise) and real atmospheric change. The atmospheric variability is represented by the maximum difference between 5 min PW averages within the 40 min window, which averages out the noise in each subinterval and reveals real change in the PW with time. For the 2-year dataset, the mean and standard deviation of the percentage change in PW during the 40 min window is $4.1\pm 4.2\%$. This analysis rejected cases where the PW changed by $>10\%$ as having too much atmospheric variability for comparison to RS92. Lower rejection thresholds were explored, but it was found that, in the mean, MWR/RS92 comparisons are insensitive to atmospheric variability, which likely means that there is symmetry between cases of increasing vs decreasing PW within the window. However, atmospheric variability is certainly an important consideration when comparing individual cases or small datasets.

The ARM MWR and RS92 PW measurements are compared in terms of their ratio, the MWR “Scale Factor” (SF). The MWR SF has been used as a height-independent correction factor that removes radiosonde bias in a way that equates the radiosonde and MWR PW [Turner *et al.*, 2003]. Although a constant correction factor does yield scaled radiosonde profiles that better reflect the overall water vapor content in the LT, it is clear from Fig. 2 that the radiosonde bias error is a function of both height and RH. This study is instead concerned with diagnosing RS92 mean bias as a function of its dependences, and complements analysis by Cady-Pereira *et al.* [2008] that evaluates RS90 mean accuracy relative to the ARM MWRRET PHYS retrievals

for the 2001-2005 time period that just precedes the 2006-2007 time period considered in this study. Figure 5 shows the MWR SF for the 2-year dataset and its dependences on: solar altitude angle (a), day of RS92 calibration (b), PW (c), and the dominant RH that contributed to the PW (d). Cases were rejected if the data quality flag was not 0 or 4, or if the liquid water path was $>200 \text{ g m}^{-2}$, or if the PW changed by $>10\%$ within the 40 min window as described above.

Figure 5a shows that nighttime (blue) RS92 PW measurements are on average 3% moister than the MWR, which represents a 3% overall RS92 moist calibration bias in the LT, and is consistent with the magnitude and trend of the comparison to CFH at 700 mb for the higher RH conditions that dominate the PW (black curve in Fig. 2f). During the day, the SF increases approximately linearly with solar altitude angle until about $\alpha=30^\circ$, and then remains nearly constant at higher solar altitude angles except for some structures (bumps at 35° and 60°). Similar bumps are also present in the nighttime data and therefore are probably not caused by SRE. When the cases are ordered by the RS92 calibration day (Fig. 5b), it reveals time-dependent variability (drift) in the accuracy of the Vaisala calibration reference by about $\pm 2\%$ relative to the long-term mean (assuming that the MWR measurements are stable over time). It seems likely that the bumps of similar magnitude in Fig. 5a represent the same time-dependent variability seen in Fig. 5b, because a bias that varies with time will also appear to vary with the solar altitude angle at launch time if the launches occur at constant synoptic times. The day/night difference in Fig. 5b remains approximately constant with time, indicating that there was no noticeable change in the reflectivity of the RH sensor coating or other factors that affect the SRE over the 2006-2007 time period. Note that the irregular spacing of points in Fig. 5 occurs because the binning algorithm attempts to keep the statistics of each bin approximately constant, and therefore the mean values are more closely spaced where the data are relatively dense.

The SF also increases somewhat with increasing PW (Fig. 5c), although both the solar altitude angle dependence and the time-dependent variability complicate interpretation of this trend. A PW dependence in the SF might be caused by a PW-dependence in the accuracy of the MWR measurements or retrieval (as in Fig. 4a), and/or by an RH-dependence in the RS92 accuracy in the LT (as seen in the CFH/RS92 comparisons at 700 mb in Figs. 2f and 3f). A

sense of how the SF varies with RH is given by the PW-weighted RH (Fig. 5d), calculated by weighting each RH value in the RS92 profile by its fractional contribution to the total PW (i.e., $\sum RH_i \cdot \Delta PW_i$, where ΔPW_i is the differential contribution of each RH_i point to the total PW). This quantity reflects the RH conditions that dominated the PW for each sounding. The SF for both nighttime and daytime cases (blue and red) increases by $\sim 3.5\%$ over the range 30-80% RH. This trend is consistent with the RH-dependence at 700 mb from the CFH comparisons in that the nighttime moist bias decreases with increasing RH and the daytime dry bias increases with increasing RH. Therefore, the PW-dependence of the SF in Fig. 5c may be a consequence of RH-dependence in the RS92 accuracy rather than PW-dependence in the accuracy of the MWR measurements or retrieval, as will be shown later to be the case.

The solar altitude angle dependence of the mean MWR SF from Fig. 5a is shown in Fig. 6a with a polynomial curve fit to the daytime data (black). The green curve shows the analogous day/night difference in MWR SF found by *Cady-Pereira et al.* [2008] for ARM SGP RS90 and RS92 radiosondes in the time period 2001-2005. Both studies show consistent nighttime time series across the 2004-2007 timeframe (Fig. 5b and their Fig. 3), where the mean bias is about +3% and it varies with time by about $\pm 2\%$. Their altitude (or zenith) angle dependence is a fit to the data using a physically-based functional form for atmospheric transmissivity, whereas the black curve in Fig. 6a is a polynomial fit to the data where the fit extrapolation is constrained by the 9% day/night difference between RS92 and CFH in the LT for high solar altitude angles found by *Vömel et al.* [2007] at a tropical site (purple dot). The two studies find very similar results for solar altitude angles $< 30^\circ$, but the present study finds a larger SRE by 1-1.5% for high solar altitude angles. The likely explanation for the difference is that ARM launched RS90 radiosondes at SGP prior to early 2005, so the results of *Cady-Pereira et al.* [2008] primarily represent RS90 sensor characteristics, whereas the present study represents RS92 sensor characteristics. The sensor heating and therefore the SRE depends on the reflectivity of the sensor coating and especially the darker sensor attachment points, and the reflectivity may differ between RS90 and RS92. We suggest that the results of *Cady-Pereira et al.* [2008] better represent the SRE for RS90 radiosondes, and the present study better represents the SRE for RS92 radiosondes in the 2006-2007 timeframe.

Later in this paper a method will be presented to remove RS92 mean bias error in a way that accounts for the height and RH dependences given by the CFH comparisons, and accounts for the solar altitude angle dependence given by the black curve fit in Fig. 6a. If the SRE is defined as the difference between the curve fit and the mean nighttime SF (0.97), and the “full sun” SRE is defined as the mean SRE for the solar altitude angle range corresponding to the WAVES CFH/RS92 soundings (62-70°), then the curve in Fig. 6b gives the fraction of the full-sun SRE as a function of solar altitude angle. For example, the SRE at a solar altitude angle of about 16° is half its magnitude at 66°. An assumption of the forthcoming bias correction method is that the solar altitude angle dependence of SRE is a constant fraction of its magnitude at 66° for any P and RH conditions (i.e., the fraction does not vary with height). The full-sun SRE at $\alpha=66^\circ$ for any given P and RH conditions is equal to the day/night difference in the RS92 bias from the CFH comparisons in Figs. 3f and 2f, so the SRE for $\alpha=16^\circ$ will be taken as half its value at $\alpha=66^\circ$ for the given P and RH conditions, thereby providing the framework for coupling the CFH and MWR assessments of the RS92 SRE in a later section. Both this study and the study of *Cady-Pereira et al.* [2008] define the SRE as the difference between the daytime and the mean nighttime MWR SF as a function of solar altitude (or zenith) angle; however, their study implicitly assumes that the SRE is constant with height and with RH, whereas this study determines the height and RH dependences of the SRE at a specific solar altitude angle (66°) from the CFH comparisons, then represents the solar altitude angle dependence as a normalized fraction of the SRE for any given P and RH conditions.

In summary, the MWR/RS92 PW comparison in Fig. 5a represents the overall RS92 performance in the LT for all RH conditions combined, and is consistent with the RS92 bias relative to CFH at the “top” of the LT region at 700 mb in Figs. 2f and 3f, demonstrating consistency between the reference sensors. The MWR comparisons indicate an RS92 mean moist bias in PW of 3% at night and a mean dry bias in PW of about 6% during the day for the high solar altitude angles represented by the CFH comparisons. A similar RH-dependent trend in the LT is indicated by both reference sensors, although the MWR has little to say about the RS92 performance below 30% RH since those conditions contribute little to the total PW for most SGP RH profiles. The RS92 performance at the surface will now be evaluated.

d. ARM SurTHref

The ARM SurTHref system (Fig. 7) consists of a standard National Weather Service naturally-ventilated instrument shelter (“Stevenson screen”), which also contains an inner fan-ventilated chamber with 6 T/RH probes (3 Vaisala HMP45D and 3 Rotronic MP100H). The ARM launch procedure involves placing the sensor arm of the active radiosonde near the reference probes prior to launch, activating a mechanical switch that causes a flag to be set in the SurTHref data indicating the presence of a radiosonde. The RS92 RAW data file contains the prelaunch RS92 measurements, which can be correlated with the SurTHref time series and used to evaluate the RS92 performance at the surface. Solar radiation error is not a factor inside the SurTHref shelter, so all comparisons represent the RS92 nighttime (calibration) accuracy regardless of whether the launch is night or day.

One strength of the RH probes as reference sensors is that their calibration uncertainty is known. Both probe manufacturers use calibration reference standards that are based on saturated salt solutions and have a NIST-traceable uncertainty of 0.6% RH. The measured probe calibration accuracy relative to the reference standard is specified in the calibration sheets at 4 RH test values of approximately 0, 11, 33, and 75% RH, and at an ambient temperature in the range 22-27°C. All Rotronic probes differed from the reference standard by <0.2% RH at all RH test values, meaning that the Rotronic probes differ from the true RH by <0.8% RH at the RH test values. The Vaisala probes were also within 0.2% RH of the reference standard at the 11 and 75% RH test values (i.e., within 0.8% RH of the true RH), but they measured moister than the reference standard by 0.6% RH at the 0% RH test value, and drier than the reference standard by 1.35% RH at the 33% RH test value. Therefore, the Vaisala probes are drier than the true RH by up to 1.95% RH at 33% RH, so the Rotronic probes are the more accurate reference sensors at 33% RH. The measured deviations of the probe measurements from the reference standard could in principle be applied as a correction to the probe measurements that would yield an absolute probe calibration uncertainty of 0.6% RH at all RH test values; however, more RH test values would be needed to do this realistically because a single point at 33% RH is not sufficient to estimate the form of the deviation over the full RH range between the nearest test values of 11 and 75% RH. Note that RS92 radiosondes are similarly calibrated at 4 RH test points represented by a second-order polynomial fit, and these RH probe calibration accuracy

considerations suggest that curve-fit error is a viable hypothesis for the source of the RH-dependence of the RS92 bias relative to CFH.

One weakness of the SurTHref system as configured during the 2006-2007 timeframe is that routine calibration checks were not performed, and therefore the calibration drift is not characterized. A second weakness is that the RH probes exhibit considerable temperature-dependence in the accuracy of the RH calibration [Miloshevich *et al.*, 2004a], and therefore the probes cannot be relied upon except near the 22-27°C temperature range of the probe calibration conditions. This study uses only cases where the surface temperature was in the range 18-30°C to evaluate the RS92 accuracy at the surface, where this slightly larger temperature range improves the statistics of the comparison and was found to produce nearly identical results to the narrower temperature range of the probe calibration conditions.

Figure 8a illustrates the methodology of comparing time series of RAW RS92 and SurTHref data during the prelaunch time period. One minute averages are calculated for the RS92 and RH probe data, then the 3 probes from a given manufacturer are averaged together because it was found that the difference between probes from the same manufacturer was small compared to the difference between manufacturers. The one-minute differences between RS92 and either the Vaisala probes or the Rotronic probes are then averaged to give a single value for the mean difference between RS92 and each manufacturers' probes, as well as the mean difference between the two manufacturers.

The mean percentage (not RH) difference between the two manufacturers' probes is shown as a function of RH in Fig. 8b. The red curve shows the 1.6% RH maximum expected difference between the two manufacturers' calibrations (0.8% RH for each manufacturer if their calibration errors are on opposite sides of the true RH). The blue dot shows the percentage difference between manufacturers after accounting for the known dry bias of 1.35% RH in the Vaisala probe calibration at 33% RH (i.e., a percentage bias of 4.1%). The two manufacturers' calibrations agree within the maximum expected difference of 1.6% RH, with the exception of conditions above 95% RH where the difference approaches 2%. However, the probe calibration accuracy at very high RH was not checked, and therefore the probes are not reliable accuracy

standards for those conditions. The SurTHref data give no information below about 20% RH, because those surface conditions do not occur often at SGP when the temperature is in the range 18-30°C. However, even if dry conditions were encountered, the RH probes would not be suitable reference sensors because the probe calibration uncertainty in terms of percentage is large for dry conditions (red curve).

The mean percentage difference between RS92 and the Vaisala or Rotronic reference sensors is shown as a function of RH in Fig. 8c. The RS92 moist bias above about 50% RH is approximately constant and equal to 3.5% for the Rotronic probes (red) and 4.5% for the Vaisala probes (black), after accounting for the known calibration bias in the Vaisala probes (blue). This is consistent with the 3% overall moist bias in the LT at night relative to MWR, and is also consistent with the magnitude and trend of the nighttime RS92 bias relative to CFH at 700 mb for moist conditions (black curve in Fig. 2f). The approximate constancy of the bias above ~50% RH relative to the SurTHref probes is also consistent with the earlier observation that the RS92 bias relative to CFH for RH>48% is independent of RH. Both manufacturers' probes indicate that the RS92 bias increases with decreasing RH below about 40% RH, as is also seen in the comparison to CFH at 700 mb in Fig. 2f. The standard deviation of RS92 percentage difference from the RH probes is about $\pm 2\%$ for conditions above 50% RH (Fig. 8c), which is consistent with the $\pm 1.5\%$ RS92 1- σ production variability estimated from the MOHAVE dual RS92 soundings, after considering that part of the standard deviation in Fig. 8c is attributable to the RH probes.

e. RS92 RAW data

The RS92 RAW data are valuable not only because they contain the prelaunch time period needed for comparison to the SurTHref sensors, but also because the RAW data contain the original measured RH values that are above 100% RH or below 1% RH (including negative values), whereas the processed EDT and FLEDT data are limited to the range 1-100% RH. Measured RH values outside this range contain information about the calibration accuracy and its variability at the expected physical limits of 0% RH and 100% RH. The maximum and minimum RH values in the troposphere for each sounding in the 2-year dataset are plotted in Fig. 9a. Minimum values are very close to 0% RH, as expected since the GC correction is applied at

ARM. Note that negative RH values in the RAW data may indicate a flawed GC correction (in which case the sounding can be reprocessed without the GC correction).

The maximum value of the true RH in liquid water clouds is presumably very close to 100% RH, but values up to 106% RH are frequently measured by the RS92. These values are set to 100% RH in the standard EDT and FLEDT data files, which masks this uncertainty and likely leads to overestimates of the thickness of presumed liquid water layers. The maximum RH for each sounding is plotted in Fig. 9b as a function of the temperature where the maximum occurs. The maximum RH decreases with decreasing temperature from 106% RH above 0°C to about 100% RH at -35°C. The distribution of points above 100% RH represents both a mean moist bias in the RS92 calibration accuracy at 100% RH, plus the random production variability. One interpretation of Fig. 9b is that it represents a mean calibration bias of +3% with a 2- σ variability of $\pm 3\%$, which is consistent with the RS92 bias at high RH in the lower troposphere relative to MWR, SurTHref, and the CFH at 700 mb. Note that the data processing limits the RH value in the RAW data files to 106% RH, but this limit has little impact on the above analysis because only 12 of the 628 data points above 100% RH were limited by the 106% RH threshold. Note also that the upper limit of the maximum RH values below -35°C is determined by the homogeneous ice nucleation process.

f. CFH in the LT during WAVES

It was stated earlier that the WAVES CFH data appear to have a moist bias prior to the forced freezing event at $T_d = -15^\circ\text{C}$, when there is liquid water on the mirror. This conclusion is based on the sum of several observations, none of which alone is conclusive. The CFH measurements during WAVES were sometimes as high as 107% RH, which exceeds water saturation by more than the estimated instrumental uncertainty of 4%. Comparison of an ensemble of WAVES nighttime RS92 measurements with PW measurements from the HU MWR (using the Radiometrics neural net retrieval algorithm) indicated that RS92 had a moist bias in the LT of 4-5%, whereas the CFH indicated that RS92 had a dry bias in the LT of about -2% (Fig. 2e). The abrupt change in the character of the difference between RS92 and CFH at 700 mb in Figs. 2b and 2e is attributed to CFH behavior, because the RS92 calibration consists of smoothly-varying polynomials and offers no mechanism to explain a discontinuity, especially

one that coincides with the CFH forced freezing event. Ensemble comparisons of CFH with two Raman lidars during WAVES indicated that the CFH was moister than the lidars by about 7% prior to the forced freezing event, but they were in good agreement above the forced freezing event.

The above observations suggested a CFH mean moist bias prior to the forced freezing event as the only single factor that would explain all the observations, and this led to analysis of the ARM MWRRET and SurTHref data as reference-quality measurements with known accuracy under moist conditions of about $\pm 2\%$. In contrast to the nighttime RS92 mean dry bias of -2% implied by CFH for moist conditions in the LT, both ARM reference instruments indicate an RS92 mean moist bias of 3-4%, and the RS92 RAW data also indicate an RS92 mean moist bias of 3% at 100% RH on physical grounds. The sum of these observations, and the consistency of the other reference sensors within 2% suggests that the CFH had an unexplained mean moist bias during WAVES of about 5% when liquid water was on the mirror, leading to exclusion of the CFH as a reference instrument when $T_d > -15^\circ\text{C}$. Inspection of the individual RS92/CFH profiles indicates that the bias is only apparent for some WAVES soundings, and it may be limited to warm, moist conditions. One possibility not related to the instrument is that there may have been a procedural problem during some WAVES launches (e.g., mirror not properly cleaned prior to launch).

3. RS92 Accuracy Assessment

This RS92 accuracy assessment is well suited to the Consensus Referencing approach (e.g., *Fitzgibbon* [2008]), where a best estimate of RS92 accuracy over its measurement range is derived from RS92 comparisons to multiple reference instruments of known accuracy in some portion of the full P-RH space. The CFH provides an RH-dependent accuracy assessment above the 700 mb level, and the SurTHref provides an RH-dependent accuracy assessment at the surface, and the MWR provides both a consistency check and an indication of RH-dependence averaged over the lower troposphere, thereby tying together the CFH and SurTHref assessments. The MWR comparisons also give the solar altitude angle dependence of SRE.

The RS92 mean bias as a function of P and RH, as informed by the 3 reference instruments, is shown in Fig. 10a for nighttime soundings and Fig. 10b for daytime, clear, high solar altitude angle soundings, in terms of polynomial curve fits to the CFH comparisons from Figs. 2f and 3f (superimposed black curves in Fig. 10). Curve fits to additional RH intervals are shown as dashed, colored curves, and are explained below. The extrapolation of the curve fits from 700 mb to 1020 mb (dashed portion of black curves) was guided by several factors:

- The SurTHref indicates a 3.5% RS92 mean moist bias at the surface that is approximately constant for $RH > 40\%$ (Fig. 8c), consistent with the 3% mean moist bias at 100% RH seen in the RS92 RAW data (Fig. 9b). The SurTHref reflects only the nighttime RS92 performance since there is no SRE.
- The MWR scale factor indicates a nighttime RS92 mean moist bias of 3% for the moister RH conditions in the lowest few km that dominate the PW (Fig. 5a), and the bias increases with decreasing RH (Fig. 5d), consistent with the magnitude and RH-dependence of the bias relative to CFH at 700 mb.
- The MWR scale factor indicates a daytime RS92 mean dry bias of about 6% for the moister conditions in the LT for high solar altitude angles (Fig. 6a), and solar radiation error (day/night difference) of 9% in the LT.
- The curve fits for the moistest 2 or 3 curves are guided smoothly to surface values indicated by the SurTHref and MWR, and for drier conditions proportional spacing between the curves is maintained relative to the RH-dependence of the CFH comparisons at 700 mb.

The curve fits in Fig. 10 represent the best estimate of RS92 mean bias relative to the reference instruments, from the surface to 75 mb at night, and to 100 mb during the day for high solar altitude angles and clear conditions. The limits of 75 and 100 mb for the curve fits (vertical black lines) represent the upper limit of reliable RS92 data for this dataset, and correspond to the location above the tropopause where dry stratospheric conditions are encountered. Curve fits were added for three additional overlapping RH intervals in Fig. 10a to better represent the strong RH-dependence for dry conditions (dashed colored curves and lower legend), where only

the curve fits but not the underlying data are shown for clarity. Two additional curve fits are added in Fig. 10b for very dry conditions of $RH < 3\%$, where these fits are guided by the data but are largely determined by the proportional spacing constraint and therefore contain additional uncertainty. The weak portion of this RS92 assessment is dry conditions in the LT, approximately $RH < 25\%$ and $P > 700$ mb, because these surface conditions did not occur in the SurTHref data in the 18-30°C temperature range, and because the MWR PW is insensitive to dry conditions if the profile also contains moister conditions in the LT.

The coefficients of the RS92 mean bias curves are given in Table 1, where each curve gives the pressure dependence of the RS92 mean bias at an RH that is nominally equal to the center of the RH interval, but differs slightly for some RH intervals to better represent the mean RH in the data. These curves represent the RS92 mean bias relative to the reference sensors, so an estimate of RS92 mean absolute accuracy must also reflect the uncertainty in the reference measurements. The RS92 mean bias curves will also be used in the next section to develop an empirical correction that removes this mean bias from RS92 measurements, which will inform the final estimate of RS92 absolute accuracy.

In the stratosphere above the 45 mb level (night) or 50 mb level (day), the RS92 measurements are almost always in the range 1-2% RH, and the mean bias above these pressure levels is given in Table 1 as a single P-dependent curve that rises steeply with decreasing P (the continuation of the solid blue data in Fig. 10). This bias becomes quite large at low pressures (in a percentage sense) because the RS92 measurement remains roughly constant in the stratosphere, whereas the CFH continues to accurately measure the generally-decreasing RH. It is important to recognize that the RS92 is insensitive to water vapor in the stratosphere, and the RS92 mean bias there merely reflects a climatology of the WAVES and MOHAVE stratospheric CFH measurements and is not suitable for stratospheric research applications, but is given for completeness to allow the correction derived in the next section to be applied to full profiles.

To complete the specification of RS92 mean bias throughout the P-RH space, a transition must be made from the set of RH-dependent curves at P1 (75 or 100 mb) to the single curve at P2 (45 or 50 mb). Detailed investigation of this dataset and another CFH/RS92 dataset from a

tropical location with a high tropopause suggests that the abrupt change at P1 that is seen in the driest RH interval(s) in Fig. 10 is also the trend for all RH conditions. Experimentation has shown that a suitable approach is to specify the RS92 bias in the transition interval, $F_T(P)$, as a parabola that connects the RH-dependent bias value at P1 to the fixed bias value at P2. So, the RS92 bias in the transition interval P1 to P2 is specified as: $F_T(P) = F1 + A(P-P1)^2$, where $A=(F2-F1)/(P2-P1)^2$, and the constants are given in Table 1. As noted above, the RS92 is insensitive to water vapor in the stratosphere ($P < P1$ for this dataset).

4. Empirical Mean Bias Correction

The RS92 mean bias curves in Fig. 10 and Table 1 describe the RS92 mean accuracy relative to the reference sensors. These same curves can be used as an empirical bias correction that removes mean bias error from RS92 measurements, with the result that the mean accuracy of corrected RS92 measurements will be approximately equal to the mean accuracy of the reference sensors. Recalling that the RS92 mean percentage bias, $F(P,RH)$, refers to time-lag-corrected RH measurements (RH_{TLAG}), then the corrected RH (RH_{CORR}) is given by:

$$RH_{CORR} = G(P,RH) \times RH_{TLAG} , \quad (1)$$

where the correction factor $G(P,RH) = 100 / (F(P,RH) + 100)$. If time-lag error is not corrected and G is applied to the measured RH (RH_{MEAS}) rather than RH_{TLAG} , then time-lag error remains in the corrected data in the UT/LS region, with characteristics to be shown later.

a. Algorithm Implementation

The correction algorithm is implemented using Eq. (1) and the RS92 bias curves from Table 1 as follows:

- The correction factor G is calculated for all RS92 measured (or preferably time-lag-corrected) RH values from the surface up to the level P1. The bias value F for a given RH value is determined by linear interpolation between the appropriate curve fits.
- The correction factor G is calculated for all measurements above the level P2 using the single (RH-independent) curve fit in Table 1 labeled “P<P2”.

- The correction factor G in the “transition region” from P1 to P2 is calculated as a parabola as described in the previous section, where the value F1 is the bias value at P1 already calculated.
- For nighttime soundings, the RH profile is multiplied by the correction factor profile $G(P,RH)$, and the procedure is complete.
- For daytime soundings, $G(P,RH)$ represents the correction when the solar altitude angle is 66° (the mean solar altitude angle for the WAVES daytime soundings), and the SRE must be adjusted to account for its dependence on the solar altitude angle. The SRE for $\alpha=66^\circ$ is given by the difference between the daytime and nighttime RS92 bias, $F(P,RH,66^\circ) - F(P,RH,night)$. The fraction of this SRE for any given solar altitude angle is shown in Fig. 6b, and the curve fit is given in Table 1 (labeled “SF(α)”). The SRE after multiplying by this fraction is added back to the nighttime (zero SRE) bias to give the daytime bias for the given solar altitude angle, $F(P,RH,\alpha)$, from which the correction factor $G(P,RH,\alpha)$ is calculated and multiplied by the time-lag-corrected (or measured) RH.
- Clouds affect the sensor heating and the SRE in a complicated and uncharacterized way, so this daytime correction that is based on clear conditions may lead to erroneous results and should not be applied to cloudy daytime soundings, as discussed later. Note that “cloudy” can be crudely and conservatively estimated as those altitudes where the corrected RH profile exceeds ice-saturation (or is 100% RH in the LT).

b. Correction Results – NIGHT

Figure 11 shows the nighttime RS92 comparison to CFH for three circumstances: using the original RS92 measurements (a and d); after applying the time-lag and nighttime empirical bias corrections (b and e); and after applying only the nighttime bias correction and not the time-lag correction (c and f). Removing both the time-lag and mean bias errors (center panels) yields a mean difference from CFH that is $<2\%$ for all RH conditions from the 700 mb level to several km above the tropopause (i.e., to level P1, the upper limit of validity for the correction). The random nature of the residual bias in Fig. 11e indicates that there is no residual mean bias

relative to CFH. The offset in the mean altitude profile below 2.5 km in Fig. 11b indicates that the CFH mean moist bias prior to the forced-freezing event is about 5% relative to corrected RS92, where corrected RS92 represents the consensus of the MWR, the SurTHref, and the CFH just after the forced freezing event. The variability (standard deviation of differences) in Fig. 11e is about $\pm 5\%$ in the UT and varies somewhat with RH in the low and mid troposphere; however, this parameter describes the sensor-to-sensor variability in both the RS92 and CFH combined, and therefore is not especially meaningful except in placing an upper limit on the RS92 variability. A more meaningful estimate of RS92 variability is given by the standard deviation of percentage differences between two RS92 radiosondes on the same payload during 9 MOHAVE soundings, stated earlier to be $\pm 1.5\%$ for conditions above 10% RH and $\pm 3\%$ for conditions below 10% RH throughout the troposphere.

It is straightforward to implement the empirical bias correction presented in this paper, but the time-lag correction is much less straightforward. The physics of the time-lag correction is simple, but the numerical algorithm described by M04 is complicated and not general in nature like the empirical bias correction, in that the details of the algorithm must be tailored to handle the particular RH resolution, time resolution, noise characteristics and data processing artifacts of a particular RS92 dataset. If only the empirical bias correction and not the time-lag correction is applied to the RS92 measurements (Figs. 11c and 11f), then the resulting mean difference from CFH is still within the $\pm 2\%$ bounds up to about the 220 mb (12 km) level, because time-lag error only becomes significant below about -45°C . Above that level, the residual mean bias and especially the large variability reflect the uncorrected time-lag error for several km above and below the tropopause, where the moist bias above the tropopause represents lag in the ability of the RH sensor to follow the sharp decrease in RH that commonly occurs at the tropopause. However, the particular residual bias and variability seen in the UT/LS region of Figs. 11c and 11f represents the impact of time-lag error on only this particular dataset and is not a general result, because time-lag error depends strongly on the local humidity gradient as well as the temperature profile and will therefore vary considerably for different types of atmospheric profiles.

The $\pm 2\%$ bounds on the residual mean difference from CFH in Fig. 11e represents two sources of uncertainty that can be interpreted as really being the same underlying uncertainty: namely, the curve-fit error seen in Fig. 10a that is error in representing the data with curve fits, and second, statistical uncertainty in the mean accuracy of the CFH measurements that is given by the sample size. The statistical uncertainty in the mean CFH accuracy for any P-RH bin is given by the CFH measurement uncertainty for any individual sensor ($\pm 4\%$ in the LT to $\pm 9\%$ in the UT) divided by the square root of the number of sensor comparisons (the number of sensors that contributed to a bin, not the number of data points in a bin). The nighttime dataset consists of 33 RS92/CFH comparisons, but the number of sensors contributing to a given bin varies from a maximum of 29 to a minimum of 3. Most bins contain contributions from 9-25 sensors, and a typical single value is 16 sensors. For the typical number of contributing sensors, the statistical uncertainty in the mean accuracy of the CFH reference measurements is $\pm 1\%$ in the LT to $\pm 2.3\%$ in the UT. These values are consistent with the $\pm 2\%$ scatter in the residual mean bias in Fig. 11e, which supports the interpretation that curve fit error does not add additional uncertainty to the results because the curve fits simply provide a reasonable guess as to the underlying smooth trend in the mean bias that the statistical uncertainty is referenced to. Thus, the $\pm 2\%$ bounds for all RH conditions above the 700 mb level in Fig. 11e can be taken as the component of uncertainty in the corrected RS92 measurements that is attributable to statistical uncertainty in the CFH accuracy, if an ensemble of about 16 (or more) soundings is considered. For an ensemble of only 4 soundings, the residual statistical uncertainty is double the above values (within the limitations of the statistics of small numbers, and the variable number of sensors contributing to each bin).

The accuracy of an individual corrected RS92 sounding must also consider the random production variability, estimated from the MOHAVE dual RS92 soundings to be $\pm 1.5\%$ for conditions above 10% RH and $\pm 3\%$ for conditions below 10% RH. The question arises as to why the range of the RS92 difference from CFH in Fig. 11b is considerably greater than these mean accuracy and variability estimates. The spikes (outlier points) arise from such things as small misalignment of the independent RS92 and CFH profiles and differences in the instrument response and data processing characteristics, which can have a large effect at times when the RH is changing rapidly, but these average out in the mean. Therefore, the spikes are more

descriptive of limitations in the comparison method than in the underlying sensor accuracy and variability. The envelope of values (apart from the spikes) is not best described by the standard deviation of differences (1- σ variability, which captures 2/3 of the measurements), but more appropriately by the 2- σ variability (which captures 95% of the measurements), and specifically is dominated by the greater RS92 variability for dry conditions. Although the 2- σ variability for RS92 sensors as estimated above from the MOHAVE dual RS92 soundings is $\pm 6\%$ for conditions below 10% RH, the appropriate variability to consider when evaluating the plot is the standard deviation of differences from CFH, since much of the variability in the comparison is attributable to the CFH. That 1- σ value is shown in Fig. 11e to be 7-10% for the dry conditions that likely form the outer edge of the envelope of points in Fig. 11b, and therefore the 2- σ variability of differences from CFH is 14-20% for dry conditions, which is consistent with the (non-spike) range of differences seen in Fig. 11b.

An additional consideration in specifying the mean accuracy of corrected nighttime RS92 measurements concerns small RH offsets in the calibration accuracy at 0% RH, attributed to variability in the quality of the GC correction (if it is applied at all), and to uncertainty over how precisely the desiccant produces an environment of 0.0% RH. We adopt an estimated RH offset uncertainty of 0.5% RH, based on the 0.5% RH mean magnitude of the GC correction for the 2-year ARM dataset. The minimum achievable RH from laboratory desiccant tests is also about 0.5% RH, but this bias is implicitly accounted for by the correction. Accounting for small RH offsets in a final specification of uncertainty in the corrected RS92 data is important because it is a large percentage uncertainty in the water vapor concentration for very dry conditions (e.g., >10% uncertainty in mixing ratio if RH<5%).

The RS92 mean bias correction in the lower troposphere is evaluated in terms of the MWR scale factor for the corrected ARM RS92 dataset (Fig. 12). Focusing on the nighttime cases in Fig. 12a (blue), the correction removes the 3% mean moist bias in PW, and therefore the overall mean accuracy of the corrected RS92 measurements in the LT is equal to the mean accuracy of the MWR retrievals, estimated earlier to be about 0.02 cm, or 2% if PW=1 cm, and <1% if PW>2 cm. These values refer to the mean accuracy of corrected RS92 measurements over the long term, and Fig. 12b reveals additional time-dependent variability of about $\pm 2\%$ with

a period of about one year, which is also evident as the bumps of magnitude $\pm 2\%$ in Fig. 12a. These uncertainty estimates for PW are consistent with the statistical uncertainty of $\pm 2\%$ above the 700 mb level from the CFH comparisons. Figures 12c and 12d show that both the PW-dependence and the RH-dependence of the scale factor that were present in the uncorrected data (Figs. 5c and 5d) are removed by the correction. The assumed RH-dependence of the RS92 bias in the LT that was given by extrapolation of the CFH curves from 700 mb to the surface (Fig. 10a) is seen to fully account for the PW and RH dependences of the bias, suggesting that the MWRRET PHYS retrieval does not contain an appreciable PW-dependence for PW values as low as 0.5 cm, unlike the MWRLOS retrieval (Fig. 4a).

The RS92 nighttime bias correction at the surface is evaluated by applying it to the RS92 RAW data for the ARM dataset and re-comparing to the SurTHref measurements (Fig. 13). With the exception of the point at 40% RH, the corrected RS92 and Rotronic probes agree within 0.5%. The Vaisala probes, after accounting for the known probe calibration bias at 33% RH (blue), measure 1% drier than the Rotronic probes, which is within the combined calibration uncertainties for the two probe types of 0.8% RH each. The anomalous behavior of the point at 40% RH cannot be explained by known calibration bias in the Rotronic probes, but the anomaly of 1.2% is $< 0.5\%$ RH at 40% RH, which is within the probe calibration uncertainty. It is reasonable to interpret Fig. 13 as indicating that the mean accuracy of the corrected RS92 data relative to the RH probes is about $\pm 1\%$ over the range 25-100% RH, which is consistent with the corrected RS92 accuracy relative to the MWR after considering that the $\pm 2\%$ time-dependent variability is averaged together in Fig. 13 over the 2-year time period and therefore affects only the variability and not the mean difference. These results only apply to the temperature range 18-30°C (near the probe calibration temperature), but there is no indication of a temperature-dependent residual error in the MWR or CFH results. The SurTHref gives no evaluation of the RS92 at the surface for conditions below 25% RH. A chilled mirror instrument would be a better reference sensor for dry conditions because its uncertainty is a constant percentage at a given temperature, whereas a calibrated RH sensor is generally a more accurate reference sensor in a percentage sense for moister conditions.

c. Correction Results – DAY

Figure 14 shows the daytime RS92 comparison to CFH for the original RS92 measurements (a and d), and after applying the time-lag and daytime empirical bias corrections (b and e), and after applying only the daytime bias correction and not the time-lag correction (c and f). Removing both the time-lag and mean bias errors (center panels) yields a mean difference from CFH that is $<3\%$ for all RH conditions up to about the 200 mb level, with some larger values above that level. The $\pm 3\%$ residual uncertainty bounds are consistent with the expected statistical uncertainty in the mean CFH accuracy for 2-7 soundings (typical value of 4), which is $\pm 2\%$ in the LT to $\pm 4.5\%$ in the UT. Above the 200 mb level, the greater residual bias and the non-random trend with pressure for some RH intervals (red, purple) speaks of residual bias error superimposed on the random sampling error in the UT. This behavior was traced to differences of about 10% in the SRE in the UT between the 2006 and 2007 WAVES experiments, leading to a bimodal distribution of RS92 bias and the larger standard deviation in the UT seen in Fig. 14e. One possible explanation is if Vaisala changed the reflectivity of the sensor coating or attachment points between the two experiments. If the time-lag correction is not done (Fig. 14f), then the $\pm 3\%$ mean accuracy bounds for corrected daytime measurements still holds to about 220 mb (12 km), but above that level, in the tropopause region, the residual moist bias and large variability caused by time-lag error is apparent.

The daytime RS92 comparison to CFH in Fig. 14 represents only high solar altitude angles and clear conditions where the SRE is a maximum. However, the MWR comparisons from the more extended ARM dataset shown in Fig. 12 indicate that in the LT the daytime correction accounts for the solar altitude angle and RH dependences of the MWR scale factor with accuracy equivalent to the nighttime correction. The daytime correction assumes that the solar altitude angle dependence of the fractional SRE (Fig. 6b) is constant with height, but there were not enough daytime clear CFH soundings at lower solar altitude angles to verify that the fractional SRE does not vary with height.

A very important caveat about the daytime correction concerns the reduction in SRE by clouds, which leads to overcorrection because the daytime correction was derived from clear soundings and therefore represents the SRE for “full-sun” conditions. Within or possibly below cloud the actual sensor heating is less than the full-sun sensor heating assumed by the correction,

and above clouds there may be additional sensor heating due to the cloud albedo. The reduction in solar heating at any point in a cloud will depend, among other things, on the cloud optical depth between the sensor and the sun, which is unique for each cloud and would vary with altitude and with position in the cloud layer. In addition, time-lag effects complicate the situation for a sensor entering or emerging from cloud, where the time scale for establishing radiative equilibrium is unknown. The overcorrection is generally not obvious in soundings, because both clear-air and in-cloud ice-supersaturation are common. However, a few comparisons of corrected daytime RS92 soundings to CFH suggest that thin cirrus layers in the UT may have little impact on the full-sun SRE either within the cloud layer or below. The few CFH comparisons also show that cloud overcorrection is more likely and more pronounced at lower altitudes, presumably due to the generally increasing cloud ice-water content (IWC) with decreasing altitude. We strongly caution against indiscriminate application of the daytime bias correction to cloudy soundings, except possibly soundings containing only thin cirrus layers. During WAVES, rather than leaving the decision to data users, it was decided to not even archive the corrected daytime cloudy soundings, and such prudence is recommended to others. The SRE could be reduced with a simple radiation shield, but the most complete and satisfying solution to SRE (and other small issues related to thermal characteristics of the RH sensor) would be to measure the actual temperature of the RH sensor for input into the Vaisala calibration equation, if such a hardware solution is technically (and economically) feasible.

d. Corrected RS92 Accuracy Specification

Several factors contribute to an uncertainty estimate and range of validity for corrected RS92 data, including whether one is concerned with the mean accuracy of an ensemble of soundings (as might be considered in a remote sensor validation study), or whether one is concerned with the accuracy of an individual corrected sounding. The mean accuracy of the ensemble of Nighttime corrected RS92 measurements analyzed in this study is primarily given by uncertainty in the mean CFH measurements ($\pm 2\%$) and uncertainty in the RS92 calibration at 0% RH ($\pm 0.5\%$ RH), and is therefore specified as $\pm(2\% + 0.5\%RH)$. However, the WAVES and MOHAVE soundings analyzed in this study represent a small number of RS92 calibration batches, and the MWR comparisons revealed that the accuracy of the Vaisala calibration reference drifts by $\pm 2\%$ in a cyclical way with a period of about a year, adding a time-dependent

variability that depends on the calibration batch of the soundings in question. The calibration dates of the WAVES and MOHAVE soundings were aligned with the timeline of the MWR comparison in Fig. 12b to show that the WAVES/MOHAVE dataset is well represented by the long-term mean over the 2-year time period of the MWR dataset. Therefore, other RS92 radiosondes in the 2006-2007 timeframe may have additional bias error up to $\pm 2\%$ if they were produced at a time corresponding to one of the peaks in Fig. 12b. Adding this 2% mean bias uncertainty to the above accuracy specification would represent an upper limit since the additional bias is smaller for most calibration batches in the timeframe, so we instead adopt more of an “expected value” given by the RMS sum of the two uncertainty components, leading to a final mean accuracy specification for nighttime corrected RS92 measurements of $\pm(3\% + 0.5\%RH)$. Daytime corrected soundings have additional statistical uncertainty due to the smaller dataset, leading to a final mean accuracy specification for daytime RS92 soundings under clear conditions of $\pm(4\% + 0.5\%RH)$.

The above mean accuracy specification corresponds to an ensemble of about 16 (or more) nighttime soundings, or 4 (or more) daytime soundings. The accuracy of an individual corrected RS92 sounding must also consider the additional 1- σ variability of about $\pm 1.5\%$ for conditions above 10% RH, and $\pm 3\%$ for conditions below 10% RH. Less accurate results are expected in the UT for individual profiles if standard 1% RH resolution EDT data are used rather than the 0.01% RH resolution FLEDT data used in this study, because the lost detail in the vertical structure will give a less accurate time-lag correction at low temperatures. If no time-lag correction is applied prior to the empirical bias correction, then the above mean accuracy specification still holds, but only to the -45 or -50°C level (about 220 mb or 12 km in Fig. 11).

Application of the empirical bias correction to RS92 radiosondes that were produced outside the 2006-2007 timeframe may have additional mean bias uncertainty if either the accuracy of the Vaisala calibration procedure changed, or if manufacturing or hardware changes are made (such as changing the reflectivity of the sensor boom). An unpublished dataset from the German Weather Service (DWD) explored the accuracy of the Vaisala calibration at 100% RH from 2005 to early 2008, using an additional ground check in an environment at 100% RH, as is done as part of the “method of standardized frequencies” radiosonde correction procedure

described by *Nagel et al.* [2001] and *Leiterer et al.* [2005]. These data revealed that the Vaisala calibration changed abruptly at the beginning of 2006 and then changed back to the 2005 behavior at the beginning of 2008. The change consisted of a 1% increase in the mean moist bias at 100% RH in the 2006-2007 timeframe, and therefore the empirical bias correction will overcorrect RS92 soundings that were produced outside the 2006-2007 timeframe by 1%.

Figure 15 shows the two example soundings from Fig. 1 after applying the time-lag and empirical bias corrections, illustrating that both the nighttime and daytime corrections effectively remove the bias relative to CFH, and the time-lag correction recovers the vertical structure in the UT/LS region. If the bias correction alone is applied (Fig. 15a, blue), then the time-lag error in the original data remains. Note that time-lag error in the daytime example is minimal (Fig. 1b) because the temperature in the UT for this summertime case is relatively warm. In the lower troposphere, the residual bias between RS92 and CFH is attributed to the mean moist bias of unknown origin in the CFH measurements prior to the forced freezing event. The MWR and SurTHref comparisons show that the corrected RS92 measurements in the LT are about the same accuracy as the rest of the profile. Figures 11e and 14e show that the accuracy specification is valid for all RH conditions and for altitudes up to several km above the tropopause where dry stratospheric conditions are encountered (level P1 on average for this dataset, or about 18 km altitude for nighttime soundings and 16.5 km for daytime soundings).

The question of why the RS92 sensor appears to become insensitive above a certain level (P1) affects how appropriate this bias correction is for tropical or polar soundings. If the upper limit of validity is truly determined by a pressure threshold (e.g., if there is a pressure-dependence in the sensor response at low pressures that is not represented in the Vaisala calibration model), then the upper limit of validity that applies to this mid-latitude dataset should also apply to tropical or polar soundings. However, if the upper limit of validity is determined by a minimum threshold water vapor concentration, then the altitude of validity depends on the RH profile and in general would be higher for tropical soundings and lower for polar soundings. The assumption in this paper that the upper limit of validity is a pressure threshold (P1) is clarified by viewing the example corrected soundings in terms of mixing ratio rather than RH (Fig. 16). After removing time-lag error in the UT/LS (red profiles), it can be seen by comparing

to the CFH profile that the RS92 responds and generally agrees with the CFH at mixing ratios of 5 ppmv and below, including at temperatures below -70°C , and therefore it is not a threshold water vapor concentration that determines the upper limit of reliable RS92 measurements. Also, the tropopause is not a limiting factor, and neither is the minimum value of mixing ratio. The divergence of the red curve from the CFH measurements occurs at an altitude (pressure) that is consistent with the upper limit of validity given earlier as 75 mb for nighttime soundings and 100 mb for daytime soundings (18 and 16.5 km respectively). Inspection of all the dual soundings in the WAVES/MOHAVE dataset in terms of mixing ratio consistently shows that the RS92 remains responsive at even the minimum mixing ratio in the LS, and that the RS92 calibration fails to be accurate and trends toward paralleling the saturation mixing ratio curve at a pressure threshold of P1. Therefore, we expect the correction to be valid up to this level for soundings in tropical and polar locations as well, although this hasn't been proven. Above the level P1, the corrected RS92 data merely represent the CFH climatology during these mid-latitude experiments.

e. Relation to MWR Scaling

The correction derived in this paper removes mean bias error from RS92 measurements, but there is still variability in the accuracy of any individual sensor. This production variability is a “random bias”, where it is random in the sense that there is a normal distribution of sensor calibration accuracies for a large batch of sensors, but for any given sensor it is a specific bias relative to the mean accuracy. If coincident MWR measurements are available, then the MWR scaling method described by *Turner et al.* [2003] provides a possible means of removing the variability by standardizing the overall RS92 bias in the LT (in terms of PW) to the accuracy and stability of the MWR. The intent of the MWR scaling procedure was to apply it directly to the original measurements, which can be envisioned as collapsing all individual MWR SF values in Fig. 5a to be identically equal to 1.000 using a constant sensor-specific scale factor, as opposed to the results of the present study (Fig. 12) where the mean bias is removed as a function of its dependences but the sensor-to-sensor variability remains (as well as the $\pm 2\%$ time-dependent bias). The empirical bias correction factor $G(P, RH, \alpha)$ is analogous to the MWR scale factor in that both are multiplied by measured RH values to yield corrected RH values, but the MWR scale factor is a constant value rather than a function of the observed dependences, so it doesn't

appropriately address the bias in the UT since it differs from the bias in the LT, and even in the LT the MWR scale factor represents an average over RH conditions. However, if the empirical bias correction first removes the mean bias as a function of its dependences, then application of a constant MWR scale factor may be an accurate method of addressing the variability and residual time-dependent bias. It remains a topic of future research to verify that the production variability is indeed a constant bias with height for a given sensor, and therefore a constant MWR scale factor is appropriate throughout the atmospheric column.

f. Comparison to TCSP Correction

It is instructive to compare the empirical bias correction derived in this paper to an earlier height-dependent bias correction derived from dual RS92 and CFH soundings in the tropics during the summer 2005 NASA TCSP experiment [Vömel *et al.*, 2007]. The time-lag and TCSP bias corrections were applied to the WAVES/MOHAVE (W/M) RS92 dataset, and the percentage difference from the W/M CFH measurements is shown as a function of altitude and as a function of P and RH in Fig. 17. The daytime TCSP correction (top row) produces an overall residual RS92 moist bias relative to CFH of about 5% in the troposphere, which increases rapidly with height above the tropopause (panel a vs b). The nighttime TCSP correction (bottom row) has a relatively smaller effect on the original data, and it increases rather than decreases the dry bias in the UT/LS (panel d vs e). The nighttime TCSP correction is problematic for several reasons, but mainly because it is based on only 4 soundings and only very moist conditions.

Differences between the TCSP results and results using the empirical bias correction derived in this paper are understood in terms of the differences between the TCSP and W/M datasets from which the corrections were derived. The TCSP correction is independent of RH, but the RS92 bias does depend on RH and leads to an RH-dependent residual bias that increases with decreasing RH (Figs. 17c and 17f). The large moist bias of some individual points in Fig. 17b and the resulting greater mean bias is caused by dry layers whose greater RS92 bias is not captured by the TCSP correction.

The TCSP correction represents the generally moist conditions that dominated the TCSP dataset, so the TCSP correction best applies to the moist RH conditions represented by the black

curves in Fig. 17. The daytime moist conditions are overcorrected (relative to the W/M-based correction) by 2-5% below the 250 mb level and by 9% above that level (Fig. 17c), and the nighttime corrected soundings have a residual dry bias of 9% in the UT for moist conditions (Fig. 17f). The TCSP nighttime soundings indicated an RS92 moist bias of 4-13% in the temperature range -60 to -70°C (Table 1 in *Vömel et al.* [2007]), as opposed to the nighttime W/M soundings that indicated an RS92 dry bias in the UT. Two factors may have led to an erroneous moist bias in the nighttime TCSP measurements. Earlier RS92 radiosondes were susceptible to sensor icing in ice-supersaturated conditions at temperatures below -45°C, because the sensor heating was turned off below this temperature (the threshold was later changed to -60°C). One such icing case was seen in a CFH and dual RS92 sounding during MOHAVE in late 2006, where one of each radiosonde type was used, and only the RS92 with the -45°C threshold displayed an obvious moist bias caused by sensor icing. The 4 nighttime TCSP soundings contained much ice-supersaturation in the UT, which may have caused sensor icing and an apparent but erroneous moist bias that is removed by the correction, leading to the residual dry bias in the UT (black curve in Fig. 17f above the 250 mb level). An erroneous RS92 moist bias in the nighttime TCSP measurements would also produce the daytime residual moist bias in the UT (Fig. 17c, black curve above the 250 mb level), as a result of the manner in which the daytime TCSP correction was derived. The RS92 calibration error as given by the nighttime mean bias was removed from the daytime measurements to isolate and specify the SRE, increasing the dry bias that is attributed to SRE and leading to overcorrection. A second factor that differs between the TCSP and W/M RS92 data is the 1% RH resolution of the TCSP data. The lost detail in the vertical structure decreases the effectiveness of the time-lag correction and leads to “smoother” time-lag-corrected data, especially at low temperatures in the tropical UT where the RH sensor has a long time-constant.

Apart from the UT, and apart from the LT where there is uncertainty about an unexplained CFH moist bias in some WAVES soundings, the daytime TCSP correction for moist conditions in the mid-troposphere is about 2% larger than the correction derived in this paper (Fig. 17c, middle portion of black curve). The higher solar altitude angles for the TCSP soundings may have led to greater SRE and a larger daytime correction, but only slightly since the SRE is nearly constant for high solar altitude angles (Fig. 6). More likely, this difference

between the corrections arises from the time-dependent variability in the Vaisala calibration accuracy of about this magnitude. However, the two corrections are consistent for moist conditions in the mid-troposphere within the stated uncertainty of $\pm(4\% + 0.5\%RH)$.

5. Conclusions

The accuracy of Vaisala RS92 radiosonde water vapor measurements has been investigated by comparing mid-latitude RS92 measurements to simultaneous measurements from three reference sensors of known accuracy (CFH, MWR, and SurTHref). The three intercomparisons are combined to produce a detailed assessment of RS92 mean bias accuracy and variability as a function of its dependences on height, RH, and solar altitude angle. The RS92 mean bias differs fundamentally between nighttime and daytime soundings, where the nighttime mean bias is taken to reflect the RS92 calibration accuracy, and the daytime mean bias reflects both the calibration accuracy plus the effect of solar radiation error.

Relative to the CFH (reference sensor above the 700 mb level), the RS92 mean bias for nighttime soundings (Fig. 2f) varies with height from a moist bias at 700 mb (3% for moist conditions and up to 20% for dry conditions) to a dry bias in the UT (5% for moist conditions and up to 20% for dry conditions). The RS92 mean bias for daytime soundings and high solar altitude angles (Fig. 3f) is a dry bias that increases from about 5% at 700 mb to 45% in the UT, and varies somewhat with RH. The RS92 mean bias in the LT is found by comparing ARM RS92 and MWR measurements of PW as a function of the solar altitude angle (using the MWRRET PHYS retrieval algorithm). The RS92 mean bias in the LT is a 3% moist bias for nighttime soundings, and a maximum dry bias of 5-6% for daytime soundings at high solar altitude angles, consistent with the RS92 bias at 700 mb relative to the CFH. The MWR comparisons also revealed time-dependent variability of $\pm 2\%$ in the accuracy of the Vaisala calibration reference relative to the long-term mean in the 2006-2007 timeframe. The RS92 mean bias at the surface is found by comparing the prelaunch RS92 RAW data to multiple calibrated RH probes in the ARM SurTHref system (Fig. 8c), which represents the nighttime RS92 accuracy since there is no solar radiation error. The RS92 is found to have a mean moist

bias of 3.5% at the surface for conditions of RH>40%, which is consistent with the nighttime RS92 mean bias relative to both the MWR and the CFH at 700 mb.

The three RS92 accuracy assessments were combined and used to derive an empirical bias correction that removes the mean bias as a function its dependences on height, RH, and solar altitude angle. The mean accuracy of the nighttime corrected measurements, if an ensemble of about 16 (or more) soundings is considered, is given by the statistical uncertainty in the reference measurements ($\pm 2\%$), and the time-dependent variability in the accuracy of the Vaisala calibration reference ($\pm 2\%$), and uncertainty related to the GC correction ($\pm 0.5\%$ RH). These components of uncertainty are combined as described earlier to yield a mean accuracy specification for nighttime corrected RS92 measurements of $\pm(3\% + 0.5\%RH)$, valid for all RH conditions from the surface to the 75 mb (18 km) level. To clarify, the uncertainty is 3% of the measured RH value plus a small RH offset component that becomes a significant percentage of the measured value for dry conditions (e.g., 0.5% RH adds >10% uncertainty for conditions below 5% RH). The smaller daytime dataset leads to a mean accuracy specification for daytime corrected RS92 measurements of $\pm(4\% + 0.5\%RH)$, valid for all RH conditions from the surface to the 100 mb (16.5 km) level for clear conditions. These accuracy estimates assume that time-lag error in the UT/LS region has been corrected. If time-lag error is not corrected, then the above accuracy specifications still hold to about the 220 mb (12 km) level, but above that level the time-lag error is seen as a residual mean bias and greater variability, and the character of time-lag error is very sounding-specific. In addition to mean bias uncertainty, individual soundings also have a sensor-specific random calibration bias from “production variability”, and the 1- σ variability is estimated from dual RS92 soundings to be $\pm 1.5\%$ for conditions above 10% RH and $\pm 3\%$ for conditions below 10% RH.

A caveat of great importance for daytime soundings is that clouds reduce solar radiation error in a complicated and uncharacterized way, and therefore the daytime empirical bias correction should only be applied to clear soundings, as it may overcorrect if applied to cloudy soundings. Clouds do not affect the nighttime results.

The RS92 and CFH data used in this study were specifically obtained as part of water vapor validation and instrument intercomparison experiments, so attention was paid to important details such as using the high RH resolution data available in the Vaisala FLEDT files, and evaluating the GC correction and state of the desiccant frequently. Such details may not be standard practice, yet are necessary to maximize the accuracy of RS92 measurements, especially if corrections are applied. Several Best Practices suggestions for maximizing the accuracy of Vaisala radiosonde data are given in the Appendix.

Acknowledgements

We thank Ricardo Forno (U. Mayor de San Andres, La Paz, Bolivia), Belay Demoz, Everette Joseph and Cassie Sterns (Howard U.), Scott Rabenhorst (U. Maryland, College Park), Eduardo Landulfo and Ani Torres (Instituto de Pesquisas Energeticas e Nucleares, Sao Paulo, Brazil), and T.D. Walsh and A.L. Grigsby (NASA/JPL) for their effort in acquiring the high-quality radiosonde and CFH data used in this project. We also thank Dave Turner (U. Wisconsin/SSEC) and Krista Gaustad (PNNL) for advance processing of the MWRRET data used in this study and for enlightening discussion about MWR retrievals. We also acknowledge with gratitude the foresight of Barry Lesht (ANL) for archiving the ARM RAW radiosonde data that is rarely used but proved important to this study. We highly value the publicly-available data that were obtained from the Atmospheric Radiation Measurement (ARM) Program sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Environmental Sciences Division. This work was funded by NASA grant #NNX06AC79G.

APPENDIX

Best Practices for Research-Quality RS92 Data

- Vaisala RH data are reported as integer values in the standard processed EDT files, which unnecessarily introduces a roundoff error of $\pm 0.5\%$ RH. Rounding also removes detail that degrades the ability of a time-lag correction to recover vertical structure in the profile at low temperatures. The relatively new Vaisala DigiCora-III data system produces a second processed output file (FLEDT) that contains RH data with a resolution of 0.01% RH, and it is recommended that the FLEDT data be used and archived for all soundings. Ideally, the high RH resolution data would be the default EDT output format, for the benefit of future global radiosonde data archives.
- The Vaisala ground check (GC) procedure, when the sensor arm is placed in a container of desiccant nominally at 0.0% RH and the radiosonde measurement in the presence of the desiccant is optionally used as an RH offset correction in the calibration equation, has high potential for “operator error”. We learned this lesson during the first WAVES experiment after tracing negative RH values in the RS92 RAW data to bad GC procedures (these soundings were reprocessed without the GC correction). The mean and standard deviation of the GC correction for the 2006-2007 ARM RS92 dataset is $-0.49 \pm 0.35\%$ RH, so GC values greater than about 1% RH should be considered suspicious and merit changing the desiccant and checking the desiccant chamber for leaks and a solid seal. Furthermore, even the best of several desiccants tested by the Australian Bureau of Meteorology achieved a minimum of only 0.5% RH after 10-15 minutes [Gorman, 2002], suggesting an RH offset uncertainty of about 0.5% RH in the calibration accuracy at 0% RH. The benefit of improved accuracy from the GC correction is comparable to the uncertainty in the true chamber RH, and since improper procedures can lead to much larger errors, it is recommended to only apply the GC correction if care and scrutiny are routinely used.
- It is not common practice to save the RAW PTU data file generated by the Vaisala data system. The RAW data files contain the prelaunch data that is acquired during the launch preparations, including the GC procedure and any comparison to reference sensors such as

SurTHref. The RAW data also contains the measurements from both RH sensors, including values below 1% RH or above 100% RH (unlike the processed EDT and FLEDT files). These values are useful for evaluating the GC correction, and for assessing the sensor calibration accuracy at 0% RH and 100% RH, and for more accurately applying the nighttime bias correction for conditions near 100% RH. It is recommended to routinely save the RAW PTU files for future reference, even if such uses as described above are not immediately relevant.

- The choice or assumption about the formula for saturation vapor pressure (SVP) over liquid water (e_w) is a source of uncertainty in water vapor measurements and calculations at low temperatures, where various e_w formulations differ by up to 7% at -70°C (see discussion of SVP formulations and issues in Appendix A of M06). The e_w formulation of *Wexler* [1976] is implicit in Vaisala's calibration and should be used when converting Vaisala RH measurements to other water vapor units or to RH with respect to ice (RH_i), or when other water vapor measurements such as mixing ratio or frostpoint temperature are converted to RH for comparison to Vaisala measurements. Other radiosonde manufacturers (and other water vapor applications) likely assume a different e_w formulation, which should be identified. Note that the parameter RH_i contains little uncertainty associated with the choice of formulation for the SVP over ice (e_i), because the common formulations all agree within 0.5% throughout the atmospheric temperature range (M06).
- A valuable auxiliary radiosonde output file that is optimized for accuracy and for research would include: 1) standard RH resolution better than 1% RH; 2) an RH parameter that is not limited to the range 1-100% RH; 3) the parameter RH with respect to ice (RH_i), to remove the dependence on e_w as discussed above; 4) GC values in the header information; and 5) a time-lag-corrected RH parameter. The time-lag correction for Vaisala radiosondes (M04) has been proven reliable (e.g., Fig. 15; M06; *Vömel et al.* [2007]), but is complicated for end-users to apply because the numerical details of the algorithm must be tailored for the particular RH resolution, sample time interval, and noise and data processing characteristics of a given type of output file. A single optimized correction algorithm applied during the original data processing would provide consistent time-lag-corrected data regardless of the output file characteristics or post-processing operations.

REFERENCES

- Cady-Pereira, K.E., M.W. Shephard, D.D. Turner, E.J. Mlawer, S.A. Clough, and T.J. Wagner (2008), Improved daytime column-integrated precipitable water vapor from Vaisala radiosonde humidity sensors, *J. Atmos. Oceanic Technol.*, 25, 873-883.
- Facundo, J. and J. Fitzgibbon (2007), Use of the Consensus Reference concept for testing radiosondes, 87th AMS annual meeting, 14th Symposium on Meteorological Observations and Instrumentation, San Antonio, Texas, 14-18 January.
- Fitzgibbon, J. (2008), The Consensus Reference methodology as it applies to a radiosonde under test, 88th AMS annual meeting, 12th Conference on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface, New Orleans, Louisiana, 20-24 January.
- Gorman, J. (2002), Instrument Test Report 2002_664: Evaluation of desiccants, Australian Bureau of Meteorology. Available at: www.wmo.int/pages/prog/www/IMOP/WebPortal-AWS/Tests/ITR664.pdf
- Leblanc, T., I.S. McDermid, T.G. McGee, L. Twigg, G. Sunnicht, D. Whiteman, K. Rush, M. Cadirola, D. Venable, R. Connell, B. Demoz, H. Vömel, and L. Miloshevich (2008), Measurements of humidity in the atmosphere and validation experiments (MOHAVE, MOHAVE II): Results overview, 24th International Laser Radar Conference, Boulder, Colorado, 23-27 July.
- Leiterer, U., H. Dier, D. Nagel, T. Naebert, D. Althausen, K. Franke, A. Kats, and F. Wagner (2005), Correction method for RS80-A humicap humidity profiles and their validation by lidar backscattering profiles in tropical cirrus clouds, *J. Atmos. Oceanic Technol.*, 22, 18-29.
- Miloshevich, L.M., H. Vömel, D.N. Whiteman, B.M. Lesht, F.J. Schmidlin, and F. Russo (2006), Absolute accuracy of water vapor measurements from six operational radiosonde types launched during AWEX-G, and implications for AIRS validation, *J. Geophys. Res.*, 111, D09S10, doi:10.1029/2005JD006083.
- Miloshevich, L.M., A. Paukkunen, H. Vömel, and S.J. Oltmans (2004), Development and validation of a time-lag correction for Vaisala radiosonde humidity measurements, *J. Atmos. Oceanic Technol.*, 21, 1305-1327.
- Miloshevich, L.M., B.M. Lesht, and M. Ritsche (2004a), New surface meteorological measurements at SGP, and their use for assessing radiosonde measurement accuracy, 14th ARM Science Team meeting, Albuquerque, New Mexico, 22-26 March. Available at http://www.arm.gov/publications/proceedings/conf14/extended_abs/miloshevich_lm.pdf

Nagel, D., U. Leiterer, H. Dier, A. Kats, J. Reichardt, and A. Behrendt (2001), High accuracy humidity measurements using the standardized frequency method with a research upper-air sounding system, *Meteor. Z.*, *10*, 395-405.

Paukkunen, A., V. Antikainen, and H. Jauhiainen (2001), Accuracy and performance of the new Vaisala RS90 radiosonde in operational use, 11th AMS Symposium on Meteorological Observations and Instrumentation, Albuquerque, New Mexico, 14-18 January.

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.

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

Vömel, H., H. Selkirk, L. Miloshevich, J. Valverde-Canossa, J. Valdes, E. Kyro, R. Kivi, W. Stolz, G. Peng, and J.A. Diaz (2007), Radiation dry bias of the Vaisala RS92 humidity sensor. *J. Atmos. Oceanic Technol.*, *24*, 953-963, doi: 10.1175/JTECH2019.1.

Vömel, H., D. David, and K. Smith (2007a), Accuracy of tropospheric and stratospheric water vapor measurements by the Cryogenic Frostpoint Hygrometer (CFH): Instrumental details and observations, *J. Geophys. Res.*, *112*, D08305, doi:10.1029/2006JD007224.

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

Whiteman, D.N., M. Adam, C. Barnet, B. Bojkov, B. Demoz, J. Fitzgibbon, R. Forno, R. Herman, R. Hoff, E. Joseph, E. Landulfo, K. McCann, T. McGee, L. Miloshevich, I. Restrepo, F. Schmidlin, B. Taubman, A. Thompson, L. Twigg, D. Venable, H. Vömel, C. Walthall (2008), The Water Vapor Variability – Satellite/Sondes (WAVES) field campaigns, 24th International Laser Radar Conference, Boulder, Colorado, 23-27 July.

Table 1: Coefficients (a_i) for the polynomial curve fits in Fig. 10 that give the RS92 mean percentage bias from the consensus of the reference sensors, $F(P)$, for the listed RH values that represents each RH interval in Fig. 10. Separate fits describe night vs day soundings, and the low-P limit of validity for the fits is given by P1. Fit coefficients are also shown for the single curve that applies to all RH conditions when $P < P2$, and the RS92 bias at P2 is F2. The polynomial fits are given by:

$$F(P) = \sum_{i=0}^N a_i \cdot P^i, \text{ where } N \text{ is the order of the fit.}$$

Fit coefficients are also shown for the daytime mean MWR scale factor (SF) and the fraction of the maximum solar radiation error (SRE) as a function of solar altitude angle (α), from Figs. 6a and 6b. Treatment of the pressure interval P1 to P2, as well as implementation of the fits as an empirical RS92 bias correction, are described in the text. Note, for example, that the nighttime fits representing 50% RH and 100% RH are identical, which indicates a constant RS92 bias over the range 50-100% RH.

NIGHT	P1=75mb	P2=45mb	F2=50				
RH / fit	a0	a1	a2	a3	a4	a5	a6
0	5.1993e+1	-7.9576e-1	3.9051e-3	-8.9666e-6	1.1825e-8	-8.4134e-12	2.4210e-15
1.5	5.1993e+1	-7.9576e-1	3.9051e-3	-8.9666e-6	1.1825e-8	-8.4134e-12	2.4210e-15
2.5	4.3729e+1	-7.8757e-1	3.8100e-3	-8.4919e-6	1.0830e-8	-7.5247e-12	2.1433e-15
3	1.0102e+1	-3.5020e-1	1.3771e-3	-1.8918e-6	1.5448e-9	-1.0460e-12	3.7543e-16
4	-1.2053e+1	-1.3963e-1	5.0608e-4	8.7142e-8	-1.1580e-9	9.6029e-13	-2.2738e-16
6	-1.9292e+1	-5.3081e-2	1.1776e-5	1.5888e-6	-3.7721e-9	3.2351e-12	-9.7876e-16
8.5	-1.4220e+1	-1.5629e-1	7.3102e-4	-5.7830e-7	-6.9512e-10	1.1583e-12	-4.3573e-16
12	-8.6609e+0	-2.3153e-1	1.1601e-3	-1.6559e-6	4.7114e-10	6.4842e-13	-3.7600e-16
20	-1.2075e+1	-9.0493e-2	4.5730e-4	-4.4334e-7	-2.5251e-10	5.6512e-13	-2.1830e-16
30	-8.4463e+0	-6.7739e-2	2.1850e-4	2.4128e-7	-1.1680e-9	1.1593e-12	-3.6948e-16
42	-7.5226e+0	-9.4287e-2	5.6012e-4	-1.0285e-6	8.1621e-10	-2.4513e-13	3.3189e-18
50	3.7854e+1	-4.9026e-1	2.0313e-3	-3.9299e-6	3.9439e-9	-1.9776e-12	3.8808e-16
100	3.7854e+1	-4.9026e-1	2.0313e-3	-3.9299e-6	3.9439e-9	-1.9776e-12	3.8808e-16
P<P2	4.3867e+3	-3.7335e+2	1.2676e+1	-1.9717e-1	1.1628e-3		
DAY	P1=100mb	P2=50mb	F2=80				
RH / fit	a0	a1	a2	a3	a4	a5	
0	6.8793e+0	1.6275e-1	-3.2097e-5	-4.1883e-7	5.0829e-10	-1.9028e-13	
1.9	-1.3058e+1	1.5405e-1	3.0599e-5	-4.9033e-7	5.4030e-10	-1.9315e-13	
2.4	-4.7161e+1	1.3916e-1	1.3784e-4	-6.1264e-7	5.9504e-10	-1.9805e-13	
3.5	-6.0069e+1	1.3320e-1	1.8078e-4	-6.6256e-7	6.1467e-10	-1.9661e-13	
5	-6.6681e+1	1.4741e-1	1.6426e-5	-1.4146e-7	8.9222e-12	4.0390e-14	
11	-6.7112e+1	1.1009e-1	3.7366e-4	-1.2284e-6	1.2520e-9	-4.3857e-13	
22	-6.6938e+1	1.1812e-1	2.8349e-4	-1.0166e-6	1.0377e-9	-3.5797e-13	
34	-6.0024e+1	1.4726e-1	-6.9462e-5	-2.0216e-7	3.1579e-10	-1.3450e-13	
100	-6.0024e+1	1.4726e-1	-6.9462e-5	-2.0216e-7	3.1579e-10	-1.3450e-13	
P<P2	5.4021e+3	-3.5312e+2	8.1766e+0	-6.4838e-2			
SF(α)	9.6886e-1	3.3717e-3	-4.2343e-5	1.7882e-7			
frac SRE(α)	-1.6061e-3	3.7746e-2	-4.7402e-4	2.0018e-6			

Example RS92/CFH Profile Comparisons

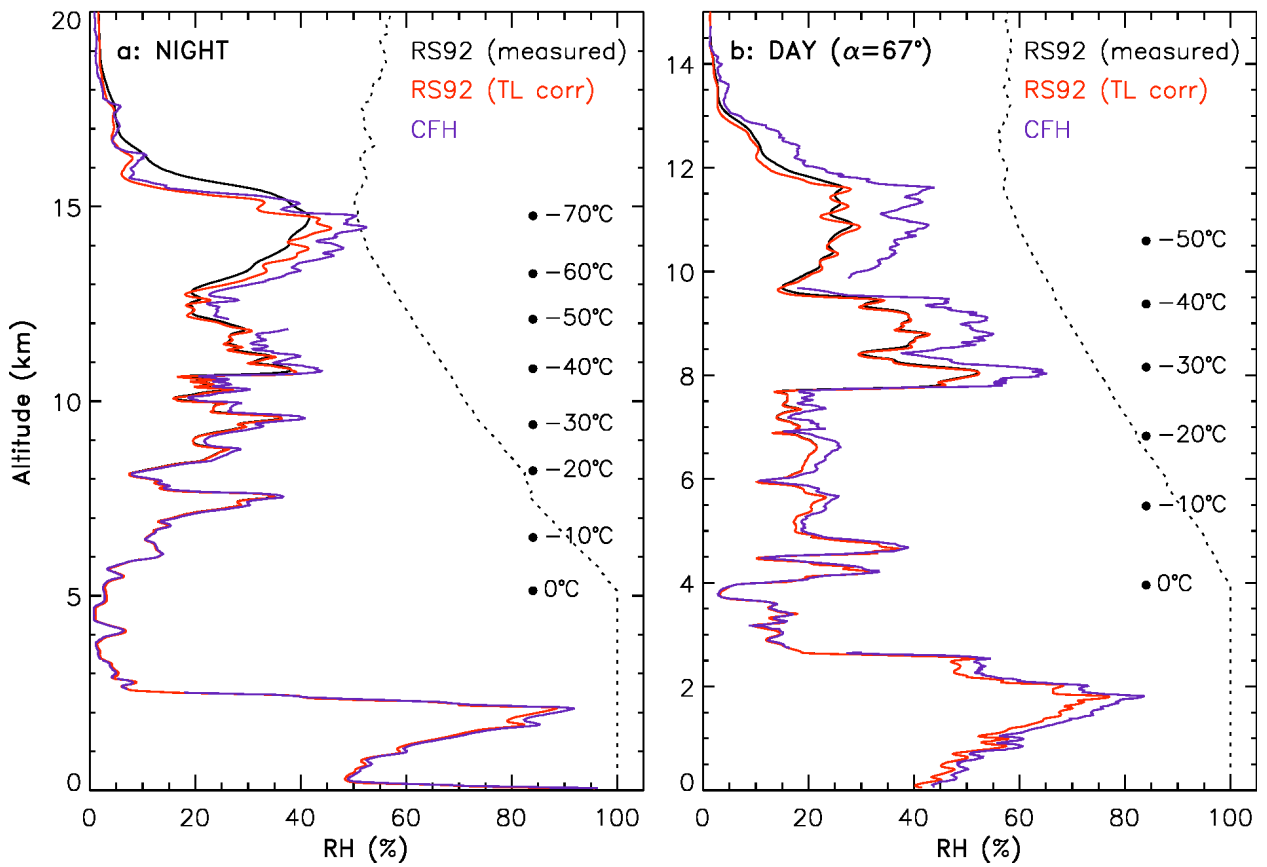


Figure 1. Example dual RS92/CFH soundings (a) at night and (b) at midday. Shown are: (black) the measured RS92 profile, (red) the RS92 profile after correcting for time-lag error, (purple) the CFH profile, and (dashed) ice-saturation. The CFH forced freezing event is indicated by the gap in the CFH data at 2.5 km altitude.

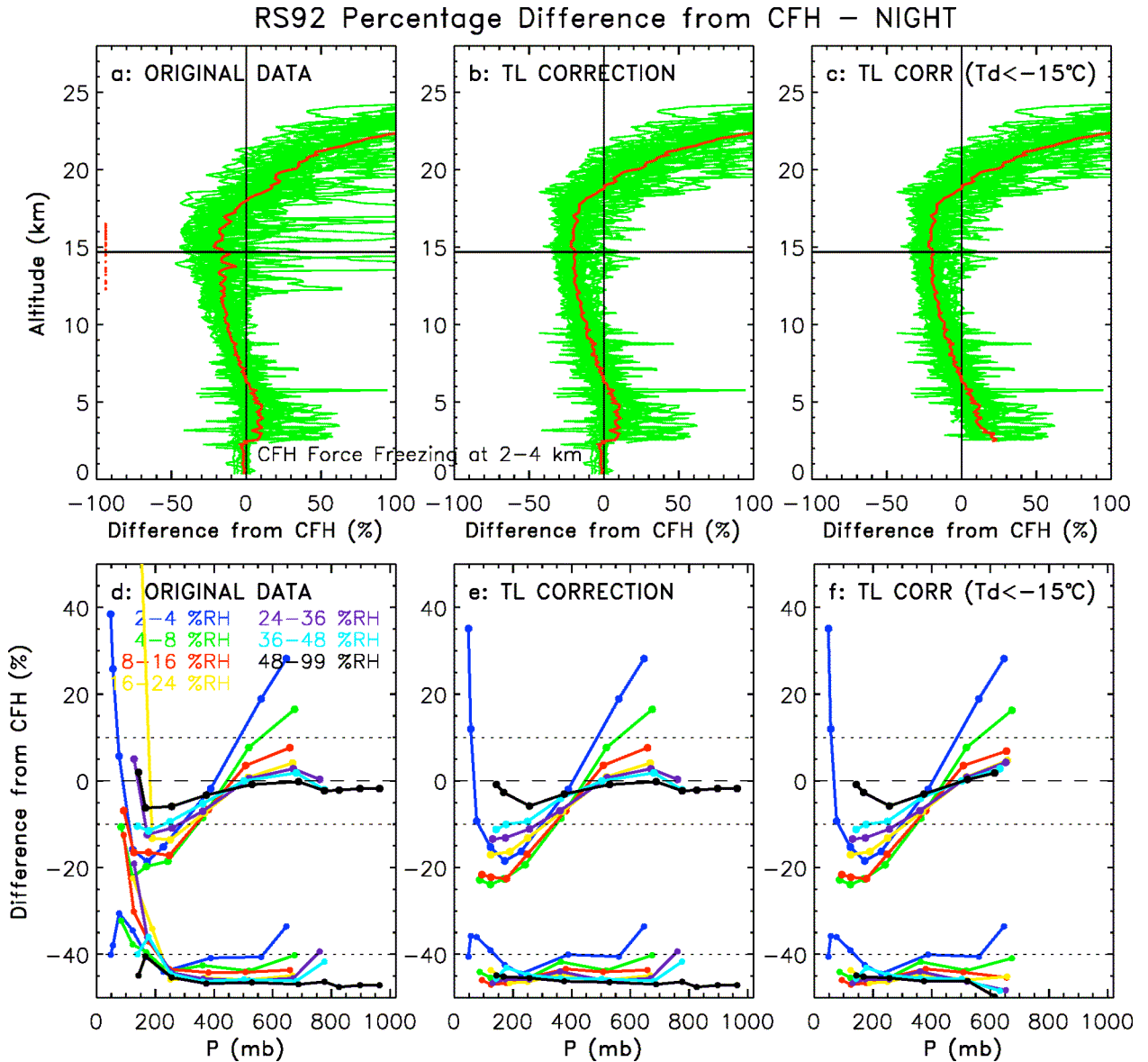


Figure 2. RS92 percentage difference from CFH for the 33 Nighttime WAVES and MOHAVE dual soundings, showing (top row) the individual altitude profiles and the mean, and (bottom row) the binned mean difference as a function of pressure in 7 RH intervals. (a and d) Left panels show the difference from CFH for the original RS92 FLEDT measurements; (b and e) center panels show the difference after correcting the RS92 data for time-lag error; and (c and f) right panels show the difference when data prior to the CFH forced freezing event are excluded. Lower curves in d-f give the standard deviation of the percentage differences, offset to zero at the bottom of the panel for clarity. Horizontal line in a-c is the mean tropopause height, and red dots are the individual tropopause estimates.

RS92 Percentage Difference from CFH – DAY

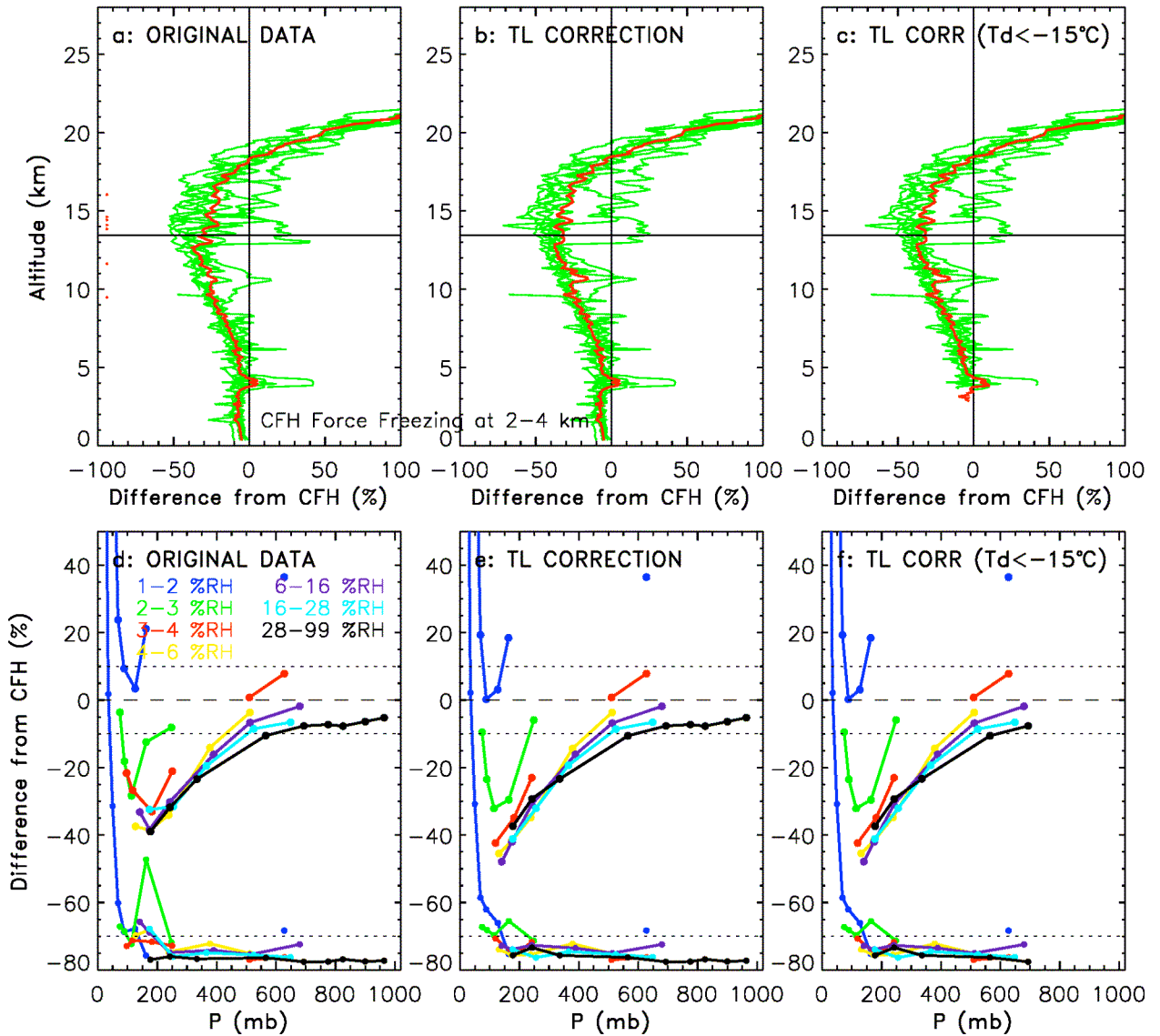


Figure 3. RS92 percentage difference from CFH for 7 Daytime WAVES and MOHAVE dual soundings, showing (top row) the individual altitude profiles and the mean, and (bottom row) the binned mean difference as a function of pressure in 7 RH intervals. (a and d) Left panels show the difference from CFH for the original RS92 measurements; (b and e) center panels show the difference after correcting the RS92 data for time-lag error; and (c and f) right panels are when data prior to the CFH forced freezing event are excluded. Lower curves in d-f give the standard deviation of the percentage differences, offset to zero at the bottom of the panel for clarity. Horizontal line in a-c is the mean tropopause height, and red dots are the individual tropopause estimates. Note that the RH intervals are different than in Fig. 2.

Comparison of MWR Retrieval Algorithms, and Seasonal PW Cycle

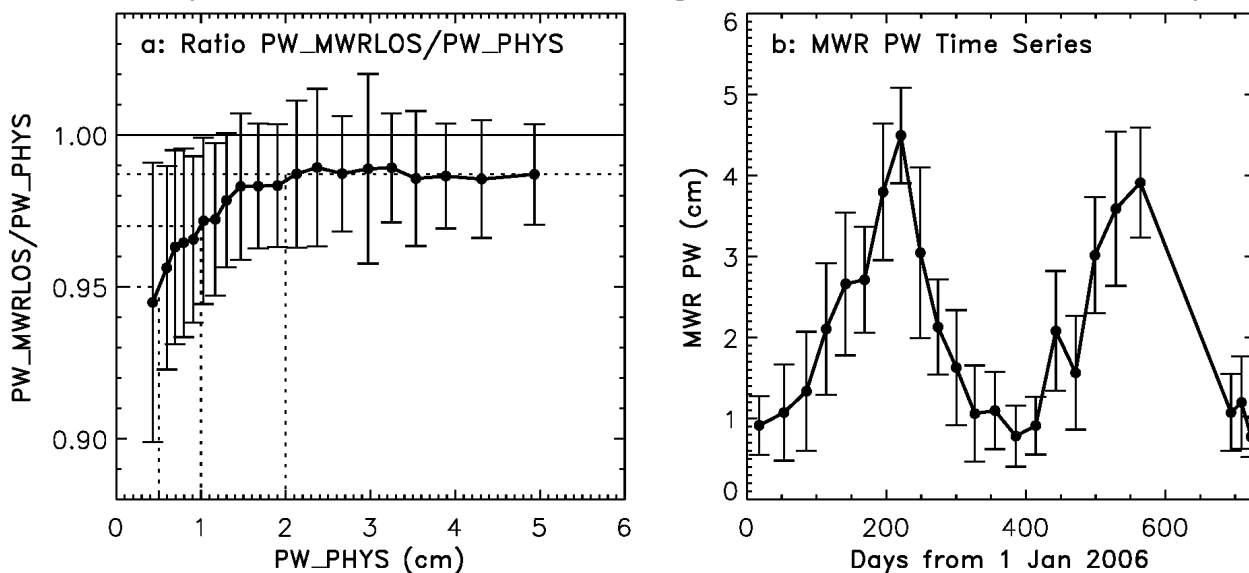


Figure 4. (a) Ratio of PW retrieved from ARM MWR measurements using the older MWRLOS algorithm to that retrieved using the more accurate MWRRET PHYS algorithm, shown as a function of PW for the 2006-2007 time period at the ARM SGP site. (b) Time series of the two year ARM dataset showing seasonal variation in the mean PW.

RS92 Comparison to ARM MWR

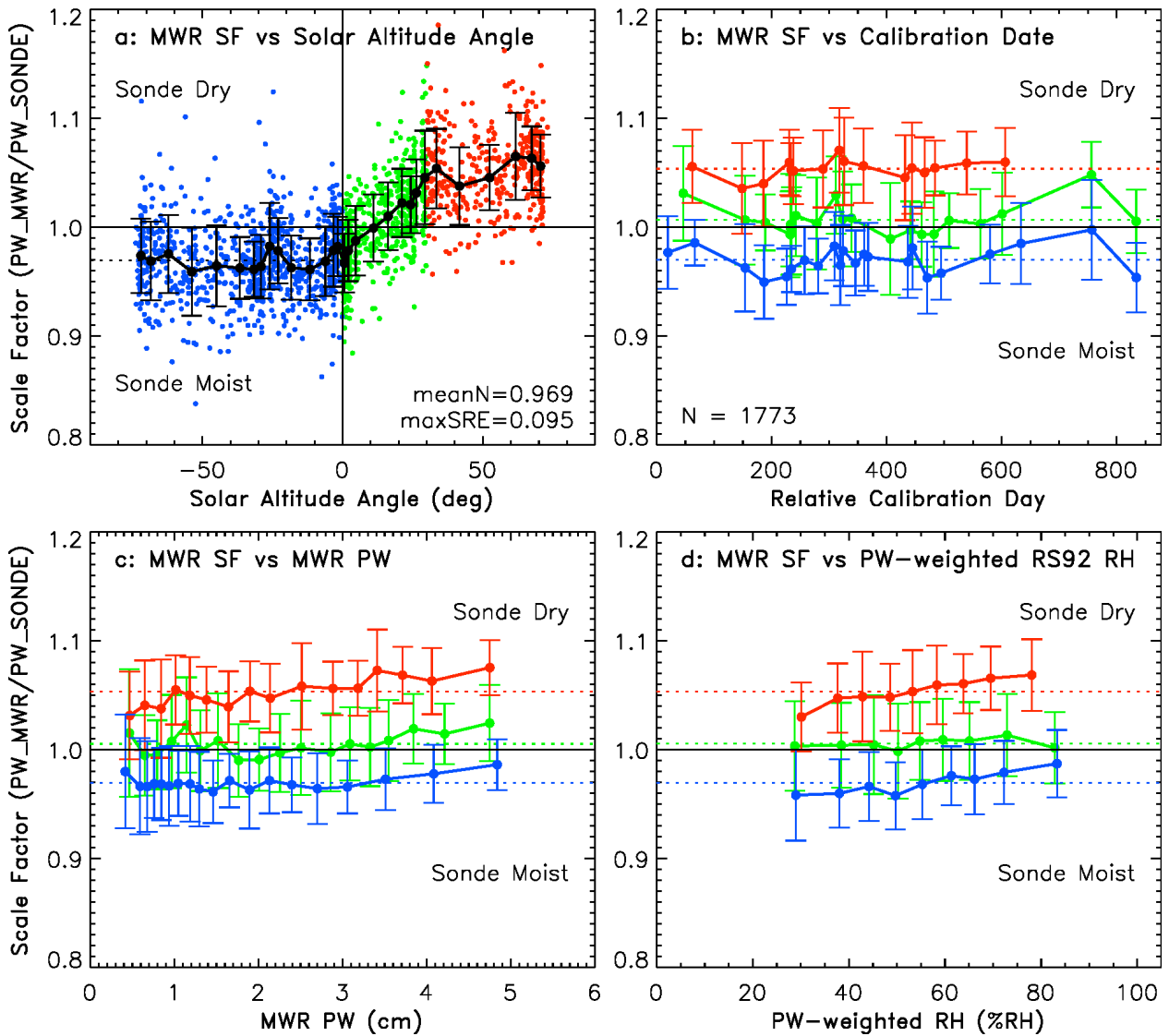


Figure 5. The MWR Scale Factor (SF) for the 2006-2007 ARM SGP dataset. Shown are the dependences of the SF on: (a) the solar altitude angle at launch; (b) on time as ordered by increasing RS92 calibration date; (c) on PW; and (d) on the PW-weighted RH calculated from the RS92 profiles as described in the text. Blue represents nighttime soundings, red represents daytime soundings with $\alpha > 30^\circ$, and green represents daytime soundings with $\alpha < 30^\circ$. The radiosonde calibration date is encoded in the serial number as described in Appendix A of M04.

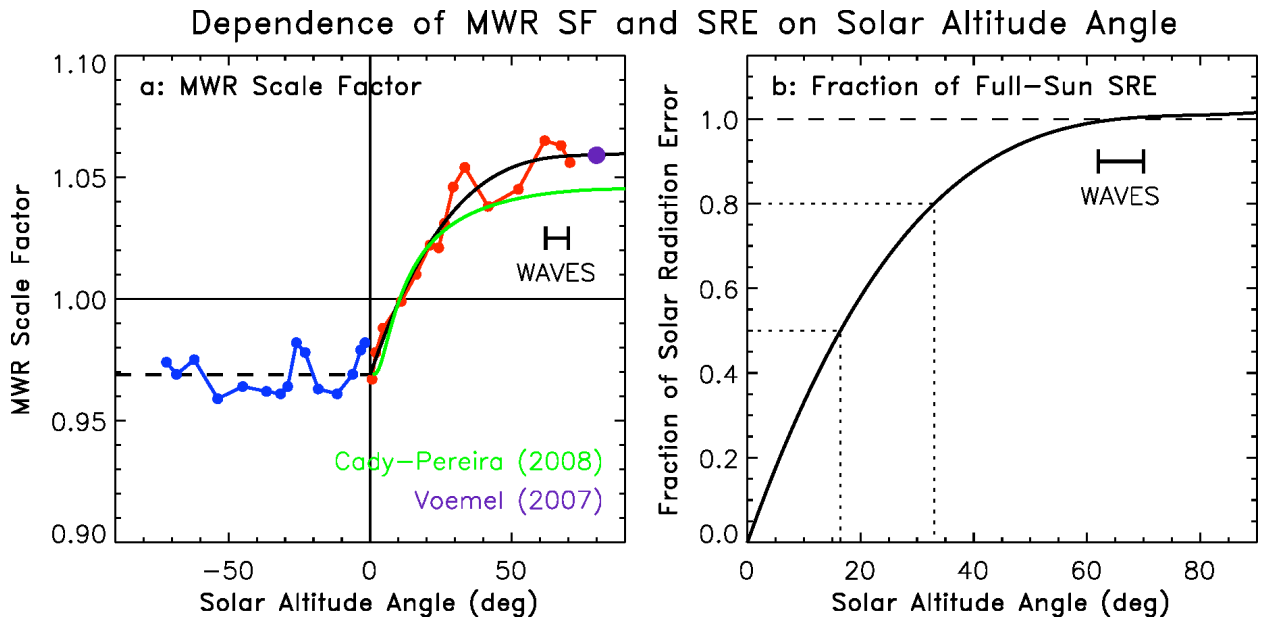


Figure 6. (a) The mean MWR Scale Factor (SF) binned as a function of solar altitude angle from Fig. 5a, with (black) a polynomial curve fit to the daytime data, (green) the results of *Cady-Pereira et al.* [2008] for earlier ARM RS90 and RS92 data, and (purple) the 9% day/night difference in the LT for high solar altitude angles from *Vömel et al.* [2007]. (b) The fractional solar radiation error (SRE), given by the difference from the mean nighttime SF, normalized to the SRE for the mean solar altitude angle from the daytime CFH soundings during WAVES shown in Fig. 3, and indicated by the horizontal bars above.



Figure 7. The ARM SurTHref system, showing the outer NWS-type instrument shelter and ventilated inner chamber containing 6 T/RH probes from 2 manufacturers, in which an active radiosonde is placed prior to launch for comparative measurements. A switch is activated by the radiosonde that sets a flag in the SurTHref datastream indicating the presence of a radiosonde.

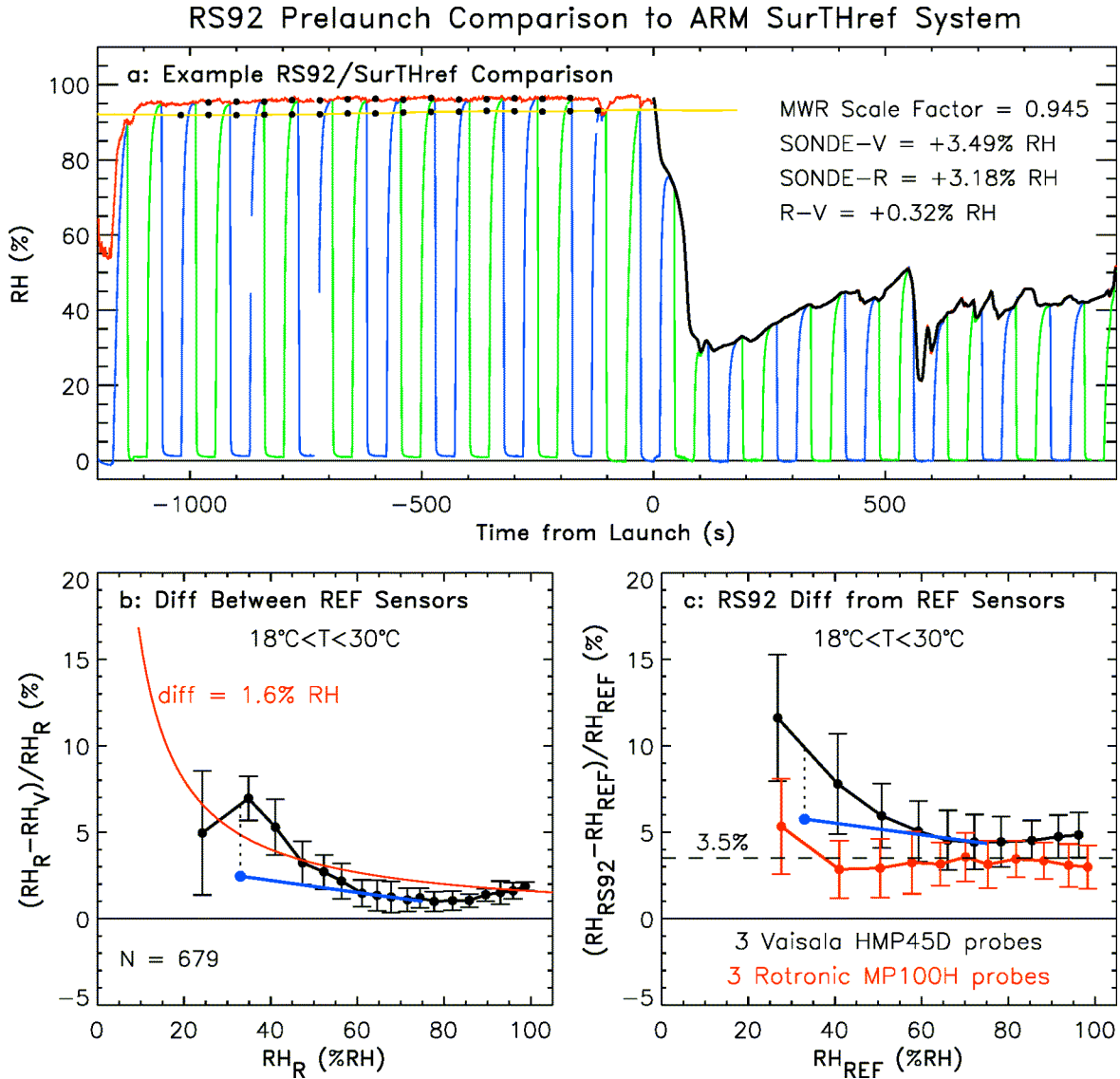


Figure 8. Summary of RS92 comparisons to the ARM SurTHref RH probes for the 2006-2007 time period at the SGP site. (a) Time series of an example comparison, showing: (red) RS92 RAW data; (yellow) average of the 3 Vaisala RH probes (“V”) and also the average of the 3 Rotronic probes (“R”), where the two are indistinguishable on this plot; (black) processed RS92 FLEDT data beginning at the launch time; (blue and green) measurements from the 2 alternately-heated RS92 RH sensors that are combined to yield the RAW and FLEDT RH data. Dots show 1-minute averages. (b) (black) Mean and standard deviation of the percentage difference between the two manufacturers’ RH probe measurements as a function of RH for the 2-year dataset, restricted to cases in the 18-30°C temperature range. (red) The maximum expected difference of 1.6% RH between the manufacturers’ NIST-traceable calibrations, and (blue) the difference after accounting for the known 1.35% RH (or 4.1%) calibration bias for the Vaisala probes at the 33% RH calibration point. (c) Percentage difference between RS92 and the Vaisala or Rotronic probes as a function of RH, also showing the effect of accounting for the known calibration bias of the Vaisala probes.

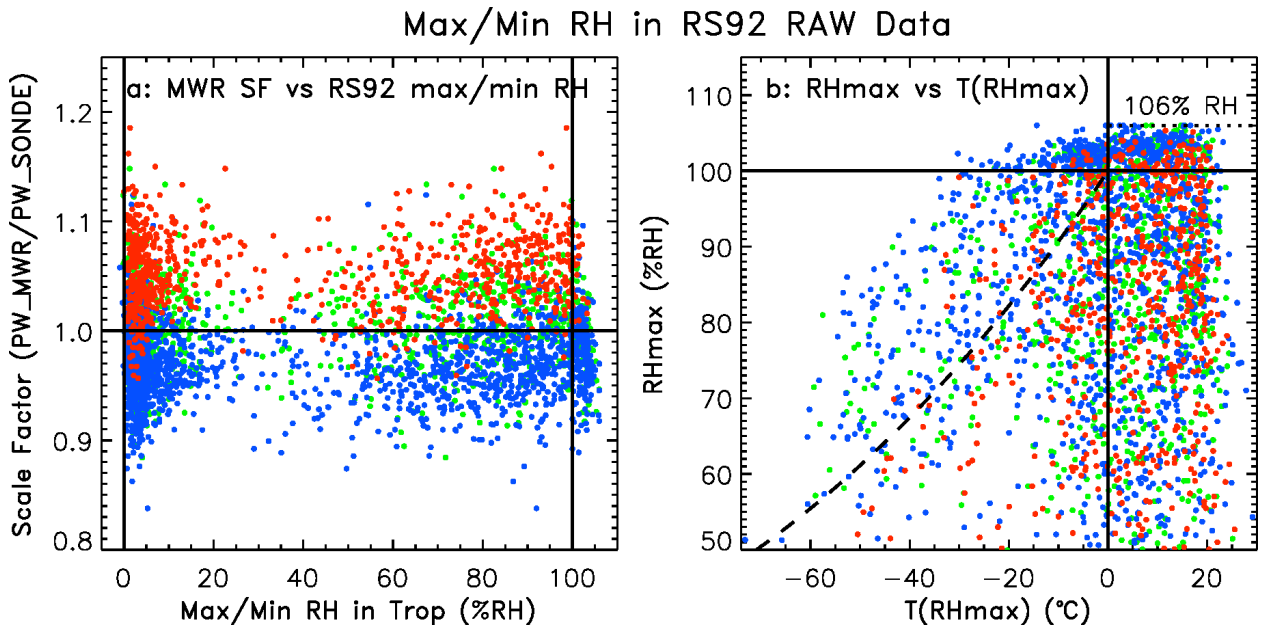


Figure 9. Maximum and minimum RH measurements in the troposphere for each sounding in the 2006-2007 ARM dataset, taken from the RS92 RAW data files that contain measured values above 100% RH or below 1% RH, unlike the EDT and FLEDT files. (a) MWR Scale Factor corresponding to the RS92 max/min RH values. (b) Temperature-dependence of the maximum RH for each sounding. Colors indicate: (blue) nighttime, (red) daytime with $\alpha > 30^\circ$, and (green) daytime with $\alpha < 30^\circ$. Dashed curve is ice-saturation.

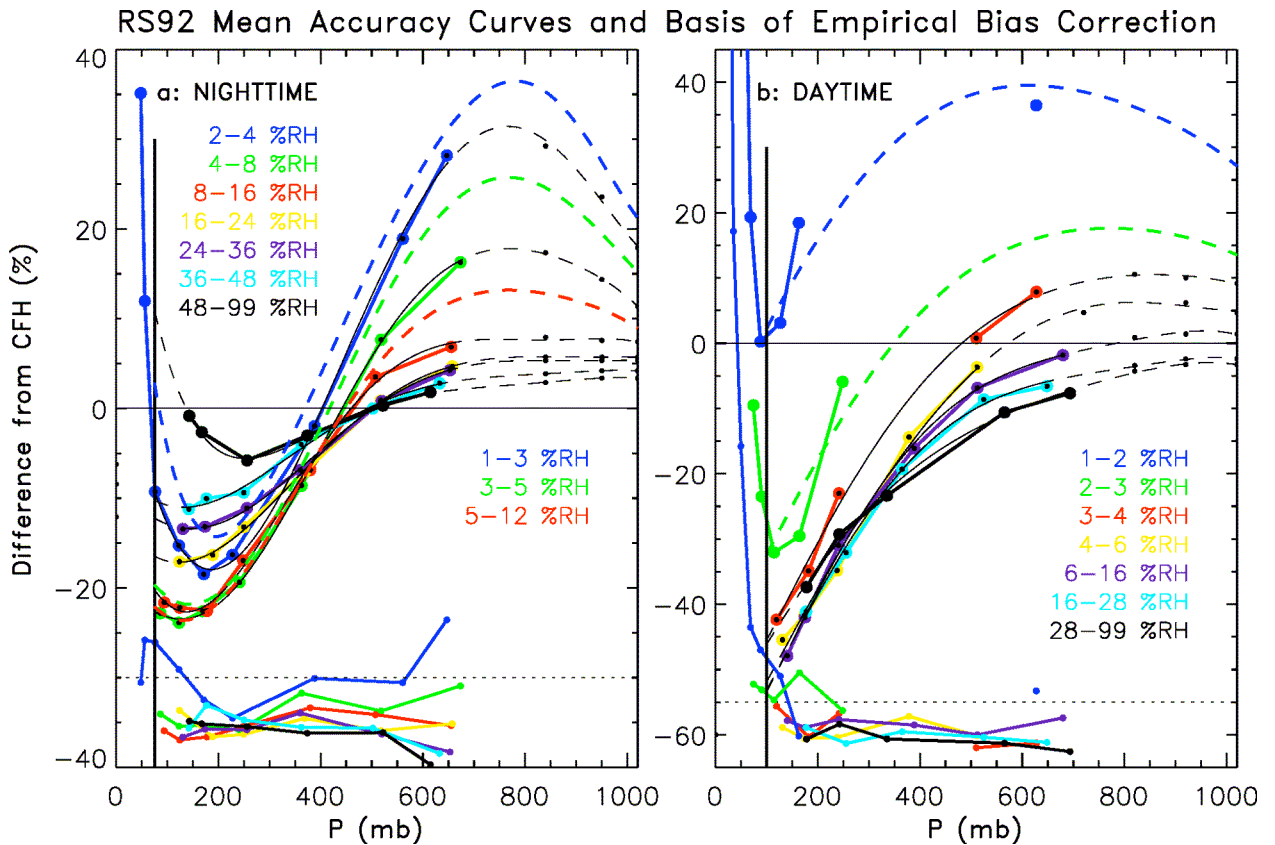


Figure 10. RS92 mean accuracy curves, given by the mean percentage difference from the consensus of the reference sensors (CFH, MWR, and SurTHref) as described in the text, for (a) nighttime soundings, and (b) daytime clear soundings with $\alpha=62-70^\circ$. Polynomials are fit to the CFH comparisons from Figs. 2f and 3f (superimposed black curves), where the fit extrapolations (dashed portion of black curves) are guided smoothly to surface values suggested by the MWR and SurTHref comparisons. Curve fits were added for additional RH intervals (dashed colored curves, and lower legend in panel a), which are described in the text. Curve fits are valid from the surface to a pressure of (night) 75 mb, and (day) 100 mb, as indicated by the vertical black lines. Lower curves give the standard deviation of the percentage difference from CFH, offset to zero at the bottom of the panel for clarity. Coefficients of the curve fits are given in Table 1.

Summary of RS92 Accuracy Relative to CFH – NIGHT

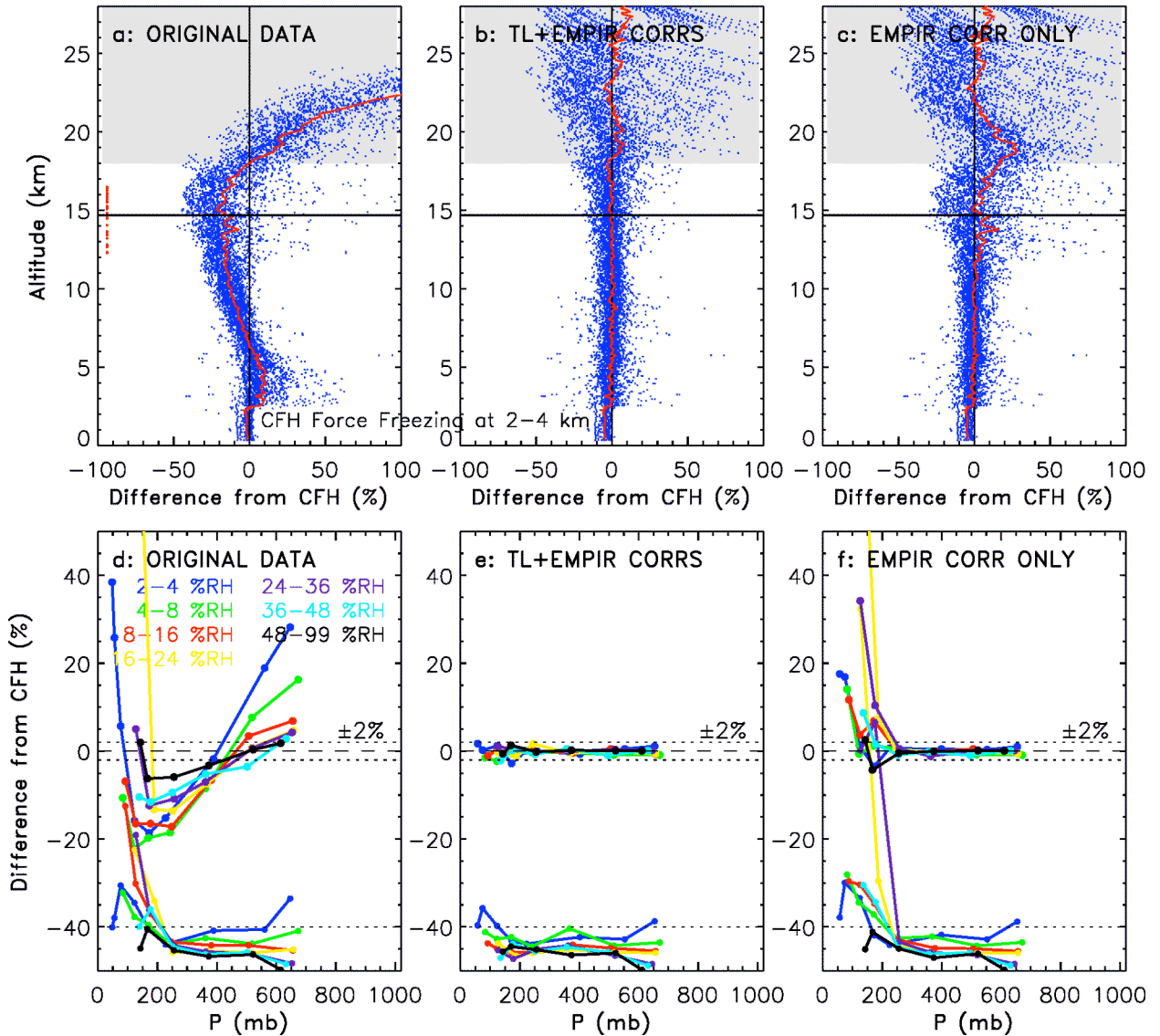


Figure 11. Effect of the time-lag and empirical bias corrections on the Nighttime CFH/RS92 comparison. The percentage difference from CFH is shown for (top row) the measurements and the mean as a function of altitude, and (bottom row) the binned mean difference as a function of pressure in 7 RH intervals. (a and d) Left panels show the percentage difference from CFH for the original RS92 measurements; (b and e) center panels show the difference after applying the time-lag and nighttime empirical bias corrections; and (c and f) right panels show the difference if only the empirical bias correction is applied. Lower curves in d-f give the standard deviation of the percentage differences, offset to zero at the bottom of the panel for clarity. Horizontal line in a-c is the mean tropopause height, red dots are the individual tropopause estimates, and the shaded region is above the upper limit of validity for the correction (level P1 in the text).

Comparison of Corrected RS92 to ARM MWR

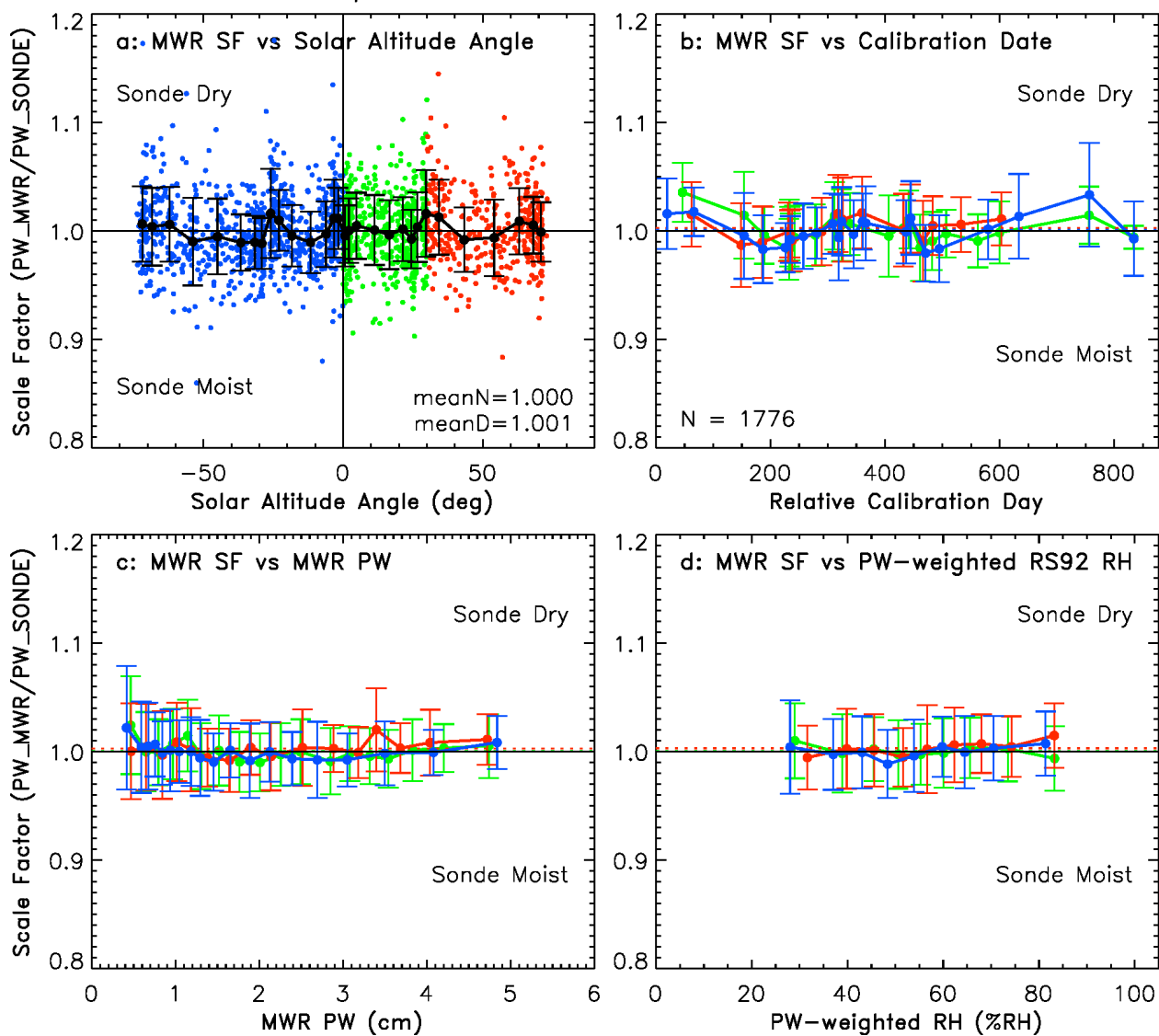


Figure 12. The MWR Scale Factor (SF) for the 2006-2007 ARM SGP dataset after applying the appropriate nighttime or daytime empirical bias correction (and the time-lag correction, although it has no impact on the PW). Shown are the dependences of the MWR SF on: (a) the solar altitude angle at launch; (b) on time as ordered by increasing RS92 calibration date; (c) on PW; and (d) on the PW-weighted RH calculated from the RS92 profile. Blue represents nighttime soundings, red represents daytime soundings with $\alpha > 30^\circ$, and green represents daytime soundings with $\alpha < 30^\circ$.

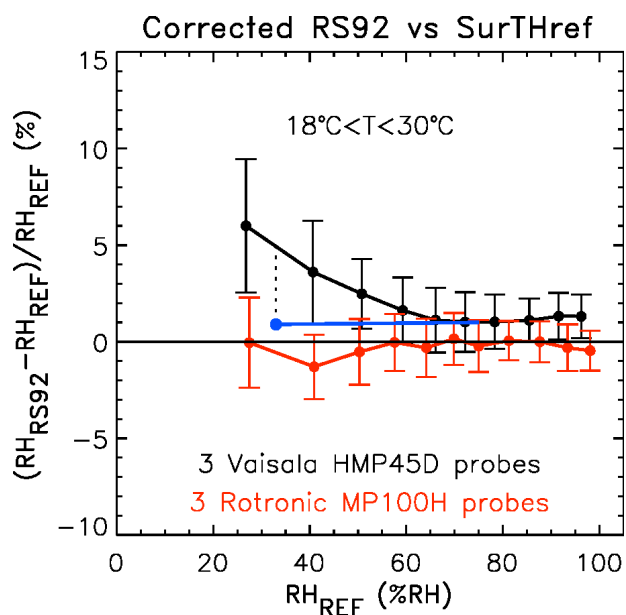


Figure 13. Mean and standard deviation of the percentage difference between RS92 and SurTHref measurements as a function of RH for the 2006-2007 time period at the ARM SGP site, after applying the nighttime empirical bias correction to the RS92 RAW data. Shown are: (red and black) percentage difference from RS92 for each manufacturers' RH probes for cases in the 18-30°C temperature range where the probe calibrations are reliable, and (blue) the difference after accounting for the known calibration bias of 1.35% RH (or 4.1%) for the Vaisala probes at the 33% RH calibration point.

Summary of RS92 Accuracy Relative to CFH – DAY

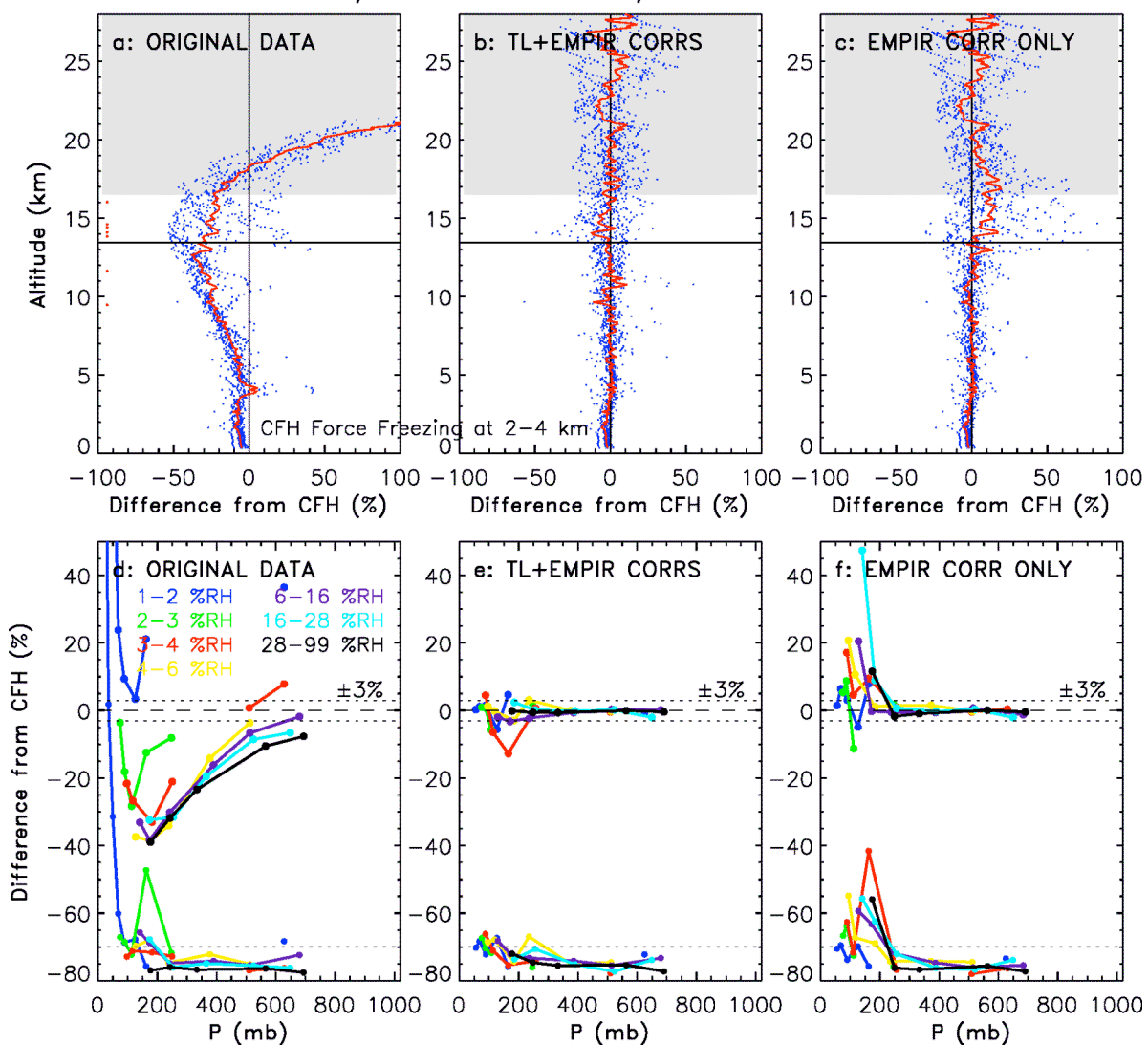


Figure 14. Effect of the time-lag and empirical bias corrections on the Daytime CFH/RS92 comparison. The percentage difference from CFH is shown for (top row) the individual measurements and the mean as a function of altitude, and (bottom row) the binned mean difference as a function of pressure in 7 RH intervals. (a and d) Left panels show the percentage difference from CFH for the original RS92 measurements; (b and e) center panels show the difference after applying the time-lag and daytime empirical bias corrections; and (c and f) right panels show the difference if only the empirical bias correction is applied. Lower curves in d-f give the standard deviation of the percentage differences, offset to zero at the bottom of the panel for clarity. Horizontal line in a-c is the mean tropopause height, red dots are the individual tropopause estimates, and the shaded region is above the upper limit of validity for the correction (level P1 in the text).

Example Corrected RS92 Soundings

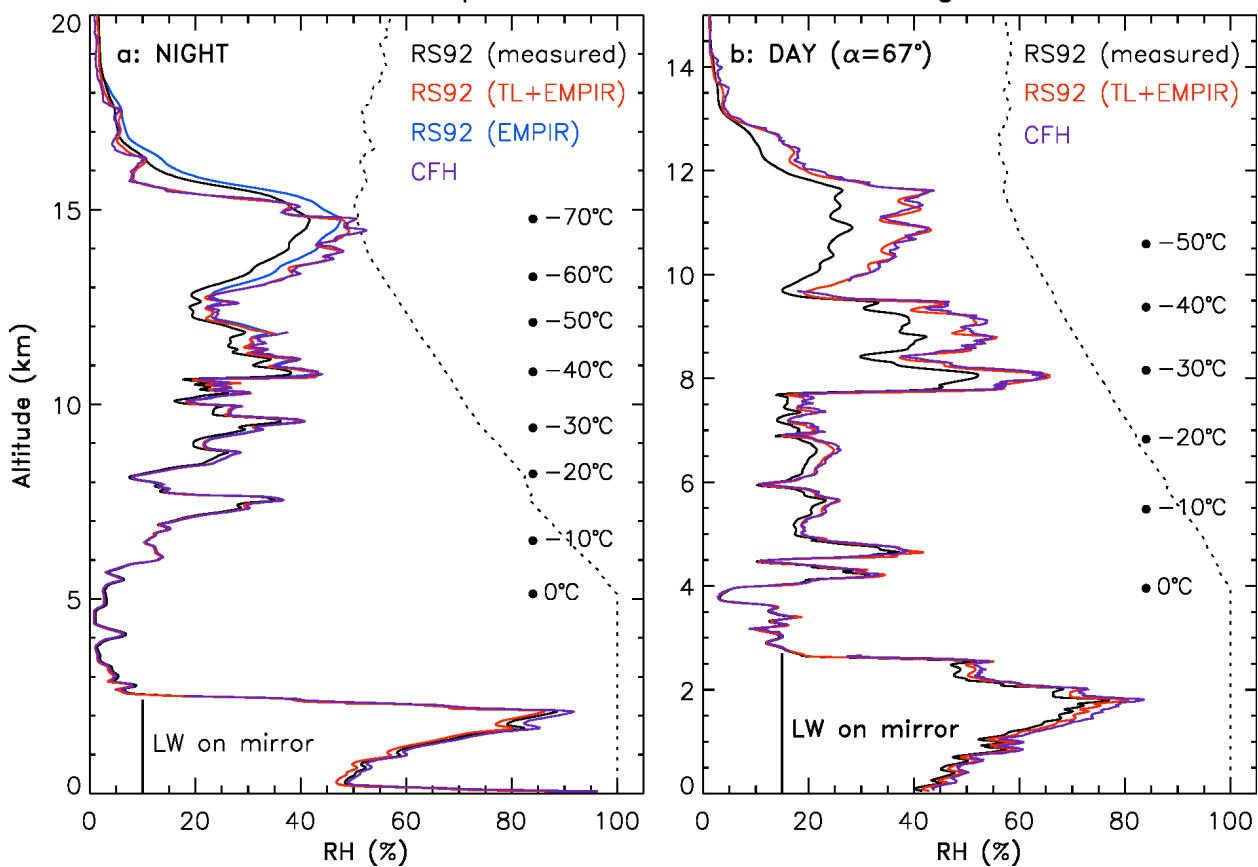


Figure 15. The example (a) nighttime and (b) daytime RS92 soundings from Fig. 1 after applying corrections. Shown are: (black) the measured RS92 profile; (red) RS92 profile after applying the time-lag and empirical bias corrections; (blue) RS92 profile if only the bias correction is applied; (purple) CFH profile; and (dashed) ice-saturation.

Example Corrected RS92 Soundings – Mixing Ratio

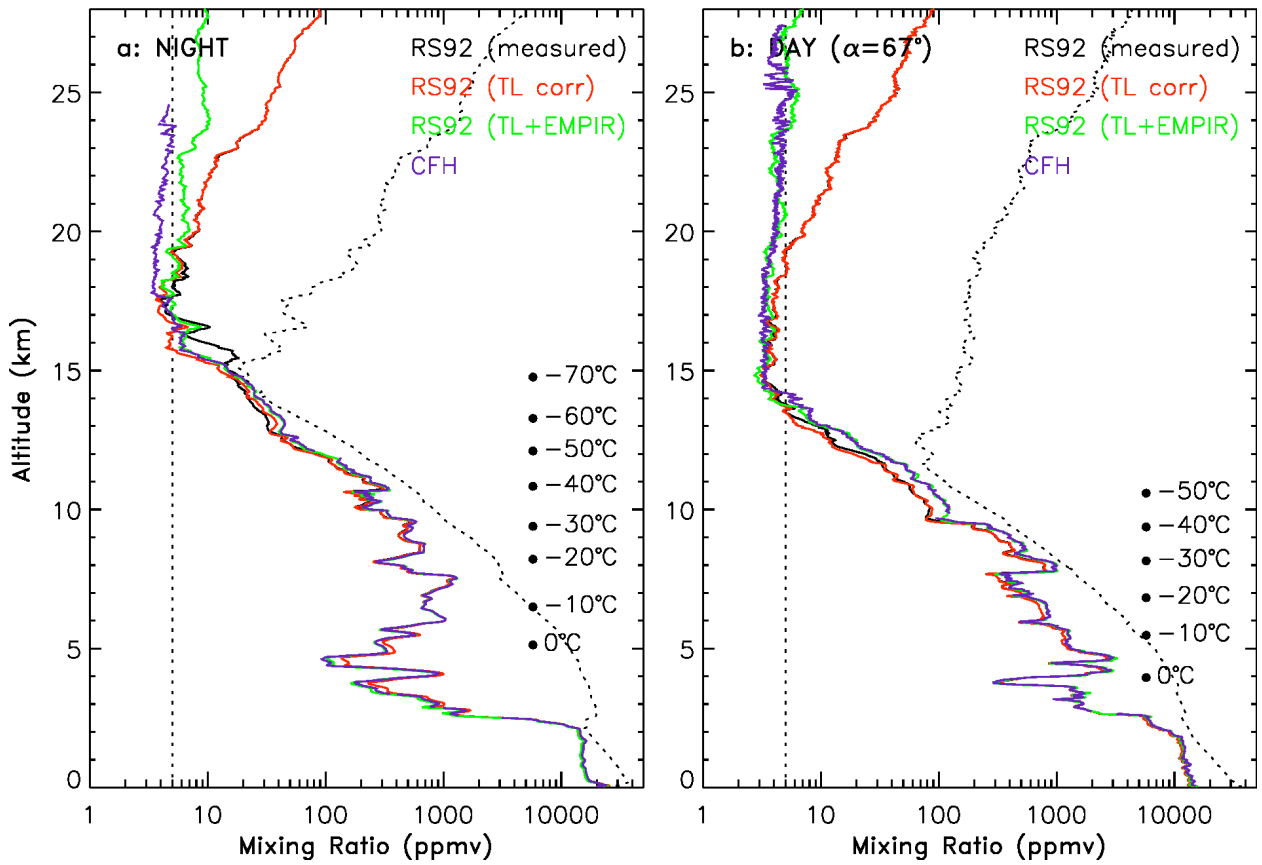


Figure 16. Same CFH and corrected RS92 soundings as in Fig. 15, except in terms of mixing ratio rather than RH. Dashed curve is the saturation mixing ratio, and the tropopause is at the minimum value.

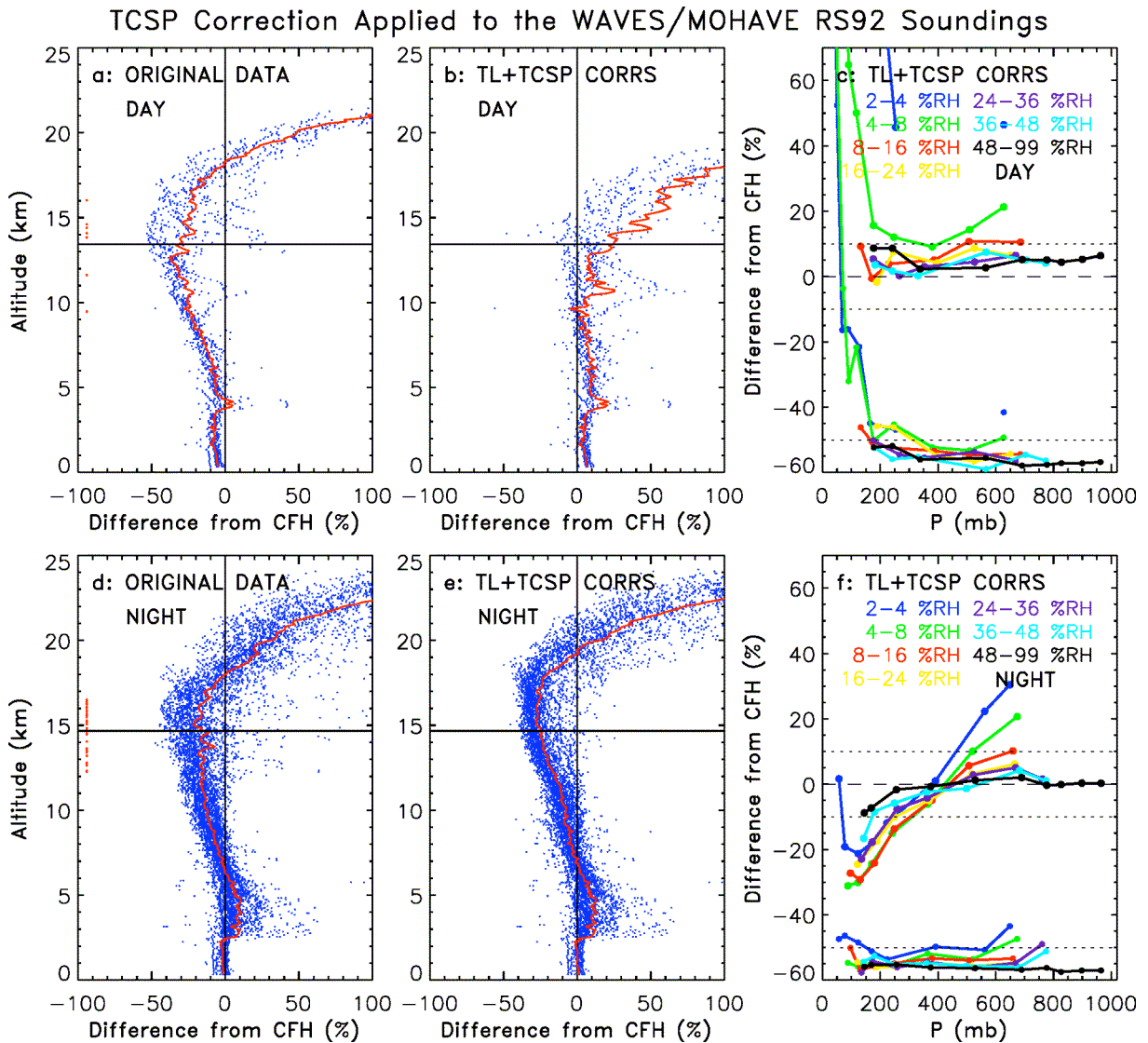


Figure 17. RS92 percentage difference from CFH for the WAVES/MOHAVE dataset after applying an earlier RS92 correction derived from TCSP dual soundings [Vömel *et al.*, 2007]. (top) Daytime comparison, and (bottom) Nighttime comparison. (a and d) Left panels show the altitude dependence of the percentage difference for the original RS92 data. (b and e) Center panels show the altitude dependence of the difference after applying the time-lag and TCSP empirical bias corrections. (c and f) Right panels show the difference after applying the corrections as a function of pressure in 7 RH intervals. Lower curves in panels c and f give the standard deviation of the percentage differences, offset to zero at the bottom of the panel for clarity. Horizontal line in the altitude panels show the mean tropopause height, and red dots are the individual tropopause estimates.