

On the Accuracy of Vaisala RS41 versus RS92 Upper-Air Temperature Observations

BOMIN SUN

IMSG, and NOAA/NESDIS/Center for Satellite Applications and Research, College Park, Maryland

TONY REALE

NOAA/NESDIS/Center for Satellite Applications and Research, College Park, Maryland

STEVEN SCHROEDER

Texas A&M University, College Station, Texas

MICHAEL PETTEY AND RYAN SMITH

IMSG, and NOAA/NESDIS/Center for Satellite Applications and Research, College Park, Maryland

(Manuscript received 21 May 2018, in final form 7 February 2019)

ABSTRACT

The accuracy of Vaisala RS92 versus RS41 global radiosonde soundings, emphasizing stratospheric temperature, is assessed from January 2015 to June 2017 using ~311 500 RS92 and ~65 800 RS41 profiles and three different reference data sources. First, numerical weather prediction (NWP) model outputs are used as a transfer medium to produce relative RS92 and RS41 comparisons by analyzing observation minus NWP model background (OB–BG) and observation minus analysis (OB–AN) differences using the NOAA Climate Forecast System Reanalysis (CFSR; both comparisons) and the operational European Centre for Medium-Range Weather Forecasts (ECMWF) model (OB–AN comparison only). Second, GPS radio occultation (GPSRO) dry temperature profiles are directly compared with radiosondes, using GPSRO data from the University Corporation for Atmospheric Research (UCAR) Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) and EUMETSAT Radio Occultation Meteorology (ROM) Satellite Application Facility (SAF). Third, dual launches (RS92 and RS41 suspended from the same balloon) at five sites allow direct assessments. Comparisons of RS92 versus RS41 from all reference data sources are basically consistent. These two sondes agree well with global average temperature differences <0.1–0.2 K in the lower stratosphere from 51.5 to 26.1 hPa based on global stations and the dual launches. RS41 appears to be less sensitive than RS92 to changes in solar elevation angle. This study indicates that nighttime RS92 and RS41 radiosonde temperature biases are negligible, but infers a stratospheric cold bias (<0.5 K) in the CFSR and ECMWF model data.

1. Introduction

Balloonborne radiosonde observations (raobs) play a critical role in upper-air climate change detection, numerical weather prediction (NWP) data assimilation and forecasting, and satellite data calibration/validation (cal/val). Vaisala RS92 is a major sonde type in the current global operational upper-air network and a reference sonde in the Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN; Bodeker et al. 2016). However, RS92 has gradually been replaced by Vaisala RS41 starting in late 2013 and managing this

transition is a high priority in GRUAN because RS92 production ended in August 2017.

Vaisala RS41 includes new sensor technologies aimed at providing improvements in measurement accuracy for temperature, humidity, and other variables throughout the atmosphere (<https://www.vaisala.com/en/products/instruments-sensors-and-other-measurement-devices/soundings-products/rs41>). Understanding the measurement accuracy of this emerging radiosonde type is of great interest to the climate trend detection, NWP, and satellite communities.

Radiation-induced error is a major issue with radiosonde temperature measurements in the upper troposphere and lower stratosphere (UTLS). Ventilation around the

Corresponding author: Dr. Bomin Sun, bomin.sun@noaa.gov

DOI: 10.1175/JTECH-D-18-0081.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

instrument diminishes with altitude due to reduced air density, causing radiosonde temperatures to be typically warm biased in daytime due to solar radiation and slightly cold biased at night due to radiation to space (Dirksen et al. 2014). Most weather stations apply radiation corrections to raobs, based on algorithms provided by radiosonde manufacturers. Raobs analyzed in this study include the manufacturer adjustments. However, because the radiation correction schemes were derived using limited data and most adjustments consider few factors [generally pressure, solar elevation angle (SEA), and balloon ascent speed], biases still remain even after correction (e.g., Gaffen 1994; Sherwood et al. 2005; Haimberger et al. 2008; Sun et al. 2013).

This study uses multiple datasets to characterize the differences between RS41 and RS92 upper-air temperature observations made at operational radiosonde stations, with standard operational Vaisala radiation corrections applied to both radiosonde types. Section 2 introduces methodology, section 3 describes the datasets used for the assessment, section 4 discusses results, and section 5 provides a summary and conclusions.

2. Methodology

As described in section 3, three different reference data sources are used for the assessment: NWP model data, GPS radio occultation (GPSRO) dry temperature, and dual-radiosonde observations. In this study, the model data are used as a transfer medium to determine the relative differences of RS92 versus RS41 using two relative comparisons:

- 1) Compare raobs with forecast model background data, where the computed observation minus forecast background (OB–BG) differences for both RS92 and RS41 are used to identify their relative difference. Basically, an operational NWP model run starts with the preceding short-term forecast valid at the starting time of the new forecast cycle as a “background” or a “first guess.” For example, with a 6-h forecast cycle system, the forecast cycle starting at 1200 UTC uses the 6-h forecast from the 0600 UTC cycle as the 1200 UTC BG. Therefore, the 1200 UTC BG is an operational model forecast and is not affected by the new OB data, including the raobs near 1200 UTC to be analyzed in this work.
- 2) Compare raobs with the corresponding forecast model analysis data. The analysis at any given time, for example at 1200 UTC, is the result of the 1200 UTC BG, adjusted by assimilating all types of new observations (received generally from 1200 UTC \pm 3 h) including radiosondes, surface data, satellite measurements, and GPSRO data. This study analyzes

the observation minus analysis (OB–AN) differences for both RS92 and RS41 in the same way as the OB–BG differences to understand further the relative difference of the two sondes. While the analysis assimilates the same raobs for which the OB–AN differences are computed, the OB and AN values are still partially independent because the BG starting field is a physically consistent short-term forecast, and assimilation that produces the AN field does not automatically accept the reported OB values, but formally balances errors of all data sources. The OB–AN differences are therefore not zero, but are expected to be smaller than the OB–BG differences, and the AN fields could be used as an alternative reference data source to estimate radiosonde data error (as discussed in section 4).

It should be mentioned that major NWP centers, including the National Centers for Environmental Prediction (NCEP) and European Centre for Medium-Range Weather Forecasts (ECMWF), have historically employed OB–BG differences to make radiosonde data radiation corrections prior to their assimilation into the AN field for operational forecasts. To develop corrections, the center uses observations and forecasts, interpolated to observation locations, to compute the OB–BG temperature differences (or increments) at various heights and solar angles (including night) for all individual sonde types. For radiosondes, each distinct type is adjusted to be equivalent to some “reference” radiosonde type. The AN fields from the NCEP model are based on assimilated raobs adjusted with the current radiation correction (RADCOR) scheme, which at NCEP has not been updated since around 2000 (Sun et al. 2013), and NCEP currently applies no correction to newer sondes including RS92 and RS41. ECMWF uses the radiosonde bias correction scheme of Agustí-Panareda et al. (2009) with the corrections being updated monthly and being applied to RS92 and RS41. So, in this study, OB temperatures of RS92 and RS41 observations compared with the NCEP model BG or AN are the same OB values as the ones compared with ECMWF model data, but corrections have been made to the radiosondes assimilated in the ECMWF AN cycle even though the corrections to those two sondes are small (Ingleby 2017).

Direct comparisons were made of RS92 minus RS41 differences (RS92–RS41) in dual launches of both types suspended from the same balloon at five sites listed in Table 1. These provide the most rigorous radiosonde comparisons because both radiosondes sample the same air column, but the comparisons are still relative differences because neither RS41 nor RS92 provides

TABLE 1. Stations with dual Vaisala RS41 and RS92 soundings. The right two columns list the starting and ending dates and number of dual launches analyzed in this study. In the right column, the four numbers are counts of dual soundings reaching at least 51.5 hPa in the SEA classes of NIGHT (SEA < -7.5°), DUSK/DAWN (-7.5° – -7.5°), DAY ($>7.5^\circ$), and ALL (shown in parentheses). Due to the small numbers of dual flights, DAY combines the SEA categories LOW (SEA 7.5° – 15°) and HIGH ($>15^\circ$) used in other analyses in this paper. Table 7 shows statistics for NIGHT, DAY, and ALL soundings at each station, except that it omits analyses where the sounding quantities are indicated by an asterisk below in the right column. Table 7 omits the NIGHT category at stations with <6 dual soundings, and the entire DUSK/DAWN category because only 2 stations performed any dual soundings in that sun angle range.

Station (WMO ID)	Lat, lon (launch elevation)	Starting and ending date	No. of dual soundings
Ny Alesund, Norway (01004)	78.92°N, 11.92°E (15.5 m)	29 Mar 2015–2 May 2017	9, 14*, 49 (72)
Payerne, Switzerland (06610)	46.81°N, 6.944°E (490.5 m)	12 Aug 2014–24 Oct 2017	28, 0*, 35 (63)
Lindenberg, Germany (10393)	52.21°N, 14.12°E (115 m)	9 Dec 2014–18 Oct 2017	53, 0*, 80 (133)
Lamont, Oklahoma (74646)	36.61°N, 97.49°W (314 m)	3–8 Jun 2014	5*, 1*, 10 (16)
Lauder, New Zealand (93817)	45.04°S, 169.68°E (370 m)	4 Nov 2015–17 Nov 2016	0*, 0*, 54 (54)

absolute accuracy. The five sites sample different climate regimes, as summarized in section 3, to assess consistency with the “global” assessment.

GPSRO provides a technique to derive atmospheric profiles of temperature that are used to independently evaluate radiosonde observations. A GPSRO is a measurement by an accurately time-synchronized satellite of the time delays (or bending angles) of a GPS satellite signal near the horizon as the signal passes through different atmospheric layers. The bending angle varies due to atmospheric density (temperature, water vapor, and air pressure) and ionospheric electron density. Assuming spherical symmetry (relative to Earth) of the air in the GPS signal path, the ionospheric corrected bending angle profile can be used to compute a refractive index (or refractivity) profile downward from the upper stratosphere. In the stratosphere and upper troposphere, the “dry temperature” (Tdry) profile can be derived from the refractivity profile by ignoring water vapor. Computing Tdry in a moist profile causes a cold bias (Ware et al. 1996), but this study extends only down to 150 hPa, where the maximum plausible cold bias in a tropical thunderstorm is ~ 0.03 K, so water vapor can be ignored. Tdry has a high vertical resolution (0.1–1 km), although its horizontal resolution is low due to the long atmospheric signal path.

Since the GPS bending angle decreases exponentially with decreasing air density, the Tdry uncertainty reportedly increases quickly above ~ 25 km or pressures lower than ~ 25 hPa (Kursinski et al. 1997; Hajj et al. 2004; Steiner et al. 2011). The Tdry profiles above ~ 35 km are highly sensitive to the a priori required to initialize the hydrostatic integration at the ionosphere due to the radio occultation (RO) measurement “null space” (Tradosky et al. 2017), and their accuracy in those upper layers is thus noticeably affected by the initialization (Steiner et al. 2013). Caution is therefore needed to analyze the radiosonde bias using GPSRO Tdry as the reference at altitudes higher than ~ 25 km.

For heights between 10 and 25–30 km (between about 100 and 25–10 hPa) where this study seeks to understand radiosonde accuracy, the average Tdry error is within 0.1–0.2 K (Hajj et al. 2004; Steiner et al. 2011), while the operational RS92 or RS41 uncertainty specified by Vaisala varies from 0.3 to 0.5 K (<https://www.vaisala.com/en/products/instruments-sensors-and-other-measurement-devices/soundings-products>; viewed October 2018). The value of GPSRO to act as an upper-air reference has been demonstrated in comparisons with radiosonde data to estimate their biases (Kuo et al. 2005; He et al. 2009; Sun et al. 2013; Ladstädter et al. 2015; Ho et al. 2017). Those studies analyze raob minus GPSRO Tdry temperature differences (OB–GPSRO), to directly compare with RS92 and RS41.

The Tdry data quality overall is consistent among different missions (satellite systems) and data processing centers, but differences among GPSRO products are still noticed particularly in regional analysis or analyses using limited data samples (Ho et al. 2012; Steiner et al. 2013). Therefore, two different GPSRO products are used in this study, as described in section 3.

The datasets used in this work, including raobs, model outputs, and GPSRO, have different vertical resolutions. For example, the dual-raob flights report data values at 1-s intervals, or usually ~ 7000 vertical levels, but standard operational raobs (reported in TEMP code format—as summarized in section 3) used in this work report ~ 50 – 150 vertical levels.

To suppress small-scale atmospheric structures in high-density vertical profiles and minimize the impact of different vertical resolutions on the raob accuracy assessment, an approach typical of satellite hyperspectral sounding retrieval validation is adopted here (Susskind et al. 2003; Tobin et al. 2006; Sun et al. 2017). Namely, radiosonde data, the NWP forecast and analysis, and GPSRO Tdry are all linearly interpolated (in the logarithmic pressure coordinate) to 100 common vertical levels from the surface to the top of the atmosphere, and

the differences relative to the temperature reported by RS92 and RS41 are computed at the 100 levels. The 100-level values are then averaged into ~ 1 -km coarse layers, and mean biases and standard deviation (SD) differences are computed in each of those layers and are used in the tables, plots, and discussions. Table 2 lists coarse layers centered from ~ 150 to ~ 10 hPa where raobs are generally subject to the strongest radiation impacts. Comparing with collocated GPSRO Tdry (see section 4c) reveals that the mean biases for RS92 and RS41 are rather small in the layers around 100 hPa, such as 92.8 and 113.9 hPa where biases for both sondes are <0.04 K. Assessment statistics are displayed in the plots up to ~ 10 hPa but only those for the lower-stratospheric layers (51.5–26.1 hPa) are included in the tables.

3. Datasets

Two and a half years (January 2015 through June 2017) of global operational RS92 and RS41 temperature observations are analyzed for their OB–BG and OB–AN differences. The NWP data fields used for the OB–BG differences are the NOAA Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) forecast background. The NWP data fields used for the OB–AN differences are the CFSR and the operational ECMWF analysis. Two GPSRO Tdry products are used as independent profiles for assessment. The first is the University Corporation for Atmospheric Research (UCAR) Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC; <http://www.cosmic.ucar.edu/cdaac>) GPSRO data using GPS receivers on the COSMIC satellites. The second is the EUMETSAT Radio Occultation Meteorology (ROM) Satellite Application Facility (SAF) Global Navigation Satellite System (GNSS) Receiver for Atmospheric Sounding data (GRAS; www.romsaf.org) using GPS receivers on the *MetOp-A* and *MetOp-B* satellites. Just as the temperature differences from the raob to target sources are abbreviated OB–BG for “observation minus background” and OB–AN for “observation minus analysis,” the raob minus Tdry differences are abbreviated OB–COSMIC and OB–GRAS.

All of those datasets are collocated in the NOAA Products Validation System (NPROVS; Reale et al. 2012; Sun et al. 2017), supported by the NOAA Joint Polar Satellite System (JPSS) and operated at NOAA NESDIS office of Satellite Applications and Research (STAR) starting 2008. NPROVS provides routine data access, collocation, and intercomparison of multiple satellite temperature and water vapor sounding product suites and NWP model profiles matched with global operational radiosonde and dropsonde observations.

TABLE 2. (left) Central pressure (hPa) and (right) bounding pressures of coarse vertical layers where raob temperature bias statistics are computed in the upper troposphere and lower stratosphere.

10.2	9.5–11.0
13.5	11.0–16.4
18.6	16.4–20.9
26.1	20.9–32.2
37.4	32.2–43.1
51.5	43.1–60.9
71.5	60.9–83.2
92.8	83.2–103.0
113.9	103.0–125.6
133.8	125.6–142.3
151.2	142.3–160.4

The collocation approach is to select the “single closest” sounding from each product suite anchored to the raob launch location. Description of the individual datasets is given below.

a. Raobs

Raobs are those assimilated operationally by NOAA’s NCEP. Those raobs are transmitted through the WMO Global Telecommunications System (GTS), and have used the alphanumeric TEMP code (WMO Code Forms FM-35 to FM-38). The TEMP code forms have used the same general format since 1968 (when teletype circuits had very limited communication capacity), so reports have much lower vertical resolution and less precision in measured values than raobs in the Binary Universal Form for Representation of Meteorological Data (BUFR) format (WMO Code Form FM-94). The rounding in the TEMP code can vary with the processing software used. According to Ingleby (2017), at temperatures below 0°C RS41 appears 0.1 K warmer than RS92 using Digicora III software due to rounding. The Digicora III RS92 reports are 60% of all RS92 profiles used in the study, and the RS92 versus RS41 differences (shown in Tables 4–6 and Fig. 3) could thus be slightly underestimated.

In the last few years, an increasing number of raobs and other observations have been transmitted in the BUFR format. According to Ingleby and Edwards (2015) and Ingleby (2017), on average those two types of raob reports are very comparable to each other (with a TEMP reporting bias $\sim 0.1^{\circ}\text{C}$ or smaller due to systematic rounding or data conversions in station or forecast center decoding). TEMP variable r_{a,r_a} or BUFR variable 002011 reports a radiosonde type code, and the code can be a typographical error (e.g., if the instrument group is omitted, and the following group is the launch time in the 2300 h, “23” can be interpreted as a Vaisala RS41 code). This analysis removed five erroneous cases of this type.

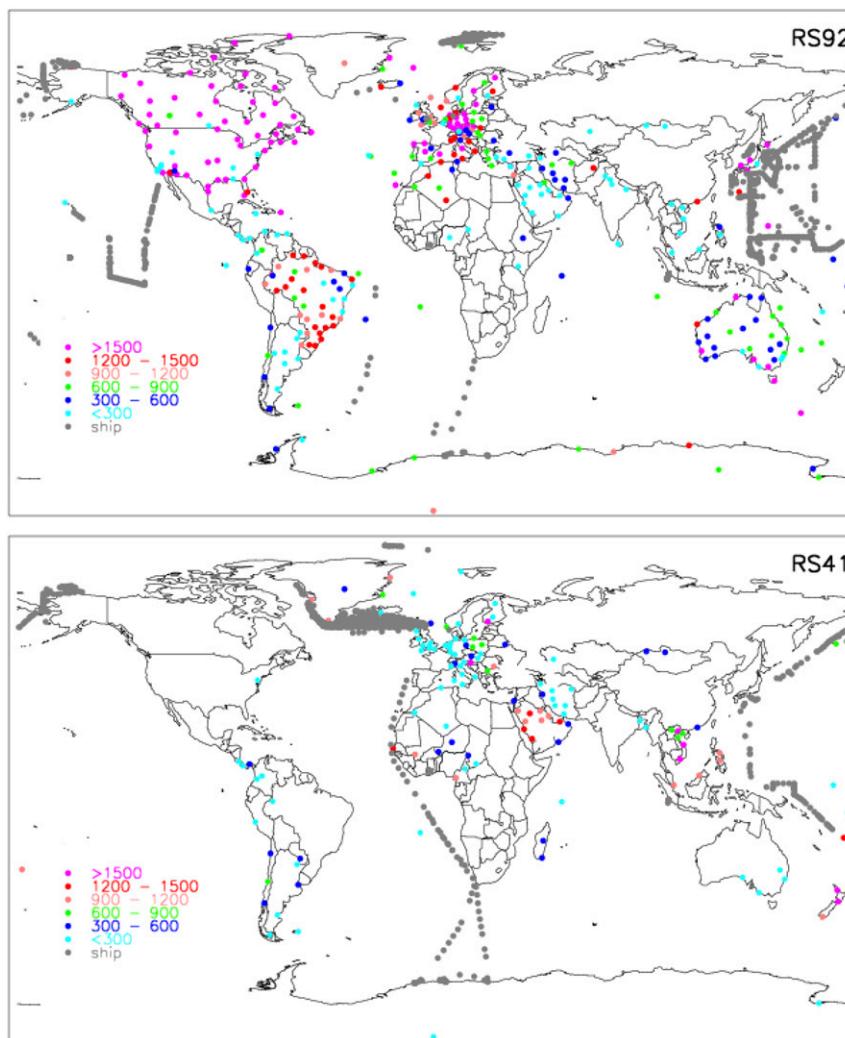


FIG. 1. Global distribution of raob sites with colors indicating different numbers of launches for January 2015 through June 2017: (top) RS92 and (bottom) RS41. Circles with gray color are individual raobs launched from ships.

As part of the data assimilation process, for archiving and later analysis NCEP stores ingested raobs and all other observations into a file called PREPBUFR, with the 6-h CFSR guess (background) and analysis fields encoded into the PREPBUFR file at radiosonde balloon locations. As stated in section 1, raobs in PREPBUFR contain only corrections applied by the stations.

It was found that 7.7% of raob profiles available for this study are associated with meaningless CFSR background temperatures at all mandatory levels. The raobs associated with those “bad” values, distributed randomly in time and across the global stations, are excluded in this study.

We rejected raobs with a “failure” quality flag in the NCEP assimilation system (Collins 2001a,b) or having temperature difference ≥ 15 K from the NCEP

background, as well as those with a vertical extent < 5 km or a vertical gap ≥ 4 km (Reale et al. 2012), totaling $\sim 1\%$ of the observations. This study uses $\sim 311\,500$ RS92 raobs from 344 fixed stations and 1101 ship launches, and $\sim 65\,800$ RS41 raobs from 135 fixed stations and 1080 ship launches.

Figure 1 shows the spatial distribution of the radiosonde stations. As of June 2017, RS41 radiosondes are most used in western Europe, the Middle East, Southeast Asia, and New Zealand, but few or no RS41 are launched from North America, Australia, and Brazil where RS92 dominates. As with all other sonde types, operational RS92 and RS41 raobs are launched primarily around 0000 and 1200 UTC. Dual launches of RS41 and RS92 made at five sites are listed in Table 1, along with launch locations, time

periods (some launches were in 2014), and number of analyzed dual launches.

b. CFSR background and analysis and ECMWF analysis

The CFSR data are those included in the PREPBUFR files generated in the NCEP data assimilation (see this section's raob data description). This study also uses ECMWF (ECMWF 2018) operational analyses available at 0000, 0600, 1200, and 1800 UTC, with 91 vertical pressure levels thinned from the 137 model sigma levels and horizontal resolution of $0.25^\circ \times 0.25^\circ$ (Eresmaa and McNally 2014). CFSR atmospheric profiles (with 64 vertical levels and ~ 38 -km horizontal resolution) are 4D interpolated to radiosonde profiles (J. Wollen 2018, personal communication; Saha et al. 2010), meaning that radiosonde balloon drift in space and time starting from the launch location is taken into account. The balloon drift information is also ingested in the ECMWF system.

c. RO Tdry data

Due to degrading COSMIC constellation satellites in this period, the average daily number of RO profiles is about 600 from COSMIC, but about 1200 from GRAS on *MetOp-A* and *MetOp-B* combined. Both sources are approximately evenly distributed across the globe.

Since CFSR profiles were spatially and temporally interpolated to radiosonde data, their collocation error is already minimized. Since ECMWF data fields are at four synoptic times, almost all raob-ECMWF collocations accepted for analysis are within 1 h (including launches around 0600 and 1800 UTC), and raobs not launched close to synoptic times are not analyzed.

GPSRO and radiosonde collocations in this study are within 3 h and 250 km, but a sensitivity test is also conducted to understand the impact of sampling or collocation errors on the raob accuracy assessment. However, since the assessment is focused on the upper troposphere and above and the balloon takes about 30 min to rise to the upper troposphere (Seidel et al. 2011), 30 min is added to the radiosonde launch time when computing the collocation mismatch with respect to GPSRO (and ECMWF). The GPSRO profile location at around 100 hPa is used to compute the GPSRO distance mismatch with the radiosonde profile.

As usual, no reference dataset is the "truth," but consistent comparison results using datasets of different sources give more confidence in the analysis, as discussed below.

4. Results

As in Sun et al. (2013), the raob temperature bias is estimated using the mean raob minus collocated target

(i.e., NWP or GPSRO) temperature, along with the SD of the differences, sorted into four SEA classes ranging from NIGHT (SEA $< -7.5^\circ$), DUSK/DAWN (SEA = -7.5° – 7.5°), LOW (SEA = 7.5° – 22.5°), and HIGH (SEA $> 22.5^\circ$). The number of soundings in DUSK/DAWN and LOW SEA classes is typically much smaller than in respective NIGHT or HIGH classes because on the average, DUSK/DAWN and LOW sun angles account for small portions of the 24-h day, and standard observation times at 0000 and 1200 UTC near 0° longitude are all near solar midnight and noon. All raobs occurring near solar noon are in the HIGH SEA class, except in winter at fairly high latitudes.

Results for ALL SEA are also computed using simple averages of all included raobs, to facilitate understanding of the overall raob bias tendency. "Global" difference statistics in this study are computed by simply averaging sounding data from all sites.

a. Analysis of OB-BG differences in CFSR

Figure 2 shows average raob minus CFSR OB-BG differences in the four SEAs for RS92 (top panels) and RS41 (bottom panels). The black curve in each left panel is the ALL SEA OB-BG difference. While it is desirable for all soundings to reach 10 hPa, the number of soundings decreases noticeably starting around 50 hPa due to balloon burst, by about 18% at 26 hPa, 38% at 19 hPa, and 59% at 10 hPa. Since balloons tend to burst at lower altitudes in winter (or at night when the balloon is not warmed by sunlight), higher-altitude unweighted statistics can be less seasonally, diurnally, and spatially representative.

In the top panels of Fig. 2, starting around 70 hPa, RS92 is systematically warmer than the CFSR BG, with warming increasing with altitude and with SEA class. The increasing warm difference as the SEA increases is a pattern of radiation-induced radiosonde bias. The BG minus COSMIC differences shown in the gray bars of Fig. 11 of Sun et al. (2013) indicate a small daytime warm bias increasing with SEA to ~ 0.16 K in the HIGH SEA class averaged for 10–70 hPa. Apparently, because NCEP applies no correction to newer radiosonde types (see the fourth paragraph of section 2), the earlier assimilation of the biased radiosonde data into the NCEP forecast model causes a small diurnal radiation bias in the BG field. Therefore, radiosonde warm biases inferred from OB-BG for both RS92 and RS41 in Fig. 2, and somewhat smaller warm biases inferred using CFSR OB-AN in section 4b could be slightly underestimated.

The standard deviations of the differences are similar among the different SEAs. They reflect the combination of random errors in both radiosonde and the model BG and the spatial and temporal mismatch between the two profiles. They are < 1.0 K throughout the troposphere

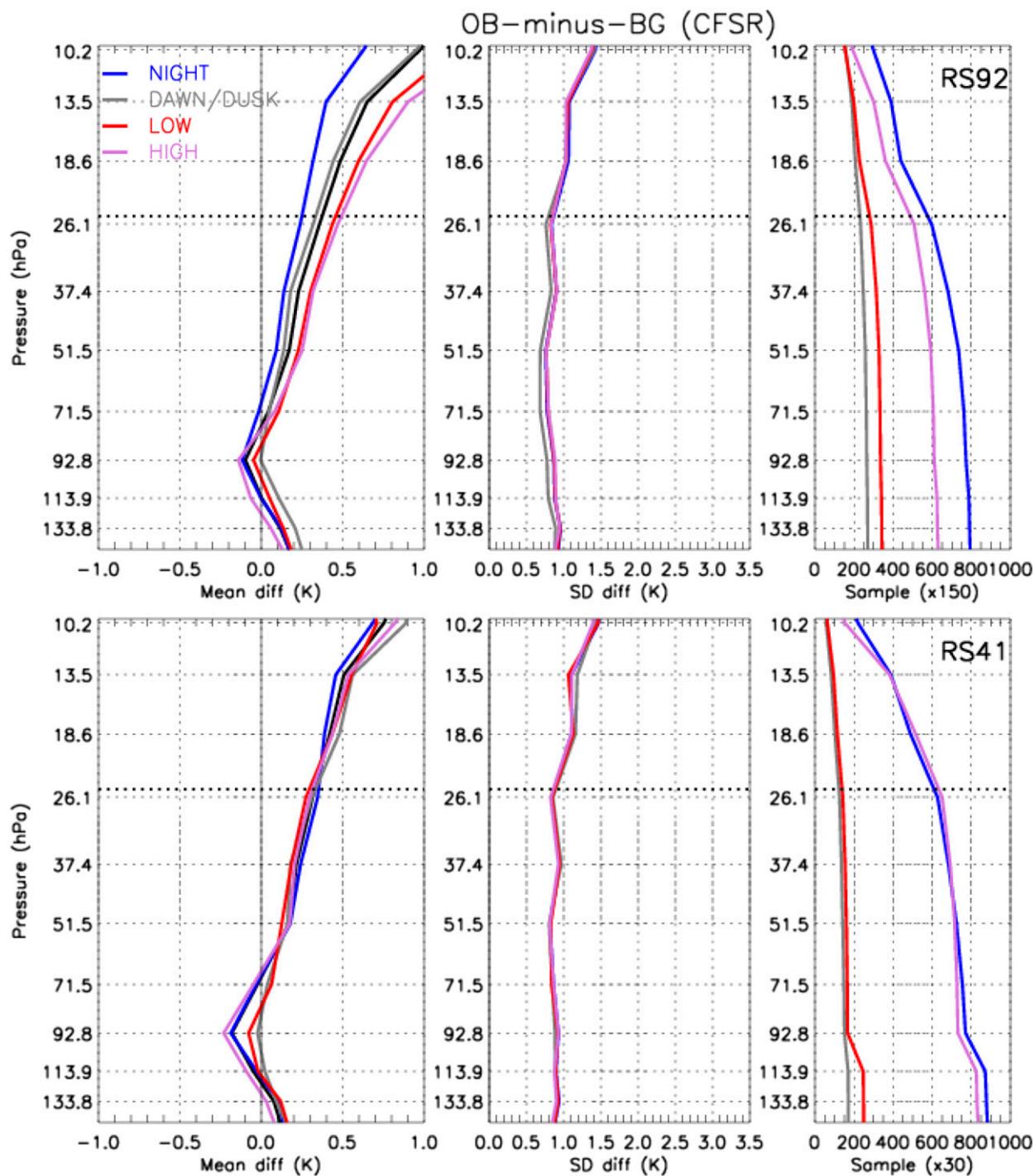


FIG. 2. Global OB-BG (CFSR) differences for (top) RS92 and (bottom) RS41. (left) Average OB-BG temperature differences in the four SEA classes (legend in the top-left panel). The four solar elevations are defined as $<-7.5^\circ$, $-7.5^\circ-7.5^\circ$, $7.5^\circ-22.5^\circ$, and $>22.5^\circ$ (see text for more information, and the black line is for ALL SEA classes). (middle) Standard deviations of OB-BG differences. (right) Numbers of observations in each SEA class (the 0–1000 range should be multiplied by the factor in parentheses after “Sample”).

TABLE 3. Mean absolute difference (MAD, K), measuring the statistical dispersion of the differences between RS92 or RS41 and the comparison data (NWP or GPSRO) among 4 SEA classes (NIGHT, DAWN/DUSK, LOW, and HIGH) for the three stated layers in the lower stratosphere. See text on how MAD values are computed.

MAD (K)	OB–BG CFSR	OB–AN CFSR	OB–AN ECMWF	OB–COSMIC Tdry	OB–GRAS Tdry
26.1 hPa	RS92: 0.098	RS92: 0.082	RS92: 0.089	RS92: 0.248	RS92: 0.169
	RS41: 0.025	RS41: 0.015	RS41: 0.026	RS41: 0.070	RS41: 0.204
37.4 hPa	RS92: 0.081	RS92: 0.070	RS92: 0.065	RS92: 0.165	RS92: 0.132
	RS41: 0.021	RS41: 0.004	RS41: 0.022	RS41: 0.110	RS41: 0.150
51.5 hPa	RS92: 0.069	RS92: 0.053	RS92: 0.054	RS92: 0.122	RS92: 0.093
	RS41: 0.010	RS41: 0.005	RS41: 0.014	RS41: 0.058	RS41: 0.097

(not shown) and stratosphere except at the surface and at altitudes above 19 hPa where the SD values are larger. In the bottom panels of Fig. 2, RS41 is also systematically warmer than the CFSR BG, but the RS41 OB–BG shows less dispersion with SEA, suggesting RS41 is less sensitive than RS92 to solar radiation errors.

An objective method to quantify the solar radiation sensitivity shown in the left panels of Fig. 2 is to compute the spread of the OB–BG differences among the four SEA classes (i.e., NIGHT, DAWN/DUSK, LOW, and HIGH). The spread, or mean absolute difference (MAD), in any pressure layer is defined as the arithmetic average of the absolute values of the mean OB–BG differences for the four SEA classes (blue, gray, red, and purple lines in the left panels of Fig. 2) from the corresponding OB–BG difference for the ALL SEA class (black line). Note that the “ALL” OB–BG difference is a measure of the “inherent” radiosonde type bias, except that it includes model biases, while MAD is a measure of the sensitivity of radiosonde temperature bias to radiation. MAD should be a good estimate of the raob solar heating error even though OB–BG differences are not absolute comparisons. MAD is computed in the same way using OB–AN or OB–GPSRO differences.

Table 3 shows the MAD values for RS92 and RS41 in three lower-stratospheric layers (26.1, 37.4, and 51.5 hPa). The “OB–BG CFSR” column shows MAD values in these three layers in the left panels of Fig. 2. The increasing MAD value with decreasing pressure for both RS92 and RS41 confirms that the radiation impact on radiosonde bias increases with height. The smaller dispersions among SEA classes in the RS41 OB–BG compared to RS92 (also seen in Fig. 2) confirm that RS41 is less sensitive to solar radiation than RS92.

Using the BG as the transfer medium to compute RS92 minus RS41 differences assumes that, because the model assimilates many observation types and establishes physical consistency, the BG field for the new forecast cycle (e.g., the preceding 6-h forecast) almost completely averages out individual instrument biases. So, if one station uses RS92 and another uses RS41, the

BG has negligibly different biases from previous RS92 and RS41 soundings at these locations. Figure 3 and the left column of Table 4 indicate that relative to RS41, RS92 tends to show a radiation-induced temperature bias pattern in the lower stratosphere, for example, with a difference in the 37.4-hPa layer of -0.10 K for NIGHT and 0.11 K for HIGH SEA. Although small, within the manufacturer-specified uncertainty (Jensen et al. 2016), the differences are statistically significant, and could pose a potential challenge for climate trend detection associated with temporal changes in instrumentation (Sherwood et al. 2005).

Note that the RS92 minus RS41 differences for ALL and particularly for HIGH shown in Fig. 3 increase with height above 26.1 hPa. The value for HIGH SEA reaches 0.37 K at 13.5 hPa and 0.52 K at 10.2 hPa. Similar tendencies are observed in the RS92 minus RS41

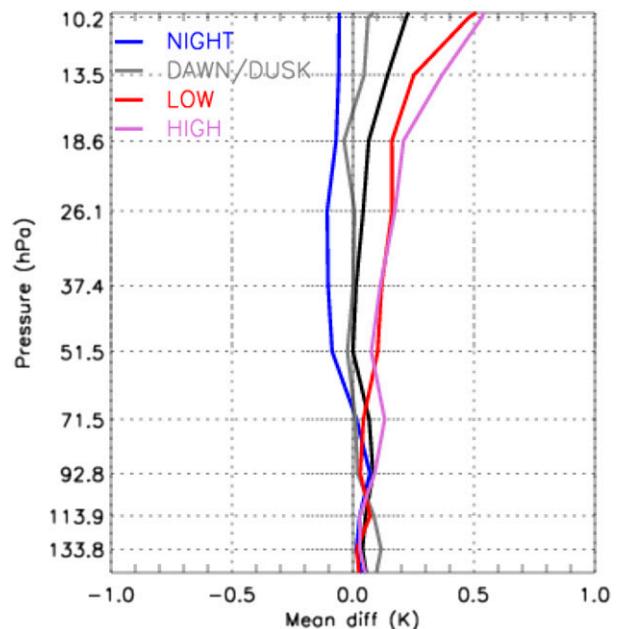


FIG. 3. Mean differences between RS92 and RS41 by using the CFSR BG as the transfer medium. Each value is the RS92 OB–BG minus the corresponding RS41 OB–BG from Fig. 2.

TABLE 4. Values of OB–BG and OB–AN for the three lower-stratospheric layers for NIGHT, HIGH, and ALL (ALL also includes cases with DAWN/DUSK and LOW sun angles). Values inside parentheses are standard deviations and numbers of samples. The third line in each table cell is the RS92 minus RS41 average difference, and italicized values are statistically not different from zero at the 0.05 level.

	OB–BG CFSR	OB–AN CFSR	OB–AN ECMWF
NIGHT			
26.1 hPa	RS92 0.239 (0.84, 89 589)	RS92 0.216 (0.68, 89 589)	RS92 0.199 (0.69, 67 782)
	RS41 0.346 (0.84, 18 813)	RS41 0.281 (0.77, 18 813)	RS41 0.203 (0.69, 14 087)
	–0.107	–0.065	–0.004
37.4 hPa	RS92 0.139 (0.90, 102 020)	RS92 0.118 (0.75, 102 020)	RS92 0.203 (0.79, 89 933)
	RS41 0.242 (0.95, 20 477)	RS41 0.183 (0.77, 20 477)	RS41 0.210 (0.81, 16 986)
	–0.100	–0.065	–0.007
51.5 hPa	RS92 0.092 (0.76, 110 275)	RS92 0.070 (0.59, 110 275)	RS92 0.192 (0.62, 98 756)
	RS41 0.178 (0.82, 21 723)	RS41 0.125 (0.61, 21 723)	RS41 0.235 (0.65, 18 211)
	–0.086	–0.055	–0.043
HIGH			
26.1 hPa	RS92 0.469 (0.83, 76 067)	RS92 0.409 (0.67, 76 067)	RS92 0.413 (0.62, 57 463)
	RS41 0.296 (0.82, 19 525)	RS41 0.253 (0.65, 19 525)	RS41 0.207 (0.64, 14 787)
	0.173	0.149	0.206
37.4 hPa	RS92 0.318 (0.91, 84 180)	RS92 0.273 (0.74, 84 180)	RS92 0.348 (0.74, 75 706)
	RS41 0.205 (0.93, 20 752)	RS41 0.177 (0.75, 20 752)	RS41 0.220 (0.78, 17 016)
	0.113	0.096	0.128
51.5 hPa	RS92 0.251 (0.76, 88 857)	RS92 0.188 (0.58, 88 857)	RS92 0.304 (0.58, 81 009)
	RS41 0.175 (0.81, 21 443)	RS41 0.129 (0.59, 21 443)	RS41 0.246 (0.62, 17 646)
	0.076	0.059	0.058
ALL			
26.1 hPa	RS92 0.357 (0.83, 243 791)	RS92 0.317 (0.67, 243 791)	RS92 0.318 (0.66, 191 651)
	RS41 0.315 (0.84, 46 379)	RS41 0.264 (0.66, 46 379)	RS41 0.221 (0.67, 35 437)
	0.042	0.053	0.097
37.4 hPa	RS92 0.229 (0.90, 270 275)	RS92 0.199 (0.75, 270 275)	RS92 0.290 (0.77, 243 016)
	RS41 0.216 (0.95, 49 926)	RS41 0.178 (0.76, 49 926)	RS41 0.228 (0.80, 41 690)
	0.013	0.021	0.062
51.5 hPa	RS92 0.171 (0.77, 286 892)	RS92 0.135 (0.58, 286 892)	RS92 0.263 (0.61, 260 773)
	RS41 0.171 (0.81, 52 251)	RS41 0.126 (0.60, 52 251)	RS41 0.249 (0.64, 43 977)
	0.000	0.009	0.014

differences using either CFSR AN or ECMWF AN as the transfer medium (figures not shown). Further investigation is needed to find out if the increasing RS92 minus RS41 daytime warm difference (with height) in the stratosphere is related to the reduction in radiosonde sample sizes, as mentioned in the first paragraph of this subsection (more balloons reach high altitudes in warm conditions, or when sunlight warms the balloon), or reflects the true nature of radiosonde biases at high altitudes.

Radiosonde biases at night tend to be much smaller than in daytime and are often considered to be negligible, or at least satisfactorily corrected by the radiation corrections applied by the stations. So, if the NIGHT OB–BG (OB–AN) difference is not close to zero, it is possible that OB is correct at night, and BG (or AN) is the actual biased quantity (when OB is considered correct, the bias becomes either BG–OB or AN–OB). The considerable magnitudes of the NIGHT biases in the left panels of Fig. 2, which are quite variable with altitude

but very similar for both RS92 and RS41, suggest that, instead of RS92 and RS41 both having a fairly large warm bias at the highest levels at night, CFSR BG has a cold bias at those levels. From Table 4, if the NIGHT raobs are the reference, then CFSR BG has a cold bias in the 26.1-hPa layer of -0.239 ± 0.84 K (\pm indicates one standard deviation in this paper) for RS92 and -0.346 ± 0.84 K for RS41, with the difference in average biases probably due to differing station locations. The CFSR biases are likely to change with altitude due to differing mixes of ingested data, and the CFSR warm bias centered at 92.8 hPa could be an artifact of interpolation of 64 coarsely spaced CFSR model levels to the radiosonde levels (Ballish and Kumar 2008; Saha et al. 2010).

The MAD values in Table 3 (and seen by the spread between the NIGHT and HIGH category lines in Fig. 2) are still a measure of the RS92 and RS41 radiative biases, especially because they increase at higher levels in Table 3, from 0.069 K at 51.5 hPa to 0.098 K at 26.1 hPa for RS92, but only from 0.010 to 0.025 K in the same layers for RS41.

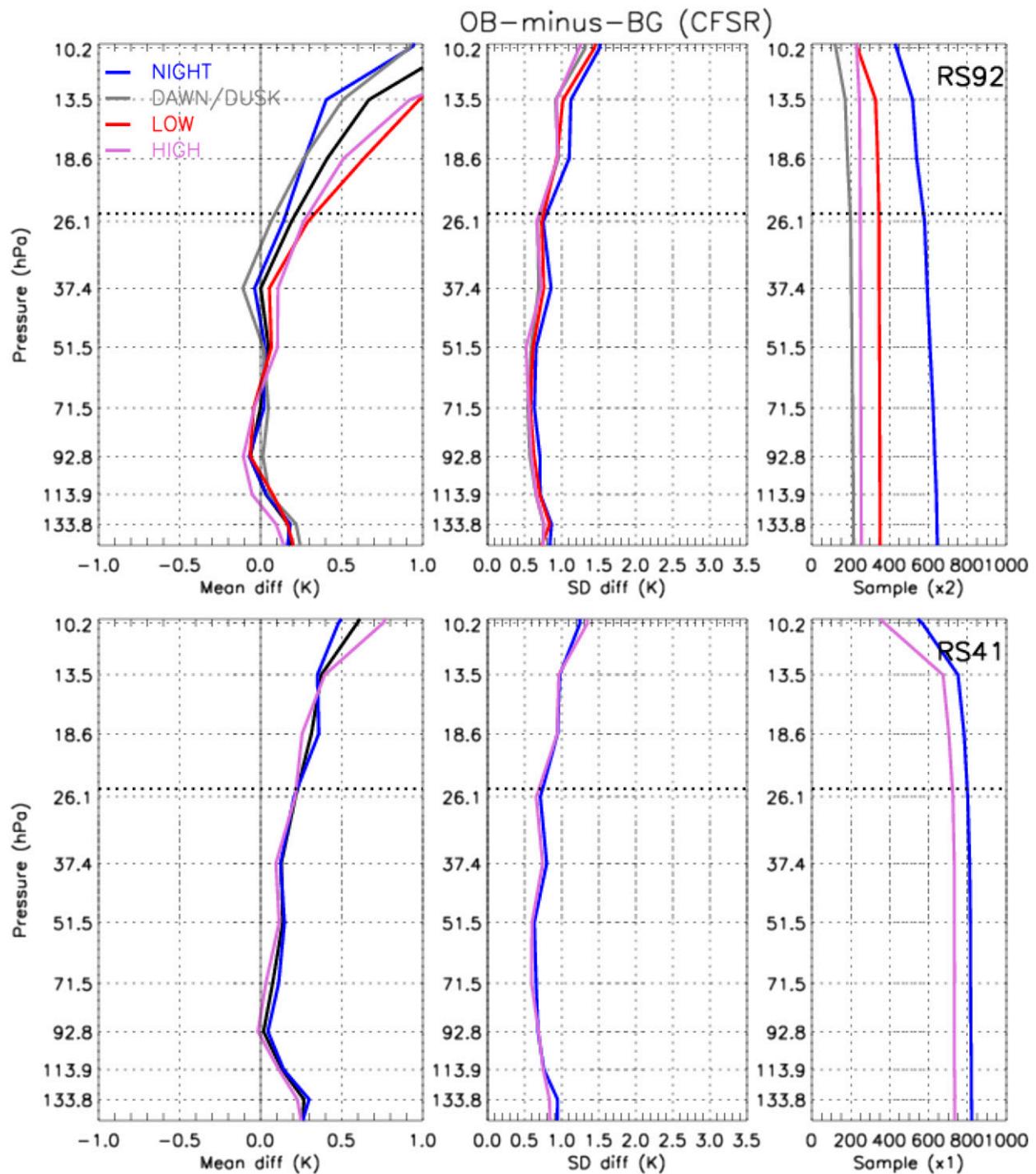


FIG. 4. As in Fig. 2, but for (top) the OB-BG (CFRS) differences for RS92 at Lindenberg and (bottom) for RS41 at Zagreb/Maksimir.

While Fig. 2 comparison statistics by SEA class are based on global data, Fig. 4 shows the comparison statistics for Lindenberg, Germany (WMO ID 10393, 52.21°N, 14.12°E, 112.0 m), launching RS92 four times per day (most 0600 and 1800 UTC soundings have

DUSK/DAWN or LOW SEA and some cold season 1200 UTC soundings have LOW SEA), and Zagreb/Maksimir, Croatia (WMO ID 14240, 45.82°N, 16.03°E, 127.6 m), launching RS41 twice a day (all 1200 UTC soundings are in the HIGH SEA class). They are about

750 km apart. The results in Fig. 4 are similar to Fig. 2 even though the difference statistics for each radiosonde are based on only one station. The MAD values in the UTLS, measuring the dispersions between SEA classes in the RS41 OB–BG at Zagreb/Maksimir are much smaller compared to RS92 at Lindenberg, for example, 0.065 K versus 0.033 K at 37.4 hPa. So the smaller radiation errors for RS41 compared to RS92 are supported by similar analyses using individual stations.

b. Analysis of the CFSR and ECMWF OB–AN differences

The CFSR OB–AN plots for RS92 and RS41 (figure not shown, but see OB–AN MAD values in Table 3 and OB–AN values in Table 4) are similar to the CFSR OB–BG. As mentioned in section 2, Table 4 shows that the CFSR OB–BG differences and standard deviations are slightly larger than the CFSR OB–AN differences and standard deviations (for RS92 at 51.5 hPa, the standard deviation decreases from 0.77 K with OB–BG to 0.58 K with OB–AN, and the corresponding RS41 change is from 0.81 to 0.60 K). This is expected because the analysis assimilates the current raobs while the background is only a forecast valid at the analysis time. The only slightly smaller OB–AN values (compared to OB–BG) indicate that the analysis process does not significantly correct any biases that are attributed to either inherent model biases or systematic biases in ingested data other than raobs. Similarly, the CFSR OB–AN MAD values in Table 3 corresponding to the same stratospheric layers as for OB–BG, indicate again that RS41 is less sensitive to SEA than RS92.

Figure 5 shows, and the right column of Table 4 lists, the ECMWF OB–AN differences, and the third data column in Table 3 lists the MAD values for RS92 and RS41. The analysis indicates that RS41 again shows improvement over RS92, but by <0.2 K for HIGH and <0.1 K for ALL in the layers of 51.5 to 26.1 hPa. The RS41 OB–AN curves for NIGHT and HIGH are almost identical except near the surface (not shown in Fig. 5) and in the highest layers. The separation of RS41 LOW and DAWN/DUSK from NIGHT and HIGH in the stratosphere in the lower-left panel of Fig. 5 could be related to small sample sizes for the former cases.

In Table 4, the NIGHT OB–AN ECMWF values are similar to the NIGHT OB–AN CFSR values in the 26.1-hPa layer, but the ECMWF values are roughly the same down to the 51.5-hPa layer while the CFSR values diminish down to that layer. Figure 5 shows statistics similar to Table 4, but for all four SEA categories (and the ALL average) in all layers from 10.2 to 151.2 hPa (the x axis is 151.2 hPa in each panel), and the NIGHT

bias for RS92 diminishes at higher pressures and for RS41 diminishes toward the highest and lowest pressures. So, in this data period, the ECMWF analysis probably has a cold bias that is less variable with UTLS altitude than the cold bias of CFSR (roughly 0.1 to 0.7 K for CFSR, but 0.1 to 0.2 K for ECMWF), and ECMWF does not show the spike to a warm bias that occurs in CFSR centered on 92.8 hPa primarily because the former has a much higher vertical resolution than CFSR.

c. Analysis of the OB–COSMIC and OB–GRAS differences

As stated in section 2, Tdry obtained from GPSRO is an independent estimate of temperature with high vertical resolution, but agreement with a single raob is expected to be degraded due to the long GPSRO horizontal signal paths. While GPSRO measurements are available worldwide, they are quite randomly distributed. The local time of an RO occultation is determined by the orbit positions of the receiver and the GPS transmitter. COSMIC occultations are more random in local time, since its satellites are on different orbit planes. The EUMETSAT *MetOp-A* and *MetOp-B* satellites are in a fixed orbit plane with local equator crossing times at 0930 and 2130. Thus the local times of GRAS occultations are scattered between about 0800 to 1100 and 2000 to 2300 equatorward of 55° latitude, and can be any local time near the poles. GPSRO collocations are allowed within 3 h and 250 km of each raob, so only about 7% (11%) of raobs have a suitable COSMIC (GRAS) GPSRO. Table 5 shows the numbers of COSMIC and GRAS collocations for NIGHT, HIGH, and ALL sun angles, along with other statistics, in the same format as Table 4.

Figure 6 shows OB–COSMIC differences and standard deviations. OB–COSMIC differences are used to infer radiosonde biases, but differences for both RS92 and RS41 become large at altitudes with pressure greater than 150 hPa (not shown in Fig. 6) from GPSRO Tdry values that are too cold due to increasing moisture at lower altitudes.

The RS92 OB–COSMIC curves in Fig. 6 show a slight radiosonde cold bias at night and a daytime warm bias increasing with SEA and altitude, similar to Table 5. This pattern was also obtained using 2008–11 data (Sun et al. 2013). However, the average (ALL SEA) RS92 biases in the UTLS layers for 113.9 to 92.8 hPa are very small, within 0.04 K.

The small RS92 OB–COSMIC difference at night and the increasing daytime warm bias continue upward even to 10 hPa, indicating that the COSMIC Tdry quality may be good even for altitudes higher than 25 km.

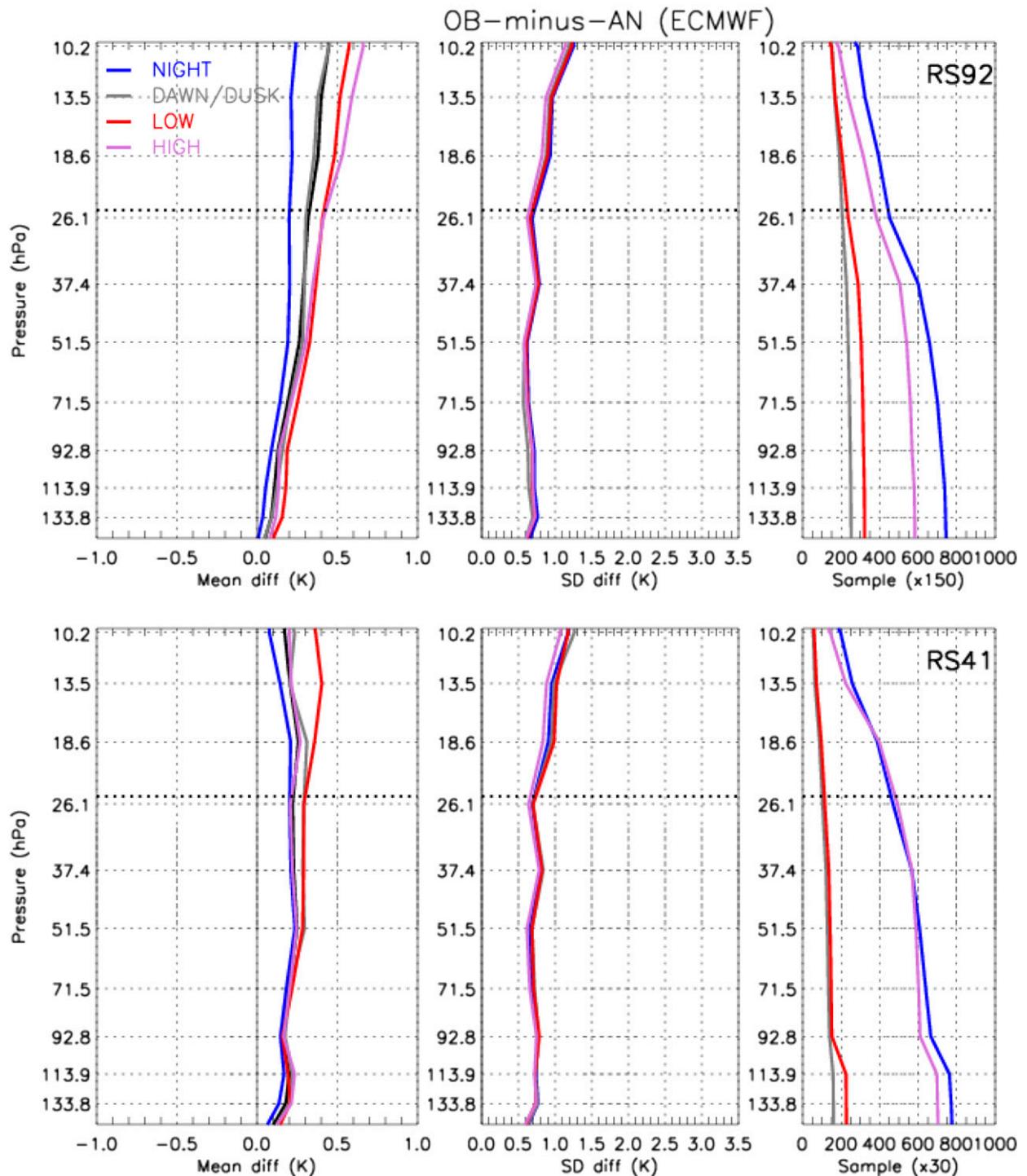


FIG. 5. As in Fig. 2, but for global average OB-AN differences for ECMWF.

Nevertheless, we need to be cautious about the bias statistics above 25 km because of the reasons mentioned in section 2, in addition to small sample sizes.

Due to the small number of RS41 collocation samples, the curves in the bottom panels of Fig. 6 are noisy, but

qualitative conclusions can still be made. Compared to the nearly zero RS92 night OB-COSMIC biases (upper-left panel of Fig. 6), RS41 shows small night warm OB-COSMIC biases, possibly related to the limited RS41 geographical regions. Daytime RS41 warm

TABLE 5. As in Table 4, but the raob bias statistics are computed using Tdry as the target. Collocations are within 3 h and 250 km.

	OB-COSMIC Tdry		OB-GRAS Tdry	
NIGHT				
26.1 hPa	RS92	-0.012 (1.57, 5859)	RS92	-0.012 (2.10, 6082)
	RS41	0.231 (1.38, 934)	RS41	-0.106 (2.10, 1093)
		-0.243		0.094
37.4 hPa	RS92	-0.009 (1.39, 7257)	RS92	-0.033 (1.82, 8322)
	RS41	0.089 (1.37, 1066)	RS41	-0.075 (1.78, 1396)
		-0.098		0.042
51.5 hPa	RS92	0.011 (1.21, 7870)	RS92	-0.043 (1.52, 9267)
	RS41	0.084 (1.23, 1126)	RS41	-0.062 (1.46, 1524)
		-0.073		0.019
HIGH				
26.1 hPa	RS92	0.531 (1.36, 4499)	RS92	0.355 (1.76, 8024)
	RS41	0.364 (1.39, 926)	RS41	0.342 (1.68, 1572)
		0.167		0.013
37.4 hPa	RS92	0.370 (1.22, 5680)	RS92	0.244 (1.53, 10 988)
	RS41	0.321 (1.29, 1085)	RS41	0.248 (1.52, 1906)
		0.049		-0.004
51.5 hPa	RS92	0.282 (1.09, 6068)	RS92	0.172 (1.30, 11 897)
	RS41	0.204 (1.14, 1117)	RS41	0.173 (1.22, 1993)
		0.078		-0.001
ALL				
26.1 hPa	RS92	0.236 (1.51, 16 126)	RS92	0.218 (1.94, 22 201)
	RS41	0.279 (1.41, 2242)	RS41	0.131 (1.92, 3572)
		-0.043		0.087
37.4 hPa	RS92	0.162 (1.34, 19 552)	RS92	0.129 (1.67, 28 920)
	RS41	0.214 (1.34, 2594)	RS41	0.100 (1.69, 4380)
		-0.052		0.029
51.5 hPa	RS92	0.141 (1.16, 20 908)	RS92	0.087 (1.41, 31 339)
	RS41	0.154 (1.18, 2704)	RS41	0.085 (1.42, 4655)
		-0.013		0.002

biases increase with altitude and sun angle, but are smaller than for RS92 (as indicated in Table 5). The smaller MAD values for RS41 than for RS92, shown in the OB-COSMIC column of Table 3, again indicate that RS41 is less sensitive to radiation impact.

Figure 7 shows OB-GRAS differences corresponding to the OB-COSMIC differences in Fig. 6, and the curves are very similar including high noise in RS41 for DAWN/DUSK and LOW due to small sample sizes. The general similarities of corresponding OB-GRAS and OB-COSMIC bias patterns for both RS92 and RS41 add to the confidence that most of the OB-GRAS biases are caused by radiosonde radiation errors. Colder RS41 OB-GRAS than OB-COSMIC biases with NIGHT and DAWN/DUSK sun angles (lower-left panels of both figures), and the similar RS92 HIGH and LOW OB-GRAS biases (Fig. 7, upper left), could be rooted in differing diurnal and geographical COSMIC and GRAS data sampling.

Note that, in Figs. 6 and 7 and Table 5, GRAS has larger standard deviations than COSMIC for both RS92

and RS41, even though the GRAS sample sizes are ~50% larger than for COSMIC. Both GPSRO products used in the analysis are nearly real time. GRAS instrument and level 2 processing was upgraded around October 2016 to include additional quality screening (S. Syndergaard 2018, personal communication), leading to smaller GRAS Tdry standard deviations in about the last third of the analysis period. For example, relative to global radiosonde data (including RS92 and RS41) collocated within 6 h and 250 km, the GRAS SD errors at 37.4 hPa average 1.91 K in February 2015 (0.29 K greater than the SD errors of COSMIC data) but decrease to 1.77 K in February 2017 (comparable to the COSMIC SD error values).

Temporal and spatial collocation mismatches impact the statistics of the observations being compared or validated (Tobin et al. 2006; Sun et al. 2010, 2017). To test the collocation sampling impact, we repeat the computations from Table 5 (collocations within 3 h and 250 km) using collocations within 1 h and 150 km in Table 6. This also reduces the sample sizes by about 80%–90%. However, the OB-COSMIC and OB-GRAS bias patterns for RS92 and RS41 are still similar, although with more noise (figure not shown), and the stratospheric standard deviations in Table 6 are about 15% smaller than in Table 5. At night, some of the small biases change sign between Tables 6 and 7, with the RS41 OB-GRAS difference changing from negative for the large collocation window to positive with the small collocation window. This exercise indicates that the error statistics obtained using GPSRO for RS92 are robust, although there is less confidence in the sign of near-zero differences with small samples, which is the case with RS41 in this study.

Overall, the OB-GPSRO biases suggest that the biases are mainly in the radiosonde data. RS92 has daytime stratospheric warm biases in both OB-COSMIC and OB-GRAS, and RS41 has smaller warm biases than RS92, with more noise in SEA classes with small samples. The nighttime stratospheric OB-COSMIC differences in Table 5 are all within ±0.1 K for both instrument types, but RS41 differences are larger than for RS92. At 26.1 hPa and lower pressures the diurnal range of raob minus GPSRO is larger than the diurnal range of raob minus NWP, the reasons for this are not currently clear.

d. Analysis of RS41 and RS92 dual-launch data

Table 7 lists the RS92 minus RS41 difference statistics from five dual-launch sites (Table 1) for SEA classes of NIGHT, DAY, and ALL. Due to the small number of launches in all sites, statistics for DAY combine LOW and HIGH SEA. “ALL” is the number of all dual

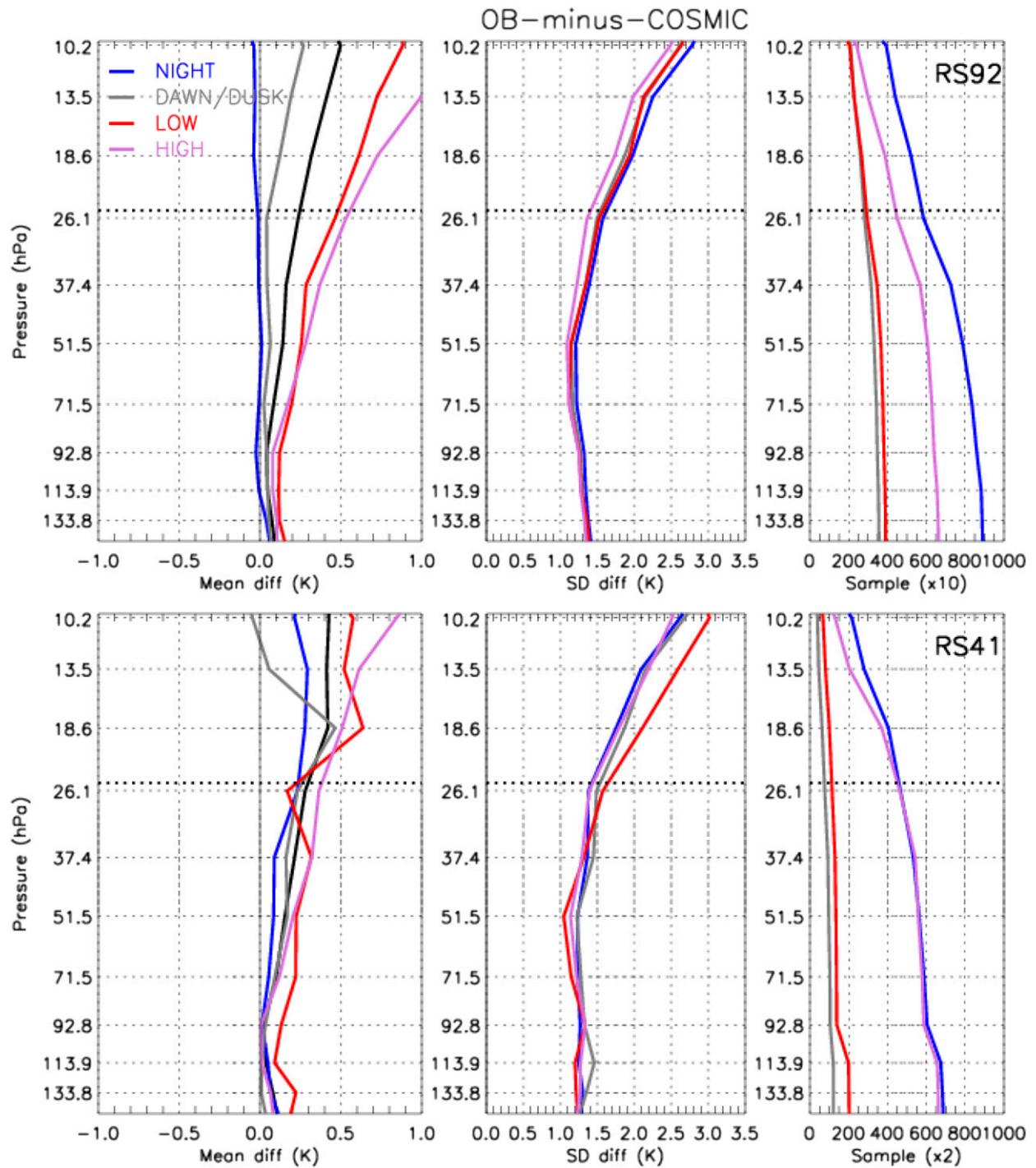


FIG. 6. OB-COSMIC differences (raob temperature minus GPSRO T_{dry}) for (top) RS92 and (bottom) RS41. Collocation window is within 3 h and 250 km. The numbers of samples in the right panels are as in Fig. 2. The horizontal dashed line at 25 hPa indicates that caution is needed to interpret the statistics above that altitude as T_{dry} accuracy could be degraded and also because there are fewer collocation samples.

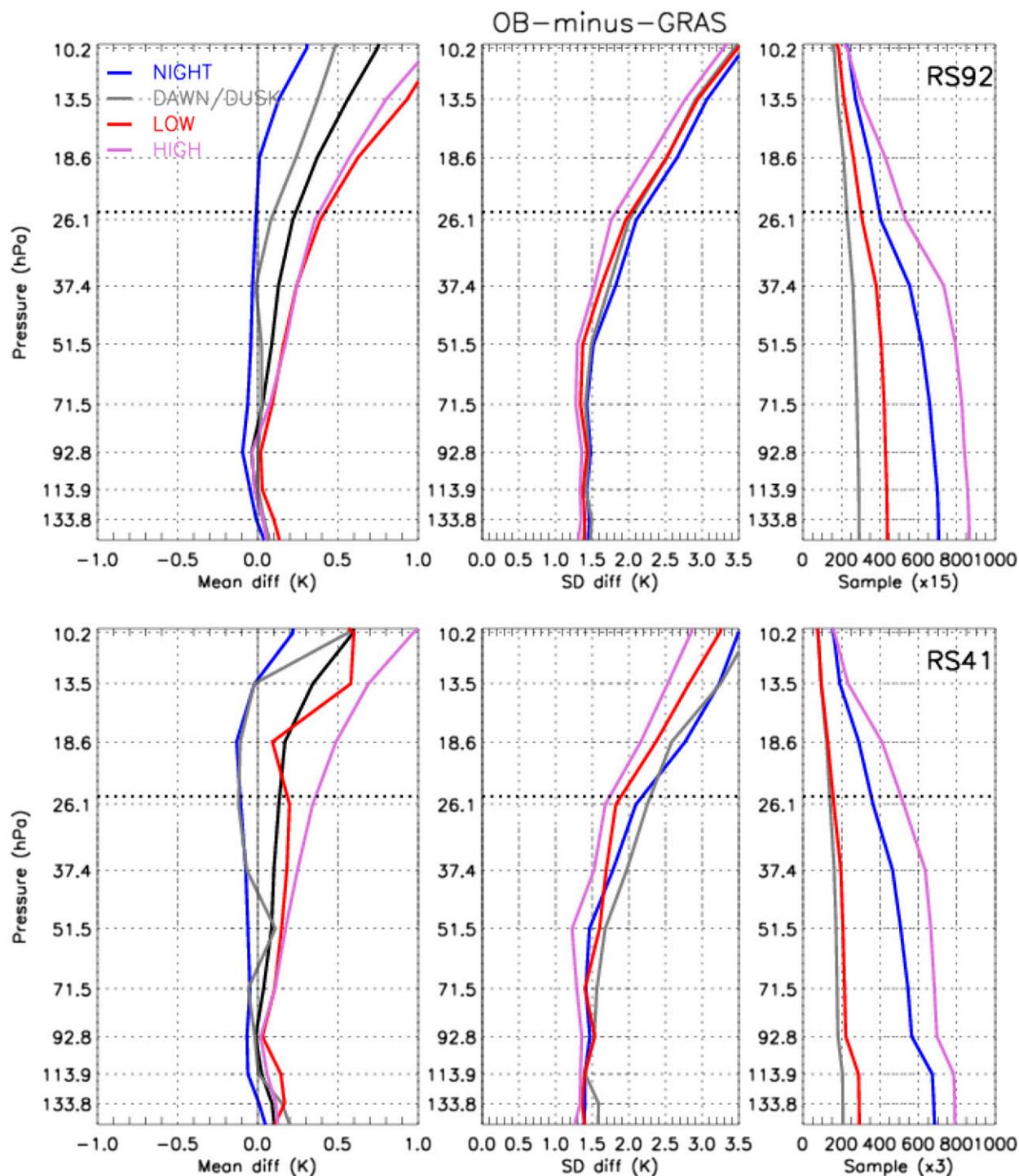


FIG. 7. As in Fig. 6, but for OB-GRAS.

soundings at all sun angles, but as described in the legend for Table 1, Table 7 omits analyses for the DUSK/DAWN category (no dual soundings except from two stations), and for the NIGHT category from Lamont and Lauder (<6 dual soundings). While these are relative

differences, within the limitations of a small number of cases, these comparisons are rigorous measurements to indicate whether RS92 or RS41 has a larger bias.

Special GRUAN data processing (GDP; GRUAN software, version 2) was performed on RS92 data from

TABLE 6. As in Table 5, but the global collocations within 1 h and 150 km are used to compute the statistics.

	OB-COSMIC Tdry		OB-GRAS Tdry	
NIGHT				
26.1 hPa	RS92	-0.006 (1.39, 769)	RS92	-0.142 (2.00, 609)
	RS41	0.082 (1.35, 116) <i>-0.088</i>	RS41	0.154 (2.11, 144) <i>-0.296</i>
37.4 hPa	RS92	0.016 (1.26, 950)	RS92	-0.068 (1.68, 873)
	RS41	0.067 (1.53, 132) <i>-0.051</i>	RS41	0.206 (1.62, 201) <i>-0.274</i>
51.5 hPa	RS92	-0.024 (1.05, 1031)	RS92	-0.089 (1.37, 991)
	RS41	0.094 (1.12, 144) <i>-0.118</i>	RS41	0.034 (1.37, 222) <i>-0.123</i>
HIGH				
26.1 hPa	RS92	0.568 (1.37, 600)	RS92	0.360 (1.78, 1250)
	RS41	0.436 (1.29, 102) <i>0.132</i>	RS41	0.221 (1.64, 209) <i>0.139</i>
37.4 hPa	RS92	0.403 (1.19, 752)	RS92	0.220 (1.46, 1811)
	RS41	0.228 (0.96, 114) <i>0.174</i>	RS41	0.306 (1.50, 261) <i>-0.086</i>
51.5 hPa	RS92	0.287 (0.99, 798)	RS92	0.157 (1.30, 1983)
	RS41	0.271 (0.93, 117) <i>0.016</i>	RS41	0.091 (1.10, 276) <i>0.066</i>
ALL				
26.1 hPa	RS92	0.207 (1.41, 2164)	RS92	0.237 (1.91, 2811)
	RS41	0.222 (1.39, 261) <i>-0.015</i>	RS41	0.088 (1.91, 475) <i>0.149</i>
37.4 hPa	RS92	0.163 (1.21, 2620)	RS92	0.143 (1.57, 3892)
	RS41	0.166 (1.36, 294) <i>-0.003</i>	RS41	0.221 (1.67, 610) <i>-0.078</i>
51.5 hPa	RS92	0.108 (1.00, 2803)	RS92	0.085 (1.30, 4262)
	RS41	0.179 (1.08, 312) <i>-0.071</i>	RS41	0.070 (1.32, 658) <i>0.015</i>

all sites except Lauder, aiming to remove systematic biases in the data with uncertainty estimates provided (Dirksen et al. 2014). At all sites, RS41 data were processed with standard Vaisala procedures (as at other synoptic stations), including standard Vaisala RS41 corrections.

RS92 agrees well with RS41 in the troposphere except for being <0.1 K colder than RS41 in the middle and upper troposphere at Lindenberg, Lamont, and Lauder (figure not shown; also see Jensen et al. 2016). RS92 shows an increasing warm bias relative to RS41 at higher altitudes. As indicated in Table 7, for layers between 51.5 and 26.1 hPa, RS92 tends to be ~ 0.05 K warmer than RS41 for NIGHT and is mostly <0.25 K warmer than RS41 for DAY. The RS92 minus RS41 DAY positive differences actually increase with height in the stratosphere through ~ 13.5 hPa at sites where the data samples are sufficient. Specifically, at 18.6 and 13.5 hPa (respectively), RS92 is warmer than RS41 by 0.36 and 0.51 K at Lindenberg, by 0.37 and 0.41 K at Ny Alesund, and by 0.21 and 0.26 K at Payerne.

A warm bias in the lower stratosphere has been noticed in RS92 with GDP, in comparisons to GPSRO Tdry (Ladstädter et al. 2015). That study pointed out that GDP tends to undercorrect RS92 data, which causes slight warming (less than 0.2 K) compared to synoptic RS92 data containing the operational Vaisala corrections. This suggests that the actual difference between the RS92 and RS41 daytime radiative heating errors may be smaller than the differences found in the direct comparisons, and solar radiative errors in operational data appear to be improved in RS41 relative to RS92 particularly for altitudes above 26.1 hPa. The smaller RS41 minus RS92 difference (by ~ 0.12 K) at 18.6 and 13.5 hPa at Lauder supports this improvement.

The ALL averages in those three lower-stratospheric layers show RS92 is generally <0.2 K warmer than RS41, but since all sites have more day than night profiles, the RS92 minus RS41 differences for ALL are weighted toward DAY. In Table 7, the standard deviations of the RS92 minus RS41 differences at all sites are greater for DAY than for NIGHT, indicating that these two sondes have larger differences during daytime than nighttime [simply, that RS92 has more radiative heating than RS41, consistent with Jensen et al. (2016)]. Table 7 also shows moderate differences among the sites in RS92 minus RS41 NIGHT or DAY difference values. That could be due to the small number of flights and local factors affecting the radiation balance surrounding the sonde such as clouds and surface albedo (Bower and Fitzgibbon 2004).

5. Summary and discussion

The accuracy of upper-air temperature observations of Vaisala RS92 versus RS41 was assessed using two and a half years (January 2015 to June 2017) of global raobs with $\sim 311\,500$ RS92 profiles and $\sim 65\,800$ RS41 profiles. This was achieved using three different data sources as references for comparison, including NWP outputs, GPSRO Tdry, and dual radiosondes. The relative differences of the two sondes are estimated by analyzing observation minus NWP model background (OB-BG) and observation minus analysis (OB-AN) differences, as well as a small number of RS92 and RS41 dual launches (RS92-RS41 differences). GPSRO Tdry profiles are used for direct assessments of radiosonde observations.

These two sondes were found to be in good agreement with average difference <0.1 – 0.2 K in the lower stratosphere from 51.5 to 26.1 hPa based on globally (but not evenly) distributed stations, and dual-launch data from five different sites. However, the daytime RS92 warm bias relative to RS41 tends to become greater at heights

TABLE 7. RS92 minus RS41 difference statistics for the three lower-stratospheric layers for NIGHT, DAY, and ALL at the RS92 and RS41 dual-launch sites listed in column 1. Values inside parentheses are standard deviations and numbers of samples. No statistics are shown with data sample <6 . The numbers of samples in the right column are equal to the numbers of dual soundings in Table 1, but counts in the 37.4- and 26.1-hPa columns are lower if any soundings did not reach that pressure layer.

	26.1 hPa	37.4 hPa	51.5 hPa
NIGHT			
Ny Alesund	0.029 (0.13, 9)	-0.014 (0.10, 9)	-0.001 (0.07, 9)
Payerne	0.049 (0.08, 26)	0.056 (0.08, 26)	0.050 (0.09, 28)
Lindenberg	0.078 (0.20, 52)	0.011 (0.10, 53)	0.002 (0.09, 53)
Lamont			
Lauder			
DAY			
Ny Alesund	0.287 (0.40, 49)	0.224 (0.33, 49)	0.179 (0.24, 49)
Payerne	0.187 (0.24, 34)	0.159 (0.21, 34)	0.116 (0.19, 35)
Lindenberg	0.227 (0.33, 78)	0.147 (0.27, 80)	0.109 (0.20, 80)
Lamont	0.221 (0.26, 10)	0.194 (0.24, 10)	0.168 (0.22, 10)
Lauder	0.036 (0.28, 52)	-0.023 (0.22, 52)	-0.054 (0.18, 54)
ALL			
Ny Alesund	0.170 (0.29, 71)	0.141 (0.21, 72)	0.123 (0.18, 72)
Payerne	0.127 (0.19, 60)	0.115 (0.17, 60)	0.087 (0.15, 63)
Lindenberg	0.167 (0.28, 130)	0.093 (0.22, 133)	0.066 (0.17, 133)
Lamont	0.221 (0.26, 16)	0.194 (0.24, 16)	0.168 (0.22, 16)
Lauder	0.036 (0.28, 52)	-0.023 (0.22, 52)	-0.054 (0.18, 54)

above 26.1 hPa. Further investigation is needed to understand if that tendency is related to radiosonde sample reduction (early balloon burst) or whether RS92 actually has a larger solar radiation error than RS41.

Accuracy of radiosonde pressure measurements can affect the bias statistics, simply by attributing temperatures to incorrect pressures. Radiosonde atmospheric pressure is measured either by direct pressure sensor measurement, or by hydrostatic computation of pressure from reported temperatures and GPS heights. All Vaisala RS92 models have a pressure sensor, but RS41 is available with or without a pressure sensor (instrument codes 123–125 with a pressure sensor, or 141–142 without). According to Vaisala (2013), GPS-based pressures from RS41 are smaller than those from RS92 on average by <0.4 hPa near the surface and <0.02 hPa above 30 hPa, and GPS-based pressures from RS41 are greater than sensor-based pressures from RS92 by <0.3 hPa near the surface and 0.2 hPa above 30 hPa. Those differences are small and within the instrument specification (Vaisala 2013).

By increasing the reported RS92 pressure by 0.15 hPa at 750 hPa to 0.02 hPa at 10 hPa in proportion to the logarithm of pressure, the raob minus Tdry biases for GRAS change slightly. For example, the RS92 bias at 37.4 hPa becomes warmer by 0.04 K for HIGH and colder by 0.02 K for NIGHT. However, the net effect of the uncertainty of pressure measurements using different methods for RS92 versus RS41 may not be

significant in the stratosphere where temperature gradients are small.

Previous studies (Sun et al. 2013; Ho et al. 2017) show a radiation-induced bias pattern in RS92 data, with a daytime warm bias increasing with altitude and with SEA. All three comparisons in this study, comparing raobs with NWP model data or GPSRO Tdry profiles, and dual RS92 and RS41 soundings, give similar results for both RS92 and RS41 with RS41 showing less difference in the warm bias with increasing SEA, indicating that RS41 is less sensitive to radiation than RS92.

While model grids are easily interpolated to the raob location and time, the collocation error introduced through spatial and temporal mismatch between GPSRO and a raob could influence the radiosonde accuracy statistics obtained. In this study, when the collocation window was reduced from 3 h and 250 km to 1 h and 150 km, the sample size is reduced by 80%–90%. The pattern and magnitude of the RS92 bias shows no pronounced change but the SD error is systematically reduced. For RS41, which has a much smaller number of launches and collocations with GPSRO Tdry profiles, the reduction in sample size using the tight collocation window does not change the overall night-versus-day bias pattern but causes more noisy differences, including some noticeably differing biases or even opposite signs in some SEA classes or pressure layers, due to suitable collocations being located in only a few climate regions (Sun et al. 2013), so this study should be updated as

RS41 observations become more numerous. However, Tradowsky et al. (2017) use NWP model data to reduce the collocation error of individual GPSRO-raob collocations, and their method is recommended to understand the raob bias when the number of collocations available is small.

The RS92 versus RS41 accuracy comparisons obtained using two different near-real-time GPSRO products, UCAR COSMIC and ROM SAF GRAS, are basically consistent, reinforcing the robustness of the sonde bias analysis. However, it is preferable to use reprocessed RO products for analysis in the future, because GPSRO changes cause discontinuities, and reprocessing allows inclusion of delayed data and consistent methods to be applied to a long data record.

NWP model forecast background or analysis biases, originating from earlier assimilation of biased observations or systematic biases in forecast models (Eyre 2016), are generally unknown. This study generally supports the hypothesis that RS92 and RS41 radiosonde biases during nighttime are negligible. Therefore, similar positive OB-BG and OB-AN differences at night (~ 0.2 – 0.5 K) in the lower stratosphere with both radiosonde types are inferred to not indicate persistent night radiosonde warm biases, but instead indicate model cold biases. The stratospheric cold biases in the models are most probably caused by radiative cooling effect of moisture leaking into the region from the upper troposphere (Shepherd et al. 2018). The spread of differences with sun angle classes indicates superimposed radiosonde daytime radiative biases, but these model biases are likely to be similar at all sun angles and are larger in CFSR than in ECMWF.

Acknowledgments. We thank Dennis Kaiser and Jack Woollen at NCEP for discussing the operational radiosonde data assimilation processing at NCEP. Comments and suggestions from Bruce Ingleby at ECMWF and Xavier Calbet at AEMET greatly improve the quality of the paper. Comments on the DMI GRAS data characteristics by Stig Syndergaard and Johannes Nielsen at DMI and Axel Von Engel at EUMETSAT are appreciated. We thank UCAR/CDAAC for COSMIC data, ROM SAF at DMI for GRAS data, and the GRUAN Lead Centre at Lindenberg, Germany, for access to the dual-launch radiosonde data and the agencies and individuals who made the dual launches. The views, opinions, and findings contained in this report are those of the author(s) and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. government position, policy, or decision.

REFERENCES

- Agusti-Panareda A., and Coauthors, 2009: Radiosonde humidity bias, correction over the West African region for the special AMMA reanalysis at ECMWF. *Quart. J. Roy. Meteor. Soc.*, **135**, 595–617, <https://doi.org/10.1002/qj.396>.
- Ballish, B. A., and V. K. Kumar, 2008: Systematic differences in aircraft and radiosonde temperatures: Implications for NWP and climate studies. *Bull. Amer. Meteor. Soc.*, **89**, 1689–1708, <https://doi.org/10.1175/2008BAMS2332.1>.
- Bodeker, G. E., and Coauthors, 2016: Reference upper-air observations for climate: From concept to reality. *Bull. Amer. Meteor. Soc.*, **97**, 123–135, <https://doi.org/10.1175/BAMS-D-14-00072.1>.
- Bower, C. A., and J. J. Fitzgibbon, 2004: National Weather Service in-situ radiation temperature correction for Radiosonde Replacement System GPS radiosondes. *Eighth Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface*, Seattle, WA, Amer. Meteor. Soc., 3.1, https://ams.confex.com/ams/84Annual/techprogram/paper_74046.htm.
- Collins, W. G., 2001a: The operational complex quality control of radiosonde heights and temperatures at the National Centers for Environmental Prediction. Part I: Description of the method. *J. Appl. Meteor.*, **40**, 137–151, [https://doi.org/10.1175/1520-0450\(2001\)040<0137:TOCQCO>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<0137:TOCQCO>2.0.CO;2).
- , 2001b: The operational complex quality control of radiosonde heights and temperatures at the National Centers for Environmental Prediction. Part II: Examples of error diagnosis and correction from operational use. *J. Appl. Meteor.*, **40**, 152–168, [https://doi.org/10.1175/1520-0450\(2001\)040<0152:TOCQCO>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<0152:TOCQCO>2.0.CO;2).
- Dirksen, R. J., M. Sommer, F. J. Immler, D. F. Hurst, R. Kivi, and H. Vömel, 2014: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde. *Atmos. Meas. Tech.*, **7**, 4463–4490, <https://doi.org/10.5194/amt-7-4463-2014>.
- ECMWF, 2018: IFS documentation CY45r1. ECMWF Rep., 103 pp.
- Eresmaa, R., and A. P. McNally, 2014: Diverse profile datasets from the ECMWF 137-level short-range forecasts. NWP Satellite Application Facilities Rep. NWPSAF-EC-TR-017, 12 pp., <https://nwpsaf.eu/oldsite/reports/nwpsaf-ec-tr-017.pdf>.
- Eyre, J. R., 2016: Observation bias correction schemes in data assimilation systems: A theoretical study of some of their properties. *Quart. J. Roy. Meteor. Soc.*, **142**, 2284–2291, <https://doi.org/10.1002/qj.2819>.
- Gaffen, D. J., 1994: Temporal inhomogeneities in radiosonde temperature records. *J. Geophys. Res.*, **99**, 3667–3676, <https://doi.org/10.1029/93JD03179>.
- Haimberger, L., C. Tavalato, and S. Sperka, 2008: Toward elimination of the warm bias in historic radiosonde temperature records—Some new results from a comprehensive intercomparison of upper-air data. *J. Climate*, **21**, 4587–4606, <https://doi.org/10.1175/2008JCLI1929.1>.
- Hajj, G. A., and Coauthors, 2004: CHAMP and SAC-C atmospheric occultation results and intercomparisons. *J. Geophys. Res.*, **109**, D06109, <https://doi.org/10.1029/2003JD003909>.
- He, W., S.-P. Ho, H. Chen, X. Zhou, D. Hunt, and Y.-H. Kuo, 2009: Assessment of radiosonde temperature measurements in the upper troposphere and lower stratosphere using COSMIC radio occultation data. *Geophys. Res. Lett.*, **36**, L17807, <https://doi.org/10.1029/2009GL038712>.
- Ho, S.-P., and Coauthors, 2012: Reproducibility of GPS radio occultation data for climate monitoring: Profile-to-profile

- inter-comparison of CHAMP climate records 2002 to 2008 from six data centers. *J. Geophys. Res.*, **117**, D18111, <https://doi.org/10.1029/2012JD017665>.
- , L. Peng, and H. Vömel, 2017: Characterization of the long-term radiosonde temperature biases in the upper troposphere and lower stratosphere using COSMIC and Met-A/GRAS data from 2006 to 2014. *Atmos. Chem. Phys.*, **17**, 4493–4511, <https://doi.org/10.5194/acp-17-4493-2017>.
- Ingleby, B., 2017: An assessment of different radiosonde types 2015/2016. ECMWF Tech. Memo. 807, 69 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2017/17551-assessment-different-radiosonde-types-20152016.pdf>.
- , and D. Edwards, 2015: Changes to radiosonde reports and their processing for numerical weather prediction. *Atmos. Sci. Lett.*, **16**, 44–49, <https://doi.org/10.1002/asl2.518>.
- Jensen, M. P., D. Holdridge, P. Survo, R. Lehtinen, S. Baxter, T. Toto, and K. L. Johnson, 2016: Comparison of Vaisala radiosondes RS41 and RS92 at the ARM Southern Great Plains site. *Atmos. Meas. Tech.*, **9**, 3115–3129, <https://doi.org/10.5194/amt-9-3115-2016>.
- Kuo, Y.-H., W. S. Schreiner, J. Wang, D. L. Rossiter, and Y. Zhang, 2005: Comparison of GPS radio occultation soundings with radiosondes. *Geophys. Res. Lett.*, **32**, L05817, <https://doi.org/10.1029/2004GL021443>.
- Kursinski, E. R., G. A. Hajj, J. T. Schofield, R. P. Linfield, and K. R. Hardy, 1997: Observing Earth's atmosphere with radio occultation measurements using the global positioning system. *J. Geophys. Res.*, **102**, 23 429–23 465, <https://doi.org/10.1029/97JD01569>.
- Ladstädter, F., A. K. Steiner, M. Schwärz, and G. Kirchengast, 2015: Climate intercomparison of GPS radio occultation, RS90/92 radiosondes and GRUAN from 2002 to 2013. *Atmos. Meas. Tech.*, **8**, 1819–1834, <https://doi.org/10.5194/amt-8-1819-2015>.
- Reale, A., B. Sun, F. H. Tilley, and M. Pettey, 2012: The NOAA Products Validation System (NPROVS). *J. Atmos. Oceanic Technol.*, **29**, 629–645, <https://doi.org/10.1175/JTECH-D-11-00072.1>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, <https://doi.org/10.1175/2010BAMS3001.1>.
- Seidel, D. J., B. Sun, M. Pettey, and A. Reale, 2011: Global radiosonde balloon drift statistics. *J. Geophys. Res.*, **116**, D07102, <https://doi.org/10.1029/2010JD014891>.
- Shepherd, T. G., I. Polichtchouk, R. J. Hogan, and A. J. Simmons, 2018: Report on Stratosphere Task Force. ECMWF Tech. Memo. 824, 32 pp., <https://www.ecmwf.int/en/elibrary/18259-report-stratosphere-task-force>.
- Sherwood, S. C., J. R. Lanzante, and C. L. Meyer, 2005: Radiosonde daytime biases and late-20th century warming. *Science*, **309**, 1556–1559, <https://doi.org/10.1126/science.1115640>.
- Steiner, A. K., B. C. Lackner, F. Ladstädter, B. Scherllin-Pirscher, U. Foelsche, and G. Kirchengast, 2011: GPS radio occultation for climate monitoring and change detection. *Radio Sci.*, **46**, RS0D24, <https://doi.org/10.1029/2010RS004614>.
- , and Coauthors, 2013: Quantification of structural uncertainty in climate data records from GPS radio occultation. *Atmos. Chem. Phys.*, **13**, 1469–1484, <https://doi.org/10.5194/acp-13-1469-2013>.
- Sun, B., A. Reale, D. J. Seidel, and D. C. Hunt, 2010: Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics. *J. Geophys. Res.*, **115**, D23104, <https://doi.org/10.1029/2010JD014457>.
- , —, S. R. Schroeder, D. J. Seidel, and B. A. Ballish, 2013: Toward improved corrections for radiation induced biases in radiosonde temperature observations. *J. Geophys. Res. Atmos.*, **118**, 4231–4243, <https://doi.org/10.1002/jgrd.50369>.
- , —, F. H. Tilley, M. E. Pettey, N. R. Nalli, and C. D. Barnett, 2017: Assessment of NUCAPS S-NPP CrIS/ATMS sounding products using reference and conventional radiosonde observations. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, **10**, 2499–2509, <https://doi.org/10.1109/JSTARS.2017.2670504>.
- Susskind, J., C. D. Barnett, and J. M. Blaisdell, 2003: Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds. *IEEE Trans. Geosci. Remote Sens.*, **41**, 390–409, <https://doi.org/10.1109/TGRS.2002.808236>.
- Tobin, D. C., and Coauthors, 2006: Atmospheric Radiation Measurement site atmospheric state best estimates for Atmospheric Infrared Sounder temperature and water vapor validation. *J. Geophys. Res.*, **111**, D09S14, <https://doi.org/10.1029/2005JD006103>.
- Tradowsky, J. S., C. P. Burrows, S. B. Healy, and J. R. Eyre, 2017: A new method to correct radiosonde temperature biases using radiosonde occultation data. *J. Appl. Meteor. Climatol.*, **56**, 1643–1661, <https://doi.org/10.1175/JAMC-D-16-0136.1>.
- Vaisala, 2013: GPS-based measurement of height and pressure with Vaisala radiosonde RS41. Vaisala Rep. B211316EN-A, 16 pp., <https://www.vaisala.com/en/products/instruments-sensors-and-other-measurement-devices/soundings-products/rs41>.
- Ware, R., and Coauthors, 1996: GPS sounding of the atmosphere from low Earth orbit: Preliminary results. *Bull. Amer. Meteor. Soc.*, **77**, 19–40, [https://doi.org/10.1175/1520-0477\(1996\)077<0019:GSOTAF>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0019:GSOTAF>2.0.CO;2).