Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements

Larry M. Miloshevich,[1] Holger Vömel,[2] David N. Whiteman,[3] and Thierry Leblanc[4]

Received 1 December 2008; revised 10 March 2009; accepted 1 April 2009; published 13 June 2009.

[1] Relative humidity (RH) measurements from Vaisala RS92 radiosondes are widely used in research and operational applications, but their accuracy is not well characterized as a function of height, RH, and time of day (or solar altitude angle). This study compares RS92 RH measurements to simultaneous water vapor measurements from three reference instruments of known accuracy. Cryogenic frost point hygrometer measurements are used to characterize the RS92 accuracy above the 700-mbar level, microwave radiometer measurements characterize the RS92 accuracy averaged over essentially the lower troposphere, and the RS92 accuracy at the surface is characterized by a system of 6 RH probes with National Institute of Standards and Technology–traceable calibrations. The three RS92 accuracy assessments are combined to yield a detailed estimate of RS92 accuracy for all RH conditions from the surface to the lowermost stratosphere. An empirical correction is derived to remove the mean bias error, yielding corrected RS92 measurements whose bias uncertainty is independent of height or RH and is estimated to be ±4% of the measured RH value for nighttime soundings and ±5% for daytime soundings, plus an RH offset uncertainty of ±0.5% RH that is significant for dry conditions. The accuracy of an individual RS92 sounding is further characterized by the 1-σ ''random production variability,'' estimated to be ±1.5% of the measured RH value. The daytime bias correction must be used with caution, as it is only accurate for clear-sky or near-clear conditions owing to the complicated effect of clouds on the solar radiation error.

Citation: Miloshevich, L. M., H. Vömel, D. N. Whiteman, and T. Leblanc (2009), Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements, *J. Geophys. Res.*, *114*, D11305, doi:10.1029/2008JD011565.

1. Introduction

[2] Atmospheric water vapor measurements are used in a wide variety of both operational and research applications, including input to forecast models and radiative transfer calculations, validation of ground-based and satellite remote sensor retrievals, and development of water vapor and cloud parameterizations, among others. The high vertical resolution of radiosonde measurements is well suited to these measurement needs, except for their 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 their uncertainty, 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 owing to hardware, manufacturing, or calibration changes. The aim of this paper is to characterize the accuracy of Vaisala RS92 radiosonde water vapor measurements as a function of height, RH, and time of day (or solar altitude angle), and then develop and evaluate an empirical correction that removes the mean bias error.

[3] 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 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.* [2007a] (V07 hereafter) characterized the SRE in daytime RS92 measurements at a tropical site, using dual RS92 and cryogenic frost point hygrometer (CFH) soundings. The RS92 measurements had a mean dry bias relative to CFH that increased with height from a 9% relative error near the surface to a 50% relative error at the tropopause for high solar altitude angles ($\alpha > 60°$), and they derived a correction that removes the mean bias as a function of pressure. *Miloshevich et al.* [2006] (M06 hereafter) compared microwave radiometer (MWR) and RS90 radiosonde measurements from the Atmospheric Radiation Measurement (ARM) program, and found that the SRE is 6–8% in terms of precipitable water vapor (PW), which

[1]National Center for Atmospheric Research, Boulder, Colorado, USA.
[2]CIRES, University of Colorado, Boulder, Colorado, USA.
[3]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[4]Table Mountain Observatory, JPL, NASA, Pasadena, California, USA.

Copyright 2009 by the American Geophysical Union.
0148-0227/09/2008JD011565

represents an average of the mean bias over all solar altitude angles and RH conditions in the lower troposphere (LT). M06 also used dual RS92 and CFH soundings to characterize the RS92 mean bias error for nighttime soundings (i.e., zero SRE conditions), and they derived a correction that is dependent on temperature and RH. *Cady-Pereira et al.* [2008] (CP08 hereafter) determined the dependence of SRE on solar altitude angle by comparing ARM MWR PW measurements to column-integrated RS90 and RS92 PW measurements, and they derived a correction that removes the mean bias in the LT as a function of the solar altitude angle. *Miloshevich et al.* [2004] (M04 hereafter) used laboratory measurements of the sensor time-constant as a function of temperature to derive a correction for sensor time-lag error that is caused by slow sensor response at low temperatures, and V07 showed that correcting time-lag error markedly improves the agreement between RS92 and CFH in the UT.

[4] This study will characterize the accuracy of RS92 RH measurements relative to three water vapor reference instruments of known accuracy. Measurements from the RS92 are compared to simultaneous measurements from the CFH, ARM MWR, and calibrated RH probes in section 2. The comparisons are synthesized in section 3 to produce an estimate of the RS92 measurement 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 RS92 data are evaluated in detail.

2. Instrumentation and Data

2.1. RS92

[5] 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, for the purpose of recovering the original factory calibration accuracy. Unlike earlier RS80 radiosondes, RS90 and RS92 radiosondes use dual RH sensors that are alternately heated while the nonheated sensor measures the ambient RH, which eliminates the problem of sensor icing in supercooled liquid water clouds or ice-supersaturated conditions. Vaisala radiosonde RH sensors are described further by M06 and references therein.

[6] The following sources of RS92 measurement error are considered in this study, including bias (systematic) errors, random error, and sensor time-lag error.

[7] 1. Mean calibration bias reflects the absolute accuracy of the Vaisala calibration references and their variation with time, which A. Paukkunen et al. (Accuracy and performance of the new Vaisala RS90 radiosonde in operational use, paper presented at 11th AMS Symposium on Meteorological Observations and Instrumentation, American Meteorological Society, Albuquerque, New Mexico, 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.

[8] 2. Random production variability is the sensor-to-sensor variability relative to the mean calibration accuracy, which reflects such things as manufacturing variability and inhomogeneous conditions within the calibration chamber. The variability is generally described by the standard deviation of differences from the mean for a batch of sensors (the "1-σ variability").

[9] 3. 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 sensor time-constant measurements.

[10] 4. Solar radiation error is a dry bias in daytime measurements caused by solar heating of the RH sensor (V07 and CP08). 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), pressure (P), the angle between the Sun and the sensor normal, the cloud optical depth along a Sun-sensor line, and the transmissivity of the air mass. 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 measured by the temperature sensor is used in the RH calibration equation, but the appropriate temperature is actually the (unknown) temperature of the RH sensor polymer itself. Other related minor measurement errors can occur in the first 100 m or so of flight or when the temperature gradient changes abruptly (e.g., at the tropopause), because the temperature sensor and the (relatively massive) RH sensor have different thermal time constants, and the measured air temperature doesn't accurately represent the temperature of the RH sensor.

[11] 5. Ground check related uncertainty can arise from both improper operator procedures and from the assumption that the desiccant creates an environment of precisely 0.0% RH. The mean and standard deviation of the GC correction for the 2006–2007 ARM RS92 data set used in this study is −0.49 ± 0.35% RH, and these are taken as reliable typical values since ARM and presumably other large operational programs have controls to minimize procedural problems. However, during less routine operations such as field experiments, 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. 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. As an estimate of this uncertainty, the best of a selection of desiccants tested by the Australian Bureau of

Meteorology (a molecular sieve type) achieved a minimum of only 0.5% RH [*Gorman*, 2002].

[12] 6. Roundoff error occurs because the RH measurements in the standard RS92 processed data files ("EDT") are reported as integers, which introduces uncertainty of ±0.5% RH that is substantial for dry conditions (e.g., >10% uncertainty for conditions below 5% RH). The 1% RH resolution of EDT data 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 from the Vaisala FLEDT (floating-point EDT) data files, which are identical to the EDT files except that the RH values are reported with two decimal places of precision (FLEDT files are only available from the Vaisala DigiCora-III data system, version 3.12 and newer).

2.2. CFH

[13] The CFH measures water vapor on the basis of the chilled-mirror principle, whereby 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 dew point temperature (T_d) if the condensate phase is liquid, or the frost point 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 Appendix A of M06). The uncertainty in CFH RH measurements is estimated by *Vömel et al.* [2007b] to be 4% of the measured RH value near the surface, increasing with height to 9% at the tropopause and no more than 10% in the stratosphere, on the basis of the separate uncertainties in the CFH frost point temperature, the RS80 air temperature, and other factors. The CFH frost point measurements are not affected by SRE because, even if the measured air is heated by the sampling tube during its 0.1-s transit time, the frost point temperature depends only on the water vapor concentration and not on the air temperature. Solar radiation only impacts the CFH RH measurements through heating of the RS80 temperature sensor, and this impact is minimized for both RS80 and RS92 temperature measurements by the Vaisala radiation correction that is applied in the standard data processing.

[14] 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%, but a difference of 50% for conditions of 2% RH. Most applications are concerned with relative (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.

[15] This study uses simultaneous measurements from CFH and RS92 launched on the same balloon during four NASA satellite and Raman lidar water vapor validation experiments. Two experiments called WAVES (Water Vapor Validation–Satellite/Sondes) were conducted at the Howard University Beltsville Campus (HUBC) in Beltsville, Maryland, in July and August of 2006 and 2007 (D. N. Whiteman et al., The Water Vapor Variability–Satellite/Sondes (WAVES) field campaigns, 24th International Laser Radar Conference, International Coordination Group for Laser Atmospheric Studies, Boulder, Colorado, 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 (T. Leblanc et al., Measurements of humidity in the atmosphere and validation experiments (MOHAVE, MOHAVE II): Results overview, paper presented at 24th International Laser Radar Conference, International Coordination Group for Laser Atmospheric Studies, Boulder, Colorado, 2008). The CFH/RS92 data set 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. These direct profile comparisons are supplemented by a much larger data set of RS92 comparisons to MWR and calibrated RH probes, and a thorough uncertainty assessment is given in section 5.

[16] 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 ±1.5% of the measured RH value for conditions above 10% RH, ±3% for conditions below 10% RH, and is 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.

[17] 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 Figure 1a have been "smoothed" by time-lag error (black), but 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 the steepness of the troposphere-stratosphere transition that is measured by the CFH (purple). The RS92 difference from CFH after removing time-lag error (red versus purple) is seen to vary as a function of height and RH, and this is the relative bias considered in this study. The relative bias is much greater for daytime than nighttime soundings owing to the SRE, and the daytime and nighttime RS92 measurements will be characterized separately. The time-lag effect in the UT is much smaller for the daytime example than the nighttime example, because the temperature in the UT is much warmer for the daytime example and the RH sensor responds more quickly.

[18] To quantitatively compare RS92 and CFH measurements that are on the same balloon, we adopt the altitude

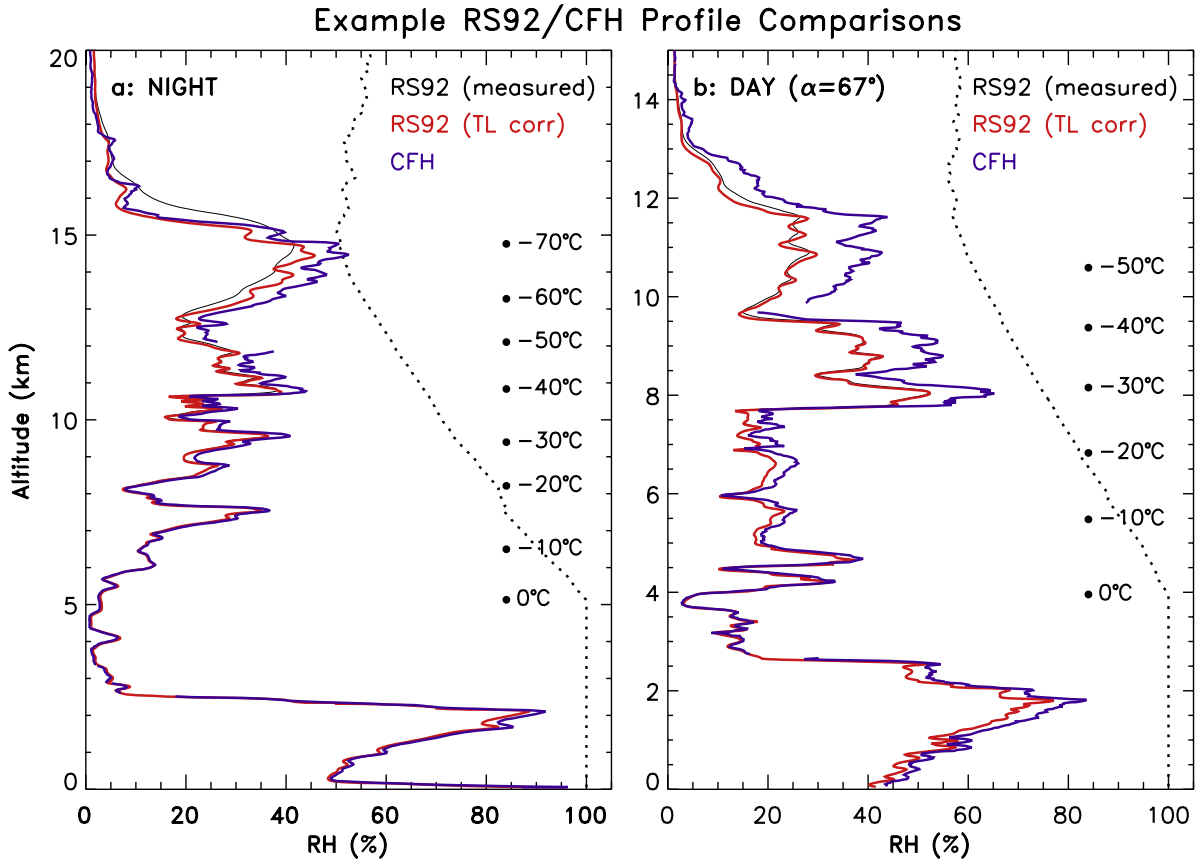

Figure 1. Example dual RS92/CFH soundings (a) at night and (b) at midday. Shown are: the measured RS92 profile (black), the RS92 profile after correcting for time-lag error (red), the CFH profile (purple), and ice saturation (dashed). The CFH forced freezing event is indicated by the gap in the CFH data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

versus time relationship, Z(t), from the CFH RS80 radiosonde as the vertical coordinate, and then align the profiles by interpolating the CFH Z(t) relationship to the RS92 time series, so that both profiles have the same altitude values at a given time from launch. Aligning the profiles in the vertical on the basis of the time from launch is more accurate than using the independent altitude profiles directly, because the clocks in the two radiosondes are more accurate and less variable than are the altitudes derived from separate pressure and temperature sensors.

[19] The RS92 bias relative to CFH is shown for all nighttime dual soundings in Figure 2, as individual altitude profiles in Figure 2 (top), and statistically in Figure 2 (bottom) as the mean and standard deviation of the RS92 relative bias as a function of pressure, shown for seven RH intervals. The RS92 relative bias is shown for the original RS92 FLEDT measurements (Figure 2, left), after applying the time-lag correction (Figure 2, middle), and showing only the data after the CFH forced freezing event (Figure 2, right), which typically occurs at 2.5 to 4.5 km altitude. The RH intervals were chosen to be as narrow as possible while still having data points from multiple sensors in all pressure bins.

[20] The relative mean bias for the original RS92 measurements contains much greater sensor-to-sensor variability (standard deviation of differences) in the UT/LS region than at lower levels (lower curves in Figure 2d, below 300 mbar). The time-lag correction recovers vertical structure in the RS92 profiles and reduces the variability in the UT/LS region to about the same as at lower levels (Figures 2b and 2e). The RS92 typically measures a nearly constant value in the range 1–2% RH in the stratosphere, 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, leading to the very large RS92 mean bias above $\sim$20 km and suggesting that the RS92 becomes insensitive to water vapor above some level. The factors that establish the upper limit for reliable RS92 measurements will be evaluated in section 4.

[21] 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 (Figure 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 with height. The data set 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 Figure 2b is also seen to occur for the separate RH intervals (Figure 2e at about 700 mbar), indicating that the abrupt

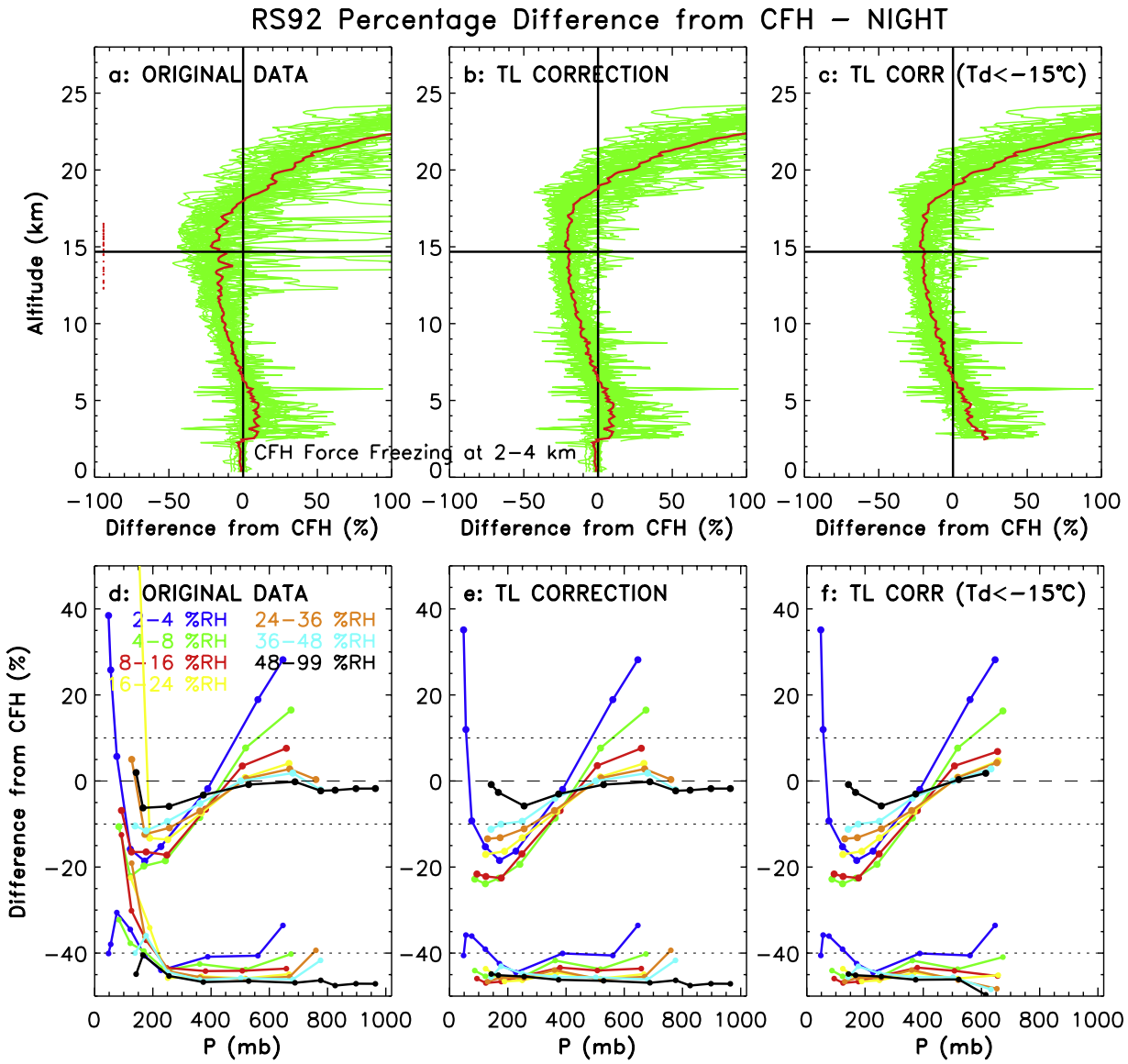

Figure 2. RS92 percentage difference from CFH for the 33 nighttime WAVES and MOHAVE dual soundings, showing (top) the individual altitude profiles and the mean and (bottom) the mean and standard deviation of the binned differences as a function of pressure in seven RH intervals. (a and d) Difference from CFH for the original RS92 FLEDT measurements; (b and e) difference after correcting the RS92 data for time-lag error; and (c and f) data after the CFH forced freezing event. The upper curves in Figures 2d–2f show the RS92 mean bias relative to CFH, and the lower curves show the standard deviation of the bias on the same scale, but offset to 0% at the bottom of the plot for clarity. Horizontal line in Figures 2a–2c is the mean tropopause height, and red dots are the individual tropopause estimates. Horizontal lines in Figures 2d–2f are reference lines.

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 coincides with the CFH forced freezing event during WAVES, and when the data prior to the forced freezing event are excluded (Figures 2c and 2f), then the mean bias above the forced freezing event varies smoothly with height and with RH, suggesting that the distinct offset is attributable to the CFH. Comparison of an ensemble of profiles measured by the CFH and by two Raman lidars during WAVES showed that the CFH measurements were moister than the lidars by a mean of about 7% up to the altitude of the forced freezing event, but above that level the CFH and Raman lidar agreed within about 2%. The CFH also measured 4–5% moister in the mean than the HUBC MWR in terms of PW, where the PW is dominated by water vapor in the lowest few kilometers. These observations are not conclusive, but analysis of data presented in this paper will show that the CFH measurements for some soundings during WAVES had 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 $T_d = -15°C$). Therefore, this RS92 accuracy assessment will use only the CFH frost point

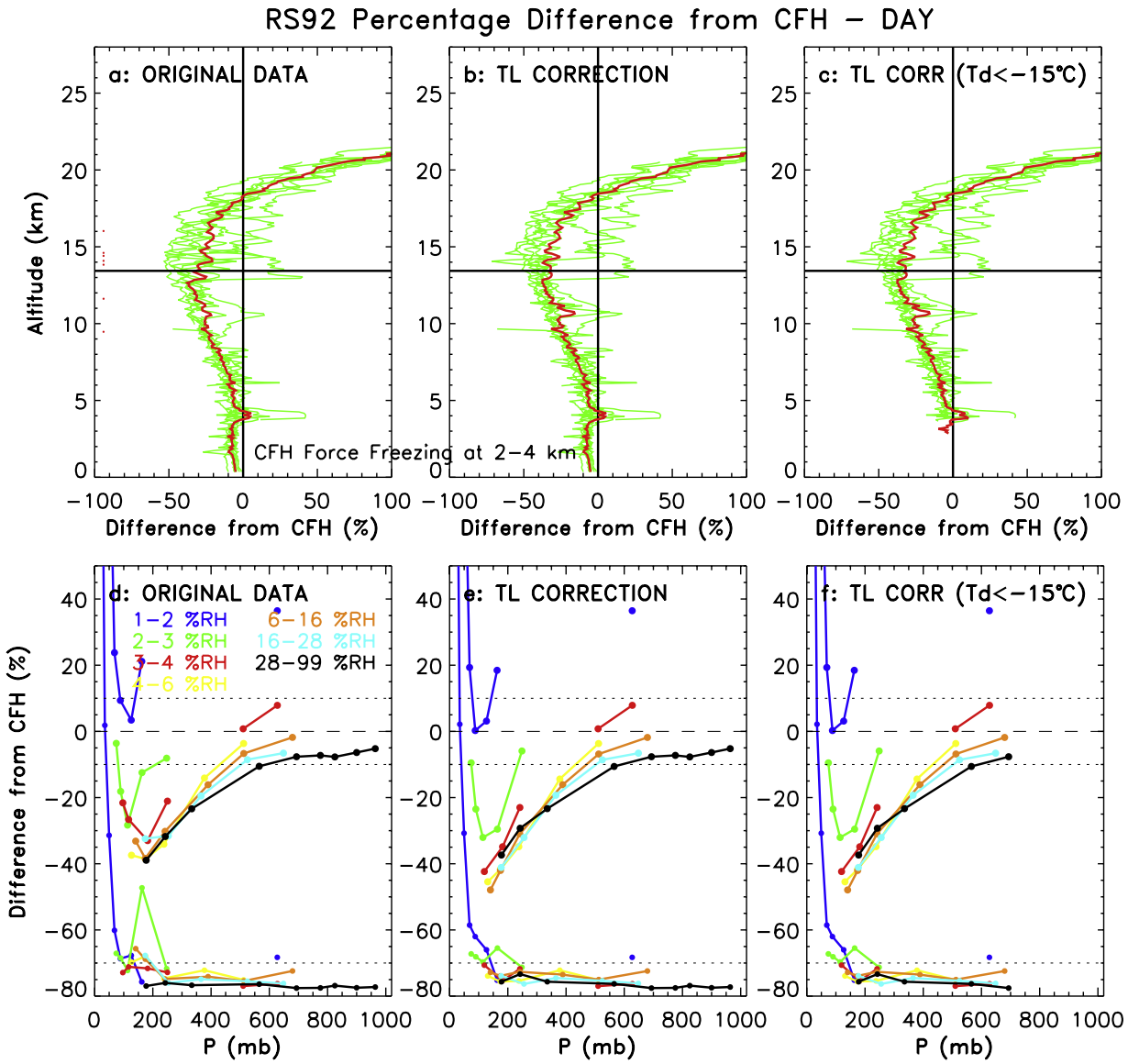

Figure 3. RS92 percentage difference from CFH for seven daytime WAVES and MOHAVE dual soundings, showing (top) the individual altitude profiles and the mean and (bottom) the mean and standard deviation of the binned differences as a function of pressure in seven RH intervals. (a and d) Difference from CFH for the original RS92 measurements; (b and e) difference after correcting the RS92 data for time-lag error; and (c and f) data after the CFH forced freezing event. The upper curves in Figures 3d–3f show the RS92 mean bias relative to CFH, and the lower curves show the standard deviation of the bias on the same scale, but offset to 0% at the bottom of the plot for clarity. Horizontal line in Figures 3a–3c is the mean tropopause height, and red dots are the individual tropopause estimates. Note that the RH intervals are different than in Figure 2.

measurements and not the dew point measurements, as shown in Figures 2c and 2f. This CFH moist bias in the LT during WAVES will be examined further at the end of this section.

[22] The pattern in Figure 2f is suggestive of underlying curve fit error in the Vaisala calibration equation. The curves exhibit an inflection point at about 400 mbar, 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 mbar 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 is used to represent the broad interval 48–99% RH, because finer subintervals of RH do not provide for data contributions in all pressure bins; however, all finer subintervals lie on top of the 48–99% RH curve and are therefore well represented by it, indicating that the calibration accuracy is constant in this RH range apart from the pressure dependence. The standard deviation of the RS92

relative bias is generally 5% or less, except for conditions below about 10% RH where even a tiny RH offset error is a substantial percentage difference.

[23] The RS92 mean bias relative to CFH for nighttime measurements (Figure 2f) represents the RS92 mean calibration accuracy, whereas the RS92 mean bias for the daytime soundings (Figure 3) represents the sum of calibration bias plus solar radiation error, for clear-sky conditions and high solar altitude angles (62°–70°). The dependence of the RS92 bias on solar altitude angle is investigated in section 2.3. The regions of individual altitude profiles that depart substantially from the mean are dry layers, and they illustrate the need to consider the RH dependence of the RS92 bias and not just the pressure dependence. The RS92 daytime bias in Figure 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 mbar rather than 400 mbar, 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-only) bias, apparently shifting the crossover point to 300 mbar and removing the upward curvature in the UT. The RS92 daytime dry bias in Figure 3f increases with decreasing pressure from 8% at 700 mbar for moist conditions (black curve) to 40–50% in the UT, and the standard deviation of the bias is about 5% in the troposphere, similar to the nighttime measurements. The daytime data set is much smaller than the nighttime data set, and therefore contains greater statistical uncertainty.

2.3. ARM MWR

[24] 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 two-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 that is based on monthly retrieval coefficients, where the retrieval is less accurate for conditions that differ significantly from the monthly mean. 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 small brightness temperature offsets that produce 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 uses the MWRRET PHYS PW retrievals, whose mean accuracy based on variability in the brightness temperature offsets at the SGP site in recent years [*Turner et al.*, 2007, Table 2], 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. Relative to the MWRRET PHYS retrievals, the MWRLOS retrievals for the 2-year data set are drier by a mean of 1.2% when PW > 2 cm, and by 3% when PW = 1 cm, and 5% when PW = 0.5 cm (further detail provided as auxiliary material).[1]

[25] 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. We calculated the fractional contribution of each RS92 altitude level to the total-column PW for each sounding in the 2-year ARM data set, and found that 50% of the PW lies below 1.5 ± 0.4 km above ground level (AGL); 75% of the PW lies below 2.8 ± 0.7 km AGL; and 90% of the PW lies below 4.2 ± 0.9 km AGL. The PW is dominated by water vapor in the lowest 1.5 km and 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, the comparison of PW between RS92 and MWR represents an average over the RH conditions in the LT, weighted toward the lower altitudes and moister portions of the profile because these dominate the PW.

[26] 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 the RS92 relative bias by equating the radiosonde and MWR PW [*Turner et al.*, 2003]. A constant correction factor yields scaled radiosonde profiles that better reflect the overall water vapor content in the LT in terms of PW, but it is clear from Figure 2 that the RS92 bias varies with both height and RH. Figure 4 shows the MWR SF for the 2-year data set, and its dependences on: solar altitude angle (Figure 4a), relative RS92 calibration day (Figure 4b), PW (Figure 4c), and the dominant RH that contributed to the PW (Figure 4d). Cases were rejected if the data quality flag in the ARM MWRRET data file was not 0 or 4, or if the LWP was >200 g m^{-2}. Cases were also rejected when atmospheric variability was high and the MWRLOS PW changed by >10% within a 40-min window centered on the radiosonde launch time.

[27] Figure 4a shows that the nighttime RS92 PW measurements (blue) are on average 3% moister than the MWR PW, which represents a 3% overall RS92 moist calibration bias in the LT. This is similar to the 2–3% RS92 moist mean bias relative to CFH at the "top" of the LT for the moister conditions that dominate the PW (i.e., the rightmost points of the black and light blue curves at the 700-mbar level in Figure 2f), indicating general consistency between the MWR and CFH reference sensors. During the day, the SF increases approximately linearly with solar altitude angle until about α = 30° and then remains nearly constant at higher solar altitude angles, except for some smaller features (bumps at 35° and 60°) that are also present in the nighttime data and therefore are probably not caused by SRE. CP08 showed that the general shape of the solar altitude angle dependence reflects the decrease in atmospheric optical depth and consequent increase in solar flux and SRE as the solar altitude angle increases.

[1]Auxiliary materials are available in the HTML. doi:10.1029/2008JD011565.

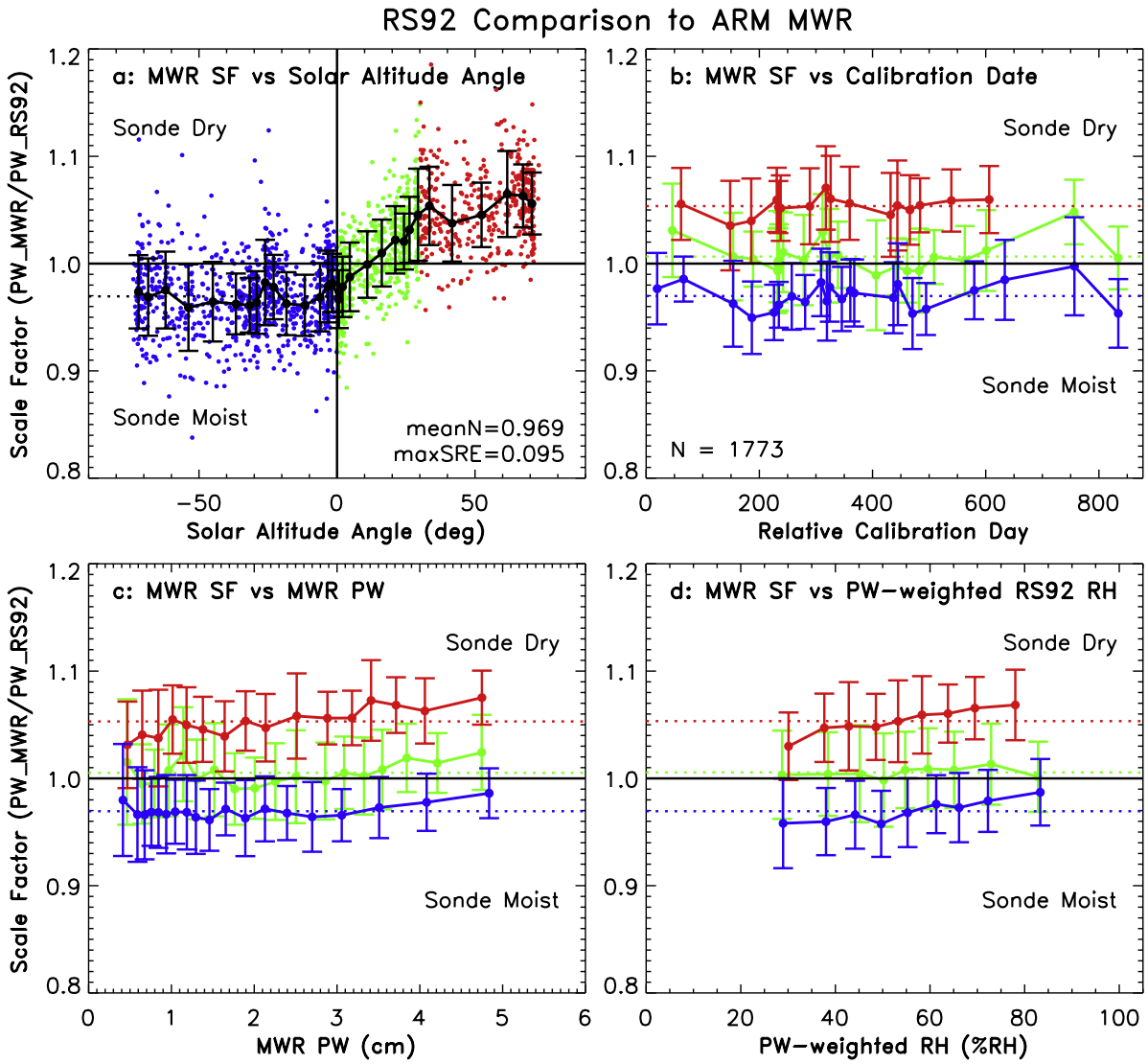

Figure 4. The MWR scale factor (SF) for the 2006–2007 ARM SGP data set. Shown are the dependences of the SF: (a) on the solar altitude angle at launch; (b) on time as ordered by increasing RS92 calibration date; (c) on PW; and (d) on the PW-weighted RH calculated from the RS92 profiles as described in the text. Blue represents nighttime soundings, red represents daytime soundings with $\alpha > 30°$, and green represents daytime soundings with $\alpha < 30°$. The radiosonde calibration date is encoded in the serial number (see Appendix A of M04).

[28] When the cases are ordered by the RS92 calibration day (Figure 4b), it reveals time-dependent variability (drift) in the accuracy of the Vaisala calibration reference by about ±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 Figure 4a represent the same time-dependent variability seen in Figure 4b, 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 Figure 4b remains approximately constant with time, indicating that there was no major change in the reflectivity of the RH sensor or other factors that affect the SRE over the 2006–2007 time period. The irregular spacing of points in Figure 4 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.

[29] The SF also increases somewhat with increasing PW (Figure 4c) and with increasing RH (Figure 4d), where the "PW-weighted RH" reflects the RH conditions that dominated the PW for each sounding, 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 RH_i). A PW dependence in the SF might be caused by a PW dependence in the accuracy of the MWR retrieval and/or by an RH dependence in the RS92 accuracy in the LT. The CFH comparison showed an RH dependence in the RS92 accuracy at the "top" of the LT (700-mbar level in Figures 2f and 3f), and indeed it is shown later that the PW depen-

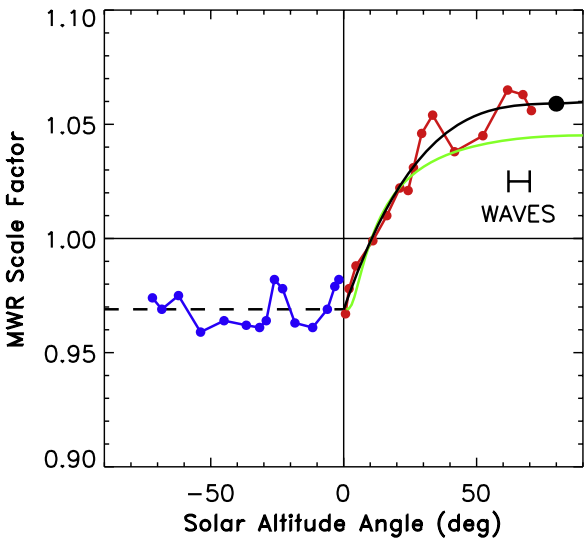

Figure 5. The mean MWR scale factor (SF) binned as a function of solar altitude angle from Figure 4a, where blue represents nighttime soundings, red represents daytime soundings, and the dashed line indicates the mean of all nighttime data. Also shown are: a polynomial curve fit to the daytime data (black curve), the results of *Cady-Pereira et al.* [2008] for earlier ARM RS90 and RS92 radiosondes (green curve), and the 9% mean day/night difference in the LT for high solar altitude angles found by V07 at a tropical site (black dot). The range of solar altitude angles for the WAVES daytime soundings is also shown.

dence of the SF in Figure 4c is a consequence of RH dependence in the RS92 accuracy.

[30] This study complements the study of CP08, which characterized the mean bias of ARM SGP RS90 and RS92 radiosondes relative to ARM MWRRET PHYS retrievals for the 2001–2005 time period that just precedes this study. Both studies considered RS92 radiosondes in the time period 2004–2007, and both find that the nighttime mean bias is about +3% and it varies with time by about ±2% (Figure 4b, and CP08's Figure 3). Figure 5 shows the dependence of the SRE (i.e., the mean day/night difference) on solar altitude angle from this study (black curve), from CP08 (green curve), and from V07 (black dot). Results from this and the CP08 study are very similar for solar altitude angles <30°, but this study finds a larger SRE by 1–1.5% for high solar altitude angles. One possible explanation for this small difference is that the CP08 study primarily represents RS90 sensor characteristics while the present study represents RS92 characteristics, and the sensors may differ in some detail that affects the SRE, such as the reflectivity of the sensor coating. CP08 also demonstrated that their results derived from SGP data also applied to the high solar altitude angles and high PW conditions of the tropics, and this conclusion presumably applies to the present study as well.

2.4. ARM Reference Humidity Probes

[31] The performance of the RS92 at the surface is evaluated by comparing RS92 measurements prior to launch with measurements from the ARM Surface Temperature/Humidity Reference (SurTHref) system (Figure 6). The SurTHref consists of a standard National Weather Service naturally ventilated instrument shelter ("Stevenson screen"), which also contains an inner fan-ventilated chamber with six T/RH probes (three Vaisala HMP45D and three Rotronic MP100H). The ARM prelaunch procedure involves placing the sensor arm of the active radiosonde near the reference probes, which activates an optical switch that causes a flag to be set in the SurTHref data stream indicating the presence of a radiosonde. The RS92 RAW data files, which contain the source measurements for the processed data files. also contain the RS92 data that precede the launch time. The RS92 prelaunch data are correlated with the SurTHref time series and used to characterize the RS92 accuracy relative to the RH probes for the ambient conditions 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.

[32] One strength of the RH probes as reference sensors is that their calibration accuracy is known. Both manufacturers characterized the probe calibration accuracy at four RH test values of approximately 0, 11, 33, and 75% RH, and at an ambient temperature in the range 22°–27°C. The RH test conditions are created using standard salt solutions that produce a specific RH environment that is known with a NIST-traceable uncertainty of ±0.6% RH. The deviation of the probe RH measurements from the four RH test values are specified in the calibration report, and those deviations represent the probe calibration bias relative to the "true RH"

Figure 6. The ARM SurTHref system, showing the outer NWS-type instrument shelter and inner ventilated chamber containing six T/RH probes from two 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 data stream that indicates the presence of a radiosonde.

(which is known within ±0.6% RH). All Rotronic probes deviated from each of the four RH test values by <0.2% RH, meaning that the Rotronic measurements differed from the true RH by <0.8% RH at the four test values, where these systematic errors are added rather than summed in quadrature. 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 at the 33% RH test value the Vaisala probes measured drier than the reference standard by 1.35% RH (i.e., up to 1.95% RH drier than the true RH), so the Rotronic probes are the more accurate reference sensors at 33% RH.

[33] A shortcoming 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 shortcoming is that the RH probes exhibit considerable temperature dependence in the accuracy of the RH calibration (L. M. Miloshevich et al., New surface meteorological measurements at SGP, and their use for assessing radiosonde measurement accuracy, paper presented at 14th ARM Science Team meeting, Atmospheric Radiation Measurements Program, U.S. Department of Energy, Albuquerque, New Mexico, 2004, available at http://www.arm.gov/publications/proceedings/conf14/extended_abs/miloshevich_lm.pdf), 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 measurements 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.

[34] Figure 7a illustrates the methodology of comparing time series of RS92 RAW data and SurTHref data during the prelaunch time period. One minute averages are calculated for the RS92 and RH probe data, then the three 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 1-min differences between RS92 and each manufacturer's probes are then averaged to give a single value for the mean difference between RS92 and each manufacturer's probes during the comparison period, as well as the mean difference between the Vaisala and Rotronic probes.

[35] The mean and standard deviation of the percentage difference between the two manufacturers' probes is shown as a function of RH in Figure 7b. 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 do not provide as reliable of an accuracy standard for those conditions. The SurTHref data give no information below about 20% RH, because those surface conditions did not occur at SGP when the temperature was 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 for dry conditions is large in terms of a percentage (red curve).

[36] The mean percentage difference between RS92 and the Vaisala or Rotronic RH probes is shown as a function of RH in Figure 7c. 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 relative to MWR at night, and also with the magnitude and trend of the nighttime RS92 mean bias relative to CFH for moist conditions (black curve in Figure 2f). The constancy of the bias above about 50% RH at the surface is also consistent with the constant bias above 48% RH relative to CFH above the 700-mbar level. Both manufacturers' probes indicate that the RS92 bias increases with decreasing RH below about 40% RH, as also seen in the comparison to CFH at 700 mbar. The standard deviation of the RS92 percentage difference from the RH probes is about ±2% for conditions above 50% RH (Figure 7c), which is consistent with the ±1.5% RS92 1-σ random production variability estimated from the MOHAVE dual RS92 soundings, after considering that part of the standard deviation in Figure 7c is attributable to the RH probes.

2.5. RS92 RAW Data

[37] The Vaisala data processing restricts RH values in the RS92 EDT and FLEDT data files to the range 1–100% RH, whereas the RAW data files contain measured RH values that are above 100% RH or below 0% RH, which gives information about the RS92 calibration accuracy at the expected physical limits of 0% RH and 100% RH. Figure 8a shows the maximum and minimum RH values in the troposphere for each sounding in the 2-year data set, and Figure 8b shows the maximum RH as a function of the temperature where it occurs. Minimum values are very close to 0% RH, as expected since the GC correction is applied at ARM, but maximum values up to 106% RH are frequently measured by the RS92. The maximum RH decreases with decreasing temperature from 106% RH above 0°C to about 100% RH at −35°C, which is consistent with the pressure-dependent trend of the RS92 mean bias relative to CFH for moist conditions. However, the maximum RH values at temperatures below about −35°C do not describe the RS92 accuracy because they are limited by the homogeneous ice nucleation process, where micron-size aerosols freeze spontaneously at humidities below water saturation, and their growth limits the vapor supply and the maximum RH [e.g., *Heymsfield and Miloshevich*, 1995].

[38] The distribution of points above 100% RH represents both a moist mean bias in the RS92 calibration accuracy at 100% RH, plus the random production variability. One interpretation of Figure 8b above 0°C is that it represents an RS92 mean calibration bias of +3% with a 2-σ variability of ±3%, which is consistent with the RS92 bias for moist conditions relative to the RH probes, the CFH at 700 mbar, and the MWR. Note that the data processing limits the RH value in the RAW data file to 106% RH, but this limit has

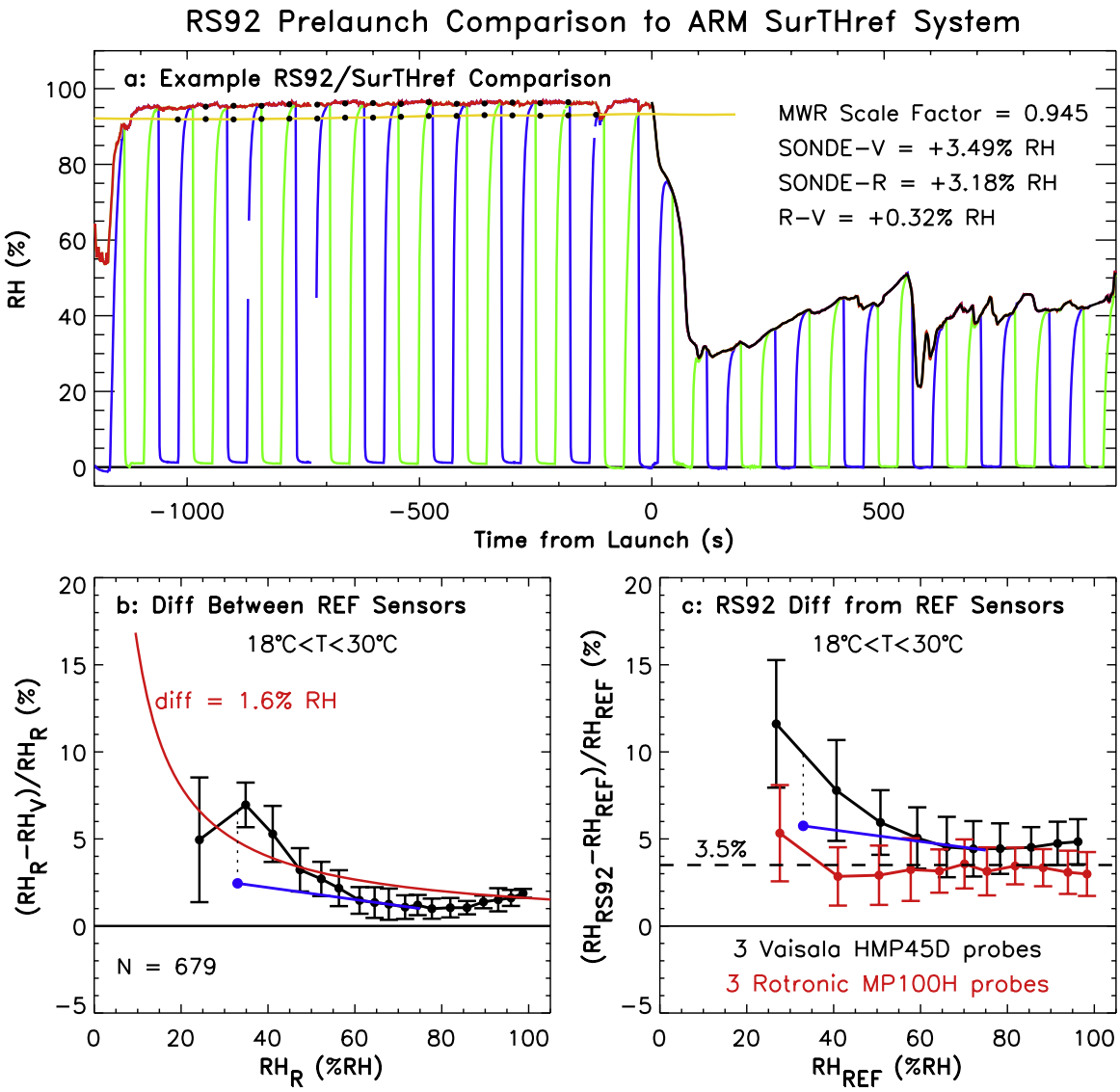

Figure 7. Summary of RS92 comparisons to the ARM SurTHref RH probes for the 2006–2007 time period at the SGP site. (a) Time series of an example comparison, showing: RS92 RAW data (red curve); average of the three Vaisala RH probes ("V") and also the average of the three Rotronic probes ("R"), where the two are indistinguishable on this plot (yellow curve); processed RS92 FLEDT data beginning at the launch time (black curve); and measurements from the two alternately heated RS92 RH sensors that are combined to yield the RAW and FLEDT RH data (blue and green curves) (the measurement portion of the cycle underlies the red and black curves). Dots show 1-min averages. (b) Mean and standard deviation of the percentage difference between the two manufacturers' RH probe measurements as a function of RH for the 2-year data set, restricted to cases in the 18°–30°C temperature range (black). The maximum expected difference of 1.6% RH between the manufacturers' NIST-traceable calibrations, shown as a percentage (red). 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 (blue). (c) Percentage difference between RS92 and the RH probes as a function of RH, also showing the effect of accounting for the known calibration bias of the Vaisala probes.

little impact on the above analysis because only 12 of the 628 data points above 100% RH were limited by the 106% RH threshold.

2.6. CFH in the LT During WAVES

[39] Strong justification is needed for the decision to include only the CFH frost point measurements and not the dew point measurements in this analysis. This decision was taken on the basis of a 4–7% CFH moist mean bias relative to the HUBC MWR and two Raman lidars during WAVES, and the distinct offset in the comparison to RS92 that coincides with the forced freezing event (Figures 2b and 2e), and the promise of further evidence that was presented in this section. The CFH measurements during WAVES were sometimes as

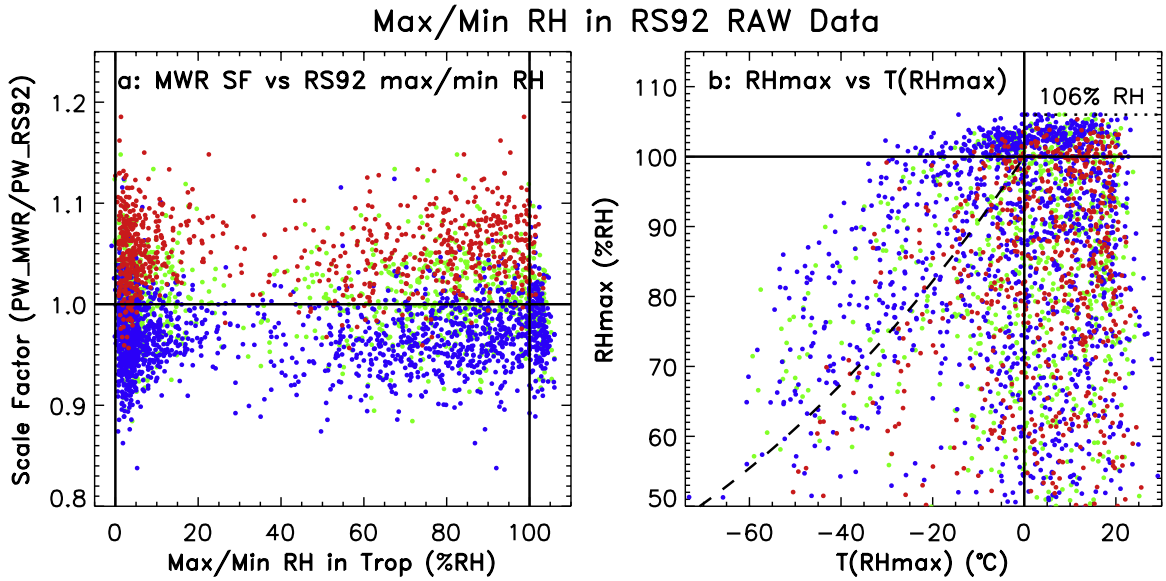

Figure 8. Maximum and minimum RH measurements in the troposphere for each sounding in the 2006–2007 ARM data set, taken from the RS92 RAW data files that also contain measured values above 100% RH or below 1% RH, unlike the EDT and FLEDT files. (a) MWR scale factor corresponding to the RS92 maximum/minimum RH values. (b) Temperature dependence of the maximum RH for each sounding. Colors indicate: nighttime (blue), daytime with $\alpha > 30°$ (red), and daytime with $\alpha < 30°$ (green). Dashed curve is ice saturation.

high as 107% RH, which exceeds water saturation by more than the estimated instrumental uncertainty of 4% in the LT. The ARM MWR and RH probes both indicated an RS92 moist mean bias of 3–4% for moist conditions, as does the RS92 RAW data at 100% RH on physical grounds, in contrast to the CFH that indicated an RS92 dry mean bias of −2% for moist conditions in the LT (Figure 2e). The other reference sensors are consistent within 1–2% about the RS92 mean bias in the LT, which points to a CFH moist mean bias of about 5% when measuring the dew point as the only single factor that would explain all the observations.

3. RS92 Accuracy Assessment

[40] A best estimate of the RS92 mean bias as a function of pressure and RH is derived in this section by combining the three RS92 accuracy assessments relative to the reference instruments. The CFH provides an RH-dependent accuracy assessment above the 700-mbar level, and the SurTHref provides an RH-dependent accuracy assessment at the surface. 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.

[41] The combined RS92 mean bias characterization for nighttime soundings is shown as a function of P and RH in Figure 9a, in terms of polynomial curve fits to the CFH comparisons (superimposed black curves). Three additional curve fits for overlapping RH intervals are included to better represent the strong RH dependence for dry conditions (dashed colored curves). The extrapolation of the curve fits from 700 mbar to the surface (dashed portion of black curves) was constrained by several factors. The RH probes indicate a nighttime RS92 moist mean bias of 3.5% at the surface for RH > 40% (Figure 7c), which is consistent with the 3% nighttime RS92 moist bias relative to MWR that represents the moister RH conditions in the LT (Figure 4a). The MWR comparisons also indicate that the RS92 nighttime bias increases with decreasing RH (Figure 4d), which is consistent with the RH dependence of the bias relative to CFH at 700 mbar and with the few SurTHref measurements for RH < 40%, indicating that this trend with RH continues in the LT. Unlike the curve fits for the moister few RH intervals, the curve fits for the drier conditions are not so directly constrained by the SurTHref and MWR comparisons, except for the general trend of increasing bias with decreasing RH. A constraint based on proportional spacing of the curves was used to maintain the RH dependence of the CFH comparisons at 700 mbar through the LT to the surface. Given the subjectivity of this constraint, the weak portion of this RS92 assessment is dry conditions in the LT (approximately RH < 25% and P > 700 mbar).

[42] The daytime RS92 mean bias curves are similarly constructed (Figure 9b). The MWR comparison at high solar altitude angles indicates that the RS92 mean dry bias is about 6% and the solar radiation error (day/night difference) is about 9%, averaged over the moister conditions in the LT. Two additional curve fits are included in Figure 9b for extremely dry conditions (RH < 3%), where these are guided by the sparse data but largely determined by the proportional spacing constraint and therefore are more uncertain.

[43] The curve fits in Figure 9 give a best estimate of the RS92 mean bias relative to the reference instruments. The coefficients of the curve fits are given in Table 1, where each curve gives the pressure dependence of the RS92 mean bias at an RH value that is nominally equal to the center of the RH

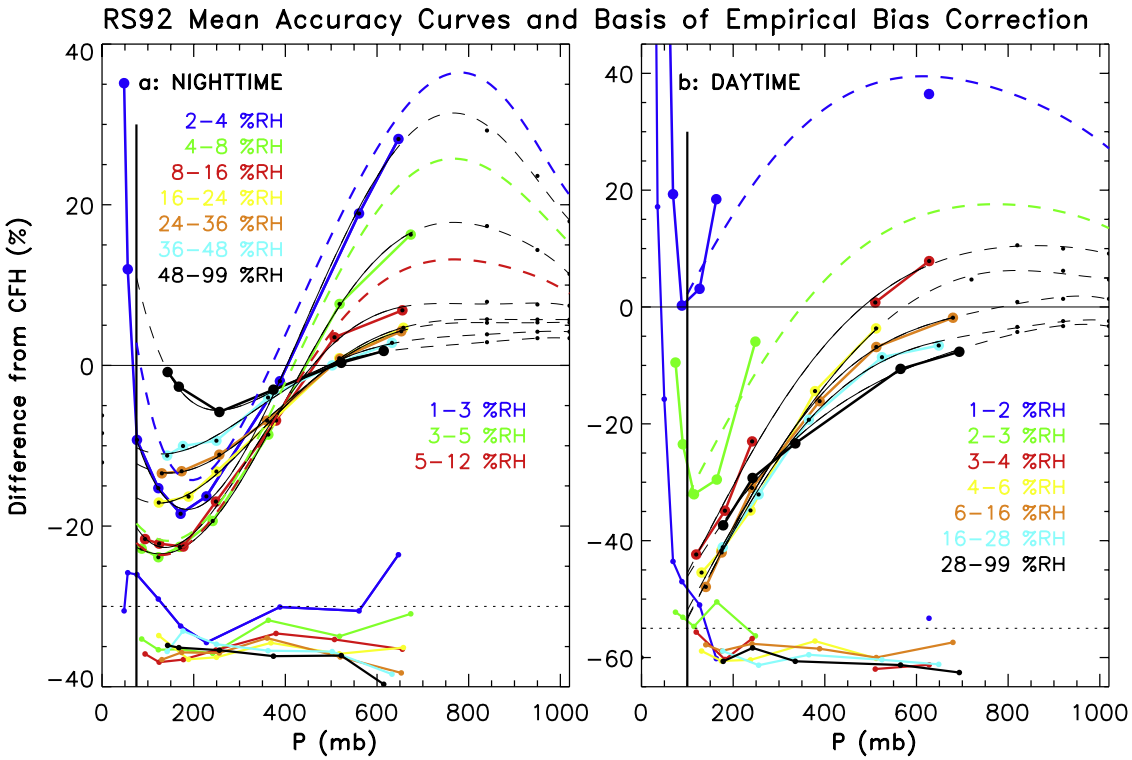

Figure 9. The RS92 mean percentage bias relative to the consensus of the reference sensors (CFH, MWR, and SurTHref) as described in the text, for (a) nighttime soundings and (b) daytime clear-sky soundings with α = 62–70°. Polynomials are fit to the CFH comparisons from Figures 2f and 3f (superimposed black curves), and the fit extrapolations (dashed portion of black curves) are constrained by the MWR and SurTHref comparisons. Curve fits were added for additional RH intervals (dashed colored curves, and lower legend in Figure 9a), which are described in the text. Curve fits are valid from the surface to a pressure of (night) 75 mbar and (day) 100 mbar (vertical black lines). Lower curves give the standard deviation of the percentage difference from CFH, offset to 0% at the bottom of the plot for clarity. Coefficients of the curve fits are given in Table 1.

interval. The curve fits are valid for all RH conditions from the surface to a pressure level, P1, equal to 75 mbar at night and 100 mbar during the day (vertical black lines in Figure 9). The level P1 is the upper limit of reliable RS92 data for this data set, and it corresponds to a location several kilometers above the tropopause. This limit and its cause are discussed at the end of section 4.

4. Empirical Mean Bias Correction

[44] The curves in Figure 9 and Table 1 describe the RS92 mean bias relative to the reference sensors. These same curves are used in this section to develop an empirical correction that removes the mean bias, with the result that the accuracy of corrected RS92 measurements will primarily be given by the random production variability, plus residual bias uncertainty related to the accuracy of the reference sensors. A detailed accuracy assessment for corrected RS92 data is given at the end of the section.

4.1. Algorithm Implementation

[45] The empirical bias correction is applied as a scale factor like the MWR SF, but this scale factor varies as a function of P and RH. The RS92 mean bias in Figure 9 and Table 1, $F(P, RH)$, refers to time-lag-corrected RH measurements (RH_{TLAG}), so 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 shown later. The correction procedure involves calculating the correction factor $G(P, RH)$ from the RS92 mean bias curves in Table 1 for all RS92 RH measurements from the surface to the level P1, where $F(P, RH)$ is given by linear interpolation of RH between the appropriate pressure-dependent curve fits.

[46] Daytime soundings require another correction step, because the curve fits represent the RS92 mean bias for just the 66° mean solar altitude angle of the WAVES CFH/RS92 soundings, or $F(P, RH, 66°)$. The SRE component of the daytime mean bias is equal to the difference between the daytime and nighttime mean biases, $SRE(66°) = F(P, RH, 66°) - F(P, RH, night)$. The dependence of SRE on solar altitude angle is calculated from the polynomial curve fit for the dependence of the MWR SF on solar altitude angle (Figure 5),

Table 1. Coefficients for the Polynomial Curve Fits in Figure 9 That Give the RS92 Mean Percentage Bias as a Function of Pressure for the Listed RH Values That Represents Each RH Interval in Figure 9[a]

	Coefficients						
RH or Fit	a0	a1	a2	a3	a4	a5	a6
				Night[b]			
≤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
for P < P2	4.3867e+3	−3.7335e+2	1.2676e+1	−1.9717e−1	1.1628e−3		
				Day[c]			
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	
for 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			

[a]Coefficients are denoted as a_i; function of pressure is F(P), in millibars. Separate fits describe night versus 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: $F(P) = \sum_{i=0}^{N} a_i \cdot P^i$, where N 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 (frac SRE) as a polynomial function of the solar altitude angle in degrees (α), from Figures 5 and 10, respectively. 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.

[b]P1 = 75 mbar; P2 = 45 mbar; F2 = 50.

[c]P1 = 100 mbar; P2 = 50 mbar; F2 = 80.

where $SRE(\alpha)$ is the day/night difference between the $SF(\alpha)$ curve fit and the mean nighttime SF of 0.97. Figure 10 shows $SRE(\alpha)$ as a fraction of (or normalized to) the SRE at $\alpha = 66°$, i.e., $SRE(\alpha) = SRE(66°) \times fraction(\alpha)$. For example, the SRE at $\alpha = 16°$ is half its magnitude at $\alpha = 66°$, so the SRE for a sounding with $\alpha = 16°$ at launch is taken as half the SRE at $\alpha = 66°$ for any given P and RH conditions, thereby coupling the CFH and MWR assessments of the RS92 SRE. Then the RS92 mean bias for daytime soundings is given by adding the solar radiation error back to the nighttime (or zero SRE) mean bias, $F(P, RH, \alpha) = F(P, RH, night) + SRE(\alpha)$, from which the correction factor $G(P, RH, \alpha)$ is calculated and multiplied by the time-lag-corrected (or measured) RH values. Note that the daytime empirical correction is based on and therefore is only valid for clear-sky soundings, because clouds affect the sensor heating and the SRE in a complicated way, as discussed later.

[47] Finally, although the RS92 accuracy assessment is only valid up to the level P1, the mean bias relative to CFH in the stratosphere does have some meaning and is also specified in the correction algorithm, in part simply to avoid outrageous results if the mean bias curves are applied above level P1. A single P-dependent curve fit to the stratospheric data (essentially the continuation of the solid blue curve in Figure 9) is given in Table 1 as the RS92 mean bias above a level, P2, equal to 45 mbar at night or 50 mbar during the day. The RS92 mean bias relative to CFH is several hundred percent at low pressures, because the RS92 is unresponsive and measures roughly constant values of 1–2% RH in the stratosphere, while the CFH continues to measure the generally decreasing RH with 10% accuracy. The RS92 mean bias above P2 merely represents a climatology of the WAVES and MOHAVE stratospheric CFH measurements, and is not meaningful for stratospheric research. To complete the correction algorithm, a transition must be made from the RH-dependent bias value at P1 to the fixed bias value at P2. An ad hoc yet suitable approach is to specify the RS92 bias in the transition interval, $F_T(P)$, as the left half of a parabola that connects the RH-dependent bias value at P1 to the fixed bias value at P2. Thus, the RS92 bias in the transition interval P1 to P2 is taken to be: $F_T(P) = F1 + A(P - P1)^2$, where $A = (F2 - F1)/(P2 - P1)^2$, and the constants are given in Table 1.

[48] The empirical bias correction was applied to the WAVES, MOHAVE, and ARM RS92 data sets. The corrected RS92 measurements are compared to the reference measurements in section 4.2, then a detailed accuracy assessment is given. The computer code for the implemented nighttime and daytime empirical bias corrections is available as auxiliary material.

4.2. Correction Results: Night

[49] Figure 11 shows the nighttime RS92 comparison to CFH for three circumstances: using the original RS92 measurements (Figure 11, left); after applying the time-lag and empirical bias corrections (Figure 11, middle); and after applying only the bias correction and not the time-lag

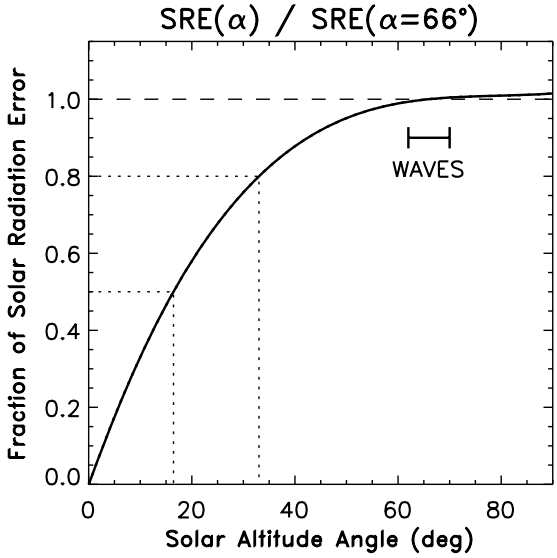

Figure 10. The dependence of the RS92 solar radiation error (SRE) on solar altitude angle (α), expressed as a fraction of the SRE at α = 66°, where 66° is the mean solar altitude angle for the daytime CFH/RS92 soundings during WAVES. The SRE is given by the difference between the polynomial curve fit for the daytime MWR SF (Figure 5 and "SF(α)" in Table 1) and the mean nighttime MWR SF of 0.97, normalized to the difference at α = 66°. Coefficients for the polynomial curve are given in Table 1 ("frac SRE(α)").

correction (Figure 11, right). Removing both the time-lag and mean bias errors (Figure 11, middle) yields a mean difference from CFH that is <2% for all RH conditions from the 700-mbar level to several kilometers above the tropopause at level P1. The random nature of the mean bias in Figure 11e indicates that there is no residual RH dependence or P dependence. The offset in the mean altitude profile below 2.5 km in Figure 11b indicates that the CFH moist mean bias prior to the forced-freezing event is about 5% relative to the consensus of the MWR, SurTHref, and CFH just after the forced-freezing event. The standard deviation of differences in Figure 11e is about ±5% and varies somewhat with RH. This parameter describes the sensor-to-sensor variability in both the RS92 and CFH combined, so it does not describe the RS92 production variability directly, but it does show that the correction removes the mean bias without adding to the variability.

[50] If the empirical bias correction alone is applied to the RS92 measurements (Figure 11, right), then the residual mean bias relative to CFH is still within the ±2% bounds up to the 220-mbar (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 uncorrected time-lag error for several kilometers above and below the tropopause. However, the impact of time-lag error seen in Figure 11f 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 be quite different for different types of atmospheric profiles.

[51] The ±2% bounds on the residual mean bias relative to CFH represents two sources of uncertainty that can be interpreted as really being the same underlying uncertainty; namely, the curve-fit error seen in Figure 9a that results from representing the data with curve fits, and statistical uncertainty in the mean accuracy of the CFH reference measurements 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 (±4% in the LT to ±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 data set 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 to 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 measurements is ±1% in the LT to ±2.3% in the UT. These values are consistent with the ±2% scatter in the residual mean bias in Figure 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 ±2% bounds in Figure 11e represent the component of residual bias uncertainty in the corrected RS92 measurements that is attributable to statistical uncertainty in the CFH accuracy. The accuracy of an individual corrected RS92 sounding must also consider the 1-σ random production variability that was estimated from the MOHAVE dual RS92 soundings to be ±1.5% for conditions above 10% RH, and ±3% for conditions below 10% RH.

[52] An additional consideration in specifying the accuracy of corrected 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 data set, and the desiccant tests of *Gorman* [2002]. Accounting for small RH offsets in a final estimate of uncertainty in the corrected RS92 data is important for very dry conditions (e.g., it is >10% uncertainty if RH < 5%).

[53] The RS92 empirical bias correction in the LT is evaluated in terms of the MWR scale factor for the corrected ARM RS92 data set (Figure 12). The correction removes the 3% nighttime moist bias in PW (Figure 12a, blue), 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 overall mean accuracy of corrected RS92 measurements over the full 2006–2007 time period; however, time-dependent variability in the PW bias of about ±2% with a time scale of about 1 year is evident in Figure 12b, and is also evident as the bumps of magnitude ±2% in Figure 12a. Figures 12c and 12d show that the correction removes both the PW dependence and the RH dependence of the scale factor that were seen in the uncorrected data.

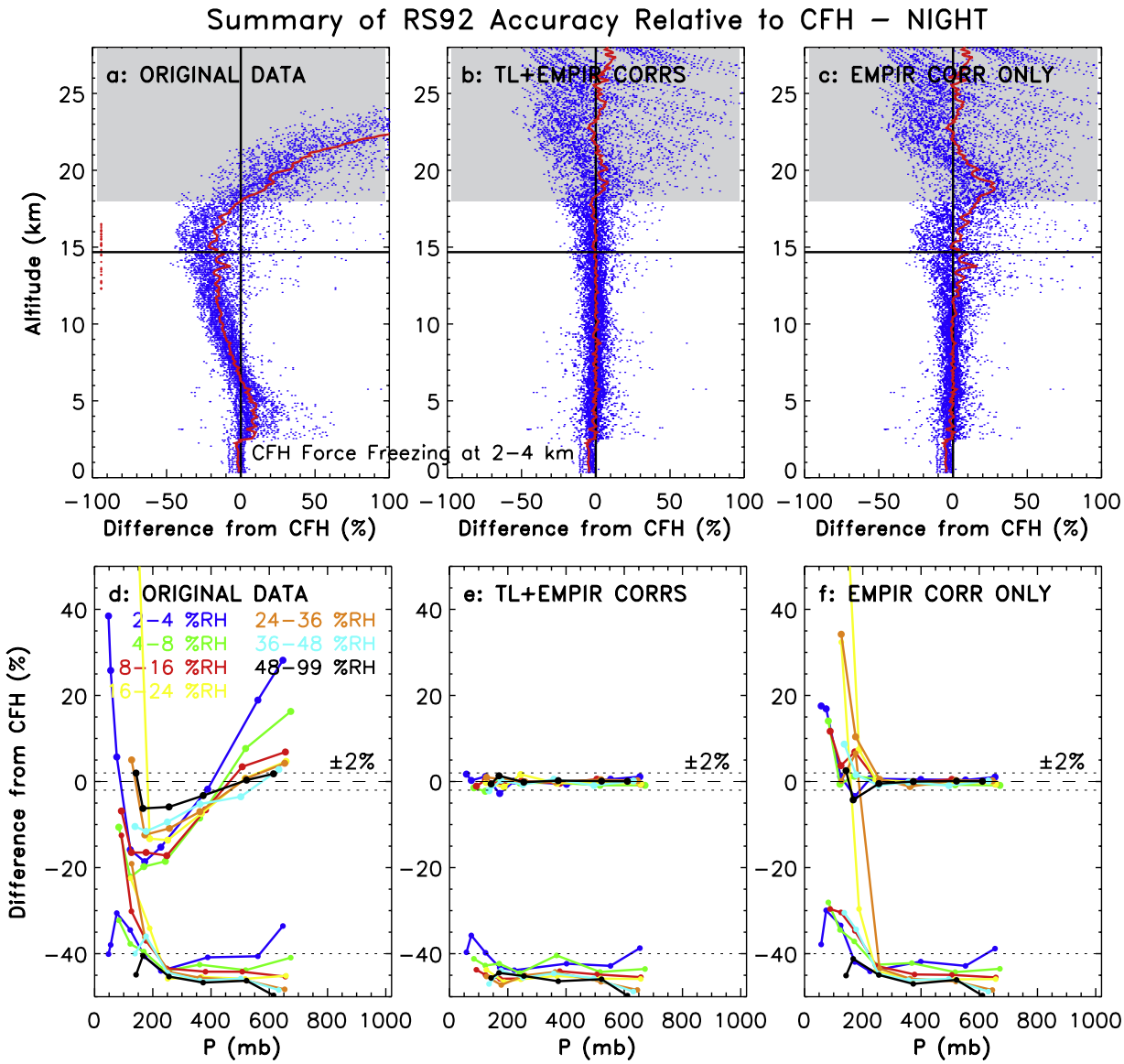

Figure 11. Effect of the time-lag and empirical bias corrections on the nighttime CFH/RS92 comparison. The RS92 percentage difference from CFH is shown for (top) the measurements and the mean as a function of altitude and (bottom) the mean and standard deviation of the binned differences as a function of pressure in seven RH intervals. (a and d) Percentage difference from CFH for the original RS92 measurements; (b and e) difference after applying the time-lag and empirical bias corrections; and (c and f) difference if only the empirical bias correction is applied. The upper curves in Figures 11d–11f show the RS92 mean bias relative to CFH, and the lower curves show the standard deviation of the bias on the same scale, but offset to 0% at the bottom of the plot for clarity. Horizontal line in Figures 11a–11c 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).

Therefore, the RS92 bias curves in the LT, which were given by extrapolation of the CFH curves from 700 mbar to the surface, fully accounts for the PW and RH dependences of the scale factor, and suggests 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.

[54] The nighttime RS92 empirical bias correction is evaluated at the surface by applying it to the ARM RS92 RAW data and comparing to the SurTHref measurements (Figure 13). The corrected RS92 and Rotronic probes agree within 0.5%, with the exception of the point at 40% RH. Measurements from the Vaisala probes, after accounting for the known probe calibration bias at 33% RH (blue), are 1% drier than the Rotronic measurements, which is within the calibration uncertainty for the two probes 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, which is within the probe calibration uncertainty. It is reasonable to interpret

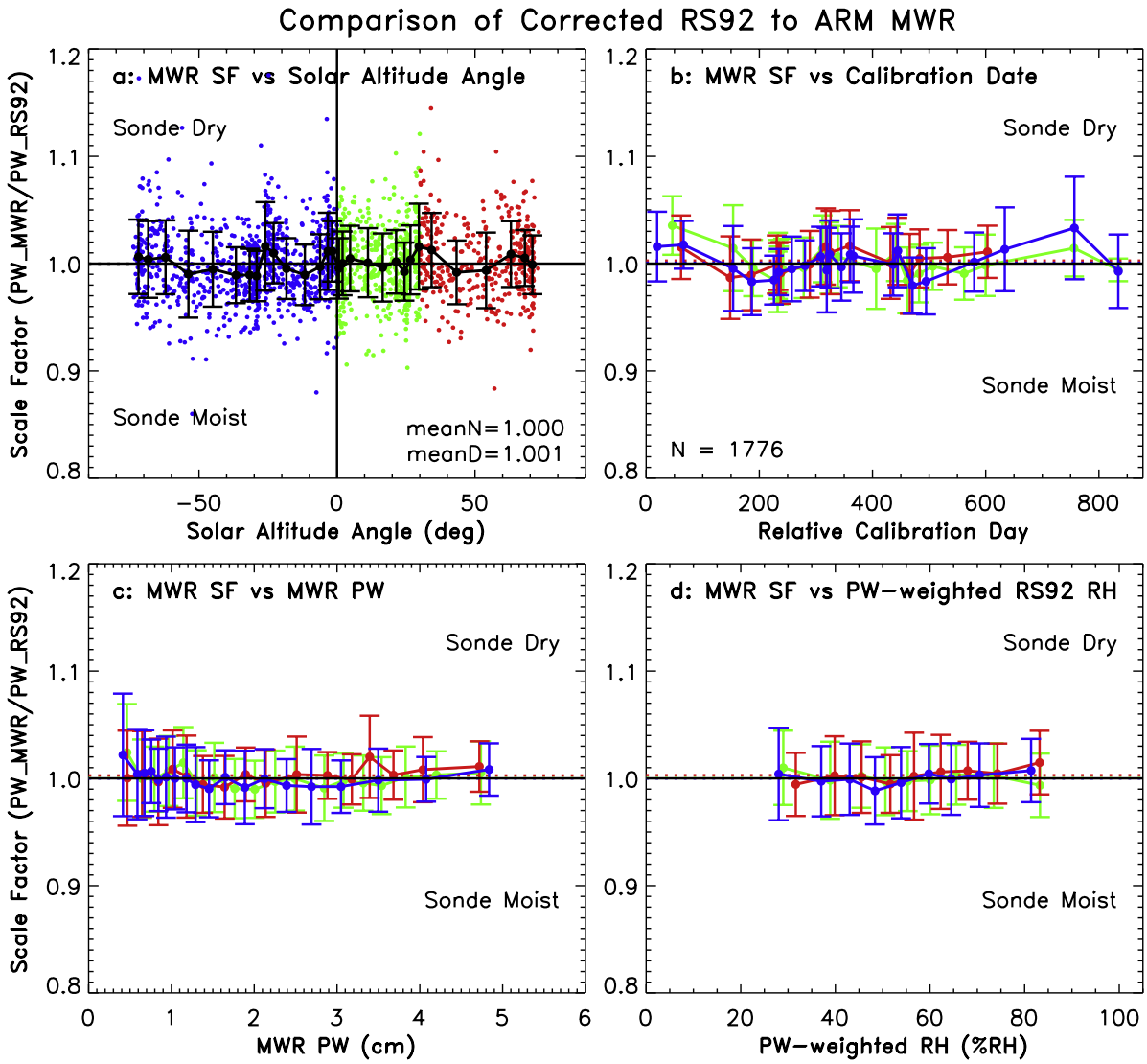

Figure 12. The MWR scale factor (SF) for the 2006–2007 ARM SGP data set after applying the appropriate nighttime or daytime RS92 empirical bias correction (and the time-lag correction, although it has no impact on the PW). Shown are the dependences of the MWR SF: (a) on the solar altitude angle at launch;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

Figure 13 as indicating that the mean accuracy of the corrected RS92 data relative to the RH probes is about ±1% over the range 25–100% RH, which is consistent with the corrected RS92 accuracy relative to the MWR after considering that the ±2% time-dependent variability in the RS92 accuracy is averaged together in Figure 13. 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 bias in the MWR or CFH results.

4.3. Correction Results: Day

[55] Figure 14 shows the daytime RS92 comparison to CFH for the original RS92 measurements (Figure 14, left), after applying the time-lag and daytime empirical bias corrections (Figure 14, middle), and after applying only the empirical bias correction (Figure 14, right). Removing both the time-lag and mean bias errors (Figure 14, middle) yields a mean difference from CFH that is <3% for all RH conditions up to the 200-mbar level, with some larger values above that level. The ±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 ±2% in the LT to ±4.5% in the UT. Above the 200-mbar level, the greater residual bias and the nonrandom trend with pressure for some RH intervals (red, orange) speaks of residual bias 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 Figure 14e. This result is suggestive of a change in the reflectivity of the sensor coating

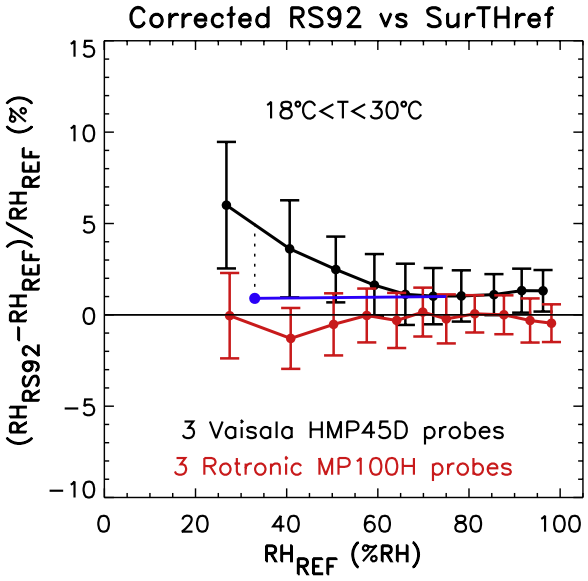

Figure 13. Mean and standard deviation of the percentage difference between RS92 and the ARM SurTHref probes as a function of RH for the 2006–2007 time period at the SGP site, after applying the nighttime empirical bias correction to the RS92 RAW data. Shown are: percentage difference from RS92 for each manufacturer's RH probes for cases in the 18°–30°C temperature range where the probe calibrations are reliable (red and black) and 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 (blue).

or other aspect of the sensor design between the two experiments (but we have no information that such a change actually occurred). If only the empirical bias correction is applied (Figure 14f), then the residual mean bias is still within the ±3% mean accuracy bounds up to the 220-mbar (12 km) level, but above that level and in the tropopause region the residual moist bias and large variability caused by time-lag error is apparent.

[56] The mean accuracy of the corrected daytime and nighttime RS92 measurements relative to CFH is nearly the same, and is independent of height or RH. The daytime comparison to CFH represents only high solar altitude angles; however, the daytime comparison to MWR (Figure 12) shows that the daytime bias correction properly accounts for the solar altitude angle and RH dependences of the MWR SF in the LT, with accuracy equivalent to the nighttime correction.

[57] The daytime bias correction was compared to the daytime correction of V07 derived from tropical soundings during the 2005 NASA Tropical Cloud Systems and Processes (TCSP) experiment, by applying the TCSP correction to the WAVES and MOHAVE RS92 soundings and characterizing the percentage difference from CFH in a manner similar to Figures 11 and 14 (not shown; provided as auxiliary material). The TCSP correction depends only on pressure and not on RH, and it represents only the moist conditions that were sampled during TCSP. As a result, TCSP-corrected RS92 measurements contain a residual moist mean bias of 5–10% for drier conditions (RH < 30%). For moist conditions, the residual RS92 moist mean bias is about 2%, which is within expected uncertainties and therefore indicates consistency between the present study and the study of V07 for the conditions where the studies are comparable.

[58] A very important caveat about the daytime bias correction concerns the effect of clouds on SRE, since the daytime correction was derived from and therefore represents clear-sky conditions. Within and probably below clouds the actual sensor heating is reduced, and above clouds there may be additional sensor heating due to the cloud albedo. The reduction in solar heating at any point within a cloud will depend on, among other things, the cloud optical depth between the sensor and the Sun, which is unique for each cloud and would vary with altitude and with position within the cloud layer. There were very few daytime cloudy RS92/CFH comparisons because clear-sky conditions were targeted, but inspection of these cases shows that the most obvious effect of applying the daytime correction to cloudy soundings is overcorrection within the cloud layer. The overcorrection is most pronounced at lower altitudes, presumably owing to the general trend of increasing cloud ice-water content (IWC) and hence cloud optical depth with decreasing altitude. Overcorrection in cirrus layers cannot be distinguished from true ice supersaturation, because both clear-sky and in-cloud ice supersaturation are common at upper levels. However, cirrus typically have low optical depths, and in these cases the sensor heating would differ little from clear-sky conditions and the daytime cloud effect should be small.

[59] We strongly caution against indiscriminate application of the daytime bias correction to cloudy soundings; however, the daytime correction is still a net benefit to accuracy for portions of the profile above clouds, or in the UT for cases of optically thin cirrus. During WAVES, rather than leaving the decision to data users, it was decided with an abundance of caution to not even archive the corrected daytime cloudy soundings.

4.4. Corrected RS92 Accuracy Specification

[60] Several factors contribute to an uncertainty estimate and range of validity for the corrected RS92 data. The nighttime residual bias uncertainty arises primarily from statistical uncertainty in the mean CFH accuracy (±2%), plus RH offset uncertainty in the RS92 calibration at 0% RH (±0.5% RH), which are combined and specified as ±(2% + 0.5% RH). However, the MWR comparison in Figure 12b also showed a time-dependent (or calibration batch-dependent) variability of ±2% relative to the long-term mean for the 2006–2007 time period, which is a bias uncertainty whose magnitude depends on the calibration batch in question. A radiosonde calibrated at one of the peaks in Figure 12b would have the maximum additional bias error of 2%, but most calibration batches would have less. We adopt more of an "expected value" by combining the uncertainty components in quadrature, leading to a bias uncertainty estimate of ±(3% + 0.5% RH) for nighttime corrected RS92 measurements in the 2006–2007 timeframe. Daytime corrected soundings have the same time-dependent variability as the nighttime soundings but greater statistical uncertainty

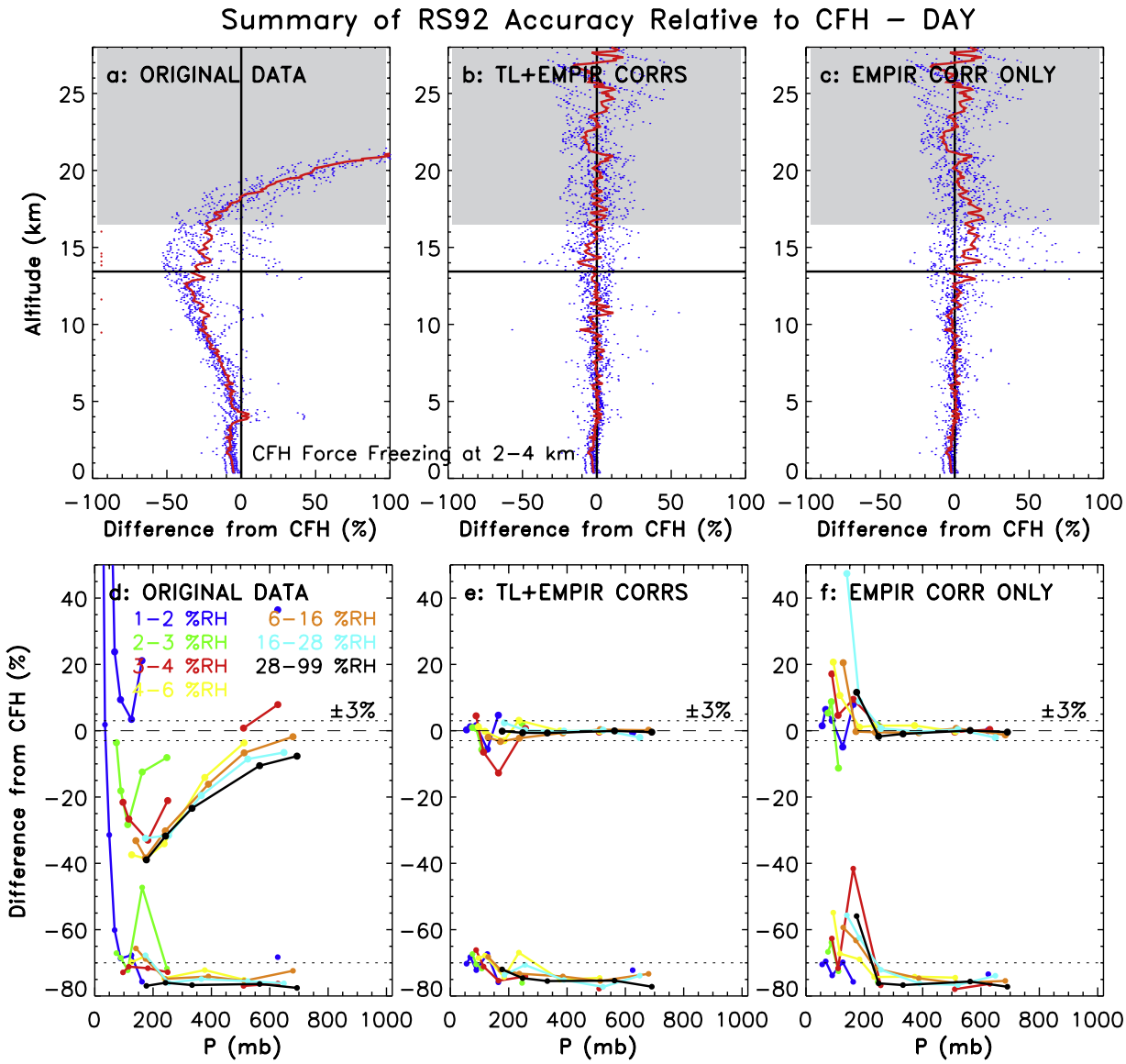

Figure 14. Effect of the time-lag and empirical bias corrections on the daytime CFH/RS92 comparison. The percentage difference from CFH is shown for (top) the individual measurements and the mean as a function of altitude and (bottom) the mean and standard deviation of the binned differences as a function of pressure in seven RH intervals. (a and d) Percentage difference from CFH for the original RS92 measurements; (b and e) difference after applying the time-lag and daytime empirical bias corrections; and (c and f) difference if only the empirical bias correction is applied. The upper curves in Figures 14d–14f show the RS92 mean bias relative to CFH, and the lower curves show the standard deviation of the bias on the same scale, but offset to 0% at the bottom of the plot for clarity. Horizontal line in Figures 14a–14c 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).

due to the smaller data set, leading to a bias uncertainty estimate for daytime corrected RS92 soundings under clear-sky conditions of ±(4% + 0.5% RH).

[61] Applying the empirical bias correction to radiosondes produced outside the 2006–2007 timeframe may have additional bias uncertainty if changes were made to the hardware or calibration procedure. An unpublished data set from the German Weather Service (DWD) shows the accuracy of the RS92 calibration at 100% RH from 2005 to early 2008, using an additional prelaunch ground check in an environment at 100% RH. The DWD data are consistent with the present study in the 2006–2007 timeframe (i.e., 3% moist mean bias at 100% RH), but also 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 moist mean 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%. This additional 1% is added to the final bias

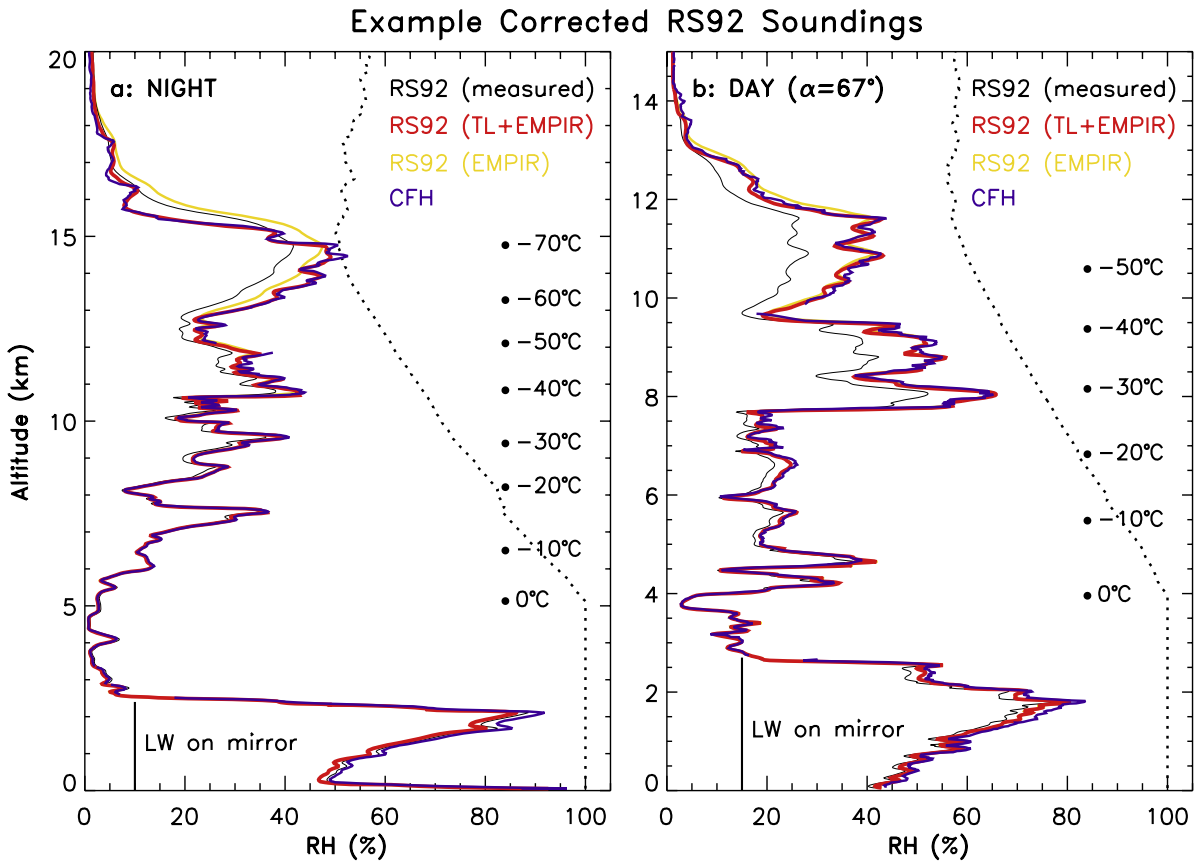

Figure 15. The example (a) nighttime and (b) daytime RS92 soundings from Figure 1 after applying corrections. Shown are: the measured RS92 profile (black curve); the RS92 profile after applying the time-lag and empirical bias corrections (red curve); the RS92 profile if only the bias correction is applied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

uncertainty estimate so that it applies to the longer time period 2005 to early 2008 (and thereafter until such time as Vaisala changes the calibration or hardware). Therefore, the final bias uncertainty estimate for nighttime corrected RS92 measurements is ±(4% + 0.5% RH), and for daytime corrected RS92 measurements it is ±(5% + 0.5% RH). The percentage component of these uncertainty estimates is reduced by 1% for radiosondes produced in the 2006–2007 time period.

[62] The accuracy of an individual corrected RS92 sounding must also consider the random error, characterized by the 1-σ variability of ±1.5% for conditions above 10% RH, and ±3% for conditions below 10% RH. The total uncertainty in an individual sounding is the RMS sum of the random and bias uncertainties. Less accurate results are expected in the UT 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 degrades the accuracy of the time-lag correction at low temperatures. If no time-lag correction is applied prior to the empirical bias correction, then the above accuracy specification is only valid up to about the −45°C level where time-lag error becomes significant.

[63] Figure 15 shows the two example soundings from Figure 1 after applying the time-lag and empirical bias corrections. 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 (yellow), then time-lag error in the original data remains, and it is more substantial for the colder nighttime case. In the lower troposphere, the residual bias between RS92 and CFH is attributed to the moist mean bias in the CFH dew point measurements during WAVES. However, the MWR and SurTHref comparisons showed that the corrected RS92 measurements in the LT are about the same accuracy as the rest of the profile.

[64] The question of why the RS92 sensor or calibration becomes insensitive to water vapor above a certain level (P1) relates to the appropriateness of the bias correction for tropical or polar soundings. If the upper limit of validity is truly determined by a pressure threshold, then its value for this midlatitude data set (P1) should also apply to tropical or polar soundings. However, if the upper limit of validity is determined by a minimum water vapor concentration, then the upper 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 RS92 RH sensor becomes nonresponsive above a pressure threshold is clarified by viewing the example corrected soundings in terms of mixing ratio rather than RH (Figure 16). After removing time-lag error (yellow) and comparing to the CFH

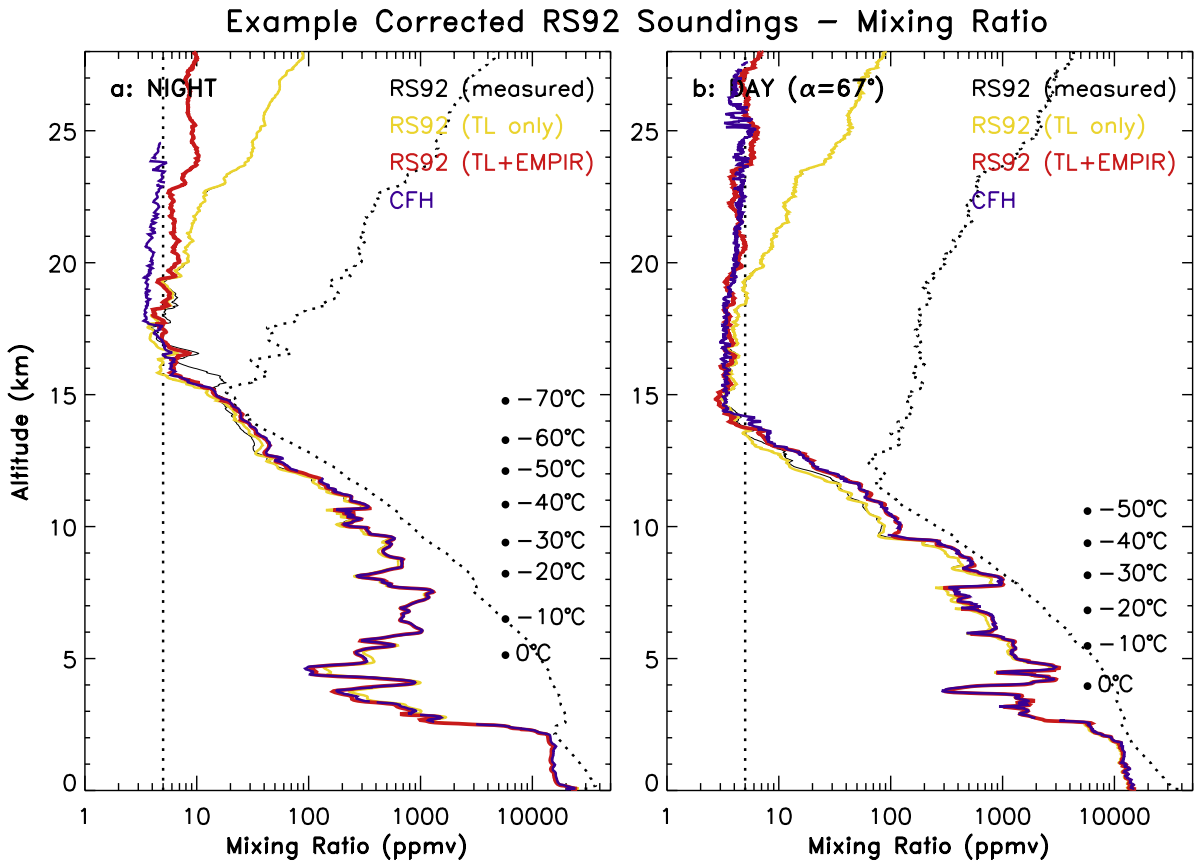

Figure 16. Same CFH and corrected RS92 soundings as in Figure 15, but shown in terms of mixing ratio rather than RH. Dashed curve is the saturation mixing ratio, and the tropopause is located at its minimum value.

profile (purple), the RS92 is seen to be responsive 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 minimum 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 yellow curve from the CFH measurements occurs at an altitude that is consistent with the upper limit of validity given earlier (75 mbar at night and 100 mbar during the day, or 18 and 16.5 km, respectively). Inspection of all the dual soundings in the WAVES/MOHAVE data set in terms of mixing ratio consistently showed that the RS92 is responsive at even the minimum mixing ratio in the LS, and that the RS92 measurements transition to paralleling the saturation mixing ratio curve above the level P1, above which the sensor calibration appears to be responding only to the changing temperature. Therefore, we expect the correction to be valid up to level P1 for tropical and polar soundings as well. Above P1, the corrected RS92 data merely represent the CFH climatology during these midlatitude experiments.

[65] Although the RS92 accuracy estimate is valid up to P1, the corrected RS92 data may or may not be useful for a given application, depending on its accuracy requirements. For example, the nighttime mean bias estimate of ±(4% + 0.5% RH) means that the uncertainty in mixing ratio for extremely dry conditions of 2% RH in the LS (as well as in the troposphere) is ±29%, and therefore corrected RS92 measurements are not sufficiently accurate for monitoring climate trends in the LS. If an ensemble of corrected RS92 soundings is used for remote sensor validation in the UT with a mean accuracy requirement of ±10%, then this requirement is met for only conditions above 8.3% RH. The contribution of the ±0.5% RH offset component of the bias uncertainty is <1% for conditions above 50% RH, so the total bias uncertainty for nighttime corrected RS92 measurements is <5% for conditions above 50% RH. These examples illustrate the distinction between the "validity" of corrected RS92 data (i.e., all conditions up to P1), and the "utility" of the corrected data, which depends on the RH and on the accuracy requirements for a given application.

5. Conclusions

[66] The accuracy of Vaisala RS92 radiosonde water vapor measurements was investigated by comparing midlatitude RS92 measurements to simultaneous measurements from three reference instruments of known accuracy (CFH, MWR, and RH probes). The three intercomparisons were combined to produce a detailed estimate of the RS92 mean bias and variability as a function of height, RH, and solar altitude angle. The RS92 mean bias differs fundamentally between nighttime and daytime soundings, where the night-

time 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.

[67] Relative to the CFH above the 700-mbar level, the RS92 mean bias for nighttime soundings (Figure 2f) varies with height from a moist bias at 700 mbar (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 (Figure 3f) is a dry bias that increases from about 5% at 700 mbar to 45% in the UT, and varies somewhat with RH. Comparison of RS92 and ARM MWR PW measurements (MWRRET PHYS retrieval) describes the RS92 mean bias averaged over all RH conditions in the lower troposphere. Nighttime RS92 measurements were found to have a moist mean bias of 3%, and daytime RS92 measurements have a mean bias that depends on the solar altitude angle and is a maximum dry bias of 5–6% for high solar altitude angles. The MWR comparisons also indicated time-dependent variability in the RS92 mean bias of ±2% relative to the long-term mean, attributed to drift in the Vaisala calibration reference. The RS92 mean bias at the surface was determined by comparing the prelaunch RS92 RAW data to measurements from calibrated RH probes in the ARM SurTHref system (Figure 7c), which always describes the nighttime RS92 performance since there is no solar radiation error. The RS92 was found to have a moist mean 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 mbar.

[68] The three RS92 accuracy assessments were combined to produce an estimate of the RS92 mean bias error as a function of pressure and RH (Figure 9 and Table 1). An empirical correction was derived to remove the mean bias, with the result that the accuracy of corrected RS92 data is primarily given by the random production variability plus residual bias uncertainty from several sources: statistical uncertainty given by the size of the CFH/RS92 data set (±2% nighttime and ±3% daytime); time-dependent variability in the accuracy of the Vaisala calibration reference (±2%); uncertainty related to the GC correction and the accuracy of the sensor calibration at 0% RH (±0.5% RH); and a 1% increase in the RS92 mean calibration bias in the 2006–2007 time period of this study relative to radiosondes produced in 2005 or 2008. These contributions were combined as described earlier to yield a residual accuracy specification for nighttime corrected RS92 measurements of ±(4% + 0.5% RH), valid for all RH conditions from the surface to the 75-mbar (18 km) level. To clarify, the uncertainty is 4% of the measured RH value plus a small RH offset component that is significant for dry conditions (e.g., it adds 10% uncertainty for conditions of 5% RH). The smaller daytime data set leads to a residual accuracy specification for daytime corrected RS92 measurements of ±(5% + 0.5% RH), valid for all RH conditions from the surface to the 100-mbar (16.5 km) level for clear-sky 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-mbar (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 residual bias uncertainty, individual soundings also have a sensor-specific random error from "production variability," where the 1-σ variability is estimated from dual RS92 soundings to be ±1.5% for conditions above 10% RH and ±3% for conditions below 10% RH.

[69] A caveat of great importance for daytime soundings is that clouds affect the solar radiation error in a complicated way, and the daytime empirical bias correction was derived from and therefore applies to clear-sky conditions. The daytime correction should be considered unreliable within and probably below optically thick clouds in the lower and middle troposphere; however, there is evidence that the cloud effect on solar radiation error is minimal compared to the magnitude of the correction for portions of soundings above clouds or within optically thin cirrus in the UT. Clouds do not affect the nighttime results.

[70] It is acknowledged that this study lacks independent validation of the correction using an independent RS92/CFH data set. Furthermore, although the RS92 was characterized over the full range of RH and P, the results have not been verified for tropical or polar soundings. The RS92 and CFH data used in this study were specifically obtained as part of water vapor validation and instrument intercomparison experiments, and attention was paid to important details such as using the high RH resolution data in the Vaisala FLEDT files, and evaluating the GC correction and state of the desiccant regularly. Several recommendations for maximizing the accuracy and utility of Vaisala radiosonde data are given in Appendix A.

Appendix A

A1. Recommendations for High-Quality RS92 Data

[71] 1. Vaisala RH data are reported as integer values in the standard EDT processed data files, which introduces a roundoff error of ±0.5% RH, and 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 two decimal places of precision, but is otherwise identical to the standard EDT files. It is recommended that the FLEDT data be routinely used and archived for all soundings, for the benefit of future global radiosonde data archives.

[72] 2. The Vaisala ground check (GC) correction is intended to recover the factory calibration accuracy by adjusting the RH measurement to be 0.0% RH when the RH sensor is sealed in a desiccated environment. However, if the desiccant is not fresh or the chamber not tightly sealed, then the environment will be $\gg$0.0% RH and the GC correction will introduce a dry bias to the measurements. Furthermore, even the best molecular sieve desiccant tested by the Australian Bureau of Meteorology achieved a minimum of only 0.5% RH [*Gorman*, 2002], suggesting an inherent bias uncertainty of about 0.5% RH in the accuracy of the Vaisala calibration at 0% RH. The mean and standard deviation of the GC correction values for the 2006–2007 ARM RS92 data set is −0.49 ± 0.35% RH, and these are

taken as reliable expected values because ARM maintains good operational GC procedures. We suggest that GC values >1% RH should be considered suspicious and merit checking the desiccant freshness and the chamber seal, and GC values >2% RH almost certainly reflect faulty GC procedures or a bad radiosonde. The benefit of improved accuracy from the GC correction is comparable to the uncertainty about the true RH in the desiccant chamber, and since improper procedures can lead to much larger errors, we recommended that the GC correction only be applied if care and scrutiny are routinely used.

[73] 3. We recommend routinely saving the RAW PTU data file generated by the Vaisala data system, which contains the prelaunch data acquired during the launch preparations, including the GC procedure and in this study the SurTHref comparisons. The RAW data also contain the measurements from both RH sensors, including values below 1% RH or above 100% RH (unlike the processed EDT and FLEDT files). This information is useful for evaluating the GC correction and the sensor calibration accuracy at 0% RH and 100% RH, and for more accurately applying corrections when values >100% RH were actually measured.

[74] 4. 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 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 frost point 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 and taken into consideration. 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.

[75] **Acknowledgments.** We thank Ricardo Forno (Universidad Mayor de San Andres, La Paz, Bolivia), Belay Demoz, Everette Joseph and Cassie Stearns (Howard University), Scott Rabenhorst (University of 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 (University of 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 the foresight of Barry Lesht (ANL) for archiving the ARM RAW radiosonde data that are 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. The National Center for Atmospheric Research (NCAR) is sponsored by the National Science Foundation. This work was funded by NASA grant NNX06AC79G.

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, doi:10.1175/2007JTECHA1027.1.

Gorman, J. (2002), Evaluation of desiccants, *Instrum. Test Rep. 2002_664*, 7 pp., Aust. Bur. of Meteorol., Melbourne, Victoria, Australia. (Available at www.wmo.int/pages/prog/www/IMOP/WebPortal-AWS/Tests/ITR664.pdf)

Heymsfield, A. J., and L. M. Miloshevich (1995), Relative humidity and temperature influences on the formation and evolution of cirrus: Observations from wave clouds and FIRE II, *J. Atmos. Sci.*, *52*, 4302–4326, doi:10.1175/1520-0469(1995)052<4302:RHATIO>2.0.CO;2.

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, doi:10.1175/1520-0426(2004)021<1305:DAVOAT>2.0.CO;2.

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.

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, doi:10.1175/1520-0426(2003)020<0117:DBAVIV>2.0.CO;2.

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, doi:10.1109/TGRS.2007.903703.

Vömel, H., H. Selkirk, L. Miloshevich, J. Valverde-Canossa, J. Valdes, E. Kyro, R. Kivi, W. Stolz, G. Peng, and J. A. Diaz (2007a), 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 (2007b), 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 to100°C: A revision, *J. Res. Natl. Bur. Stand. U. S., Sect. A*, *80*, 775–785.

T. Leblanc, Table Mountain Observatory, JPL, NASA, 4800 Oak Grove Drive, Pasadena, CA 91109, USA.

L. M. Miloshevich, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, USA. (milo@ucar.edu)

H. Vömel, CIRES, University of Colorado, Boulder, CO 80309, USA.

D. N. Whiteman, NASA Goddard Space Flight Center, Code 613.1, Building 33, Room D404, Greenbelt, MD 20771, USA.