9.4

IMPACT OF TAMDAR ON THE RUC MODEL: A LOOK INTO SOME OF THE STATISTICS

Edward J. Szoke[1,2], Stan Benjamin, Randall S. Collander[1], Brian D. Jamison[1], William R. Moninger, Thomas W. Schlatter[3], and Tracy L. Smith[1]
NOAA/Earth System Research Laboratory/Global Systems Division (ESRL/GSD), Boulder, Colorado
[1]Collaboration with the Cooperative Institute for Research in the Atmosphere (CIRA), Fort Collins, Colorado
[3]Collaboration with the Cooperative Institute for Research in the Environmental Sciences

1. INTRODUCTION

AMDAR (Aircraft Meteorological Data Relay) is a worldwide program providing automated real-time reports of atmospheric conditions from commercial airliners. AMDAR data have been available, and used by weather forecasters and in weather models, for nearly two decades (Moninger, et al. 2003). The most applicable AMDAR information used by forecasters has been wind and temperature data. Vertical soundings of wind and temperature are available from the major airports and some smaller airports. Otherwise, most of the traditional in-flight AMDAR is at higher levels, generally above 30,000 ft AGL.

While AMDAR has improved weather forecasting over the years, two limitations have existed; the lack of data below 20,000 ft between major airline hubs, and no measurement of moisture. The amount and distribution of moisture in the lower troposphere is critical for many weather forecasts, but accurate measurements of point observations of moisture above the surface have remained elusive, and the rawinsonde (hereafter, RAOB) remains the reliable standard for above-surface point moisture measurements. Unfortunately, RAOBs are generally available only twice per day, and the distance between RAOB sites is substantial.

In an attempt to fill both of these gaps, the NASA Aviation Safety Program funded the development of a sensor called TAMDAR (Tropospheric AMDAR) by AirDat, LLC, of Raleigh NC, designed for deployment on aircraft flown by regional airlines (Daniels et al. 2006). Beginning in mid January 2005, with the support of NASA and the FAA, these sensors have been deployed on 63 aircraft flying over the U. S. Midwest in an experiment called the Great Lakes Fleet Evaluation (GLFE, Daniels et al. 2006, website at http://www.crh.noaa.gov/tamdar/).

In addition to the added measurement of moisture, the aircraft taking part in the GLFE fly out of many smaller airports (in addition to major hubs) that typically do not have coverage from the current aircraft data, adding a considerable number of ascent/descent soundings. Furthermore, many of the flights are at levels well below the jet stream level of typical AMDAR aircraft, adding much data in the level between approximately 14 to 20 kft AGL. Typical coverage for TAMDAR flights is shown in Fig. 1, compared with non-TAMDAR flights.

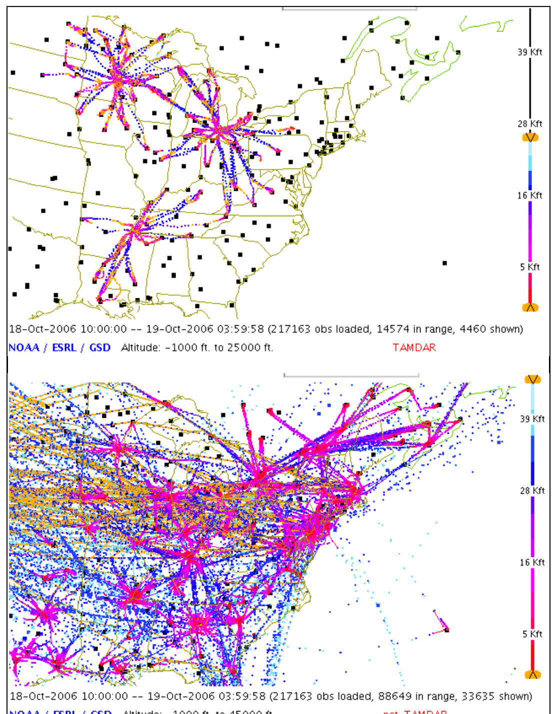

Fig. 1. Comparison of typical flight coverage between TAMDAR (top) and non-TAMDAR (bottom) aircraft. 16-h period from 1000 UTC on 18 Oct to 0400 UTC on 19 Oct 2006 is shown. Airports are represented by the small black squares.

It is apparent from Fig. 1 that the coverage of TAMDAR at this time is only a small fraction of the AMDAR coverage. Nonetheless, although difficult to detect in Fig. 1, there are a number of regional airports served by TAMDAR that are not in the current AMDAR coverage. And as noted earlier, the flight levels of the TAMDAR aircraft are lower

[2]*Corresponding author address:* Ed Szoke, NOAA/ESRL/GSD,R/E/GSD7, 325 Broadway, Boulder, CO 80305-3328;
e-mail: Edward.J.Szoke@noaa.gov

than for non-TAMDAR aircraft. Compare Fig. 1 (bottom) with Fig. 2, which depicts the non-TAMDAR data coverage for the same period but only for altitudes up to 25,000 ft AGL, matching the height restriction for Fig. 1 (top). Fig. 2 demonstrates how much of the dense coverage occurs at fairly high altitudes, with TAMDAR (Fig. 1, top) nicely filling this gap in vertical coverage. In addition, Fig. 2 gives a much better depiction of the airports served by non-TAMDAR aircraft, mainly the larger airports. Note, as an example, that only two airports are served by non-TAMDAR aircraft in Minnesota, compared with nine airports for TAMDAR aircraft. And this is with the limited TAMDAR aircraft participating in the GLFE. The airport coverage shown in Fig. 2 limits the number of vertical ascent/descent sounding locations for the current AMDAR fleet, a type of display often used by forecasters in many applications.

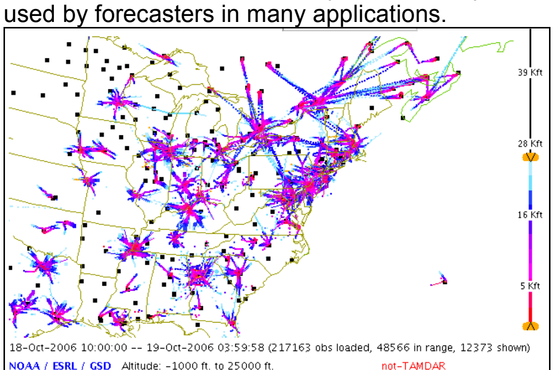

Fig. 2. Non-TAMDAR coverage for the same period as in Fig. 1b, but restricted to altitudes at and below 25,000 ft AGL.

NOAA/Global Systems Division (GSD) has been part of an extensive effort to evaluate the experimental TAMDAR data, both in terms of the quality of the data, and its impact on weather forecasting. The impact on forecasting has been approached on two levels; the quantitative impact of TAMDAR on numerical weather prediction (NWP), and a more subjective examination of the potential impact on various weather forecast problems that face forecasters at the National Weather Service (NWS). Others have been involved in this later effort, with a lead role taken by the NWS Weather Forecast Office (WFO) at Green Bay, Wisconsin (Mamrosh et al. 2006; Brusky et al. 2006a, b; Kurimski et al. 2006). AirDat has also examined the impact of TAMDAR on weather problems (Druse and Jacobs, 2007). The use of TAMDAR at the NWS Aviation Weather Center (AWC) has also been documented (Fischer 2006).

The impact of TAMDAR on numerical weather prediction has been examined at NOAA/GSD through the use of the Rapid Update Cycle (RUC, Benjamin et al. 2004) assimilation and model system. In order to test the impact of TAMDAR, two versions of the RUC model have been run in real-time for the last two years at 20-km horizontal grid resolution (Moninger et al. 2006, Benjamin et al. 2006). The RUC analysis independently assimilates observations to include TAMDAR data for one model run but excludes them for the other run. Model forecasts are made at 1-h intervals to 3 h, and at 3-h intervals out to at least 12 h. A number of pregenerated images are made for each run, as well as several other RUC real-time runs, at the GSD, and are available online at http://ruc.noaa.gov/. The RUC runs with TAMDAR are labeled “20 km dev2 RUC”, and those without TAMDAR “20 km dev RUC”.

A long-term set of statistics has been generated from this pair of RUC runs for wind, temperature, and relative humidity by comparing the model forecasts to RAOBs at 0000 and 1200 UTC. The latest results are discussed in companion papers at this conference (Benjamin et al. 2007, Moninger et al. 2007). Subjective evaluations of the RUC forecasts with and without TAMDAR through case studies examining various weather phenomenon (Szoke et al. 2006) have also been conducted. This paper focuses on the statistics that have been generated from the companion RUC runs, in particular the humidity verification.

2. METHODOLOGY

In general, the long-term statistics for the companion RUC runs have shown that the TAMDAR data have a positive impact for temperature, wind, and humidity. The impact for humidity, however, has been the most variable, with a fairly high negative impact (worse forecast with TAMDAR) on some of the days. The attempt to understand the cause of this negative impact led to the study reported in this paper.

First, some of the more noteworthy negative impact days from the statistics were identified. The long-term statistics were generated for the RAOB significant levels, 850 and 700 mb being the focus. Most of the TAMDAR aircraft fly between 1100 UTC and 0200 UTC, so impact on the RUC model should be most apparent for forecasts made after 1200 UTC. Thus, forecasts valid at the 0000 UTC RAOB time were examined, and shorter-term 3-h and 6-h forecasts from the last six months were the focus. Statistics were calculated for a limited domain comprising the most dense TAMDAR data, a larger domain containing all the TAMDAR data, and a full CONUS domain (Fig. 3). This study was restricted to the more manageable limited domain, which comprises 13 RAOB sites. For an identified case, each of the 13 RAOB sites was examined and compared to the RUC forecast sounding with and without TAMDAR, using a sounding display tool developed at NOAA/GSD.

Initial examination of some of the days in which the RUC run with TAMDAR scored worse than the

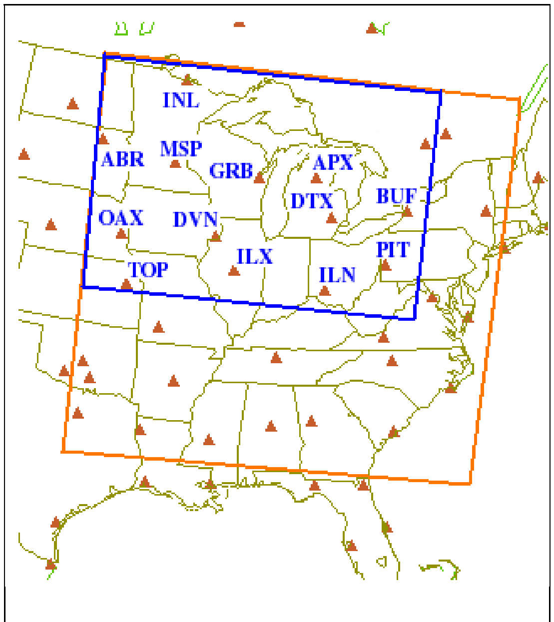

Fig. 3. Small (blue box) and large (red) TAMDAR domains. RAOB sites used within the small domain are labeled.

run without TAMDAR suggested that most of the error for the 14 RAOB TAMDAR domain could come from one or two RAOB sites. Furthermore, at these sites the large humidity error sometimes resulted from a slight shift in the vertical of the forecast moisture profile compared to that observed. In such cases, the model forecast could be penalized quite severely if the moisture profile was off by only a small vertical amount but happened to fall at the verification level. Based on these preliminary results, we hypothesized that at least some of the poor scores were not representative of a bad overall performance by the RUC run with TAMDAR, but reflected a penalty arising from the calculation of error at just a few specific heights. A goal of this study then was to determine if this hypothesis could be generalized, and, if so, should the manner in which the statistics are calculated be changed to better reflect the overall error between forecast and observed humidity.

3. CASES

The study focused on statistics from mid-June through mid-October 2006, picking cases to scrutinize. Obvious outliers were identified, using the 3-h and 6-h forecasts valid at 0000 UTC, and the 700 mb and 850 mb levels. The daily RMS errors for the above four combinations are shown in Fig. 4, with cases of interest identified. The cases of interest are discussed below, including two cases where dev (RUC without TAMDAR) performs notably worse than dev2 (RUC with TAMDAR).

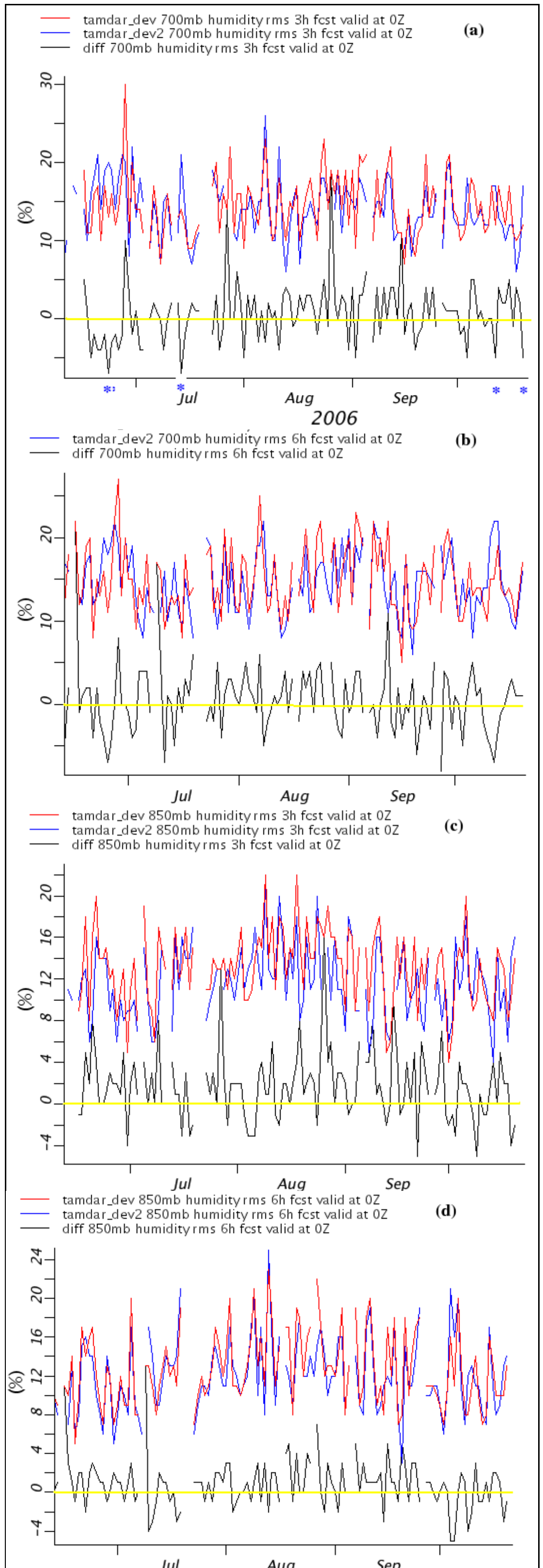

Fig. 4. RMS relative humidity error for dev and dev2, and difference (dev-dev2, with zero line highlighted) for mid Jun-mid Oct 2006.

3.1 *23 Jun 2006: 3-h forecasts at 700 mb.*

The RMS error for the 3-h dev2 700 mb humidity forecasts ending at 0000 UTC on 23 June 2006 was close to 7% worse than for dev. An error of this magnitude was infrequent (Fig.4a). Investigation of the 13 RAOB comparisons revealed that dev2 was substantially worse than dev at the 700 mb level at only two of the RAOB sites, Peoria (ILX), Illinois (Fig. 5), and Pittsburgh (PIT), Pennsylvania (Fig. 6). The comparison at ILX illustrates the points raised earlier in the hypothesis. Qualitatively, the dev2 forecast appears to match the RAOB through the 400 mb level better than the dev forecast. In particular, the dry layer centered near 750 mb in the RAOB is pretty well captured by the dev2 forecast, and better than in the dev forecast. However, the driest point within the dry layer in the dev2 forecast is about 50 mb too high, which happens to put it right at the 700 mb level. Consequently, a very substantial error in relative humidity (RH) results when calculated at the 700 mb level. In this instance, the dev2 RH at 700 mb is 34%, dev is 94%, and the RAOB is 74%, yielding a 60% error for dev2, three times that for dev. This contributes to most of the overall RMS error on this day.

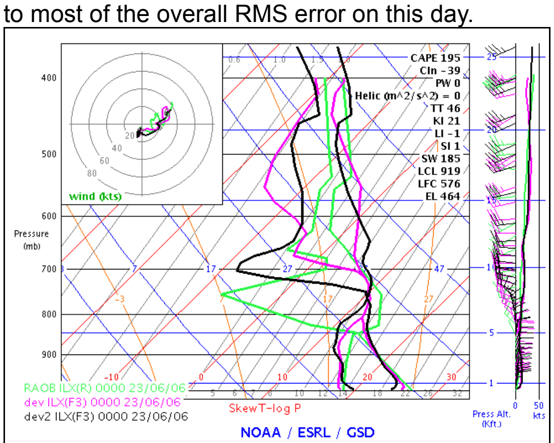

Fig. 5. Comparison of 3-h forecasts from dev (blue) and dev2 (black) RUC runs, valid at 0000 UTC on 23 June 2006 for Peoria, Illinois, with the Peoria 0000 UTC RAOB (green).

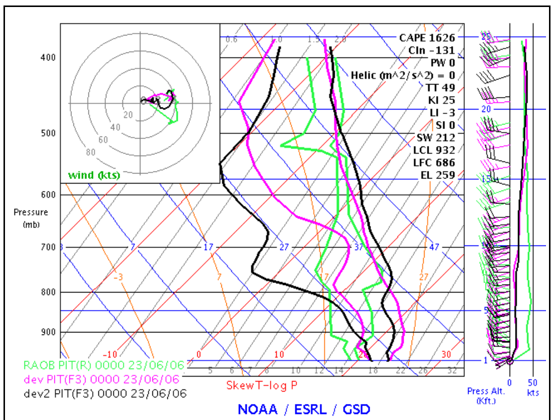

Fig. 6. As in Fig. 5, for Pittsburgh, Pennsylvania.

At Pittsburgh (Fig. 6), the verification issue is different. The overall shape of the two dewpoint profiles from the RUC forecasts is relatively similar, but the dev2 profile is shifted to the left. This excessive drying in the dev2 forecasts exists over a deep layer, so at this RAOB site the error at 700 mb (39% too dry for dev2 vs. 12% too moist for dev) is a good representation of the overall dev2 forecast sounding error.

3.2 *14 Jul 2006: 3- and 6-h forecasts at 700 mb.*

Large forecast errors were present for both the 3-h and 6-h forecasts from dev2 ending at 0000 UTC on 14 July 2006, with dev2 approximately 7% worse than dev for the 3-h forecasts and 5% worse for the 6-h forecasts. For the 3-h forecasts, four of the 14 verifying RAOB sites had notably greater errors in the moisture profile for the dev2 forecast than for the dev forecast. One of these sites (Buffalo (BUF), New York, Fig. 7) displayed some of the verification issues that were seen on 23 Jun. In this case a very sharp but vertically limited dry layer is observed near 750 mb. It would be unrealistic to expect the RUC model to be able to predict such a dry layer that only extends over ~30 mb in depth. Of the two forecasts, however, the dev2 run does a better job than the dev run of matching the drying beginning near the 800 mb level. The dev2 run is also closer to matching the magnitude of the drying, but is off by ~20 mb in the vertical. This leads to an RH of 12% at 700 mb for dev2, compared to 22% for the RAOB. Meanwhile, it happens that the moisture profile for the dev model forecast matches the RAOB exactly at 700 mb. The RH is off slightly for dev (2%) owing to a temperature difference at 700 mb, but ends up 8% better than for dev2. In this case the RH differences are not as great as in the previous case, in part because the overall conditions are drier, so there is less of a change in RH for a given change in dewpoint.

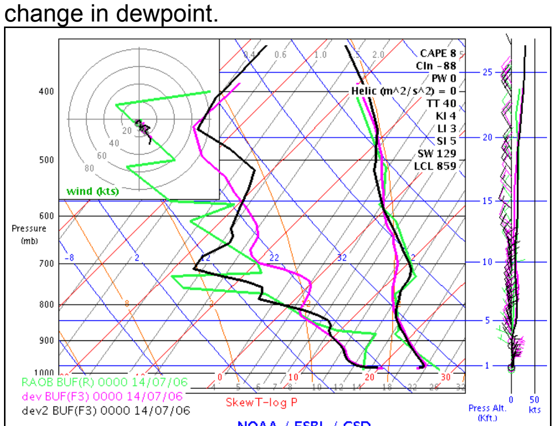

Fig. 7. As in Fig. 5, for Buffalo (BUF), New York, 3-h forecasts valid 0000 UTC 14 July 2006.

The RAOB comparison at Aberdeen (ABR), South Dakota (Fig. 8) is representative of the error for the 3-h dev2 forecasts with the other three

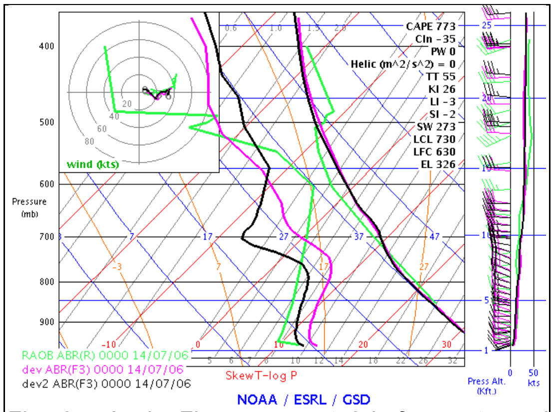

Fig. 8. As in Fig. 7, except 3-h forecasts and RAOB for Aberdeen (ABR), South Dakota.

soundings that had larger dev2 3-h forecast errors. In this case, the shape of the two moisture forecasts is similar, but shifted erroneously towards a drier solution for dev2. The overall comparison of the dev forecast to the RAOB is not particularly good, but at 700 mb the dev dewpoint forecast happens to nearly agree with the RAOB, while the shifted dev2 forecast has a relative minimum in dewpoint at 700 mb. The result is an exaggerated poor verification for dev2 and the opposite for dev at the 700 mb level (32% RH for dev2, 50% for dev, and 55% for the RAOB). Note that below ~780 mb the dev2 moisture forecast is actually closer to the RAOB than the dev forecast.

Investigation of the 6-h forecasts found that the biggest RH error at 700 mb actually occurred for a dev forecast, but the overall RMS RH dev2 error was 5% worse than dev owing to smaller errors for dev2 at close to half of the RAOB sites. An example of one of the more interesting comparisons, at BUF, is shown in Fig. 9. The RAOB at BUF has a vertically narrow layer of higher moisture centered at 700 mb within an overall dry layer. This feature is not captured by either RUC model forecast, which have a dry layer near 700 mb, though less dry for the dev forecast. There was nothing in the synoptic data to suggest

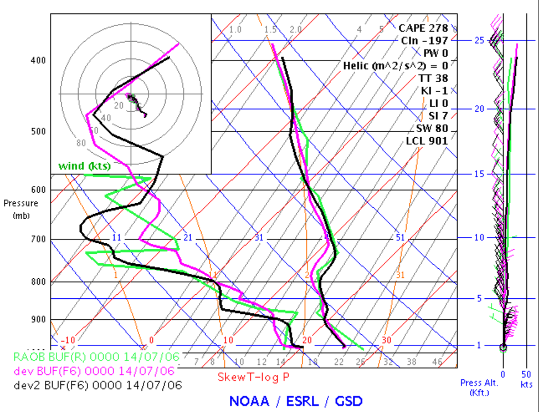

Fig. 9. As in Fig. 8, but for 6-h forecasts.

why this more moist layer was present, but if it did not exist, the dewpoint trace would look much like the dev2 forecast. However, because it is present, the RH error at 700 mb for dev2 is 15%, versus 7% for dev.

3.3 *12 Oct 2006: 3- and 6-h forecasts at 700 mb.*

Forecasts valid at 0000 UTC on 12 October are examined at the 700 mb level. The RMS RH error for the 6-h dev2 forecasts at 700 mb was 7% worse than for dev, and for the 3-h forecasts 5% worse. For the 6-h forecasts, one large error at 700 mb at a single site, Green Bay (GRB), Wisconsin, was responsible for most of the 7% error. The forecast comparison with the RAOB for GRB in Fig. 10 shows the much drier conditions in the dev2 forecast for a deep layer, yielding a huge error at 700 mb, where the RAOB is nearly saturated (88% RH), compared to 83% for dev and 22% for dev2. The RH difference between the two forecasts is 61%, the highest observed for this study.

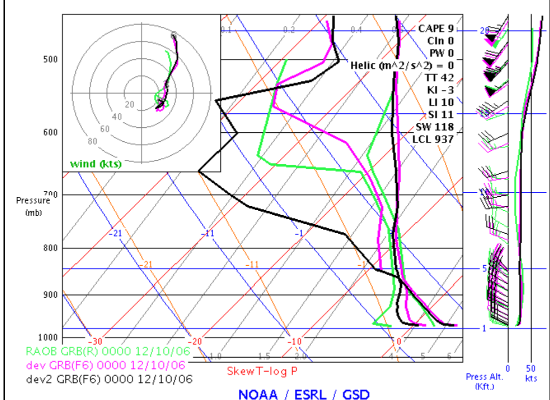

Fig. 10. RUC 6-h forecasts from dev and dev2 compared to 0000 UTC 12 Oct 2006 GRB RAOB.

Though certainly reflecting a huge error, under closer examination, Fig. 10 does indicate a much drier layer was present in reality, but the drying began around 670 mb. Thus, the dev2 forecast may not have been as bad as that error would suggest, but instead erred more due to having a deeper dry layer that was not as sharp as what was present. Note that the dev forecast also has the dry layer, but it begins above where it was observed. The synoptic conditions at 700 mb (Fig. 11) help explain the differences, as the trough axis is just passing GRB at 0000 UTC. With this position of the trough axis, slight timing differences between the two forecasts, with the dev2 run a bit faster, could have yielded the very large RH differences.

For the 3-h forecasts, most of the error differences between the two forecasts were relatively small, but added up to favor dev. The worst verification for dev2 at 700 mb occurred at Minneapolis (MPX), Minnesota (Fig. 12). The

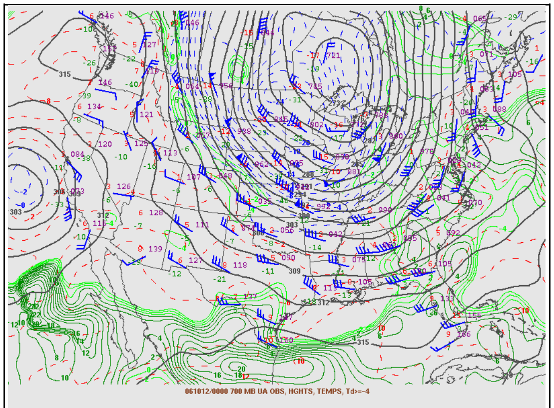

Fig. 11. Height and temperature analysis at 700 mb with plots for 0000 UTC 12 October 2006.

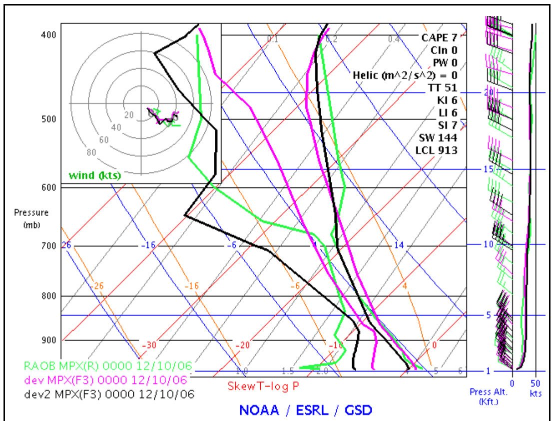

Fig. 12. As in Fig. 10, but 3-h forecasts for MPX.

excessive drying through a deeper layer than was present in the RAOB is consistent with the verification problems discussed above for the 6-h forecast at GRB. As with the GRB RAOB, there was, in fact, a dry layer just above 700 mb of similar magnitude to that forecast by dev2, but the transition from moist to dry conditions is much sharper in reality than in the forecast. Note that the dev forecast, while verifying better at 700 mb (74% RH, versus 78% observed and 45% for dev2), does not capture the observed drying.

3.4 *9 Oct 2006: 3- and 6-h forecasts at 850 mb.*

The RMS errors at 850 mb for RH were 5% worse for dev2 for the 3-h forecasts and 4% worse for the 6-h forecasts. At some of the sounding sites it was apparent that the dev forecast was a better forecast through a relatively deep layer. But at other RAOB sites the problem of verifying sharp RH changes in the vertical was apparent, and is illustrated for a representative RAOB at Davenport (DVN), Iowa in Fig. 13. Four very sharp relative maximum and minimum RH points are observed in the lowest 700 mb. Aside from the relative maximum near 750 mb, the dev2 forecast captures the structure of the RAOB RH better than the dev forecast. However, at the 850 mb level, dev2 ends up verifying worse than the inferior dev forecast

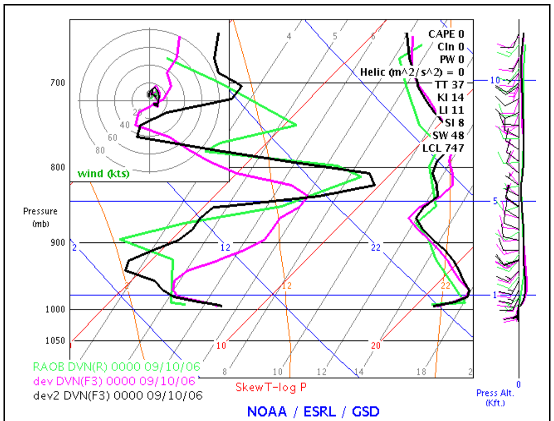

Fig. 13. RUC 3-h forecasts from dev and dev2 compared to 0000 UTC 9 Oct 2006 DVN RAOB.

(42% RH for dev2, 59% for dev, and 52% for the RAOB), even though the dev2 forecast does a much better job of capturing the RH maximum near 810 mb.

3.5 *20 Oct 2006: 3-h forecasts at 700 mb.*

Two (ABR and International Falls (INL), Minnesota) of the 14 RAOB comparisons accounted for most of the 4.5% RMS RH difference between the dev2 and dev forecasts valid at 0000 UTC on 20 Oct. Here the forecasts for INL, shown in Fig. 14, are compared. A qualitative look at the two forecasts indicates the dev2 forecast has a better representation of the observed RH, at least below ~650 mb. The dev forecast fails to capture any of the rather deep dry layer between 800 and 700 mb. Dev2 has a vertically sharper dry layer than was observed, and ends up being more moist at the 700 mb layer than dev and the RAOB (70% RH for dev2, versus 42% RH for dev, and only 12% for the RAOB). The result is dev2 scoring 28% worse than dev at the 700 mb level.

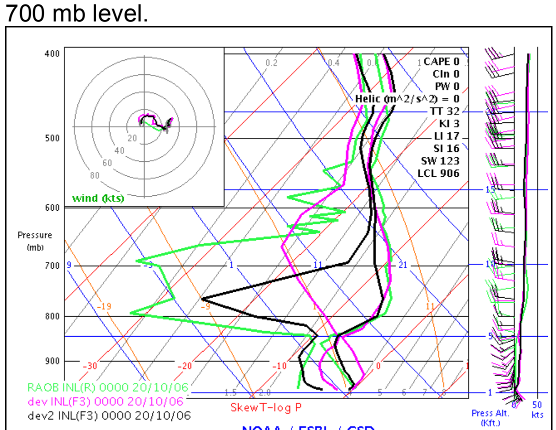

Fig. 14. RUC 3-h forecasts from dev and dev2 compared to 0000 UTC 20 Oct 2006 INL RAOB.

As with the 12 October case, a trough was passing the sounding site of interest (here, INL) at

0000 UTC (Fig. 15). The RAOB from INL in Fig. 14 indicates a deep layer of drying behind the trough axis. This drying is better captured by the dev2 forecast, yet the dev forecast ends up verifying better at the 700 mb level.

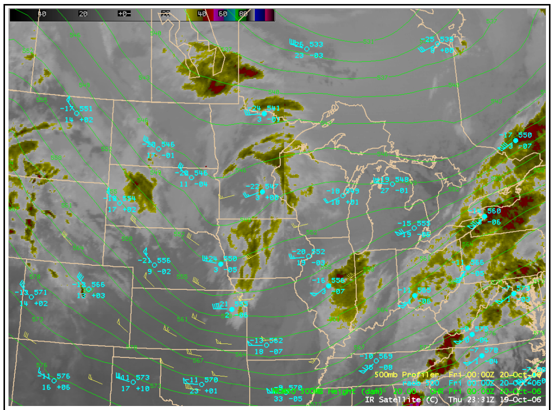

Fig. 15. 500 mb plot with RUC height analysis (dm) at 0000 UTC 20 Oct, with 2331 UTC 19 Oct IR satellite image.

3.6 *Examples of better dev2 verification*

Of course, there are many days when the RUC run with TAMDAR (dev2) verified better than the run without TAMDAR (dev), and a couple of examples are shown here to illustrate that similar verification issues occur. The first case examines the 3-h forecasts at 700 mb valid on 0000 UTC 28 June 2006, when the RH RMS difference was 10%, favoring dev2. Several of the verifying RAOB sites had errors favoring either dev or dev2, and at one site (GRB) dev2 actually verified considerably worse than dev. As in some of the cases that verified better for dev, the favorable overall error for dev2 came down mostly to the verification at a single site, Wilmington (ILN), Ohio (Fig. 16). The 700 mb RH observed at ILN was 78%, compared to 61% for dev2, and a very dry 13% for dev. giving a 48% greater RH error for the dev run. Note that both runs have the same decrease in moisture with height, but for the dev

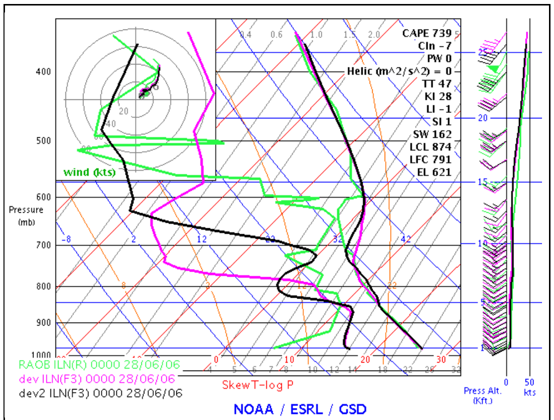

Fig. 16. RUC 3-h forecasts from dev and dev2 compared to 0000 UTC 28 Jun 2006 ILN RAOB.

forecast this occurs ~80 mb lower than for dev2, so that at 700 mb dev2 ends up with a much better verification. Both would have verified equally as poorly at a higher level, for example near 630 mb.

A similar case is shown in Fig. 17, comparing the two model forecasts to the Minneapolis, Minnesota RAOB (here labeled MSP). The overall RMS RH error at 700 mb was ~4% better for dev2 over dev, with most of the differences coming down to the comparison at MSP. The better verification at 700 mb for this case results from a similar displacement of the dry layer near 700 mb that was found for the 28 June case, except that the shift is only ~30 mb. Both forecasts attempt to resolve the very sharp dry layer above 700 mb in the RAOB, but the dev2 forecast happens to verify much better at 700 mb. The RH error for dev2 was 31% better than dev for this case.

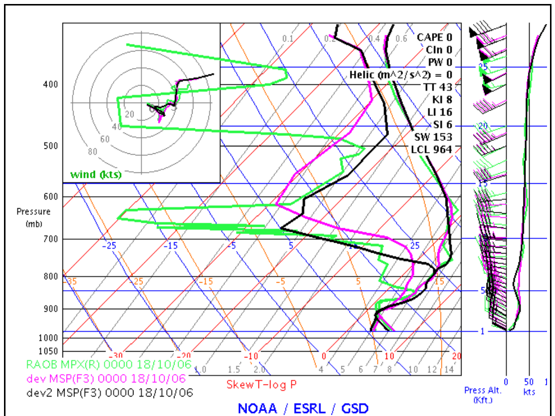

Fig. 17. RUC 3-h forecasts from dev and dev2 compared to 0000 UTC 18 Oct 2006 MSP RAOB.

4. DISCUSSION

NOAA/GSD has been undertaking a long-term ongoing effort to demonstrate the potential impact of TAMDAR on numerical model forecasts via companion runs of the RUC model with (dev2) and without (dev) TAMDAR. Objective verification has generally shown consistent positive results for temperature and wind, but it has been more difficult to find consistent positive impact for RH, which is the new measurement made by TAMDAR aircraft. This study was undertaken to further understand why this variability existed, and in particular, what accounted for the days when the RUC forecasts with TAMDAR verified considerably worse than those not using the data. Using short-term forecasts verifying at 0000 UTC for a subset of the overall GLFE region, each of the 13 RAOBs within this region for each case were examined .

Examples from the overall collection of cases were shown in Section 3. It was found that:

- Outlier cases had typical RH RMS differences between 4% and 8%. Several such cases were identified at both the 700 mb and 850 mb level over the Jun-Oct 2006 period, either favoring or not favoring the RUC run with TAMDAR (dev2).

- Generally a large, obvious error at a single (out of 13) RAOB site accounted for the verification differences.
- More rarely, an accumulation of smaller errors at a number of sites added up to the overall error.
- The single large errors at a given site often resulted from a small vertical shift in the moisture profile, rather than an overall poor forecast.
- In such situations as outlined above, the practice of scoring at a single level yielded a poor verification that did not represent the real quality of the forecast. In fact, in some cases the other (dev, for example, for the cases when dev2 scored worse) forecast could be scored far higher than seemed appropriate, based on an overall subjective assessment of the forecast versus the RAOB.

As a result of these conclusions, other verification ideas were explored. This led to a verification tool that compared the RUC forecasts to the RAOB at 10 mb vertical intervals. The verification interface allows one to choose any layer (or still a single layer) and apply the 10 mb increment verification. A comparison of this new method with our previous verification will be shown in more detail at the conference. A preliminary comparison with a few of the cases shown in Section 3 is discussed here.

The effect of the new scoring method is shown for the October cases in Fig. 18. For all the cases (highlighted by the yellow vertical lines) the RMS error differences that were present at 700 or 850 mb are reduced. For 9 Oct and 18 Oct the new scoring produces no difference between the dev and dev2 forecasts. For 9 Oct the 4% RMS difference that was present at 850 mb (Fig. 18c) is gone when the new method is employed (Fig. 18a). The 18 Oct case was used as an example of dev2 scoring considerably *better* than dev at 700 mb, but under the new scoring method this difference is gone. It was shown (Section 3.6) how a vertically narrow layer in the verifying RAOB happened to favor the dev2 forecast, and this effect is removed with the new scoring method.

For 12 Oct there was a collection of smaller errors for several sites and then one example shown for MPX (Fig. 12). The new scoring method reduces the overall error by about half. Examination of the RAOB/forecast comparison in Fig. 12 suggests that the new method of verifying at 10 mb intervals would still have penalized dev2 (correctly) for being too dry in a deep layer near 700 mb, but it would have improved the score at other levels not accounted for in the old method.

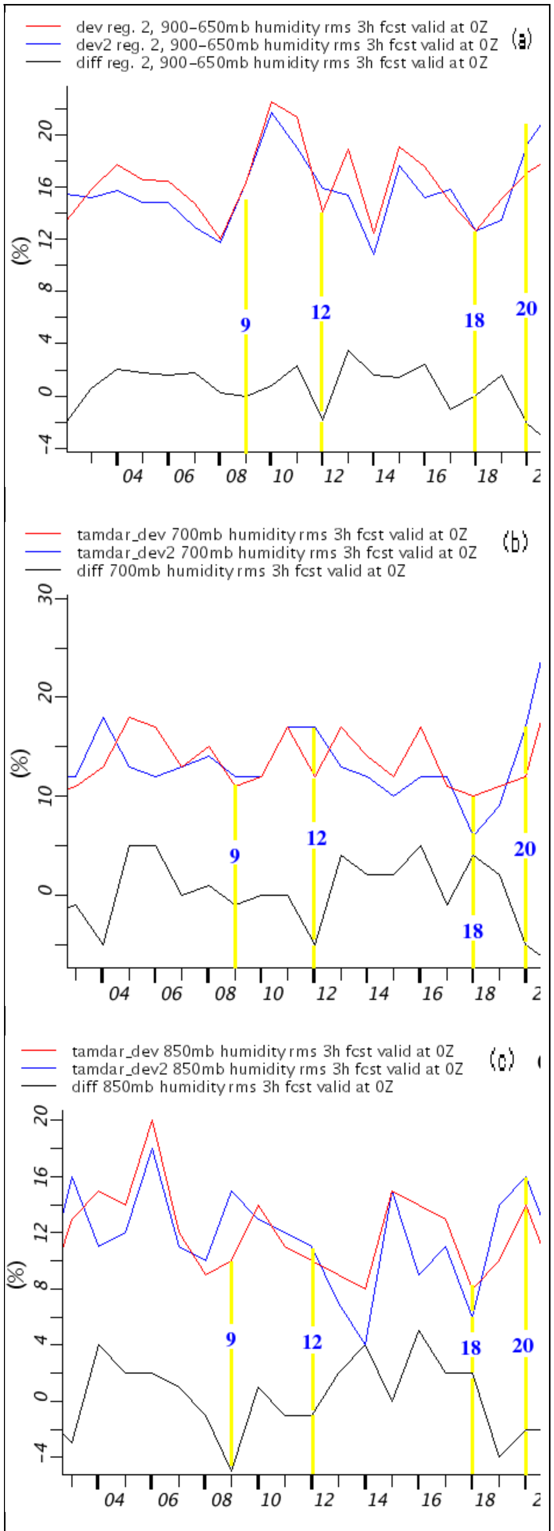

Fig. 18. RMS RH scores (%) for 3-20 Oct 2006, for 3-h forecasts valid 0000 UTC, with the days of interest highlighted. New scoring method for 900 to 650 mb layer in (a), and older scoring method for single levels of 700 mb (b) and 850 mb (c).

Similarly, the RMS RH error for 20 Oct was cut in half using the new scoring method, but would still favor dev as the better overall forecast. Looking back at the key INL comparison for this case (Fig. 12) suggests that while the new method would have resulted in dev2 getting more credit for trying to resolve the dry layer than scoring the runs just at 700 mb, dev2 would have scored worse in the 700 to 650 mb layer where it was too moist.

Early results, then, indicate that the new method of scoring helps remove some of the verification issues that resulted in worse scores for forecasts that subjectively appeared to be superior. Whether or not this is the best method to use remains to be tested. The new method may still be subject to unrepresentative errors when the RAOB has rapidly fluctuation changes in RH in the vertical, for example, as in the upper levels in Fig. 17. Whether such changes are real or not, they probably are often not representative of the overall environment, and it might be unrealistic to expect a model with the resolution of the RUC to resolve such features. In fact, one might not really be interested in trying to verify such features if they do not represent the applicable scales for the model. This issue suggests that some type of averaging or smoothing of the RAOB RH before applying a scoring technique might be avenue to examine. Another idea might be to calculate layer RH averages for both the forecast and the RAOB and use these for verification. Potential alternatives will be explored and the new scoring method further evaluated. Any progress on these investigations will be discussed at the conference.

5. ACKNOWLEDGMENTS

This research is in response to requirements and funding by the Federal Aviation Administration (FAA) under interagency agreement DTFAWA-03-X-02000. The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA. We thank Chris Anderson of GSD for an internal scientific review and Annie Reiser of GSD for a technical review.

6. REFERENCES

Benjamin, S., W. Moninger, T. L. Smith, B. Jamison, and B. Schwartz, 2006: TAMDAR aircraft impact experiments with the Rapid Update Cycle. *10th Symposium on IOAS-AOLS*, Atlanta, GA, Amer. Meteor. Soc., Paper 9.8.

Benjamin, S., W. Moninger, T. L. Smith, B. Jamison, and E.J. Szoke, 2007: 2006 TAMDAR impact experiment results using the RUC at NOAA's Global Systems Division. *11th Symposium on IOAS-AOLS*, San Antonio, TX, Amer. Meteor. Soc., Paper 9.1.

Benjamin, S., D. Devenyi, S. S. Weygandt, K. J. Brundage, J. M. Brown, G. A. Grell, D. Kim, B. E. Schwartz, T. G. Smirnova, and T. L. Smith, 2004: An Hourly Assimilation-Forecast Cycle: The RUC. *Mon. Wea. Rev.,* **132**, 495-518.

Brusky, E.S., and P. Kurimski, 2006a: The Utility of TAMDAR Regional Aircraft Sounding Data in Short-term Convective Forecasting. *10th Symposium on IOAS-AOLS*, Atlanta, GA, Amer. Meteor. Soc., Paper 9.5.

Brusky, E.S., and R.D. Mamrosh, 2006b: The Utility of Aircraft Soundings in Assessing the Near Storm Environment. *23rd Conference on Severe Local Storms*, St. Louis, MO, Amer. Meteor. Soc., Paper 2.1.

Daniels, T.S., W.R. Moninger and R.D. Mamrosh, 2006: Tropospheric Airborne Meteorological Data Reporting (TAMDAR) Overview. *10th Symposium on IOAS-AOLS*, Atlanta, GA, Amer. Meteor. Soc., Paper 9.1.

Druse, C.M., and N.A. Jacobs, 2007: Tropospheric Airborne Meteorological Data Reporting (TAMDAR) Overview. *11th Symposium on IOAS-AOLS*, San Antonio, TX, Amer. Meteor. Soc., Paper 9.5.

Fischer, A., 2006: The Use of TAMDAR (Tropospheric Airborne Meteorological Data Reporting) as a Convective Forecasting Supplement in the Northern Plains and Upper Midwest. *10th Symposium on IOAS-AOLS*, Atlanta, GA, Amer. Meteor. Soc., Paper 9.6.

Kurimski, P. and E.S. Brusky, 2006b: Application of Aircraft Sounding Data in Short-term Convective Forecasting. *23rd Conference on Severe Local Storms*, St. Louis, MO, Amer. Meteor. Soc., Paper 2.2.

Mamrosh, R.D., E. S. Brusky, J. K. Last, E. J. Szoke, W. R. Moninger, and T. S. Daniels, 2006: Applications of TAMDAR Aircraft Data Reports in NWS Forecast Offices. *10th Symposium on IOAS-AOLS*, Atlanta, GA, Amer. Meteor. Soc., Paper 9.4.

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

Moninger, W.M., F. Barth, S. G. Benjamin, R. S. Collander, B. D. Jamison, P. A. Miller, B. E. Schwartz, T. L. Smith, and E. Szoke, 2006: TAMDAR evaluation work at the Forecast Systems Laboratory: an overview. *10th Symposium on IOAS-AOLS*, Atlanta, GA, Amer. Meteor. Soc., Paper 9.7.

Moninger, W.M., F. Barth, S. G. Benjamin, R. S. Collander, B. D. Jamison, P. A. Miller, B. E. Schwartz, T. L. Smith, and E. Szoke, 2007: TAMDAR evaluation work at the Forecast Systems Laboratory: an overview. *10th Symposium on IOAS-AOLS*, Atlanta, GA, Amer. Meteor. Soc., Paper 9.7.

Szoke, E.J., B.D. Jamison, W.R. Moninger, S.G. Benjamin, B.E. Schwartz, and T.L. Smith, 2006: Impact of TAMDAR on RUC forecasts: Case studies. *10th Symposium on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*, Atlanta, GA, Amer. Meteor. Soc., Paper 9.9.