

Comparisons of temperature, pressure and humidity measurements by balloon-borne radiosondes and frost point hygrometers during MOHAVE-2009

D. F. Hurst^{1,2}, E. G. Hall^{1,2}, A. F. Jordan^{1,2}, L. M. Miloshevich³, D. N. Whiteman⁴, T. Leblanc⁵, D. Walsh⁵, H. Vömel⁶, and S. J. Oltmans²

¹Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA

²NOAA Earth System Research Laboratory, Global Monitoring Division, Boulder, Colorado, USA

³Milo Scientific LLC, Lafayette, Colorado, USA

⁴NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

⁵Jet Propulsion Laboratory, Table Mountain Facility, Wrightwood, California, USA

⁶Meteorologisches Observatorium Lindenberg, Deutscher Wetterdienst, Lindenberg, Germany

Received: 27 April 2011 – Published in Atmos. Meas. Tech. Discuss.: 11 July 2011

Revised: 5 October 2011 – Accepted: 25 November 2011 – Published: 16 December 2011

Abstract. We compare coincident, in situ, balloon-borne measurements of temperature (T) and pressure (P) by two radiosondes (Vaisala RS92, Intermet iMet-1-RSB) and similar measurements of relative humidity (RH) by RS92 sondes and frost point hygrometers. Data from a total of 28 balloon flights with at least one pair of radiosondes are analyzed in 1-km altitude bins to quantify measurement differences between the sonde sensors and how they vary with altitude. Each comparison (T , P , RH) exposes several profiles of anomalously large measurement differences. Measurement difference statistics, calculated with and without the anomalous profiles, are compared to uncertainties quoted by the radiosonde manufacturers. Excluding seven anomalous profiles, T differences between 19 pairs of RS92 and iMet sondes exceed their measurement uncertainty limits (2σ) 31 % of the time and reveal a statistically significant, altitude-independent bias of $0.5 \pm 0.2^\circ\text{C}$. Similarly, RS92-iMet P differences in 22 non-anomalous profiles exceed their uncertainty limits 23 % of the time, with a disproportionate 83 % of the excessive P differences at altitudes >16 km. The RS92-iMet pressure differences increase smoothly from -0.6 hPa near the surface to 0.8 hPa above 25 km. Temperature and P differences between all 14 pairs of RS92 sondes exceed manufacturer-quoted, reproducibility limits (σ) 28 % and 11 % of the time, respectively. About 95 % of the

excessive T differences are eliminated when 5 anomalous RS92-RS92 profiles are excluded. Only 5 % of RH measurement differences between 14 pairs of RS92 sondes exceed the manufacturer's measurement reproducibility limit (σ). RH measurements by RS92 sondes are also compared to RH values calculated from frost point hygrometer measurements and coincident T measurements by the radiosondes. The influences of RS92-iMet T and P differences on RH values and water vapor mixing ratios calculated from frost point hygrometer measurements are examined.

1 Introduction

Vertical profile measurements of essential climate variables T , P and RH have been made around the globe for decades. These measurements were predominantly made for weather forecasting and other short-term investigations, hence the collected data often lack the long-term stability and traceability necessary for climate research (Thorne et al., 2005; Titchner et al., 2009; Immler et al., 2010). Long-term measurement records from balloon-borne radiosondes are particularly questionable because of poorly documented instrument and procedural changes over the years (Titchner et al., 2009; Seidel et al., 2004).

This paper presents statistical evaluations of the differences between coincident, in situ, vertical profile measurements of T and P by two types of radiosondes; Vaisala RS92



Correspondence to: D. F. Hurst
(dale.hurst@noaa.gov)

and Internet iMet-1-RSB. We also examine differences between RH measurements by Vaisala RS92 radiosondes and two frost point hygrometers, the cryogenic frost point hygrometer (CFH; Vömel et al., 2007a) and the NOAA frost point hygrometer (FPH; Mastenbrook and Oltmans, 1983; Hurst et al., 2011). These balloon-borne measurements were made as part of the MOHAVE-2009 campaign, 11–27 October 2009 (Leblanc et al., 2011). We compare only measurements made from the same balloons, eliminating any concerns about spatial and temporal differences between the measurements by different sensors.

2 Experimental

A total of 44 balloons were launched during MOHAVE-2009 from the Table Mountain Facility (34.4° N, 117.7° W, 2285 m a.s.l.) near Wrightwood, California. Twenty-eight of these balloons were instrumented with two or more radiosondes, and 20 of the 28 balloons also carried a CFH (16) or NOAA FPH (4) frost point hygrometer (Table 1). Two RS92 sondes flown on the same balloon were designated as primary and secondary RS92 sondes for the purpose of comparison. This distinction was unnecessary for Internet sondes because the balloon payloads never included more than one iMet-1-RSB sonde.

All but 2 of the 28 balloon flights analyzed here were launched at night to compare water vapor measurements by the balloon-borne sensors with those by ground-based Raman lidars. Two balloons launched during daytime (Table 1) provided comparison data for solar FTIR spectroscopic measurements of water vapor. Each balloon reached an altitude of at least 27 km except for four that burst prematurely. The maximum altitude and minimum pressure attained by a balloon were 32.6 km and 8 hPa.

2.1 Radiosondes

The Vaisala RS92 sondes used in this campaign were models -SGP and -K, the only difference being the GPS signal-receiving capability of the -SGP. Both models had a capacitive wire temperature sensor, a pair of thin film capacitive polymer humidity sensors, a capacitive silicon pressure sensor and a 403 MHz-band transmitter. Sensor data from each RS92 sonde were received and recorded by one of two Vaisala DigiCORA MW31 systems. As specified by the manufacturer, RS92 sondes were removed from their original packaging less than one hour before launch and ground checked using one of two Vaisala GC25 units. The GC25 units also reconditioned the RS92 RH sensors. Though these pre-launch checks were performed as recommended by the manufacturer to promote high data quality, no RS92 sondes were rejected during MOHAVE-2009. Only ascent data of the RS92 sondes were recorded because their data transmission ceased shortly after the start of descent.

Table 1. Balloon launches and payloads.

Date, Time (UTC*)	Flight	RS92	iMet	Frost Point
11 Oct, 08:23	TF022	2	1	CFH
15 Oct, 05:01	Oct15	2		
16 Oct, 04:19	TF024	1	1	FPH
16 Oct, 07:58	Oct16	2		
17 Oct, 04:47	TF025	2	1	CFH
17 Oct, 08:31	Oct17	2		
18 Oct, 02:55	TF026	1**	1	FPH
18 Oct, 06:45	Oct18	2		
18 Oct, 21:11	TF027	1	1	CFH
19 Oct, 03:31	TF028	2	1	CFH
19 Oct, 07:33	Oct19	2		
20 Oct, 05:11	TF029	1	1	CFH
20 Oct, 05:26	TF030	1	1	FPH
20 Oct, 08:11	TF031	1	1	CFH
20 Oct, 10:49	Oct20	2		
21 Oct, 06:08	TF033	2**	1	CFH
21 Oct, 09:25	TF034	1	1	CFH
22 Oct, 02:58	TF035	1	1	CFH
22 Oct, 03:17	TF036	1**	1	FPH
22 Oct, 10:34	TF037	1	1	CFH
22 Oct, 08:12	Oct22	2		
24 Oct, 03:21	TF038	1	1	CFH
24 Oct, 05:56	Oct24	2		
25 Oct, 03:55	TF039	2**	1	CFH
25 Oct, 20:30	TF040	1	1	CFH
26 Oct, 05:59	TF041	2**	1	CFH
27 Oct, 05:17	TF042	1	1	CFH
27 Oct, 08:35	TF043	1	1	CFH

* UTC = Pacific daylight time + 8 h. All flights except TF027 and TF040 were conducted at night. ** Every RS92 sonde was model -K except for 5 RS92-SGP.

Vaisala quotes two types of measurement uncertainties pertinent to this comparison study: total measurement uncertainties in soundings, given at the 95 % confidence level (2σ), and reproducibility values for flights with two RS92 sondes, given as one standard deviation (σ) of their measurement differences. Total uncertainties and reproducibility values for P measurements are quoted in two different pressure regimes, 3–100 and 100–1080 hPa, while T reproducibility values are quoted in three pressure intervals: 3–20, 20–100 and 100–1080 hPa. The specifications for RS92-K and RS92-SGP measurements of P , T and RH are exactly the same because they have the same sensors. Manufacturer-quoted measurement uncertainty and reproducibility values are summarized in Table 2.

The Internet North America model iMet-1-RSB radiosondes used in this campaign had a bead thermistor temperature sensor, a piezo-resistive silicon pressure sensor, a capacitive polymer humidity sensor, a 12-channel GPS receiver, and a 403 MHz band data transmitter. This model should not be confused with the Internet South Africa iMet-2

Table 2. Manufacturer-specified and combined uncertainties for radiosonde measurements.

<i>P</i> Range ^a hPa	Manufacturer-Specified Uncertainties									Combined Uncertainties		
	RS92 total ^b			RS92 reproducibility ^c			iMet accuracy ^b			RS92-iMet ^d		
	<i>P</i> hPa	<i>T</i> °C	RH %	<i>P</i> hPa	<i>T</i> °C	RH %	<i>P</i> hPa	<i>T</i> °C	RH %	<i>P</i> hPa	<i>T</i> °C	RH %
3–20					0.5							
4–100	0.6			0.3						0.78		
4–400							0.5					
4–1070		0.5	5			2		0.3	5		0.58	7.1
20–100					0.3							
100–400										1.12		
100–1080	1			0.5	0.2							
400–1070							1.8			2.06		

^a Pressure ranges over which the specified measurement uncertainties are valid. RS92 manufacturer's quoted pressure range limits of 3 and 1080 hPa are aligned with iMet's pressure range limits of 4 and 1070 hPa to facilitate combinations of RS92 and iMet uncertainties.

^b RS92 total uncertainties and iMet accuracy values are 2 standard deviations (2σ) of measurement errors.

^c RS92 reproducibility is one standard deviation (σ) of measurement differences for dual soundings.

^d Calculated from RS92 total uncertainties and iMet accuracy values (2σ), in quadrature.

radiosonde. In addition to transmitting their own sensor data, the iMet-1-RSB sondes also relayed data from all connected CFHs, FPHs and ozonesondes. Sensor data from the iMet sondes and connected instruments were received by an Icom IC-R8500 receiver and recorded by a PC running custom software “STRATO” (iMet + CFH) or “SkySonde” (iMet + FPH). Data from iMet sondes and connected instruments were received and recorded during both the ascent and descent of balloons.

Internet quotes *P*, *T* and RH measurement uncertainties for its iMet-1-RSB radiosondes as 2σ accuracy limits of ± 1.8 hPa (400–1070 hPa), ± 0.5 hPa (4–400 hPa), ± 0.3 °C, and ± 5 % RH. We calculate 2σ uncertainty limits for the differences between coincident RS92 and iMet measurements by combining these values with the Vaisala RS92 total measurement uncertainties, in quadrature (Table 2).

Data from a total of 37 RS92-K, 5 RS92-SGP, and 20 iMet-1-RSB radiosondes are included in this comparison. Sonde measurement data were used as received and recorded (i.e., raw) except for the RS92 RH data (all flights) and the iMet and RS92 *T* data for the 2 daytime flights. Solar radiation corrections of the daytime *T* data from the RS92 and iMet sondes were made by the RS92 DigiCORA software and by applying Internet-provided correction algorithms, respectively. These altitude-dependent adjustments reduced *T* by a maximum of ~ 1 °C at the highest altitudes. Corrections to the raw RS92 RH data were applied post-flight using established algorithms that are based on a large body of research (Miloshevich et al., 2004; Vömel et al., 2007b; Miloshevich et al., 2009). The RH corrections include adjustments for solar radiation effects (daytime flights), mean calibration bias, and *T*-dependent sensor time-lag errors ($T < -45$ °C). For the relatively dry tropospheric conditions

observed during most MOHAVE-2009 balloon flights, these corrections increased RS92 RH values by 1–3 % RH in the middle and upper troposphere. Note that Vaisala recently implemented their own RH corrections for solar radiation and time-lag errors in their new version of the DigiCORA software (v3.64, December 2010).

Relative humidity measurements by iMet-1-RSB sondes are not compared in this paper because of sensor problems during the campaign. Measurement glitches and negative RH values presented the greatest difficulties. Since October 2009, Internet has worked to improve the implementation and calibration of their RH sensors, and recent flights with the iMet-1-RSB demonstrate that most of these problems have been eliminated.

2.2 Frost point hygrometers

The two frost point hygrometers flown during MOHAVE-2009 (CFH and FPH) utilize the same measurement principle and have many similar design features. Both instruments rely on the growth and maintenance of a stable frost layer on a temperature-controlled mirror positioned within a steady air stream. A stable frost layer on the mirror implies equilibrium between the ice surface and water vapor in the overlying air stream. At equilibrium the partial pressure of water vapor in the air stream is directly related to the mirror temperature (Brewer et al., 1948) through the Goff-Gratch formulation of the Clausius-Clapeyron equation (Goff, 1957). The frost point measurement technique relies on first principles and accurate calibrations of thermistors embedded in the mirrors; no water vapor calibration standards or scales are required.

Both frost point hygrometers stem from the fundamental design and operational principles of the NOAA FPH that was first flown over Boulder, Colorado in 1980 (Mastenbrook and Oltmans, 1983). CFH technology diverged from the FPH in 2003 through efforts to reduce instrument size and weight, improve frost layer stability and eliminate the need for a sun shield (Vömel et al., 2007a). Both hygrometers were significantly improved after 2003, but since these improvements were not exactly the same, the instruments diverged in subtle ways. Though different, the CFH and FPH produce very similar water vapor mixing ratio data. Stratospheric water vapor mixing ratios measured by the FPH and CFH during five dual flights over Boulder in 2008–2009 differed by an average 0.1 ± 0.2 ppmv, roughly 2 ± 4 % of stratospheric mixing ratios (Hurst et al., 2011).

2.3 Data matching

Every measurement comparison presented here is strictly between multiple instruments carried by the same balloon. Radiosondes, frost point hygrometers and ozonesondes were all packaged together as a single payload tethered 30–40 m below the balloon. The bundling of multiple instruments into a single payload guarantees spatial coincidence of their measurements, but temporal coincidence must also be ensured for these comparisons to be meaningful. We use the measurement time stamps of the multiple instruments on each balloon to synchronize their data. Measurement time stamps are recorded by a sounding system when it receives data packets transmitted by its radiosonde. The different receiving systems (i.e., combinations of radiosonde, receiver, PC and data recording software, see Sect. 2.1) all provide time stamps as elapsed times since balloon launch. Other methods of data synchronization that require the matching of altitudes or sensor data are prone to temporal offsets, especially if altitude or sensor biases exist. Time stamp synchronization of the sounding data is free from such biases, but only after it is made certain that the elapsed times recorded by the different sounding systems correspond to the same moments in time. Often they do not because each sounding system uses its own algorithm and radiosonde data to detect the exact moment of balloon launch and record it as $t = 0$. Any launch time ($t = 0$) “mismatch” can be remedied by applying a small temporal offset to the elapsed time stamps of one sounding system. Once the correct temporal offset is applied the time stamps of the two sounding systems are synchronized in real time.

We determined the time stamp offsets required to synchronize data from the 2 or 3 sounding systems used for each flight by comparing the T measurements by their radiosondes. Temperatures are well suited for this task because there are often features, abrupt and significant departures of T from the standard lapse rate model, measured with high precision by the radiosondes during each flight. The correlation between two radiosondes’ temperature data is naturally reduced by any temporal mismatches in their time stamps,

so for each flight we find the time stamp offset that maximizes this correlation. The final data files for MOHAVE-2009 present RS92 data at regular intervals of 2 s and iMet data at irregular intervals of 1–3 s, so we apply offsets in uniform 2-s increments to the RS92 time stamps until the maximum correlation coefficient (r) is found. This method cleanly determined the optimum offsets for 85 % of the T comparisons, but for the other 15 % was unable to conclusively identify one of two sequential time stamp offsets as optimum. For these cases we visually inspected the two T time series and selected the offset that produced the best temporal overlap of the measured T features. This visual method also confirmed the optimum time stamp offsets determined by the correlation method for the other 85 % of T comparisons.

Time stamp offsets for each of the 14 dual RS92 soundings were mostly zero or 2 s, however one RS92 pair required an 8 s offset. Offsets for the 26 RS92 sondes flown alongside 20 iMet sondes averaged 2.5 ± 6.0 s, with several flights requiring offsets > 10 s. Again, these offsets were necessary only because the different sounding systems detected balloon launch ($t = 0$) at different times. In most cases these small time stamp offsets also significantly improved correlations between the P and RH measurements of the different radiosondes.

After applying offsets to the RS92 time stamps the RS92 data were temporally matched to iMet time stamps with a tolerance of ± 0.5 s. Though some pairs of RS92 sondes did require small time stamp offsets, time matching the offset data was simplistic because the uniform 2-s data intervals were unaltered by the time stamp offsets (in multiples of 2 s). Though these data synchronization and matching procedures minimized temporal differences between the measurements by different sensors, systematic differences of up to 1 s may remain between the matched RS92 and iMet measurements. We reiterate here that RS92 sonde data were not transmitted or recorded during balloon descent so the following measurement comparisons include only ascent data.

3 Measurement comparisons

Each measurement difference profile was calculated from the time-matched measurements of two sensors on the same balloon. Difference profiles between two RS92 sondes on the same balloon were consistently calculated as primary RS92 minus secondary RS92 and labeled with the flight number (Table 1). Data from an iMet sonde or frost point hygrometer were always subtracted from RS92 sonde data. If a balloon carried two RS92 sondes, profiles of differences from the primary (secondary) RS92 sonde were labeled with the flight number (flight number plus suffix “b”).

For many comparisons we identify several difference profiles as “anomalous” because they do not conform to the majority of difference profiles. Though not strictly quantitative, the process of identifying anomalous difference profiles

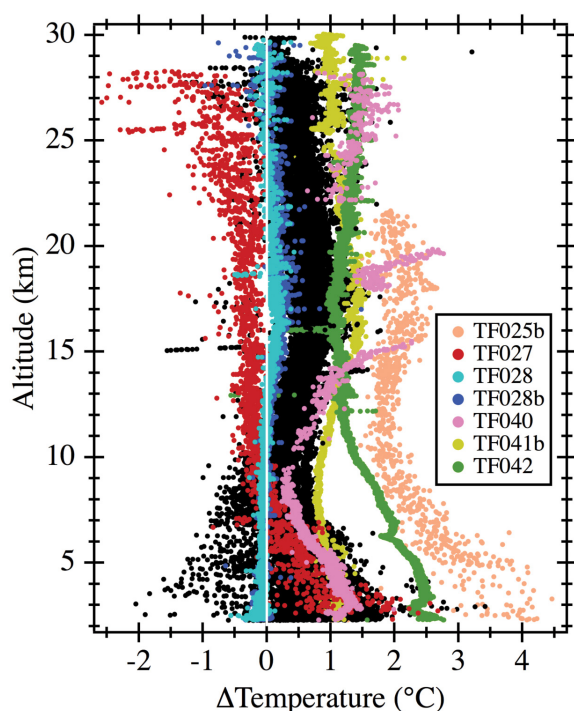


Fig. 1. Profiles of T differences between RS92 and iMet radiosondes. Seven of the 26 profiles are identified as anomalous (i.e., not conforming to the majority of difference profiles), colored and listed in the legend. Profiles denoted by the flight number plus suffix “b” present differences between secondary RS92 sondes and iMet sondes. The white vertical zero line serves as a visual guide.

was very straightforward and unambiguous. We exclude the anomalous difference profiles when calculating measurement difference statistics to assess biases between different sensors, including altitude-dependent biases, and to gauge measurement uncertainties quoted by the radiosonde manufacturers. We also report measurement difference statistics that include the anomalous profiles to demonstrate the realistic results of assuming every sonde will perform to its specifications without pre-flight evaluation.

3.1 Temperature

Vertical profiles of the differences between coincident T measurements by RS92 and iMet radiosondes depict considerable biases as well as some random disparities caused by sporadic sensor noise (Fig. 1). Seven of these 26 difference profiles, those plotted with colored markers and identified in the figure legend, do not conform to the majority of difference profiles and are deemed anomalous. From here forward we use colored markers in difference profile plots to identify anomalous profiles, and black markers to represent the majority of profiles.

To focus more on measurement biases and less on sensor noise we prefer to plot profiles of median measurement

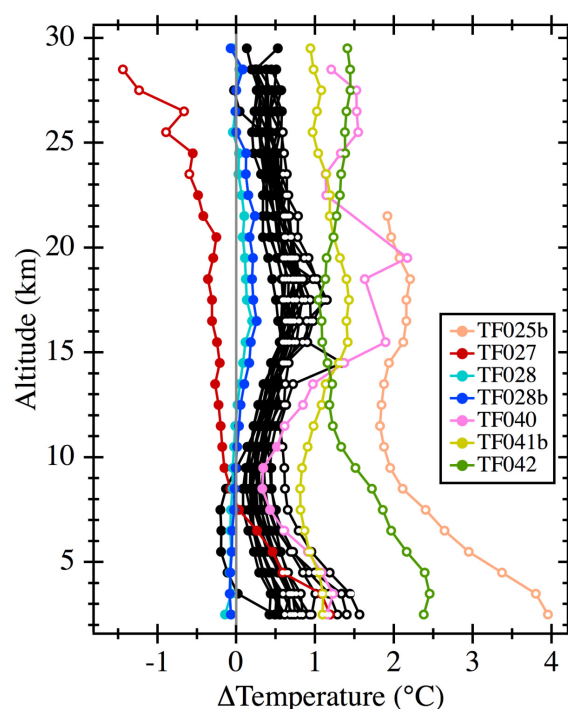


Fig. 2. Same as Fig. 1 except the RS92-iMet T differences are presented as median differences in 1-km altitude bins (ΔT). Open circles depict the 38 % of ΔT that exceed the 2σ combined measurement uncertainty of ± 0.58 °C. Of all the excessive ΔT , 51 % are located above 16 km.

differences in 1-km altitude bins (Fig. 2) instead of profiles that include all measurement differences (Fig. 1). In general, plots of median difference profiles are much less visually cluttered and have smaller x-axis ranges that increase figure clarity. Medians were chosen over averages because they reduce the influences of large, random measurement differences on the statistics we employ to assess measurement uncertainties and biases. Hereinafter let ΔX denote the 1-km altitude-binned median differences of parameter X .

As expected the same 7 anomalous profiles in Fig. 1 stand out from the majority of ΔT profiles in Fig. 2. One anomalous difference profile (TF025b) has ΔT as large as 4 °C near the surface, while above 9 km the largest ΔT is 2.3 °C. The majority of profiles suggest a positive bias in RS92-iMet ΔT (Fig. 2, black markers) that is discussed in detail below.

The ΔT profiles for pairs of RS92 sondes on the same balloons also portray several anomalous difference profiles around a well-defined majority of profiles (Fig. 3). Five anomalous RS92-RS92 ΔT profiles are identified. Except for profile TF025 all ΔT are within ± 1 °C and all but two ΔT in the majority of profiles are within ± 0.25 °C. The majority of profiles suggest no bias in ΔT for paired RS92 sondes.

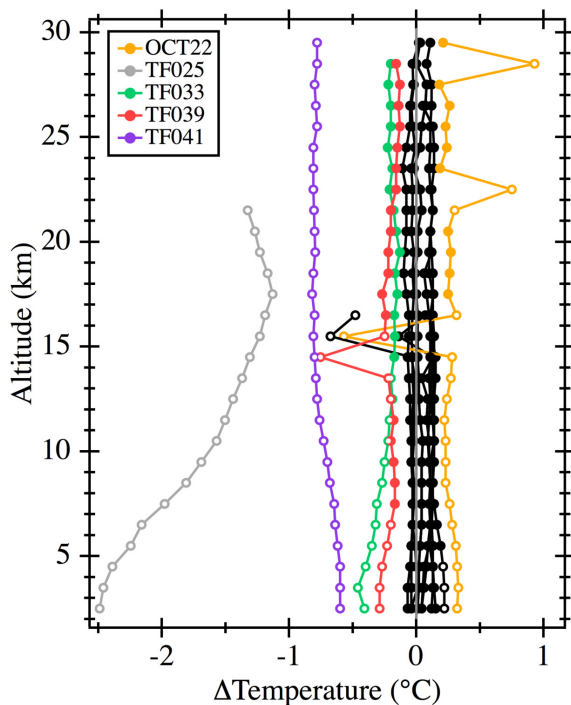


Fig. 3. Profiles of ΔT for paired RS92 sondes. Five anomalous difference profiles are identified by their flight numbers. Open circles denote the 28 % of ΔT that exceed manufacturer-quoted limits (1σ) for T measurement reproducibility (Table 2). All but 5 of these excessive ΔT belong to the five anomalous profiles.

Cumulative distribution functions (CDFs) of ΔT are constructed for both the RS92-iMet and RS92-RS92 comparisons (Fig. 4) to help evaluate biases between the temperature sensors. CDFs display the fractions of ΔT in all profiles (dashed curves), and in the majority of profiles (solid curves), that are less than the ΔT values specified by the bottom axis. In general, CDFs computed from all difference profiles (dashed) are skewed from normal by large differences in the anomalous profiles, while those determined from the majority of difference profiles (solid) portray more normal distributions. In some cases even near-normal CDFs are offset positively or negatively from zero, reflecting potential biases between sensors.

To test the statistical significance of the ΔT we examine the mean, median, standard deviation range (mean $\pm 1\sigma$) and inner-68 % range (16th to 84th percentiles) for each sensor pair (Table 3). The inner-68 % range, equal to the standard deviation range of a normal distribution of ΔT , can be read directly from a CDF (or from Table 3). The standard deviation ranges are displayed as horizontal error bars to the right of the CDFs (Fig. 4). Good agreement between the inner-68 % and standard deviation ranges, and between the median and mean ΔT (Table 3) indicates that the information garnered from CDFs and Gaussian statistics can be meaningfully meshed together. In no way do these agreements infer

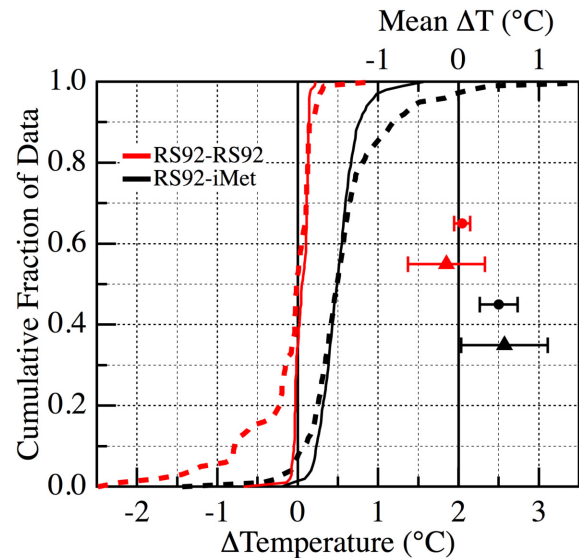


Fig. 4. Cumulative distribution functions and Gaussian statistics for RS92-RS92 ΔT (red) and RS92-iMet ΔT (black). CDF data fractions (left axis) are associated with the bottom axis scale. CDFs of RS92-iMet ΔT in all profiles (dashed) and in the majority of profiles (solid) are positively offset from zero and show inner-68 % ranges that do not include zero, signifying that the median ΔT values ($\sim 0.5^\circ\text{C}$) are statistically different from zero. To the right of the CDFs, means (markers) and standard deviations (error bars) of ΔT are presented for all profiles (triangles) and for the majority of profiles (dots), using the top axis scale. The standard deviation ranges for RS92-iMet ΔT (black) do not include zero, revealing mean ΔT values ($\sim 0.5^\circ\text{C}$) that are statistically different from zero. Black vertical lines at $\Delta T = 0$ (for both axes) are included as visual guides.

that the distribution of ΔT is normal. Using CDFs and Gaussian statistics, the median ΔT is significantly different from zero when the inner-68 % range does not include zero, and the mean ΔT is statistically different from zero when the standard deviation range does not include zero (Fig. 4). If the median and mean ΔT are both statistically different from zero the T bias is considered statistically significant. If the mean ΔT is statistically different from zero but the median ΔT is not (or vice-versa) the T bias is considered to have marginal statistical significance.

For the majority of profiles of RS92-iMet ΔT and RS92-RS92 ΔT there is good agreement between the means and medians, and between the standard deviation ranges and inner-68 % ranges (Table 3). CDFs for RS92-iMet ΔT are positively offset by $\sim 0.5^\circ\text{C}$ (Fig. 4), similar to the mean (0.49°C) and median (0.48°C) differences (Table 3) and suggestive of a bias. The fact that neither the standard deviation range nor the inner-68 % range includes zero confirms there is a statistically significant bias between the RS92 and iMet T sensors. The mean and median of RS92-RS92 ΔT for the majority of profiles are both near zero

Table 3. Measurement difference statistics.

Parameter	Sensor 1 ^a	Sensor 2	<i>N</i> ^b	Mean	Median	1 σ Range		Inner-68% Range		Units ^c
Temperature	RS92	iMet	26	0.57	0.49	0.03	1.11	0.21	0.94	°C
Temperature	RS92	iMet	19	0.49	0.48	0.26	0.73	0.28	0.69	°C
Temperature	RS92	RS92b	14	−0.15	−0.02	−0.63	0.33	−0.43	0.13	°C
Temperature	RS92	RS92b	9	0.04	0.05	−0.06	0.14	−0.04	0.13	°C
Pressure	RS92	iMet	26	−0.16	0.38	−2.35	2.04	−0.64	0.98	hPa
Pressure	RS92	iMet	22	0.33	0.43	−0.37	1.04	−0.30	1.00	hPa
Pressure	RS92	RS92b	14	−0.14	−0.14	−0.37	0.08	−0.35	0.04	hPa
RH	RS92	RS92b	14	−0.5	−0.3	−1.6	0.5	−0.9	0.1	% RH
RH	RS92	RS92b	11	−0.3	−0.3	−0.7	0.1	−0.7	0.1	% RH
RH	FP _{RS92}	FP _{IMET}	26	−0.9	−0.5	−2.1	0.3	−1.6	−0.1	% RH
RH	FP _{RS92}	FP _{IMET}	19	−0.6	−0.4	−1.2	0.1	−1.0	−0.1	% RH
RH	RS92	FP _{RS92}	26	−1.0	−0.7	−3.9	2.0	−2.1	0.5	% RH
RH	RS92	FP _{RS92}	21	−0.6	−0.6	−2.0	0.7	−1.6	0.2	% RH
RH	RS92	FP _{IMET}	26	−1.9	−1.2	−5.0	1.3	−3.0	−0.1	% RH
RH	RS92	FP _{IMET}	22	−1.2	−1.1	−2.8	0.3	−2.5	−0.1	% RH
H ₂ O	FP _{RS92}	FP _{IMET}	26	−0.5	0.0	−2.0	1.0	−1.6	0.2	%
H ₂ O	FP _{RS92}	FP _{IMET}	22	−0.7	−0.1	−2.0	0.5	−1.7	0.0	%
Altitude	RS92	iMet	26	−0.04	0.00	−0.15	0.08	−0.13	0.04	km
Altitude	RS92	RS92b	14	0.00	0.00	−0.03	0.04	−0.01	0.02	km

Statistics were computed from median measurement differences in 1-km altitude bins for each profile. Means, medians and statistical ranges printed in bold typeface represent measurement differences that are statistically different from zero.

^a Sensors FP_{RS92} and FP_{IMET} indicate the parameter was calculated from frost point hygrometer measurements using coincident data from the RS92 and iMet radiosondes, respectively. RH calculations require coincident temperature data while H₂O mixing ratio calculations require coincident pressure data. Only RH differences below 20 km altitude were used to compute the RH difference statistics.

^b The number of difference profiles included. Rows with repeated parameters and sensor pairs but a reduced number of profiles indicate that the anomalous difference profiles were excluded.

^c Units for H₂O mixing ratio differences are percentages because they are relative to the mixing ratio.

(0.04 and 0.05 °C) and are not statistically different from zero (Table 3).

We now compare ΔT in all difference profiles and in the majority of profiles to manufacturer-quoted measurement uncertainty and reproducibility values for T measurements (Table 2). The combined uncertainty limits for RS92-iMet T differences are given at the 2σ level, so it is expected that only $\sim 5\%$ of the observed ΔT should exceed these limits. If the 7 anomalous ΔT profiles are included (excluded), 38 % (31 %) of the RS92-iMet ΔT exceed the combined uncertainty limits (Table 4). The “excessive” differences are clearly not limited to the anomalous difference profiles (Fig. 2), but are predominantly attributable to the RS92-iMet T bias (0.5 °C) being nearly as large as the upper limit of uncertainty (0.58 °C). A disproportionate 55 % of the excessive RS92-iMet ΔT in the majority of profiles are concentrated above 16 km (Fig. 2). Interestingly, every ΔT in profiles TF028 and TF028b is well within the ± 0.58 °C uncertainty limits, and in fact within ± 0.3 °C (Fig. 2), but these two profiles are anomalous because they do not exhibit the 0.5 °C bias common to the majority of difference profiles.

Vaisala’s T measurement reproducibility limits are quoted as 1σ uncertainties, hence we expect $\sim 32\%$ of the RS92-RS92 ΔT to exceed these limits. The reproducibility limits are quoted in three different P regimes separated by 20 and 100 hPa (Table 2), so we transform these into altitude intervals separated by 26 and 16 km, respectively, to coincide with the altitude-binned measurement differences. Median RS92-RS92 ΔT in all 14 profiles exceed the reproducibility limits 28 % of the time (Table 4), with 95 % of the excessive ΔT belonging to the anomalous profiles. Only 3 % of ΔT in the majority of profiles exceeds the combined reproducibility limits (Table 4).

Flights with three radiosondes and at least one anomalous difference profile can be qualitatively assessed to ascertain which radiosonde performed irregularly. Three of the 6 flights with anomalous ΔT profiles also had payloads with 3 radiosondes: TF025, TF028 and TF041. For TF025, ΔT for the secondary RS92 and iMet (TF025b in Fig. 2) and for the two RS92 sondes (Fig. 3) are anomalous and roughly mirror images, while the primary RS92-iMet ΔT profile (TF025) conforms to the majority of profiles. These

Table 4. Percentages of differences exceeding radiosonde measurement uncertainties.

<i>P</i> Range ^a hPa	Alt Range ^b km	All profiles					Majority of profiles				
		RS92-iMet ^c		RS92-RS92 ^d			RS92-iMet ^c		RS92-RS92 ^d		
		<i>P</i> hPa	<i>T</i> °C	<i>P</i> hPa	<i>T</i> °C	RH %	<i>P</i> hPa	<i>T</i> °C	<i>P</i> hPa	<i>T</i> °C	RH %
3–20	>26				17					0	
4–100	>16	41		8			40		8		
4–1070	>0	30	38	11	28	5	23	31	11	3	0
20–100	16–26				19					2	
100–400	7–16	21					10				
100–1080	0–16			12	34				12	3	
400–1070	0–7	18					4				

Percentages of 1-km bin median differences that exceed measurement uncertainties or repeatability values determined from manufacturers' specifications in the designated pressure ranges (Table 2).

^a Pressure ranges over which the specified percentages are valid.

^b Altitude ranges over which the specified percentages are valid in this study.

^c Percentages exceeding the combined measurement uncertainty values (2σ).

^d Percentages exceeding the measurement reproducibility values (σ).

observations point to irregular data from the secondary RS92. The same arguments for flight TF041 also reveal irregular data from the secondary RS92. For flight TF028 the RS92-RS92 ΔT conforms to the majority of profiles (Fig. 3) while the two RS92-iMet ΔT profiles TF028 and TF028b are anomalous and similar (Fig. 2), implicating irregular data from the iMet sonde as the reason for the anomalous profiles.

CDFs and Gaussian statistics can expose a significant bias between two sensors, but a lack of overall bias for the whole profile does not preclude there being a significant altitude-dependent bias. Statistics in Table 3 show that RS92-RS92 ΔT are not significantly different from zero, but is this the fortuitous cancellation of negative biases at low altitudes and positive biases at high altitudes (or vice-versa)? Interestingly, some anomalous RS92-iMet ΔT profiles have conforming values at the surface that grow large and anomalous at higher altitudes (e.g., TF027 in Fig. 2). Other RS92-iMet anomalous profiles have large ΔT at the surface that decrease with altitude but never conform to the majority of profiles (e.g., TF042). In contrast, all 5 anomalous RS92-RS92 ΔT profiles have non-conforming differences all the way from the surface to the highest altitudes (Fig. 3). Reasons for the anomalous ΔT profiles are not definitively known but may include sensor production variability including poor calibration, sensor damage before launch, or sensor measurements being compromised by in-flight radio frequency interferences from radiosonde transmitters. Regardless, to achieve the highest quality sounding data, radiosondes with large T measurement offsets at the surface should have been rejected if pre-flight checks revealed them.

To further explore T measurement differences and their behavior with altitude we binned the ΔT for the majority of

profiles in 5-km altitude bins and computed CDFs and Gaussian statistics for each bin. A layer width of 5 km was chosen to populate each bin with 50–100 ΔT values and to simplify the plots and interpretation of the altitude-dependent statistics. As with the CDFs and Gaussian statistics for all ΔT , agreement between the mean and median, and between the standard deviation and inner-68 % ranges in each 5-km altitude bin are considered indicators that CDF and Gaussian statistics can be combined to evaluate T biases in each bin. Similar to our assessments of full profile biases, we link the statistical significance of biases in 5-km bins to standard deviation ranges and inner-68 % ranges that do not include zero (Fig. 5). For the RS92-iMet ΔT , statistically significant biases are exposed in every 5-km bin, with bin means and medians ranging from 0.3 to 0.7 °C. Though there is a visual suggestion that the RS92-iMet ΔT are altitude dependent (Fig. 2), overlap between the wide standard deviation and inner-68 % ranges in all six 5-km altitude bins (Fig. 5) precludes statistically significant altitude dependence. The same analysis of RS92-RS92 ΔT reveals no significant biases in any of the six 5-km layers (Fig. 5).

3.2 Pressure

Four of the 26 RS92-iMet ΔP profiles stand out as anomalous relative to the majority of profiles (Fig. 6). Just over 30 % of ΔP for all 26 RS92-iMet profiles exceed the combined 2σ uncertainty limits (Tables 2 and 4). Exclusion of the 4 anomalous profiles reduces this fraction to 23 %. Combined uncertainties for the RS92 and iMet sondes include an additional P boundary of 400 hPa (Table 2) that translates to an altitude boundary of 7 km. A disproportionate 83 % of

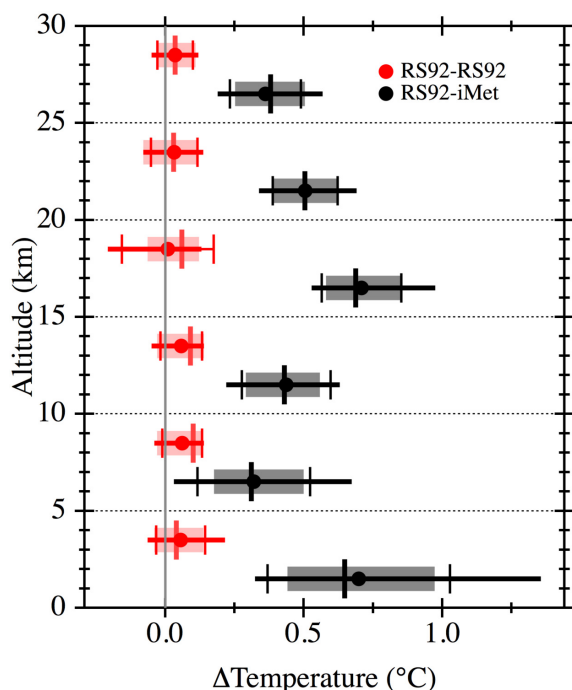


Fig. 5. Statistics of ΔT for RS92-iMet (black) and RS92-RS92 (red) in 5-km altitude bins. The mean (dot), median (thick vertical dash), inner-68 % range (shaded bar), standard deviation range (bounded by two thin vertical dashes) and 5th to 95th percentile range (horizontal line) of ΔT for the majority of profiles are presented. In each altitude bin both the standard deviation range and inner-68 % range of RS92-iMet ΔT do not include zero, revealing statistically significant biases over the entire altitude range. For RS92-RS92 ΔT both the standard deviation range and inner-68 % range of every altitude bin include zero, exposing no significant biases.

the excessive ΔP are located above 16 km (Fig. 6) where the uncertainty limits shrink from ± 1.12 to ± 0.78 hPa (Table 2). The majority of difference profiles (Fig. 6) suggest there may be a small negative bias near the surface and small positive bias above ~ 16 km.

The ΔP for all 14 dual-RS92 profiles are smaller and more consistent (Fig. 7) than those for RS92-iMet profiles (Fig. 6). None of the RS92-RS92 ΔP profiles are deemed anomalous though there is curious ~ 0.2 hPa split in the profiles above 14 km. We explored but subsequently dispelled the idea that this split was caused by small differences in P measurements by RS92-K and -SGP sondes during their 3 flights together (Table 1). Though interesting, the vast majority of ΔP that create this high-altitude split are within the manufacturer's reproducibility limits (σ), and for all 14 profiles of RS92-RS92 ΔP only 11 % of the differences exceed these limits.

For the RS92-iMet majority of ΔP profiles there is fairly good agreement between the mean and median differences, and between the standard deviation and inner-68 % ranges. Though the mean (0.33 hPa) and median (0.43 hPa) suggest

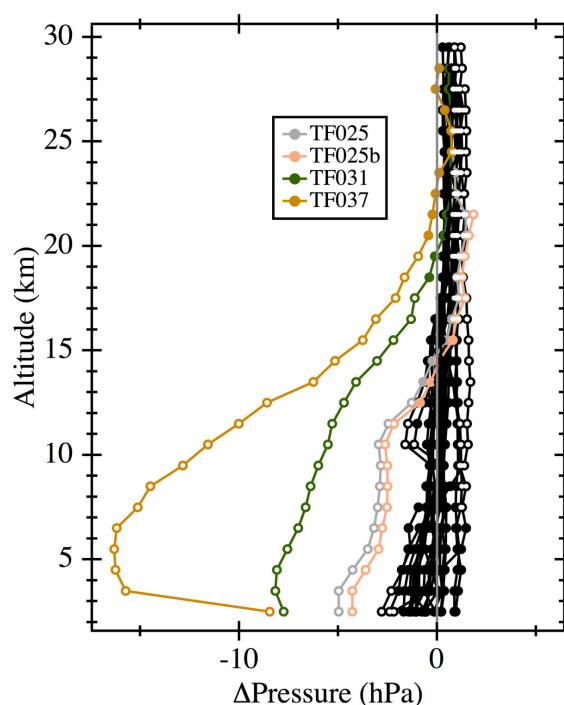


Fig. 6. The 26 RS92-iMet ΔP profiles, of which 4 are anomalous. If the anomalous profiles are included (excluded) 30 % (23 %) of the ΔP (open circles) exceed the combined measurement uncertainties (2σ) presented in Table 2, and 83 % of these excessive ΔP are located above 16 km.

a bias, both the standard deviation and inner-68 % ranges (Fig. 8) are wide enough to include zero. Overall, the variability in RS92-iMet ΔP , even for the majority of profiles, is large enough to negate the statistical significance of these small differences.

The CDFs and Gaussian statistics of RS92-RS92 ΔP in all 14 profiles (Fig. 8, Table 3) portray a distribution of differences that is more normal than for the RS92-iMet ΔP , albeit with a slight negative offset. The median and mean differences of -0.14 hPa are suggestive of a small negative measurement bias, but the inner-68 % and standard deviation ranges signify they are not statistically different from zero.

As was done above for anomalous ΔT profiles, we can identify which radiosonde produced irregular P data if three radiosondes were flown. For ΔP there is only one such flight (TF025) with anomalous profiles of RS92-iMet ΔP (TF025 and TF025b in Fig. 6). The similarity of these two anomalous profiles, along with a RS92-RS92 profile that conforms to the majority of profiles (Fig. 7), point to erratic iMet pressure data for this flight. Note that all of the anomalous RS92-iMet ΔP profiles have large differences near the surface that diminish with altitude and conform to the majority of profiles above 20 km.

We also examine the anomalous ΔP profiles by comparing the pressures measured directly by radiosondes with

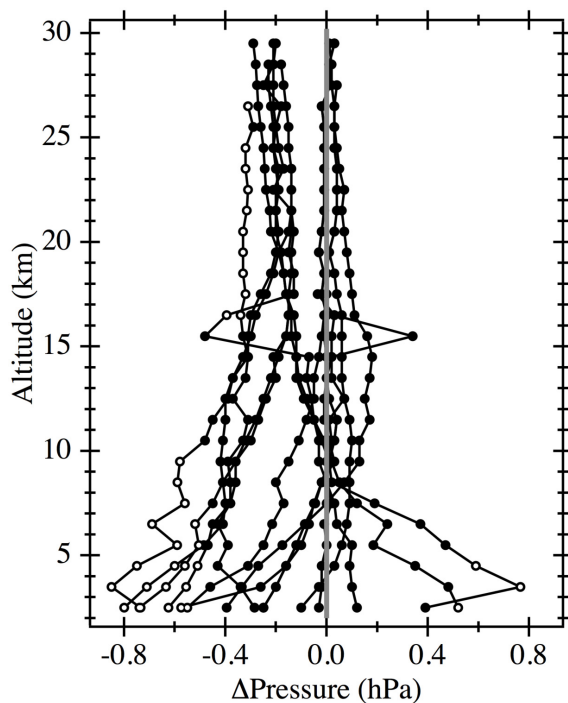


Fig. 7. The 14 ΔP profiles for paired RS92 sondes. None stand out as anomalous and only 11 % of ΔP exceed the manufacturer-quoted limits (1σ) for P measurement reproducibility in Table 2.

geopotential pressures calculated from the iMet GPS altitudes. We use the hypsometric equation that relates changes in air density, calculated from sequential radiosonde measurements of P , T , and RH, to changes in geopotential height, a proxy for altitude. Rearrangement of this equation permits the calculation of incremental changes in pressure from measured incremental differences in altitude, T , and RH. Employing GPS altitudes to derive pressures is preferable to using geopotential heights because the latter are calculated from P , T and RH measurements that may be potentially biased, while GPS altitudes are free from these biases. The geopotential pressures derived from GPS altitudes do, however, depend on T and RH measurements, and on a measured pre-launch surface pressure, so they are not completely independent of potential measurement biases. Nonetheless, the geopotential pressures derived from GPS altitudes provide two additional sets of P data that can be used to assess the quality of radiosonde P measurements.

Differences between the geopotential pressures derived from GPS altitudes and the pressures measured directly by the iMet sondes, calculated for each profile as median differences in 1-km altitude bins (ΔP_{GPS}), average -0.9 ± 1.7 hPa over all flights. Excluding the three flights with anomalous RS92-iMet ΔP profiles, the average ΔP_{GPS} becomes -0.6 ± 1.0 hPa. A similar analysis of all 26 profiles of ΔP_{GPS} for RS92 pressures yields an average of -0.8 ± 0.9 hPa. Though somewhat noisy, the ΔP_{GPS} are

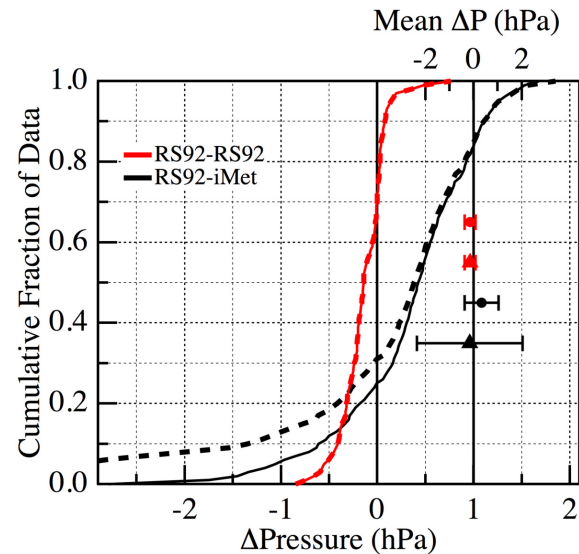


Fig. 8. CDFs and Gaussian statistics for RS92-RS92 ΔP (red) and RS92-iMet ΔP (black) that present the same information as for ΔT in Fig. 4. The red dashed and solid curves are identical because none of the RS92-RS92 ΔP profiles are anomalous. Though the CDFs of RS92-iMet ΔP are offset positively by about 0.5 hPa, their standard deviation ranges and inner-68 % ranges indicate no significant biases.

useful in assessing the four anomalous RS92-iMet ΔP profiles (Fig. 6). The four anomalous ΔP profiles in Fig. 6 are nearly identical to those of ΔP_{GPS} for iMet pressure data (not shown) but bear no resemblance to the ΔP_{GPS} profiles for RS92 pressure data (not shown).

The necessity of examining measurement differences for altitude-dependent biases is demonstrated by the RS92-iMet ΔP statistics for each 5-km altitude bin (Fig. 9). The overall statistics for ΔP in the majority of profiles exposed no significant bias (Table 3), but these 5-km altitude-binned statistics certainly do (Fig. 9). Mean and median P differences in the 5-km altitude bins increase smoothly with altitude from -0.6 ± 0.9 hPa (0–5 km) to 0.8 ± 0.3 hPa (25–30 km), but only above 15 km are the biases statistically significant (Fig. 9). Good agreement between the means and medians, and between the inner-68 % and standard deviation ranges attests to near-normal distributions of ΔP in most of the 5-km bins. Without several large ΔP in the 0–5 km bin the negative bias would be statistically significant. The RS92-RS92 standard deviation ranges for ΔP in the two highest 5-km bins do not include zero but the inner-68 % ranges do, signifying marginally significant biases of -0.01 and -0.02 hPa. It is not clear why a P bias with even marginal significance exists between the primary and secondary RS92 sondes, but we reiterate this is not caused by P differences between the 3 paired K and SGP sondes.

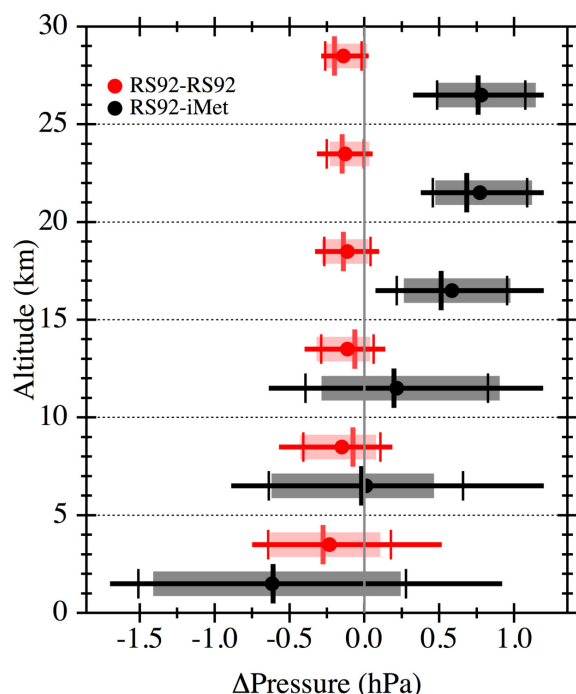


Fig. 9. Statistics of ΔP for RS92-iMet (black) and RS92-RS92 (red) in 5-km altitude bins. The markers, bars, dashes and lines represent the same statistics of differences described for Fig. 5. Statistically significant biases in RS92-iMet ΔP are revealed above 15 km. For RS92-RS92 ΔP only the standard deviation ranges in the 20–25 and 25–30 km bins do not include zero.

3.3 Relative humidity

First we compare the direct RH measurements by pairs of RS92 sondes on the same balloons. It is noted here that all RH values are presented in units of % RH with respect to the saturation vapor pressure over liquid water (not ice), as these are the RH values reported by the DigiCORa MW31, STRATO and SkySonde sounding systems. As described earlier, the RS92 RH values compared in this work have been corrected using well-documented algorithms (Miloshevich et al., 2004; Vömel et al., 2007b; Miloshevich et al., 2009). The impacts of these RH corrections for the MOHAVE-2009 campaign are illustrated in Fig. 1 of Leblanc et al. (2011). And again, RH measurements by the iMet sondes are not compared here because of their poor quality.

Three of the 14 vertical profiles of ΔRH for paired RS92 sondes are anomalous (Fig. 10). For these 3 sonde pairs there is little indication from the near-surface ΔRH that differences would be anomalously large at higher altitudes. Only 5 % of ΔRH for all 14 profiles exceed the measurement reproducibility limits of ± 2 % RH (σ), and exclusion of the three anomalous profiles removes every excessive ΔRH . Since 95 % of the RH values measured above 20 km are < 3 % RH (Fig. 10, top axis), ΔRH at these altitudes are unlikely to exceed the ± 2 % RH reproducibility limits so we

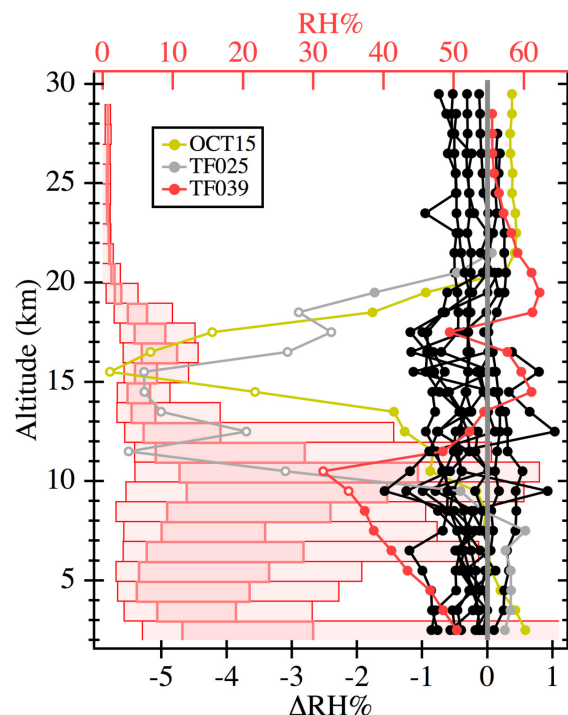


Fig. 10. Profiles of ΔRH for 14 pairs of RS92 sondes, of which 3 are anomalous. Only 5 % of ΔRH (open circles) exceed the manufacturer-quoted limits of ± 2 % RH (1σ) for measurement reproducibility. The lighter and darker red horizontal bars depict the inner-90 % ranges and interquartile ranges, respectively, of RS92 RH measurements in 1-km altitude bins, referenced to the top axis scale. These data illustrate how RH values become very small above ~ 20 km. With 95 % of RH < 3 % RH above 20 km, ΔRH are constrained to be small so these high-altitude data are excluded from the difference statistics.

exclude them from RH measurement comparison statistics except those for the 5-km altitude bins.

In lieu of RS92-iMet RH measurement comparisons we instead examine differences between frost point hygrometer-based RH values calculated using RS92 temperatures (FP_{RS92}) and those calculated using iMet temperatures (FP_{iMet}). Frost point hygrometers make direct measurements of the partial pressure of water vapor (P_w) in air and require coincident T measurements by a radiosonde to calculate the saturation vapor pressures over liquid water (P_{sat}) needed to convert P_w to % RH ($= 100 \times P_w/P_{sat}$). Differences between the frost point hygrometer-based RH values calculated using independent T measurements by RS92 and iMet sondes quantitatively demonstrate the influences of RS92-iMet T biases on the RH values.

Differences between the two sets of frost point hygrometer-based RH values ($FP_{RS92} - FP_{iMet}$) are displayed as vertical profiles of 1-km bin medians (ΔRH). Seven of the 26 ΔRH profiles are anomalous (Fig. 11), and not surprisingly 4 of these were also anomalous for ΔT (Fig. 2).

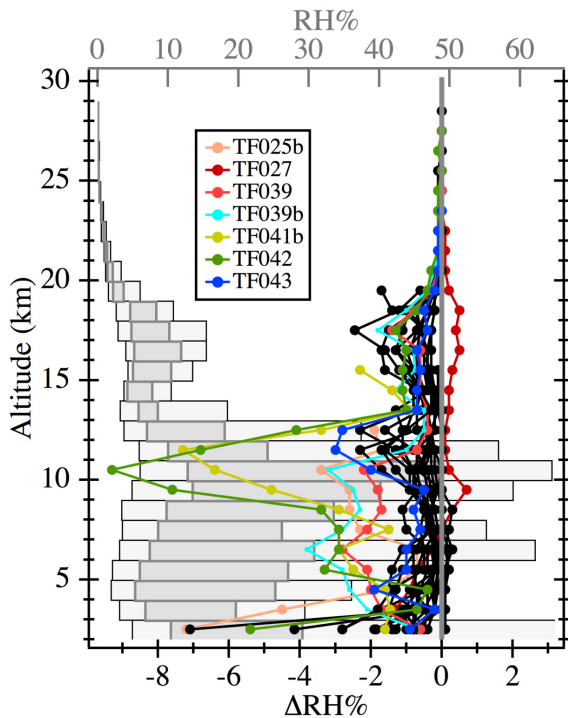


Fig. 11. Profiles of differences between the RH values calculated from frost point hygrometer measurements using independent temperature measurements by RS92 sondes (FP_{RS92}) and iMet sondes (FP_{iMet}). Four of the 7 anomalous ΔRH profiles identified here were also anomalous for RS92-iMet ΔT (Fig. 2). Similar to Fig. 10, the lighter and darker gray horizontal bars depict the inner-90 % and interquartile ranges of FP_{RS92} RH values, demonstrating the rapid decline of RH above 20 km.

The anomalous profiles in Figs. 2 and 11 are not exactly the same because the Clausius-Clapeyron equation defines P_{sat} as a non-linear function of T . The dependence of RH on T is inverse, meaning a positive ΔT (Fig. 2) will produce a negative ΔRH (Fig. 11), but the same ΔT at high and low temperatures will have very different effects on ΔRH .

Figure 12 presents the CDFs of ΔRH for paired RS92 sondes and for $FP_{RS92}-FP_{iMet}$. All ΔRH above 20 km are excluded for the reason given above. For paired RS92 sondes both the mean and median ΔRH for the majority of profiles suggest a bias of -0.3% RH (Table 3) but neither is statistically different from zero. The median $FP_{RS92}-FP_{iMet}$ ΔRH of -0.4% RH for the majority of profiles (Table 3) is statistically different from zero but the mean is not, signifying a bias of marginal significance. This is an unexpected result given the statistically significant $0.5 \pm 0.2^\circ\text{C}$ bias in RS92-iMet ΔT .

Statistics for the RS92-RS92 ΔRH in 5-km altitude bins reveal only marginally significant negative biases in the 0–5 and 25–30 km bins (Fig. 13). For $FP_{RS92}-FP_{iMet}$ ΔRH there are statistically significant negative biases in the 10–15 and 15–20 km bins, with mean and median differences ranging

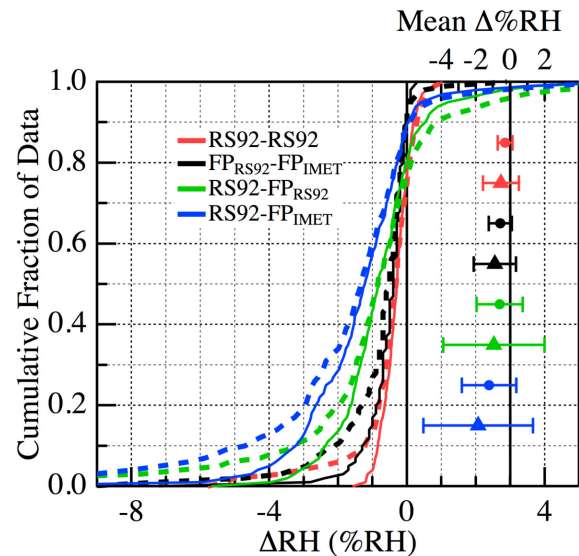


Fig. 12. CDFs and Gaussian statistics of ΔRH for RS92-RS92, $FP_{RS92}-FP_{iMet}$, RS92- FP_{RS92} and RS92- FP_{iMet} . The latter two sets of RH reflect differences between direct RH measurements by RS92 sondes and RH values calculated from frost point hygrometer measurements using T data from the RS92 and iMet sondes, respectively. ΔRH above 20 km were excluded from each comparison. Though the CDF for every ΔRH data set is shifted negatively, only the inner-68 % ranges of $FP_{RS92}-FP_{iMet}$ and RS92- FP_{iMet} expose median ΔRH values that are statistically different from zero. All of the standard deviation ranges include zero.

from -0.5 to -0.7% RH. Below 10 km the median $FP_{RS92}-FP_{iMet}$ ΔRH of -0.7 and -0.2% RH are statistically different from zero, but the mean differences are not. It is again surprising that $FP_{RS92}-FP_{iMet}$ ΔRH biases are statistically significant only between 10 and 20 km while RS92-iMet ΔT are significant in every 5-km altitude bin (Fig. 5).

Armed with a better quantitative understanding of the frost point hygrometer-based ΔRH induced by RS92-iMet ΔT , we now compare frost point-based RH values with direct RH measurements by the RS92 sondes. Even the largest RH differences induced by ΔT (Fig. 11) are dwarfed by ΔRH of $>10\%$ RH in both the RS92- FP_{RS92} and RS92- FP_{iMet} comparisons (Figs. 14 and 15). Except for TF042 the four anomalous profiles in both Figs. 14 and 15 are similar, indicating that RS92-iMet ΔT are only weak contributors to these much larger ΔRH . The majority of profiles of RS92- FP_{RS92} ΔRH suggest negative biases (Fig. 14), but even when the anomalous profiles are excluded the standard deviation and inner-68 % ranges of ΔRH both include zero (Fig. 12, Table 3). For RS92- FP_{iMet} ΔRH in the majority of profiles only the median of -1.1% RH is statistically different from zero.

Our method of identifying which instrument produced irregular data is applicable for two flights with anomalous ΔRH profiles. For flight TF041 the RS92- FP_{RS92} ΔRH is similar for both the primary and secondary RS92 sondes

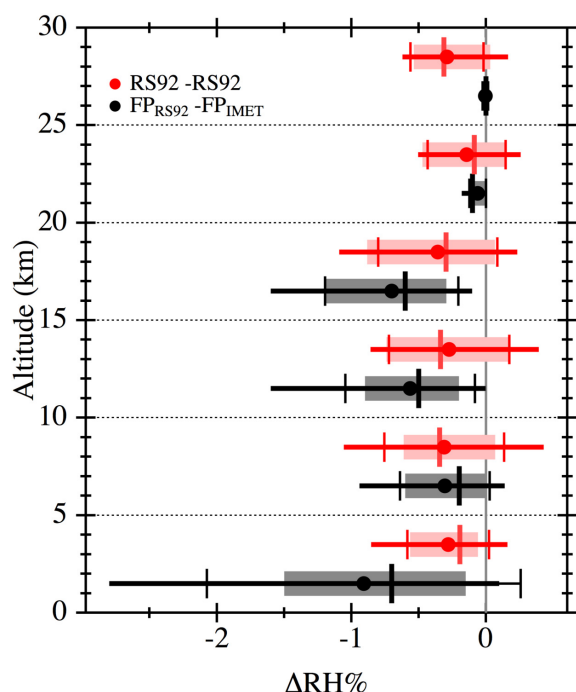


Fig. 13. Statistics of ΔRH for RS92-RS92 (red) and $FPRS92-FP_{IMET}$ (black) in 5-km altitude bins. Each of the symbols represents the same statistics of differences described for Fig. 5. For $FPRS92-FP_{IMET}$ ΔRH there are statistically significant biases in the 10–15 and 15–20 km bins and median differences in the 0–5 and 5–10 km bins that are statistically different from zero. Median and mean RS92-RS92 ΔRH are statistically different from zero in the 0–5 and 25–30 km bins, respectively.

(TF041 and TF041b in Fig. 14), as is the case for RS92- FP_{IMET} ΔRH in these same profiles (Fig. 15). The ΔRH appear to be largely independent of the choice of primary or secondary RS92 RH data and the choice of RS92 or iMet T data. These observations together imply that the CFH produced irregular data between 5 and 12 km during flight TF041. For flight TF025 there are large ΔRH for RS92-RS92 (Fig. 10) and RS92b-CFH (TF025b, Figs. 14 and 15), but not for RS92-CFH (TF025). The evidence points to high-biased RH measurements by the secondary RS92 sonde (TF025b) in the 10–13 km layer. The source of irregular RH data for TF026 (Figs. 14 and 15) cannot be identified because there was only one RS92 sonde on the balloon, but the large ΔRH certainly do not stem from RS92-iMet T differences.

Altitude-dependent statistics for the RS92- $FPRS92$ and RS92- FP_{IMET} ΔRH (Fig. 16) reveal significant negative biases of -0.9 and -1.6 % RH for the 15–20 km bin, respectively, and significant positive biases of 0.6 % RH (for both) in the 25–30 km bin. Strangely, neither bias in the 20–25 km bin is even marginally significant, giving the appearance of a smooth transition in ΔRH biases from negative at 15–20 km to positive at 25–30 km. RS92- FP_{IMET} biases in the 0–5 and 10–15 km bins are marginally significant and statistically

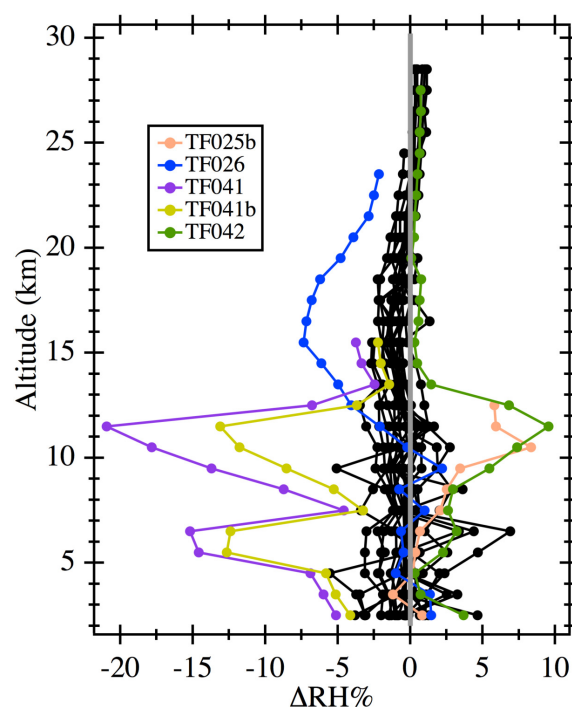


Fig. 14. Profiles of differences between RH measured directly by RS92 sondes and RH calculated from frost point hygrometer measurements using coincident RS92 temperature data (RS92- $FPRS92$). Five profiles are identified as anomalous. The mean \pm standard deviation of ΔRH for the majority of difference profiles is -0.6 ± 1.3 % RH. The standard deviation range is much narrower than the RS92 manufacturer's total RH measurement uncertainty range of ± 5 % RH.

significant, respectively. We note that none of these biases are outside the ± 2 % RH limits of measurement reproducibility quoted by Vaisala.

3.4 Water vapor mixing ratios

Water vapor volume mixing ratios are calculated from the frost point hygrometer measurements of P_w and coincident P measurements by the RS92 and iMet sondes. As we did for frost point hygrometer-based RH values, we demonstrate how the RS92-iMet ΔP quantitatively influence water vapor mixing ratio differences ($FPRS92-FP_{IMET} \Delta H_2O$). Given that water vapor mixing ratios during MOHAVE-2009 ranged from 10 000 ppmv at the surface to as low as 3 ppmv in the stratosphere, statistics for ΔH_2O are calculated relative to the mixing ratios themselves, in percentage units.

In simple terms, volume mixing ratios are inversely proportional to P such that a positive ΔP will produce a negative ΔH_2O . Since RS92-iMet ΔP are the only sources of H_2O mixing ratio differences examined in this comparison, vertical profiles of ΔH_2O (not shown) are roughly mirror images of the RS92-iMet ΔP profiles (Fig. 6). Statistics for all 26 RS92-iMet ΔH_2O profiles and for 22 profiles that exclude

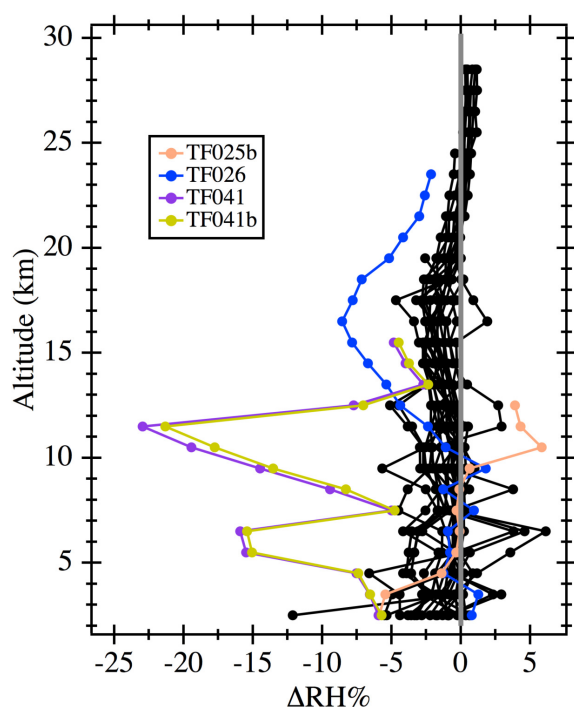


Fig. 15. Profiles of differences between RH measured directly by RS92 sondes and RH calculated from frost point hygrometer measurements using coincident iMet temperature data (RS92-FP_{IMET}). Each of the four anomalous profiles identified here were also anomalous in Fig. 14. The mean \pm standard deviation of Δ RH for the majority of difference profiles is -1.2 ± 1.5 RH, portraying a standard deviation range that fits wholly within the RS92 manufacturer's ± 5 % RH total measurement uncertainty range.

the 4 anomalous RS92-iMet ΔP profiles (Fig. 6) are presented in Table 3. These statistics reveal no significant biases in ΔH_2O . The median ΔH_2O is not statistically different from zero because the inner-68 % range for the 22 ΔH_2O profiles (-1.7 to 0.02 %) just barely includes zero.

The 5-km altitude bin statistics for FP_{RS92}-FP_{IMET} ΔH_2O reveal significant negative biases above 15 km that increase with altitude from -1 % to nearly -4 % (Fig. 17). Profiles of RS92-iMet ΔP (Fig. 9) and ΔH_2O (Fig. 17) are not perfect mirror images because small absolute ΔP become large relative ΔP (and therefore large ΔH_2O) at low pressures. The mean and median ΔH_2O above 15 km (Fig. 17) are sizeable fractions of the $< \pm 10$ % estimated uncertainty for CFH and FPH measurements of stratospheric water vapor, illustrating the need for accurate radiosonde P measurements to convert frost point hygrometer measurements into accurate water vapor mixing ratios.

3.5 Altitude

Here we compare the payload altitudes during flights calculated by the Vaisala DigiCORA software using RS92 P , T and RH measurements, and those computed by both the

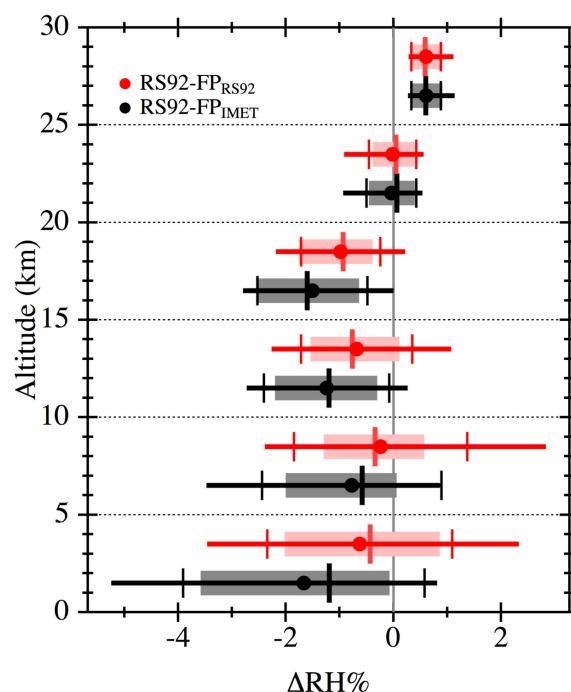


Fig. 16. Statistics of Δ RH for RS92-FP_{RS92} (red) and RS92-FP_{IMET} (black) in 5-km altitude bins. Each of the symbols represents the same statistics of differences described for Fig. 5. Statistically significant biases are revealed for both RS92-FP_{RS92} and RS92-FP_{IMET} in the 15–20 and 25–30 km bins. There is also a significant RS92-FP_{IMET} bias at 10–15 km, and the median RS92-FP_{IMET} RH difference for the 0–5 km bin is statistically different from zero.

STRATO and SkySonde programs using similar measurements by the iMet sondes. These calculated altitudes, more correctly called geopotential heights, are determined independently by each sounding system using the hypsometric equation. As described before, this equation relates small changes in the geopotential height to small changes in air density calculated from sequential differences in P , T and RH as the balloon rises or falls. The initial geopotential height at launch is anchored to the known launch site elevation, but otherwise the geopotential height calculations depend solely on measured changes in P , T and RH. Sequential changes in geopotential height from the launch site elevation are cumulatively summed to estimate the balloon altitude at each timestamp after launch.

Biased measurements of P , T or RH may or may not induce biases in the calculated geopotential heights because only the differences between sequential measurements of these parameters are considered. Sequential differences computed from a measurement time series afflicted by a constant bias are exactly the same as those computed from the same measurements without the bias. Of course, a measurement bias that changes with altitude will induce an altitude-dependent bias in the computed geopotential heights. Since

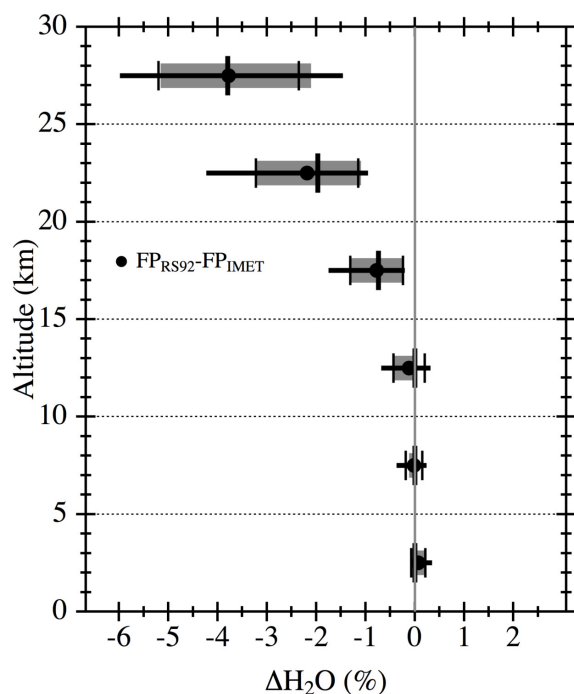


Fig. 17. Statistics of ΔH_2O present the relative differences between water vapor mixing ratios calculated from frost point hygrometer measurements using independent P measurements by RS92 sondes (FP_{RS92}) and iMet sondes (FP_{iMET}), in 5-km altitude bins. Each of the symbols represents the same statistics of differences described for Fig. 5. The ΔH_2O have statistically significant negative biases above 15 km that result from positive biases in RS92-iMet ΔP above 15 km (Fig. 9).

these calculations are incremental the cumulative errors induced by small, altitude-dependent measurement biases can become large.

Statistics for the differences between geopotential heights calculated for the RS92 and iMet sondes and between those calculated for paired RS92 sondes (Table 3) reveal no significant biases. As expected, the ΔAlt for RS92 sonde pairs are near zero; both the median and mean ΔAlt are within ± 0.004 km and the standard deviation is ± 0.033 km. The RS92-iMet ΔAlt are more variable, with standard deviation and inner-68 % ranges of -0.15 to 0.08 km and -0.13 to 0.04 km, respectively. Overall, the agreement in geopotential heights calculated by the different sounding systems is very good, and somewhat surprising given that the RS92 DigiCORA may use a slightly different geopotential height algorithm than that employed by both the STRATO and SkySonde systems.

Altitude-dependent statistics in 5-km bins for both sets of ΔAlt are shown in Fig. 18. None of the bin statistics for RS92-RS92 ΔAlt reveal a significant bias (Fig. 18) while the RS92-iMet ΔAlt are significantly biased above 20 km. We presume that these high-altitude biases result from the high-altitude biases in RS92-iMet ΔP (Fig. 9), but we cannot

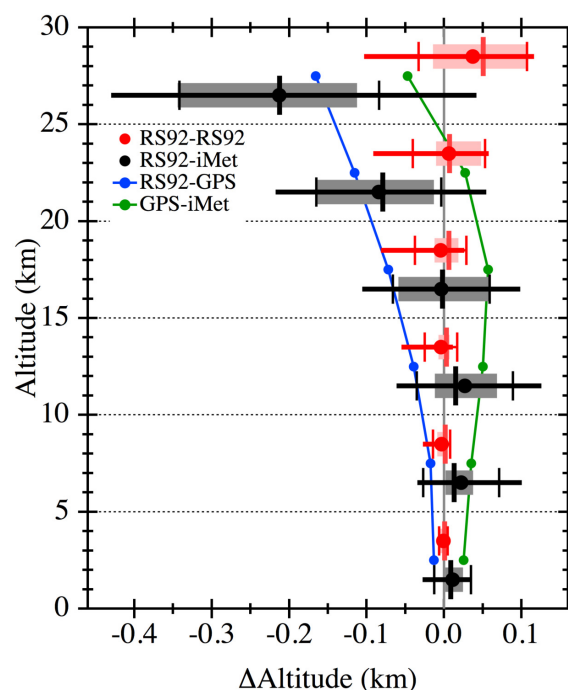


Fig. 18. Statistics of ΔAlt reflect differences in geopotential heights for paired RS92 sondes (red) and for RS92-iMet sondes (black) in 5-km altitude bins. Each of the symbols represents the same statistics of differences described for Fig. 5. Geopotential heights are calculated using P , T , and RH measurements by the radiosondes. The statistically significant negative biases in RS92-iMet ΔAlt above 20 km may result from disparities in RS92 and iMet P , T and RH measurements, or from possible differences in the geopotential height algorithms employed by the different sounding systems. Median differences between GPS altitudes and geopotential altitudes for the RS92 sondes (blue markers) and iMet sondes (green markers) in 5-km altitude bins, when combined, are consistent with the mean and median RS92-iMet geopotential height differences in every altitude bin.

rule out algorithmic differences between the RS92 and iMet sounding systems. Comparing the RS92 and iMet geopotential heights to the iMet GPS altitudes demonstrates that the combination of RS92-GPS and GPS-iMet altitude differences is consistent with the RS92-iMet ΔAlt biases (Fig. 18). The GPS altitudes agree better with geopotential heights from the iMet than with those from RS92 sondes. From this comparison we conclude that the geopotential heights and GPS altitudes are different by < 0.1 km below 20 km, < 0.2 km for 20–25 km, and < 0.4 km for 25–30 km.

4 Conclusions

We have rigorously compared coincident, in situ, balloon-borne measurements of T and P by RS92 and iMet-1-RSB radiosondes, and measurements of RH by RS92 sondes with RH values calculated from frost point hygrometer measurements. All sensor data compared were obtained from the same balloons, during ascent only. Profiles of median differences in 1-km altitude bins were constructed for each sensor pair, and for every comparison (except RS92-RS92 ΔP) we identified several anomalous difference profiles that do not conform to the majority of profiles (e.g., Fig. 2). Though the anomalous profiles do not always show poor sensor pair agreement at the surface, the rejection of sondes that performed poorly during pre-flight checkout would have reduced the number of anomalous profiles.

The measurement differences were analyzed with cumulative distribution functions (CDFs) and Gaussian statistics (e.g., Fig. 4). For each parameter we considered all profiles and only the majority of profiles in these analyses to gauge the magnitudes of differences with and without the influences of the anomalous profiles (Table 3). Statistics that include the anomalous profiles provide realistic estimates of sensor pair agreement if every sonde is trusted to perform within its measurement specifications. Statistics that exclude the anomalous profiles are more suitable for the detection of measurement biases, including altitude-dependent biases, and for comparison to manufacturers' measurement uncertainty and reproducibility limits.

A statistically significant, altitude-independent bias of $0.5 \pm 0.2^\circ\text{C}$ is revealed for RS92-iMet T differences (Table 3). This T bias propagates a median difference of -0.5% RH between frost point hygrometer-based RH values calculated independently using RS92 and iMet temperatures ($\text{FP}_{\text{RS92}} - \text{FP}_{\text{iMet}}$). Differences between direct RH measurements by RS92 sondes and frost point hygrometer-based RH values calculated using iMet temperatures ($\text{RS92} - \text{FP}_{\text{iMet}}$) also reveal a median difference of -1.1% RH that is statistically different from zero (Table 3). None of the RH comparisons exposed measurement biases with full statistical significance. We did not attempt to compare RH measurements by the iMet because of their poor quality during this campaign.

Altitude-dependent biases for RS92-iMet ΔP range smoothly from -0.6 hPa near the surface to 0.8 hPa at 25–30 km (Fig. 9). Above 15 km we find altitude-dependent biases of -1% to -4% between frost point hygrometer-based H_2O mixing ratios calculated independently using P measurements from RS92 and iMet sondes (Fig. 17). These biases stem from the altitude-dependent RS92-iMet P biases (Fig. 9) and illustrate how the accuracy of frost point hygrometer-based stratospheric H_2O mixing ratios very much depends on the accuracy of radiosonde P measurements.

Altitude-dependent differences of -0.1 to -0.2 km are found above 20 km between the geopotential altitudes

calculated by the RS92 DigiCORA sounding system and the iMet STRATO and SkySonde sounding systems (Fig. 18). These biases may result from differences in the radiosonde measurements of P , T and RH, or in the sounding system algorithms used to calculate geopotential heights.

The measurement differences observed during MOHAVE-2009 are compared to measurement uncertainty and reproducibility limits determined from manufacturers' specifications (Table 2). Differences between paired RS92 sondes exceeded the manufacturer-quoted reproducibility limits (Table 2) only 11, 28 and 5 % of the time for P , T and RH, respectively (Table 4), less than the 32 % expected for 1σ limits. Exclusion of the anomalous ΔT and ΔRH profiles reduces these fractions to 3 % (T) and 0 % (RH). RS92-iMet P and T differences exceeded their combined measurement uncertainty limits 30 and 38 % of the time, respectively, much more frequently than the 5 % expected for 2σ limits. Exclusion of the anomalous difference profiles reduces these percentages to 23 (P) and 31 % (T). The combined RS92-iMet measurement uncertainties for these parameters are clearly too small, especially above 16 km (<100 hPa) where 41 and 42 % of all ΔP and ΔT exceed the measurement uncertainties. Excluding the anomalous profiles slightly reduces these percentages to 40 (ΔP) and 36 % (ΔT). Disproportionate fractions of the excessive ΔP (83 %) and the excessive ΔT (55 %) are found above 16 km.

Acknowledgements. Thanks go to the entire staff of the magnificent JPL Table Mountain Facility for hosting the MOHAVE-2009 campaign. We are grateful for the assistance of T. Grigsby, M. Calhoun, D. Venable and A. Dirisu during the campaign. NOAA provided the financial support for D. Hurst, E. Hall and A. Jordan to participate in this campaign and for the 4 NOAA FPHs that were flown. Two anonymous reviewers provided well-conceived comments that helped strengthen this paper.

Edited by: H. Smit

References

- Brewer, A. W., Cwilog, B., and Dobson, G. M. B.: Measurements of absolute humidity in extremely dry air, *Proc. Phys. Soc.*, 60, 52–70, 1948.
- Goff, J. A.: Saturation pressure of water on the new Kelvin temperature scale, *Trans. Am. Soc. Heat. Ventil. Eng.*, 63, 347–354, 1957.
- Hurst, D. F., Oltmans, S. J., Vömel, H., Rosenlof, K. H., Davis, S. M., Ray, E. A., Hall, E. G., and Jordan, A. F.: Stratospheric water vapor trends over Boulder, Colorado: Analysis of the 30 year Boulder record, *J. Geophys. Res.*, 116, D02306, doi:10.1029/2010JD015065, 2011.
- Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, *Atmos. Meas. Tech.*, 3, 1217–1231, doi:10.5194/amt-3-1217-2010, 2010.

- Leblanc, T., Walsh, T. D., McDermid, I. S., Toon, G. C., Blavier, J.-F., Haines, B., Read, W. G., Herman, B., Fetzer, E., Sander, S., Pongetti, T., Whiteman, D. N., McGee, T. G., Twigg, L., Sumnicht, G., Venable, D., Calhoun, M., Dirisu, A., Hurst, D., Jordan, A., Hall, E., Miloshevich, L., Vömel, H., Straub, C., Kampf, N., Nedoluha, G. E., Gomez, R. M., Holub, K., Guttman, S., Braun, J., Vanhove, T., Stiller, G., and Hauchecorne, A.: Measurements of Humidity in the Atmosphere and Validation Experiments (MOHAVE)-2009: overview of campaign operations and results, *Atmos. Meas. Tech.*, 4, 2579–2605, doi:10.5194/amt-4-2579-2011, 2011.
- Mastenbrook, H. J. and Oltmans, S. J.: Stratospheric water vapor variability for Washington, D.C./Boulder, CO; 1964–82, *J. Atmos. Sci.*, 40, 2157–2165, 1983.
- Miloshevich, L. M., Paukkunen, A., Vömel, H., and Oltmans, S. J.: Development and validation of a time-lag correction for Vaisala radiosonde humidity measurements, *J. Atmos. Ocean. Tech.*, 21, 1305–1327, 2004.
- Miloshevich, L. M., Vömel, H., Whiteman, D. N., and Leblanc, T.: Accuracy, assessment and correction of Vaisala RS92 radiosonde water vapor measurements, *J. Geophys. Res.*, 114, D11305, doi:10.1029/2008JD011565, 2009.
- Seidel, D. J., Angell, J. K., Christy, J., Free, M., Klein, S. A., Lanzante, J. R., Mears, C., Parker, D., Schabel, M., Spencer, R., Sterin, A., Thorne, P., and Wentz, F.: Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature datasets, *J. Climate*, 17, 2225–2240, doi:10.1175/1520-0442(2004)017<2225:UISOLC>2.0.CO;2, 2004.
- Thorne, P. W., Parker, D. E., Christy, J. R., and Mears, C. A.: Uncertainties in climate trends: Lessons from upper-air temperature records, *B. Am. Meteorol. Soc.*, 86, 1437–1442, doi:10.1175/BAMS-86-10-1437, 2005.
- Titchner, H. A., Thorne, P. W., McCarthy, M. P., Tett, S. F. B., Haimberger, L., and Parker, D. E.: Critically reassessing tropospheric temperature trends from radiosondes using realistic validation experiments, *J. Climate*, 22, 465–485, 2009.
- Vömel, H., David, D. E., and Smith, K.: Accuracy of tropospheric and stratospheric water vapor measurements by the cryogenic frost point hygrometer: Instrumental details and observations, *J. Geophys. Res.*, 112, D08305, doi:10.1029/2006JD007224, 2007a.
- Vömel, H., Selkirk, H., Miloshevich, L., Valverde-Canossa, J., Valdes, J., Kyro, E., Kivi, R., Stolz, W., Peng, G., and Diaz, J. A.: Radiation dry bias of the Vaisala RS92 humidity sensor, *J. Atmos. Ocean. Tech.*, 24, 953–963, doi:10.1175/JTECH2019.1, 2007b.