Examining the Compatibility of Aircraft Moisture Observations and Operational Radiosondes

SKYLAR S. WILLIAMS, TIMOTHY J. WAGNER, AND RALPH A. PETERSEN

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

(Manuscript received 14 April 2020, in final form 28 January 2021)

ABSTRACT: The addition of moisture observations via the Water Vapor Sensing System (WVSS) from about 150 aircraft available operationally through the World Meteorological Organization’s (WMO) Aircraft Meteorological Data Relay (AMDar) program now provides highly reliable thermodynamic profiles of the troposphere. The nearly 900 profiles available daily provide greater temporal and spatial density than the operational radiosonde network over many parts of the United States. Previous studies comparing WVSS reports with specially collocated radiosondes have documented the quality and consistency of the WVSS observations. These studies, however, have been limited for short periods at a single location. This study expands on the earlier evaluations by using operational U.S. radiosondes from 2015 in a variety of locations, seasons, and climates. Comparison profiles at radiosonde sites were calculated in pressure layers and then interpolated to terrain-following sigma coordinates to account for the differences in elevations of comparison sites and provide a better means of integrating the higher vertical resolution of AMDAR observations taken in the boundary layer. Overall, systematic differences between the WVSS and radiosondes are smallest just above the surface, with the WVSS observations being slightly moister than the radiosondes aloft, with WVSS reports being moister during ascent than descent—possibly the result of small hysteresis effects. Standard deviations averaged 1.3 g kg⁻¹ near the surface over the yearlong period. Differences varied by season and region. Overall, the results indicate that WVSS observations are compatible with radiosonde reports and can be used with high confidence to fill temporal and spatial data gaps.

KEYWORDS: Water vapor; Aircraft observations; In situ atmospheric observations; Instrumentation/sensors; Profilers, atmospheric; Radiosonde/rawinsonde observations

1. Introduction

In 2015, 680,000 observations of geolocated wind and temperature were collected daily by 3500 aircraft from 39 airlines globally as part of routine commercial aircraft operations. These data are coordinated through the World Meteorological Organization’s (WMO) Aircraft Meteorological Data Relay (AMDar) program, which monitors global aircraft data. While routine monitoring of environmental temperature and wind conditions is necessary for safe operation of a commercial jetliner, airlines do not achieve an immediate operational benefit from adding moisture observations to their existing sensor suites. Therefore, only planes that are specially equipped with moisture sensors can monitor moisture conditions; approximately 150 aircraft, mainly in the United States, are capable of providing moisture data (Petersen 2016).

The second and current generation of the Water Vapor Sensing System (WVSS) uses a laser diode that directly measures the water vapor mixing ratio by counting the individual molecules moving past the sensor in a specific volume of air, thus providing a direct measurement of water vapor specific humidity that is independent of any errors in the aircraft temperature and wind observations (Fleming and May 2004; Petersen et al. 2016). The diode operates at a single wavelength of 1.37 µm, a strong water vapor absorption channel that is insensitive to ice crystals or aerosols. Comparisons of the first versions of the WVSS to radiosondes begun in 2005 showed slightly negative biases but, more importantly, inconsistent random errors uncovered flaws in the system’s electronics design that needed correction (Petersen et al. 2006a,b).

These problems were resolved in a subsequent version of WVSS that has been deployed operationally since 2009. The moisture data measured by WVSS are transmitted to the ground via the Aircraft Communications Addressing and Reporting Systems (ACARS) and are included with independent measurements of temperature, wind, flight level, and aircraft position. The sensor weighs 3.5 kg and has a range of detectable signal of 50 to 40,000 ppmv with an accuracy of ±50 ppmv or ±5% of the reading (whichever is greater). The operating temperature range for the WVSS-II is −65° to +50°C while the sampling pressure ranges from the surface to 200 hPa (SpectraSensors 2010). The optical chamber is heated to 38°C. In the rare event any liquid water or ice particles get into the WVSS, they are converted into vapor and can alter the readings to some degree. AMDAR observations are made at different intervals during different phases of flight [see Petersen (2016) for details]. In the lowest 100 hPa of the atmosphere, ascending aircraft nominally provide records approximately once every 100 m while descending aircraft report roughly every 300 m. The high sampling frequency allows for vertical moisture profiles to be collected with a vertical...
resolution more similar to radiosondes. Between this level and final cruise altitude, reports are made approximately every 300 m, while at flight level, reports are made at least once every 100 km. In the United States, the sensor is deployed primarily on Boeing 757 aircraft operated by United Parcel Service (UPS) and Boeing 737 aircraft operated by Southwest Airlines, while in Europe it is deployed aboard Airbus A320 aircraft flying in short-haul service for Lufthansa. Additional information on the WVSS can be found in WMO (2019).

The quantity of data collected from aircraft reflect the diurnal cycle of aviation traffic as a whole, with a primary maximum in the evenings and a secondary maximum in the mornings with the primary minimum overnight and a secondary minimum midday (Benjamin et al. 2010). The number of reports also decreases during weekends. High-resolution vertical profiles are concentrated near major hubs (Jamison and Moninger 2002). The number of observations at any particular airport can vary when certain weather conditions force flight cancelations. Although convective weather is the greatest weather-based cause of flight delays overall, the greatest disruptions in WVSS observations are caused by ice and snow as adverse winter weather tends to force cancellations while convection merely tends to delay them. A canceled flight results in missed profiles at both the origin and destination airport.

Although airborne observations are generally restricted to the troposphere and lower stratosphere, they have the advantage of being extremely cost-effective with respect to the global observing system (Eyre and Reid 2014), with an entire aircraft profile costing less than 5% of that for a complete radiosonde launch. A WMO study found that, over a 10-yr period, a hypothetical radiosonde network of 10 locations compared to a fleet of 30 aircraft (assuming the aircraft generated eight profiles per day) would cost 5–8 times more than that of the aircraft observations yet record an order of magnitude fewer profiles (WMO 2014).

In the past, several validation studies have documented the quality of WVSS observations by comparing to specially launched radiosondes (e.g., Fleming 2000; Moninger et al. 2003; Petersen et al. 2006a, 2011, 2016). The operational utility of these studies was limited by being performed over short time periods at a single location and relying on observations from nonoperational radiosondes, which also prevented the examination of seasonality in the WVSS observations and the instrument’s performance in different climate regimes and geographical areas. As the use of these data increases, both operationally and in climate study applications (e.g., Rahn and Mitchell 2016), it is necessary to properly characterize the accuracy of the WVSS sensor with a broader scope. The present work demonstrates how the WVSS performs compared to the operational National Weather Service (NWS) radiosonde network for an entire year at all locations in the continental United States where comparisons are feasible, enabling the evaluation of the accuracy of the WVSS from two airlines and aircraft types against two different radiosonde systems in diverse seasons and water vapor environments. Examining the geographical and seasonal variations in accuracy allows for an increase in confidence in this instrument and also enables correction schemes to be implemented for any observed differences.

2. Data and methodology

Operational NWS radiosondes were chosen as the comparison standard “truth” for this study since they are the only moisture observation at a sufficiently high vertical resolution available operationally across the CONUS. It is important to note that this does not imply that radiosondes provide more accurate observations of moisture profiles than WVSS. In fact, the WVSS may be more accurate than the radiosondes for several reasons (Petersen 2016). First, radiosondes are designed to be one-time use instruments while the WVSS systems are used repeatedly which allows the latter to be engineered to higher standards; the first ones deployed have outperformed their 10-yr life expectancy. In addition, the WVSS has a much faster response time, as it also directly counts the number of water vapor molecules in the atmosphere instead of deriving relative humidity indirectly from changes in capacitance.

During 2015, the target year of this study, the NWS operated two different models of radiosondes: the LMS-6 by Lockheed Martin Sippican (hereafter called Sippican) and the RS92-NGP by Vaisala (hereafter called Vaisala). Both models report pressure, temperature, dewpoint temperature (which is converted to relative humidity in postprocessing), latitude, longitude, and altitude. The measurement accuracy achievable by Vaisala is discussed by Dirksen et al. (2014). Comparisons of Vaisala radiosondes and satellite hyperspectral infrared measurements can be found in Sun et al. (2021). While some studies comparing the WVSS and radiosondes assumed the radiosonde location at all levels to be the launch site location (e.g., Mamrosh et al. 2006), this study accounts for the drift in the radiosonde by using the GPS coordinates at each level of the radiosonde which allows more accurate matching with airplane locations, an approach similar to that of Schwartz and Benjamin (1995). Both the Sippican and Vaisala radiosondes have a thin-film capacitance relative humidity sensor with a range of 0%–100% with a resolution of 1% and an accuracy of 5% over a pressure range of 1080 to 3 hPa. The Sippican radiosonde has a pressure accuracy better than 0.5 hPa and a resolution of 0.1 hPa while the Vaisala radiosondes have an accuracy of 1 hPa and resolution of 0.1 hPa between 1080 and 100 hPa (Vaisala 2015; WMO 2011).

To enable a direct comparison between the instruments, radiosonde humidity observations were converted to water vapor mixing ratio to match the observation taken directly from the WVSS; the conversion was made in this direction instead of the opposite (converting the WVSS specific humidity to humidity) as this would introduce additional biases into the analysis since aircraft temperature observations with their own set of biases not evaluated here (Mamrosh et al. 2006; Helms et al. 2010; Petersen et al. 2006a) would be required for such a conversion. This also closely approximated the humidity variable used in many NWP data assimilation (DA) systems.

Figure 1 shows the locations of airports with WVSS profile observations relative to NWS radiosonde launch sites. Locations in red are within 50 km of a radiosonde location and are used to
compare WVSS and NWS radiosonde moisture observations; blue locations are greater than 50 km from a radiosonde site and are not used in this study. The area of each circle is proportional to the number of WVSS observations at each site. Note that the locations with the greatest number of observations are not necessarily the busiest airports overall. Since the WVSS is primarily mounted aboard aircraft operated by Southwest and UPS, the greatest observational density is found at their hubs, like Dallas Love Field and Chicago Midway Airport (for Southwest) and Louisville, Kentucky (for UPS), and not the busier nearby airports like Dallas–Fort Worth International Airport or Chicago O’Hare International Airport. There are more comparison profiles in the eastern United States due to the fact that there are more flights occurring at the 1200 UTC (0700/0800 local time) radiosonde launch compared to the West Coast where 1200 UTC is much earlier relative to local time (0400/0500 local time).

Comparisons between the WVSS and NWS radiosondes were made for the entire year of 2015 against a variety of classifications: for both radiosonde types across the CONUS, for ascending and descending aircraft, over different climate regions determined by the Köppen–Geiger climate class, and for different airlines/airframes. These comparisons allow for a fuller understanding of how the WVSS responds for aircraft during the different flight patterns during ascent and descent as well as how the sensor responds for different climate regimes and seasons. By characterizing how the WVSS performs in these various situations, its utility for meteorological and climatological study as well as DA is greatly enhanced.

The criteria for evaluating observations for comparison are similar to those used in previous studies (e.g., Petersen et al. 2006a, 2016). To ensure that a set of observations are observing similar conditions, each WVSS observation must

- be within 30 min of the radiosonde observation,
- be within 10 hPa of the radiosonde observation,
- be within a 50 km radius of the radiosonde observation, and
- for low- and midlevel tropospheric comparisons, originate from an airplane taking off or landing at an airport near the radiosonde launch location.

This last criterion ensures that an airplane is creating a profile for comparison rather than only briefly flying over the comparison location while headed to a different destination. This is important because aircraft that are ascending/descending are traveling through rapidly changing moisture environments whereas aircraft traveling at altitude are in a more homogeneous moisture environment. Studying differences at high vertical resolution is important to determine how well WVSS responds to rapidly changing moisture environments.

To perform the comparisons, observations from the WVSS and the radiosondes were first interpolated onto a common vertical grid of 10 hPa from 1050 to 200 hPa, consistent with the highest vertical reporting interval of the AMDAR reports. Differences were then calculated by subtracting the calculated radiosonde water vapor mixing ratio value from the WVSS observed water vapor mixing ratio value. A positive value shows that the WVSS report is moister than the radiosonde while a negative value shows the WVSS is drier. When averaging profiles from multiple locations together, all profile comparisons were treated individually.

Merging comparisons from sites with greatly different surface elevations, however, precluded using the isobaric results directly. To guarantee that WVSS observations made within the boundary layer at all stations were treated together, the vertical comparison results were then converted from pressure space to a modified sigma space. For each individual comparison profile, the radiosonde pressure at the base of the lowest elevation 10 hPa intercomparison was used to define the bottom of the sigma domain. For example, if the water vapor measurement from the radiosonde closest to the surface was in the 1000–990 hPa bin, then 1000 hPa was used as the surface pressure. Note, because AMDAR reports sometimes begin after an aircraft has left the ground, WVSS measurements were
not always available at the surface as they were with the radiosonde profile. The sigma domain had a vertical spacing of 0.0125\(\sigma\) and the top of the domain was defined at 200 hPa, slightly above the highest flight level used by commercial aircraft over the United States. Intermediate levels were defined using the formula

\[ \sigma_{200} = (p - p_{200})/(p_S - p_{200}). \]

The advantage of using a sigma display system is illustrated in Fig. 2, which shows the average differences when combining all difference profiles between the WVSS and the radiosondes for 2015 from two locations, Fort Worth, Texas (FTW), at an elevation of approximately 200 m, and Denver, Colorado (DEN), located nearly 1400 m higher. The average individual profiles from each city in pressure coordinates (Fig. 2a) show the boundary layer much higher in DEN. When merged at common pressure levels (Fig. 2b), it is difficult to distinguish the boundary layer affecting the combined profile below 800 hPa. In fact, the lowest 150 hPa are only influenced by the comparisons from FTW, whereas the negative difference near the surface at DEN is reflected as reduction in the bias in the combined results below 800 hPa, well above the apparent boundary layer levels in the merged profile. When viewed in sigma coordinates (Fig. 2c), the influence of observations from the highest pressures is greatly reduced as the annual averaging done in conjunction with the isobaric-to-sigma coordinate transformation at each site combines the small number of comparisons made during periods of extremely high surface pressure (where the WVSS–radiosonde differences are negative) with at least 6 times as many cases made when the surface pressure was near average or lower (where the observation differences were positive). This results in positive differences.

**Fig. 2.** An example of individual and combined AMDAR–radiosonde difference profiles from Denver (DEN) and Fort Worth (FTW) for 2015 using (top) isobaric and (bottom) \(\sigma_{200}\) coordinates for vertical displays.
in the boundary layer that have been tempered slightly by the limited number of drier WVSS reports in cases of extremely high surface pressure. The sigma analysis thus more clearly separates boundary layer effects from the free atmosphere (Fig. 2d).

3. Results and discussion

The following section is broken into several subsections: comparisons in the lower half of the troposphere, ascent versus descent profiles, seasons, radiosonde model, upper-tropospheric comparisons, regional comparisons, and comparisons by aircraft type. This is done to easily examine and discuss results of each subsample of data together. All comparisons made, except for the upper-tropospheric comparisons, are limited by aircraft taking off or landing at airports near the radiosonde location. For lower-tropospheric comparisons, this ensures that aircraft are ascending or descending and have not reached cruising altitude. For upper-tropospheric comparisons, this allows for aircraft that have reached cruising altitude to be included in the comparisons.

a. Comparisons in the lower half of the troposphere

Comparisons made in the lower half of the atmosphere ($\sigma_{200} = 1.0$ to $\sigma_{200} = 0.5$) required the WVSS measurements be provided by an airplane that was taking off or landing from an airport near the comparison radiosonde location. The average of all the WVSS–radiosonde differences in the lower half of the troposphere was 0.15 g kg$^{-1}$. The median of the differences was 0.075 g kg$^{-1}$ with a slight positive skewness in the distribution of differences. Performing a one-sample $t$ test with the differences resulted in rejecting the null hypothesis of the mean is equal to zero at the 5% significance level. Figure 3a shows the average of all the comparison profiles from across the entire CONUS for the entire year. The comparison profile was created by taking all the individual WVSS profiles and subtracting the matched radiosonde’s water vapor mixing ratios then averaging the differences at all stations together. Because each individual profile is a unique sample of a different moisture environment, no weighting was done by location. Near the surface, the average differences between the WVSS and radiosonde observations was slightly negative at the surface. Since radiosondes are rarely launched at major airports, this could reflect local differences in surface cover.

The number of WVSS comparisons per level (Fig. 3b) shows a maximum in collocated data points between $\sigma_{200} = 0.94$ and $\sigma_{200} = 0.92$ and then a steady decrease with increasing height. This is not due to fewer WVSS observations but results from the fact that aircraft are likely to be separated from a radiosonde by more than 50 km at these higher altitudes as they travel away from their departure airports with greater horizontal displacement than the radiosondes.

The standard deviation of all WVSS differences was 1.16 g kg$^{-1}$. The profile of standard deviations decreased steadily from approximately 1.3 g kg$^{-1}$ near the surface to less than 0.9 g kg$^{-1}$ in the middle of the atmosphere. This variation in random differences could be the result because the total amount of water vapor decreases with height and because the tightest horizontal gradients in moisture tend to be within the boundary layer closest to the surface. If the WVSS equipped aircraft is flying toward one side of a moisture gradient while the radiosonde remains on the other side, the difference could be large.
b. Comparisons by ascent versus descent profiles

Similar to previous studies (Petersen et al. 2011, 2016), the cause of the apparent moist bias with respect to the radiosondes in the WVSS observations can be investigated further by separating results between ascending and descending aircraft (Fig. 4a). Throughout most of the troposphere, the WVSS observations are on average slightly moister (0.24 g kg$^{-1}$) from ascending airplanes than for descending flight (0.13 g kg$^{-1}$). The differences for ascending aircraft can be explained in part by hysteresis effects as ascending aircraft typically move from moister levels to drier ones (see inset in Fig. 4a) allowing some of the moist air to lag in the sensor for a short period. Because aircraft ascend at steeper rates during takeoff than during landing, the hysteresis effects could be amplified. An opposite lag would occur for descending aircraft leading to drier reports, although descending WVSS observations still show more moisture than radiosondes on average. Below approximately $\sigma_{200} = 0.98$, WVSS reports near airports are generally drier than at nearby radiosonde sites while the smaller number of descending aircraft become drier earlier (below $\sigma_{200} = 0.96$), with these differences again consistent with the same hysteresis effects.

The skewing of the overall average mean profile toward results from ascending aircraft can be traced to differences in the number of observations per layer available from ascending and descending airplanes (Figs. 4b,c). Twice as many WVSS observations are made during ascent as descent, a natural result of the varying rates at which observations are recorded for ascending and descending aircraft (100 vs 300 m, respectively). Overall, the regular structures of the systematic difference profiles should simplify the application of bias correction schemes essential to the overall impact of combining radiosonde and WVSS datasets.

c. Comparisons between seasons

The yearlong scope of this study enables the investigation of performance of the WVSS in different seasons, allowing the assessment of the WVSS sensor in environments with varying mixing ratios. Seasons are defined as December–February (DJF) for winter, March–May (MAM) for spring, June–August (JJA) for summer, and September–November (SON) for fall. Figure 5a indicates a strong seasonal signal that is obscured when only the entire year is assessed. Overall, all seasons above $\sigma_{200} = 0.95$ show the WVSS slightly moister than the radiosondes. Summer had the greatest differences with a maximum average difference over 0.45 g kg$^{-1}$ and winter with the smallest with the average differences below 0.1 g kg$^{-1}$ at all levels. During spring, when the atmosphere over the CONUS...
contains more water vapor than in winter, WVSS reports are moister than the radiosondes throughout the rest of the profile, with a maximum difference under 0.22 g kg\(^{-1}\) near the top of the boundary layer. Fall shows the WVSS being moister above with smaller positive average differences than Spring, exceeding 0.1 g kg\(^{-1}\) at several points. Near the surface, all seasons show the WVSS drier than the radiosondes. The profiles of standard deviation are similar by seasons with the smallest standard deviation profile for winter and the largest for summer.

Because the amount of water vapor present in the atmosphere varies throughout the year, it is also useful to look at the percent difference rather than the pure differences. Figure 5b shows the average percent difference for all the seasons. Percent differences relative to the average of both radiosonde and WVSS reports were calculated for each individual profile using the following formula and then averaged across the season:

\[
\text{Percent Difference} = \frac{(\text{WVSS} - \text{Radiosonde})}{(\text{WVSS} + \text{Radiosonde})/2}.
\]

This approach was chosen because it assumes neither observation is error-free. For all the seasons except fall, the lowest levels show negative differences of less than 5%, which increase with height, staying within 10% below \(\sigma_{200} = 0.6\) and then increasing above that in winter and spring as water vapor content becomes very small. During fall, the percent differences remain negative below \(\sigma_{200} = 0.8\), and then slowly become positive throughout the rest of profile, staying with 5% at all levels.

To normalize the seasonal variation in water vapor content, relative humidity differences were also examined (Fig. 5c). Similar to previous comparisons, the difference was calculated by subtracting the radiosonde relative humidity from the WVSS relative humidity. Fall and winter had the smallest differences with all differences being less than 5%. Spring and summer had the largest differences but still was majorly below 5% as well. A future publication will examine differences between WVSS and radiosondes in relative humidity more thoroughly.

d. Comparisons by radiosonde model

Because each radiosonde model that the NWS operates has its own set of observational biases, comparisons were also performed between the WVSS and each individual radiosonde model (Fig. 6). Vaisala radiosondes account for approximately

**Fig. 5.** The average of all comparison profiles by season for the CONUS (a) with differences (solid) and standard deviations (dashed) for winter (blue), spring (green), summer (yellow), and fall (red). (b) Percent difference. (c) Average relative humidity differences.
20% of the more than 7500 radiosondes used in this study with Sippican radiosondes accounting for the remainder. All reports used were processed by the NWS and did not undergo any further modification.

Figure 6a shows the WVSS differences over the full CONUS for both the Sippican and Vaisala radiosondes. Below \( \sigma_{200} = 0.7 \), the WVSS is drier than the Vaisala radiosondes. From \( \sigma_{200} = 0.7 \) through \( \sigma_{200} = 0.575 \), there is very little difference and above \( \sigma_{200} = 0.65 \), the WVSS is just slightly moister than these radiosondes. Comparisons to Sippican radiosondes show that at the surface, the WVSS is slightly drier and quickly become slightly moister throughout the remainder of the lower half of the troposphere. The Sippican comparisons very closely follow the CONUS composite in Fig. 3, reflecting the fact that Sippican radiosondes make up approximately 80% of the radiosondes used in this study.

Because there are differences in comparisons between the radiosonde models, it is useful to identify where these radiosonde models were being launched (Fig. 6b). In general, the location of the Vaisala radiosondes are along the perimeter of the CONUS, with five of the eight locations along the warmer coastal areas where more moisture resides. Qualitatively, this could have introduced a bias when compared to the Sippican radiosondes and accounted for part of the 20%–30% larger random differences seen at all levels in the Vaisala comparisons. Similar increases in bias and standard deviation differences were observed between cold (drier) and warm (moister) seasons in Fig. 5. However, the fact that many of the same planes fly into airports near both types of radiosondes (thereby providing a consistent standard for comparison between the two systems across all CONUS sites) suggests that the random observational moisture errors from Vaisala radiosondes should be expected to be larger than from Sippican reports. This result differs from than that found in WMO (2011) and deserves further investigation.

e. Upper-tropospheric comparisons

Although this paper is intended to focus primarily on comparisons of moisture observations in the lower half of the troposphere, it is also important to examine results from higher levels. Unlike low-level comparisons, comparisons between \( \sigma_{200} = 0.5 \) and aircraft flight levels (200 hPa) involve not only WVSS observations that originated from an aircraft that was taking off or landing from an airport near the radiosonde launch site but also situations where a WVSS equipped aircraft flying near the radiosonde ascent path fell within the established time, space, and time collocation requirements. These upper-tropospheric comparisons encompass \( \sigma_{200} = 0.5 \) through \( \sigma_{200} = 0 \). For a surface pressure of 1000 hPa, \( \sigma_{200} = 0.5 \) is equivalent to 600 hPa.

Figure 7c shows the upper-level average differences between the radiosondes and the WVSS. As with the lower-level comparisons, the WVSS is slightly moister than the radiosondes. The average differences increase only slightly with altitude to a maximum just above 0.25 g kg\(^{-1}\) at \( \sigma_{200} = 0.2 \), above which the average differences rapidly decrease. By contrast, random differences decrease with altitude throughout the upper troposphere due to the decrease of water vapor available in the atmosphere. The number of WVSS observations per layer (Fig. 7d) shows a steady decrease in the bottom
half of this layer then an increase above $\sigma_{200} = 0.15$ where aircraft are near cruising altitude.

In a WMO report comparing radiosonde systems, they reported "At temperatures higher than $-40^{\circ}C$, all the relative humidity sensors had good reproducibility" (WMO 2011). When the radiosonde temperature sensors are below $-40^{\circ}C$, the response time of the relative humidity sensors becomes longer and causes less accurate measurements. Figure 7b shows that the average temperature profile from the radiosondes drops below $-40^{\circ}C$ near $\sigma_{200} = 0.12$ ($-300$ hPa). A simple filtering procedure was applied that removed radiosonde-based observations (7a) that exceeded the possible saturation mixing ratio (calculated with respect to water) calculated using the following formula for saturation vapor pressure (Petty 2008):

$$w_s(T, p) = \varepsilon_e e^{\frac{T_{\text{Radiosonde}}}{243.5}}$$

where $\varepsilon_e = 0.622$, $p$ is the pressure, and the vapor pressure with respect to water was calculated with

$$e_s(T) = 6.112 \times T_{\text{Radiosonde}}^{0.378}$$

where $T_{\text{Radiosonde}}$ is the temperature of the radiosonde in degrees Celsius. The equations used for this calculation (as well as the equations for the original conversion from relative humidity to water vapor mixing ratio for the radiosondes) are accurate with any calculations over $-40^{\circ}C$ and are accurate within a few percent below $-40^{\circ}C$ down to $-70^{\circ}C$. However, at low temperatures, the choice of saturation vapor pressure equation does matter and future work to investigate the sources of differences between the radiosonde and WVSS reports should be done using equations described Wexler (1976) or Sonntag (1994) as recommended by WMO (2011, 2018).

The filtered data were used for the comparisons shown in Fig. 7c. From WMO (2011), "The upper troposphere and lower stratosphere present the greatest challenge to radiosonde relative humidity measurement, because of the low temperatures involved, down to $-80^{\circ}C$ in Yangjiang, China, and because of the problems of contamination of water vapour picked up passing through upper cloud, or in the lower layers of the atmosphere at warmer temperatures" and "Unlike temperature measurements, the best performance is often found in the day, suggesting that contamination after passing through moist levels is worse at night." These statements show the difficulties of radiosonde observations and associated calculations of water vapor at these levels could be contaminated from passing through clouds. The fact that an additional filtering procedure was necessary to remove questionable radiosonde reports in this comparison provides further evidence that WVSS observations of moisture made directly at the molecular level may be more reliable, especially at the highest aircraft altitudes.

**f. Comparisons by regions**

WVSS reports were also separated into several geographical groups to allow regional differences to be examined in areas
visited by WVSS-equipped aircraft. Four regions were used (Fig. 8b) that consisted of climatologically similar locations from the Köppen–Geiger climate classification scheme (Kottek et al. 2006) and contained enough intercomparison locations to be scientifically useful. The western region included four locations from the warm temperate, dry hot summers (Csa), warm temperate, dry warm summers (Csb), and snow, dry warm summers (Dsb) regions including Oakland, California; and Spokane, Washington. The mountain/desert region included 6 locations from the cold arid, steppe (BSk) and cold arid, desert (BWk) regions including Denver; Las Vegas, Nevada; and Albuquerque, New Mexico. The eastern regions consisted of 16 locations composed from warm temperate, fully humid, hot summer (Cfa) and snow, fully humid, warm summer (Dfb) regions. This region was subsequently separated into the northeast and the southeast regions by visual inspection to create two smaller groups because of its larger spatial extent. The northeastern region was composed of Buffalo, New York; Chanhassen, Minnesota; and four other cities while the southeastern region was composed of Dallas, Texas; Jacksonville, Florida; and eight other cities. Regional comparisons were made by identifying the location within the region with the lowest number of comparisons and randomly selecting that number of comparison profiles from each of the other locations within the region to prevent any bias toward one location within the region. All selected comparisons were then averaged across the region to form the regional average. Due to the fact that more locations are located in the southern half of the CONUS than the northern half, overall statistics mentioned before may have been skewed toward the southern states. Because the maximum possible moisture content is larger farther south, the range of differences could be larger there as well.

The regional results (Fig. 8a) show increased vertical variability compared to the results for the WVSS–radiosonde comparisons made over the full CONUS. The southeastern region had the largest average differences of the four regions at $\sigma_{200} = 0.95$ through $\sigma_{200} = 0.85$ which would be due in part to the fact that this was the most climatologically diverse region of the four. The WVSS in the mountain/desert region and the western region were drier than radiosondes at the surface through approximately $\sigma_{200} = 0.9$ but then became moister than the radiosondes similar to the eastern region above. Around $\sigma_{200} = 0.7$, all regions converge and have average differences about the same. For the western region above $\sigma_{200} = 0.7$, the average differences become drier and have negative average differences between $\sigma_{200} = 0.65$ and $\sigma_{200} = 0.57$. However, the WVSS was almost always within 0.25 g kg$^{-1}$ of the radiosonde no matter which region or where in the atmosphere the difference is computed; the only exception is the southeast region in the lower half of the troposphere. Considering the climatic variability of very moist in the east, to very dry in the mountain/desert and west regions the small average differences found between the WVSS and the radiosondes in all these regions show that the WVSS observations are compatible with conventional in situ moisture observations all environments.

g. Comparisons by aircraft type

UPS and Southwest are the two largest carriers of WVSS, and the differences between these two carriers in terms of flight patterns and airframe mounting practices are also significant.
enough to warrant analysis. First, each carrier has mounted the sensor to different aircraft models: UPS has deployed them on their Boeing 757 fleet, while Southwest exclusively flies the Boeing 737. The different aircraft could experience different airflow patterns which could impact air sampling and induce biases. In addition, UPS flights use much sharper takeoff and landing profiles for its freight-only flights. These sharper profiles mean that the temporal rate of change of the moisture environments during ascent and descent are much greater for UPS aircraft. Finally, UPS flights tend to operate at night to facilitate overnight shipping and to enjoy reduced traffic at airports. However, this means that these aircraft are sampling different boundary layer conditions than they would if they flew during the day. This section examines the overall average differences of all UPS and non-UPS profiles, the ascent and descent differences of both, and the average differences in the profiles at 0000 and 1200 UTC.

Because publicly accessible AMDAR datasets contain an encrypted unique identifier for each individual airplane rather than airline and aircraft tail numbers to maintain corporate security, it is not possible to directly determine which airline has produced a particular observation. Therefore, an alternative method of discriminating between airlines had to be developed for this study. The largest hub for UPS operations is Louisville, where on most nights hundreds of aircraft converge for a brief period so that packages can be exchanged in time for next-day deliveries. This happens during the overnight hours when very few airplanes from other airlines are arriving or departing from Louisville. To identify UPS airplanes, the total number of flights that each different aircraft identifier made arriving at or departing from Louisville between the hours of 0500 and 0900 UTC was examined for the full year. The top three quartiles of airplanes making those flights were considered to be from UPS. It is possible that this technique to distinguish UPS from non-UPS aircraft does not capture all of the aircraft correctly, but this subset of data will be largely different from the total CONUS dataset.

The number of comparison profiles for the UPS aircraft are approximately 10% of the non-UPS comparisons, while UPS aircraft account for about 22% of the total WVSS equipped aircraft in the United States (Brusky and Mamrosh 2015). The difference in percentage of UPS profiles and aircraft is likely due to the fact that this study is limited to use WVSS profiles where radiosondes are launched nearby and UPS major hubs at Louisville and Rockford, Illinois, are outside of these limits. As a result, matches between colocated radiosonde and non-UPS reports outnumber those from UPS and therefore dominate any of the combined statistics. In addition, Southwest (i.e., non-UPS) aircraft also make more ascents and descents in a day than a typical UPS plane.

Overall, the average difference between aircraft and radiosonde moisture reports from UPS aircraft are less positive than the non-UPS aircraft (Fig. 9a). Below \( \sigma_{200} = 0.82 \), the WVSS on UPS aircraft are systematically slightly drier than the radiosondes while non-UPS aircraft are only drier just immediately near the surface, above which they remain slightly moister than the radiosondes at all levels. By contrast, the UPS reports become only slightly moister than radiosondes above \( \sigma_{200} = 0.82 \) with values about half those of non-UPS aircraft. The random differences from the non-UPS aircraft are consistently less than from UPS aircraft, with reductions of 20%–30% throughout the bottom third of the troposphere.

Ascending and descending UPS/non-UPS aircraft were also examined separately as a means of investigating differences in hysteresis effects associated with airframe and/or WVSS instrument mounting location (Fig. 9b) on the UPS Boeing 757s and Southwest 737s. The results show that on average WVSS reports from ascending UPS aircraft are slightly drier than the radiosondes close to the surface but then quickly becomes moister than the radiosondes at all levels above \( \sigma_{200} = 0.98 \), reaching maxima near 0.30 g kg\(^{-1}\). The differences for non-UPS aircraft are very similar. During descent, the non-UPS aircraft exhibit a small reduction in the moist bias relative to nearby radiosonde observations, another indication of the hysteresis effects noted earlier. For descending UPS aircraft, the WVSS moisture profiles are much drier than the radiosondes throughout the entire profile, possibly influences of faster descent rates used by the UPS freight aircraft. Although reliable calculations of aircraft ascent and descent rates were not possible with the available dataset, both the enhanced rate at which average difference go from negative to positive during takeoff and negative bias during descent provide evidence of amplified hysteresis effects during more rapid ascents and descents.

In addition to addressing the influences of aircraft type, the dataset also fostered investigation of diurnal incompatibilities between WVSS platforms and nearby radiosonde reports. The average differences between the 0000 and 1200 UTC comparisons were also computed excluding the low number of asynoptic launches (Fig. 9c). Twice as many UPS profiles are available at 1200 UTC compared to 0000 UTC while the opposite is true for non-UPS flights. Overall, the average differences for UPS comparisons made at 0000 and 1200 UTC show no marked differences, although both profiles reflect a dry hysteresis in the boundary layer noted above. The lack of differences between the average comparison profiles from both UPS and non-UPS aircraft show that time of day does not affect how the WVSS responds. Overall, average differences between WVSS sensors mounted on UPS 757s and the other aircraft (most likely Southwest 737s) support the robust nature of WVSS observations.

4. Summary

With the development and rollout of the WVSS sensor on AMDAR equipped aircraft, water vapor measurements have been added to the routine observations of pressure, wind, and temperature taken by nearly 150 commercial aircraft each day, most of which are reported as profiles during aircraft ascent and descent. Unlike previous studies which used specialized research radiosonde observations as a comparison standard, the present work examined how the sensor performed in different regions and seasons by using operational NWS radiosondes. Comparisons were made throughout the troposphere (up to 200 hPa) and were presented using a terrain-following vertical coordinate to assure that all of the
higher-vertical-resolution WVSS in the boundary layer were viewed together. Comparisons were also conducted to investigate any differences for ascending and descending flights, radiosonde models, and aircraft/airline type. The results showed that WVSS moisture profiles reported by commercial aircraft during takeoff and landing are fully compatible with nearby radiosonde observations.

Comparisons from the lower half of the troposphere show the greatest agreement between the WVSS and the radiosondes just above the surface, but with only small average differences throughout the rest of the profile. The WVSS observations were in general slightly moister than the radiosondes, but with differences no greater than 0.2 g kg\(^{-1}\). Random differences were largest near the surface and reduced from 1.3 to 0.9 g kg\(^{-1}\) aloft. Similar to previous studies, the WVSS observations appeared drier (with smaller differences vs the radiosondes) on descending aircraft than on ascending ones, although still slightly moister than nearby radiosonde observations. These differences are likely hysteresis effects caused by residual moisture remaining in the sensor as ascending aircraft fly from moister to drier levels with the opposite happening for descending aircraft. Differences are also seen when the different radiosonde models are compared to the WVSS, with Vaisala radiosondes moister than the WVSS in the bottom quarter of the troposphere. The Sippican radiosondes were only positive near the surface, but then became consistently drier throughout most of the lower troposphere. The upper-tropospheric comparisons showed a similar trend of the WVSS being moister than the radiosondes. Below ~40°C, radiosonde humidity sensors suffer from slower response and occasional contamination problems; however, WVSS may also suffer from overreading of small humidity amounts in some instances. Regional comparisons showed differences between the southeastern, northeastern, mountain/desert, and western regions. The mountain/desert and western regions were in general agreement at most levels, while the WVSS reports from the eastern regions were moister than nearby radiosondes in the lowest quarter of the atmosphere. WVSS aboard UPS aircraft showed smaller average differences above 500 hPa than non-UPS aircraft. For both UPS and non-UPS aircraft, ascending observations were moister than the radiosondes but only UPS descending observations were drier. Time of day did not appear to affect the average differences for WVSS aboard UPS aircraft.

A number of factors could be responsible the differences between the WVSS and the radiosonde reports. Because water vapor is highly variable in space and time, gradients could still be present within the 50 km and 10 hPa collocated boundaries. As demonstrated by Petersen et al. (2016), if the WVSS and the radiosondes are sampling on different sides of the gradient, the result would produce large local differences. Already known biases within the radiosondes could also contribute to the

**Fig. 9.** (a) CONUS composite of comparison profiles using only WVSS-II observations for UPS aircraft (blue solid), UPS standard deviation (blue dashed), non-UPS aircraft (red solid) and standard deviation (red dashed). (b) Comparison profiles split up by observations from ascending UPS aircraft (blue solid), descending UPS aircraft (blue dashed), ascending non-UPS aircraft (red solid), and descending non-UPS aircraft (red dashed). (c) Comparison profiles from UPS aircraft at 0000 (blue solid) and 1200 UTC (blue dashed) and non-UPS aircraft at 0000 (red solid) and 1200 UTC (red dashed).
differences when compared to the WVSS. Fleming and May (2004) and Helms et al. (2010) both noted that the WVSS laser diode technology has a very fast response time which is needed for the quickly changing environments aircraft experience. Another study, Vance et al. (2015), compared the WVSS using two different air intakes to several chilled mirror hygrometers, which showed that the WVSS responded faster than the chilled mirror devices, however, when comparing the standard flush inlet (used by UPS and Southwest aircraft), it did overread except at higher humidities. This study also found that from the two ways to mount the WVSS, the standard flush inlet and the Rosemount inlet, the standard flush inlet was not susceptible to either liquid or solid water from cloud contamination and that it responded faster than the Rosemount inlet.

A future publication by Wagner and Petersen examines the differences in relative humidity to normalize comparisons over seasons and different locations. Additional future work includes examining WVSS experiencing any form of precipitation since excess moisture within the sensor could lead to increased differences when compared to radiosondes. The examination of the differences of temporal and spatial distance between the radiosondes and WVSS observations should also be completed to investigate the moisture variability in more depth and see how the differences change. By reducing the temporal and spatial extent of the comparison criteria, confidence that the two systems are observing identical conditions increases but the number of valid comparisons necessarily decreases. This would confirm whether or not the current differences are due to the sampling of different moisture environments. Additional future work could involve calculating the total precipitable water from these moisture profiles to use for validation of satellite observations and for future climate studies.

Overall, it has been shown that WVSS are compatible and agree well with operational radiosonde observations over the CONUS throughout the entire troposphere. WVSS observations have already been found to be impactful for forecasters and NWP. Hoover et al. (2017) found the assimilation of AMDAR moisture observations improved forecasts and the observation-minus-background bias and RMSE was lowered in the warm season experiment and James and Benjamin (2017) found that aircraft water vapor observations from WVSS were responsible for over 50% of the relative humidity forecast improvement for aircraft observations overall in the hourly updated Rapid Refresh NWP assimilation forecasts. Increasing deployment of this sensor would allow the continuation of gap-filling in our upper-air observation network, and greater NWP skill—all for less cost and greater accuracy than by expanding the radiosonde network.

Acknowledgments. The authors would like to thank the four anonymous reviewers for their formal reviews, which improved the manuscript. Funding provided by the National Weather Service Aircraft Based Observations Program through the Cooperative Institute for Meteorological Satellite Studies under Cooperative Agreement NA15NES4320001.

Data availability statement. WVSS data used were provided by NCAR/EOL under the sponsorship of the National Science Foundation (available at https://data.col.uncar.edu/). NWS radiosonde data used were accessed via NCEI (available at ftp://ftp.ncdc.noaa.gov/pub/data/rss-data/).

REFERENCES


