Evaluation of Regional Aircraft Observations Using TAMDAR

WILLIAM R. MONINGER, STANLEY G. BENJAMIN, BRIAN D. JAMISON,* THOMAS W. SCHLATTER,[+]
TRACY LORRAINE SMITH,* AND EDWARD J. SZOKE*

NOAA/Earth System Research Laboratory, Boulder, Colorado

ABSTRACT

A multiyear evaluation of a regional aircraft observation system [Tropospheric Aircraft Meteorological Data Reports (TAMDAR)] is presented. TAMDAR observation errors are compared with errors in traditional reports from commercial aircraft [aircraft meteorological data reports (AMDAR)], and the impacts of TAMDAR observations on forecasts from the Rapid Update Cycle (RUC) over a 3-yr period are evaluated. Because of the high vertical resolution of TAMDAR observations near the surface, a novel verification system has been developed and employed that compares RUC forecasts against raobs every 10 hPa; this revealed TAMDAR-related positive impacts on RUC forecasts—particularly for relative humidity forecasts—that were not evident when only raob mandatory levels were considered. In addition, multiple retrospective experiments were performed over two 10-day periods, one in winter and one in summer; these allowed for the assessment of the impacts of various data assimilation strategies and varying data resolutions. TAMDAR's impacts on 3-h RUC forecasts of temperature, relative humidity, and wind are found to be positive and, for temperature and relative humidity, substantial in the region, altitude, and time range over which TAMDAR-equipped aircraft operated during the studied period of analysis.

1. Introduction

As of late 2009, commercial aircraft provide more than 239 000 observations per day of wind and temperature aloft worldwide (Fig. 1). The general term for these data is aircraft meteorological data reports (AMDAR). These data have been shown to improve both short- and long-term weather forecasts and have become increasingly important for regional and global numerical weather prediction (Moninger et al. 2003). Figure 2 shows the AMDAR coverage over the contiguous United States.

Two shortfalls of the current AMDAR dataset have been the near absence of data below 20 000 ft between major airline hubs (Fig. 3) and the almost complete absence of water vapor data at any altitude. To address these deficiencies, a sensor called the Tropospheric AMDAR (TAMDAR), developed by AirDat, LLC, under the

sponsorship of the National Aeronautics and Space Administration's (NASA) Aviation Safety and Security Program, was deployed on approximately 50 regional turboprop commercial aircraft flying over the north-central United States and lower Mississippi Valley (Daniels et al. 2006). These turboprops are operated by Mesaba Airlines (doing business as ''Northwest Airlink''). The aircraft cruise at lower altitudes (generally below 500 hPa) than traditional AMDAR jets, and fly into regional airports not serviced by AMDAR-equipped jets. Figure 4 shows TAMDAR data along with traditional AMDAR data, and shows how TAMDAR fills in the region between major hubs in the U.S. midwest. For example, in the Great Lakes region, traditional AMDAR-equipped aircraft serve 23 airports—providing ascent and descent atmospheric soundings at each, while TAMDAR-equipped aircraft serve 62 airports.

Like the rest of the AMDAR fleet, TAMDAR measures wind and temperature. But unlike most of the rest of the fleet, TAMDAR also measures humidity, turbulence, and icing. [The Water Vapor Sensing System-II (WVSS-II) sensor (Helms et al. 2005) also provides water vapor measurements from several commercial aircraft and is scheduled to expand substantially in the near future. But the then-current version of the WVSS-II provided
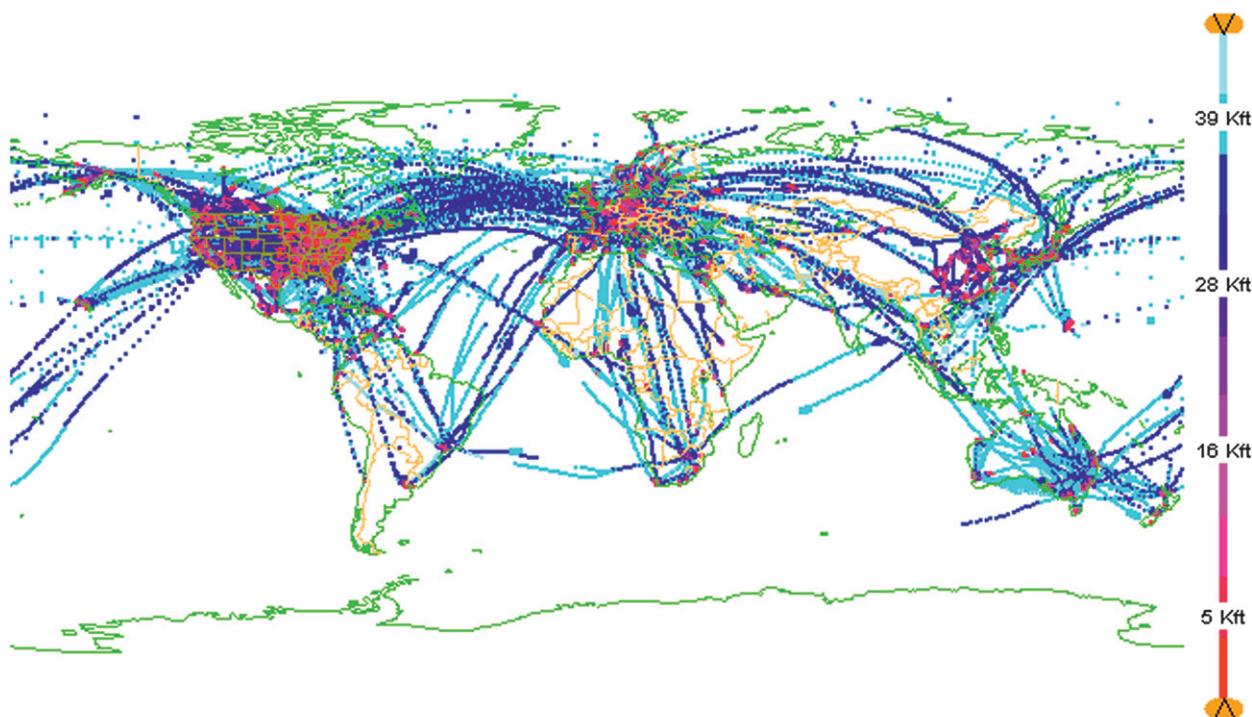
* In collaboration with the Cooperative Institute for Research in the Atmosphere, Fort Collins, Colorado.
[+] Retired.

Corresponding author address: William R. Moninger, ERSL, 325 Broadway, Boulder, CO 80304.
E-mail: bill.moninger@noaa.gov

13-Oct-2009 00:00:00 -- 13-Oct-2009 23:59:59 (288074 obs loaded, 239865 in range, 18260 shown)

NOAA / ESRL / GSD   Altitude: -1000 ft. to 45000 ft.   Good w and T

FIG. 1. Worldwide AMDAR reports, Tuesday, 16 Oct 2009. There were 239 865 observations of wind and temperature.

relatively few reliable water vapor measurements during the time period studied here.]

The National Oceanic and Atmospheric Administration/ Earth System Research Laboratory/Global Systems Division (NOAA/ESRL/GSD) has built an extensive system for evaluating the quality of TAMDAR and AMDAR data, and has applied this system for the 4 yr that TAMDAR has been in operation. This evaluation system relies on the Rapid Update Cycle (RUC) numerical model and data assimilation system (Benjamin et al. 2004a,b, 2006a).

Under FAA sponsorship, NOAA/ESRL/GSD performed careful TAMDAR impact experiments. The RUC is well suited for regional observation impact experiments due to its complete use of hourly observations and diverse observation types.

## 2. RUC experiments to study TAMDAR data quality and forecast impacts

Between February 2005 and December 2008, we ran two real-time, parallel versions of the RUC with the following properties:

- Dev (or ''development version 1'') assimilated all hourly non-TAMDAR observations.

- Dev2 is the same as dev but also assimilated TAMDAR wind, temperature, and relative humidity observations.

- The same lateral boundary conditions, from the National Centers for Environmental Prediction's (NCEP's) North American Model (NAM; Rogers et al. 2009), were used for both dev and dev2 runs.

- These RUC experiments are run at 20-km resolution, but using more up-to-date 13-km-version code.

In February 2006 and subsequently in April 2007, the analysis and model code in the dev–dev2 versions of the RUC used for the TAMDAR impact experiments were upgraded to improve the observation quality control and precipitation physics. These modifications were generally the same as those implemented into the operational NCEP 13-km RUC, with the exception that dev and dev2 do not ingest radar data (implemented in the NCEP RUC in November 2008).

The studies herein focus on these real-time model runs and, also, on retrospective runs (also at 20-km resolution) over two 10-day periods: one in winter and one in summer. These same test periods were used in a broader set of observation sensitivity experiments (OSEs) for eight different observation types described by Benjamin et al. (2010).
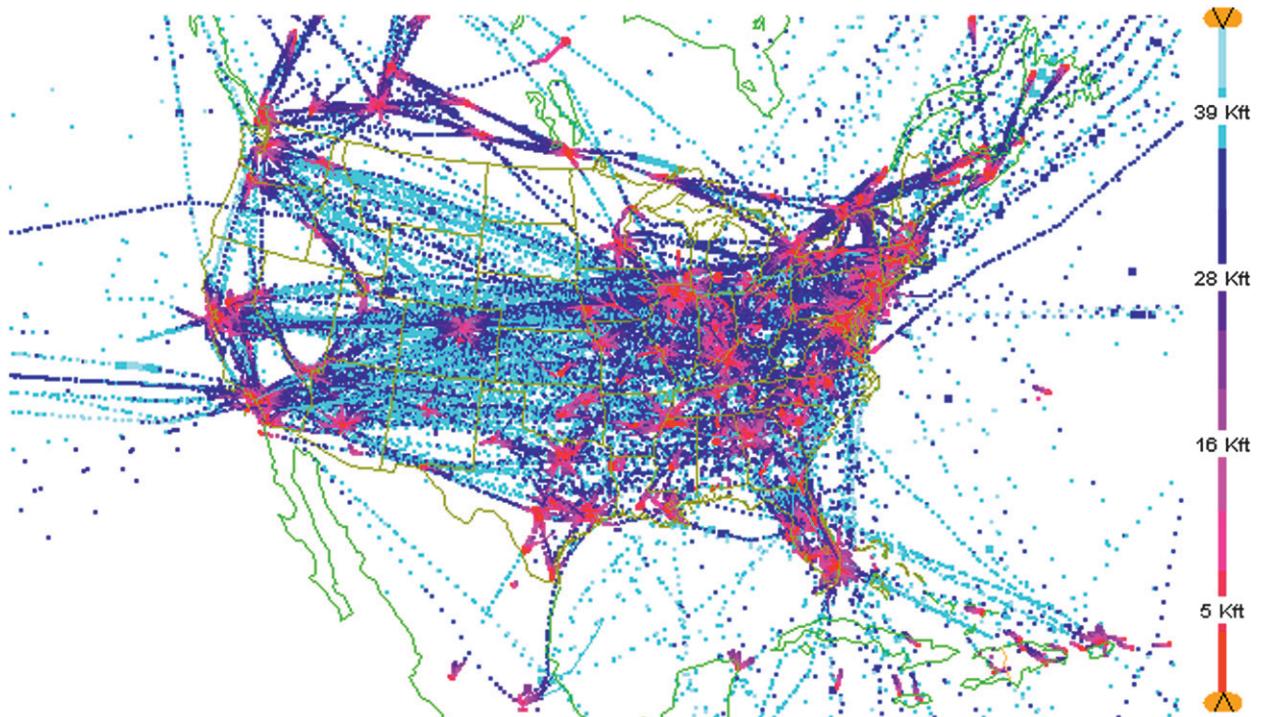
13-Oct-2009 00:00:00 -- 13-Oct-2009 23:59:59 (288074 obs loaded, 126665 in range, 28849 shown)

NOAA / ESRL / GSD   Altitude: -1000 ft. to 45000 ft.   Good w and T not-TAMDAR

FIG. 2. Traditional AMDAR (i.e., non-TAMDAR) reports over the contiguous United States, on Tuesday, 13 Oct 2009, for a total of 126 665 observations.

The 20-km RUC version used for the TAMDAR experiments includes complete assimilation of nearly all observation types (as used in the operational RUC), including cloud analysis [Geostationary Operational Environmental Satellite (GOES) and aviation routine weather reports (METARs)], full METAR assimilation (temperature, dewpoint, winds, pressure, cloud, visibility), GPS precipitable water, GOES precipitable water, all other aircraft, profilers, mesonets, and raobs. A summary of the characteristics of the June 2006 operational RUC is available online (http://ruc.noaa.gov/ruc13_docs/RUC-testing-Jun06.htm). More details on the RUC assimilation cycle and the RUC model are available in Benjamin et al. (2004a,b). Other details on the RUC TAMDAR experimental design are described in Benjamin et al. (2006a,b, 2010).

3. TAMDAR data quality

To evaluate the *quality* (as opposed to the *model forecast impacts*) of TAMDAR, ESRL/GSD maintains a database of AMDAR (including TAMDAR) observations, and 1-h forecasts interpolated to the AMDAR observation point from the RUC dev and dev2 cycles.

This allows us to calculate the mean and RMS differences between RUC 1-h forecasts and aircraft-observed temperature, wind, and relative humidity.

Model data are interpolated vertically (linear in $\log p$) and horizontally to the locations of the observations. No temporal interpolation is performed; observations are compared with the 1-h forecast valid at the nearest hour.

For each observation time and location, we store the observed and forecasted temperature, relative humidity, wind direction and speed, and phase of flight (ascent, descent, or en route). In addition, the RUC quality control disposition of each observation (independent QC for each variable) was stored between December 2005 and December 2008, as well as which variable(s) were actually used in the RUC analysis.

a. Web-based access to the AMDAR-RUC database

Access to the AMDAR-RUC database is available online (http://amdar.noaa.gov/RUC_amdar/). Because access to real-time (i.e., less than 48 h old) AMDAR data is restricted to NOAA and selected other users, access to the real-time portions of this site is restricted (information online at http://amdar.noaa.gov/FAQ.html).

Database access is provided in the following forms:

13-Oct-2009 00:00:00 -- 13-Oct-2009 23:59:59 (288074 obs loaded, 78754 in range, 7338 shown)

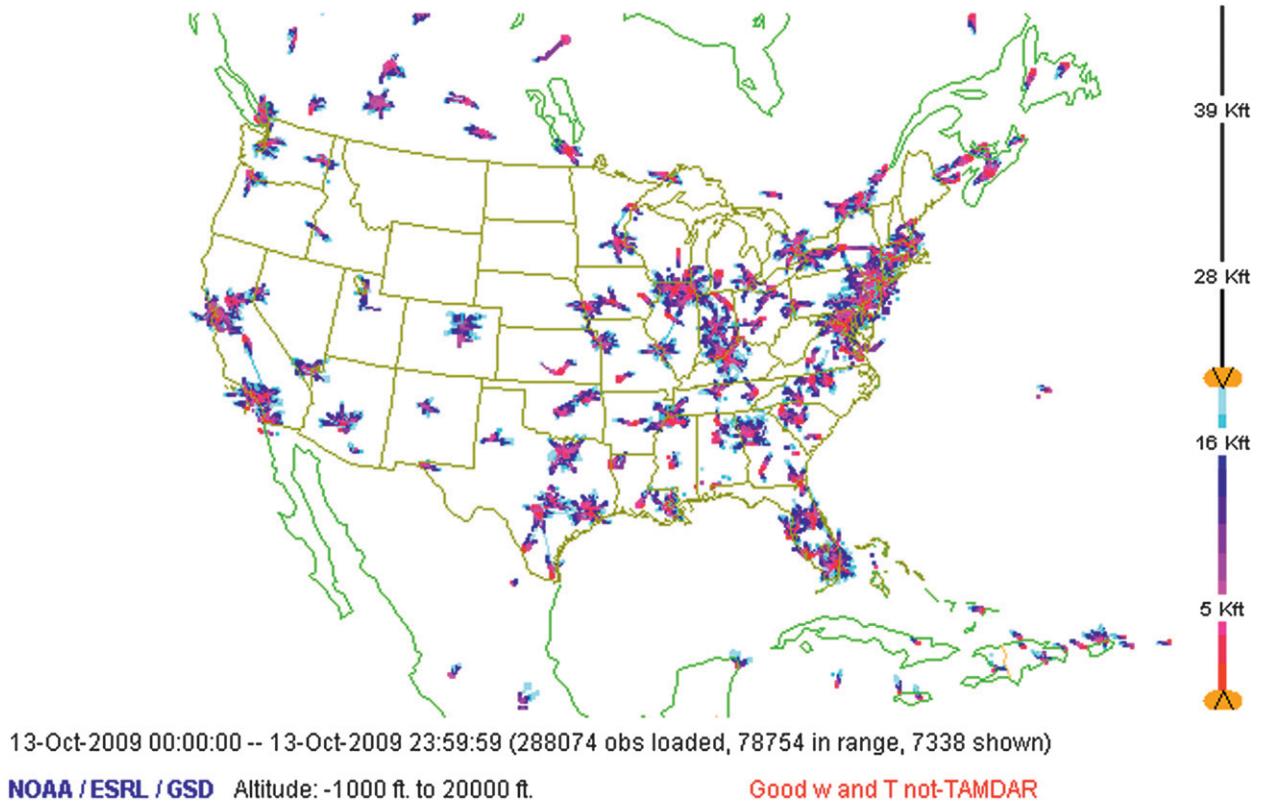NOAA / ESRL / GSD   Altitude: -1000 ft. to 20000 ft.   Good w and T not-TAMDAR

FIG. 3. As in Fig. 2 but below 20 000 ft. There were 78 754 observations.

- 7-day statistical summaries for each aircraft for four altitude ranges (all altitudes, surface to 700 hPa, 700–300 hPa, and above 300 hPa), sortable by a variety of values, and

- time series data for any aircraft (restricted).

b. Error characteristics of the TAMDAR/AMDAR fleet

In this section we look at aircraft differences with respect to the RUC dev2 cycle. We do not consider the RUC dev2 1-h forecasts to be "truth"; rather, we use them to be a common benchmark (assimilating all aircraft types) with which to compare the error characteristics of various aircraft fleets. The 1-h forecast through the RUC forecast model and initialization (Benjamin et al. 2004a,b) forces some (but not total) independence from any particular observation type. We focus on 1–30 October 2006, a reasonably representative month in terms of RUC forecast error and TAMDAR impacts, as will be discussed further in section 4b.

We look at aircraft–RUC differences over the TAMDAR Great Lakes region (the small rectangle shown in Fig. 5), which includes the upper midwestern region of the United States, for "daylight" hours (1200–0300 UTC) when TAMDAR-equipped aircraft generally fly.

In our analyses of aircraft–RUC differences, we found it useful to stratify the data by phase of flight (descent and en route/ascent) as well as altitude. There are enough TAMDAR data that each point we show in this section is the average of at least 100 observations; in most cases, especially lower in the atmosphere, each data point represents the average of more than 1000 observations.

Temperature bias relative to the RUC 1-h forecast for traditional AMDAR jets and TAMDAR turboprops is shown in Fig. 6. The jets show a small warm bias at most altitudes peaking at 0.3–0.4 K between 800 and 500 hPa, and descents (blue) show less warm bias than enroute–ascent (red) data above 600 hPa. Below 800 hPa, descents show a slightly warmer bias than ascents for this time period. TAMDAR shows a smaller temperature bias (nearer zero) than AMDAR from 800 to 500 hPa. In general, both AMDAR and TAMDAR temperature biases with respect to the RUC are small—less than 0.4 K in absolute magnitude.

The temperature RMS difference from RUC 1-h forecasts for TAMDAR and AMDAR (Fig. 7) is about 1 K at most levels, with TAMDAR RMS being generally equivalent to that of AMDAR jets. Some of this difference is attributable, of course, to RUC forecast error, which would affect TAMDAR and AMDAR equally.
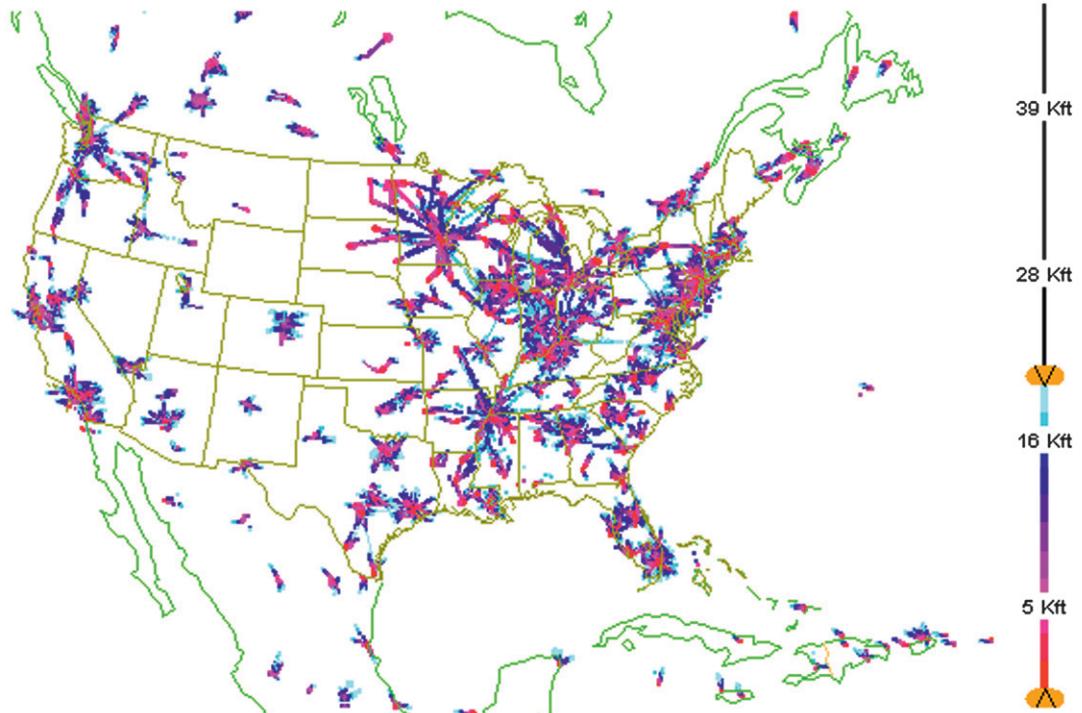
FIG. 4. As in Fig. 3 but with TAMDAR observations included. There were 93 207 observations.

The RMS vector wind differences between aircraft-measured winds and RUC 1-h forecast winds (Fig. 8), in contrast to temperature, are considerably larger for TAMDAR (turboprops) than for AMDAR jets, and TAMDAR's differences on descent are larger than those on ascent and en route. The lower quality of the wind data from TAMDAR is likely due to the less accurate heading information provided to TAMDAR by the Saab-340b avionics system. Accurate heading information is required for the wind calculation, and the Saab heading sensor is magnetic and known to be less accurate than the heading sensors commonly used on large jets.

The greater error on descent is due, we believe, to aircraft maneuvers, which occur more often on descent than on ascent. In response to this, we eliminated TAMDAR wind measurements taken on descent in the RUC experiments described here. Also, in our current versions of the RUC, we have also implemented a larger observational error estimate for TAMDAR turboprop winds in the RUC analysis, where code allows for different error estimates for each aircraft fleet.

We also examine the relative humidity bias (observation − forecast; Fig. 9) for TAMDAR but not for other aircraft, because most traditional AMDAR jets do not measure moisture. [A few WVSS-II moisture sensors

(Helms et al. 2005) were flying at this time; we do not consider them in our analysis.] The humidity bias is generally below 5% RH.

The relative humidity RMS differences for TAMDAR from RUC 1-h forecasts (Fig. 10) are generally similar for the ascent/en route versus descent reports, and increase from ~9% RH near the surface to ~20% RH at 500 hPa. To put this statistic in perspective, the assumed raob RMS observational error used by the North American Mesoscale (NAM) model, run operationally at NCEP in its assimilation cycle (D. Keyser 2006, personal communication), is shown in black. This error varies from ~8% RH near the surface to ~16% above 600 hPa. Note that assumed RH errors for raobs (often taken as a rough data standard, reflecting the observation error only from the measurement and spatial representativeness) do not differ greatly from the RH errors shown by TAMDAR (reflecting the combination of TAMDAR observation error, representativeness error, and RUC 1-h forecast error).

We summarize this section as follows:

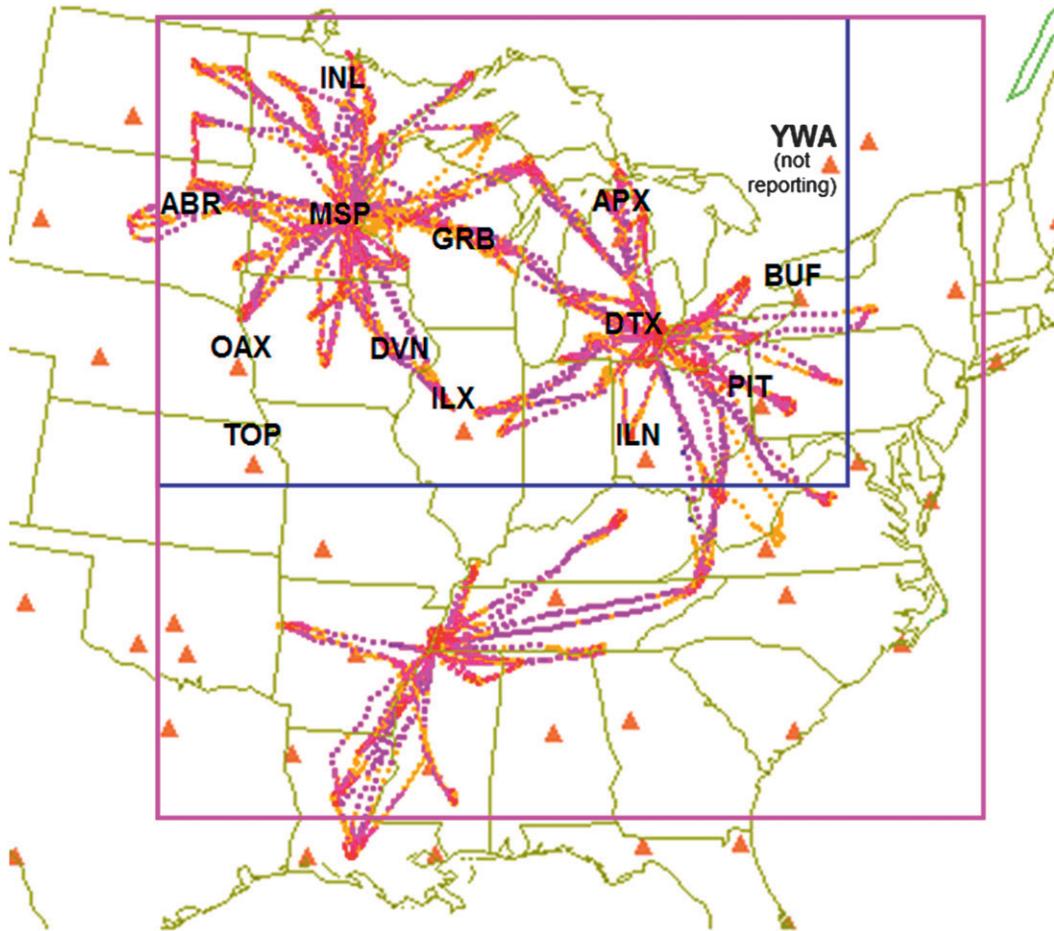- Temperature measurements from TAMDAR turboprops and AMDAR jets are approximately equally accurate.

FIG. 5. TAMDAR observations typical for a 24-h period in 2007. Verification areas are shown by the blue rectangle (Great Lakes region, 13 raobs) and magenta rectangle (eastern U.S. area, 38 raobs).

- Wind measurements from TAMDAR turboprops are worse than for AMDAR jets, due to less-accurate heading information from the turboprops. (The Saab-340B turboprops flown by Mesaba, like many turboprop aircraft, use flux-gate heading sensors, which are less accurate than the laser-gyroscopic sensors used on most jets.)
- On descent, wind measurements from TAMDAR turboprops are significantly degraded over those taken on ascent, likely due to enhanced heading errors during maneuvers.
- Relative humidity measurement errors from TAMDAR are commensurate with assumed errors for raobs.

4. TAMDAR's impact on RUC forecasts

The forecast skill of the RUC is evaluated against raobs. Figure 5 shows the specific regions for which we generate results (the eastern United States and the Great Lakes).

In studying TAMDAR's impacts, we take a two-pronged approach. First, we consider the two real-time RUC cycles discussed above (dev and dev2), to see the long-term stability and trends. Second, we consider two 10-day intensive study periods: one in winter and one in summer. For each of these periods, we performed several retrospective experiments, with and without TAMDAR, and with varying data assimilation strategies. These experiments complement a broader set of observation impact experiments for the same summer and winter retrospective periods described in Benjamin et al. (2010).

a. Forecast verification procedure

In 2006 we developed a new raob verification procedure for these evaluations. Under the previous verification procedure, the following conditions applied:

- RUC–raob comparisons were made only at mandatory sounding levels (850, 700, and 500 hPa in the TAMDAR altitude range).
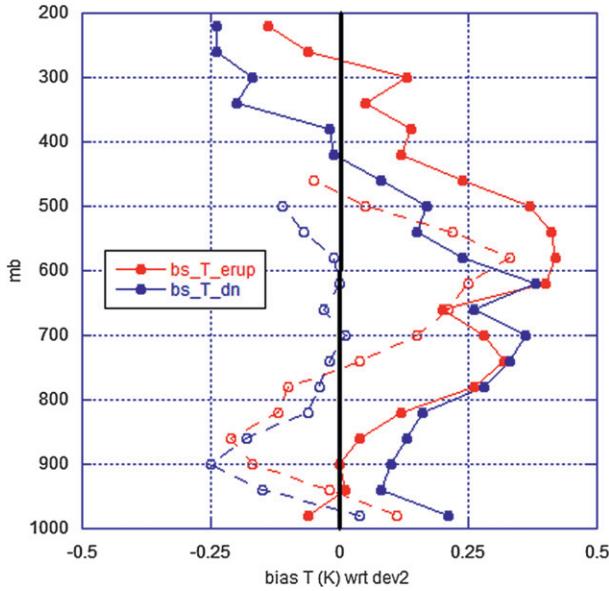
FIG. 6. TAMDAR (open circles) and AMDAR (solid circles) temperature "biases" (aircraft − RUC dev2 1-h forecasts) for October 2006. Observations taken during descent are shown in blue and labeled with the suffix "_dn"; observations taken during ascent and en route are shown in red and labeled with the suffix "_erup."

- Verification used RUC data interpolated horizontally and vertically to 40-km pressure-based grids from the RUC native coordinate (isentropic-sigma 20 km) data.
- Raob data that failed quality control checks in the operational RUC analyses were not used.
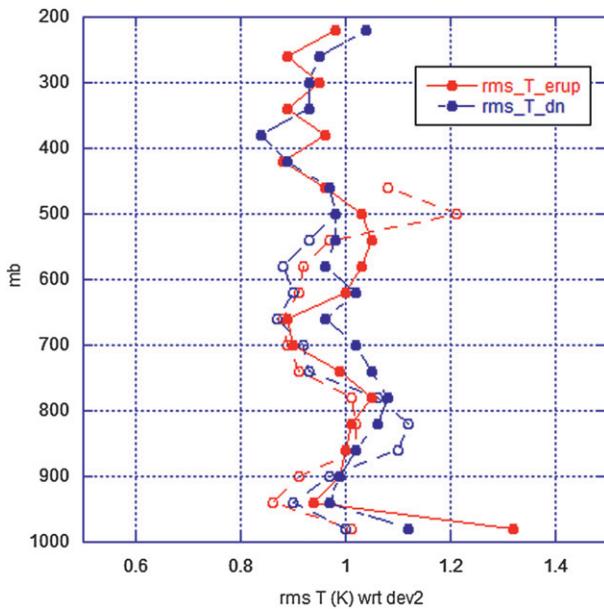
FIG. 7. As in Fig. 6 but for the temperature RMS difference from the RUC dev2 1-h forecasts for October 2006.
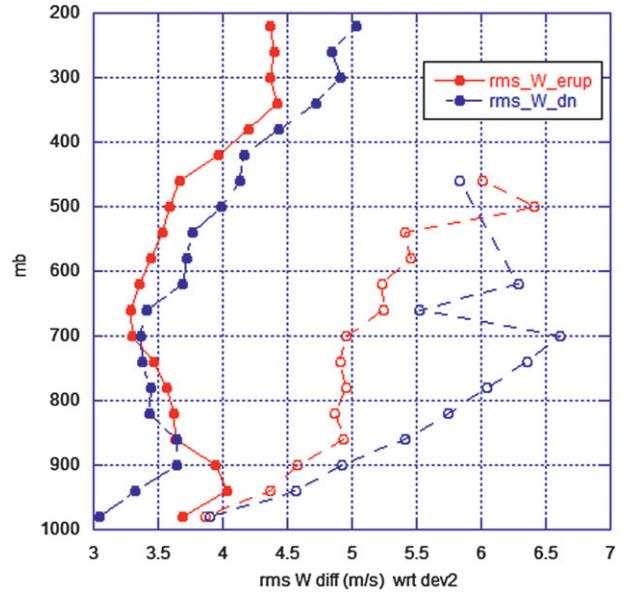

FIG. 8. As in Fig. 6 but for the RMS wind vector differences.

Under the new verification system, the following conditions apply:

- Full raob soundings, interpolated to every 10 hPa, are compared with model soundings.
- Model soundings, interpolated to every 10 hPa, are generated directly from RUC native files (20-km resolution, isentropic-sigma native levels).
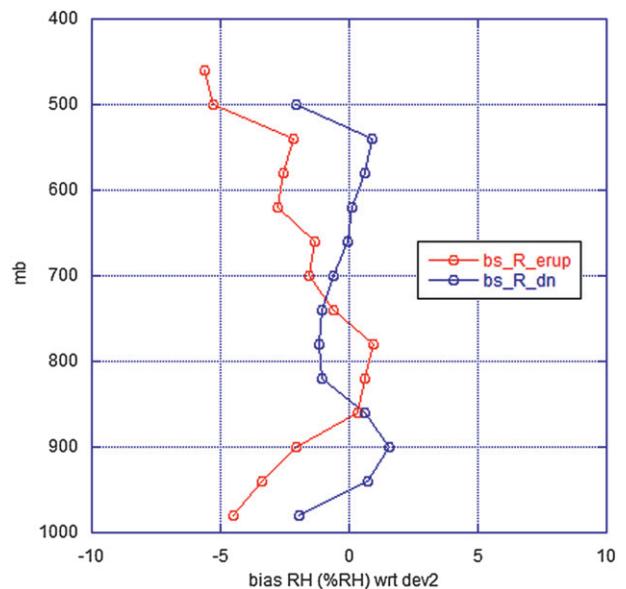- Comparisons are made every 10 hPa up from the surface.

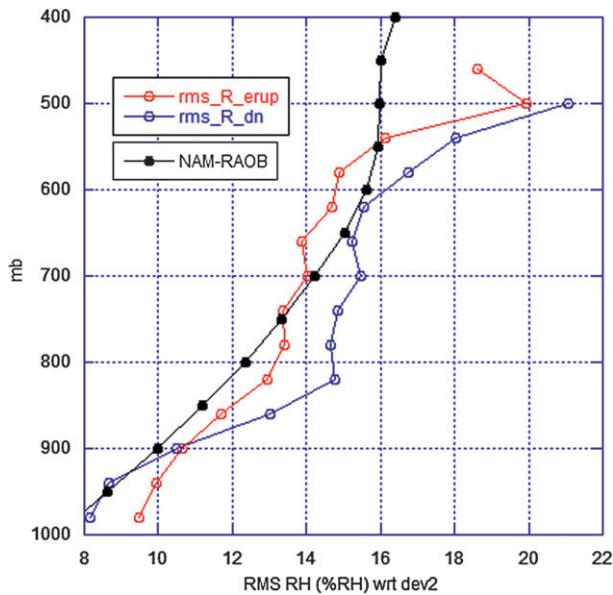FIG. 9. As in Fig. 6 but for the RH, for TAMDAR only.

FIG. 10. TAMDAR (open circles) RH RMS with respect to RUC dev2 1-h forecasts for October 2006. Solid black circles show the raob RH error assumed by the operational NAM model run at NCEP.

• No raob data are automatically eliminated based on differences from the operational RUC analysis data. (Fifteen obviously erroneous raobs were eliminated by hand between 23 February 2006 and December 2008.)

To compare the old and new verification methods, we look at the temperature impacts from TAMDAR at 850 hPa [discussed further in section 4b(1) below]. For most of the verified variables at various levels, the old and new methods give nearly identical answers, as shown in Figs. 11a and 11b for 850-hPa temperature. For this variable and level, the differences in QC screenings between the old and new verification methods made almost no difference. Almost identical results were evident, with an average 0.2-K improvement from dev2 (assimilating TAMDAR) over dev (no TAMDAR) 3-h forecasts in the Great Lakes region for the April–October 2006 period. As will be discussed in section 6, this is generally consistent with the current results.

The new verification system has allowed us more vertical precision; we can now inspect TAMDAR's impacts in the lowest 1500 m above the surface, below 850 hPa. Moreover, inclusion of more raob data has revealed previously obscured positive TAMDAR impacts on the relative humidity forecasts. These impacts were also obscured because some correct raob data were rejected by the old verification system—primarily at 500 hPa—and inclusion of these data resulted in greater calculated skill for dev2 with respect to dev, and hence
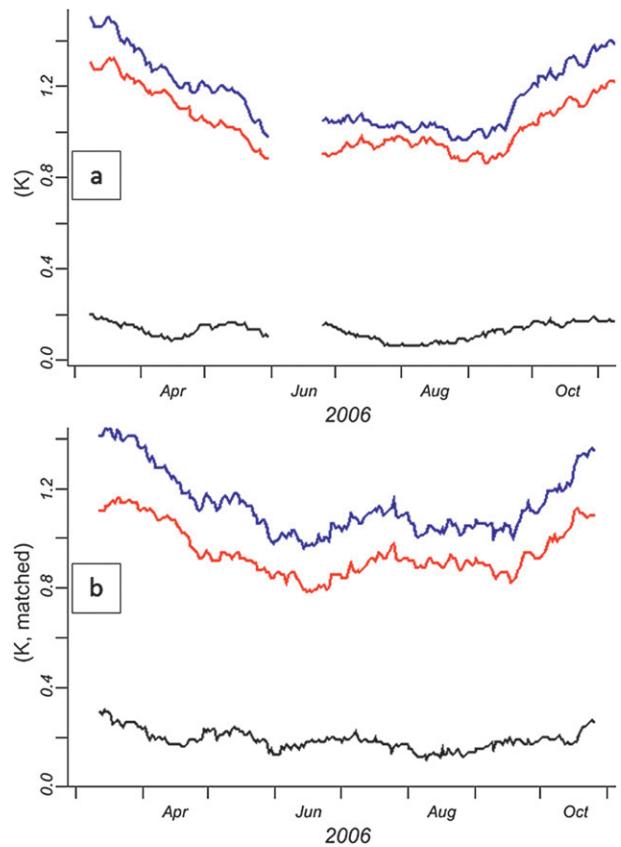

FIG. 11. The 850-hPa temperature 3-h forecast (valid at 0000 UTC, Great Lakes region) RMS difference between the model and raobs for 3-h RUC dev (blue) and RUC dev2 (red), for the (a) old and (b) new verifications. Thirty-day running averages were used.

a greater TAMDAR impact, especially for RH in the middle troposphere. No longer excluding raob data based on their differences from the operational RUC values has made a substantial difference in the new verification of the 600–400-hPa RH forecasts, as shown in the next example.

A comparison using the old and new verifications for the 500-hPa RH RMS errors for dev and dev2 forecasts is presented in Figs. 12 and 13. The new verification (Fig. 13) yields higher RMS error because of the use of *all* raob RH values. However, the new verification also shows a much greater *difference* between dev and dev2, indicating that the previously missing raob data have affected the verification of the two cycles unequally. Apparently, assimilation of TAMDAR RH observations improves the RUC RH forecasts in cases with large errors in the middle troposphere where raob values were being flagged using the old verification method. Note that the spacing on the vertical axis is equal, even though the magnitude of the error is larger with the new verification.

To see why this is so, we look at a particular case. Table 1 shows 500-hPa RH values for raob observations
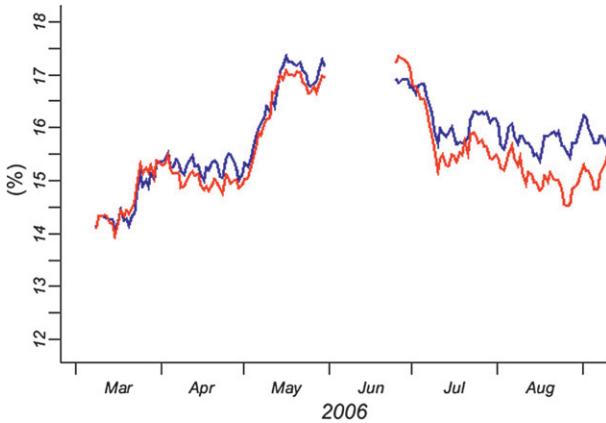
FIG. 12. RMS RH 500-hPa 3-h forecast error for RUC forecasts for dev (no TAMDAR, blue) and dev2 (TAMDAR, red) against raobs for the old verification system (centered at 15% RH). Thirty-day running averages were used.
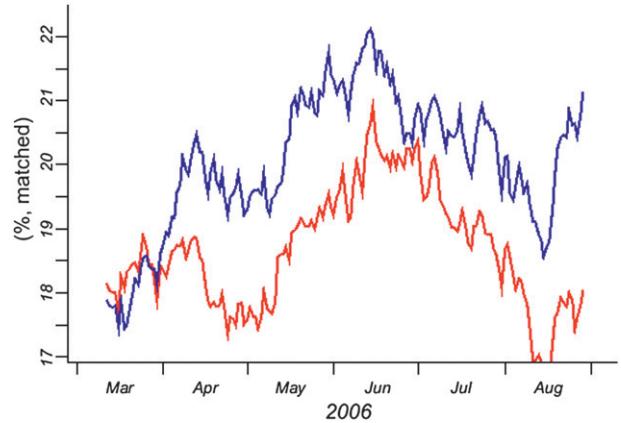

FIG. 13. As in Fig. 12 but for the new verification system (centered at 20% RH).

and the 3-h dev and dev2 RUC forecasts, all valid at 0000 UTC 1 July 2006. The old verification did not use the 500-hPa RH raobs at Pittsburgh, Pennsylvania (PIT) and Lincoln, Illinois (ILX). In both cases (see soundings in Figs. 14 and 15), strong subsidence layers were evident, with very dry air with bases just below 500 hPa, accompanied by sharp vertical moisture gradients in the 500–520-hPa layer. The QC screening algorithm used in the previous verification method flagged the 500-hPa RH observations at these two stations since the operational RUC analysis did not maintain this vertical gradient quite as sharply as in the full raob data. In both of these cases, the TAMDAR data led the dev2 RUC to better capture this vertical moisture gradient.

Figure 14 shows the observed raob and 3-h forecasts for RUC dev and dev2 soundings at ILX. The dev2 forecast sounding suggests that TAMDAR had detected a dry layer at 500 hPa. Nearby raobs (not shown) also suggest that the observed dry layer at and above 500 hPa was real.

Figure 15 shows the soundings for PIT. In this case, the accuracy of the dry raob observation at 500 hPa is less clear, but is not obviously wrong. Apparently, the much stronger TAMDAR impacts shown in Fig. 13 between the dev and dev2 500-hPa RH forecasts with the *new* verification screening are attributable to these cases with very sharp vertical moisture gradients near 500 hPa, also suggested by Szoke et al. (2007). Assimilation of the TAMDAR data allows the dev2 RUC forecasts to better capture these features.

In fact, the TAMDAR impacts on RH forecasts are potentially larger than this: in section 5a, we describe that a change in the RH error characteristic used in the dev2 assimilation of TAMDAR data—not implemented until 26 April 2007—increases the TAMDAR RH impacts.

This new verification system also provides much finer vertical resolution than the old and provides data below 850 hPa. Figure 16 shows vertical profiles of the RH biases for dev and dev2. Note that the RH biases of both models starts positive (more moist than the raobs) near the surface, become negative between approximately 900 and 700 hPa, then become increasingly positive with increasing altitude. The old verification system produced data on only three levels at and below 500 hPa (500, 700, and 850 hPa), thereby obscuring vertical variations such as these.

Since some of the finer resolution results from interpolating linearly in log p between significant levels, we investigated the extent to which this interpolation might differ from the actual atmospheric values. One-second-resolution data now available from the Radiosonde Replacement System (Facundo 2004) allowed us to study this. To test the effects of interpolation over relatively large pressure ranges, we chose a sounding with relatively few significant levels (Grand Junction, Colorado, at 0000 UTC

TABLE 1. RH values at 500 hPa at 0000 UTC 1 Jul 2006.

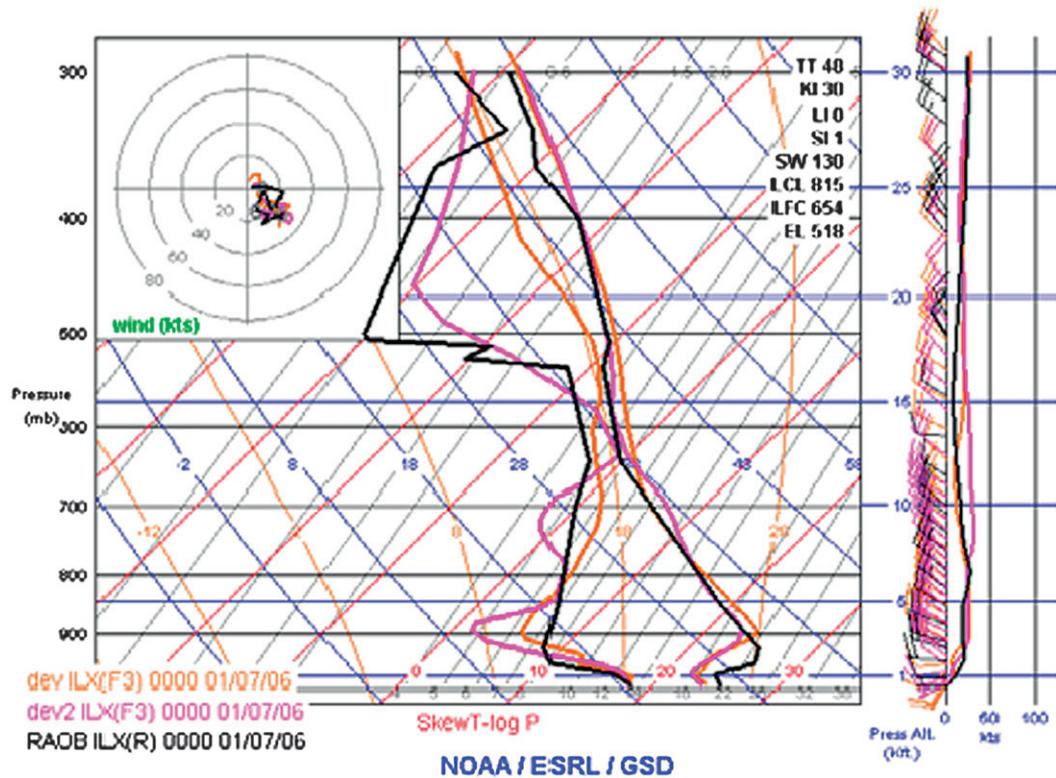| ID	Name	Raob	3h-Dev	3h-Dev2
ABR	Aberdeen, SD	85	90	87
APX	Alpena, MI	6	6	9
BUF	Buffalo, NY	8	37	7
DTX	Detroit, MI	14	15	11
DVN	Davenport, IA	16	39	41
GRB	Green Bay, WI	30	18	31
ILN	Cincinnati, OH	33	61	48
ILX	Lincoln, IL	19	84	40
INL	International Falls, MN	26	10	21
MPX	Minneapolis, MN	9	28	33
OAX	Valley, NE	15	53	41
PIT	Pittsburgh, PA	3	76	33
TOP	Topeka, KS	57	83	75

FIG. 14. Soundings at ILX at 0000 UTC 1 Jul 2006. Raob is shown in black, RUC dev 3-h forecasts in orange, and RUC dev2 3-h forecasts in magenta.

29 June 2009; Fig. 17). In this case, the interpolation extended over pressure ranges up to 120 hPa (between 820 and 700 hPa). For this sounding we calculated the average and RMS difference for temperature, relative humidity, and wind, between the 1-s data and the 10-mb interpolated sounding. Results are shown in Table 2 for various pressure bands; they are lower by a factor of 3–10 than the RMS values in the various data deprivation experiments we discuss below, particularly for temperature. Thus, we are confident that our interpolation scheme is not obscuring or skewing our forecast impact results and that the linear approximation between raob significant levels agrees well with the 1-s data, especially for temperature.

b. Effect of TAMDAR assimilation on RUC forecast skill

1) TEMPERATURE

TAMDAR impacts on 3-h RUC temperature forecasts averaged over the 1000–500-hPa layer for the 2006–08 period are shown in Fig. 18. The temperature RMS errors for both RUC dev (without TAMDAR) and dev2 (including TAMDAR) show the common seasonal variations with larger errors in winter and smaller errors in

summer, when the lower troposphere is more commonly well mixed with a deeper boundary layer. We consider only 0000 UTC raobs because this is the time when we expect to see the maximum TAMDAR impact, given the schedule (1200–0300 UTC, primarily daylight hours) of the Mesaba TAMDAR fleet.

The TAMDAR impacts (black line in Fig. 18) are always positive and are largest in winter, when the temperature forecast errors are themselves largest. In winter, TAMDAR reduces the 3-h temperature forecast error by an average of 0.2 K over the entire 1000–500-hPa depth.

Figure 18 can help put the results discussed in section 3b into climatological perspective. Note that October 2006—the period over which TAMDAR's errors were evaluated in section 3b—is a transition period between the relatively lower RUC RMS errors and TAMDAR impacts of summer and the larger errors and impacts of winter, but is otherwise generally consistent with the RUC behavior and TAMDAR impacts over the entire 3-yr period, and is consistent with the behavior in the autumns of 2007 and 2008. Thus, we are confident that the results in section 3b are reasonably representative of any fall period.

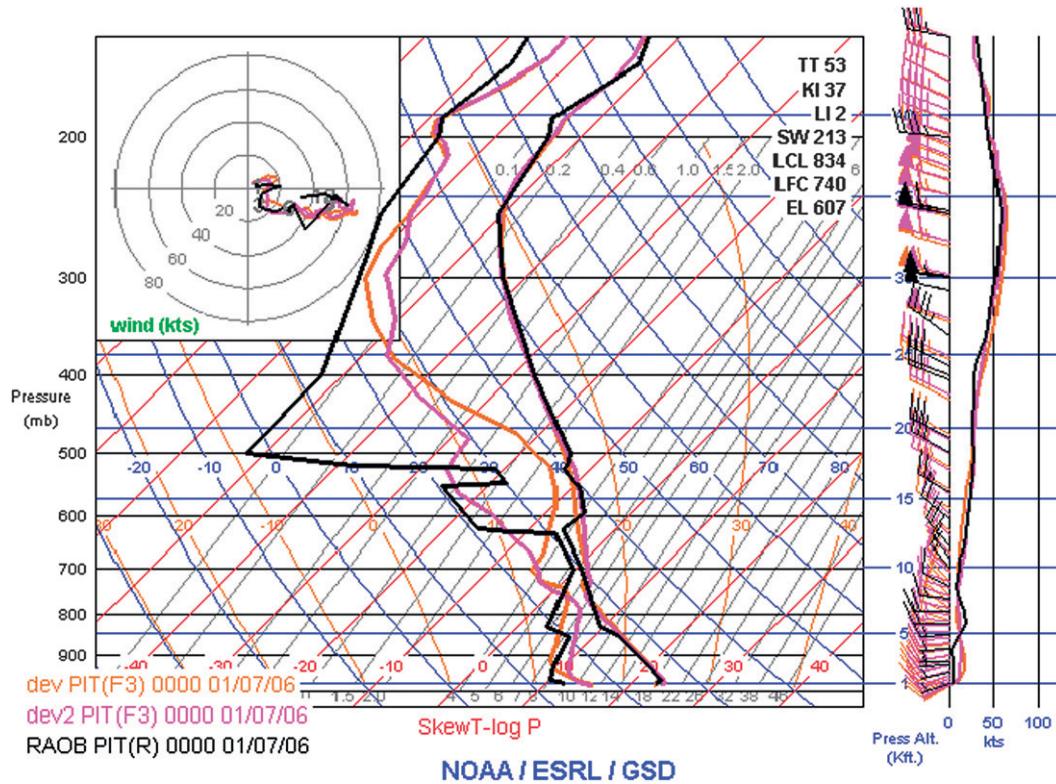A vertical profile of the temperature RMS error for the RUC dev and dev2 3-h forecasts for March 2008 is

FIG. 15. As in Fig. 15 but for PIT.

2) WIND

shown in Fig. 19. Figure 18 suggests that this is a typical spring period in terms of RUC temperature error and TAMDAR impacts. Inaccuracy in forecast temperatures associated with errors in planetary boundary layer (PBL) depth results in a maximum from 950 to 800 hPa in the vertical profile for temperature errors in the RUC dev model forecasts (without TAMDAR). The dev2 has lower errors for all levels between the surface and 320 hPa but especially between 850 and 950 hPa. We interpret this as TAMDAR's ascent–descent profiles being particularly important in defining the PBL depth more accurately. The maximum RMS error difference between dev and dev2 occurs at 900 hPa and is about 0.4 K.

To put this TAMDAR impact into perspective, we present profiles of 3-h temperature forecast errors (Fig. 20) from the November–December 2006 retrospective period (see section 5) for 1) all AMDAR data (including TAMDAR), equivalent to the dev2; 2) no TAMDAR data, equivalent to the dev; and 3) no AMDAR data at all.

The TAMDAR impacts (the black curve) peaks at 0.4 K at 900 hPa, just as they do in Fig. 19. The AMDAR impacts also peak at 900 hPa and have a value of nearly 1.1 K. AMDAR impacts also have an additional peak at 500 hPa, above the region where Mesaba turboprop aircraft (carrying TAMDAR) fly most of the time.

TAMDAR impacts on 3-h wind forecasts, also averaged over the surface–500-hPa layer (Fig. 21) were consistently positive, although small. This indicates that even though TAMDAR wind errors are greater than those of the traditional AMDAR jet fleet, as discussed in
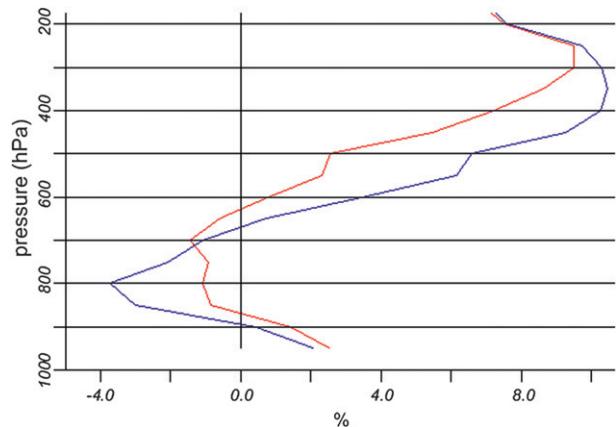

FIG. 16. Vertical profile for RH bias (model − raob) for dev (no TAMDAR, blue) and dev2 (TAMDAR, red) 3-h forecasts with respect to raobs at 0000 UTC in the Great Lakes region for the April–August 2006 period.
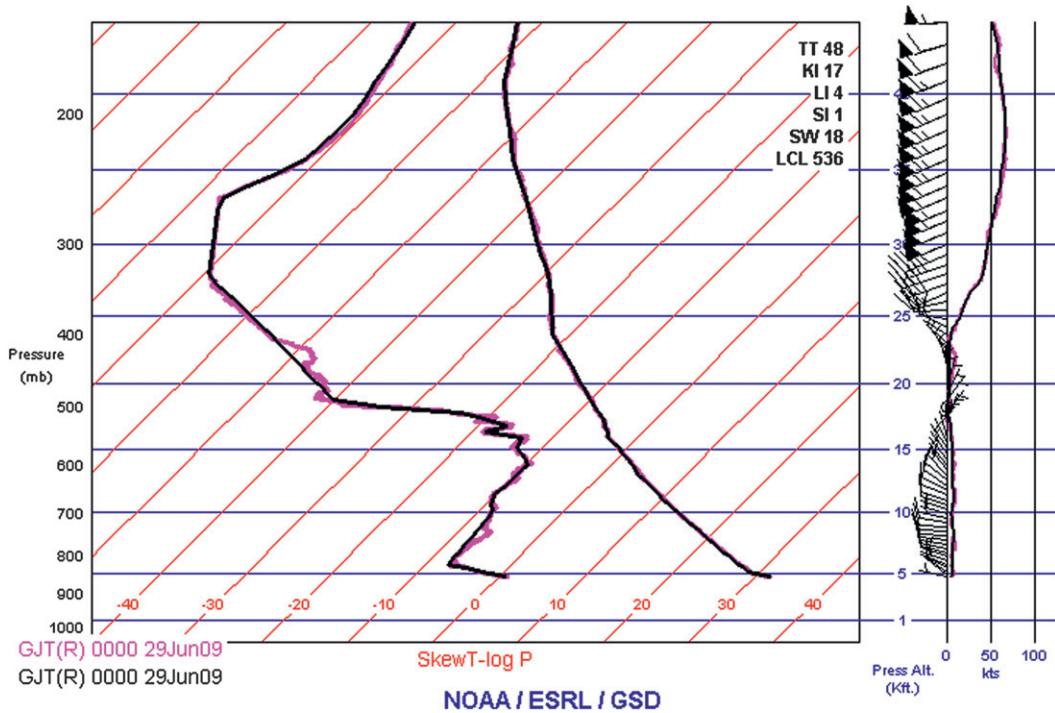
FIG. 17. Grand Junction, CO, raob at 0000 UTC 29 Jun 2009. Data interpolated to 10 hPa from mandatory and significant levels are shown in black; data from 1-s data are shown in magenta. (One-second wind barbs are not shown.)

section 3, they nonetheless provide additional value for wind forecasts in the RUC.

In the vertical profile for 3-h wind forecast error (RMS of the vector wind difference between the model and raobs) for March 2008 (Fig. 22), TAMDAR impacts on winds show a double peak, with a maximum at 700 hPa. At this level, the RMS reduction due to TAMDAR is about 0.25 m s^{-1}.

The vertical profile for the 3-h wind forecast error (Fig. 23) from the November–December 2006 retrospective period (see section 5) compares the TAMDAR wind impacts with the impacts of all of the aircraft (AMDAR, which includes TAMDAR). The heavy gold curve in Fig. 23 shows the maximum AMDAR wind impact to be at 450–500 hPa, with an RMS error re-

duction of 0.7 m s^{-1}. The TAMDAR impacts peak at about 600 hPa and are about 0.2 m s^{-1}. Below 550 hPa, the similarities of the TAMDAR and the AMDAR impact curves (heavy) indicate that TAMDAR is responsible for most of the (small) AMDAR wind impacts in this altitude range. Above 550 hPa, AMDAR jets provide most of the impacts on the RUC 3-h wind forecasts.

3) RELATIVE HUMIDITY

A 3-yr history for TAMDAR impacts on lower-tropospheric RH forecasts (RUC dev versus RUC dev2) is shown in Fig. 24. The impacts are generally between 1% and 2% when averaged between the surface and 500 hPa. A change was made on 26 April 2007 to the

TABLE 2. Bias and RMS differences between 10-hPa interpolated and 1-s raob data from the GJT sounding at 0000 UTC 29 Jun 2009. N is the number of observations in each pressure band.

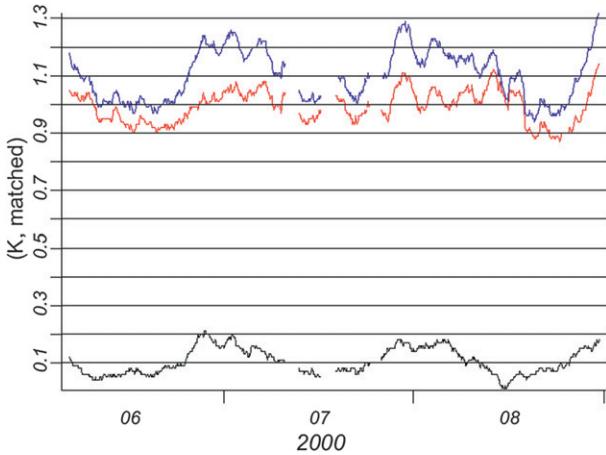| Pressure (hPa)	N	T bias (RMS) (°C)	RH bias (RMS) (%)	Speed bias (vector wind RMS) (m s^{-1})
1000–800	122	−0.02 (0.02)	−0.00 (0.24)	−0.24 (0.7)
800–700	260	−0.02 (0.02)	−0.52 (0.84)	0.10 (0.93)
700–600	278	−0.02 (0.02)	0.37 (0.95)	−0.44 (0.80)
600–500	338	−0.00 (0.01)	−0.57 (2.52)	0.10 (1.26)
500–400	398	0.00 (0.01)	−0.30 (0.95)	−1.09 (2.85)
400–300	521	−0.00 (0.01)	0.09 (0.10)	0.04 (0.87)
300–200	633	−0.02 (0.03)	−0.18 (0.34)	−0.62 (1.28)
200–100	935	0.03 (0.06)	0.07 (0.57)	−0.08 (2.7)
100–0	1960	−0.01 (0.08)	−0.01 (0.59)	−1.35 (3.94)

FIG. 18. Time series of 3-h temperature forecast errors (RMS difference from 0000 UTC raobs) for dev (no TAMDAR, blue) and dev2 (TAMDAR, red), and the dev–dev2 difference (black), for the Great Lakes region, in the layer between the surface and 500 hPa. Thirty-day running averages were used. Positive differences indicate a positive TAMDAR impact.

FIG. 20. As in Fig. 19 but for 27 Nov–5 Dec 2006: all-data run, light red; no-TAMDAR run, light blue; no-aircraft run, gold; TAMDAR impact, black; and aircraft impact, heavy gold.

specific observation error for TAMDAR RH (see section 5a). Although we know from reprocessing a 10-day period that the new RH observation error increases TAMDAR's RH impacts, the increase is small enough that it is not clearly evident compared with the seasonal variations shown in Fig. 24.

The corresponding vertical profile for the RH forecast impacts from TAMDAR (Fig. 25) is relatively uniform from the surface to 700 hPa, from 1% to 3%. An enhancement in RH impacts from TAMDAR around 600 hPa is also evident. This enhancement is consistent over the seasons (not shown). We speculate that the surface observations limit the impacts of TAMDAR at

altitudes below this level, and there are relatively few TAMDAR observations above.

4) TAMDAR IMPACTS AS A FRACTION OF ESTIMATED MAXIMUM POTENTIAL IMPROVEMENT (EMPI)

To put these error reductions in perspective, it is worth considering what the minimum model–raob differences ("errors") might be, given a perfect model. Raobs have instrument errors and also exhibit representativeness errors because they provide in situ point observations, whereas a model provides an average over the area of a grid cell (20 km^2 in the case of the dev and dev2). To account for these inherent differences between the model and verifying observations, we can take the *analysis* error as an approximate measure of the minimum forecast error to be expected, similar to the normalization of
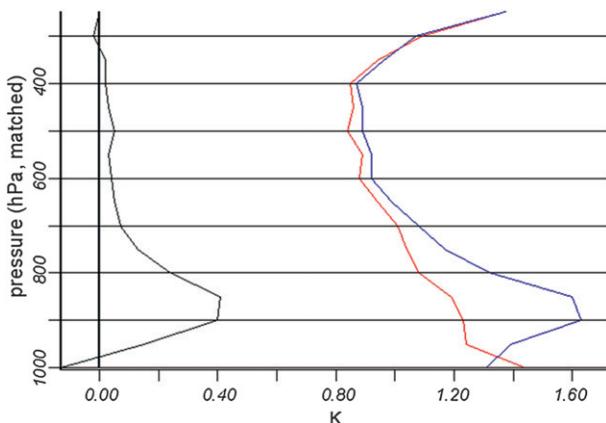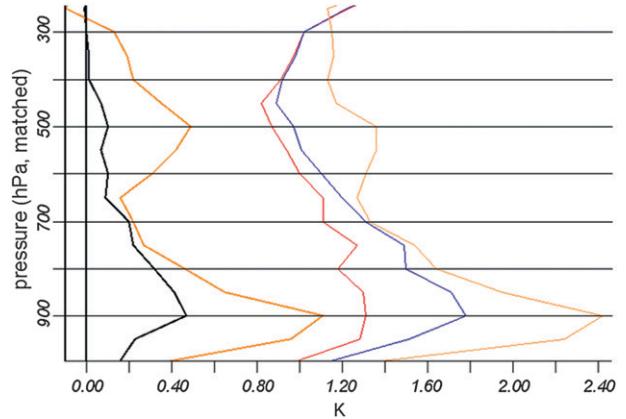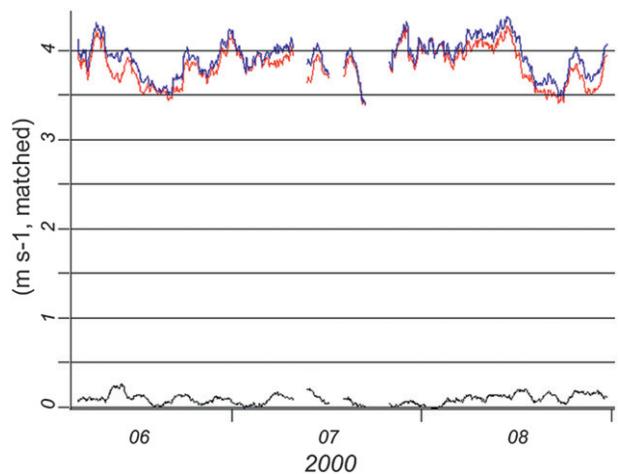

FIG. 19. Vertical profile of 3-h temperature forecast errors (RMS difference from 0000 UTC raobs) for dev (no TAMDAR, blue) and dev2 (TAMDAR, red), and the dev–dev2 difference (black), for the Great Lakes region, during March 2008.

FIG. 21. As in Fig. 18 but for wind forecasts (RMS vector difference from 0000 UTC raobs).
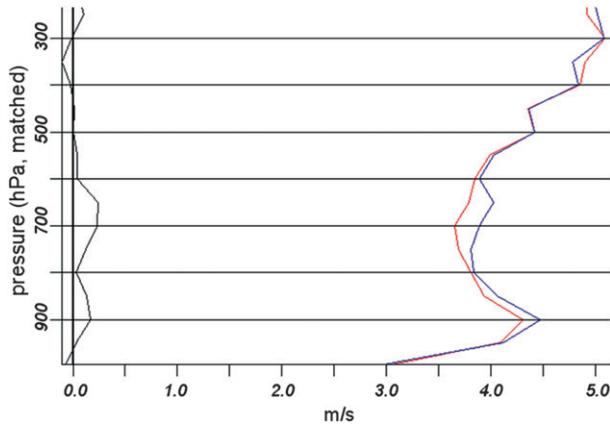
FIG. 22. As in Fig. 19 but for 3-h wind forecasts.

FIG. 24. As in Fig. 18 but for RH forecasts.

forecast impacts shown by Benjamin et al. [(2004c), their Eq. (3)].

As a specific example, Fig. 26 shows profiles of the dev2 and dev 3-h RH forecast errors along with the dev2 *analysis* errors. The RMS for the analysis varies between 6% RH at 950 hPa and about 14% RH at 700 hPa. The *difference between* the dev and dev2 3-h RH forecast error (i.e., the TAMDAR impact) varies between 1% RH and 3% RH. At 600 hPa, the TAMDAR impact is 2.8% RH, and the difference between the dev 3-h forecast curve and the dev2 analysis curve is 5.7% RH. Assuming that this 5.7% RH error reduction at 600 hPa is the best we can hope for (the EMPI), the TAMDAR impact is about 50% of the EMPI. Over all altitudes, TAMDAR provides 15%–50% of the EMPI.

For temperature, we reason similarly. Because the analysis fit to the raob verification data is about 0.5 K, as described in Benjamin et al. (2006a,b, 2007), the maximum possible reduction in RMS error difference would

be about 1.1 K (the difference between the ~1.6 K RMS shown for dev in Fig. 19 at 900 hPa and the 0.5-K analysis fit). Therefore, TAMDAR's impact is about 35% of the EMPI for the 3-h temperature forecast error at 900 hPa [35% ≈ 0.4/(1.6–0.5)].

For wind, the analysis fit to the raobs is about 2.2 m s^{-1} near 600 hPa (not shown). Thus, TAMDAR's impact on 3-h wind forecasts in this altitude range is about 15% of the EMPI.

5. Further applications of retrospective runs

To study TAMDAR's impacts in more detail, and determine how these new data are best assimilated in the RUC, we saved all data for two 10-day periods: 1200 UTC 26 November–1200 UTC 5 December 2006 and 0000 UTC 15 August–0000 UTC 25 August 2007. We then reran the RUC with a variety of different assimilation schemes and TAMDAR data variations over these periods.
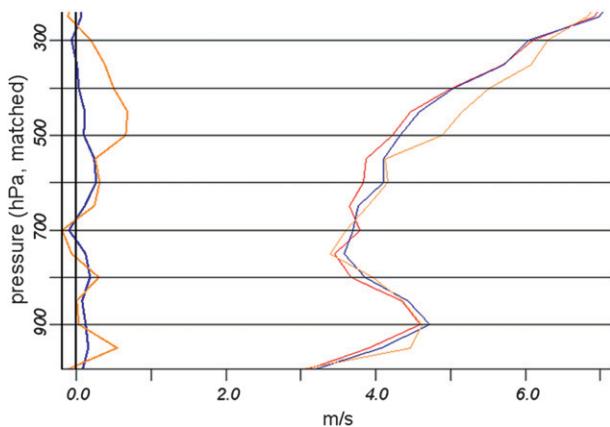

FIG. 23. As in Fig. 20 but for 3-h wind forecasts: all-data run, light red; no-TAMDAR run, light blue; no-aircraft run, light gold; TAMDAR impact, heavy blue; and aircraft impact, heavy gold.
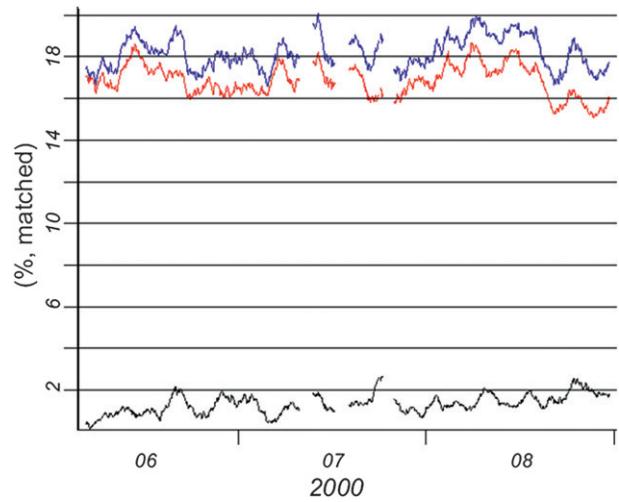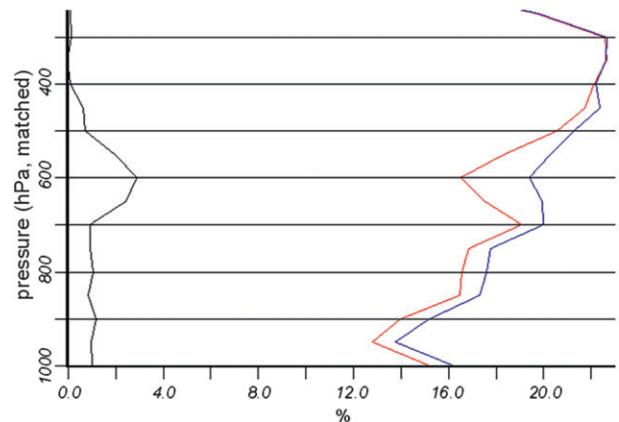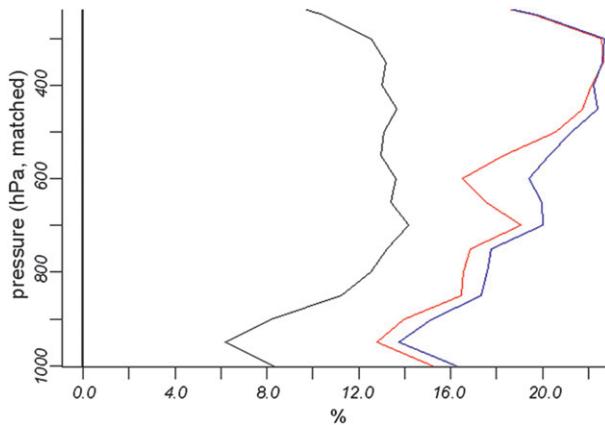

FIG. 25. As in Fig. 19 but for 3-h RH forecasts.

FIG. 26. RMS difference between RUC grids and raobs in the Great Lakes region at 0000 UTC for the dev2 RH analysis (black), dev 3-h forecasts (blue), and dev2 (red) 3-h forecasts, during March 2008.

FIG. 27. Vertical profile of 3-h RH forecast errors (RMS difference from 0000 UTC raobs) for "old RH processing" (blue, RH errors divided by 4), and "new RH processing" (RH errors corrected), without TAMDAR data, for the Great Lakes region. The black curve shows the difference; negative values indicate that the new processing has lower RMS errors.

We chose these periods because they included intense weather events. The 2006 period includes a potent early winter storm that featured a band of heavy snow and ice through the heart of the TAMDAR network, mainly from 30 November through 1 December, and includes more typically moderate weather in the later portion of the period. The August 2007 period includes a variety of weather systems as well. Since results from this summer period generally corroborate the winter results, we do not include them here.

These periods were chosen primarily in support of our TAMDAR investigations. However, they have served as a basis for additional experiments denying other data sources and are discussed in detail by Benjamin et al. (2010).

a. Relative humidity observation error specification for assimilation

Because high temporal and spatial resolution RH measurements have been unavailable in the past, we had no firm guidance for choosing the appropriate error for these observation measurements other than from engineering-based estimates by manufacturers. Both instrument errors and representativeness errors must be accounted for, so that the importance of each observation relative to the model background field is correctly assessed. Estimating an RH observation error that is too large will result in less-than-optimal TAMDAR impacts. Choosing a value that is too small will result in overfitting, causing numerical noise that will degrade forecasts.

We experienced overfitting when, during the fall of 2005, the TAMDAR RH error was inadvertently set to 1% RH. During this period, TAMDAR's impacts on the 3-h RH forecasts were negative (Benjamin et al. 2007,
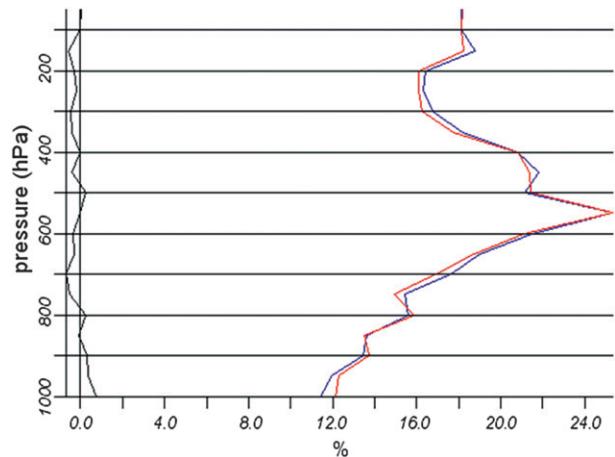
their Figs. 9 and 10). However, for most of the time we that have assimilated TAMDAR's data, we have run the system with RH observation errors for TAMDAR that were between 3% and 12%. With these errors, TAMDAR has had a positive impact of reducing the subsequent RUC RH forecast errors by 1%–3% RH [see section 4b(3)].

In April 2007, we discovered that the observation errors for all RH observations [TAMDAR, surface observations, raobs, and integrated precipitable water data from the Meteorological Applications of GPS experiment (GPS-Met; Smith et al. 2007)] had been inadvertently set too low since the start of our TAMDAR experiments. We corrected this in a retrospective run and found that the correction (called "new RH processing" below) resulted in slightly increased model skill (decreased RMS) for RH forecasts at nearly all levels, as Fig. 27 shows, even in the absence of TAMDAR.

When TAMDAR data are included, the new processing increased the TAMDAR impact, as shown in Fig. 28.

Each curve in Fig. 28 indicates the difference between the RMS errors of the TAMDAR and no-TAMDAR runs (with respect to 0000 UTC raobs in the Great Lakes verification region shown in Fig. 5). The blue curve shows the impacts under the old RH processing; the red curve shows the impacts with the new RH processing. The larger values for the red curve demonstrate that the TAMDAR impacts in RH forecasts increase substantially at levels between 850 and 450 hPa with the new processing using a more appropriate observation error for the TAMDAR RH observations.
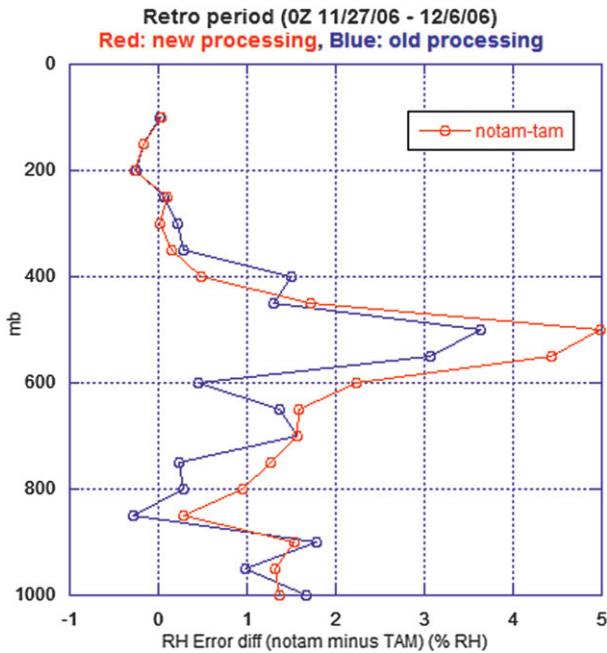
FIG. 28. TAMDAR impacts on 3-h RH forecasts (see text for explanation) for ''new RH processing'' (12% TAMDAR RH error, red) and ''old RH processing'' (12%/4 = 3% TAMDAR RH error, blue) for the retrospective time period.

Additional retrospective runs using TAMDAR RH observation errors of 18% and 25% showed that these values resulted in slightly less TAMDAR impact than the 12% value. Therefore, we implemented the 12% RH error, and the correction of the other RH observation errors, in our real-time dev2 runs on 26 April 2007. Although TAMDAR's RH impact was less than it might have been before this date, our long time series show that TAMDAR's impact on RH forecasts was notable even before this change was implemented.

b. Indirect relative humidity impacts

There has been some speculation that improved resolution in temperature and wind data alone will indirectly improve RH forecasts, because better wind and temperature fields will result in better placement of humid areas. We therefore performed a retrospective run in which we included TAMDAR wind and temperature observations, but no TAMDAR RH observations. (All other data were included.)

When TAMDAR RH observations are excluded, TAMDAR has virtually no impact on 3-h forecasts of RH (Fig. 29). However, TAMDAR wind and temperature data alone do have some impact on the longer-range forecasts, such as the 9-h RH forecasts shown in Fig. 30. In that case, the blue curve between 500 and 450 hPa shows RH errors about halfway between the all-TAMDAR
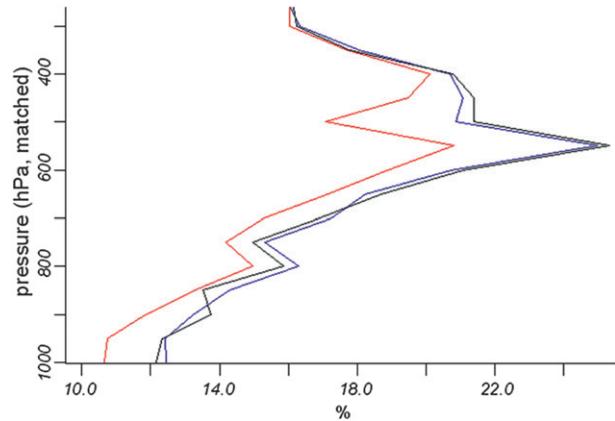

FIG. 29. The 3-h RH forecast errors (RMS difference from 0000 UTC raobs) for the Great Lakes region, for the retrospective period, for three cases: all-TAMDAR data, red; no-TAMDAR data, black; and TAMDAR wind and temperature data only, blue.

(red) and no-TAMDAR (black) runs. Interestingly, this is at a higher altitude than TAMDAR generally flies. This suggests that the model vertical motion is improved by the temperature and wind data, thereby improving the subsequent RH forecasts.

Thus, we can conclude that for 3-h forecasts, RH observations are needed to improve RH forecasts, at least on the 20-km scale of our RUC model runs. However, at longer forecast projections such as 9 h, a small improvement in the RH forecasts is apparent solely from the TAMDAR temperature and wind observations.

c. Vertical resolution

During the retrospective time period, AirDat provided high vertical resolution data [10 hPa in the lowest 200 hPa (for both ascents and descents), and 25 hPa above that]. At other times, to save communication costs, they have provided data at lower vertical resolution. To study
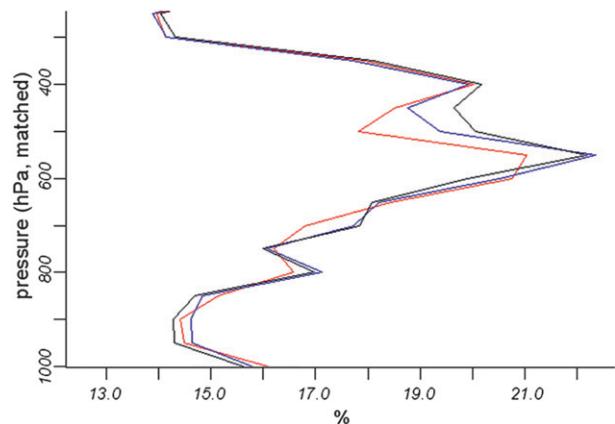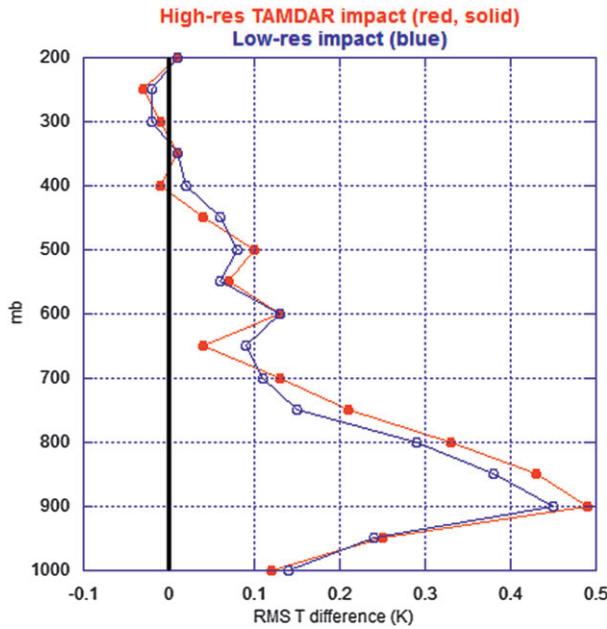

FIG. 30. As in Fig. 29 but for 9-h RH forecasts.

FIG. 31. TAMDAR 3-h temperature forecast impacts (see text for explanation) for the full vertical resolution run (red) and the low-vertical resolution run (blue), for the retrospective period.

the impacts of using different vertical resolutions, we artificially degraded the resolution above the lowest 100 hPa AGL to 50 hPa; we kept the 10-hPa resolution in the lowest 100 hPa. This removed about one-half of the TAMDAR observations.

The curves in Fig. 31 may be compared to the black curve in Fig. 19. That is, each is the difference in the RMS temperature error between an all-TAMDAR run and the no-TAMDAR run. The results indicate that the lowered vertical resolution does indeed reduce TAMDAR's impact on 3-h temperature forecasts below 700 hPa. TAMDAR's impact is reduced by about 10% at 900 hPa, growing to a 30% reduction at 750 hPa. For RH forecasts, reducing the vertical resolution had little consistent impact (not shown). However, for all variables, the impact of reduced vertical resolution is certainly larger in certain situations—often related to adverse weather conditions. We note that higher vertical resolution has been very useful in some critical weather situations for human forecasters who look directly at the TAMDAR soundings (Szoke et al. 2006).

6. Recent developments

Recently, additional TAMDAR fleets have started reporting to ESRL/GSD. Currently (fall 2009), the four commercial air carrier fleets providing TAMDAR data are

- Mesaba, data first received in 2004, and reported on above;

- PenAir, data first received in late 2007; PenAir flights connect Anchorage with smaller cities in southwestern Alaska and the Aleutian Islands—a generally data-poor region;
- Chautauqua, data first received in April 2008; this fleet of regional jets flies higher and faster than the turbo-props in the other fleets and, therefore, can potentially provide valuable data at higher altitudes than available from turboprops; and
- Horizon, data first received in December 2008.

A more recent (April 2009) horizontal distribution of TAMDAR data reported to ESRL/GSD for a 24-h period is shown in Fig. 32. Reports from the PenAir fleet are evident over Alaska. Horizon reports are over the western United States; Chautauqua reports are now made over Mexico, the lower Midwest, and the east coast. Note the data points coded in light blue, representing data taken above 28 000 ft by Chautauqua jets.

Our initial studies of data from the Chautauqua jets indicate that the quality of the temperature, wind, and relative humidity data is as good or better than that produced by the Mesaba turboprops (not shown). We started ingesting Chautauqua data into the dev2 on 30 April 2008 and have seen a notable increase in TAMDAR's impact—particularly on relative humidity forecasts—since that time.

Figure 33 shows TAMDAR's impact on 3-h relative humidity forecasts for the entire eastern U.S. region (the violet rectangle in Fig. 5). This geographic and altitude (up to 400 hPa) region was not densely covered by the initial TAMDAR Mesaba fleet alone (Fig. 5). The increased TAMDAR impacts on RH forecasts for this region since late April 2008 are evident in the difference curve.

A December 2008 vertical profile of TAMDAR impacts on RH 3-h forecasts (Fig. 34) in the eastern U.S. region includes the effects of the Chautauqua fleet. Comparing this with Fig. 25, which shows the corresponding pre-Chautauqua vertical RH impacts for the Great Lakes region, reveals that TAMDAR's impact now extends higher—to above 300 hPa.

7. Summary and a look ahead

The TAMDAR sensor provides meteorological data on a regional scale over the U.S. midwest (and now over most of the United States). By equipping regional aircraft, TAMDAR provides ascent–descent profiles at regional airports for which traditional AMDAR profiles were not available. Moreover, TAMDAR also reports relative humidity, a variable not generally or reliably available previously from commercial aircraft. We have evaluated the impacts of TAMDAR's wind, temperature, and relative
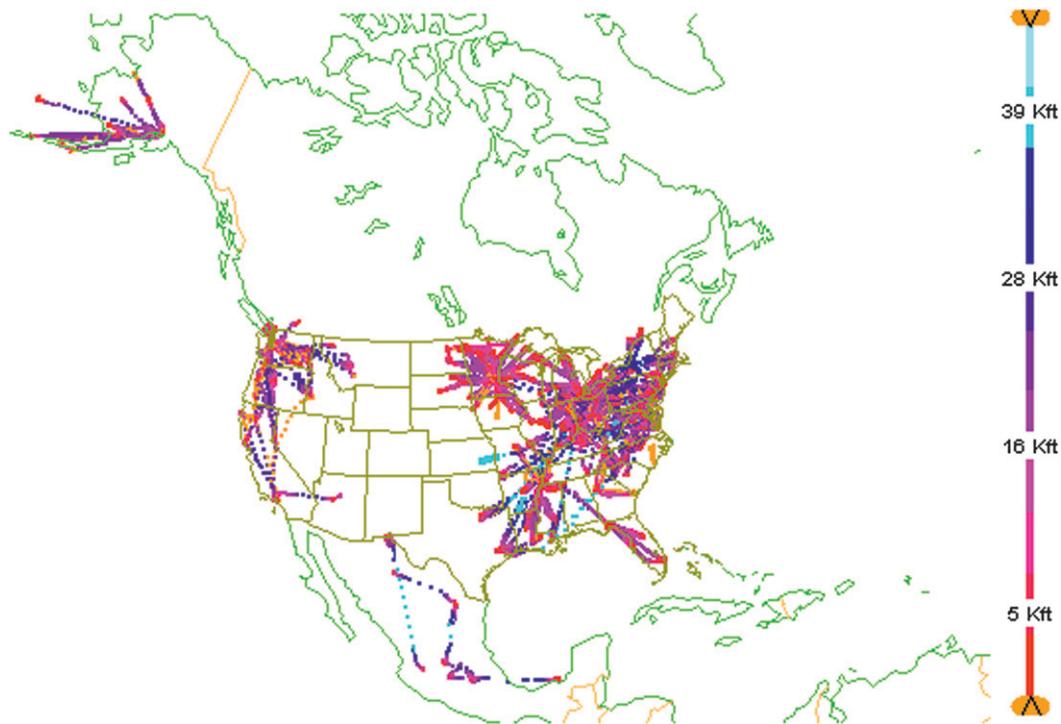
FIG. 32. TAMDAR observation reports received over a 24-h period at ESRL/GSD on 29 Apr 2009. There were 30 877 reports.

humidity data on the RUC model–assimilation system with 1) real-time matched TAMDAR and no-TAMDAR runs for the past 3 yr and 2) retrospective runs over two 10-day active weather periods during the winter of 2006 and summer of 2007.

We have shown that assimilation of TAMDAR observations improves 3-h RUC forecasts in the region and altitude range in which TAMDAR flies. We estimate the TAMDAR's impact as follows:

- The 3-h temperature forecast errors are reduced by up to 0.4 K, dependent on vertical level.
- The 3-h wind forecast errors are reduced by up to 0.25 m s^{-1}.
- The 3-h relative humidity forecast errors are reduced by up to 3% RH.

As discussed in section 4b(4), we can cast these error reductions into fractions of the *estimated maximum potential improvement* (EMPI). In these terms, TAMDAR results in these impacts:

- The 3-h temperature forecast errors are reduced by up to 35% of the EMPI.
- The 3-h wind forecast errors are reduced by up to 15% of the EMPI.
- The 3-h relative humidity forecast errors are reduced by 15% to 50% of the EMPI.

Retrospective runs have revealed the following:

- The optimal TAMDAR RH observational error specification is 12% for assimilation impacts. Both lower and higher values resulted in lower RH forecast impacts.
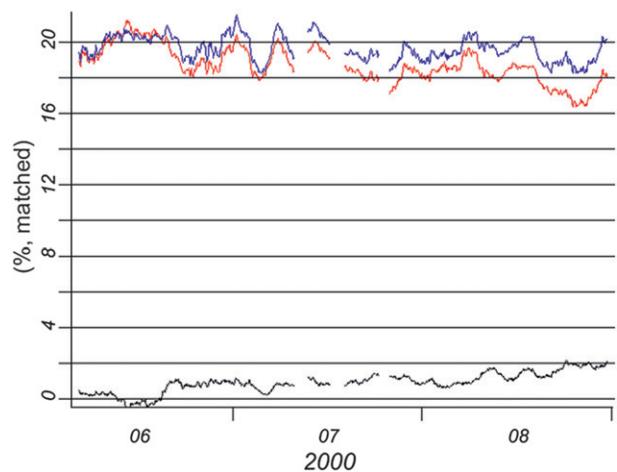

FIG. 33. Time series of 3-h RH forecast errors (RMS difference from 0000 UTC raobs) for dev (no TAMDAR, blue) and dev2 (TAMDAR, red), and the dev–dev2 difference (black), for the eastern U.S. region, in the layer between the surface and 400 hPa. Thirty-day running averages were used. Positive differences indicate a positive TAMDAR impact.
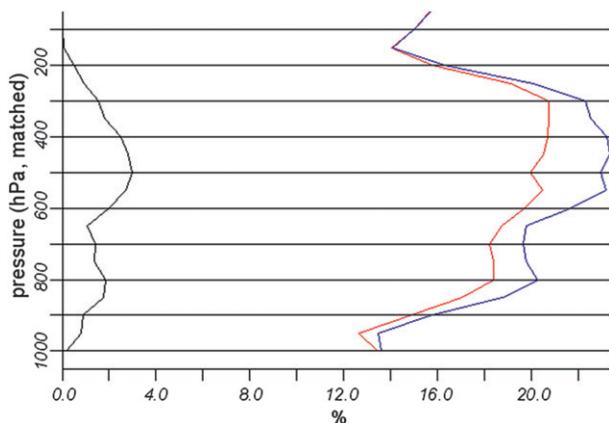
FIG. 34. Vertical profile of 3-h RH forecast errors (RMS difference from 0000 UTC raobs) for dev (no TAMDAR, blue) and dev2 (TAMDAR, red), and the dev-dev2 difference (black), for the eastern U.S. region, during December 2008.

The 12% RH error is now being used in real-time RUC cycles.

- RH observations are generally required to improve 3-h forecast skill. However, for longer forecasts, wind and temperature observations alone, at sufficiently fine resolution, can improve the RH forecasts indirectly.
- Lowered vertical resolution reduces the TAMDAR-related forecast improvement from 10% to 30% for temperature forecasts, but in individual cases this reduced accuracy may cause important meteorological conditions to be unobserved or inadequately resolved.

REFERENCES

Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004a: Mesoscale weather prediction with the RUC hybrid isentropic–terrain-following coordinate model. *Mon. Wea. Rev.,* **132,** 473–494.

——, and Coauthors, 2004b: An hourly assimilation–forecast cycle: The RUC. *Mon. Wea. Rev.,* **132,** 495–518.

——, B. E. Schwartz, E. J. Szoke, and S. E. Koch, 2004c: The value of wind profiler data in U.S. weather forecasting. *Bull. Amer. Meteor. Soc.,* **85,** 1871–1886.

——, W. R. Moninger, T. L. Smith, B. D. Jamison, and B. E. Schwartz, 2006a: TAMDAR aircraft impact experiments with the Rapid Update Cycle. Preprints, *10th Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land*

Surface, Atlanta, GA, Amer. Meteor. Soc., 9.8. [Available online at http://ams.confex.com/ams/pdfpapers/105064.pdf.]

——, ——, ——, ——, and ——, 2006b: Impact of TAMDAR humidity, temperature, and wind observations in RUC parallel experiments. Preprints, *12th Conf. on Aviation, Range, and Aerospace Meteorology,* Atlanta, GA, Amer. Meteor. Soc., 4.5. [Available online at http://ams.confex.com/ams/pdfpapers/105028.pdf.]

——, ——, ——, ——, E. J. Szoke, and T. W. Schlatter, 2007: 2006 TAMDAR impact experiment results for RUC humidity, temperature, and wind forecasts. Preprints, *11th Symp. on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface,* San Antonio, TX, Amer. Meteor. Soc., 9.2. [Available online at http://ams.confex.com/ams/pdfpapers/119903.pdf.]

——, B. D. Jamison, W. R. Moninger, S. R. Sahm, B. E. Schwartz, and T. W. Schlatter, 2010: Relative short-range forecast impact from aircraft, profiler, radiosonde, VAD, GPS-PW, METAR, and mesonet observations via the RUC hourly assimilation cycle. *Mon. Wea. Rev.,* **138,** 1319–1343.

Daniels, T. S., W. R. Moninger, and R. D. Mamrosh, 2006: Tropospheric Airborne Meteorological Data Reporting (TAMDAR) overview. Preprints, *10th Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface,* Atlanta, GA, Amer. Meteor. Soc., 9.1. [Available online at http://ams.confex.com/ams/pdfpapers/104773.pdf.]

Facundo, J., 2004: Update on the implementation of the National Weather Service's Radiosonde Replacement System. Preprints, *Eighth Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface,* Seattle, WA, Amer. Meteor. Soc., 3.2. [Available online at http://ams.confex.com/ams/pdfpapers/69730.pdf.]

Helms, D., and Coauthors, 2005: NOAA's deployment of the Water Vapor Sensor System (WVSS II). Preprints, *21st Conf. on Weather Analysis and Forecasting/17th Conf. on Numerical Weather Prediction,* Washington, DC, Amer. Meteor. Soc., P1.45. [Available online at http://www.nws.noaa.gov/ost/wvss/AMS_NWP-WVSSII_Poster_JUL05.pdf.]

Moninger, W. R., R. D. Mamrosh, and P. M. Pauley, 2003: Automated meteorological reports from commercial aircraft. *Bull. Amer. Meteor. Soc.,* **84,** 203–216.

Rogers, E., and Coauthors, 2009: The NCEP North American Mesoscale Modeling System: Recent changes and future plans. Preprints, *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction,* Omaha, NE, Amer. Meteor. Soc., 2A.4. [Available online at http://ams.confex.com/ams/pdfpapers/154114.pdf.]

Smith, T. L., S. G. Benjamin, S. I. Gutman, and S. Sahm, 2007: Short-range forecast impact from assimilation of GPS-IPW observations into the Rapid Update Cycle. *Mon. Wea. Rev.,* **135,** 2914–2930.

Szoke, E., R. Collander, B. Jamison, T. Smith, T. Schlatter, S. Benjamin, and W. Moninger, 2006: An evaluation of TAMDAR soundings in severe storm forecasting. Preprints, *23rd Conf. on Severe Local Storms,* St. Louis, MO, Amer. Meteor. Soc., 8.1 [Available online at http://ams.confex.com/ams/23SLS/techprogram/paper_115158.htm.]

——, ——, ——, ——, S. G. Benjamin, W. R. Moninger, T. W. Schlatter, and B. Schwartz, 2007: Impact of TAMDAR data on RUC short-range forecasts. Preprints, *22nd Weather Analysis and Forecasting Conf.,* Park City, UT, Amer. Meteor. Soc., 4A.4. [Available online at http://ams.confex.com/ams/22WAF18NWP/techprogram/paper_124715.htm.]