

Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics

Bomin Sun,^{1,2} Anthony Reale,² Dian J. Seidel,³ and Douglas C. Hunt⁴

Received 6 May 2010; revised 18 August 2010; accepted 15 September 2010; published 3 December 2010.

[1] Collocated global atmospheric temperature, humidity, and refractivity profiles from radiosondes and from Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) radio occultation data for April 2008 to October 2009 are compared for two purposes. The first is to quantify the error characteristics of 12 radiosonde types flown in the global operational network, as a function of height and for both day and nighttime observations, for each of the three variables. The second is to determine the effects of imperfect temporal and spatial collocation on the radiosonde-COSMIC differences, for application to the general problem of satellite calibration and validation using in situ sounding data. Statistical analyses of the comparisons reveal differences among radiosonde types in refractivity, relative humidity, and radiation-corrected temperature data. Most of the radiosonde types show a dry bias, particularly in the upper troposphere, with the bias in daytime drier than in nighttime. Weather-scale variability, introduced by collocation time and distance mismatch, affects the comparison of radiosonde and COSMIC data by increasing the standard deviation errors, which are generally proportional to the size of the time and distance mismatch within the collocation window of 6 h and 250 km considered. Globally, in the troposphere (850–200 hPa), the collocation mismatch impacts on the comparison standard deviation errors for temperature are 0.35 K per 3 h and 0.42 K per 100 km and, for relative humidity, are 3.3% per 3 h and 3.1% per 100 km, indicating an approximate equivalence of 3 h to 100 km in terms of mismatch impact.

Citation: 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, doi:10.1029/2010JD014457.

1. Introduction

[2] Radiosonde observations (raobs) are a key data set in operational weather forecasting and upper-air climate research. Due to their global distribution and high vertical resolution, raobs have also been used as “ground truth” for calibration and validation of satellite temperature and water vapor retrievals [e.g., Divakarla *et al.*, 2006; Tobin *et al.*, 2006; Reale *et al.*, 2008].

[3] With the advent of the new generation of high-spectral resolution sounding instruments (such as the Atmospheric Infrared Sounder (AIRS) onboard NASA’s Aqua satellite and Infrared Atmospheric Sounding Interferometer (IASI) onboard the EUMETSAT MetOp-A satellite), renewed interest in sat-

ellite retrievals and assessing their accuracy using collocated radiosonde measurements has emerged. In pursuit of these goals, concerns have been raised regarding the quality of radiosonde data [McMillin *et al.*, 1988; Miloshevich *et al.*, 2006; McMillin *et al.*, 2007; Adam *et al.*, 2010] and the effects of temporal and spatial mismatch (or noncoincidence) between radiosonde launch and satellite overpass [Tobin *et al.*, 2006; Pougatchev, 2008; Pougatchev *et al.*, 2009; Adam *et al.*, 2010]. This paper addresses both of those concerns through analysis of a global set of “collocated” satellite and radiosonde profiles.

[4] The quality of raobs suffers from measurement biases due to sensor limitations [e.g., Miloshevich *et al.*, 2006; Wang and Zhang, 2008]. Moreover, biases differ between stations according to sonde type and/or national practices [e.g., Soden and Lanzante, 1996; Christy and Norris, 2009], and vary with time due to changes in sensors and reporting practices [Gaffen, 1994; Luers and Eskridge, 1998; Lanzante *et al.*, 2003; Christy and Norris, 2009]. Using 1979–1991 water vapor channel data from the TIROS Operational Vertical Sounder on NOAA polar orbiting satellites as the reference, Soden and Lanzante [1996] revealed differences in global radiosonde instruments in their upper tropospheric

¹LM. Systems Group, Rockville, Maryland, USA.

²Center for Satellite Applications and Research, NESDIS, NOAA, Camp Springs, Maryland, USA.

³Air Resources Laboratory, NOAA, Silver Spring, Maryland, USA.

⁴University Corporation for Atmospheric Research, Boulder, Colorado, USA.

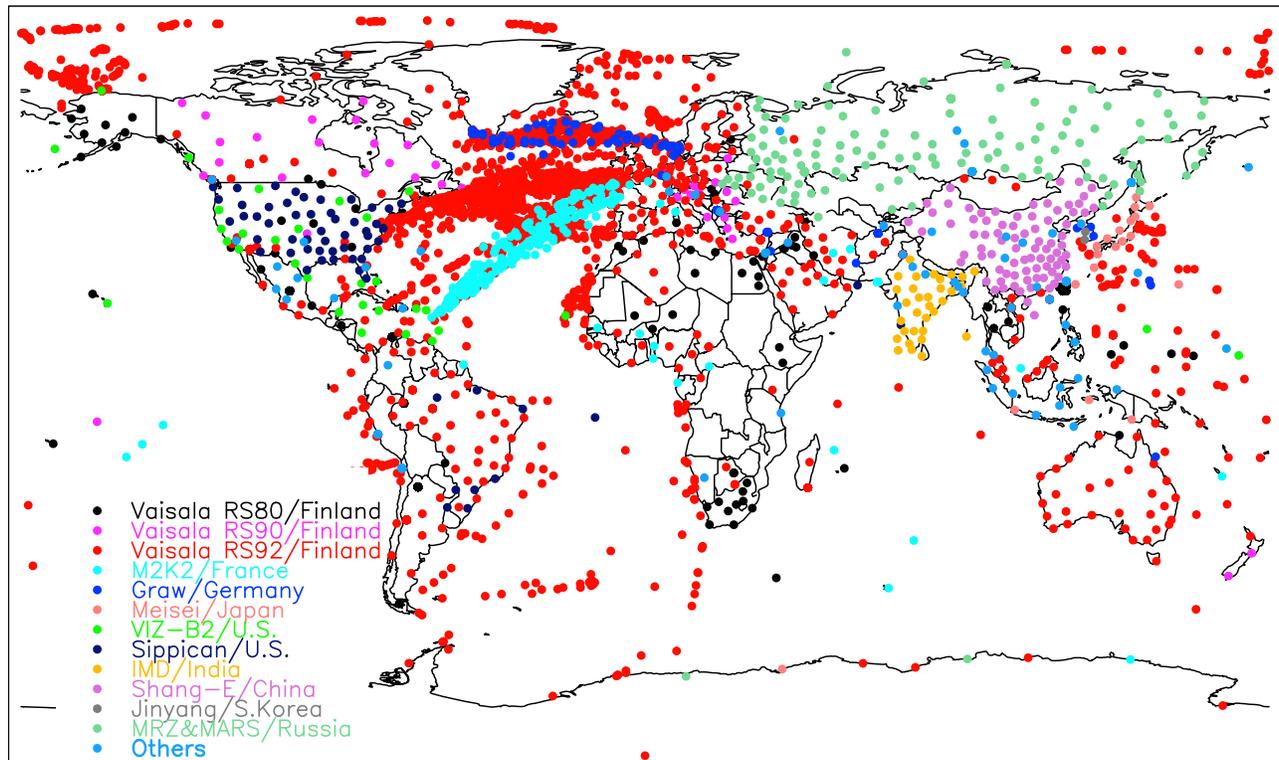


Figure 1. Global distribution of radiosonde stations and ship reports collocated within 7 h and 250 km of COSMIC soundings for April 2008 to October 2009. See Table 1 for more information on 12 radiosonde types of interest (denoted with different colors).

water vapor measurements. *Kuo et al.* [2004] compared Global Positioning System (GPS) radio occultation (RO) refractivity profiles from the CHAMP (CHALLENGING Minisatellite Payload) satellite mission with radiosondes in five countries, each using a different sonde type, and suggested that the RO soundings are of sufficiently high accuracy to differentiate their performance. Comparing the upper tropospheric and lower stratospheric temperature profiles derived from GPS RO data from the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) mission [*Anthes et al.*, 2008] with those from four major sonde types, *He et al.* [2009] indicated the COSMIC temperature data are extremely useful as benchmark data for evaluating radiosonde instrument performance. Currently there are dozens of sonde types in the global network (Figure 1), and one goal of this work is to identify differences among those sonde types to better inform their use in characterizing satellite retrievals. In this study, we examine atmospheric temperature (T), relative humidity (RH), and refractivity (N) profiles, by comparing radiosonde profiles of 12 sonde types flown in the global operational radiosonde network with COSMIC RO sounding data, for which the error characteristics have been studied in detail [e.g., *Kursinski et al.*, 1997; *Kursinski and Hajj*, 2001; *Kuo et al.*, 2004; *Sokolovskiy et al.*, 2006; *Ao*, 2007; *Schreiner et al.*, 2007; *Neiman et al.*, 2008; *Yunck et al.*, 2009].

[5] Current raobs certainly do not constitute a perfect reference system for anchoring satellite products, but even with much more accurate reference-quality soundings (as we expect from the Global Climate Observing System's Refer-

ence Upper-Air Network [*Seidel et al.*, 2009]), the collocation issue will remain a source of uncertainty in comparisons. So the second goal of this study is to quantify the impact of temporal and spatial mismatches between satellite and radiosonde data on the accuracy assessment of satellite retrievals. Atmospheric variability on the space and time scales of such collocation mismatches lowers the perceived accuracy of satellite retrievals when they are evaluated against the raobs, as shown by *Tobin et al.* [2006]. This problem is compounded by radiosonde balloon drift, which can result in ~ 200 km horizontal displacements during the ~ 100 min ascent from the surface to stratosphere, and this problem has not been evaluated in previous investigations.

[6] We employ collocated raob and COSMIC soundings to address both goals. Launched in April 2006, COSMIC currently yields around 2000 all-weather soundings per day distributed around the globe from which N, T, and RH profiles can be extracted. Because biases in COSMIC sounding data are relatively small and are expected to be stable in time and space [*Kuo et al.*, 2004], COSMIC serves as a globally consistent reference system for our comparison of radiosonde types T, RH, and N profiles, for both day and night. The high vertical resolution (from ~ 100 m in the lower troposphere to ~ 1 km in the lower stratosphere) and relatively even temporal and spatial distribution of COSMIC observations should allow more detailed analysis than previous studies using Advanced TIROS Operational Vertical Sounder (ATOVS), with its lower vertical resolution and regional sampling biases (due to the Sun-synchronous polar satellite orbit and synoptic radiosonde observation times)

Table 1. Characteristics of Major Radiosonde Types Used in the Study^a

Type	Manufacturer or Country	Temperature Sensor	Humidity Sensor	BUFR Code (Subtype)	Number of Reports
RS80	Vaisala (Finland)	Bead thermocap	A/H-Humicap	037, 052, 060, 061, 062, 063, 067	2478
RS90	Vaisala (Finland)	Thin wire F- thermocap	Twin H-Humicap	071	1371
RS92	Vaisala (Finland)	Thin wire F- thermocap	Twin H-Humicap	034, 079, 080, 081	14,130
M2K2 and M2K2-DC	Modem (France)	Bead or Chip thermistor	Capacitive polymer	056, 057	949
DFM-06 and DFM-90	Graw (Germany)	Bead thermistor	Capacitive polymer	017, 050	207
RS2-91 and RS-01G	Meisei (Japan)	Rod thermistor	Capacitive polymer	047, 055	749
VIZ-B2	VIZ (U.S.)	Rod thermistor	Carbon hygristor	049, 051	1276
Sippican Mark II and IIA	Sippican (U.S.)	Chip thermistor	Carbon hygristor	084, 085, 086, 087	4003
Mark IV	IMD (India)	Rod thermistor	Carbon hygristor	020	287
Shang-E	Shang (China)	Rod thermistor	Carbon hygristor	032	2674
VIZ Mark II	Jinyang (Korea)	Rod thermistor	Carbon hygristor	021	165
MRZ and MARS	(Russia) ^b	Rod thermistor	Goldbeater's skin hygrometer	027, 028, 029, 058, 068, 069, 075, 088, 089	8969

^aRadiosonde report counts are global totals for April 2008 to October 2009 (see Figure 1).

^bThe Russian sonde types are made by different manufacturers. Very recent preliminary information indicates that about 36 percent of these sounding reports were made by capacitive polymer humidity sensors.

[Reale *et al.*, 2008] or from intensive, but spatially and temporally restricted, research programs [Tobin *et al.*, 2006; Adam *et al.*, 2010; Pougatchev *et al.*, 2009].

[7] Section 2 describes the raob-COSMIC collocation data set, and section 3 presents the methodology of assessing radiosonde type characteristics and mismatch impacts. Characteristics for 12 major sonde types are presented in section 4 by comparing their observations with collocated COSMIC data, followed by analysis of the impacts of the time and distance mismatch. Discussion and a summary of findings are given in sections 5 and 6, respectively.

2. Collocated Raob and COSMIC Profile Data

[8] The raob-COSMIC collocation data set is extracted from the NOAA Products Validation System (NPROVS) operated by the Office of Satellite Applications and Research (STAR) at NESDIS. NPROVS routinely assembles collocated radiosonde and derived satellite atmospheric sounding products from a constellation of seven environmental satellites (including NOAA-18, NOAA-19, MetOp, Aqua-AIRS, GOES, DMSP-F16, and COSMIC). The system supports the National Polar-orbiting Operational Environmental Satellite System calibration and validation of the Cross-track Infrared Sounder (CrIs)—Advanced Technology Microwave Sounder (ATMS) Environmental Data Records [Reale *et al.*, 2010].

[9] Radiosonde data used in NPROVS collocations are those routinely utilized during the NOAA/National Centers for Environmental Prediction (NCEP) operational Numerical Weather Prediction (NWP) assimilation. Quality control (QC) markers of key variables determined during the NWP assimilation and ancillary information, such as balloon drift and collocated NWP data, are appended to the radiosonde data. NPROVS applies quality assurance procedures to these NCEP-provided radiosonde data, which are then treated as the “anchor” for compiling collocated satellite observations. The procedures include two steps: rejecting the data values at levels recommended by NCEP QC markers and excluding remaining outliers detected in tests against climatological values. The balloon drift data are computed from radiosonde wind and height data by NCEP (D. Keyser, personal commu-

nication, 2010). It should be noted that radiosonde observations do not include station or balloon launch locations. Station locations come only from catalogs and each data center or agency has its own catalog, updated independently. Uncertainties are not uncommon in many catalog station locations, but the average errors are no more than a few kilometers. The balloon drift path is always computed relative to its launch location, so the uncertainty in balloon drift location is not large enough to alter any basic conclusions drawn from this collocation mismatch impact study.

[10] Like Kuo *et al.* [2005] and He *et al.* [2009], collocation mismatch values in this study are computed from the times and locations of the radiosonde balloon launch and COSMIC occultation point, i.e., the point on Earth's surface to which the retrieved refractivity profile is assigned, which is generally located at 2–4 km above the surface along the refractivity profile trajectory [Kuo *et al.*, 2004]. The maximum mismatch allowed for these collocations is 7 h and 250 km, although drift of either the radiosonde or the COSMIC profile trajectory can increase (or decrease) the mismatch. If multiple COSMIC soundings are within the window, only the one that is closest in distance and time to the raob is collocated. For the temporal and spatial mismatch impact analysis, we required raob drift information, which was not available at all stations; in China only 10 (of the total 90) stations provide drift information.

[11] NPROVS began routine collocations of global radiosonde reports with satellite sounding products in April 2008. For the period April 2008 to October 2009, this procedure resulted in ~205,000 raob-COSMIC collocations, from 721 stations and 27 ships, distributed over the globe. Most (~68%) collocations are in Northern Hemisphere midlatitude land areas where the raobs are concentrated, as shown in Figure 1, which also shows the reported major radiosonde types. Most raob stations used a single radiosonde type during this period, but some stations in Russia and elsewhere used multiple types. The Vaisala RS92 was the most widely used type, flown over many parts of the world, including Australia, South America, Europe, Canada, West Africa, and many islands of the global oceans. In the U.S. network, Sippican Mark IIA, VIZ-B2, and Vaisala RS80 were the major types flown. In China, starting January 2002, the digital Shang-E

with a rod thermistor and carbon hygistor was introduced into the radiosonde network to replace the Shang-M with a bimetal coil temperature sensor and goldbeater's skin humidity sensor; by August 2008, the Shang-E was used in 82 of the 90 Chinese stations. In Russia, more than 10 radiosonde types were flown; we selected nine types with rod thermistor and grouped them into a single category for analysis.

[12] Table 1 gives details of the 12 radiosonde types shown in Figure 1, including their subtypes (indicated by the three-digit sonde codes transmitted with the data). Following Wang and Zhang [2008], the 12 types can be grouped into three main categories according to humidity sensor type: capacitive polymer (Vaisala RS80, RS90, and RS92, M2K2 and M2K2-DC, DFM-90 and DFM-06, and RS2-91 and RS-01G), carbon hygistor (VIZ-B2, Sippican Mark II and Mark IIA, IMD Mark IV, Shang-E, and Jinyang VIZ Mark II), and goldbeater's skin. Column-integrated tropospheric water vapor data for capacitive polymer sensors have been found to have a dry bias while the other two a moist bias [Wang and Zhang, 2008]. Because of their poor response in low T and when subject to daytime solar heating, most of the radiosonde types listed in Table 1 show a dry bias in the upper atmosphere [Wang et al., 2003; Nakamura et al., 2004; Miloshevich et al., 2006] except the Russian sondes which have a tropospheric moist bias [Soden and Lanzante, 1996; Wang and Zhang, 2008].

[13] The NCEP-provided raob T data used in this analysis experience two stages of radiation correction. The first stage is a correction applied at the field station using schemes provided by radiosonde vendors or specified by the country. Depending on radiosonde types or stations, the correction can be made for solar heating bias only or the combination of solar heating and infrared cooling bias. The second stage is made by NCEP during their data assimilation after station data are transmitted over the Global Telecommunication System (GTS). The NCEP correction schemes are based on comparison of raob data with NCEP's Global Data Assimilation System 6 h forecast, for different solar elevation classes and for different radiosonde types. Most of the data corrected by NCEP are for types that were already in use before around 1999, and no corrections are made on more recent types, including Vaisala RS92, Sippican Mark series and others, because the NCEP correction schemes have not been updated since 1999 (B. Ballish and S. Schroeder, personal communication, 2010). The NCEP corrections address only T, not dew point temperature or RH data.

[14] To compare raob and COSMIC T, RH, and N profiles, we compute raob N from raob T, humidity, and pressure data using the formula [Smith and Weintraub, 1953]

$$N = 77.6 \frac{P}{T} + 3.73 \times 10^5 \frac{P_w}{T^2}, \quad (1)$$

where P is atmospheric pressure (hPa) and P_w atmospheric vapor pressure (hPa), T is in Kelvin, and refractivity N is defined as $10^6 \times (\text{refractivity index} - 1)$. COSMIC N, T, P and P_w profiles data are the near-real-time L2 products obtained from the COSMIC Data Analysis and Archive Center (CDAAC). COSMIC T and P_w profiles were generated from N profiles with a one-dimensional variational scheme that uses analyses of the NCEP Global Forecast System (GFS) model as its first guess. (For details, see the

COSMIC CDAAC website: <http://cosmic-io.cosmic.ucar.edu/cdaac/doc/index.html>.) COSMIC RH profiles are computed from COSMIC T, P_w , and P profiles.

[15] High precision and accuracy of the RO soundings are the basis for the radiosonde data error characteristics evaluation to be discussed in sections 3.1 and 4.1. The RO N error obtained from the analysis of real N data against NWP analyses, which involves both measurement and representativeness errors, is 0.3–0.5% for the altitudes of 5–25 km [Kuo et al., 2004]. The theoretical estimate of the RO N measurement error is smaller (i.e., 0.2% for the altitudes of 5–30 km [see Kursinski et al., 1997]). In the mid-upper troposphere and stratosphere, where the moisture contribution to N is negligible, the retrieved T is supposed to be very close to the dry T, and our calculation (not shown) appears to confirm that (i.e., the difference between the UCAR COSMIC dry and retrieval T is small, < 0.03 K). N and dry T profiles in the upper troposphere and lower stratosphere have been considered accurate enough to assess the quality of other types of observational data [Kursinski and Hajj, 2001; Kuo et al., 2005; Ho et al., 2007; He et al., 2009] compared RO moisture data with ECMWF analyses and suggested that the derived moisture in the mid-upper troposphere should be accurate to about 0.1 g/kg. The high vertical resolution of COSMIC moisture retrievals is also considered to be useful in monitoring atmospheric mesoscale phenomena in data-spare regions [Neiman et al., 2008]. Our calculation (not shown) finds the COSMIC RH data in the mid-upper troposphere are comparable to the NCEP GFS forecast data (which is also shown by Neiman et al. [2008]) with the former moister than the latter by 0.5%, suggesting the COSMIC humidity retrievals are largely influenced by NCEP GFS first guess, but both are moister than the radiosonde measurements by 2% at 400 hPa, increasing to 7% at 200 hPa.

[16] Because RO signals do not penetrate into the lower troposphere, particularly in moist regions, and because of signal blockage in high terrain, the availability of RO soundings in the lower troposphere is limited. Therefore, we limit our analysis to 29 predefined pressure levels from 850 to 10 hPa.

3. Methodology

[17] The mean and standard deviation of satellite-minus-raob differences are commonly used to evaluate derived satellite product performance. Here we instead calculate mean and standard deviation of raob-minus-satellite differences to focus on radiosonde bias characteristics and on the effects of collocation mismatch,

$$\overline{\Delta X} = \frac{1}{n} \sum_{i=1}^n (X_i^r - X_i^c) \quad (2a)$$

$$SD_{\Delta X} = \sqrt{\frac{1}{n} \sum_{i=1}^n ((X_i^r - X_i^c) - \overline{\Delta X})^2}, \quad (2b)$$

where X^c and X^r represent the respective COSMIC and raob T or RH, the subscript i the ith raob-COSMIC collocation, and n the number of collocations.

[18] For refractivity N, $\overline{\Delta X}$ and $SD_{\Delta X}$ represent the mean fractional difference and standard deviation of the fractional

difference between raob and COSMIC, respectively. They are computed using the following formula:

$$\overline{\Delta X} = \frac{1}{n} \sum_{i=1}^n \frac{(X_i^r - X_i^c)}{X_i^c} \quad (3a)$$

$$SD_{\Delta X} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\left(\frac{(X_i^r - X_i^c)}{X_i^c} \right) - \overline{\Delta X} \right)^2} \quad (3b)$$

3.1. Comparison of Radiosonde Types

[19] Characteristics of 12 radiosonde types with different humidity and temperature sensors are identified by comparing their measurements with the collocated COSMIC data within 3 h and 150 km windows using $\overline{\Delta X}$ and $SD_{\Delta X}$. Collocation mismatch impacts are not considered in this analysis because (1) $\overline{\Delta X}$ does not vary significantly with the size of time and distance mismatch, (2) limited collocation samples for some radiosonde types preclude a statistically meaningful evaluation of mismatch impacts, and (3) data for some radiosonde types do not include balloon drift information, as mentioned above.

[20] To cross check the radiosonde type characteristics identified using COSMIC data, observations of radiances, from polar-orbiting satellites, are compared with radiances calculated from collocated raobs using the Community Radiative Transfer Model (CRTM) [Han *et al.*, 2006] (see section 4.1).

3.2. Evaluation of Imperfect Collocation

[21] We assess collocation mismatch impacts as follows. First, we found all of the raob-COSMIC pairs within 6 h and 250 km, based on times and locations of the radiosonde balloon launch and COSMIC occultation point (as stated in section 2). Second, for those selected pairs, time and distance mismatch values are computed for each height level by using the radiosonde balloon time and location data at that level and COSMIC time at the occultation point and location at that level. Third, for each level, based on the computed mismatch values, raob-COSMIC pairs are divided into bins of 1 h time intervals centered at 0.5, 1.5, ..., and 5.5 h and 50 km distance intervals centered at 25, 75, ..., and 275 km. Then $\overline{\Delta X}$ and $SD_{\Delta X}$ are computed for each of these 36 bins using equations (2a) and (3a) and equations (2b) and (3b), respectively. Finally, impacts of collocation mismatch on satellite retrieval validation statistics are assessed by quantifying the changes of $\overline{\Delta X}$ and $SD_{\Delta X}$ with time mismatch at constant distance, and with distance mismatch at constant time.

[22] $\overline{\Delta X}$ was found to not change significantly with time or distance mismatch, suggesting the atmospheric variability introduced by collocation mismatch for mean raob-minus-COSMIC difference is random and averages to zero for a sufficiently large sample. The mismatch impact assessment to be discussed in section 4 is therefore focused on $SD_{\Delta X}$ as described by equations (4a) and (4b).

$$SD_{\Delta X}(d) = C(t \approx 0) + \frac{\partial(SD_{\Delta X})}{\partial(t)} * \Delta t \quad (4a)$$

$$SD_{\Delta X}(t) = C(d \approx 0) + \frac{\partial(SD_{\Delta X})}{\partial(d)} * \Delta d. \quad (4b)$$

[23] In equation (4a), $C(t \approx 0)$ and $\partial(SD_{\Delta X})/\partial(t)$ are the $SD_{\Delta X}$ value associated with zero time mismatch (perfect temporal match) and the $SD_{\Delta X}$ sensitivity to time mismatch, respectively, both corresponding to the distance mismatch of d . Similarly, $C(d \approx 0)$ and $\partial(SD_{\Delta X})/\partial(d)$ in equation (4b) are the $SD_{\Delta X}$ value associated with zero distance mismatch (perfect spatial collocation) and the $SD_{\Delta X}$ sensitivity to distance mismatch, respectively, both corresponding to the time mismatch of t .

[24] To remove the distance mismatch impact, $C(t \approx 0)$ and $\partial(SD_{\Delta X})/\partial(t)$ corresponding to zero distance mismatch are linearly extrapolated from values of $C(t \approx 0)$ and $\partial(SD_{\Delta X})/\partial(t)$, respectively, which are estimated at distance mismatch of 25 km, 75 km, ..., and 275 km, and an analogous procedure is applied to remove the time mismatch impact. These intercept and slope values are used to assess collocation mismatch impacts as will be discussed in section 4.2.

4. Results

4.1. Biases in Radiosonde Profiles

[25] In this section, we present radiosonde error characteristics by sonde type. Results are shown both for combined daytime and nighttime data, and for daytime and nighttime data separately, for T, RH and N profiles. Results based on "All" types give estimates of the overall performance of the global operational network.

4.1.1. Temperature

[26] Mean T differences, $\overline{\Delta T}$, between raobs and COSMIC vary according to sonde type. However, results for the overall global network (thick black curves in Figure 2) show agreement within 0.15 K (Figure 2a), with $SD_{\Delta T}$ of 1.5–2.0 K (Figure 2b) throughout the troposphere and lower stratosphere. The sign of ΔT , however, changes with height: positive $\overline{\Delta T}$ below 500 hPa for most radiosonde types; slightly negative $\overline{\Delta T}$ (~ -0.15 K) from the upper troposphere to ~ 50 hPa; positive $\overline{\Delta T}$ (< 0.15 K) in the stratosphere above 50 hPa for several types, probably reflecting solar radiation errors in the raob data.

[27] Note that the error statistics of individual sonde types for temperature in this section, and for relative humidity and refractivity in sections 4.1.2 and 4.1.3, were computed based on the samples listed in Table 1, whose sizes vary significantly among sonde types. Temperature error characteristics for individual radiosonde types include the following.

[28] 1. The Graw DFM-06 and DFM-90, IMD Mark IV, and Jinyang VIZ Mark II have anomalously large $\overline{\Delta T}$ and $SD_{\Delta T}$ (Figure 2), which is also seen in the RH results (Figures 3a and 3b). Similar error characteristics in GRAW DFM-97 were also noticed in the World Meteorological Organization (WMO) radiosonde intercomparisons conducted at Alcântara, Maranhão, Brazil, 29 May to 10 June 2001 and Vacoas, Mauritius, 7–27 February 2005 [Nash *et al.*, 2006]. IMD Mark IV and Jinyang VIZ Mark II were not included in those experiments.

[29] 2. Vaisala RS80 and Meisei RS2-91 and RS-01G sondes have warm biases of 0.2–0.4 K, and Modem M2K2

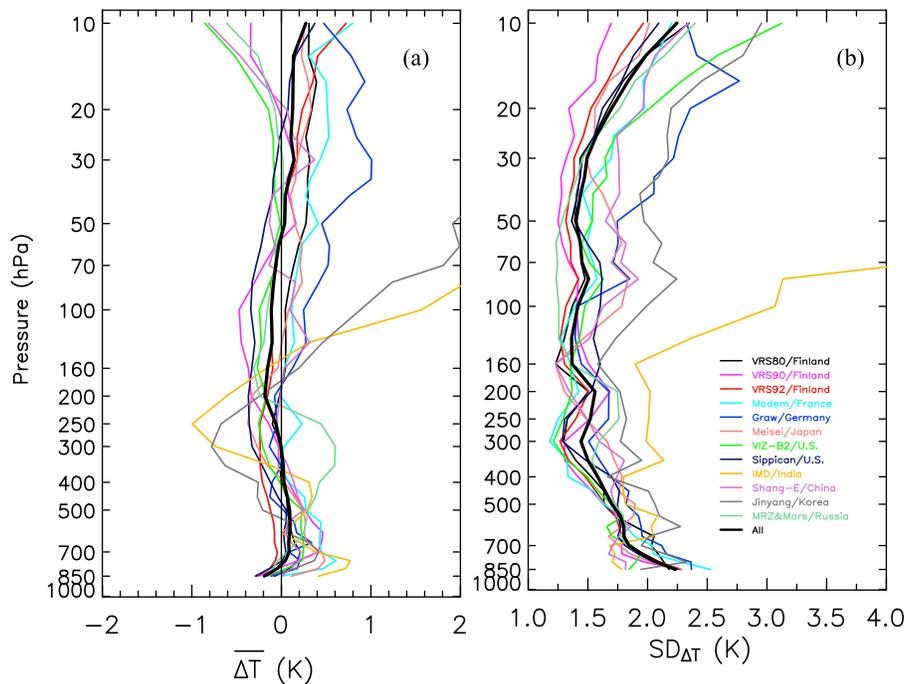


Figure 2. Comparison of radiosonde and COSMIC temperature profiles, by radiosonde type (colored curves) and for the network average (thick black curves). (a) Mean raob-minus-COSMIC temperature difference. (b) Standard deviation of the raob-minus-COSMIC temperature difference.

and M2K2-DC have warm biases of 0.2–0.5 K, compared with COSMIC in the upper troposphere and lower stratosphere, possibly due to undercorrected solar radiation error.

[30] 3. For Vaisala RS92, we find an increasing warm bias (from ~zero to ~0.4 K) with altitude above 50 hPa. Since no radiation correction for this sonde was done at NCEP, this result suggests the solar heating might be insufficiently corrected at the sites.

[31] 4. MRZ and MARS, VIZ-B2, and Sippican Mark II and Mark IIA show a cool bias of <0.4 K for the upper troposphere and lower stratosphere, with the Sippican sondes showing larger biases. This may be due to over-correction of solar radiation errors and/or inadequate treatment of errors due to long-wave radiative cooling. The MRZ & MARS sondes have warm biases of 0.3–0.6 K in the mid-to-upper troposphere.

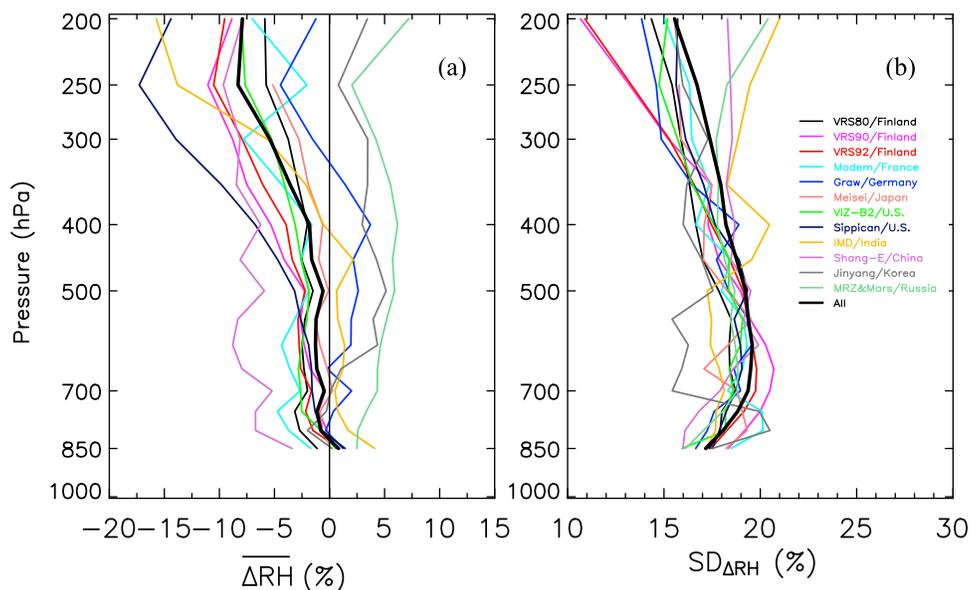


Figure 3. Same as Figure 2 but for relative humidity.

[32] *He et al.* [2009] presented radiosonde-COSMIC temperature differences in the height interval 12–25 km for Vaisala RS92, MRZ, Shang-E, and VIZ-B2 sondes. We attempted to replicate their results, with mixed success. Our results are in agreement with *He et al.* [2009] for the VIZ-B2 sonde ($\overline{\Delta T} = -0.20$ K, $SD_{\Delta T} = 1.68$ K) and for the Shang-E sonde ($\overline{\Delta T} = 0.05$ K, $SD_{\Delta T} = 1.83$ K). However, for Vaisala RS92, we obtain $\overline{\Delta T} = -0.12$ K, compared with 0.03 K in *He et al.* [2009], which may be due to our use of global data, including ship measurements, while *He et al.* [2009] used only land data, and/or our use of COSMIC retrieval T while *He et al.* [2009] used dry T (even though they differ by < 0.03 K on global average; see section 2). For the Russian radiosondes, we obtain biases of opposite signs: $\overline{\Delta T} = -0.20$ K in our analysis and $\overline{\Delta T} = 0.26$ K from *He et al.* [2009]. The reason for the difference is likely because *He et al.* [2009] used only one MRZ type, for which radiation correction was made only at the site, while this analysis used nine MRZ and MARS subtypes, and more importantly, some of the subtype data have experienced radiation correction both at the field site and at NCEP as well. This highlights the sensitivity of these bias statistics to choices of radiosonde group definitions, and to prior corrections to the sonde data. Comparing $SD_{\Delta T}$ results, we obtain values on average 13.5% smaller than those of *He et al.* [2009] for all of four radiosonde types analyzed. This is probably because their analysis allowed collocations within 300 km, while our collocations window was only 150 km. As shown below (section 4.2), $SD_{\Delta T}$ increases significantly with the increase in mismatch.

[33] Although most of the radiosonde types used in this study have undergone radiation error correction either at the site, or by NCEP during their assimilation process, and so should have no day-night difference in T error characteristics, this is not what we found. For most radiosonde types, daytime ΔT tends to be larger than nighttime for the upper troposphere and stratosphere (not shown). Average (over all types) day-minus-night $\overline{\Delta T}$ increases from about zero at 350 hPa, to 0.10 K at 50 hPa, to 0.20–0.30 K at 10–20 hPa. These findings suggest a residual, uncorrected daytime radiation error in NCEP raob data.

4.1.2. Relative Humidity

[34] Relative to COSMIC, most of the radiosonde types show a dry bias which increases with altitude from the middle troposphere to the upper troposphere (Figure 3a), consistent with *Soden and Lanzante* [1996], who compared radiosonde data with satellite measurements, *Wang and Zhang* [2008], who compared radiosonde data with ground-based GPS total precipitable water data, and other studies [e.g., *Wang et al.*, 2003; *Nakamura et al.*, 2004; *Vömel et al.*, 2007]. The moist bias of $\sim 5\%$ throughout the troposphere in Russian sensors is an exception among the radiosonde types discussed (Figure 3a). On average for the whole network (thick black curve in Figure 3a) $\overline{\Delta RH}$ is $< 2\%$ in the low troposphere and increases to 5–8% in the upper troposphere.

[35] Major characteristics of RH measurements for individual radiosonde types are summarized as below.

[36] 1. Vaisala RS90 and RS92 show similar dry biases (e.g., of $\sim 10\%$ at 250–300 hPa) and $SD_{\Delta RH}$ values throughout the troposphere, which is consistent with the fact that they carry equivalent sensors in terms of calibration accuracy and time response [*Miloshevich et al.*, 2006].

In the upper troposphere, the dry bias for RS90 and RS92 is greater than for Vaisala RS80. One might argue that this is a result of solar radiative heating of the RS90 and RS92 twin H-Humicap sensor both prior to launch and during flight [*Vömel et al.*, 2007], and an aluminized plastic shield over the RS80 A/H Humicap reduces this effect [*Wang and Zhang*, 2008]. However, even the nighttime dry bias for RS92 and RS92 is greater than for RS80 (Figure 6). The smaller dry bias in RS80 may be fortuitous: On average the RS80 stations are located in moister regimes than RS90 and RS92. Average 250 hPa RH is 29.4% for RS80, 27.3% for RS90, and 25.3% for RS92. So the RS80 sondes are subject to less challenging measurement environments.

[37] 2. Data from the VIZ-B2 and Sippican carbon hygrometers have a dry bias from 700 hPa to the upper troposphere, where the Sippican bias reaches 10–15% RH (Figure 3a). The cause for the dry bias is not known, but these types of sensors have known time lags and poor response to humidity changes at low temperature [*Wang et al.*, 2003].

[38] 3. The Shang-E carbon hygrometer shows a dry bias of 6–10% from 850 to 300 hPa. This may be because the Shang-E carbon hygrometer was calibrated against Vaisala RS80-A which has its own dry bias; Y. Chu et al. (Dry bias in the Beijing radiosonde soundings as revealed by GPS and MWR measurements, 2006, <https://www.Eol.ucar.edu/icmcs/presentations>) report a 5–15% bias compared with dew point hygrometers. However, the Shang-E dry bias appears to be greater than that of RS80-A.

[39] 4. The Russian MRZ and MARS sondes have a moist bias (of 2–3%) in the mid-upper troposphere, as does Jinyang VIZ Mark II.

[40] To further investigate radiosonde type humidity characteristics, we performed a similar analysis with independent satellite humidity observations from the MetOp Microwave Humidity Sensor (MHS). Figure 4a shows the radiosonde type RH bias at 300 hPa relative to the COSMIC. Figure 4b, on the other hand, shows the difference (radiosonde minus MHS) in upper tropospheric humidity-sensitive channel 3 brightness temperature (BT). The raob BT is calculated from temperature and water vapor profiles using CRTM and compared with collocated satellite BT observations. Figure 5 shows similar cross-validation results for 550 hPa. The qualitative agreement between these two sets of independently computed biases, for both pressure levels, enhances our confidence in the results.

[41] Comparison of daytime versus nighttime $\overline{\Delta RH}$ indicates daytime radiosonde biases tend to be drier (or less moist) than nighttime ones. Figure 6 shows an example of daytime and nighttime $\overline{\Delta RH}$ at 300 hPa. Statistically significant daytime versus nighttime $\overline{\Delta RH}$ differences (at the 0.05 or better levels) are found for all radiosonde types except Graw DFM-06 and DFM-90 and IMD Mark IV. Note for some sonde types the daytime and nighttime samples sizes differ significantly. For example, for Graw/Germany, the daytime and nighttime reports are 58 and 138, respectively, and the bias averaged from the daytime and night differences shown in Figure 6 is -3.1% ; this is larger than the all-day bias of -1.6% (shown in Figure 4), which is averaged from values of the all-day samples. On average for the whole network, the daytime and nighttime $\overline{\Delta RH}$ at 300 hPa are

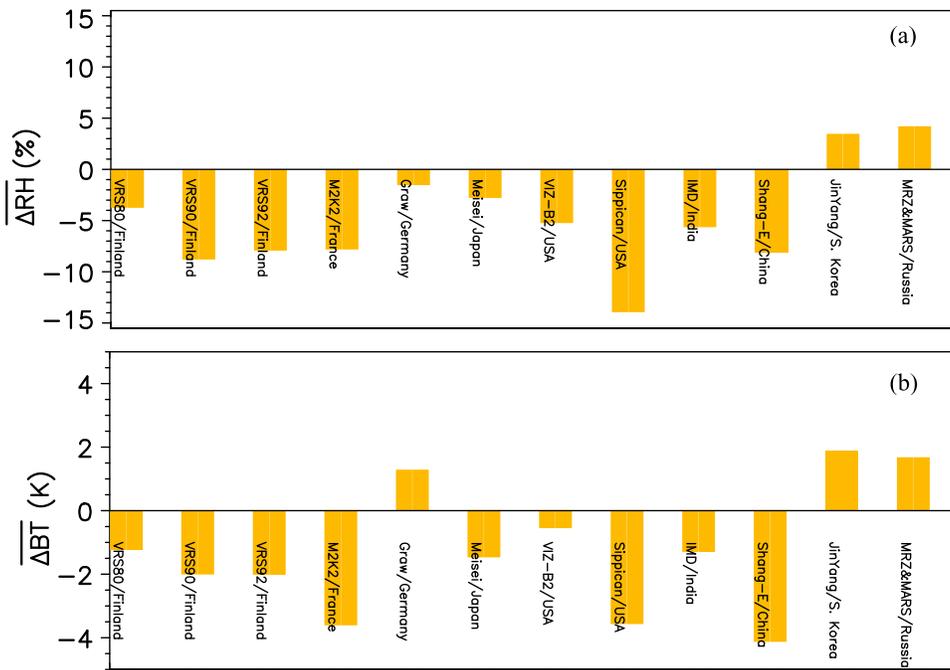


Figure 4. Comparison of raob relative humidity biases based on COSMIC and MetOp Microwave Humidity Sensor (MHS) data, for different radiosonde types. (a) Mean raob-minus-COSMIC relative humidity difference at 300 hPa. (b) Mean raob-minus-MHS upper tropospheric brightness temperature. The MHS data is for channel 3 (183.31 ± 1 GHz).

-7.2% and -3.3%, respectively, and are statistically significantly different at the 0.001 level.

4.1.3. Refractivity

[42] Atmospheric refractivity N depends directly on T and water vapor, so we expect raob-minus-COSMIC N differ-

ences (which we express as fractional differences) to be related to results for T and RH discussed above. Fractional N differences are generally negative (within -0.3%) in the lower troposphere and are slightly positive (average 0.1%) in the upper troposphere (Figure 7a). The negative N bias

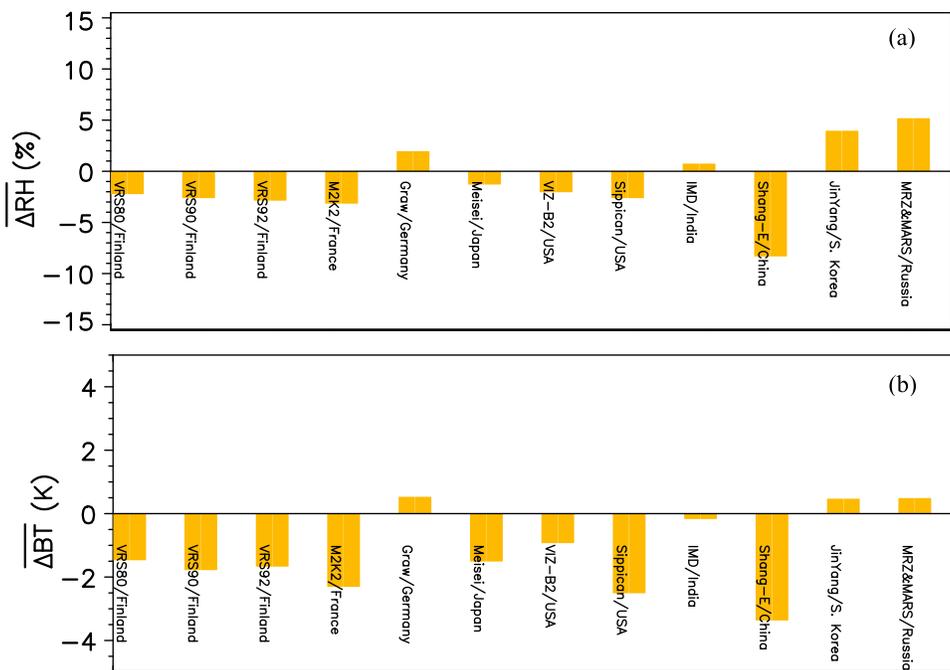


Figure 5. Same as Figure 4 but for (a) 550 hPa and (b) brightness temperature for the MHS channel 4 middle troposphere spectral interval (183.31 ± 3 GHz).

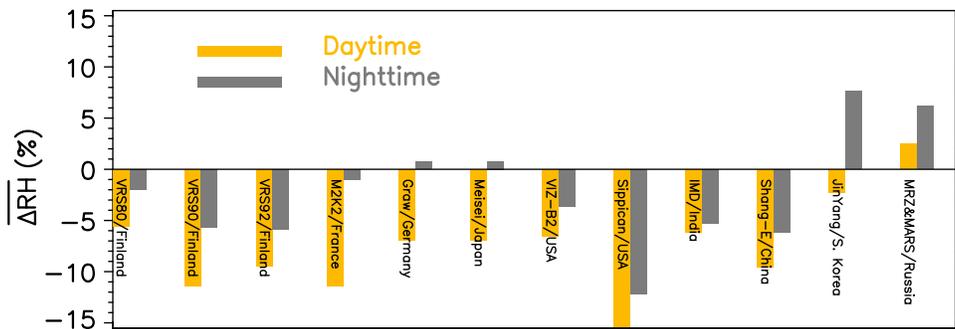


Figure 6. Mean 300 hPa raob-minus-COSMIC relative humidity difference for different radiosonde types for daytime and nighttime.

in the lower troposphere is consistent with the dry RH (Figure 3a), and the positive N bias in the upper troposphere is consistent with the warm T bias (Figure 2a).

[43] Consistent with their relatively large T and RH biases, the Graw, Jinyang, and IMD radiosondes show large N biases as well (Figure 7). The Shang-E radiosondes’ pronounced negative refractivity biases (i.e., 0.5–1.0% in the low-mid troposphere) likely are result of its dry RH bias, as its T bias is small. For the VIZ-B2 and Sippican radiosondes, the dry and cold biases both contribute to negative N bias from the low troposphere up to 350 hPa. Averaged over the 700 to 200 hPa layer, the Vaisala RS92 N, with mean fractional bias of -0.05% and standard deviation of 1.22% , is in closest agreement with COSMIC, despite the dry bias discussed above (Figure 4). Daytime $\overline{\Delta N}$ is less than the nighttime for most of the radiosonde types, consistent with the radiosonde daytime RH bias discussed above.

[44] To summarize section 4.1, radiosonde type characteristics revealed through comparing raob with collocated COSMIC data are basically consistent with, and extend,

results from field experiments and other comparisons. They indicate the value of COSMIC data for use as a relative reference to bring data from different radiosonde types into relative agreement for their better use in satellite sounding validation.

4.2. Assessment of Temporal and Spatial Collocation Mismatch Impacts

[45] Section 4.1 focused on mean differences (biases) between raob and COSMIC, which on global average are not particularly sensitive to temporal and spatial collocation mismatches within the mismatch window of 6 h and 250 km. However, the standard deviations of the differences, $SD_{\Delta X}$, are dependent on the closeness of the collocation, as would be expected. This section quantifies those dependencies for T, RH, and N (Figures 8, 9, and 10), using the methodology described in section 3.2 to evaluate both temporal and spatial mismatch effects. We first present results for each variable using 300 hPa data, to illustrate the method.

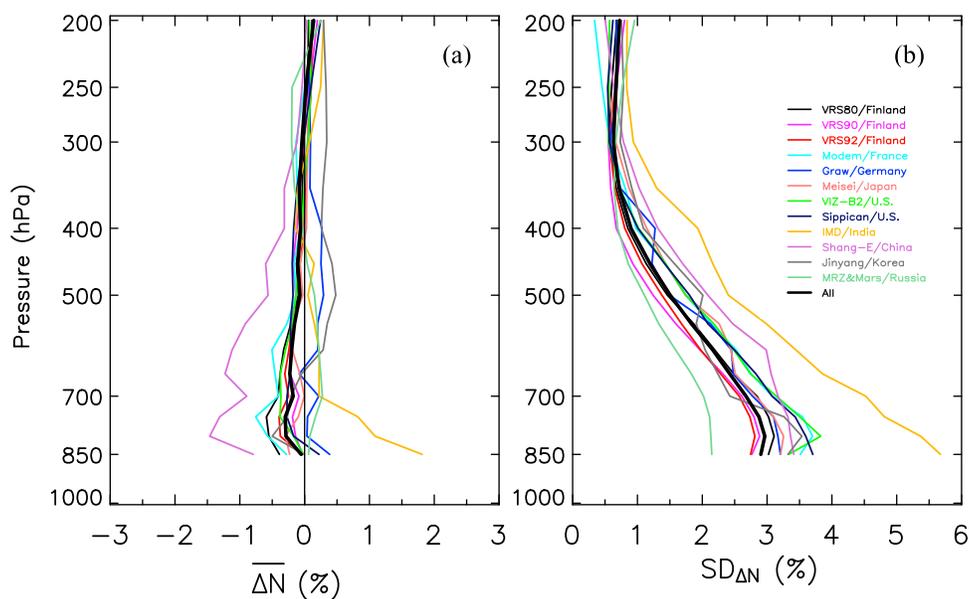


Figure 7. Same as Figure 2 but for (a) mean raob-minus-COSMIC fractional refractivity difference and (b) standard deviation of fractional difference.

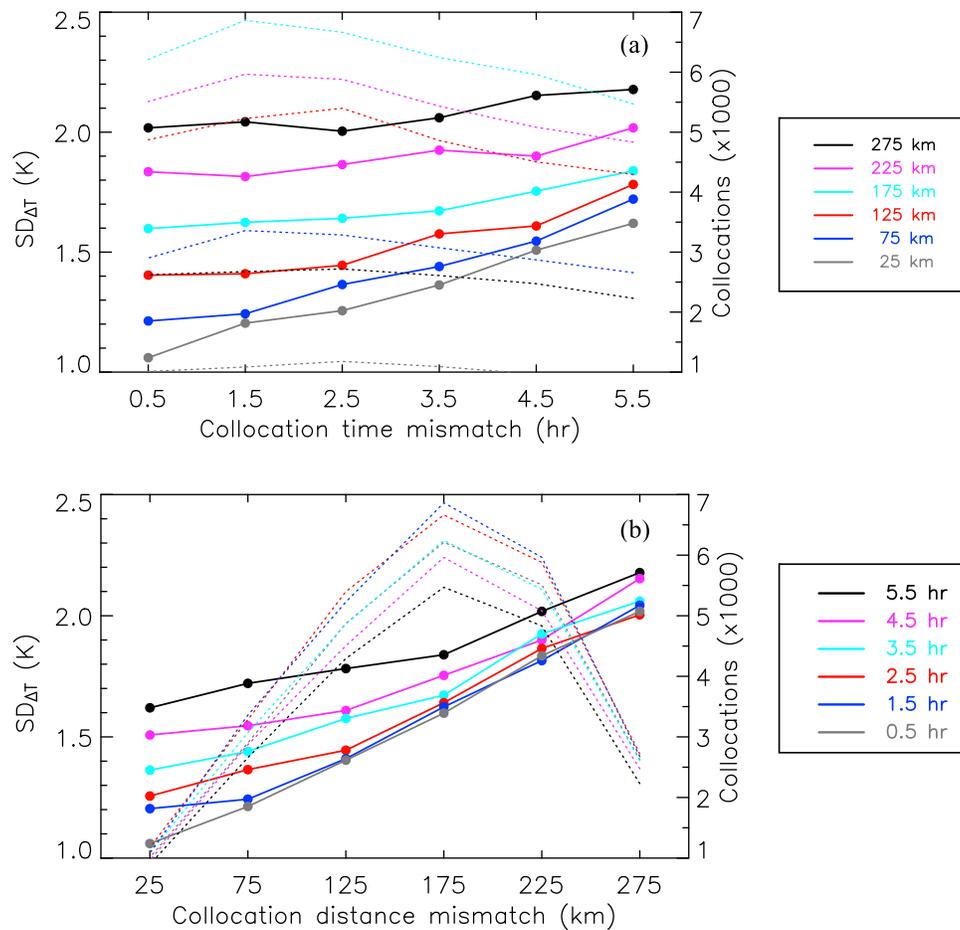


Figure 8. Dependence of 300 hPa standard deviation of raob-minus-COSMIC temperature difference on collocation mismatch for (a) time mismatch and (b) distance mismatch. Solid curves represent $SD_{\Delta T}$, and dotted curves denote the number of collocations used to compute $SD_{\Delta T}$.

[46] Figure 8a shows the increase in 300 hPa $SD_{\Delta T}$ with increasing time mismatch, for sets of collocations with distance mismatch centered at 25, 75, ..., and 275 km. (Dotted curves indicate corresponding collocation sample sizes.) For each distance mismatch bin (25, 75, ..., 275 km), we computed the time mismatch regression intercept values. They are 1.00, 1.12, 1.31, 1.55, 1.78, and 1.97 K, showing increasing impact with increasing distance mismatch. Extrapolating from these values, we obtain 0.85 K as the intercept value associated with zero distance mismatch. The corresponding regression slopes are 0.33, 0.30, 0.22, 0.14, 0.11, and 0.10 K/3 h, and the extrapolated slope associated with zero distance mismatch is 0.36 K/3 h. The decrease in slope values with increasing distance mismatch suggests the time mismatch impact is easier to detect for smaller distance mismatch.

[47] Figure 8b shows the increase in 300 hPa $SD_{\Delta T}$ with increasing distance mismatch, for sets of collocations with time mismatch centered at 0.5, 1.5, ..., and 5.5 h. The associated $SD_{\Delta T}$ versus distance mismatch regression intercept values are 0.93, 1.03, 1.13, 1.24, 1.37, and 1.54 K, and the extrapolated intercept value associated with zero time mismatch is 0.85 K. The regression slopes are 0.39, 0.35, 0.31, 0.29, 0.25, and 0.21 K/100 km, and the extrapolated slope associated with zero time mismatch is 0.40 K/100 km.

Again, this indicates the competing impacts of time and distance mismatch on $SD_{\Delta T}$.

[48] Similar to the mismatch impact on $SD_{\Delta T}$, $SD_{\Delta RH}$ generally increases with the increase in time mismatch (Figure 9a) and distance mismatch (Figure 9b). However, the relationship between $SD_{\Delta RH}$ and time or distance mismatch is not as strongly linear as for $SD_{\Delta T}$ (Figure 8). Due to poor humidity sensor performance in cold and dry environments, some radiosonde stations do not provide humidity measurements in the upper troposphere, so that collocation sample sizes for some bins, particularly those with small distance mismatch, (e.g., 25 km, the gray curve in Figure 9a) are too small for statistical analysis. Also, the noisiness of upper-tropospheric RH data may worsen the relationship between $SD_{\Delta RH}$ and time or distance mismatch. This appears to be confirmed by the stronger linear relationship between the mismatch impact and time or distance mismatch at lower than at upper levels (not shown).

[49] The intercept $SD_{\Delta RH}$ and slope $\partial(SD_{\Delta RH})/\partial(t)$ values associated with zero distance mismatch, as extrapolated from regression intercept values for distance mismatch of 25, 75, ..., and 275 km (Figure 9a), are 14.5% and 2.12%/3 h, respectively. The intercept $SD_{\Delta RH}$ and slope $\partial(SD_{\Delta RH})/\partial(t)$ values associated with zero time mismatch extrapolated

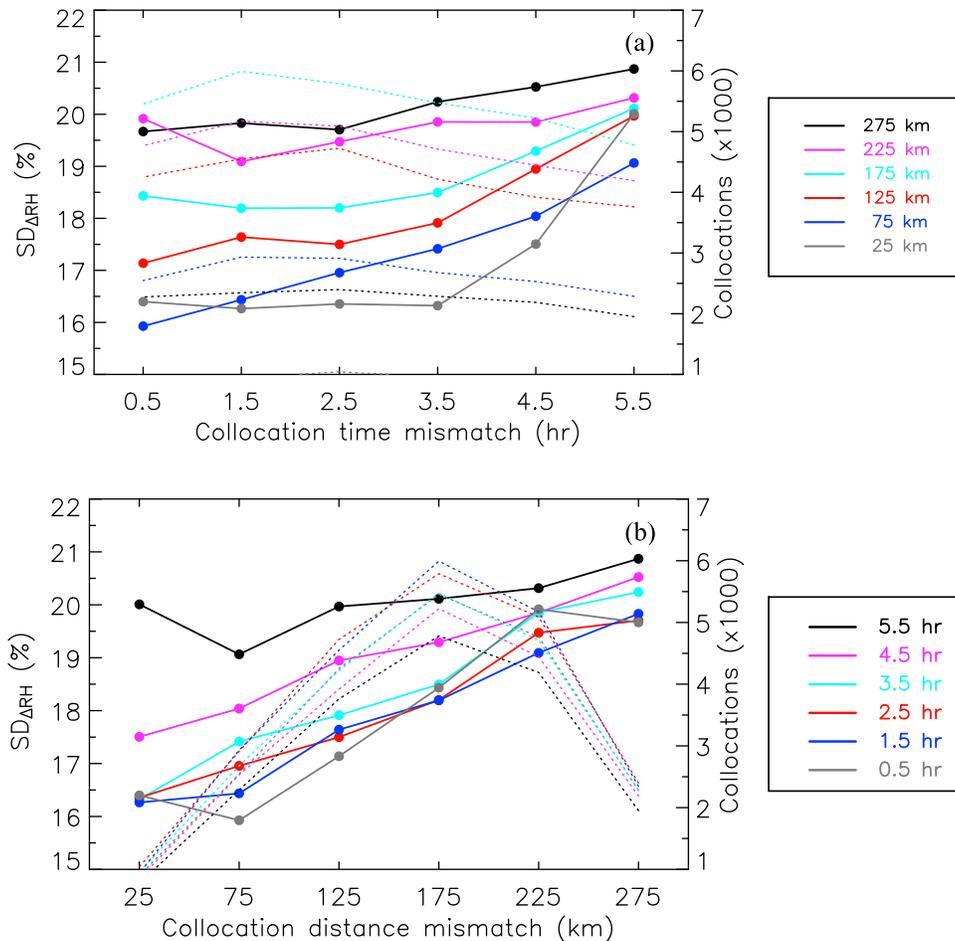


Figure 9. Same as Figure 8 but for relative humidity at 300 hPa.

from time mismatch of 0.5, 1.5, ..., and 5.5 h (Figure 9b) are 14.5% and 1.90%/100 km, respectively.

[50] Because N is more sensitive to T than RH in the upper troposphere, the relationship between $SD_{\Delta N}$ and time or distance mismatch at 300 hPa (Figures 10a and 10b) is similar to that for $SD_{\Delta T}$ (Figures 8a and 8b). The extrapolated intercept $SD_{\Delta N}$ and slope $\partial(SD_{\Delta N})/\partial(t)$ values associated with zero distance mismatch are 0.36% and 0.16%/3 h, respectively, and the extrapolated intercept $SD_{\Delta N}$ and slope $\partial(SD_{\Delta N})/\partial(d)$ values associated with zero time mismatch are 0.36% and 0.17%/100 km, respectively.

[51] We conducted similar analyses for all other levels and show vertical profiles of the regression intercept values for T , RH , and N in Figures 11, 12, and 13. In these plots, the dotted curves show the vertical profiles of the intercept values of $SD_{\Delta X}$ profiles (associated with zero time and distance mismatch), computed in the manner discussed above for Figures 8–10. Table 2 gives average values for vertical layers representing the troposphere (850–200 hPa) and stratosphere (200–10 hPa).

[52] As we found at 300 hPa, $SD_{\Delta T}$ increases with increasing time and distance mismatch (Figure 11) throughout the atmospheric profile, with maximum sensitivities at 200 hPa, the approximate level of the jet streams. Here the $SD_{\Delta T}$ sensitivity to time mismatch reaches 0.53 ± 0.046 K/

3 h and to distance mismatch reaches 0.49 ± 0.016 K/100 km (where uncertainties are given in terms of standard error).

[53] Using these sensitivity estimates, we can assess SD errors due to radiosonde balloon drift. For example, at 100 hPa, the $SD_{\Delta T}$ sensitivity to time mismatch is 0.27 ± 0.017 K/3 h and to distance mismatch is 0.34 ± 0.025 K/100 km. On global average, by 100 hPa, radiosonde balloons drift 0.90 h and 48 km from the launch time and location. Therefore, the $SD_{\Delta T}$ introduced at 100 hPa is approximately 0.16 K. This average value does not take into account the variability of drift time and distance associated with local wind conditions.

[54] The dotted curves in Figures 11a and 11b are identical, because they are projected $SD_{\Delta T}$ values for complete coincidence in time and space of radiosonde and COSMIC profiles. These $SD_{\Delta T}$ values can be considered zero-mismatch SD errors, as distinct from the $SD_{\Delta T}$ values associated with nonzero mismatches, solid curves in Figure 10. The zero-mismatch temperature SD error is minimum (0.84 K) around 200 hPa and gradually increases to ~ 2.0 K toward the low troposphere and stratosphere.

[55] For RH (Figure 12), the $SD_{\Delta RH}$ sensitivity to time and distance mismatch is highest at 550 hPa, where for $SD_{\Delta RH}$ it is $5.26 \pm 0.576\%/3$ h and $4.57 \pm 0.568\%/100$ km, respectively. The zero-mismatch SD error is $\sim 14\%$ through-

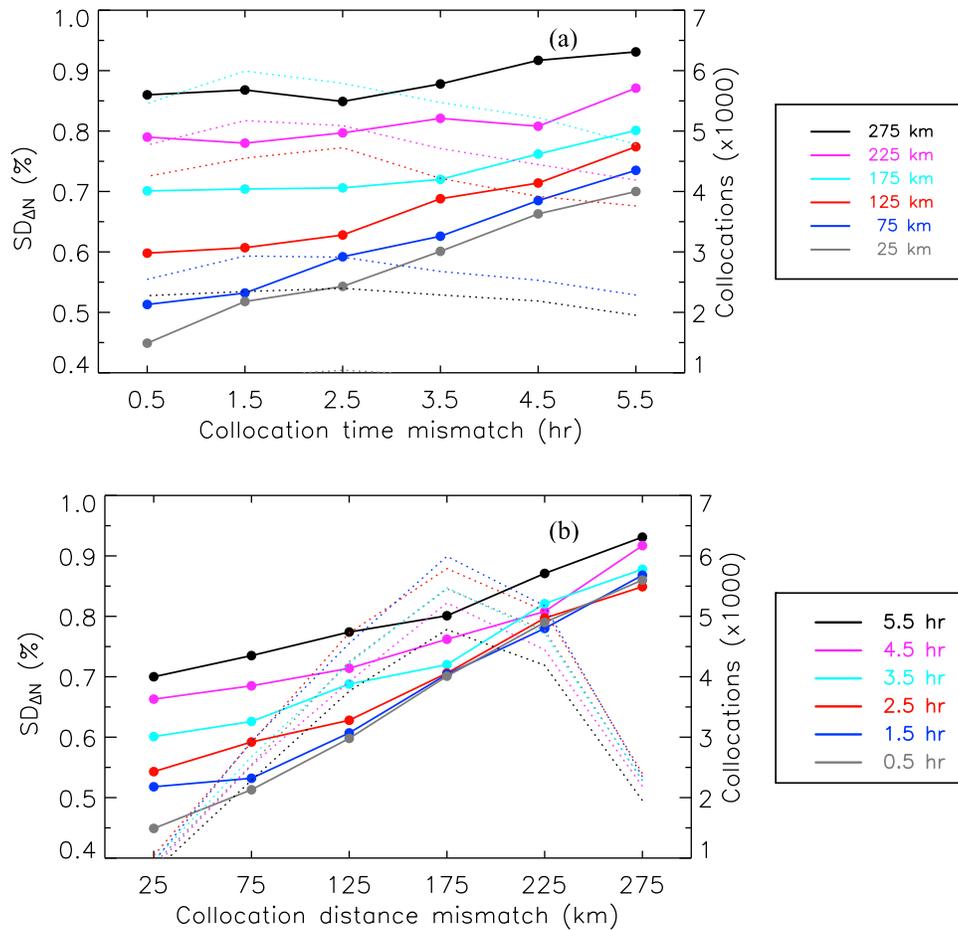


Figure 10. Same as Figure 8 but for standard deviation of raob-minus-COSMIC fractional refractivity difference at 300 hPa.

out the troposphere (dotted curves in Figure 11), but since the upper troposphere is generally relatively drier, the zero-mismatch RH SD error is greater in a relative sense at those levels. For N (Figure 13), the sensitivity of $SD_{\Delta N}$ to time and distance mismatch minimizes around 350 hPa. The zero-mismatch SD error (dotted curves in Figure 13) decreases gradually from 2.28% in the low troposphere to 0.36% from about 300 to 200 hPa.

[56] The zero-mismatch SD errors presented above can be considered a baseline summation of random errors in GPS RO measurement, COSMIC retrieval algorithm, NWP background, or radiosonde data. The increase of zero-mismatch SD N error with height starting at 400 hPa toward the lower troposphere (dashed curves in Figure 13) is similar to the ones obtained through comparing RO N data with ECMWF and NCEP global analyses [see Kuo *et al.*, 2004, Figure 13]. The N SD error increasing trend toward the lower troposphere is likely related to the increasing RO N error over that altitude range [Kursinski *et al.*, 1997; Kuo *et al.*, 2004; and Schreiner *et al.*, 2007]. Measurement and representativeness errors in radiosonde data, which were suggested to be greater than the RO error [Kuo *et al.* [2004], and the variability among sonde types (see section 4.1) are likely the more important factors contributing to the profiles of SD errors (dashed curves of Figures 11–13). One application of

these computations is that similar methods could be applied to other satellite retrievals (e.g., AIRS and IASI) to obtain baseline errors for those observing systems [Maddy and Barnett, 2008; Pougatchev *et al.*, 2009], for comparison.

[57] The mismatch impact results discussed above were computed using global raob-COSMIC collocation data. We performed the analysis for mid-high-latitude and low-latitude regions separately, with results shown in Table 2. In general, the $SD_{\Delta X}$ due to time or distance mismatch is greater in mid-high latitudes than in low latitudes for T, RH, and N, as expected due to atmospheric weather-scale variability [Tobin *et al.*, 2006]. However, for stratospheric $SD_{\Delta T}$, sensitivity to time mismatch is larger at low latitudes than mid-high latitudes (0.47 K/3 h versus 0.27 K/3 h). The standard errors (values within the parentheses) in the low latitudes are greater than those in the mid-high latitudes, reflecting smaller raob-COSMIC collocation samples in low latitudes. Analysis of Figures 11–13 and Table 2 also suggest that globally, in the troposphere, the mismatch impact due to 3 h difference is approximately equivalent to that due to 100 km difference, and this result should be considered in selecting collocation windows for in situ and satellite data comparisons.

[58] In this analysis the SD statistics are computed at profile levels using level data. It is anticipated that values of

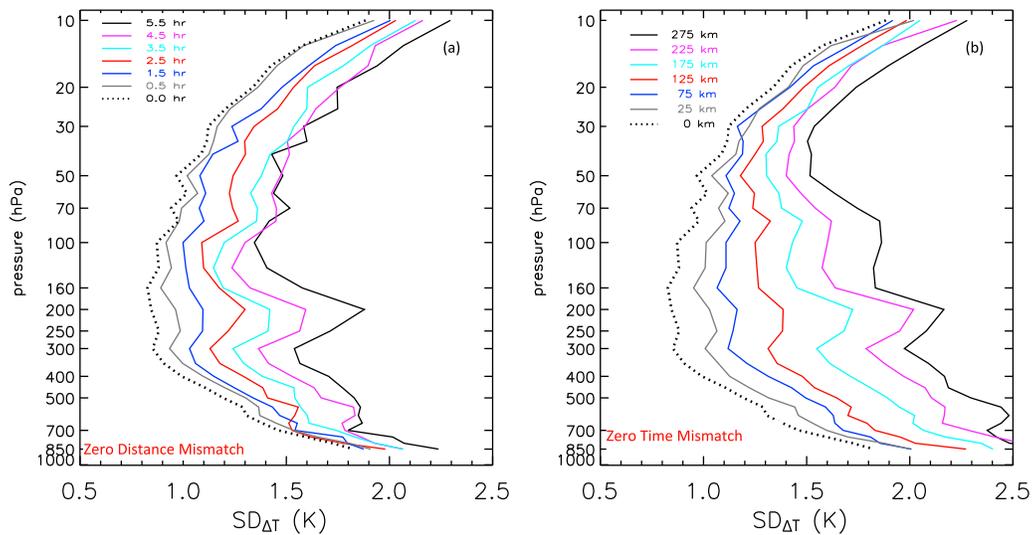


Figure 11. Dependence of standard deviation of raob-minus-COSMIC temperature difference ($SD_{\Delta T}$) on (a) time collocation mismatch and (b) distance collocation mismatch. The $SD_{\Delta T}$ values for the solid curves are the regression intercepts (see section 3.2 for detail), corresponding to zero distance mismatch (Figure 11a) and zero time mismatch (Figure 11b). The $SD_{\Delta T}$ values associated with zero time mismatch (dotted curve in Figure 11a) and zero distance mismatch (dotted curve in Figure 11b) are linearly extrapolated from $SD_{\Delta T}$ values associated with nonzero time mismatches (Figure 11a) and nonzero distance mismatches (Figure 11b), respectively.

these statistics would be smaller if computed from layer-mean data, where fine structure errors are greatly suppressed [Tobin *et al.*, 2006].

5. Discussion

[59] The quantitative results presented in section 4 have several practical applications, which we summarize here. The radiosonde error characteristics revealed in this study through comparison with COSMIC data are generally consistent with results of previous published studies, but they include more radiosonde types and additional variables. Furthermore, the radiosonde humidity error characteristics are supported by comparison with collocated satellite MHS brightness temperature observations. The robustness of these

results recommends their application to (1) efforts that employ raobs to calibrate and validate satellite retrievals and (2) refining the adjustment of radiosonde observations in the NWP data assimilation process.

[60] Our results can be usefully applied to several important aspects of the satellite data validation and calibration process, including: correcting radiosonde measuring biases and selecting specific sonde types and times of day for comparison; refining and quantifying the impact of temporal and spatial collocation criteria; differentiating calibration/validation statistics at different vertical levels; and differentiating among T, RH and N as variables for comparison.

[61] Examples are given here to illustrate the applications mentioned above. Given the measurement biases and their dependence on sonde types (see section 4.1), using radio-

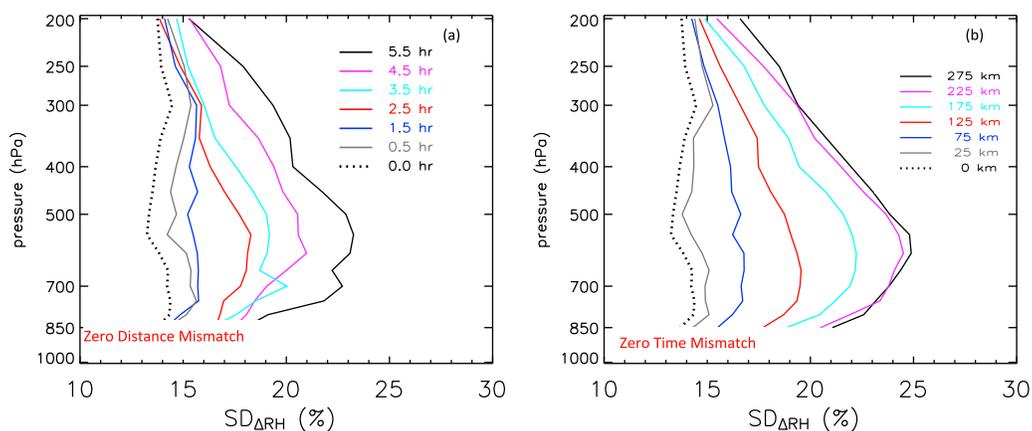


Figure 12. Same as Figure 11 but for standard deviation of raob-minus-COSMIC relative humidity difference ($SD_{\Delta RH}$).

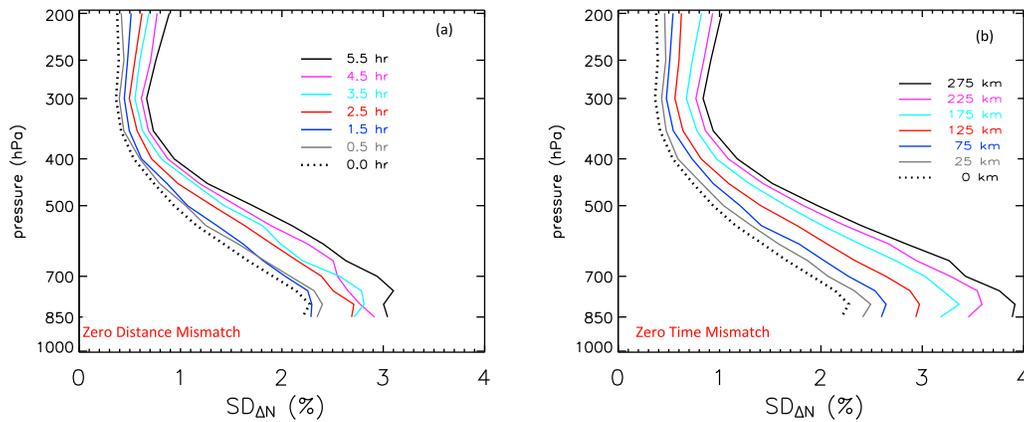


Figure 13. Same as Figure 11 but for standard deviation of raob-minus-COSMIC fractional refractivity difference ($SD_{\Delta N}$).

sonde data without correction could introduce artificial errors into the satellite data evaluated. For instance, satellite retrievals in the upper troposphere would appear to be biased too moist because most of the sonde types have dry biases (Figure 3). Also, the mismatch impact equivalence of 3 h to 100 km (see section 4.1) should be used to pick the right satellite sounding for collocating to a raob, when multiple satellite soundings are within the mismatch window of interest, and the mismatch impact, for example, 0.35 K per 3 h and 0.42 K for 100 km for global tropospheric temperature (see Table 2) is recommended to be taken into account when the weather-monitoring performance of satellite retrievals is evaluated.

[62] As discussed in sections 3 and 4 and *Tobin et al.* [2006], the collocation mismatch impact presented in this study is caused fundamentally by the atmospheric weather-scale variability introduced by the temporal and spatial mismatch. Because short-term weather-scale variability varies with region, the mismatch impact should differ correspondingly, as shown in Table 2 for mid-high latitude versus tropical region. For the same reason, it is speculated the mismatch impact particularly in the lower troposphere may differ with areas which have different terrain types, such as land, sea, and mountain. And the sensitivity of mean biases and SD errors to temporal or spatial mismatch over those areas can be estimated using the methodology described in section 3, but the statistical analysis technique requires samples from a much longer time period to be accumulated for the small spatial-scale analysis, compared to the ones for the global analysis discussed in this work.

[63] Note different sonde types have different levels of measurement noise which could increase the mismatch impact statistics noise. When enough samples are accumulated, it is desirable to repeat the mismatch impact analysis for individual sonde types even though it is not expected that the conclusion drawn from this work on the relation of mismatch impact to temporal and spatial mismatch will change.

[64] We also note that COSMIC retrievals differ from satellite observations based on radiance measurements with respect to geometry and the resulting vertical and horizontal data resolution. We plan to conduct similar analyses with

AIRS and IASI retrievals to see if our results regarding time and distance mismatch impacts are valid for those systems.

[65] It is also important to recognize that the results here are based on “weather-scale” atmospheric variability and may not be directly applicable to the use of satellite measurements for detecting long-term climate change, for which issues of intersatellite calibration, orbital decay and diurnal drift are important [*Christy et al.*, 2000; *Zou et al.*, 2006]. Ideally, high-accuracy radiosonde measurements could be used to calibrate and adjust satellite data for climate monitoring. Whether mismatch between radiosonde launch and satellite overpass affects calibration of the satellite observations, and thus the detection of climate trends, has not been the focus of this analysis. To address that question, we must understand: What is the impact of collocation mismatch on the SD error computed from monthly data (as opposed to the synoptic data used in this analysis)? The answer could be relevant to requirements for sampling strategies for the GCOS Reference Upper-Air Network [*Seidel et al.*, 2009].

[66] With regard to the second application, recall that the raob T data used in this analysis had undergone radiation corrections at the station and/or by NCEP. Nevertheless,

Table 2. Standard Deviation Errors Introduced by Time Mismatch (per 3 h) and Distance Mismatch (per 100 km) for T, RH, and N for the Troposphere (850 to 200 hPa Average) and Stratosphere (200 to 10 hPa)^a

	$SD_{\Delta T}$ (K) Troposphere	$SD_{\Delta T}$ (K) Stratosphere	$SD_{\Delta RH}$ (%) Troposphere	$SD_{\Delta N}$ (%) Troposphere
<i>Time Mismatch Impact</i>				
Globe	0.35 (0.042)	0.30 (0.042)	3.30 (0.506)	0.33 (0.050)
Mid-high latitude	0.40 (0.049)	0.27 (0.053)	3.52 (0.551)	0.35 (0.049)
Low latitude	0.11 (0.121)	0.47 (0.139)	2.32 (0.929)	0.24 (0.105)
<i>Distance Mismatch Impact</i>				
Globe	0.42 (0.030)	0.22 (0.025)	3.05 (0.290)	0.36 (0.023)
Mid-high latitude	0.46 (0.031)	0.22 (0.025)	3.19 (0.298)	0.35 (0.021)
Low latitude	0.20 (0.048)	0.22 (0.052)	2.58 (0.461)	0.32 (0.050)

^aValues within the parentheses are the standard errors of the estimates. The low-latitude region is defined as 30°N and 30°S, and the mid-high-latitude region is the rest of the world.

there remained differential biases with respect to COSMIC T profiles, which suggests that those corrections are not wholly adequate. The NCEP corrections are for radiation error only, and they have not been updated since 1999. Therefore, it is likely that newer instruments, such as Vaisala RS92 and Sippican radiosondes, require different adjustments, as do any instruments for which instrument type codes are incorrectly reported. Results of this work prompt us to more thoroughly examine radiation biases in NCEP adjusted and unadjusted radiosonde data, by using the GPSRO data as globally consistent transfer standard. Such a study should help improve or update NCEP corrections to radiosonde T profiles, as well as extend them to humidity observations.

6. Summary

[67] Using 19 months of NPROVS global raob-COSMIC collocation data, this study has identified the characteristics of major radiosonde types, using the COSMIC data as a transfer standard, and has quantified the impact of time and distance collocation mismatch on the accuracy assessment of satellite retrievals of temperature, relative humidity and refractivity. Our main results are as follows.

[68] 1. Although on global average, raob and COSMIC-derived temperature profiles agree within 0.15 K, there are differences among sonde types, which vary with height and from day to night. Average differences tend to be < 0.5 K for most types at most levels, with standard deviations of 1.2–2 K.

[69] 2. Most of the radiosonde types show a dry bias, particularly in the upper troposphere where the bias reaches 5–8% in RH, and this bias is found in comparison both with COSMIC and with independent satellite moisture-sensitive radiance measurements. For most radiosonde types, the daytime RH bias is greater than for nighttime.

[70] 3. Most sonde types have refractivity biases < 0.5% in the low troposphere and < 0.2% in the upper troposphere.

[71] 4. Weather-scale variability, introduced by collocation time and distance mismatch, affects the comparison of radiosonde and COSMIC data by increasing the standard deviation errors. The errors were found to be generally proportional to the size of time and distance mismatch.

[72] 5. The T standard deviation errors are most sensitive to collocation mismatch in the upper troposphere, and RH standard deviation errors are most sensitive to collocation mismatch at 550 hPa. Errors in N are more sensitive to T errors in the upper troposphere and to moisture errors in the lower troposphere. In the troposphere, collocation mismatch impacts are greater in the mid-high latitudes than in low latitudes.

[73] 6. Globally, in the troposphere (850–200 hPa), the collocation mismatch impacts on the comparison standard deviation errors for temperature are 0.35 K per 3 h and 0.42 K per 100 km, and for relative humidity are 3.3% per 3 h and 3.1% per 100 km, indicating an approximate equivalence of 3 h to 100 km in terms of mismatch impact.

[74] **Acknowledgments.** The authors are grateful to Steve Schroeder (Texas A&M University), Bradley Ballish and Dennis Keyser (NOAA NCEP), and Carl Bower (NOAA NWS) for discussions on radiosonde data and instrument characteristics. Melissa Free and Christoph A. Vogel (NOAA ARL) provided valuable comments on the revision of this paper.

Three anonymous reviewers suggested important revision to the original manuscript. Thanks also go to Michael Pettey and Frank Tilley (I.M. Systems Group) for their technical support. This work was funded by the NOAA Integrated Program Office (IPO) for Cross-Track Infrared Microwave Sounder Suite (CrIMSS) Environmental Data Record (EDR) calibration and validation (Cal/Val) in support of the Joint Polar Satellite System (JPSS).

References

- Adam, M., et al. (2010), Water vapor measurements by Howard Univ. Raman lidar during the WAVES 2006 Campaign, *J. Atmos. Oceanic Technol.*, 27, 42–60, doi:10.1175/2009JTECHA1331.1.
- Anthes, R. A., et al. (2008), The COSMIC/FORMOSAT-3 Mission: Early results, *Bull. Am. Meteorol. Soc.*, 89, 313–333, doi:10.1175/BAMS-89-3-313.
- Ao, C. O. (2007), Effect of ducting on radio occultation measurement: An assessment based on high-resolution radiosonde soundings, *Radio Sci.*, 42, RS2008, doi:10.1029/2006RS003485.
- Christy, J. R., and W. B. Norris (2009), Discontinuity issues with radiosonde and satellite temperatures in the Australian region 1979–2006, *J. Atmos. Oceanic Technol.*, 26, 508–522, doi:10.1175/2008JTECHA1126.1.
- Christy, J. R., R. W. Spencer, and W. D. Braswell (2000), MSU tropospheric temperatures: Dataset construction and radiosonde comparisons, *J. Atmos. Oceanic Technol.*, 17, 1153–1170.
- Divakarla, M. G., C. D. Barnet, M. D. Goldberg, L. M. McMillin, E. Maddy, W. Wolf, L. Zhou, and X. Liu (2006), Validation of Atmospheric Infrared Sounder temperature and water vapor retrievals with matched radiosonde measurements and forecasts, *J. Geophys. Res.*, 111, D09S15, doi:10.1029/2005JD006116.
- Gaffen, D. J. (1994), Temporal inhomogeneities in radiosonde temperature records, *J. Geophys. Res.*, 99(D2), 3667–3676, doi:10.1029/93JD03179.
- Han, Y., F. P. van Delst, Q. Liu, F. Weng, B. Yan, R. Treadon, and J. Derber (2006), JCSDA Community Radiative Transfer Model (CRTM)–Version 1, *NOAA Tech Rep.* 122, NOAA, Silver Spring, Md.
- 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, doi:10.1029/2009GL038712.
- Ho, S.-P., Y.-H. Kuo, Z. Zeng, and T. C. Peterson (2007), A comparison of lower stratosphere temperature from microwave measurements with CHAMP GPS RO data, *Geophys. Res. Lett.*, 34, L15701, doi:10.1029/2007GL030202.
- Kuo, Y.-H., T.-K. Wee, S. Sokolovskiy, C. Rocken, W. Schreiner, D. Hunt, and R. A. Anthes (2004), Inversion and error estimation of GPS radio occultation data, *J. Meteorol. Soc. Jpn.*, 82, 507–531, doi:10.2151/jmsj.2004.507.
- 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, doi:10.1029/2004GL021443.
- Kursinski, E. R., and G. A. Hajj (2001), A comparison of water vapor derived from GPS occultations and global weather analyses, *J. Geophys. Res.*, 106(D1), 1113–1138, doi:10.1029/2000JD900421.
- 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(D19), 23,429–23,465, doi:10.1029/97JD01569.
- Lanzante, J. R., S. A. Klein, and D. J. Seidel (2003), Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology, *J. Clim.*, 16, 224–240, doi:10.1175/1520-0442(2003)016<0224:THOMRT>2.0.CO;2.
- Luers, J. K., and R. E. Eskridge (1998), Use of radiosonde temperature data in climate studies, *J. Clim.*, 11, 1002–1019, doi:10.1175/1520-0442(1998)011<1002:UORTDI>2.0.CO;2.
- Maddy, E. S., and C. D. Barnet (2008), Vertical resolution estimates in Version 5 of AIRS operational retrievals, *IEEE Trans. Geosci. Remote Sens.*, 46, 2375–2384, doi:10.1109/TGRS.2008.917498.
- McMillin, L. M., M. E. Gelman, A. Sanyal, and M. Sylva (1988), A method for the use of satellite retrievals as a transfer standard to determine systematic radiosonde errors, *Mon. Weather Rev.*, 116, 1091–1102, doi:10.1175/1520-0493(1988)116<1091:AMFTUO>2.0.CO;2.
- McMillin, L. M., J. Zhao, M. K. Rama Varma Raja, S. I. Gutman, and J. G. Yoe (2007), Radiosonde humidity corrections and potential Atmospheric Infrared Sounder moisture accuracy, *J. Geophys. Res.*, 112, D13S90, doi:10.1029/2005JD006109.
- Miloshevich, L. M., H. Vömel, D. N. Whiteman, B. M. Lesht, F. J. Schmidlin, and F. Russo (2006), Absolute accuracy of water vapor measurements from six operational radiosonde types launched during

- AWEX-G and implications for AIRS validation, *J. Geophys. Res.*, *111*, D09S10, doi:10.1029/2005JD006083.
- Nakamura, H., H. Seko, and Y. Shoji, and the Aerological Observatory, and Meteorological Instrument Center (2004), Dry biases of humidity measurements from the Vaisala RS80-A and Meisei RS2-91 radiosondes and from ground-based GPS, *J. Meteorol. Soc. Jpn.*, *82*(1B), 277–299, doi:10.2151/jmsj.2004.277.
- Nash, J., R. Smout, T. Oakley, B. Pathack, and S. Kurnosenko (2006), Intercomparison of radiosonde systems, Vacoas, Mauritius, 2–25 February 2005, *WMO/TD 1303*, 115 pp., WMO, Geneva.
- Neiman, P. J., F. M. Ralph, G. A. Wick, Y.-H. Kuo, T.-K. Wee, Z. Ma, G. H. Taylor, and M. D. Dettinger (2008), Diagnosis of an intense atmospheric river impacting the Pacific northwest: Storm summary and off-shore vertical structure observed with COSMIC satellite retrievals, *Mon. Weather Rev.*, *136*, 4398–4420, doi:10.1175/2008MWR2550.1.
- Pougatchev, N. (2008), Validation of atmospheric sounders by correlative measurements, *Appl. Opt.*, *47*, 4739–4748.
- Pougatchev, N., T. August, X. Calbet, T. Hultberg, O. Oduleye, P. Schlüssel, B. Stiller, K. S. Germain, and G. Bingham (2009), IASI temperature and water vapor retrievals—error assessment and validation, *Atmos. Chem. Phys. Discuss.*, *9*, 7971–7989, doi:10.5194/acpd-9-7971-2009.
- Reale, A., F. Tilley, M. Ferguson, and A. Allegrino (2008), NOAA operational sounding products for ATOVS, *Int. J. Remote Sens.*, *29*, 4615–4651, doi:10.1080/01431160802020502.
- Reale, T., B. Sun, M. Pettey, and F. Tilley (2010), The NOAA products validation system (NPROVS) and archive summary system (NARCSS), paper presented at 6th Annual Symposium on Future Operational Environmental Satellite Systems-NPOESS and GOES-R, Am. Meteorol. Soc., Atlanta, Ga.
- Schreiner, W., C. Rocken, S. Sokolovskiy, S. Syndergaard, and D. Hunt (2007), Estimates of the precision of GPS radio occultations from the COSMIC/FORMOSAT-3 mission, *Geophys. Res. Lett.*, *34*, L04808, doi:10.1029/2006GL027557.
- Seidel, D. J., et al. (2009), Reference upper-air observations for climate: Rationale, progress, and plans, *Bull. Am. Meteorol. Soc.*, *90*, 361–369, doi:10.1175/2008BAMS2540.
- Smith, E. K., and S. Weintraub (1953), The constants in the equation for atmospheric refractive index at radio frequencies, *J. Res. Natl. Bur. Stand. U. S.*, *50*, 39–41.
- Soden, B. J., and J. R. Lanzante (1996), An assessment of satellite and radiosonde climatologies of upper tropospheric water vapor, *J. Clim.*, *9*, 1235–1250, doi:10.1175/1520-0442(1996)009<1235:AAOSAR>2.0.CO;2.
- Sokolovskiy, S., Y.-H. Kuo, C. Rocken, W. S. Schreiner, D. Hunt, and R. A. Anthes (2006), Monitoring the atmospheric boundary layer by GPS radio occultation signals recorded in the open-loop mode, *Geophys. Res. Lett.*, *33*, L12813, doi:10.1029/2006GL025955.
- Tobin, D. C., H. E. Revercomb, R. O. Knuteson, B. M. Lesht, L. L. Strow, S. E. Hannon, W. F. Feltz, L. A. Moy, E. J. Fetzer, and T. S. Cress (2006), Atmospheric radiation measurement site atmospheric state best estimates for Atmospheric Infrared Sounder temperature and water vapor validation, *J. Geophys. Res.*, *111*, D09S14, doi:10.1029/2005JD006103.
- Vömel, H., D. E. David, and K. Smith (2007), Accuracy of tropospheric and stratospheric water vapor measurements by a cryogenic frost point hygrometer: Instrument details and observations, *J. Geophys. Res.*, *112*, D08305, doi:10.1029/2006JD007224.
- Wang, J., and L. Zhang (2008), Systematic errors in global radiosonde precipitable water data from comparisons with ground-based GPS measurements, *J. Clim.*, *21*, 2218–2238, doi:10.1175/2007JCLI1944.1.
- Wang, J., D. J. Carlson, D. B. Parsons, T. F. Hock, D. Lauritsen, H. L. Cole, K. Beierle, and N. Chamberlain (2003), Performance of operational radiosonde humidity sensors in direct comparison with a chilled mirror dew-point hygrometer and its climate implication, *Geophys. Res. Lett.*, *30*(16), 1860, doi:10.1029/2003GL016985.
- Yunck, T. P., E. J. Fetzer, A. M. Mannucci, C. O. Ao, F. W. Irion, B. D. Wilson, and G. J. M. Manion (2009), Use of radio occultation to evaluate atmospheric temperature data from spaceborne infrared sensors, *Terr. Atmos. Oceanic Sci.*, *20*, 71–85, doi:10.3319/TAO.2007.12.08.01(F3C).
- Zou, C.-Z., M. D. Goldberg, Z. Cheng, N. C. Grody, J. T. Sullivan, C. Cao, and D. Tarpley (2006), Recalibration of microwave sounding unit for climate studies using simultaneous nadir overpasses, *J. Geophys. Res.*, *111*, D19114, doi:10.1029/2005JD006798.

D. C. Hunt, University Corporation for Atmospheric Research, Boulder, CO 80307, USA.

A. Reale, Center for Satellite Applications and Research, NESDIS, NOAA, Camp Springs, MD 20746, USA.

D. J. Seidel, Air Resources Laboratory, NOAA, Silver Spring, MD 20910, USA.

B. Sun, I.M. Systems Group, 3206 Tower Oaks Blvd., Ste. 300, Rockville, MD 20852, USA. (bomin.sun@noaa.gov)