On the Performance of Airborne Meteorological Observations against Other In Situ Measurements

TIMOTHY J. WAGNER[a] AND RALPH A. PETERSEN[a]

[a] *Cooperative Institute for Meteorological Satellite Studies, Space Science and Engineering Center, University of Wisconsin–Madison, Madison, Wisconsin*

(Manuscript received 10 November 2020, in final form 19 March 2021)

ABSTRACT: Routine in situ observations of the atmosphere taken in flight by commercial aircraft provide atmospheric profiles with greater temporal density and, in many parts of the country, at more locations than the operational radiosonde network. Thousands of daily temperature and wind observations are provided by largely complementary systems, the Airborne Meteorological Data Relay (AMDAR) and the Tropospheric Airborne Meteorological Data Reporting (TAMDAR). All TAMDAR aircraft also measure relative humidity while a subset of AMDAR aircraft are equipped with the Water Vapor Sensing System (WVSS) measure specific humidity. One year of AMDAR/WVSS and TAMDAR observations are evaluated against operational National Weather Service (NWS) radiosondes to characterize the performance of these systems in similar environments. For all observed variables, AMDAR reports showed both smaller average differences and less random differences with respect to radiosondes than the corresponding TAMDAR observations. Observed differences were not necessarily consistent with known radiosonde biases. Since the systems measure different humidity variables, moisture is evaluated in both specific and relative humidity using both aircraft and radiosonde temperatures to derive corresponding moisture variables. Derived moisture performance is improved when aircraft-based temperatures are corrected prior to conversion. AMDAR observations also show greater consistency between different aircraft than TAMDAR observations do. The small variability in coincident WVSS humidity observations indicates that they may prove more reliable than humidity observations from NWS radiosondes.

KEYWORDS: Aircraft observations; In situ atmospheric observations; Radiosonde/rawinsonde observations

1. Introduction

Upper-air observations are critical for many aspects of operational meteorology. Forecasters, for example, rely on atmospheric profiles to assess atmospheric stability or the potential of hazardous precipitation types, while analyses of data assimilation (D/A) systems used in numerical weather prediction (NWP) models show that in situ profiles continue to play a prominent role in facilitating accurate forecast simulations (Eyre and Reid 2014). Since their operational introduction in the 1930s (DuBois et al. 2002), radiosondes (including rawinsondes) have provided the standard by which other vertical profiling systems, both in situ and remotely sensed, have been judged. Operational radiosonde networks, however, have a number of well-known limitations. Chief among these is the lack of temporal and spatial density. For example, even over the densely observed contiguous United States (CONUS), outside of infrequent significant weather events radiosondes are seldom launched outside the internationally agreed upon synoptic times of 0000 and 1200 UTC. These hours are not necessarily well aligned with local forecaster needs. Over North and South America, the 0000 UTC observations occur well after daytime convection usually initiates and offer little prognostic insight into storm strength or mode, while the 1200 UTC observation mostly misses the effects of solar heating that can subsequently remove nocturnal inversions. Furthermore, even over the CONUS, the approximately 400 km spatial density of the network is too coarse to directly resolve many subsynoptic-scale phenomena, especially those related to moisture. Satellites can only partially fill these gaps, as the vertical resolution in the planetary boundary layer is still too coarse to resolve many important characteristics and continuous profiling over CONUS with hyperspectral sounders in geostationary orbit remains years away. If alternative types of profile observation were available to augment the operational radiosonde network, significant additional understanding of the atmosphere can be gained.

Over the past several decades, commercial aircraft have become a significant new source of in situ profiles of the troposphere. By either providing existing routine observations of temperature and winds that are necessary for safe and efficient operations, or by outfitting aircraft with special sensors to collect these or additional parameters, automated aircraft-based observations (ABOs) are now routinely obtained and distributed to national weather services and other stakeholders with high accuracy and low latency. These observations have significant positive impact on numerical weather prediction globally and can exceed radiosondes in importance in areas of highest reporting density (Petersen 2016; James and Benjamin 2017; James et al. 2020).

The Airborne Meteorological Data Relay (AMDAR; Painting 2003) program was established by the World Meteorological Organization (WMO) to collect and to distribute the full set of global ABOs (Moninger et al. 2003). AMDAR is a voluntary program that has varying levels of participation globally. Most major airlines in the United States and Europe contribute to the program and provide high spatial and temporal density of coverage over the populated areas of those regions. Coverage is less dense over the oceans (where observations, where they

Corresponding author: Timothy J. Wagner, tim.wagner@ssec.wisc.edu

DOI: 10.1175/JTECH-D-20-0182.1

exist, are concentrated in narrow layers at cruising altitudes) but increasing in portions of the Southern Hemisphere [WMO Integrated Global Observing System (WIGOS); WIGOS 2014]. AMDAR is sometimes referred to as Aircraft Communication Addressing and Reporting System (ACARS), which is the name of one of the communications protocols that transmits the AMDAR observations to the surface. A portion of the costs associated with transmitting, stripping out proprietary information, and distributing the data is borne by some governments (including in the United States), but the per-profile cost is substantially less than that of radiosondes. The data are made available in real time to global NWP centers via standard WMO communications protocols. For nonoperational interests and the general public, archived AMDAR data are generally available 48 h after collection.

Pressure, temperature, and wind data collected through the AMDAR program have immediate operational utility to the pilots and ground-based support staff while the airplanes are in flight. Some AMDAR aircraft also provide a derived turbulence parameter. Although a number of studies have shown that assimilation of in situ ABOs of water vapor can have significant benefits for the airlines through the improved numerical weather forecasts of events that disrupt normal airline operations (Petersen et al. 2016), the lack of instrumentation to make moisture observations as part of standard equipment on all aircraft makes routine observation more difficult. As means of obtaining humidity information, a subset of AMDAR aircraft has been outfitted using the Water Vapor Sensing System II (WVSS), a laser diode that is capable of obtaining the specific humidity (hereafter denoted SH) by directly counting water vapor molecules. Petersen et al. (2016) contains more details including observation and reporting strategies. In the United States, approximately 120 aircraft from Southwest Airlines and United Parcel Service (UPS) provide observations from these sensors. A rigorous evaluation of WVSS performance against operational radiosondes across the full United States presented by Williams et al. (2021) showed excellent agreement between the two systems. NOAA supports the WVSS program directly and all meteorological reports from aircraft flying these sensors are freely available to the public in real time.

The private sector has recognized that the existing AMDAR network, with its concentration of observations near the tropopause and near major airline hubs, could be augmented with an alternate observing system that focused on different aircraft types. One such system, the Tropospheric Airborne Meteorological Data Reporting (TAMDAR; Daniels et al. 2006) observation and communications system was developed through a public–private partnership between NASA, NOAA, the Federal Aviation Administration (FAA), and a private corporation. While AMDAR observations tend to come from aircraft carrying out longer-haul flights (and thus experiencing between four and eight ascents and descents per day), TAMDAR systems were designed to be installed on shorter-haul regional passenger aircraft (including both propeller-driven and smaller jet planes) that required additional onboard instrumentation and communications capabilities commonly available on larger AMDAR aircraft. TAMDAR aircraft provide coverage to smaller airports and have the potential for providing more soundings daily via shorter, more frequent flights. As a private enterprise, however, TAMDAR data are not generally available for free and open access by all NWP centers. Instead, weather agencies must purchase access to the data.

TAMDAR reports include a superset of the AMDAR parameters. In addition to temperature and winds, TAMDAR reports also provide relative humidity (hereafter denoted RH), icing and turbulence. Since the majority of TAMDAR-equipped aircraft fly short distances between regional airports, they typically do not reach the same altitudes as the long-haul AMDAR aircraft. Daniels et al. (2004) indicated that observations were originally made every 10 hPa for the first 100 hPa, then every 50 hPa thereafter, which was upgraded later (Jacobs et al. 2006) to 10 hPa intervals throughout the lowest 200 hPa of the atmosphere and every 25 hPa above that during ascent. After the ascent is completed, observations are made every 3 min expect during the last 200 hPa of descent, when the 10 hPa interval returns.

This paper provides the first long-term joint assessment of AMDAR and TAMDAR observations against collocated radiosondes. AMDAR temperature and wind observations have previously been compared to both model output (e.g., Moninger et al. 2010), radiosondes (e.g., Schwartz and Benjamin 1995; Petersen et al. 2016), or other aircraft (e.g., Benjamin et al. 1999; Drüe et al. 2008; Petersen 2016). AMDAR WVSS measurements have been evaluated against radiosondes (Petersen et al. 2016; Williams et al. 2021) and within NWP systems (Hoover et al. 2017). Zhang et al. (2018) provides a table outlining these previous studies and others in their multiyear study of AMDAR performance against radiosondes at selected locations. Although fewer studies of the quality of TAMDAR observations are available, TAMDAR reports have been evaluated against both models (Daniels et al. 2006; Moninger et al. 2010; Gao et al. 2012; James et al. 2020) and radiosondes (Gao et al. 2012). The present study serves two purposes: first, to expand previous observation validation studies mentioned above by evaluating system performance over a large portion of CONUS for an entire year; and second, to evaluate which of the ABO systems performs better when in the same environment. Moisture observations are examined in special detail, as relative humidity is a key parameter for data assimilation but has received little attention in previous ABO studies.

2. Methodology

This study was conducted using observations made during the full calendar year 2018. The first step in evaluating the quality of the ABOs was to compare them to collocated radiosonde observations, which for this work were considered to provide a true and unbiased assessment standard. The implications and appropriateness of such an assumption are discussed later in this paper. The high-resolution Radiosonde Replacement System (RRS) archive of 1 s data from the National Centers for Environmental Information (NCEI) was used as was chosen as the primary radiosonde dataset for

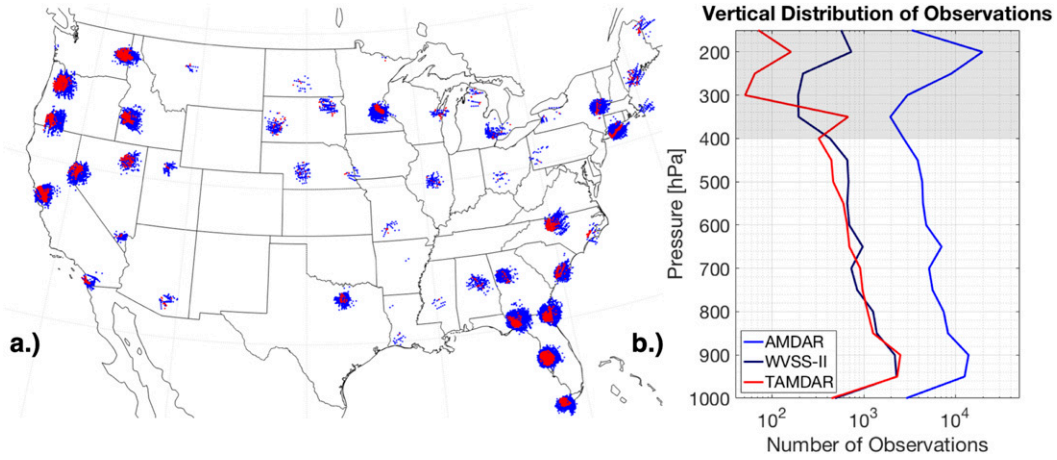

FIG. 1. (a) Map of the locations of radiosonde observations with AMDAR/WVSS observations (blue) and TAMDAR observations (red) between the surface and 400 hPa and (b) vertical profile of the number of observations collocated with radiosondes for the study period from AMDAR (blue), WVSS (dark blue), and TAMDAR (red) per 50 hPa layer. Only observations from below 400 hPa were used in this study; observations from higher altitudes are included in (b) to illustrate the peak in AMDAR observations at cruising altitude but are shaded to denote their elimination from further use in this analysis.

this study because it contains the latitude and longitude data necessary to geolocate all observation throughout a balloon's ascent (NCDC 2008). Other NCEI archives of radiosonde profiles like the Integrated Global Radiosonde Archive (IGRA) only retain the latitude and longitude of the launch site. Unfortunately, not all radiosonde sites were available from the high-resolution RRS archive for the period of analysis (including, among others, the profiles from Norman, Oklahoma; and Dulles, Virginia). As a result, some spatial gaps exist where the aircraft observations could not be validated. Finally, while AMDAR observations cover almost all of CONUS, the TAMDAR observations available during the study period were concentrated primarily near the Atlantic Ocean and in the western United States. To prevent the unequal distribution of TAMDAR observations from biasing the analysis, only radiosonde observations that matched at least both one AMDAR and TAMDAR data record were used in the study, as seen in Fig. 1a.

All AMDAR and TAMDAR data were obtained from the Meteorological Assimilation Data Ingest System (MADIS) archive, which is the WMO global data center for ABOs. In addition to the meteorological data values, every observation includes altitude, latitude, longitude, and origin and destination airports. A number of criteria were applied to ensure that the ABO and radiosonde observations were well matched, including a spatial and temporal limit of 50 km and 30 min, respectively. To account for radiosonde drift, these criteria were not limited to just the launch time and position but instead moved with the radiosonde throughout its ascent. Vertical matching required two separate procedures depending on the instrument type as the archived AMDAR and WVSS observations were recorded in pressure altitude while the TAMDAR observations used GPS altitude. For the AMDAR observations, the pressure altitudes were converted to pressure using the standard atmosphere and matched to the closest radiosonde pressure. For the TAMDAR observations, the GPS altitude was compared to the radiosonde altitudes and the associated pressure from the radiosonde was assigned to that TAMDAR observation. This permitted all comparisons to be carried out in pressure space. Statistics were then calculated over 50-hPa-deep layers from Earth's surface to 400 hPa to provide statistical information from the regions where we had similar reporting frequencies; very few TAMDAR observations are found at altitudes higher than 400 hPa (Fig. 1b).

Automated quality control flags determined by MADIS contained within the datasets were also applied. These tests include assuring that 1) each observation has a value that is reasonable (e.g., temperature within a certain range given its altitude), 2) different parameters observed at the same time are physically consistent (e.g., ensuring that the dewpoint temperature does not exceed the air temperature), and 3) the rate of change in observations from one time reporting time to the next are within a range that would be expected for an airplane in its current phase of flight (ascent, cruise, or descent). All observations that failed any of these tests were excluded from the analysis. Despite these automated checks, some outliers remained clearly present in the MADIS datasets, especially in the TAMDAR records. Therefore, an additional quality control measure was implemented independently that rejected any ABO report that differed from its radiosonde counterpart by more than three standard deviations (3σ) away from the mean for each individual computed difference, under the assumption that such obviously erroneous observations would also be rejected by NWP data assimilation systems. Theoretically, over 99% of the observations should have been retained using this approach if the differences were Gaussian. The rejection of one

TABLE 1. The mean and standard deviation of the airborne observation minus radiosonde differences before and after a three standard deviation (3σ) gross quality check is applied. AMDAR values are shown in boldface while TAMDAR values are in regular type. Humidity values are calculated using aircraft-observed temperatures for parameters not measured directly. The percentage of data removed by this check is shown in the rightmost column. The parenthetical values in italics represent the percentage of TAMDAR observations that would have been rejected if subjected to the 3σ values of the AMDAR observations.

	Before 3σ check		After 3σ check		
	Mean	Std dev	Mean	Std dev	Percent rejected
Wind speed ($m\,s^{-1}$)	**−0.36**	**1.97**	**−0.38**	**1.77**	**1.3**
	1.15	6.44	0.90	3.37	0.8 (*7.8*)
Wind direction (°)	**1.76**	**29.83**	**1.51**	**22.60**	**2.8**
	5.94	46.95	4.50	39.54	2.7 (*8.2*)
Vector difference ($m\,s^{-1}$)	**2.32**	**1.64**	**2.21**	**1.36**	**1.6**
	4.48	7.14	3.97	3.16	1.0 (*7.0*)
Temperature (K)	**−0.16**	**0.99**	**−0.20**	**0.80**	**1.8**
	−0.25	5.35	0.25	2.19	2.1 (*9.7*)
RH (%)	**3.02**	**13.13**	**3.38**	**10.16**	**2.2**
	4.45	12.41	4.56	10.58	1.7 (*1.4*)
SH ($g\,kg^{-1}$)	**0.29**	**1.28**	**0.31**	**0.99**	**1.6**
	0.61	2.76	0.53	1.58	2.4 (*6.1*)

observation of one parameter at a particular level was not grounds for rejecting other parameters observed by the same aircraft at the same time nor the same parameter at other levels; each individual observation was evaluated independently. The AMDAR standard deviations were generally much smaller than the TAMDAR ones, so the 3σ filter permitted a much wider range of TAMDAR values than it did for AMDAR. For example, TAMDAR temperature reports with an error greater than 16.06 K were flagged for removal compared to only 2.97 K for AMDAR temperature reports. After applying this gross error check, the range of acceptable AMDAR–radiosonde differences changed only slightly (to 2.39 K) while that for TAMDAR decreased more dramatically (to 6.58 K). The averages and standard deviations of the observation-minus-radiosonde differences for the analyzed variables, both before and after the 3σ error check is applied, are recorded in Table 1, along with the percentage of TAMDAR observations that would have been rejected if they were subject to the stricter AMDAR limits. For example, the rejection rate for TAMDAR temperature observations would have increased substantially from approximately 2% to nearly 10%.

Figure 1a shows the locations of the AMDAR and TAMDAR observations below the 400 hPa level that met the matching requirements and were used in this analysis. Note that some radiosonde sites were missing from the archive while many ABOs made during ascent and descent at major airport hubs (like Chicago, Illinois) are too far away from radiosondes to be captured in this study. An additional requirement was included to assure that the AMDAR and TAMDAR datasets were evaluated under similar atmospheric conditions. To do this, every AMDAR/radiosonde comparison had to have at least one contemporaneous TAMDAR/radiosonde comparison, and vice versa. As a result, several locations with hundreds of satisfactory AMDAR–radiosonde matches but no TAMDAR observations (e.g., Denver, Colorado) were excluded from the study.

Figure 1b contrasts the vertical distribution of the observations used in this study. Although AMDAR shows a strong peak in observations near 200 hPa cruise levels, very few TAMDAR observations in this study are found at altitudes higher than the 400 hPa level; that is why this pressure level was chosen as the upper boundary for this study. Note that this study contains many more AMDAR observations than TAMDAR ones; if the requirement for observed environments to contain both a TAMDAR and an AMDAR observation were removed, this discrepancy would be substantially larger.

3. Aircraft to radiosonde intercomparisons

After the aircraft data were quality controlled and properly matched to corresponding radiosonde observations, the differences between the ABOs and the radiosondes were calculated. The mean and standard deviation of the difference of the ABO reports minus the collocated radiosonde observation were the primary means of assessing the performance of the two ABO systems, with the mean representing the systematic difference between the observing systems while the standard deviation represents the random differences. Proper characterization of both types of the differences is needed to maximize their utility of both systems in NWP data assimilation systems. Although statistical techniques can be used to moderate the effects of the systematic differences in individual profiles, random differences cannot be reduced as easily.

Results for each variable are presented in a variety of ways: first, as histograms of the magnitude of differences between the ABO and radiosonde observations for all levels; second, as vertical profiles of the differences; and third, as differences as a function of the magnitude of the parameter being observed. Unlike Williams et al. (2021), where use of a sigma display coordinate was essential in assuring that boundary layer information was synthesized properly between radiosonde sites with widely varying surface pressures, previously discussed data limitations constrained this study to locations with more homogeneous elevations, resulting in very small variations between the isobaric and sigma displays for all parameters

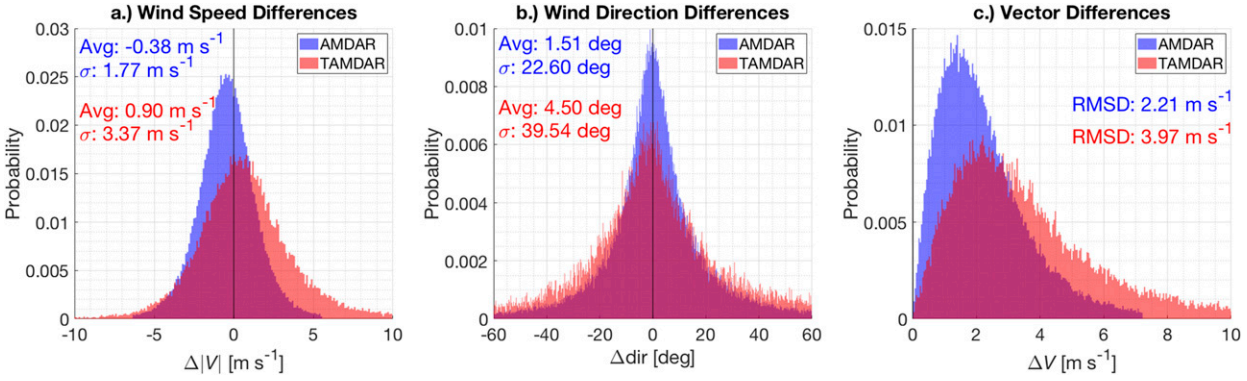

FIG. 2. (a) Histograms of the airborne minus radiosonde differences for AMDAR (blue) and TAMDAR (red) for wind speed (in m s^{-1}) and (b) wind direction (in degrees) from the surface through 400 hPa. The bin sizes are 0.1 m s^{-1} and 0.25°, respectively. (c) The histogram of the vector difference between the two observing systems and the corresponding radiosonde observations is also shown (in m s^{-1}) in 0.04 m s^{-1} bins.

studied. As such, only isobaric results are presented here. Parameters from the AMDAR and TAMDAR systems that can be compared directly are discussed first, followed by an assessment of different forms of humidity reported by the two systems.

a. Winds

Figure 2 displays differences between two aircraft observations methods and collocated radiosondes for wind speed (Fig. 2a) and direction (Fig. 2b), as well as the vector difference (Fig. 2c). For all parameters, AMDAR wind observations (usually obtained from larger aircraft) outperform TAMDAR reports collected from smaller aircraft by nearly a factor of 2. The mean wind speed differences are of opposite signs, with AMDAR having an average difference of −0.38 m s^{-1} and TAMDAR having an average difference of +0.90 m s^{-1} while the wind direction shows positive average differences of 1.51° and 4.50°, respectively. Random differences for speed and direction (measured as standard deviations σ) show large discrepancies between the two ABO systems. AMDAR observations appear nearly twice as consistent with radiosonde reports as TAMDAR, with standard deviations of 1.77 m s^{-1} (22.60°) and 3.37 m s^{-1} (39.54°), respectively, for wind speed (direction). As a result, the root-mean-squared (RMS) vector differences (RMSD), which combines the effects of both wind speed and directions differences) between the ABO reports and collocated radiosonde observations from TAMDAR (3.97 m s^{-1}) were almost double that from AMDAR (2.21 m s^{-1}).

Vertical profiles of the wind observation differences are presented in Fig. 3. It should be noted that by aggregating results into 50 hPa bins, the variability associated with bins with few observations is reduced, especially between 300 and 800 hPa. When viewed separately as wind speed and direction, AMDAR speed biases and standard deviations are nearly constant with height. By contrast, TAMDAR biases are noticeably larger and less vertically consistent. AMDAR wind direction observations exhibit small average differences at all evaluated levels, although the sign of the TAMDAR average difference switches from positive near the surface to negative above 750 hPa. Standard deviation for both observing systems decrease in magnitude with increasing height, with AMDAR maintaining a substantial advantage at all levels. AMDAR wind vector RMSDs are nearly constant at all elevations, ranging from 2.4 to 2.6 m s^{-1}, while the differences between TAMDAR and radiosondes decrease from 5.4 m s^{-1} in the lowest 100 hPa of the profiles to a minimum of 4.3 m s^{-1} at 500 hPa. By contrast, Moninger et al. (2010) found somewhat different results when comparing TAMDAR and AMDAR winds to 1-h forecasts from a 20 km version of the Rapid Update Cycle (RUC) model. Although their overall conclusion that TAMDAR reports from turboprop aircraft had larger errors than for AMDAR jets when using a smoother reference dataset is similar to that presented here using point-specific radiosonde reports, the vertical structure of the TAMDAR errors differed. The earlier study, which used observations from a different set of TAMDAR aircraft, revealed that RMSDs for AMDAR increased from about 3.5 m s^{-1} near the surface to over 4.5 m s^{-1} above 400 hPa, while overall TAMDAR RMSDs were considerably larger, increasing rapidly from approximately 3.9 m s^{-1} near the surface and to approach 6 m s^{-1} at 500 hPa and above.

The ABO minus radiosonde differences were further evaluated to determine if the differences noted above were associated with along-track and cross-track wind components. No statistically significant difference (at the 95% confidence level) were found between the average difference of the along-track and cross-track wind components for either the AMDAR or TAMDAR observations. This implies that within each ABO source, observations are similarly affected by headwinds, tailwinds, and crosswinds.

b. Temperature

The distribution of temperature differences between radiosondes and both AMDAR (blue) and TAMDAR (red) are shown in Fig. 4a. These results (aggregated into 0.1 K bins) show that the AMDAR temperature observations are systematically cooler than radiosondes (−0.20 K) while the TAMDAR temperature average differences are both warmer and larger (+0.25 K). The TAMDAR-observed temperatures also exhibit nearly 3 times more random differences when

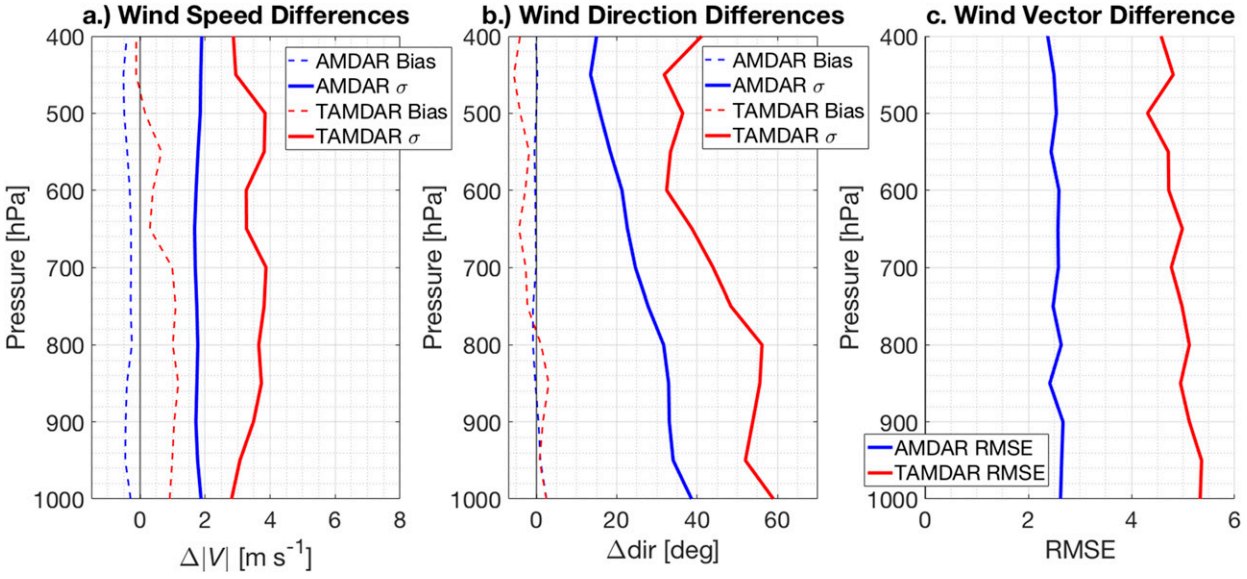

FIG. 3. (a) Vertical profiles of the aircraft minus sonde average difference (dashed lines) and standard deviation (solid lines) for AMDAR (blue) and TAMDAR (red) wind speed (in m s^{-1}) in pressure coordinates. (b) As in (a), but for wind direction (in degrees). (c) As in (a), but for the RMS of the vector difference (RMSD). All data have been aggregated into layers with a thickness of 50 hPa to allow for a sufficient number of observations in each bin.

compared to radiosondes than the AMDAR observations, with a standard deviation of 2.19 K for the former against 0.80 K for the latter. It should be noted that some operational assimilation schemes are capable of identifying aircraft that are providing bad observations and rejecting all observations from those aircraft, therefore the actual uncertainties associated the assimilated observations can be lower than what is reported here.

Vertical profiles (Fig. 4b) show the AMDAR temperature reports perform better than the TAMDAR data at all levels from the surface to 400 hPa. Average AMDAR temperature differences remain fairly uniform, ranging from −0.2 to −0.3 K. Random differences between AMDAR and radiosondes are also small and vertically consistent, with the largest variability (near 1.0 K) in the lowest levels of the atmosphere, decreasing to a consistent value of 0.7 K above the lowest 200 hPa. By contrast, both measures of TAMDAR differences are substantially larger. Average differences increase slightly with elevation from +0.3 K near the surface to about 0.4 K at 500 hPa, whereas standard deviations show a much larger and irregular vertical structure, with values ranging from 2.1 K near the surface to 2.7 K aloft. Larger random errors near the surface may be expected for both systems due to larger variability in the planetary boundary layer as opposed to the free troposphere; aircraft also undergo more substantial maneuvering at lower levels, which can disrupt airflow and cause nonrepresentative sampling. Unlike the average differences, which can be corrected, random differences will persist event after bias correction schemes are developed and applied.

c. Humidity

Unlike the wind and temperature assessments, direct comparison of humidity observations between the two ABO systems is more challenging as the AMDAR and TAMDAR provide fundamentally different measurements, with AMDAR/WVSS detecting SH and TAMDAR measuring RH. Conversions between variables require contemporaneous temperature observations, which can be problematic as the use of possibly biased ABO temperature values to convert between various measures of humidity can introduce additional errors that could obscure the quality of the moisture observations themselves. To address this issue, three different ABO-minus-sonde intercomparisons are presented for each ABO system. The first uses only direct aircraft observations of temperature and moisture, applying the aircraft-based temperature in the conversion to convert the specific humidity natively measured by AMDAR to relative humidity (and vice versa for TAMDAR). Although this approach provides a measure of the utility of observations made by each aircraft, it can particularly obfuscate the quality of the TAMDAR humidity sensor as converting from the TAMDAR RH to SH allows the relatively large TAMDAR temperature deviations to influence the results. It is useful to evaluate these comparisons in terms of RH as those values (and their associated average differences and standard deviations) do not trend toward zero with increasing altitude in the way that SH does.

In an attempt to assess the different ABO humidity sensors more directly and to remove the adverse effects of temperature differences between the aircraft and the radiosonde comparison standard, a second set of intercomparisons evaluated the measured WVSS SH reports against the TAMDAR SH calculated using temperatures observed by collocated radiosondes, as was done in Petersen et al. (2016). This approach provides an independent method of diagnosing the optimal quality possible from either ABO humidity sensor. In operations, however, this approach may overrepresent the utility of

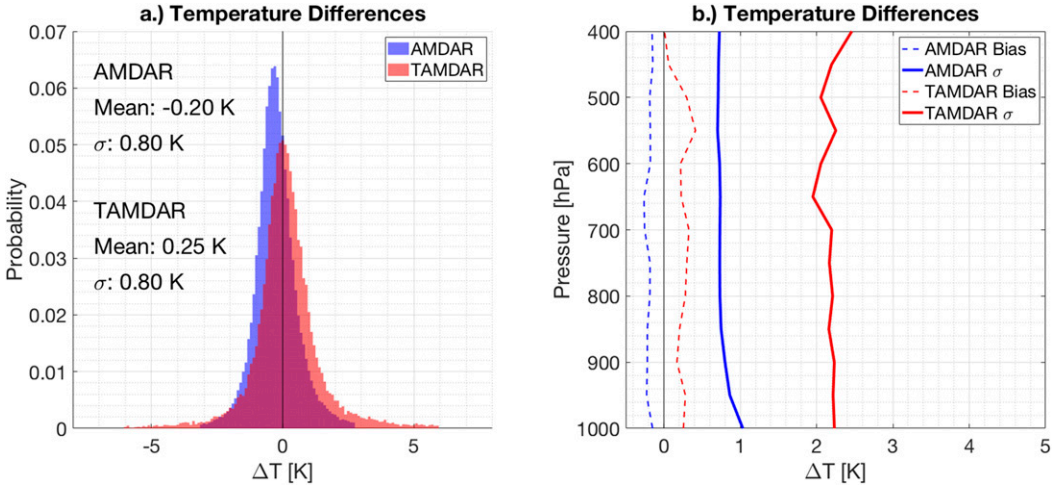

FIG. 4. (a) As in Fig. 2, but for temperature in 0.1 K bins. (b) As in Fig. 3, but for temperature (in K).

assimilating the ABO reports into NWP as most airborne observations are not found at locations and times that radiosonde temperature reports would be available. Corresponding intercomparisons made by contrasting directly measured TAMDAR RH with derived WVSS RH obtained using the radiosonde temperature facilitate the assessment of WVSS accuracy in RH space without undue influence from aircraft temperature inaccuracies.

Finally, a third set of tests were used in which profiles of the previous assessment of ABO temperature performance (as seen in Fig. 4b) was used to difference correct the ABO temperature observations in the moisture conversions. This was done in an attempt to optimize the accuracy of the majority of aircraft observations that are not closely matched with radiosondes. The difference correction was accomplished by linearly interpolating the vertical average difference profiles shown in Fig. 4 to the pressure level of the aircraft observations, then subtracting that value from the observed temperatures. This process removed systemic differences but retained the random error structures.

Figure 5 presents histograms of differences between airborne observations and radiosondes for the various humidity measures outlined above. Overall AMDAR performs as well as or better than TAMDAR for mean and random differences regardless of the humidity measure being used. Both sensors tend to produce observations that are moister than the radiosondes. For example, the aircraft-derived AMDAR RH average difference in Fig. 5a (using aircraft temperatures in converting SH to RH) is +3.38% with a standard deviation of 10.16%, while the directly observed TAMDAR measurements have slightly greater average differences and standard deviations of +4.56% and 10.58%, respectively. When radiosonde temperatures are used in the humidity conversion (Fig. 5b), the AMDAR average RH statistics drop to +3.12% and 9.15%, respectively (the TAMDAR values remain unchanged as no conversion was done). Because the conversion with the aircraft temperature introduces additional uncertainties into the AMDAR RH observation, it is not surprising that the average difference and standard deviation are larger when the aircraft temperature is used than when the radiosonde temperature is used.

For SH (Figs. 5d,e), the directly observed WVSS measurements show average differences of $+0.31\,g\,kg^{-1}$ and standard deviations of 0.99. $g\,kg^{-1}$, which in this case are 30%–50% smaller than the corresponding TAMDAR statistics of +0.53 and $1.58\,g\,kg^{-1}$ (using aircraft temperature) and +0.51 and $1.30\,g\,kg^{-1}$ (using radiosonde temperature). Again, using radiosonde observations in the TAMDAR RH-to-SH conversion provides an optimal estimate of the random SH differences. However, if the AMDAR and TAMDAR temperature reports that have been difference corrected are used for the conversions (Figs. 5c,f), the systematic differences between the ABO and radiosonde show some improvement relative to the plane-only values. Based on this result and to improve consistency between the two different sources of moisture information, it is recommended that, at a minimum, all ABO temperature observations be subjected to bias removal procedures before any moisture parameter conversions are attempted in operations.

As before, it is instructive to investigate vertical profiles of the accuracy and utility of the ABO humidity data. As shown in Fig. 6, both ABO systems exhibit positive moist average differences at all levels. AMDAR/WVSS almost always outperforms TAMDAR at all levels for both statistical measures regardless of which humidity parameter is used. The clearly identifiable decrease in SH average difference and standard deviation with height for both instruments is due to a large degree to the reduction in absolute water vapor at increasing altitudes due to decreasing saturation vapor pressures. The impact of applying average difference corrections to aircraft temperature profiles (e.g., the values shown in Fig. 4b) in humidity parameter conversion is also apparent. The corrected humidity measures (AMDAR in Fig. 6c and TAMDAR in Fig. 6f) show a reduced average difference relative to the values obtained using plane temperatures (corresponding Figs. 6a,d), though random errors change only slightly. Reductions in differences are most evident for TAMDAR

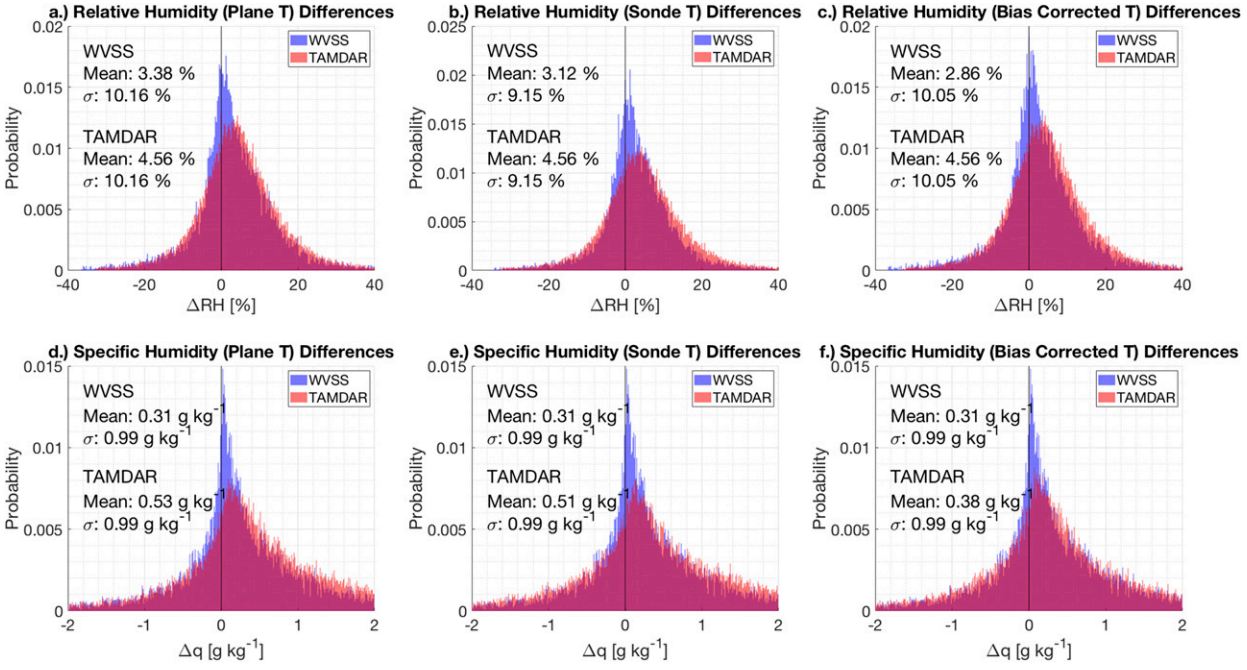

FIG. 5. (a) As in Fig. 2, but for RH (in %). AMDAR observations of SH are converted to RH using the aircraft temperature observation. Data are separated into 0.25% bins. (b) As in (a), but for AMDAR observations converted to RH using the collocated radiosonde temperature. (c) As in (a), but using difference-corrected AMDAR temperatures to convert WVSS SH to RH. (d) As in (a), but for SH (in $g\,kg^{-1}$). Data are separated into 0.01 $g\,kg^{-1}$ bins. TAMDAR observations of RH are converted to SH using the aircraft temperature observation. (e) As in (d), but for TAMDAR observations converted to SH using the collocated radiosonde temperature. (f) As in (d), but using difference-corrected TAMDAR temperatures to convert TAMDAR RH to SH.

SH in the 1000–800 hPa layer. Random differences remain higher for TAMDAR at the vast majority of levels, independent of moisture parameter or conversion choice. It should be noted that there is negligible correlation between the temperature and moisture differences relative to radiosonde observations for both the AMDAR and TAMDAR observations.

What is not immediately clear is whether the persistent moist average difference of the aircraft observations is the result of deficiencies in the ABO instruments or a reflection of known dry biases in the radiosondes caused by direct insolation of the humidity sensor (e.g., Miloshevich et al. 2001; Wang et al. 2013; Dzambo et al. 2016). If the radiosonde is underreporting atmospheric humidity, then an accurate airborne sensor will appear to have a moist bias relative to the radiosonde, and this bias would be larger during the day than at night when solar heating is present. One way to evaluate this is to separately quantify the biases of the aircraft against daytime and nighttime radiosonde launches. Unfortunately, the synoptic launch times of 0000 and 1200 UTC are difficult to classify into daytime and nighttime categories in the CONUS region due to their proximity to sunset and sunrise. While it is possible to parse the radiosonde dataset for launches that are known to be at night to conduct a deeper analysis of this issue (e.g., although most 1200 UTC launches in the western United States happen at night, many of the 0000 UTC launches in the east happen at night during the times of the year with shorter days), such an analysis is beyond the scope of the present work. As an alternative, asynoptic sondes (those launched at times outside of 0000 and 1200 UTC, often timed to capture preconvective environments for days in which significant severe weather is anticipated) were used to provide a preliminary look at this issue. When compared only to the midday asynoptic launches, the RH average differences and standard deviations (not shown) calculated using radiosonde temperatures for AMDAR were +3.12% and 9.99%, respectively, and for TAMDAR were +3.03% and 9.90%. For both ABO systems, these differences are smaller during midday than they were for the full dataset, which is dominated by synoptic-time reports. One would expect that if the radiosonde dry bias were an issue, ABO systems would have a stronger moist bias relative to the midday radiosondes than relative to the entire dataset, but this was not the case.

d. Differences as a function of radiosonde values

The analysis of instrument performance discussed above was carried out across all levels for all valid observations regardless of their value; for example, so long as a low-humidity and high-humidity observation occurred at the same pressure level, they were analyzed together. However, it is not certain that a given instrument will experience the same performance for a given variable when those observations are low as it will when those observations are large. Figure 7 shows how the ABO minus radiosonde differences vary as a function of the observed radiosonde value. Calculations were made using all available parameter matchups. Results are

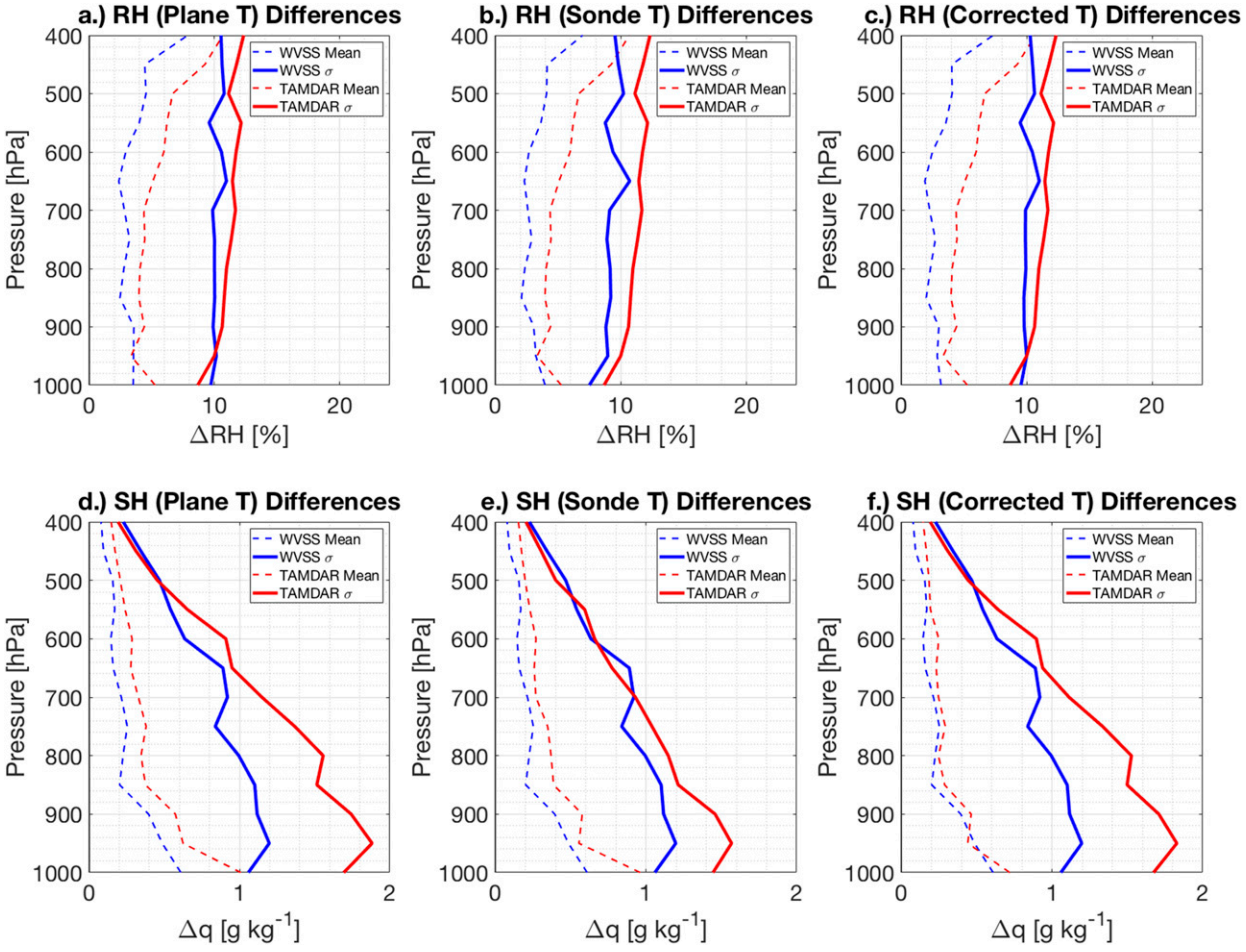

FIG. 6. (a) Vertical profiles of the aircraft minus sonde average difference (solid lines) and standard deviation (dashed lines) for AMDAR (blue) and TAMDAR (red) RH observations in percent in pressure coordinates. WVSS observations of SH are converted to RH using the aircraft temperature observation. (b) As in (a), but for AMDAR observations converted to RH using the collocated radiosonde temperature. (c) As in (a), but using difference-corrected AMDAR temperatures to convert WVSS SH to RH. (d) As in (a), but for SH. TAMDAR observations of RH are converted to SH using the aircraft temperature observation. (e) As in (d), but for TAMDAR observations converted to SH using the collocated radiosonde temperature. (f) As in (d), but using difference-corrected TAMDAR temperatures to convert TAMDAR RH to SH.

displayed as box-and-whisker plots for both AMDAR (blue) and TAMDAR (red). Boxes cover the range between the 25th and 75th percentiles, while whiskers extend to the 5th and 95th percentiles. Data are sorted into bins of 4 K, 4 m s^{-1}, 5%, and 1 g kg^{-1} for temperature, wind speed, RH, and SH, respectively. The number of observations in each bin are also shown as line plots at the bottom of each panel, which allows for an at-a-glance assessment of how the instruments are performing across the observation space.

The comparative values for wind speeds are shown in Fig. 7a. Both systems observe the same modal value for wind speed, but a greater proportion of the AMDAR observations are at higher wind speeds than the TAMDAR observations, a consequence of the lack of TAMDAR-equipped aircraft flying at higher altitudes (see Fig. 1b). While the same-environment criterion discussed earlier meant that every TAMDAR difference was paired with at least one AMDAR difference, at upper altitudes a balloon observation could be matched to many AMDAR observations but as few as one TAMDAR observation, resulting in differences and uncertainties that may not be fully representative. In the lowest wind speed range, both AMDAR and TAMDAR exhibited a positive average difference. This should be expected since any nonzero wind observation can only be evaluated as being faster than a calm radiosonde wind report, as no negative speed observations are possible. For speeds below 8 m s^{-1}, the median AMDAR difference is closer to zero than the median TAMDAR observation, but that relationship is reversed at higher wind speeds, with both systems experience a slow (negative) median difference, although the mean TAMDAR difference is closer to zero. Earlier it was noted that for the entire dataset the mean TAMDAR difference was positive while these results indicate that the majority of the observable ranges experiences negative average differences for TAMDAR. This is due to the large

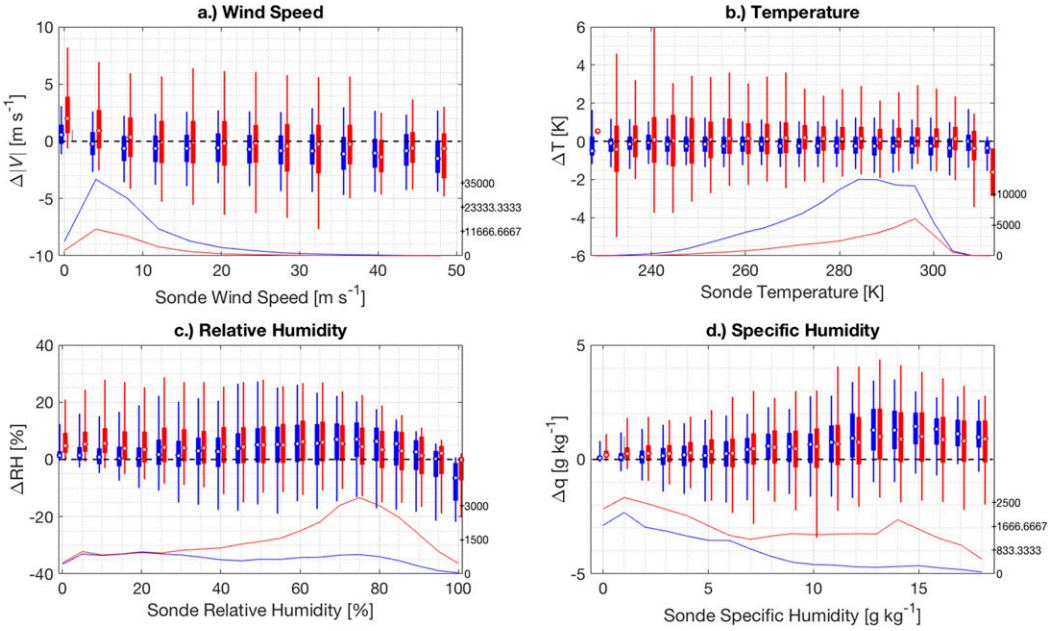

FIG. 7. Differences between ABOs and radiosondes for selected variables for both AMDAR (blue) and TAMDAR (red) grouped by observed radiosonde values. Bin sizes are 4 K, 4 m s^{-1}, 5%, and 1 g kg^{-1} for temperature, wind speed, relative humidity, and specific humidity, respectively. Both the AMDAR relative humidity and the TAMDAR specific humidity were calculated using plane, not sonde, temperatures. The boxes span the 25th–75th percentiles of differences in each bin, while the whiskers extend to the 5th and 95th percentiles. The white dot denotes the median. To avoid overlap of the plotted lines, the AMDAR bins are displaced slightly to the left while the TAMDAR bins are displaced slightly to the right; however, the underlying radiosonde-based bins are identical for both observing systems. For reference, the number of observations used in determining the distributions in each bin is plotted at the bottom and labeled along the lower-right side of each panel.

number of observations concentrated at the low wind speeds where the differences are largely positive, but it illustrates the importance of investigating the performance of these systems as a function of both pressure and observed value. In contrast to the median values, overall uncertainties in wind speed (noted by the length of the individual boxes) remain less for AMDAR observations than for TAMDAR. If the 400 hPa vertical limit were removed (thus including many more AMDAR cruise level observations), AMDAR average difference show nearly constant value around 0.2 m s^{-1} for all wind speeds faster than 10 m s^{-1}.

For the temperature observations (Fig. 7b), the AMDAR median difference and uncertainty are consistent across the full observed temperature range (from 245 to 305 K). While the TAMDAR median values exhibit similar uniformity throughout that range, their associated uncertainties are larger and have greater variability from one bin to the next than AMDAR. The distribution of observations indicates that the modal AMDAR temperature is cooler than it is for TAMDAR, since the overall AMDAR dataset contained a greater proportion of its observations at higher altitudes. For all observed temperatures, the uncertainties are larger for TAMDAR than they are for AMDAR, the only exceptions being in bins at the extreme ends of the range that have very few observations.

Results for RH and SH (Figs. 7c and 7d, respectively) were derived using the aircraft temperature observations to represent how real-time users of the data would view the data before NWP-based quality control corrections were applied. For RH values less than 45%, the derived AMDAR values exhibit smaller median differences than the direct TAMDAR observations, with both being positive (moist). From 45% to 65%, the median differences for the two systems are effectively the same. Between 65% and 85%, TAMDAR tends to have smaller median differences while the uncertainties are similar. Above 85%, the trend in both systems is for the median difference to become less dry with increasing RH. Both systems exhibit a dry bias for values above 95% as there is a finite upper limit on observable RH.

For the SH, both systems tend toward moister observations than measured by the radiosondes (although the AMDAR median difference is 0.2 g kg^{-1} or less for radiosonde-observed values less than 5 g kg^{-1}). Median AMDAR differences remain smaller than TAMDAR below 7 g kg^{-1}, but that relationship is reversed for moister environments. These differences increase with increasing SH up to maxima near 1.5 g kg^{-1} for values approaching 13 g kg^{-1} (TAMDAR) to 15 g kg^{-1} (AMDAR), at which point they begin to decrease. However, because the AMDAR dataset is dominated by observations in environments with low overall absolute moisture, overall median differences of SH measured directly by WVSS are smaller than those derived from TAMDAR RH and temperature reports.

Further investigation of the results in high-humidity regimes revealed that some of the discrepancy in differences may be attributed to differences in the biases between the two different types of radiosonde that were used for validation at various sites and the fact that the ratio of TAMDAR to AMDAR flights into the nearby airports were not uniform. For example, if more TAMDAR flights frequented airports in more tropical environments that used sondes that were biased moist, that could skew the comparison results and result in a reduced overall average difference for this humidity range, since few high reports of humidity would be available from more temperate regions. Further investigation of this factor is underway but beyond the scope of this paper.

4. Aircraft-to-aircraft observations

With the high density of flights in and out of major airports, direct comparisons between aircraft were also possible. Quantifying these differences is crucial for knowing how repeatable the ABO observations are; that is, if two different aircraft observe the same environment simultaneously, what is the likelihood that both aircraft return the same observation? For this analysis, similar matching criteria as before were used, including the same spatial, vertical, and temporal thresholds. To facilitate direct comparison with the aircraft-to-radiosonde analysis described above, this analysis also included an additional restriction requiring that aircraft from both AMDAR and TAMDAR needed to provide observations coincident with each radiosonde report. Although this approach limits the overall number of plane-to-plane matchups for each of the two ABO systems, the stricter matching criteria used here assures that the results are representative of the same atmospheric conditions.

In this section, the RMS of the differences is used as a measure of agreement among observations instead of the average and standard deviation of the differences since one cannot determine which in a pair of aircraft observations should be used as the baseline for comparison. This provides a consistent standard for assessing the relative spread of the observations within each of the two ABO systems and between ABOs and radiosondes. Although RMS does not necessarily provide full details about the accuracy of a given observing system, it does provide a quantitative estimate of the repeatability of an observation: smaller values indicate that differences between two aircraft are more likely to be due to differences in the environment and not due to differences in instrumentation.

As in results shown earlier, AMDAR provided more consistent observations than corresponding TAMDAR observations when compared either with radiosondes or against other similar types of aircraft. Aircraft-to-aircraft comparisons are summarized in italics in the top row of each cell in Table 2, which shows that across the board, AMDAR observations made by nearby aircraft are more consistent with each other (lower values for the RMS) for every observed parameter than TAMDAR. RMS statistics for the radiosonde-aircraft intercomparison, displayed in the bottom row of each cell, quantify the superiority of AMDAR noted previously in all categories. It is worth noting that overall, RH observations made directly by the TAMDAR instrument, as well as RH values using AMDAR SH and radiosonde temperatures, show better agreement in the plane-to-plane comparisons than in the plane-to-radiosonde results. Parameter conversion errors introduced by AMDAR temperature observations, however, negate this advantage. Although the overall plane-to-plane RMS values for SH are slightly larger than the plane-to-radiosonde values (1.07 vs 1.03 for AMDAR), the fact that SH values trend toward zero with increasing altitude means that numerous near-zero differences could affect these results.

With this in mind, the performance of all observation types was also assessed in the lower regions of the atmosphere (between 750 hPa and the surface). This is of particular interest for humidity parameters, since that is where water vapor content has significant impacts on stability and precipitation. Comparisons for all parameters for the lower troposphere are shown in the right columns of Table 2. In the lower troposphere, plane-to-plane RMS for AMDAR SH remain similar to plane-to-radiosonde values but increase in magnitude from 1.07 to 1.28 $g\,kg^{-1}$. By contrast, plane-to-plane values for TAMDAR SH derived using radiosonde and aircraft temperatures increased to 1.70 and 2.14 $g\,kg^{-1}$, respectively. Expansion of these statistics to include all possible collocations within the full ABO datasets could provide even more information about expected spatial and temporal variability in observation accuracy and representativeness.

5. Summary and conclusions

ABOs are a crucial part of the global observing network, greatly increasing the density of upper-air observations at a much lower cost-per-observation-impact than other observing systems. Two largely complementary ABO systems currently operating within CONUS: freely available AMDAR, which provides observations of temperature and winds (120 of which include WVSS SH observations); and commercially available TAMDAR, which observes temperature, RH, and wind from a distinct set of mostly smaller aircraft generally flying shorter routes at lower altitudes than the AMDAR aircraft. In this work, both observing systems were compared to operational radiosondes in order to assess their statistical uncertainties and compatibilities.

For this study, AMDAR and TAMDAR observations were compared against full-resolution NWS radiosonde observations over the CONUS for all of 2018. To remove obvious outliers that had passed internal system quality control checks, a gross error check was also included. Differences between the ABO and the corresponding radiosonde that were more than three standard deviations away from the average difference for that observation type were rejected from this analysis. While this could be due to instrumentation issues, there are also meteorological reasons why such large differences could exist between two properly operating collocated measurement systems, such as frontal boundaries or scattered cloudiness. It should be noted, however, that the data rejection had a greater impact on the TAMDAR

TABLE 2. The root-mean-square (RMS) differences for plane-to-plane (top row of each cell; italics) and plane-to-sonde (bottom row of each cell; regular type) comparisons for the AMDAR and TAMDAR datasets. Boldface values indicate where plane-to-plane comparisons have lower RMS differences than plane-to-sonde comparisons. Values are shown for all study levels in the left two columns and from the surface through 750 hPa in the right two columns.

	AMDAR RMS 400 hPa upper limit	TAMDAR RMS 400 hPa upper limit	AMDAR RMS 750 hPa upper limit	TAMDAR RMS 750 hPa upper limit
Wind speed ($m\,s^{-1}$)	*1.82*	***3.43***	*1.83*	*3.40*
	1.81	3.49	1.81	3.33
Wind direction (°)	*22.96*	*42.97*	*27.88*	*47.83*
	22.85	39.79	27.68	44.27
Wind vector difference ($m\,s^{-1}$)	*2.84*	*7.67*	*2.87*	*6.64*
	2.21	3.97	2.24	3.96
Temperature (K)	***0.82***	*2.94*	*0.90*	*3.08*
	0.82	2.20	0.89	2.31
RH (plane temperature) (%)	*10.91*	***11.52***	*10.57*	*11.39*
	10.71	11.90	10.39	10.81
RH (sonde temperature) (%)	***9.66***	***11.52***	*9.27*	*11.39*
	9.82	11.90	9.09	10.81
SH (plane temperature) ($g\,kg^{-1}$)	*1.07*	*1.91*	*1.28*	*2.14*
	1.03	1.66	1.24	1.92
SH (sonde temperature) ($g\,kg^{-1}$)	*1.07*	*1.52*	*1.28*	*1.70*
	1.03	1.40	1.24	1.61

observations than the AMDAR or WVSS datasets and cannot be explained using the same rationale. For example, the observed average and standard deviation of the aircraft-minus-radiosonde temperature for the AMDAR dataset after applying a 3σ gross error check was -0.20 ± 0.80 K while for TAMDAR it was 0.25 ± 2.19 K. If a gross check had not been used, the AMDAR values shifted slightly to -0.16 ± 0.99 K. However, the TAMDAR values shifted substantially to -0.25 ± 5.35 K. Similar shifts can be seen across all variables examined here. Therefore, when assessing the relative performance of AMDAR and TAMDAR, one must remember that the AMDAR observations tend to have better quality control in their as-delivered state while additional steps to control the quality of the TAMDAR observations may need to be taken by the end user.

Overall, AMDAR observations largely performed better than the TAMDAR. Differences between AMDAR and closely collocated NWS radiosondes exhibited smaller average and standard deviations of the aircraft-minus-sonde differences than TAMDAR observations at the same radiosonde locations and times for wind, temperature, and multiple measures of atmospheric humidity. AMDAR temperatures and wind speeds both showed small negative average differences while all other AMDAR measures and all TAMDAR observations exhibited positive average differences. Standard deviations of the TAMDAR–radiosonde matchup differences for wind and temperature were 2–3 times larger than those from AMDAR. For moisture, AMDAR showed reductions in random differences when compared with TAMDAR reports for both moisture parameters, 30%–50% for SH and 5%–15% for RH. The performance benefits of AMDAR generally extended throughout the depths of the troposphere over which the analysis was conducted and throughout the full parameter ranges, with only isolated instances where the TAMDAR averages or standard deviations of the differences were smaller than the corresponding value from AMDAR. It should be noted that well-characterized biases can be easily corrected, thus enhancing these observations' utility for forecasting, NWP assimilation, satellite product validation, and other uses.

Because AMDAR and TAMDAR provide fundamentally different measurements of humidity, the observation from each ABO system sets were each converted to the parameter provided by the other system using both the aircraft temperature and the temperature from a collocated radiosonde. In that way, both the real-world utility of the measured quantity and the performance of the moisture sensor could be assessed. In general, AMDAR outperformed TAMDAR in both humidity measures, with TAMDAR's SH performance hampered by larger systematic and random temperature uncertainties. This was true even when, as recommended in the text, average difference corrections were applied to the TAMDAR temperature reports. Obviously, most ABOs do not benefit from a collocated radiosonde observation that can reduce the impact of temperature errors further. For any numerical weather prediction systems that are assimilating TAMDAR observations in SH format, the inaccuracies in the TAMDAR temperature could contribute to incompatible moisture observations unless other forms of bias removal are applied.

It is worth noting that the analysis presented here implicitly assumed that radiosondes represented unbiased truth and that deviations from the radiosonde value were due to issues with the aircraft observations. However, any of the comparisons presented here, especially moisture measurements, could have been affected by inaccuracies in the radiosonde measurements. A rudimentary analysis indicated that known dry biases in daytime radiosonde observations may not be fully responsible

for the moist average differences seen in the ABO systems. A future pathway for analysis could be to rigorously separate the sounding dataset both by daytime and nighttime soundings and by radiosonde type to see how the moist average difference varies in those environments—effectively using the most reliable ABO profiles to validate and intercalibrate operational radiosondes. Small hysteresis effects and different performance as a function of NWS radiosonde type have been noted previously in ABOs (e.g., Williams et al. 2021), and an additional task for future analysis would be to quantify the differences between ascending and descending AMDAR and TAMDAR aircraft and as a function of the two major radiosonde types used by the NWS. Other potential explanations for observed differences included discrepancies in the temporal averaging and reporting frequency of the two ABO observing systems.

A measure of the consistency among the two different ABO reporting systems was obtained by including aircraft-to-aircraft intercomparisons. In all cases, AMDAR observations showed greater consistency between aircraft in the same environments than TAMDAR, with smaller RMS differences for every observation type analyzed here. For winds, temperatures, and SH, AMDAR plane-to-plane agreements were roughly 2–3 times better than TAMDAR. For humidity, AMDAR aircraft-to-aircraft comparisons showed better agreement than TAMDAR for both RH and SH at lower altitudes (where SH is usually higher). This consistency from one airplane to the next could be an indicator that reliability and reproducibility of AMDAR SH observations.

As this study illustrates the accuracy and repeatability of AMDAR/WVSS as compared to TAMDAR, and since these observations benefit from a worldwide network to acquire and disseminate them, it is clear that the global operational meteorological community would be well served by expanding the reach of AMDAR and WVSS observations. As noted by WMO (2017) and International Air Transport Association (IATA; IATA 2020), national meteorological and transportation services would do well to join with air carriers in their regions to explore mutually beneficial partnerships that would increase the number of airborne observations (especially from WVSS) in otherwise underobserved regions, such as much of the Southern Hemisphere.

Acknowledgments. Support for this research was provided by the National Weather Service Office of Observations through the Airborne Observation Program and was distributed through a cooperative agreement via the Cooperative Institute for Meteorological Satellite Studies at the University of Wisconsin–Madison. Stan Benjamin and one anonymous reviewer offered helpful comments that improved the readability and utility of this paper.

Data availability statement. All airborne data were obtained from the MADIS/point/acars/ data stream; more details are available at https://madis.ncep.noaa.gov. All radiosonde data were obtained from the Radiosonde Replacement System (RRS) Archive at the National Centers for Environmental Information.

REFERENCES

Benjamin, S. G., B. E. Schwartz, and R. E. Cole, 1999: Accuracy of ACARS wind and temperature observations determined by collocation. *Wea. Forecasting*, **14**, 1032–1038, https://doi.org/10.1175/1520-0434(1999)014<1032:AOAWAT>2.0.CO;2.

Daniels, T., G. Tsoucalas, M. Anderson, D. J. Mulally, W. Moninger, and R. Mamrosh, 2004: Tropospheric Airborne Meteorological Data Reporting (TAMDAR) sensor development. *11th Conf. on Aviation, Range, and Aerospace Meteorology*, Hyannis, MA, Amer. Meteor. Soc., 7.6, https://ams.confex.com/ams/11aram22sls/techprogram/paper_81841.htm.

——, W. R. Moninger, and R. D. Mamrosh, 2006: Tropospheric Airborne Meteorological Data Reporting (TAMDAR) overview. *10th Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface*, Atlanta, GA, Amer. Meteor. Soc., 9.1, https://ams.confex.com/ams/Annual2006/techprogram/paper_104773.htm.

Drüe, C., W. Frey, A. Hoff, and T. Hauf, 2008: Aircraft type-specific errors in AMDAR weather reports from commercial aircraft. *Quart. J. Roy. Meteor. Soc.*, **134**, 229–239, https://doi.org/10.1002/qj.205.

DuBois, J. L., R. P. Multhauf, and C. A. Ziegler, 2002: *The Invention and Development of the Radiosonde, with a Catalog of Upper-Atmospheric Telemetering Probes in the National Museum of American History*. Smithsonian Press, 82 pp.

Dzambo, A. M., D. D. Turner, and E. J. Mlawer, 2016: Evaluation of two Vaisala RS92 radiosonde solar radiative dry bias correction algorithms. *Atmos. Meas. Tech.*, **9**, 1613–1626, https://doi.org/10.5194/amt-9-1613-2016.

Eyre, J., and R. Reid, 2014: Cost-benefit studies of observing systems. Met Office Forecasting Research Tech. Rep. 593, 11 pp.

Gao, F., X. Zhang, N. Jacobs, X.-Y. Huang, X. Zhang, and P. Childs, 2012: Estimation of TAMDAR observational error and assimilation experiments. *Wea. Forecasting*, **27**, 856–877, https://doi.org/10.1175/WAF-D-11-00120.1.

Hoover, B. T., D. A. Santek, A. Daloz, Y. Zhong, R. Dworak, R. A. Petersen, and A. Collard, 2017: Forecast impact of assimilating aircraft WVSS-II water vapor mixing ratio observations in the Global Data Assimilation System (GDAS). *Wea. Forecasting*, **32**, 1603–1611, https://doi.org/10.1175/WAF-D-16-0202.1.

IATA, 2020: WMO and IATA agree to improve aircraft meteorological reporting. IATA Press Release 85, 2 pp., https://www.iata.org/en/pressroom/pr/2020-10-27-01/.

Jacobs, N. A., Y. Liu, and C. M. Druse, 2006: Evaluation of temporal and spatial distribution of TAMDAR data in short-range mesoscale forecasts. *10th Symp. on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface*, Atlanta, GA, Amer. Meteor. Soc., 9.11, https://ams.confex.com/ams/Annual2006/techprogram/paper_103388.htm.

James, E. P., and S. G. Benjamin, 2017: Observation system experiments with the hourly updating Rapid Refresh model using GSI hybrid ensemble–variational data assimilation. *Mon. Wea. Rev.*, **145**, 2897–2918, https://doi.org/10.1175/MWR-D-16-0398.1.

——, ——, and B. D. Jamison, 2020: Commercial-aircraft-based observations for NWP: Global coverage, data impacts, and COVID-19. *J. Appl. Meteor. Climatol.*, **59**, 1809–1825, https://doi.org/10.1175/JAMC-D-20-0010.1.

Miloshevich, L. M., H. Vömel, A. Paukkunen, A. J. Heymsfield, and S. J. Oltmans, 2001: Characterization and correlation of relative humidity measurements from Vaisala RS80-A radiosondes at cold temperatures. *J. Atmos. Meteor. Technol.*, **12**, 135–156, https://doi.org/10.1175/1520-0426(2001)018<0135:CACORH>2.0.CO;2.

Moninger, W. R., R. D. Mamrosh, and P. M. Pauley, 2003: Automated meteorological reports from commercial aircraft. *Bull. Amer. Meteor. Soc.*, **84**, 203–216, https://doi.org/10.1175/BAMS-84-2-203.

——, S. G. Benjamin, B. D. Jamison, T. W. Schlatter, T. L. Smith, and E. J. Szoke, 2010: Evaluation of regional aircraft observation using TAMDAR. *Wea. Forecasting*, **25**, 627–645, https://doi.org/10.1175/2009WAF2222321.1.

NCDC, 2008: National Climatic Data Center Data documentation for dataset 6213: Radiosonde Replacement System—1 second BUFR. NCDC Doc., 22 pp.

Painting, J. D., 2003: AMDAR reference manual. WMO Doc. 958, https://library.wmo.int/doc_num.php?explnum_id=9026.

Petersen, R. A., 2016: On the impact and future benefits of AMDAR observations in operational forecasting. Part I: A review of the impact of automated aircraft wind and temperature reports. *Bull. Amer. Meteor. Soc.*, **97**, 585–602, https://doi.org/10.1175/BAMS-D-14-00055.1.

——, L. Cronce, R. Mamrosh, R. Baker, and P. Pauley, 2016: On the impact and future benefits of AMDAR observations in operational forecasting. Part II: Water vapor observations. *Bull. Amer. Meteor. Soc.*, **97**, 2117–2133, https://doi.org/10.1175/BAMS-D-14-00211.1.

Schwartz, B., and S. G. Benjamin, 1995: A comparison of temperature and wind measurements from ACARS-equipped aircraft and radiosondes. *Wea. Forecasting*, **10**, 528–544, https://doi.org/10.1175/1520-0434(1995)010<0528:ACOTAW>2.0.CO;2.

Wang, J., L. Zhang, A. Dai, F. Immler, M. Sommer, and H. Vömel, 2013: Radiation dry bias correction of Vaisala RS92 humidity data and its impacts on historical radiosonde data. *J. Atmos. Oceanic Technol.*, **30**, 197–214, https://doi.org/10.1175/JTECH-D-12-00113.1.

WIGOS, 2014: The benefits of AMDAR data to meteorology and aviation. WMO Tech. Rep. 2014-01, 89 pp., https://library.wmo.int/doc_num.php?explnum_id=10548.

Williams, S. S., T. J. Wagner, and R. A. Petersen, 2021: Examining the compatibility of aircraft moisture observations and operational radiosondes. *J. Atmos. Oceanic Technol.*, **38**, 859–872, https://doi.org/10.1175/JTECH-D-20-0053.1.

WMO, 2017: Guide to aircraft-based observations. WMO Rep. 1200, 141 pp., https://library.wmo.int/doc_num.php?explnum_id=4120.

Zhang, Y., D. Li, Z. Lin, J. A. Santanello, and Z. Gao, 2018: Development and evaluation of a long-term data record of planetary boundary layer profiles from aircraft meteorological reports. *J. Geophys. Res. Atmos.*, **124**, 2008–2030, https://doi.org/10.1029/2018JD029529.