Relative short-range forecast impact from aircraft, profiler, radiosonde, VAD, GPS-PW, METAR and mesonet observations via the RUC hourly assimilation cycle

Stanley G. Benjamin, Brian D. Jamison[1], William R. Moninger, Susan R. Sahm, Barry E. Schwartz[2], and Thomas W. Schlatter[3]

NOAA Earth System Research Laboratory, Boulder, CO

Revised manuscript for publication in

Monthly Weather Review

25 September 2009

Initial submission - 10 June 2009

Notification of acceptance – 2 August 2009

[1] Collaboration with the Cooperative Institute for Research in the Atmosphere (CIRA), Fort Collins, CO, USA

[2] Affiliated with Systems Research Group, Colorado Springs, CO, USA

[3] Retired

Abstract

An assessment is presented on the relative forecast impact on the performance of a numerical weather prediction model from eight different observation data types (aircraft, profiler, radiosonde, VAD (velocity azimuth display), GPS-derived precipitable water, METAR (surface), surface mesonet, and satellite-based atmospheric motion vectors. A series of observation sensitivity experiments was conducted using the Rapid Update Cycle (RUC) model/assimilation system in which various data sources were denied to assess the relative importance of the different data types for short-range (3-12h) wind, temperature, and relative humidity forecasts at different vertical levels and at the surface. These experiments were conducted for two 10-day periods, one in November-December 2006 and one in August 2007.

These experiments show positive short-range forecast impacts from most of the contributors to the heterogeneous observing system over the RUC domain. In particular, aircraft observations had the largest overall impact for forecasts initialized 3-6 h before 0000 or 1200 UTC, considered over the full depth (1000-100 hPa), followed by radiosonde observations, even though the latter are available only every 12h. Profiler data, GPS-precipitable water estimates, and surface observations also led to significant improvements in short-range forecast skill.

1. Introduction

An increasing number of atmospheric observation systems are used to initialize operational numerical weather prediction (NWP) models. Observation system experiments (OSEs) have been found very useful in determining the impact of particular observation types on operational NWP systems (e.g., Graham et al. 2000; Bouttier and Kelly 2001; Zapotocny et al. 2002, 2007; Lord et al. 2004; Cardinali 2009). OSEs can provide a basis for decisions regarding the design and implementation of current and future observing systems.

Such studies have provided valuable guidance on relative expenditures for different observational systems and where expansions of current limited areal deployments for certain observing systems (e.g., NOAA Profiler Network - NPN) might be most helpful toward improved NWP guidance. As heterogeneity of the overall composite observing system increases and as data assimilation and modeling techniques are improved, new OSEs will be needed to evaluate these new configurations.

This study uses a commonly used OSE design, with different observation types being excluded from the data assimilation system for separate experiments to measure effects on subsequent NWP forecasts, with the *control* experiment using all available observation types. This study differs from adjoint-based observation sensitivity experiments (e.g., Cardinali 2009; Zhu and Gelaro 2008; Baker and Daley 2000, others). As described by Cardinali (2009), the adjoint-

based sensitivity method tests the impact of all observations from a given time (or a short time over which a tangent linear model is run). An OSE by comparison shows impact over a longer period and also requires a much larger number of experiments (a separate experiment for each denial of a given observation type or subset, as done here). In contrast to those experiments mentioned above, this OSE study is performed using a regional model/assimilation system, whereas those previous listed (except for Zapotocny et al 2002) were performed using global systems. Finally, the OSE data *denial* approach used in this study also differs from a data addition approach using a baseline control with, for instance, radiosondes only and adding other observation types to this control, one at a time.

This new OSE study is also unique in that it considers the very short-range forecast (3-h to 12-h) effects from most of the currently assimilated high-frequency observing systems in a 1-h assimilation cycle, the Rapid Update Cycle (RUC, Benjamin et al. 2004a), which runs at the highest assimilation frequency of operational NWP models operated at the NOAA National Centers for Environmental Prediction (NCEP). RUC short-range forecasts are heavily used as guidance for aviation, severe weather, energy, and other applications, some applying automated decision support algorithms suitable for hourly-updated NWP systems. Therefore, consideration of observation impact on very short-range (1-h to 12-h) forecasts is important in considering investment in these observation systems, both from regional and global perspectives.

The new study is similar to the previous observation impact experiments also using the RUC reported in Benjamin et al. (2004c), which consider only wind forecast impact from wind profilers over a 13-day winter period. This new impact study is much broader than the previous study, now for a greater number of observation types over both summer and winter experiment periods, and for three fields: wind, temperature, and moisture. Other previous work on effects of high-frequency (hourly) observations on short-range forecasts include those reported by Smith et al. (2007) for GPS precipitable water observations and Weygandt et al. (2004) for simulated lidar wind observations (a regional observing system simulation experiment – OSSE).

The observation sensitivity experiments reported here were carried out using a 2007 version of the RUC, including both assimilation system and forecast model components. The observing systems considered in this study include seven primary wind/temperature observation types over the United States (US): radiosonde observations (raobs), aircraft, METAR (surface aviation reports), mesonet, wind profilers, VAD (velocity azimuth display) wind profiles from WSR-88D radars, and satellite AMVs (atmospheric motion vectors, sometimes less precisely called cloud-drift winds). All these observing systems except radiosondes provide hourly data. This study also includes the primary tropospheric moisture observation types (radiosondes, GPS ground-based precipitable water (PW, Smith et al. 2007)). Relative effects of METAR and mesonet surface observations are also considered. We do not consider effects of satellite-measured radiances or retrieved soundings from satellite radiances in

this study (neither are assimilated in the RUC – Benjamin et al. 2004b, section 2). Finally, this study also includes the relative impact of actual 16-km (full depth) NOAA profilers versus hypothetical profilers with only a 8-km vertical range.

This paper accompanies a report by Moninger et al. (2009) that focuses on a multiyear data impact study specifically for TAMDAR-based (Tropospheric Aircraft Meteorological Data and Recording) observations from regional commercial aircraft. The companion paper includes results from a TAMDAR-denial experiment to measure TAMDAR impact during the same test periods used in this paper.

2. RUC version used for OSEs

The version of the RUC used in these experiments employs the same code as the 13-km version run operationally at NCEP as of March 2007, including 50 hybrid isentropic-sigma vertical levels and model physical parameterizations as described by Benjamin et al. (2004b), including five-species mixed-phase bulk cloud microphysics, Grell-Devenyi convective parameterization, and RUC-Smirnova land-surface model. For computational efficiency, these experiments were run at 20-km resolution with no other modifications except for this resolution modification via a single parameter. The hourly intermittent assimilation cycle in the RUC (Benjamin et al. 2004a) allows full use of hourly observational data sets. The analysis method is the three-dimensional variational (3dVAR) technique implemented in the operational RUC in May 2003 (Devenyi and Benjamin 2003,

Benjamin et al. 2004a, section 4), but with subsequent improvements listed below.

The key RUC modifications used in these OSE experiments made since the version of the RUC described by Benjamin et al. (2004a, b) include:

- Modification of moisture analysis variable from ln q (natural logarithm of water vapor mixing ratio) to pseudo relative humidity (pseudo-RH), defined as q / q-saturation-background (Dee and da Silva 2003). Assimilation of all integrated precipitable water observations (GPS-PW and GOES) was applied to the RUC 3dVAR using a forward model for vertically integrated pseudo-RH with respect to precipitable water. (Benjamin et al. 2004d). A small modification in moisture background error specification was made between the winter and summer seasons that did not appear to modify observation impact results (Moninger et al. 2009).
- Assimilation of GPS precipitable water data added in 2005 (Smith et al. 2007)
- Fractional application of lowest temperature analysis increment to top two levels in soil-vegetation-snow model used in RUC.
- Assimilation of METAR ceiling and visibility observations modifying the 3-d RUC 3-d hydrometeor (five species) and 3-d water vapor mixing ratio fields (Benjamin et al. 2004e)
- Assimilation of pseudo-residuals for surface observations distributed within the planetary boundary layer (PBL) using the background (RUC 1h

forecast) PBL depth, using certain constraints. (Benjamin et al. 2004f)

- Extension of digital filter initialization (DFI) used in RUC model to a two-pass diabatic DFI.

Changes were also made in RUC model physics using the Thompson mixed-phase cloud microphysics and Grell-Devenyi convective parameterization as described by Benjamin et al. (2004f).

Observational data assimilated in the version of the RUC used in this OSE study are listed in Table 1. GOES-based cloud-top temperature/pressure retrievals, AMVs, and precipitable water are also assimilated in the RUC 1-h cycle.

3. Experiment design for observation impact experiments

A series of experiments was conducted using the Rapid Update Cycle (RUC) model/assimilation system in which various data sources were denied to assess relative importance of the different data types for short-range (3-h to 12-h duration) wind, temperature, and relative humidity forecasts at different vertical levels. This assessment was carried out for 10-day periods in cold season (November-December 2006) and warm season (August 2007).

The same boundary conditions were used in all experiments, damping the signal in differences between experiments, more than might be expected in similar OSEs performed with global assimilation and models. The damping effect by lateral boundary conditions becomes stronger as the model/assimilation domain is reduced, and therefore is larger in this study using the RUC domain than that for the NAM-based (larger regional domain) described by Zapotocny et al. (2002). Nutter et al. (2004) also show a similar effect from lateral boundary conditions limiting the spread of regional ensemble forecasts; the same effect occurs in the OSEs described here limiting variation between experiments more than in global OSEs. In a regional OSSE study for simulated lidar wind observations, Weygandt et al. (2004) found the observational impact from simulated lidar winds interior to the RUC regional domain about equal to that from variations in lateral boundary conditions from associated global OSSE experiments with and without lidar. In this study, the observations considered are generally denser over the US than over oceans and other land areas, but the

actual impact is underestimated in this study because of the common lateral boundary conditions prescribed in these experiments.

a. **Experiments performed**

A control experiment was performed for each of two seasonal 10-day test periods in which all available observations were used, similar to the operational RUC. In subsequent experiments, different observation types were withheld, as shown in Table 2. Most of these observation types were available over the full RUC horizontal domain covering the lower 48 United States and adjacent Canada and Mexico (approximately that shown in Fig. 1). As shown in Fig. 2, some of the observation types (profiler, TAMDAR aircraft) were available only in the midwestern US, motivating us to also employ a verification subregion in that area, as discussed in the next section. In the RUC, GOES AMVs are assimilated only over oceanic areas, since aircraft data (generally of higher quality) are predominant over land area in the RUC domain. Impact experiments for AMVs, 8-km profilers, and 12-km profilers were performed only for the winter period (Table 2).

Lateral boundary conditions were specified from the NCEP NAM (North American Mesoscale) model, initialized every 6 h and available with 3-h output frequency. NAM boundary conditions were specified in the same delayed manner as with the operational RUC: RUC model runs at 00, 06, 12, and 18

UTC use NAM boundary conditions from the previous NAM cycle (18, 00, 06, and 12 UTC, respectively).

The experiments for the winter and summer 10-day data assimilation periods are shown in Table 3. The November-December 2006 winter period was synoptically active in the northern United States, especially in the upper Midwest and Great Lakes area. An example of surface conditions during this period (1200 UTC 1 December 2006) is shown in Fig. 1, with a strong winter storm centered over Indiana. The 10-day summer experiment period spanned 15-25 August 2007, and was chosen because it included considerable intense weather in the Great Lakes region. The period started with a warm front producing heavy precipitation in that region; later, flooding occurred in MN and WI. Severe storms continued to appear, and generally move toward the east, throughout the period.

b. Verification

We verified model forecasts against conventional, twice-daily radiosonde data over the two domains depicted in Fig. 3. The first domain contains all the radiosonde sites located within the RUC domain; the second (the red rectangle) is a limited area over the data-rich Midwest U.S.

Verification results for the national region reflect the impact of observations over the full RUC domain, covering the lower 48 contiguous US and significant

proportions of Canada and Mexico. The Midwest verification region shown in Fig. 3 has special interest because of the NOAA Profiler Network (marked in green) and TAMDAR aircraft coverage at that time (see Moninger et al. 2009 for TAMDAR coverage). With the US considering expenditures for wider deployment of profilers and regional aircraft observations, the Midwest verification domain corresponds to the density that might be expected nationally over the next few to several years.

For each RUC experiment, residuals (forecast minus observed (*f-o*) differences) for temperature (T), relative humidity (RH), and wind (V) were computed at all radiosonde locations located within each verification domain. These *f-o* residuals were calculated for 3-, 6-, 9-, and 12-h forecasts. The rms (root mean square) difference between forecasts and observations was computed for each 12-h radiosonde verification time (0000 and 1200 UTC). This difference is sometimes referred to below as 'forecast error', or 'RMSE', but in fact also contains a contribution from observation error (including representativeness "error" from the inability of a grid to resolve sub-grid variations sometimes evident in observations).

In the following results, increase in forecast error from denying a given observation type can be considered equivalent to the added forecast skill when that observation type is added to other existing observations. Benjamin et al. (2004c) explain this verification procedure.

Verification in this paper uses 10-hPa vertical resolution, including significant-

level radiosonde observations and native model levels interpolated to that resolution, for calculating *f-o* differences using a verification capability explained in Moninger et al. (2009). This high vertical resolution of forecast errors allows clearer attribution of differences to physical mechanisms than verifying against radiosonde data only at mandatory levels (e.g., 850, 700, 500, 400, 300, 250, 200, and 150 hPa). For example, higher vertical resolution in verification revealed a peak near 900 hPa in temperature forecast error and aircraft impact at that level, subsequently related to boundary-layer depth as described by Moninger et al. (2009). The 10-hPa verification also increases the number of *f-o* data points over what would have been available with mandatory level raob data only, numbering about 5200 for a 200-hPa layer in the Midwest domain (10 days x 2 times/day x 13 raob sites x 20 vertical points) and about 32,000 *f-o* points for the national domain (80 raobs), increasing significance of results shown later.

For quality control of radiosonde data used in verification, *f-o* values from the control experiment were subjectively screened for egregiously large values and removed when found. While some erroneous values may have escaped detection, they were used uniformly in verifying all experiments and therefore do not contribute to the *relative* impact results shown below.

We looked for impact on precipitation forecasts in control vs. denial experiments for the two observation types most likely to show them, GPS-PW, and TAMDAR aircraft observations, and found only negligible effect. Assimilation of radar reflectivity data, by contrast, has shown a strong effect on RUC precipitation

forecasts (Weygandt et al 2008).

c. Statistical significance of results

Results that follow present differences in rms forecast error (RMSE) for model runs with and without specific observation types. Each estimate has an associated uncertainty due, in part, to the small number of days we examined. We present *overall* RMSE differences for each period (winter and summer), but we can estimate the uncertainty in these RMSE differences by considering the variations in RMSE over *each of the 20 raob times* in each seasonal period. The uncertainty on the mean ("standard error") is estimated as:

$$\text{Standard Error} = \sigma/\sqrt{(n-1)(1-\varphi)}$$

where σ is the standard deviation, n is the number of RMSE forecast values, x is the set of RMSE forecast differences, and φ Is the lag one autocorrelation derived from the time series x. This is empirically derived from the RMSE values with the approximation:

$$\varphi \cong \mathrm{cor}\left(x_{1\ldots(n-1)}, x_{x\ldots n}\right).$$

The estimate of the standard error on the mean is distinct and separate from the standard deviation from the sample. The standard error is an estimate of how well we understand the underlying, fundamental differences in RMSE between using the additional data and ignoring the additional data. The standard

deviation is an estimate of how far off the mean value any one RMSE forecast may be (Weatherhead et al., 1998). Thus, were our experiment to be repeated in a similar season and for a similar duration, we can say that the mean RMSE difference has a 67% likelihood of being within one standard error of our results, and a 95% likelihood of being within two standard errors of our results.

This approach at least partially accommodates the fact that the pair-wise differences in RMSE are auto-correlated. Physically, this implies that in some situations, the added data have more influence than others. Those situations can last for more than one day, thus the sequential forecast RMSE differences are not independent estimates of the effect of the added data, but represent an over-sampling of the system. The standard error equation above accounts for these. It should be noted that the lag 1 Auto-Regressive assumption in this case refers to a 12-hour lag as the most significant approximation to the autocorrelation. It should also be noted that aggregation of 00Z and 12Z results allows for a larger sample size, but may result in combining different physical causes of differences as well as different statistical properties of the time series. These more fine scale effects are beyond the scope of the research presented in this paper.

In the figures that follow, standard errors are indicated where relevant. Differences of 1 standard error are significant at the 67% confidence level; differences of 2 standard errors are significant at the 95% confidence level.

d. Procedure for 8-km (quarter-scale) vs. full-scale profiler experiments

For these experiments, we extracted 8-km (quarter-scale) and 12-km profiler data from actual 16-km full-scale profiler data by removing data for all vertical gates higher than 8 km (or 12 km) above station elevation. The hypothetical 8-km profilers provide half the vertical coverage of wind observations compared with the full-scale 404 MHz profilers and approximate data from proposed 8-km 449 MHz profilers. With the 16-km full-scale profilers, winds are available from 36 high-mode gates and 36 low-mode gates, with a slight overlap near 8-km elevation above ground level (AGL). We extracted quarter-scale (8-km) and 12-km profiler data at the 30 profiler sites shown in Fig. 3. In removing winds above 8 km AGL, data were left for 33 low-mode gates, and 5 high-mode gates. We assumed that the observation error in these hypothetical future profilers would be the same as that used for the existing 16-km profilers.

The profiler stations assimilated by the RUC in these experiments included about 21 CAP (Cooperative Agency Profiler) sites operating at 915-MHz with a vertical range of about 4 km, and these data were not truncated in 8-km and 12-km profiler experiments. These CAP profilers include 12 in CA, 4 in TX, and 1 each in NM, AZ, MN, NJ, and Nova Scotia), as shown by White et al. (2007, Fig. 3).

4. Results for impact from existing observations

a. Stratification

To summarize the complexity of the OSE results from this study, we considered the five verification stratifications: experiment, regions, layers, seasons, and

forecast duration, as shown in Table 4. Rather than examine detailed vertical profiles of forecast errors (e.g., Benjamin et al. 2004c, Moninger et al 2009), we found it effective to break down the full 1000- to 100-hPa vertical domain into three layers: 1000-800 hPa (dominated by boundary-layer and surface effects), 800-400 hPa (middle troposphere), and 400-100 hPa (upper troposphere to lower stratosphere including tropopause and upper-level jet maxima).

We developed a composite graphical format used throughout the rest of this section that is introduced in Fig. 4 to summarize results for all OSE experiments for a given domain (national or Midwest) and vertical layer (1000-100, 1000-800, 800-400, or 400-100 hPa). Due to known raob moisture sensor limitations above approximately 400 hPa, where temperatures are commonly below -30 °C, we used 1000-400 hPa for the "full troposphere" results for RH.

b. National, full troposphere

We begin with the broadest view by examining results on the NATIONAL domain for vertically integrated layers: 1000-100 hPa for temperature and wind, and 1000-400 hPa for RH. In section 4.c, we shall show stratifications over different vertical layers, and in section 4.d, results specifically from the Midwest region where observations are dense.

In the first graphical composite (Fig. 4), we consider impact results for RH for 1000-400 hPa. Results are for differences between experiments in which various

observation types were withheld, as well as the control experiment in which all observations were assimilated (similar to the operational RUC). We use different colors to depict results for each of the eight observation denial experiments. Results in the top graphs are for *winter*; those in the bottom two boxes are for *summer*. For graphs at left, the three adjacent columns for each OSE are for 3-h, 6-h, and 12-h forecasts, respectively. The graphs on the right show the same information as those on the left, but organized by forecast projection to allow easier interpretation from that perspective. Again, as stated in the last section, increase in forecast error from denying a given observation type can be considered equivalent to the added forecast skill ("forecast impact") when that observation type is added to other existing observations.

The black bars indicate +/- one standard error (sec. 3c) from the forecast impact of each observation type. Differences of one standard error are significant at the 67% confidence level; differences of two standard errors are significant at the 95% confidence level.

For RH over the 1000-400 hPa layer for 3-h to 12-h RUC forecasts over the full domain (Fig. 4), the observation type with the largest impact is clearly raobs, for which the impact is 1-2% RH for all forecast durations (3h, 6h, and 12h) in both summer and winter. GPS-PW (Smith et al. 2007) had the second largest impact especially in winter (0.6-0.9% for 3-h, 6-h forecasts). The third most important observation source is aircraft in winter (<0.5%, presumably primarily from TAMDAR reports (including moisture) in the Midwest) and surface observations

in summer. In the summer period, each of the five observation types that provide moisture observations are shown to have varying degrees of at least small positive impacts on the short-range RUC RH forecasts over the full national domain.

The impact of raobs at 12h on RH forecasts is large, sometimes even larger than that at 3h and 6h. This is so because 12-h forecasts valid at 0000 and 1200 UTC have the direct benefit of raob data in the initial conditions whereas 3-h and 6-h forecasts do not. Even so, the impact at 3h and 6h is substantial. We attribute this to the "memory" in the assimilation system of raobs incorporated several cycles before the start of these forecasts.

For temperature forecasts over the full (1000-100 hPa) atmospheric depth (Fig. 5), in winter, on average, raobs and aircraft observations had about equal effect (0.05-0.15K) on average over 3-h to 6-h impact, more from aircraft at 3h, equal at 6h, and much more impact from radiosondes at 12h. In summer, surface observations have nearly equal impact as both aircraft and raobs over the full 1000-100 hPa. The explanation is that a deeper mixed layer in summer extends the potential vertical influence of surface observations. This PBL effect is accounted for in the RUC 3dvar design, as discussed in section 2 and in Benjamin et al. 2004f.

For the vector wind difference (Fig. 6), aircraft observations have the strongest overall impact for 3-h and 6-h forecast projections for both for the summer (0.3-0.6 m/s) and winter periods (0.15-0.20 m/s) and for 12-h forecasts in summer

season. Radiosondes have the greatest impact for winds at 12 h in winter only. Satellite AMVs provide a small positive impact (<0.05 m/s) at 12 h in winter, in third place after radiosondes and aircraft. All observation types tested showed at least a small positive impact except for VAD winds in summer at 12-h duration (perhaps due to bird migration problems not detected by the RUC bird detection algorithm (Benjamin et al. 2004a, section 4.e) and mesonet observations, which frequently have unrepresentative siting for wind measurement (Benjamin et al. 2007).

Our results indicate that aircraft observation impact (Fig. 6) was stronger in summer (0.3-0.6 m/s over full layer) than winter ((0.15-0.20 m/s), surprising to us since upper-level wind forecast errors are usually larger in winter than in summer over the US. To examine this behavior a bit further, we first looked at seasonal variations of upper-level wind (400-200 hPa) forecast error for the RUC at 9-h, 3-h, and 1-h forecast duration (Fig. 7, 30-day running mean) over a period from January 2007 to May 2009. For RUC 9-h wind forecasts for the 400-200 hPa layer, error (vs. raobs) was about 5.8-6.0 m/s in winter for 2007-2009 and lower, about 5.2-5.3 m/s on average, in summer for 2007-2008. However, the short-range increment in forecast skill (e.g., 9-h to1-h forecast skill difference, bottom in Fig. 7), largely from assimilation of recent observations (Benjamin et al. 2004a), does not vary drastically over season, although the 30-day running mean show some apparent shorter-period regime-dependent variations. Therefore, we consider the larger aircraft impact for wind forecasts in the summer August 2007 period than in winter to be slightly unusual but plausible, consistent with the

particular synoptic-scale regimes of those separate 10-day periods.

c. National, but stratified into three layers – for wind only.

Next, we stratify the OSE results within three layers as described in section 4a, 1000-800 hPa (near-surface), 800-400 hPa (middle troposphere), and 400-100 hPa. We start with the lower tropospheric (1000-800 hPa) layer for wind forecasts (Fig. 8).

For the lower tropospheric 1000-800 hPa layer (Fig. 8), aircraft, VAD, and surface observations have about equal impact for 3-h wind forecasts in winter, when the PBL is typically shallow and inversions are common. In summer, surface observations have the most impact. We attribute this to deeper PBL mixing and the addition of PBL-depth pseudo-residuals for surface observations in the RUC 3-d assimilation (discussed in section 2 and Benjamin et al. 2004f). Mesonet observations were found to add little or no impact to 3-h or 6-h lower-tropospheric wind forecasts to other surface observations (primarily METARs) even in summer when stronger effects are shown from surface observations, but have a very small positive effect at 12h in summer and winter. The increasing impact of aircraft observations with forecast projection in summer may result from better midtropospheric winds (next section) that are mixed down over time in the typically deeper PBL.

For midtroposphere winds over the national verification domain (Fig. 9), aircraft observations had the strongest impact overall, especially in summer (0.25-0.40 m/s), followed by raobs. Raobs had the strongest impact for 12-h forecasts in

winter for midtropospheric winds. Profilers, VAD winds, and AMVs all have a small positive effect for midtropospheric winds. [WRM1]The slight positive impact from GOES AMVs shows a slight increase with forecast projection as its offshore effect (assimilated only over water) propagates inland.

Fig. 10 indicates that aircraft have a pronounced impact on upper-level wind forecast accuracy in the RUC domain for all forecast projections and both seasons, consistently larger than that for any other observation type, ~0.3 m/s in this particular winter period and 0.8-0.9 m/s in the summer period for 3-h and 6-h forecasts. Hourly automated aircraft reports over the U.S. were the original primary justification for the development and implementation of a rapidly updated data assimilation cycle to improve short-range upper-level wind forecasts (Benjamin et al. 1991). The results depicted in Fig. 10 are still consistent with that justification, despite the addition of many other observation types since 1991. Raob data had the second largest impact on upper-level wind forecasts over the national verification domain, with profiler also making a very small positive impact over this larger domain. As with the 800-400-hPa layer, AMVs ("cloud-drift" winds) had a small but positive effect on upper-level winds, larger at 12 h than at 3 h, again a consequence of their assimilation in RUC only over oceanic regions.

d. Midwest (very data rich area)

The Midwest region in the US has exceptional upper-air observational coverage,

denser than any other area in the US because of the region's proximity to the NOAA Profiler Network, and the initial deployment of TAMDAR sensors on regional aircraft in this area (see Moninger et al. 2009). Therefore, we considered it useful to examine the relative impact of different observation types for short-range RUC forecasts specifically in this region.

4.d.1. Relative humidity impact in Midwest region

In the Midwest verification region, we start again with the overall observational impact on relative humidity forecasts starting with the 1000-400 hPa layer (Fig. 11). Here, radiosondes still show the largest impact in winter (1-2% RH), but with nearly equal impact from aircraft observations in summer (all forecast projections) and in winter 3-h forecasts. The availability of aircraft-based moisture observations from TAMDAR clearly contributed strongly in this region, comparing Fig. 11 with corresponding RH impact for the national domain (Fig. 4) showing much less aircraft impact. OSE results (control – noTAMDAR) in Moninger et al. (2009) confirm its very large impact, averaging about 2% for the 1000-400 hPa layer in both the Nov-Dec 2006 winter period and from fall 2008 onward. RH impact from GPS-PW observations followed closely that from aircraft data in both summer and winter test periods. Even profilers made a positive contribution to RH forecasts (0.2-0.5%) although they do not measure moisture, presumably due to improved vertical motion and horizontal transport fields.

4.d.2. Temperature forecast impact over Midwest domain

For temperature impact over the full depth (1000-100 hPa) in the Midwest region (Fig. 12), results were similar to those shown in Fig. 5 for the national domain, with aircraft showing the greatest impact for 3-h and 6-h temperature forecasts in both winter and summer. However, the average impact from assimilation of aircraft observations in winter in 3-h temperature forecasts was significantly stronger in the Midwest (0.25 K) than over the full national domain (0.15 K), likely due to the higher density of aircraft data in this region.

Regarding temperature forecasts in the lower troposphere (1000-800 hPa, Fig. 13), aircraft reports have the strongest impact (0.3-0.56 K) by far in winter for 3- to 6-h forecasts. In summer, the aircraft data have a slightly but significantly larger impact (0.12-0.20 K) than surface observations. The extra spatial density provided by TAMDAR aircraft observations, which include data from frequent ascents and descents into regional airports (Moninger et al. 2009), contributes to the forecast impact for lower-tropospheric temperatures, and especially during periods of wintertime lower-tropospheric temperature inversions.

Aircraft observations also exhibit the largest impact in the middle troposphere for both winter and summer seasons (Fig. 14), although the impact for all observation types is quite low in summer (<0.1 K for all types, all forecast projections), presumably due to a relatively low thermal stability and a general absence of midlevel frontal zones.

4.d.3 Wind forecast impact over Midwest domain

The observation impact results for winds integrated over the full 1000-100 hPa layer within the data-rich Midwest domain is shown in Fig. 15, indicating a nearly equal impact for aircraft and profiler observations in winter at 3 h and 6 h (0.13-0.20 m/s). In summer, aircraft observations had the greatest overall impact at 3h (0.35 m/s) and 6h (0.25 m/s), followed by profiler, surface, and raobs, all with about the same effect. Note that surface observations have such a large effect on the fully integrated 1000-100 hPa layer, again indicating their representativeness in the deeper summertime boundary layer and the effectiveness of the RUC PBL-based pseudo-residual assimilation technique. The addition of mesonet observations, by contrast, had a slight negative effect on 1000-100 hPa 3-h wind forecasts in both winter and summer, again presumably due to widespread siting issues.

For lower-tropospheric (1000-800 hPa) wind forecasts in the Midwest domain (Fig. 16), it is not surprising that surface observations had the largest positive effect on both 3-h and 6-h forecasts in both winter and summer periods. After surface observations, the largest effects in lower-tropospheric 3h wind forecasts were from aircraft in winter, and from profiler in summer. As with the national domain (Fig. 9), aircraft data have very little effect in summer near-surface (1000-800 hPa) winds for 3-h and 6-h duration, but have the largest effect by 12h. VAD wind observations had the third-largest impact at 3h in summer and for 3-6h in winter.

The lowest gate in NOAA profilers is 500 m AGL, and sites within the Midwest verification domain are at 170-300 m elevation, limiting the profiler impact below 800 hPa, and perhaps contributing to negligible impact in winter in this layer. Over a different regional domain centered directly on the NOAA Profiler Network but using a similar RUC observation impact strategy for a profiler-only OSE with a 14-day test period in February 2001, Benjamin et al. (2004c) showed a larger 0.3 m/s impact from assimilation of profiler winds for 3-h wind forecasts at 850 hPa and a 0.1 m/s impact over a larger eastern US verification domain. The smaller impact in this study is attributed to a shift in the Midwest domain not centered on the NPN (Fig. 3) and averaging over the 1000-800 hPa layer, essentially down to the surface.

For middle-tropospheric (800-400 hPa, Fig. 17) wind forecasts in the data-rich Midwest verification domain, aircraft, followed by profiler observations had the greatest impact. Clearly, these two observation types in the Midwest are not redundant, but together produce a larger reduction in forecast error. For upper-level winds (400-100 hPa, Fig. 18), profiler observations had the largest positive impact (reduction in forecast error) in winter at 3h and 6h, followed by aircraft observations in winter. In the summer period, the opposite was true, with aircraft showing the greatest effect (e.g., > 0.5 m/s at 3 h), followed by radiosondes and profilers at 3h. At 6h, aircraft had the most impact, with profiler and raobs second. Profiler impact for the NPN-centered verification domain shown by Benjamin et al. (2004c) from the February 2001 period was approximately 0.5 m/s for the 800-400 hPa layer for 3-h forecasts and about 0.1 m/s for 12h

forecasts. We are unable to explain the small negative impact at 12h from aircraft in the Midwest layer for the 400-100 hPa layer evident in both winter and summer periods, except that it may be a sampling anomaly. The dropoff of profiler impact with projection time is attributed to propagation of that impact quickly outside of the limited extent of the NOAA Profiler Network (Fig. 3). A similar dropoff with time was shown in Benjamin et al. (2004c) for profiler impact.

5. Results from profiler height experiments – impact from vertically truncated profiler heights

We added two additional experiments (Table 2), involving hypothetical 8-km and 12-km profilers, referring to the vertical reach of the profiler antenna. The additional experiments give us a quantitative measure of the impacts that potential reduction in the vertical reach of existing NOAA network profilers will have on forecast accuracy. The operating frequency of network profilers must soon be changed from 404 to 449 MHz. Larger (and more expensive) antennas are required to reach 16 km than to reach lower altitudes. The experiment also relates to the cost of a possible expansion of the current network from the mid-US to the entire lower 48 states

The 8-km profilers have half the vertical reach of the full-scale (16-km) network profilers. The former are often called “quarter-scale profilers” because their antennas occupy only one quarter of the area. To manufacture data from quarter-scale wind profilers, we merely extracted actual data (surface to 8 km AGL) from the full-scale profilers.

The results for these experiments are depicted in Figs. 19-21 for both 3-h and 6-h forecast projections in the Midwest region for the 10-day winter period (Table 2). The vector wind error difference, No-Profiler minus Control (N-C, blue line in figures), shows the impact of profiler data themselves (equivalent to degradation if profiler data are missing). When profiler data are denied for the 10-day test period, 3-h forecasts of winds aloft (Fig. 19) from 600-300 hPa are worse by ~0.4 m/s in the Midwest region and by 0.2-0.3 m/s for the same layer for 6-h forecasts (Fig. 20). The greatest improvement from the inclusion of profiler winds for 3-h wind forecasts was ~0.55 m/s at 350 hPa (Fig. 19). These results are similar to those shown by Benjamin et al. (2004c) in a previous profiler impact study for the profiler (Midwest) domain and downscale domains, respectively.

The vector wind error difference, Quarter-scale minus Control (Q-C, red line in Figs. 19,20), shows the value of using 16-km full-scale profilers versus 8-km quarter-scale profilers. The Q-C difference is near zero at most altitudes, indicating that most of the value added to tropospheric wind forecasts from full-scale profilers is also added by quarter-scale profilers up to the jet levels where aircraft data are plentiful. Only at 200 hPa and above do full-scale profilers show value added (more accurate 3-h forecasts) that are not available with 8-km profilers. Quarter-scale profilers actually delivered somewhat better 3-h wind forecasts in the 900-750 hPa layer than full-scale profilers, possibly resulting from less lower-tropospheric geostrophic wind adjustment without the stratospheric (200 hPa and above) wind observations available only with 16-km profilers.

The final experiment was performed to simulate the inclusion of 12-km profilers, extracted from the actual 16-km profiler data. The results for 12-km profilers in Fig. 21 are very similar to those for the 8-km (quarter-scale) profilers up to the 250-hPa level. However, as might be expected, the 12-km profilers do add forecast skill improvement over 8-km profilers for 3-h wind forecasts for the 150- to 200-hPa layer. The 12-km profilers do not capture the extra improvement in the 50- to 100-hPa layer available from the full 16-km profilers.

6. Conclusions

We performed extensive observation system experiments (OSE) involving data denial for two 10-day periods, one in winter and one in summer, using the hourly-updating Rapid Update Cycle model/assimilation system. We examined forecast impact for relative humidity, temperature and wind at 3, 6, 9, and 12h.

We conclude from these experiments that the heterogeneous atmospheric observing system in US is effective for short-range (3-12h) 1000-100 hPa forecasts for all three variables studied: relative humidity, temperature, and wind.

Overall, aircraft data were found to have the most impact on reducing error in short-range forecasts over the US from the lower stratosphere down to the surface, but they are strongly and necessarily augmented by other observing systems. As shown by Moninger et al. (2009) in a companion article, TAMDAR aircraft observations (also including moisture) clearly improved forecast accuracy in the Midwest and eastern US area when added to all other observations in a

complementary experiment to those shown in this paper.

Radiosonde observations were second in importance overall, within the parameters defining this OSE for 3-6h, and arguably most important for 12-h forecast impact on the national scale followed closely by aircraft. GPS-PW, surface, profiler, and VAD all provided value added to forecast accuracy, in roughly that order. GPS-PW was similar to raob contributions for short-range RH forecasts. Given that surface observations showed a significant additional value to lower tropospheric forecasts, especially for the 1000-800 hPa layer and in summer time, we conclude that the RUC assimilation and use of PBL depth for pseudo-residuals is effective for 3-d assimilation of these surface variables. The impact of profiler wind data was notably higher in the Midwest domain, where the NOAA Network is located, than in the national domain, where their effect is heavily diluted. The relatively small impact from AMVs (used only over ocean areas) is attributable to the relatively small extent of the RUC domain over oceanic areas, limiting the possible AMV-related effect. Generally, the relative impact for profiler, aircraft, and raobs in this experiment was similar to that shown by Schwartz and Benjamin (2004) for an OSE using an earlier version of the RUC for a February 2001 test period. Middle troposphere wind forecast impact from profiler data in that earlier study was larger with a verification domain centered directly over the NOAA Profiler Network than in this study for the Midwest verification domain, shifted from the NPN area.

Experiments using hypothetical vertically-truncated profiler data (with a vertical reach of 8-km and 12-km AGL) were performed for the winter (November-December 2006) period. These experiments showed that 8-km (quarter-scale) profilers provide 3-h and 6-h wind forecast improvement about equal to that from full-scale (16-km) profilers from the surface up to 250 hPa, suggesting that 8-km profilers would complement aircraft data for short-range tropospheric forecasts.

We note once again that the magnitude of forecast impacts from different observation denial experiments is damped by the same lateral boundary conditions used in all experiments for the regional RUC domain. The 10-day periods used in this study for winter and summer seasons are barely long enough for robust results, but were limited by the logistics for the 1-h update cycle environment (unique to this study) and available computing and storage resources. These limitations were partially mitigated by performing verification every 10-hPa using full significant level radiosonde data, adding considerable data points. Standard error calculations for each experiment indicate that, even for our relatively short 10-day summer and winter periods, results are statistically robust, with many forecast impacts being significant at more than the 95% confidence level.

The observation impact results in this study often showed a decrease with forecast projection (example in Fig. 15). This was evident, in general, for aircraft and profiler data, for which this effect was enhanced by regional concentrations of those observations (aircraft enhanced by TAMDAR in Midwest area, profiler

primarily in NOAA Profiler Network). The impact of raob data was a prominent exception, showing an apparent increase with time, a statistical quirk from our verification only at 0000 and 1200 UTC for forecasts valid at those times, initialized at 0900 and 2100 for 3-h forecasts, for instance. Of course, in general, raobs are available only every 12 h, so their impact on the analyses that create the less-than-12-h forecasts valid at 0000 and 1200 UTC is indirect (only through the hourly-cycled background field) and degrades as the analysis time moves away from 0000 and 1200 UTC. Also, the overall impact of high-frequency observations is somewhat larger at analysis times when not competing with raobs.

Conducting OSEs can sometimes reveal flaws in the assimilation system from forward models or observation-error specification. In this study, relatively consistent positive (sometimes very small) or near-zero impacts were shown for nearly all observation types, presumably indicating no major flaws in treatment in the RUC for any observation types. But in initial experiments performed for this OSE, some counter-intuitive results arose, leading to detection of assimilation design flaws for aircraft moisture observation error, moisture assimilation design, and too-small observation errors specified for radiosonde RH and wind observations. The results shown in this paper are dependent, for better or for worse, on the design of the RUC 3-d variational analysis and modeling system as described in section 2, and we cannot rule out remaining design flaws or outright errors.

This OSE study included vertically stratified results for the data-rich Midwest verification domain. Even here, nearly all observation types contributed positive impact, with clear, positive, and complementary effects from profiler and aircraft data, indicating that this region is not oversampled by observations. A strong positive effect from surface observations over surprisingly deep layers was shown, especially but not solely in summer, for temperature, wind, and RH, but very little positive impact was shown when mesonet observations were added to METAR observations.

We intend to add new observation impact experiments using high-frequency assimilation of radar reflectivity (Weygandt et al. 2008), added to the operational RUC at NCEP in November 2008, and using hydrometeor assimilation from GOES and METAR cloud/visibility data (Benjamin et al. 2004e). We also intend to identify diurnal variations in observation impact (1200 vs. 0000 UTC) and repeat similar OSEs with the upcoming Rapid Refresh replacing the RUC hourly assimilation/model cycle.

7. Acknowledgments

We thank Betsy Weatherhead of NOAA/ESRL/GSD for her thoughtful and effective analysis of the statistical significance of our results. We also thank Annie Reiser and Dezso Devenyi of NOAA/ESRL/GSD for their excellent reviews of this manuscript.

8. References

Baker, N.L., and R. Daley 2000: Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Q. J. Roy. Meteor. Soc.*, **126**, 1431-1454.

Benjamin, S.G., K.A. Brewster, R.L. Brummer, B.F. Jewett, T.W. Schlatter, T.L.

Smith, and P.A. Stamus, 1991: An isentropic three-hourly data assimilation system using ACARS aircraft observations. *Mon. Wea. Rev.*, **119**, 888-906.

Benjamin, S.G., D. Devenyi, S.S. Weygandt, K.J. Brundage, J.M. Brown, G.A. Grell, D. Kim, B.E. Schwartz, T.G. Smirnova, T.L. Smith, and G.S. Manikin, 2004a: An hourly assimilation/forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495-518.

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

Benjamin, S.G., 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.

Benjamin, S.G., T.G. Smirnova, K. Brundage, S.S. Weygandt, T.L. Smith, B. Schwartz, D. Devenyi, J.M. Brown, G.A. Grell, 2004d: A 13-km RUC and beyond: Recent developments and future plans. Preprints, 11th Conf. Aviation, Range, Aerospace Meteor., AMS, Hyannis, MA, October.

Benjamin, S.G., J.M. Brown, S.S. Weygandt, T.L. Smith, B. Schwartz, W.R. Moninger, 2004e: Assimilation of surface cloud, visibility, and current weather observations in the RUC. Preprints 16th Conf. Num. Wea. Pred., Seattle.

Benjamin, S.G., S. Weygandt, D. Devenyi, J.M. Brown, G. Manikin, T.L. Smith, and T. Smirnova, 2004f: Improved moisture and PBL initialization in the RUC using METAR data. Preprints 11th Conf. on Aviation, Range, and Aerospace Meteorology, Hyannis, MA, Amer. Meteor. Soc., CD-ROM, 17.3.

Benjamin, S.G., W.R. Moninger, S.R. Sahm, T.L. Smith. 2007: Mesonet wind quality monitoring allowing assimilation in the RUC and other NCEP models. Preprints, 22nd Conf. Wea. Analysis Forecasting / 18th Conf. Num. Wea. Pred., Park City, UT, Amer. Meteor. Soc.

Bouttier, F., and G. Kelly 2001. Observing-system experiments in the ECMWF 4D-Var data assimilation system. *Q. J. R. Meteor. Soc.*, **127**, 1469-1488.

Cardinali, C., 2009: Monitoring the observation impact on the short-range forecast. *Q. J. Roy. Meteor. Soc.*, **135,** 239-258.

Dee, D.P., and A.M. da Silva, 2003: The choice of variable for atmospheric moisture analysis, *Mon. Wea. Rev.*, **131**, 155-171.

Devenyi, D., and S.G. Benjamin, 2003: A variational assimilation technique in a hybrid isentropic-sigma coordinate. *Meteor. Atmos. Phys.,* **82,** 245-257.

Graham, R.J., S.R. Anderson, and M.J. Bader, 2000. The relative utility of current observations systems to global-scale NWP forecasts. *Q. J. Roy. Meteor. Soc.*, **126**, 2435-2460.

Lord, S.J., T. Zapotocny, and J. Jung 2004: Observing system experiments with NCEP's Global Forecast System. Proceedings 3rd WMO Workshop on Impact of Various Observing Systems on Numerical Weather Prediction. Alpbach, Austria, WMO/TD No. 1228, 56-62.

Moninger, W.R., S.G. Benjamin, B.D. Jamison, T.W. Schlatter, T.L. Smith, and E.J. Szoke, 2009: Evaluation of TAMDAR: A new atmospheric sensing system. *Mon. Wea. Rev.*, submitted to *Mon. Wea. Rev.*.

Nutter, P., D. Stensrud, and M. Xue, 2004: Effects of coarsely resolved and temporally interpolated lateral boundary conditions on the dispersion of limited-area ensemble forecasts. *Mon. Wea. Rev.*, **132**, 2358-2377.

Schwartz, B., S. Benjamin, 2004: Observation sensitivity experiments using the Rapid Update Cycle. Preprints 16th Conf. Num. Wea. Pred., AMS, Seattle. [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.

Weatherhead, E. C., G. C. Reinsel, G. C. Tiao, X. Meng, D. Choi, W. Cheang, T. Keller, J. DeLuisi, D. J. Wuebbles, J. B. Kerr, A. J. Miller, S. J. Oltmans, and J. E. Frederick, 1998: Factors affecting the detection of trends: Statistical considerations and applications to environmental data. Journal of Geophysical Research, 103, 17149-17161.

Weygandt, S.S., T.W. Schlatter, S. E. Koch, S.G. Benjamin, A. Marroquin, J. R. Smart, M. Hardesty, B. Rye, A. Belmonte, G. Feingold, D. M. Barker, Q. Zhang, and D. Devenyi, 2004: Potential forecast impacts from space-based lidar winds: results from a regional observing system simulation experiments. *Preprints 8th Symp. Integ. Obs. Systems.*, Seattle.

Weygandt, S. S., S. G. Benjamin, T. G. Smirnova, J. M. Brown, 2008: Assimilation of radar reflectivity data using a diabatic digital filter within the Rapid Update Cycle, 12th Conference on IOAS-AOLS, Amer. Meteor. Soc., New Orleans, LA

White, A.B., F.M. Ralph, J.R. Jordan, C.W. King, D.J. Gottas, P.J. Neiman, L. Bianco, and D.E. White, 2007: Expanding the NOAA Profiler Network: Technology evaluation and new applications for the coastal environment. Proc. 7th Conf. Coastal Atmos. and Oceanic Pred. and Processes, 10-13 Sept, San

Diego, AMS.

Zapotocny, T.H., W.P. Menzel, J.P. Nelson, and J.A. Jung, 2002: An impact study of five remotely sensed and five in situ data types in the Eta Data Assimilation System. *Wea. Forecasting*. **17**, 263-285.

Zapotocny T., J. Jung, J. Le Marshall, and R. Treadon, 2007: A two-season impact study of satellite and in situ data in the NCEP Global Data Assimilation System. Wea. Forecasting, 22, 887–909.

Zhu, Y., and R. Gelaro, 2008: Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interpolation (GSI) analysis system. *Mon. Wea. Rev.*, **136**, 335-351.

Figure Captions

Fig. 1.

Surface analyses for a) 1200 UTC 1 December 2006, in middle of winter experiment period, and b) 1200 UTC 20 August 2007, in the middle of the summer experiment period. (Courtesy of NOAA/NCEP/HPC).

Fig. 2.

Distribution of data sources. Sample data from 8 Sep 2009 (data available in retrospective periods is similar). Color coded as in the histogram plots below. a) AMDAR, including TAMDAR (only the TAMDAR fleets available during the retrospective period are shown); b) Profilers (National network plus cooperating agency profilers); c) VAD, from NEXRAD radars; d) raobs; e) GPS; f) AMV; g) All surface (METARS plus mesonet); h) METAR, color coded by altitude.

Fig. 3.

Midwest/Great Lakes (red rectangle) and National (gray area) verification regions. Also shown is the location of the NOAA profilers (blue rectangles) and verifying raobs (brown triangles).

Fig. 4.

*Differences in rms error (vs. radiosonde) between observation denial experiments listed in Table 2 and control run for **1000-400 hPa** for **relative humidity (in %RH)** for **national domain.** Results for each of the eight observational denial experiments are coded with a different color (aircraft – red, profiler – blue, VAD – pink, RAOB – tan, surface – light blue, GPS-PW – green, mesonet – black, satellite AMV winds – brown). Graphs at top are for winter results; those at bottom are for summer. For graphics at left, the three adjacent columns for each OSE are for 3-h, 6-h, and 12-h forecasts. For graphs at right, the same results are organized by observation type for each forecast projection (3h, 6h, and 12h). The impact of AMV data was not tested during the summer period. Statistical uncertainties are indicated for each observation denial experiment by narrow black lines showing +/- 1 standard error from the mean impact.*

Fig. 5.

*Same as Fig. 4 but for **temperature error (units – K)**, for **1000-100 hPa** layer over **national** domain.*

Fig. 6.

*Same as Fig. 4 but for **wind vector difference (units – ms^{-1})**, still for 1000-100 hPa layer over national domain.*

Fig. 7.

Vector wind error (vs. radiosondes) for RUC 9h (black), 3h (blue), and 1h (red) forecasts averaged over 400-200 hPa layer and for a 30-day running mean for January 2007 – April 2009. Also shown are the forecast error increments (between 0 and 1) from assimilation of recent observations for 9-h to 1-h (red, near lower axis) and 3-h to 1-h (tan) pairs.

Fig. 8. *Same as Fig. 6, but now for **1000-800 hPa** only (still wind, national domain)*

Fig. 9.

*Same as Fig. 8, but for **800-400 hPa** layer, still **winds** for **national domain***

Fig. 10.

*Same as Fig 8 but for **400-100 hPa** layer, still **winds** for **national domain***

Fig. 11.

*Same as Fig.4 (RH, 1000-400 hPa), but now for **Midwest (or, Great Lakes)** regional verification domain.*

Fig. 12.

*Same as Fig. 5 (**temperature**, full depth, **1000-100 hPa**) but for **Midwest** verification region.*

Fig. 13.

*Same as Fig. 12 (temperature, Midwest domain), but for **1000-800 hPa layer** only*

Fig. 14.

*Same as Fig. 12 (temperature, Midwest domain), but for **800-400 hPa layer**.*

Fig. 15.

*Same as Fig. 6 (wind, 1000-100 hPa layer) but for **Midwest** domain.*

Fig. 16.

*Same as Fig. 15 **(wind, Midwest** domain) but for **1000-800 hPa** layer.*

Fig. 17.

*Same as Fig. 15 (**wind, Midwest** domain) but for **800-400 hPa** layer.*

Fig. 18.

*Same as Fig 15 (**Wind, Midwest** domain), but for **400-100 hPa** layer*.

Fig. 19. *Differences in 3-h vector wind errors between No-profiler-minus-Control (N-C, blue line) and 8-km (quarter-scale) minus Control (8-C, red line) over the Midwest region.*

Fig. 20. *Same as Fig. 19 but for **6-h** forecasts instead of 3-h forecasts.*

Fig. 21. *Similar to Fig. 20 for 3-h wind forecast impact in Midwest domain, but with* ***results added for 12-km profilers****. Green line (12-C) is for difference between experiments with 12-km profilers (12) vs. Control (C).*

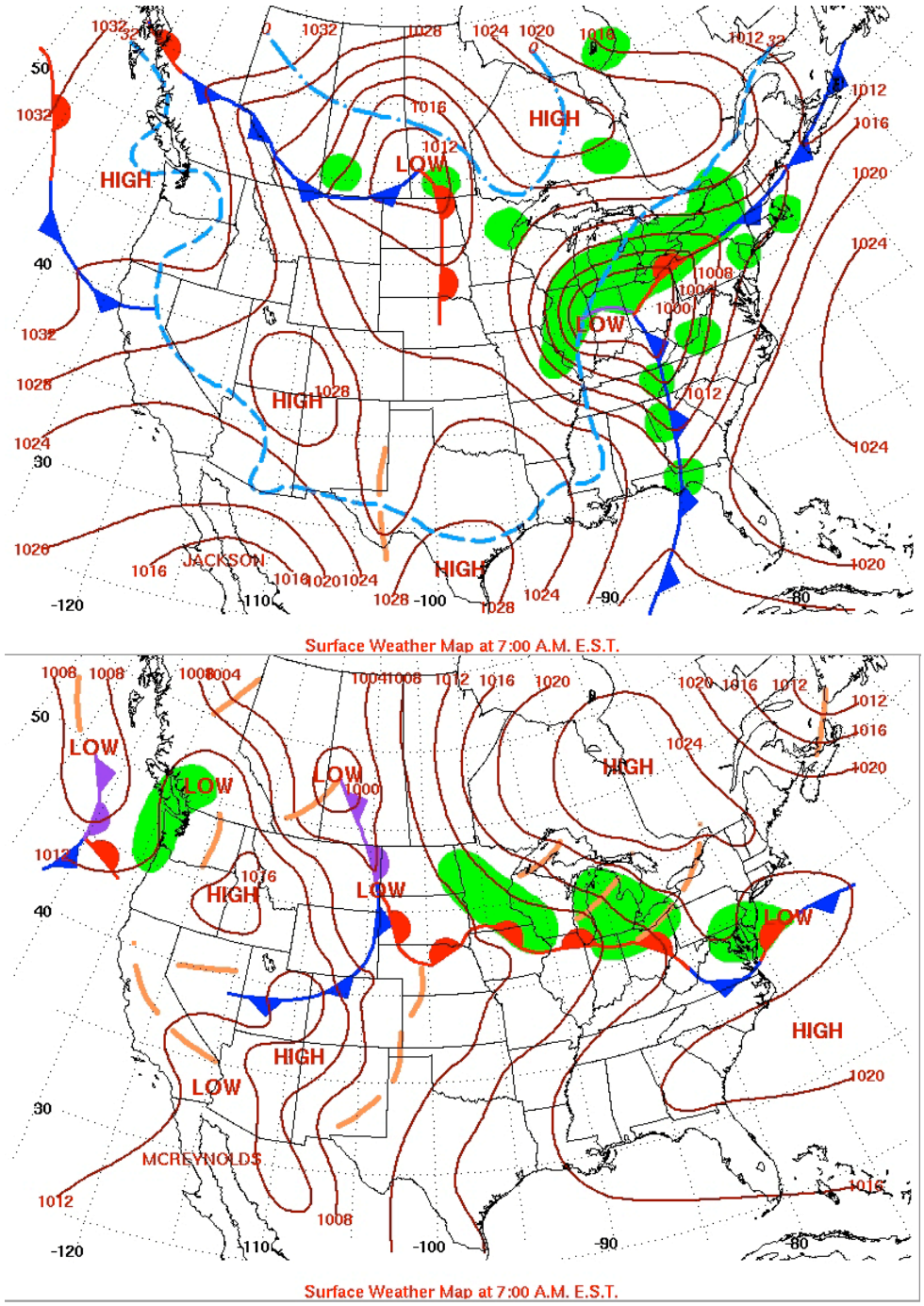

Fig. 1.

Surface analyses for a) 1200 UTC 1 December 2006, in middle of winter experiment period, and b) 1200 UTC 20 August 2007, in the middle of the summer experiment period. (Courtesy of NOAA/NCEP/HPC).

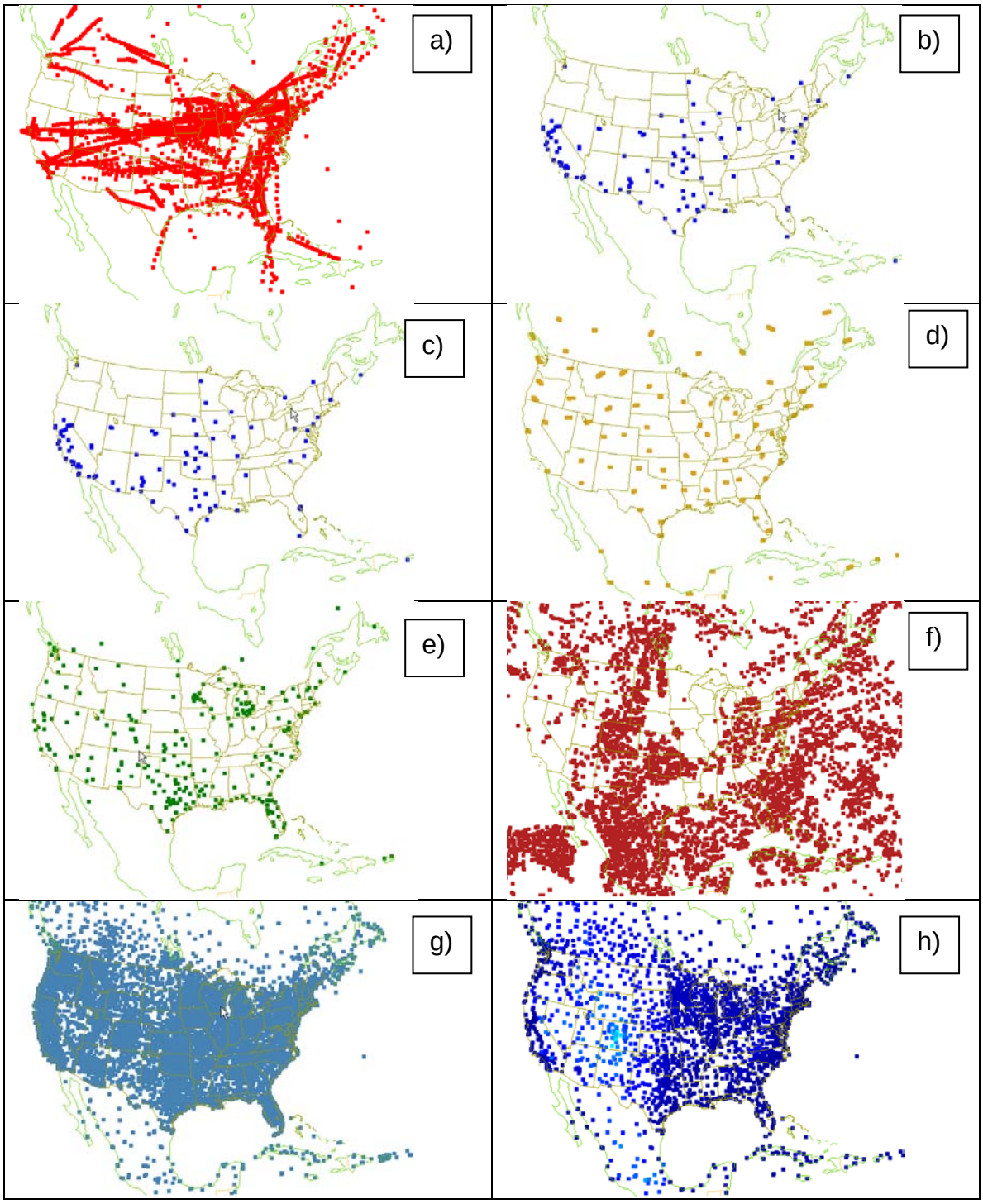

Fig. 2. Distribution of data sources. Sample data from 8 Sep 2009 (data available in retrospective periods is similar). Color coded as in the histogram plots below. a) AMDAR, including TAMDAR (only the TAMDAR fleets available during the retrospective period are shown); b) Profilers (National network plus cooperating agency profilers); c) VAD, from NEXRAD radars; d) raobs; e) GPS; f) AMV; g) All surface (METARS plus mesonet); h) METAR, color coded by altitude.

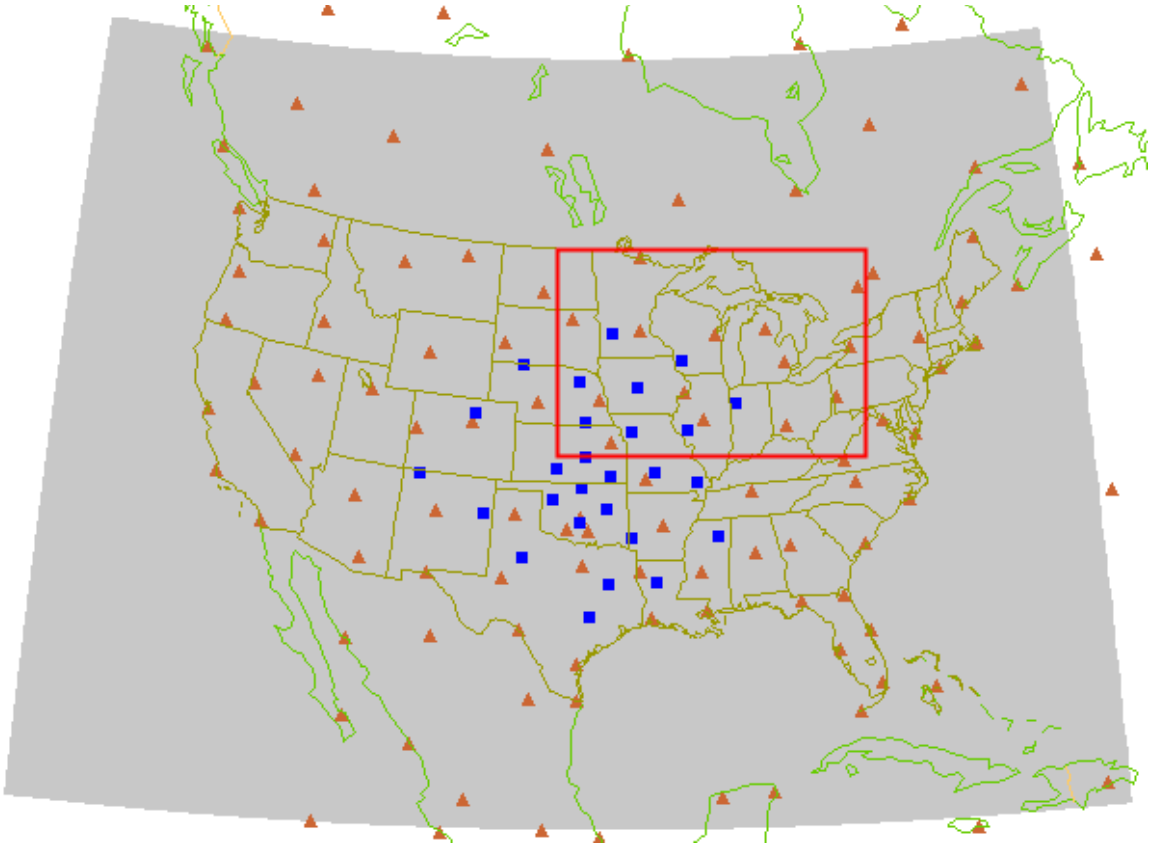

Fig. 3.

Midwest/Great Lakes (red rectangle) and National (gray area) verification regions. Also shown is the location of the NOAA profilers (blue rectangles) and verifying raobs (brown triangles).

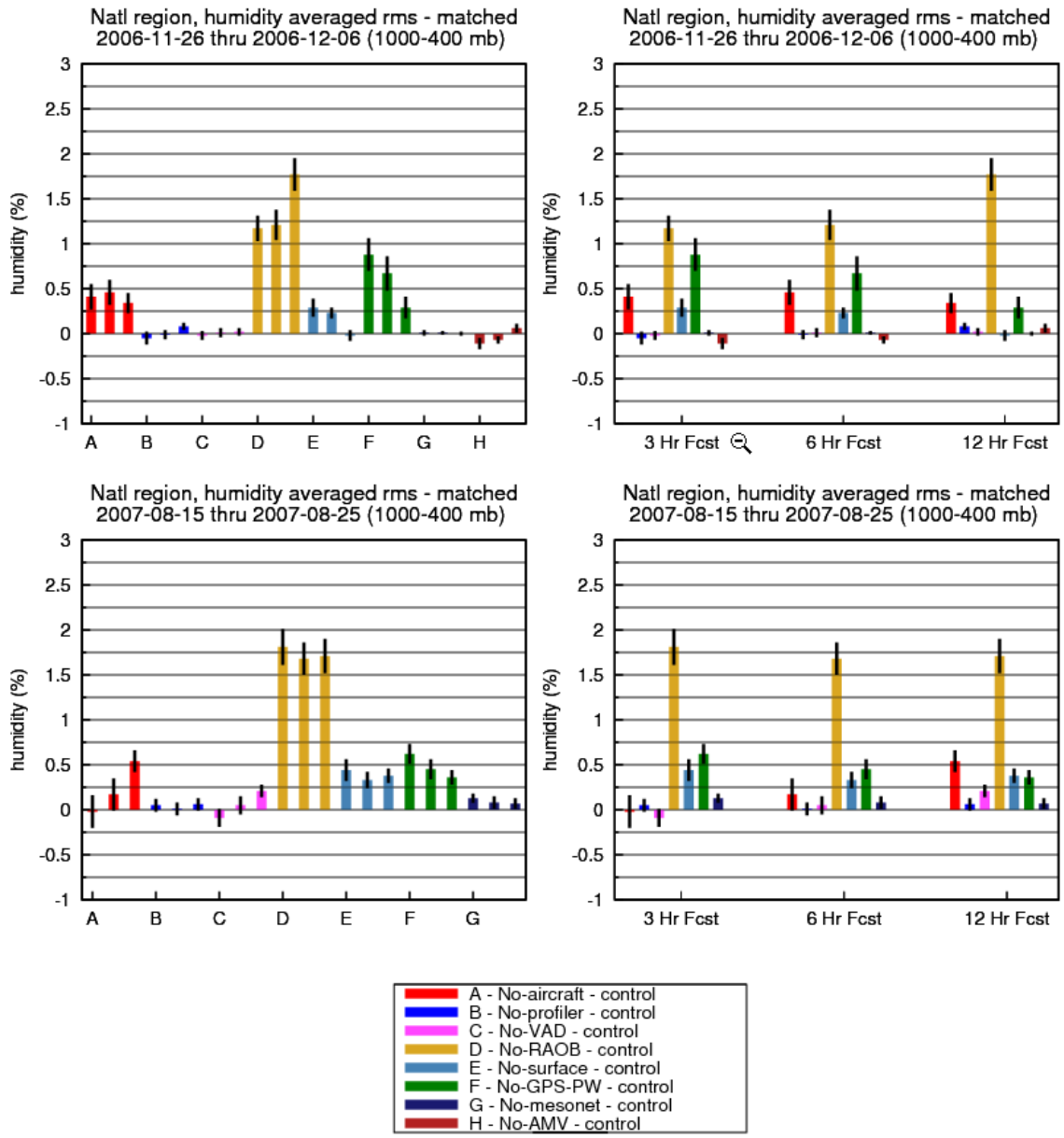

Fig. 4.

Differences in rms error (vs. radiosonde) between observation denial experiments listed in Table 2 and control run for ***1000-400 hPa*** *for* ***relative humidity (in %RH)*** *for* ***national domain.*** *Results for each of the eight observational denial experiments are coded with a different color (aircraft – red, profiler – blue, VAD – pink, RAOB – tan, surface – light blue, GPS-PW – green, mesonet – black, satellite AMV winds – brown). Graphs at top are for winter results; those at bottom are for summer. For graphics at left, the three adjacent columns for each OSE are for 3-h, 6-h, and 12-h forecasts. For graphs at right,*

the same results are organized by observation type for each forecast projection (3h, 6h, and 12h). The impact of AMV data was not tested during the summer period. Statistical uncertainties are indicated for each observation denial experiment by narrow black lines showing +/- 1 standard error from the mean impact.

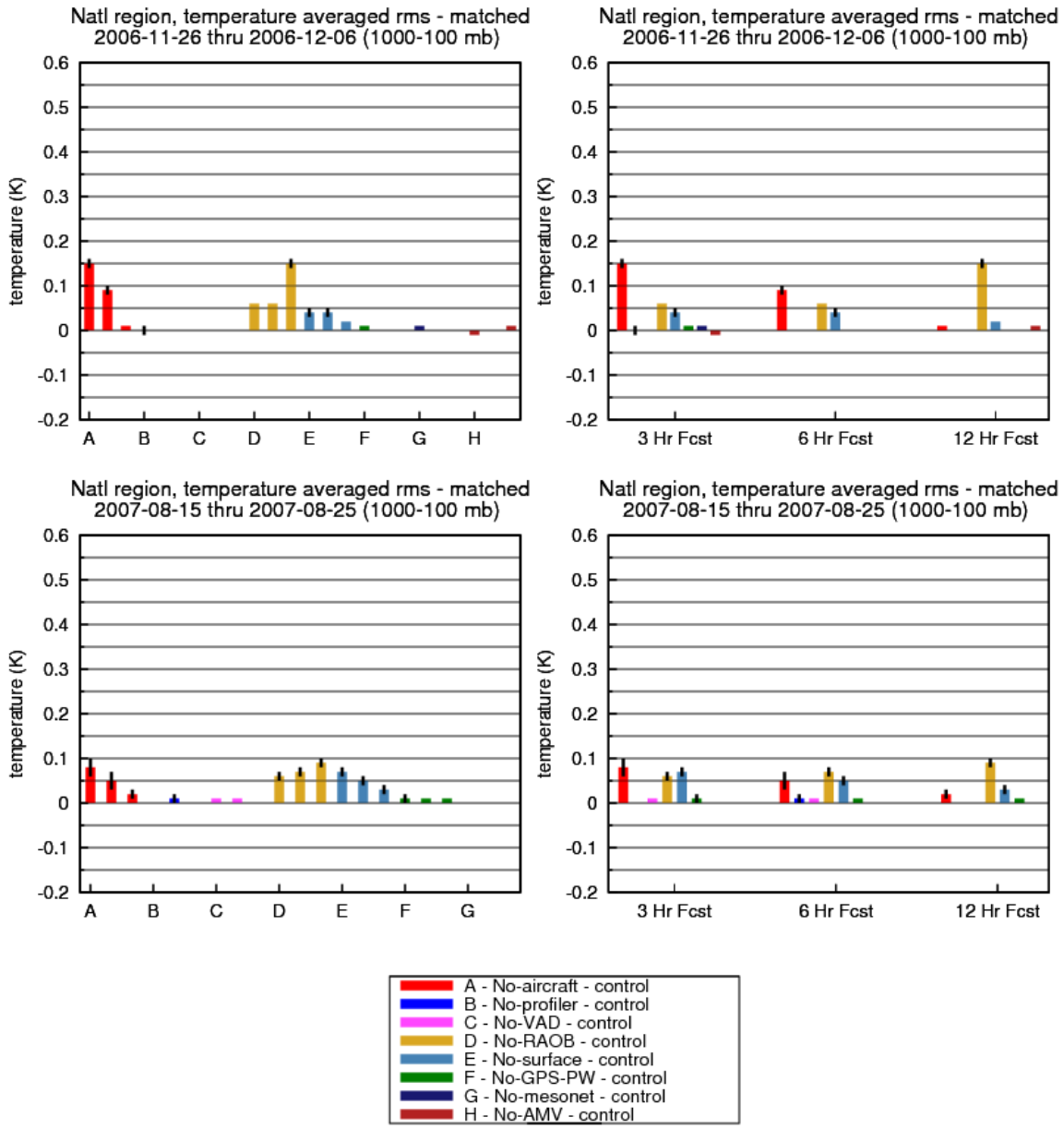

Fig. 5.

Same as Fig. 4 but for ***temperature error (units – K)****, for* ***1000-100 hPa*** *layer over* ***national*** *domain.*

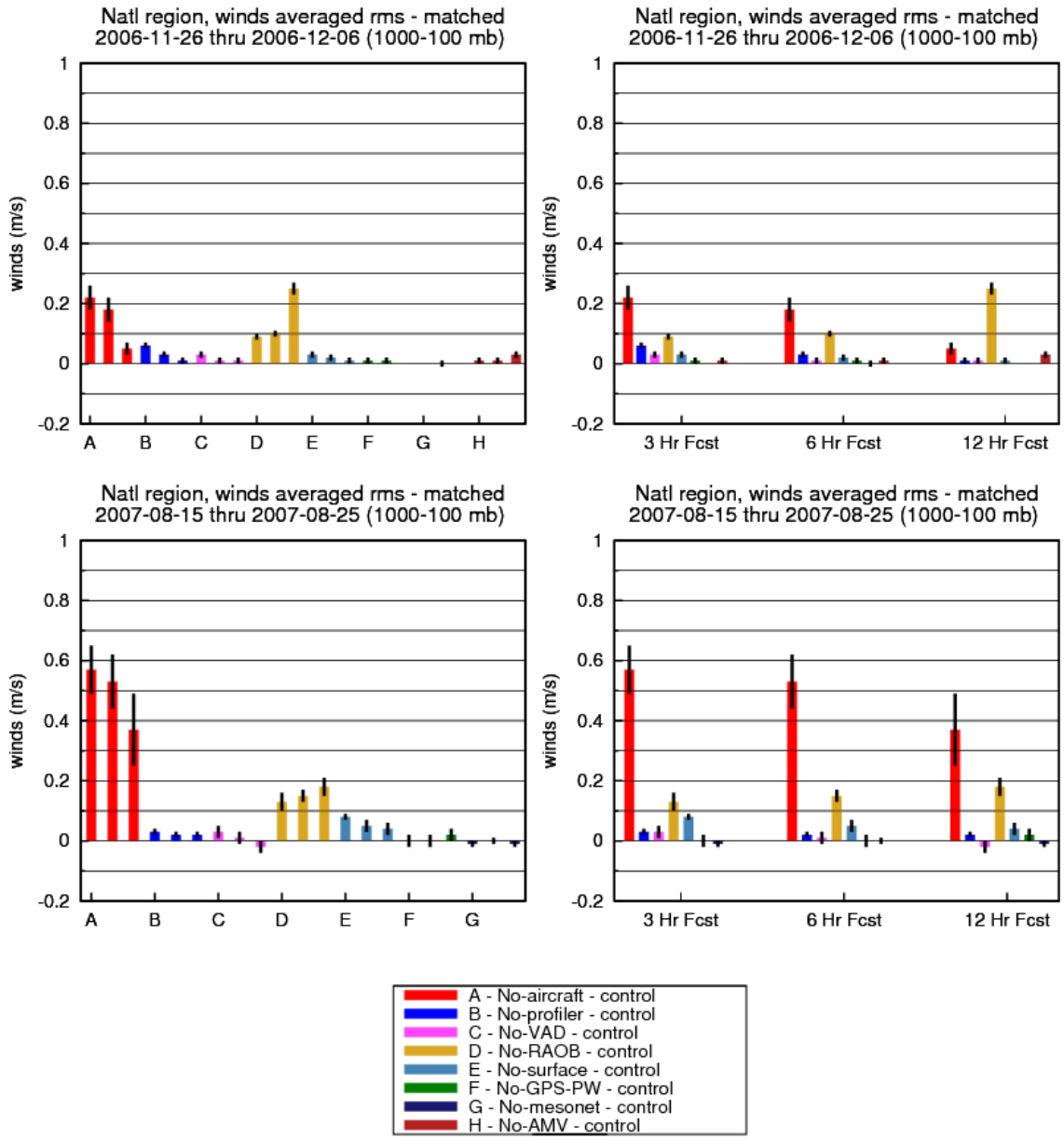

Fig. 6.

Same as Fig. 4 but for ***wind vector difference (units – ms^{-1})****, still for 1000-100 hPa layer over national domain.*

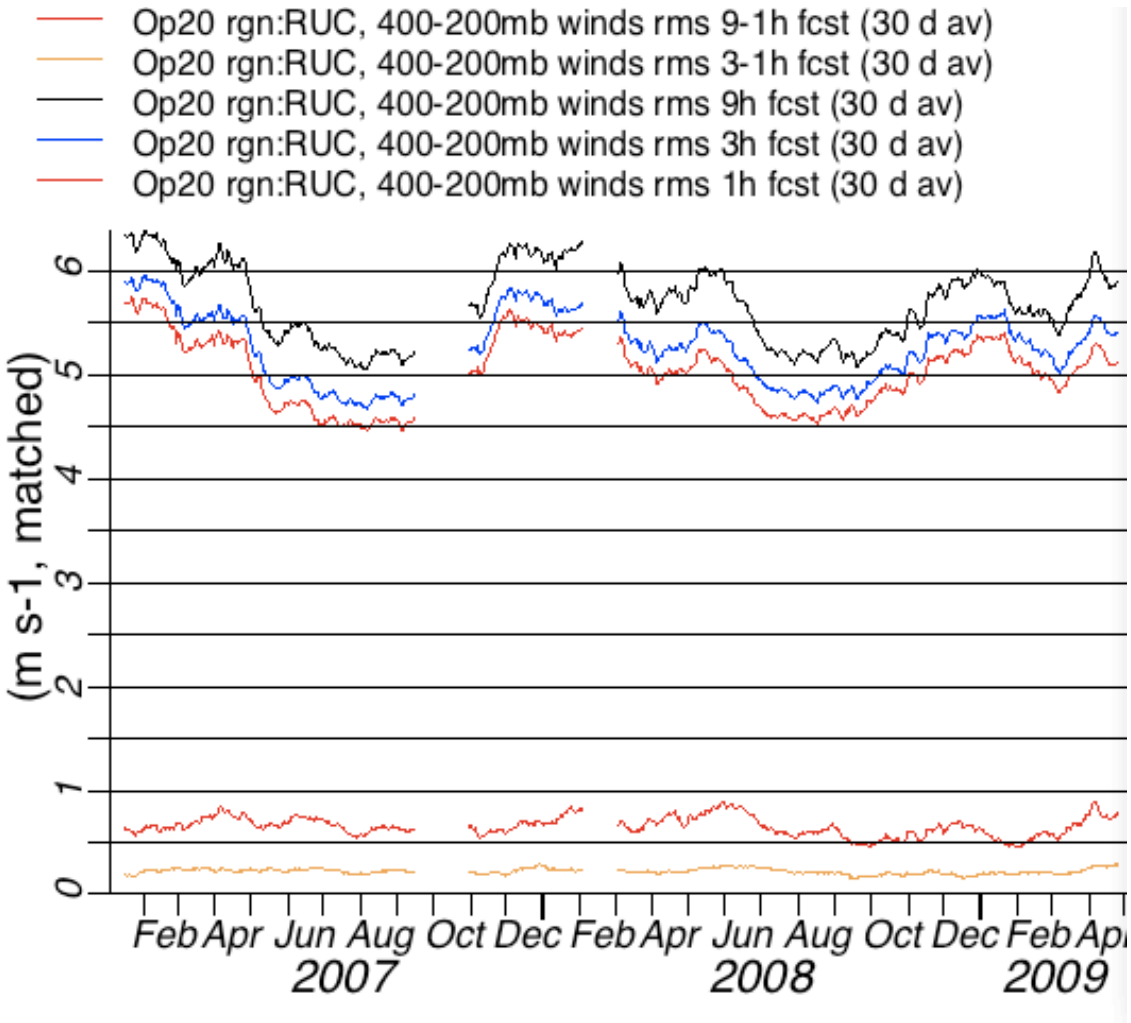

Fig. 7.

Vector wind error (vs. radiosondes) for RUC 9h (black), 3h (blue), and 1h (red) forecasts averaged over 400-200 hPa layer and for a 30-day running mean for January 2007 – April 2009. Also shown are the forecast error increments (between 0 and 1) from assimilation of recent observations for 9-h to 1-h (red, near lower axis) and 3-h to 1-h (tan) pairs.

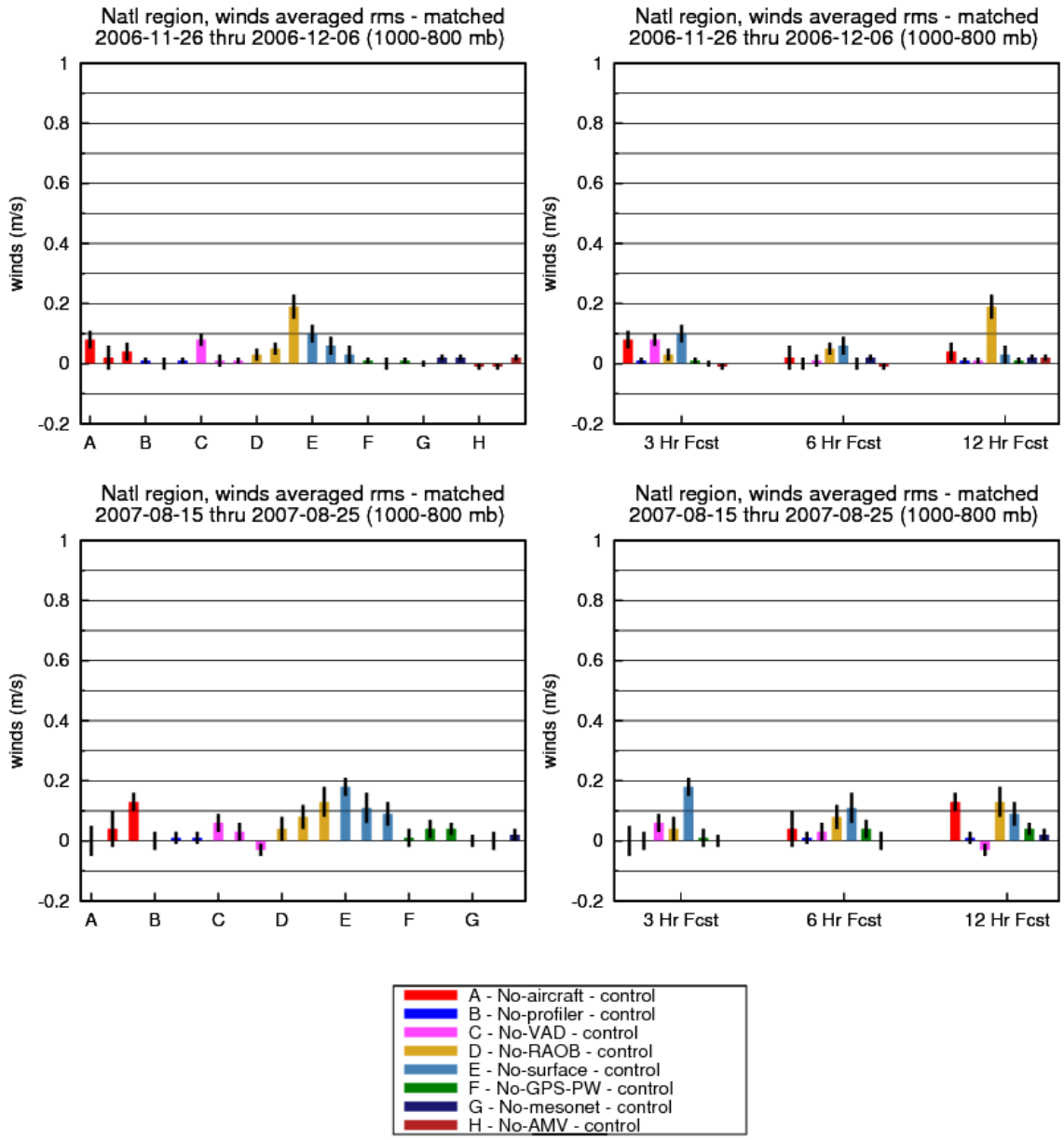

Fig. 8. *Same as Fig. 6, but now for **1000-800 hPa** only (still wind, national domain)*

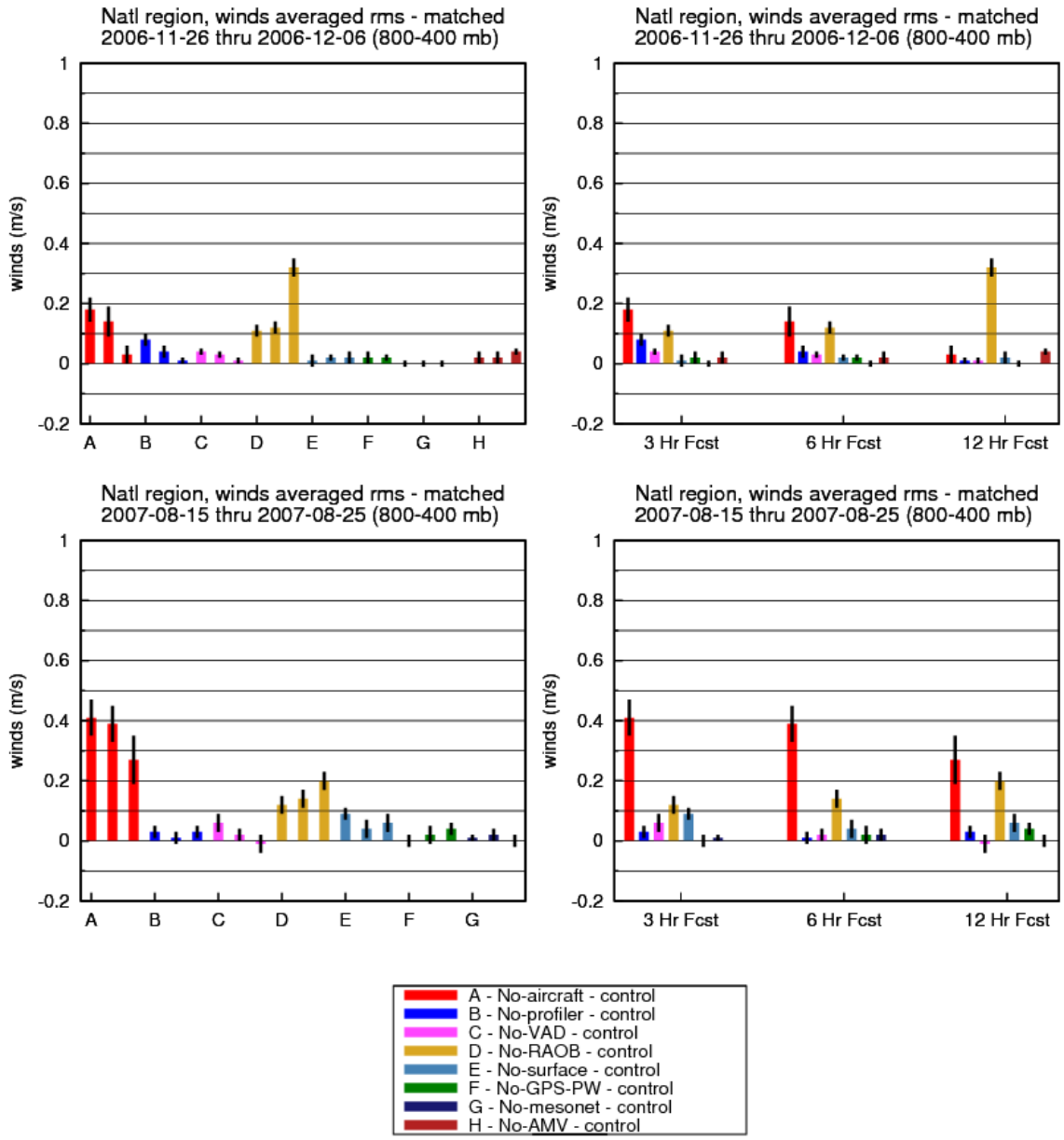

Fig. 9.

Same as Fig. 8, but for ***800-400 hPa*** *layer, still* ***winds*** *for* ***national domain***

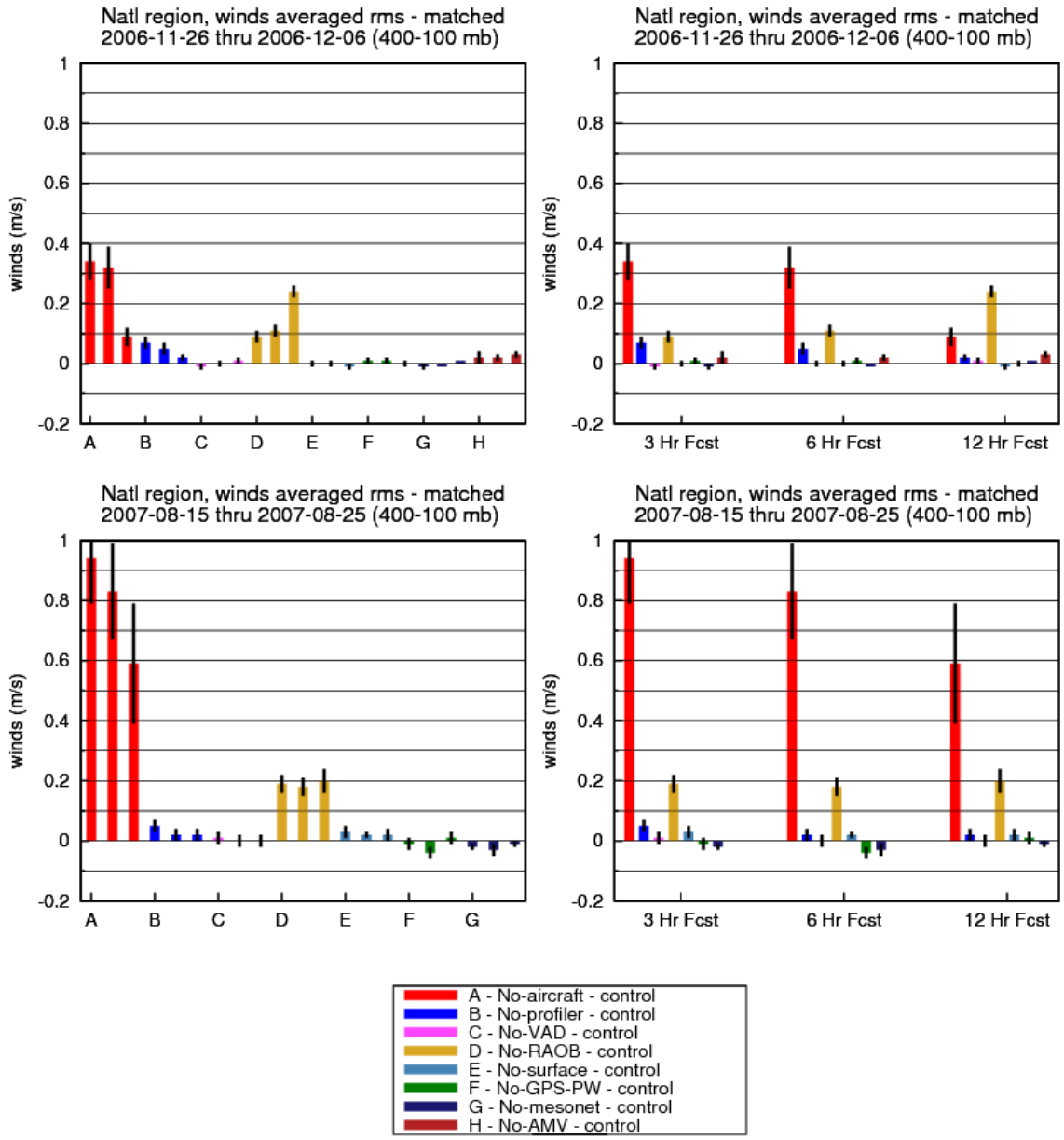

Fig. 10.

*Same as Fig 8 but for **400-100 hPa** layer, still **winds** for **national domain***

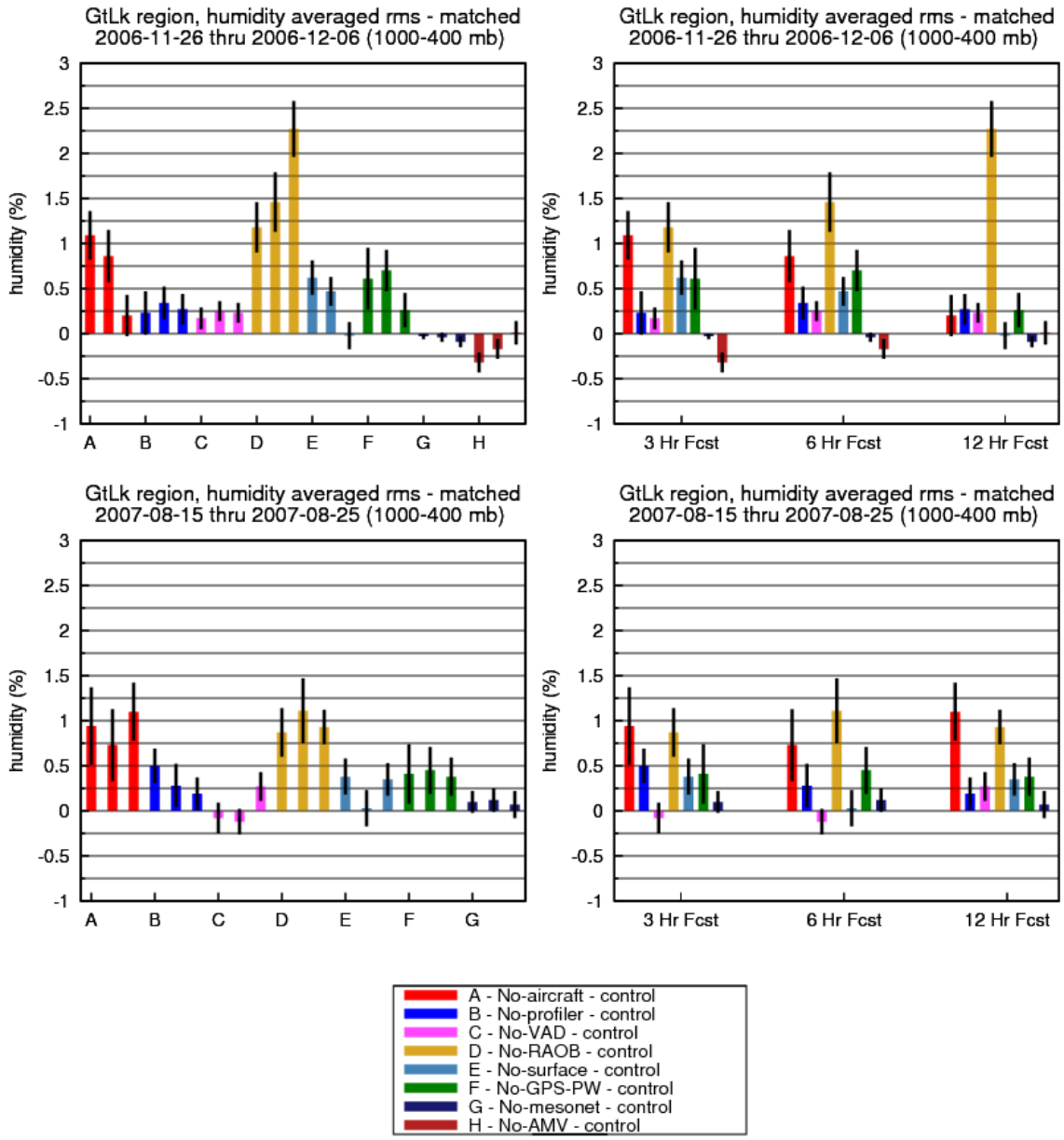

Fig. 11.
Same as Fig.4 (RH, 1000-400 hPa), but now for ***Midwest (or, Great Lakes)*** *regional verification domain.*

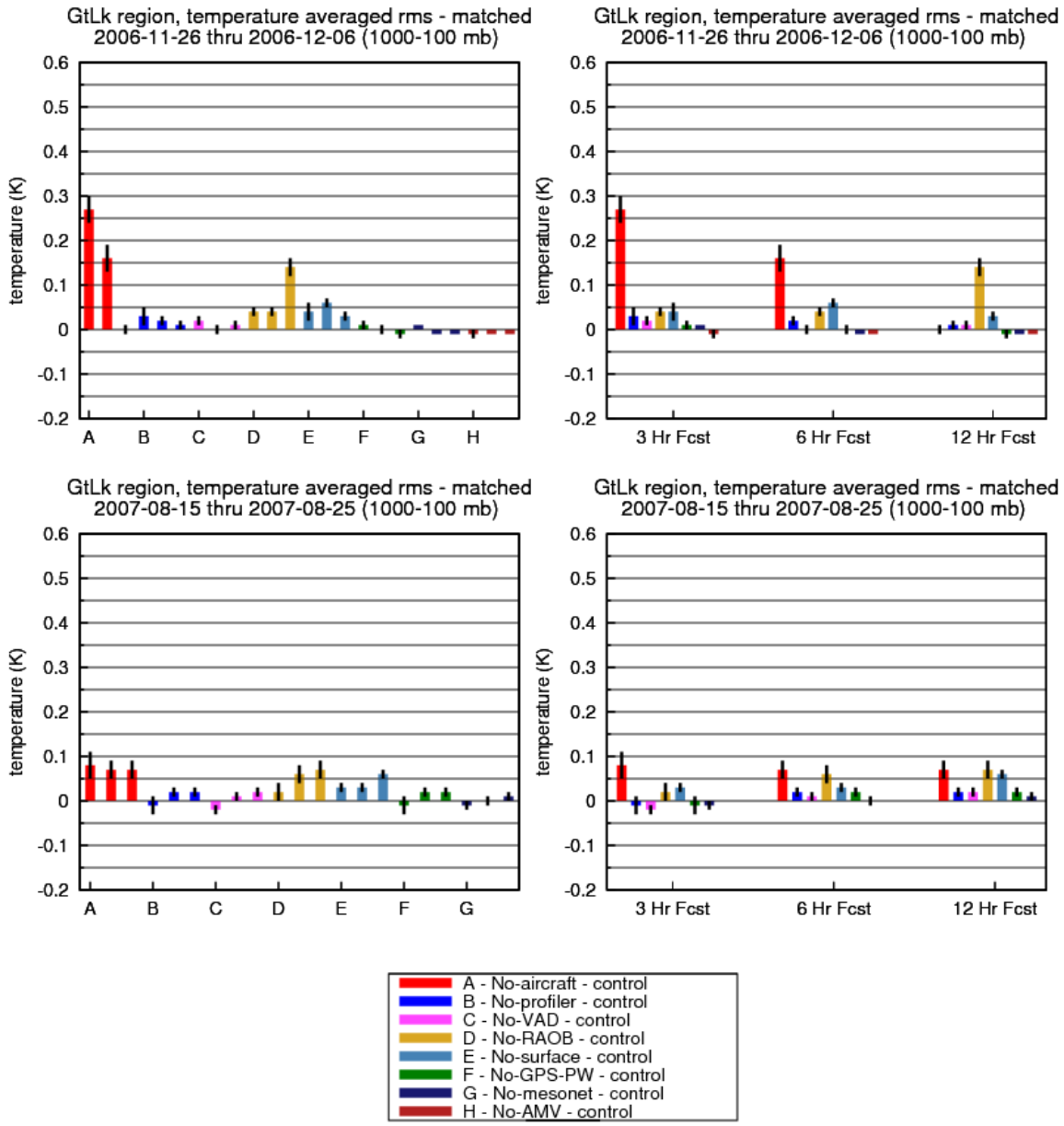

Fig. 12.

*Same as Fig. 5 (**temperature**, full depth, **1000-100 hPa**) but for **Midwest** verification region.*

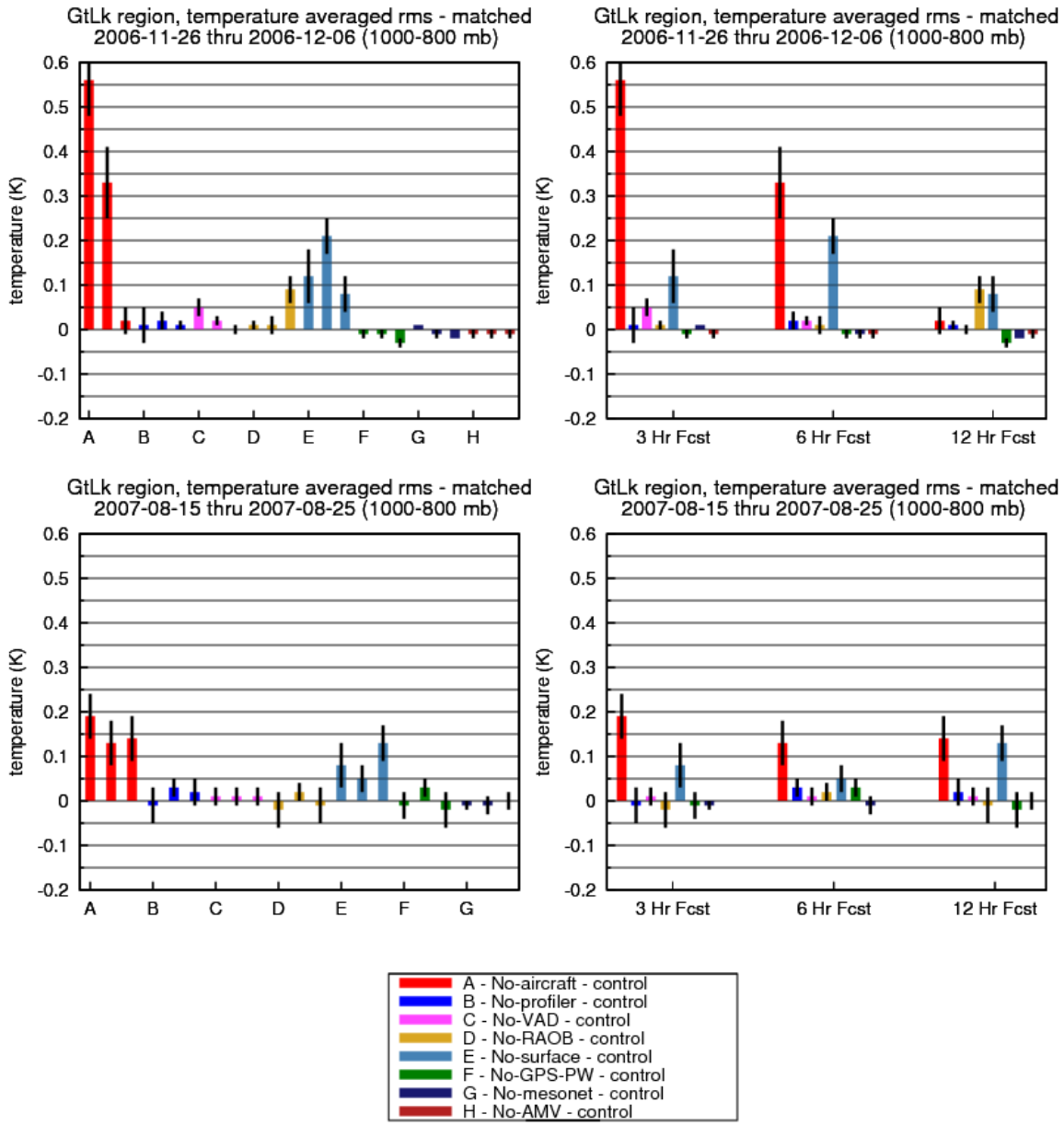

Fig. 13.

Same as Fig. 12 (temperature, Midwest domain), but for ***1000-800 hPa layer*** *only*

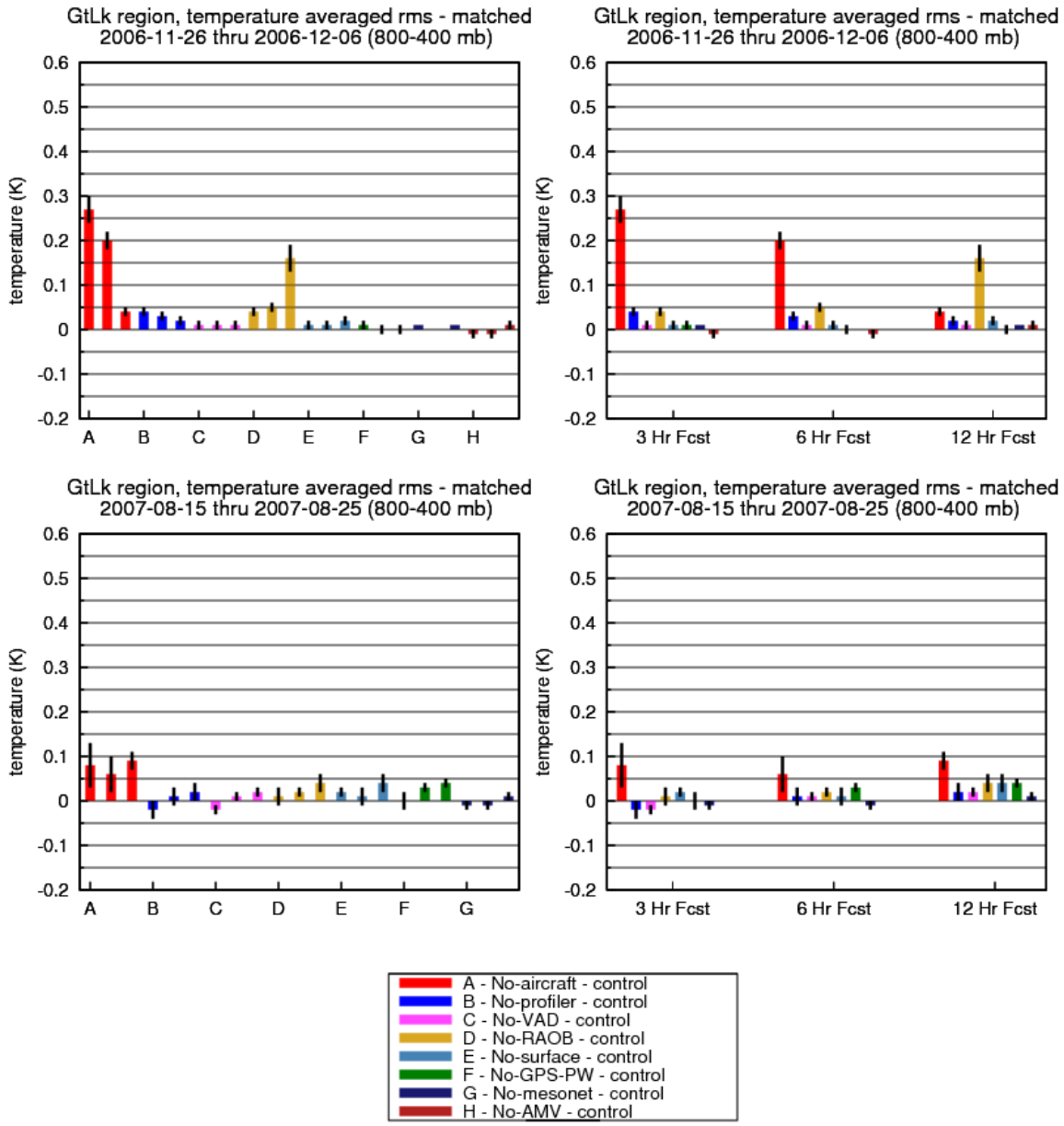

Fig. 14.

Same as Fig. 12 (temperature, Midwest domain), but for ***800-400 hPa layer****.*

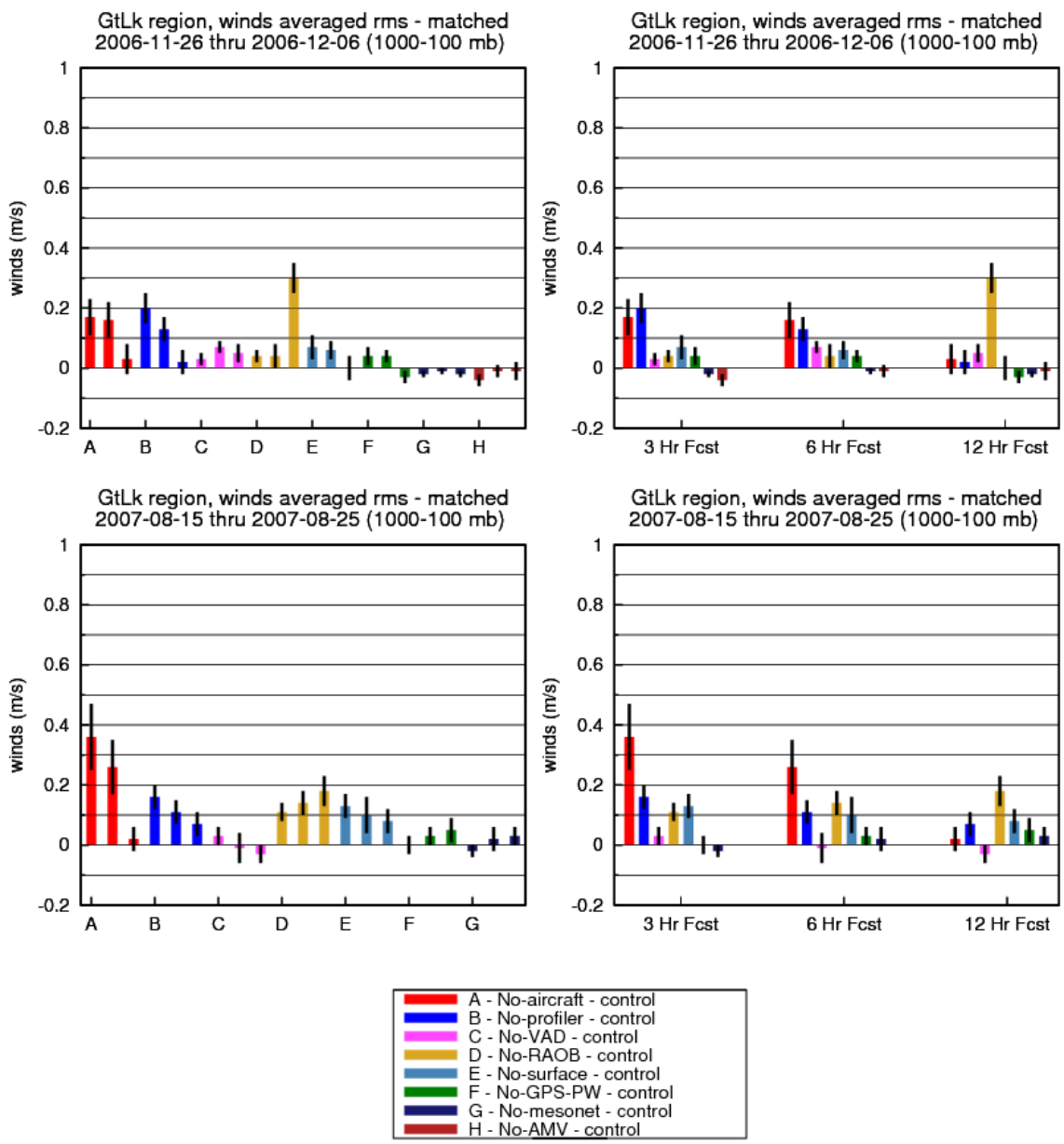

Fig. 15.

Same as Fig. 6 (wind, 1000-100 hPa layer) but for ***Midwest*** *domain.*

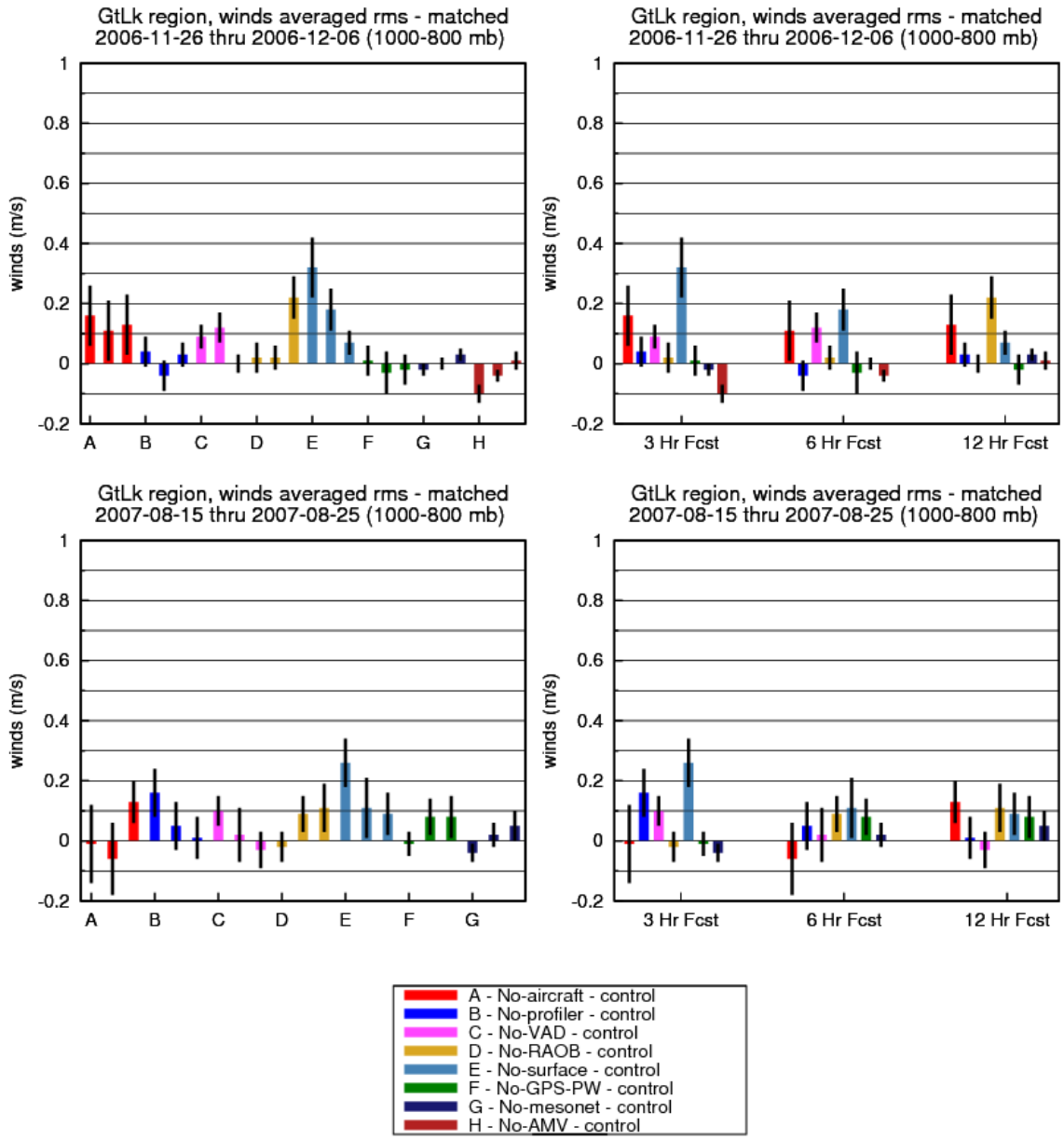

Fig. 16.

Same as Fig. 15 ***(wind, Midwest*** *domain) but for* ***1000-800 hPa*** *layer.*

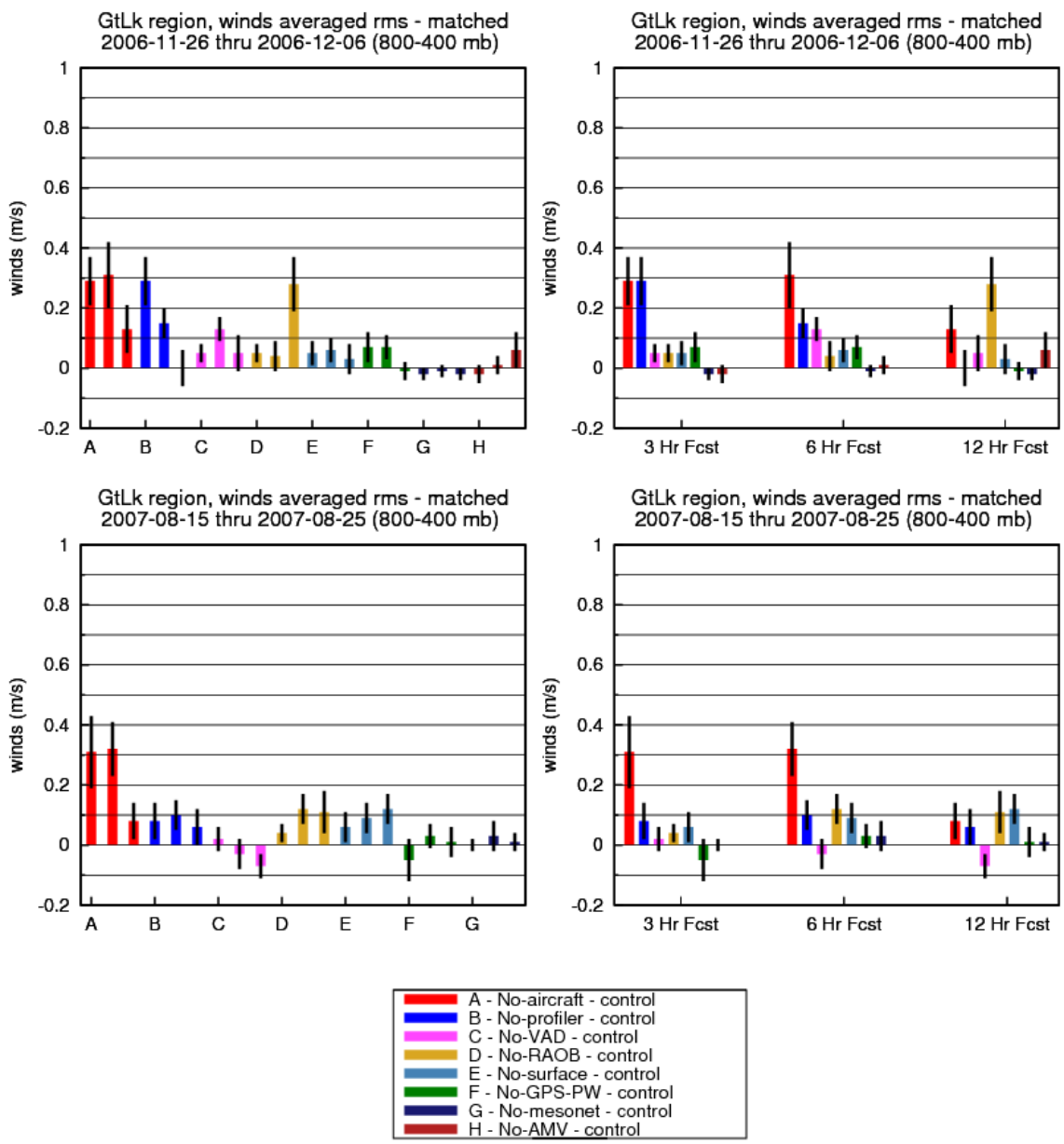

Fig. 17.

*Same as Fig. 15 (**wind, Midwest** domain) but for **800-400 hPa** layer.*

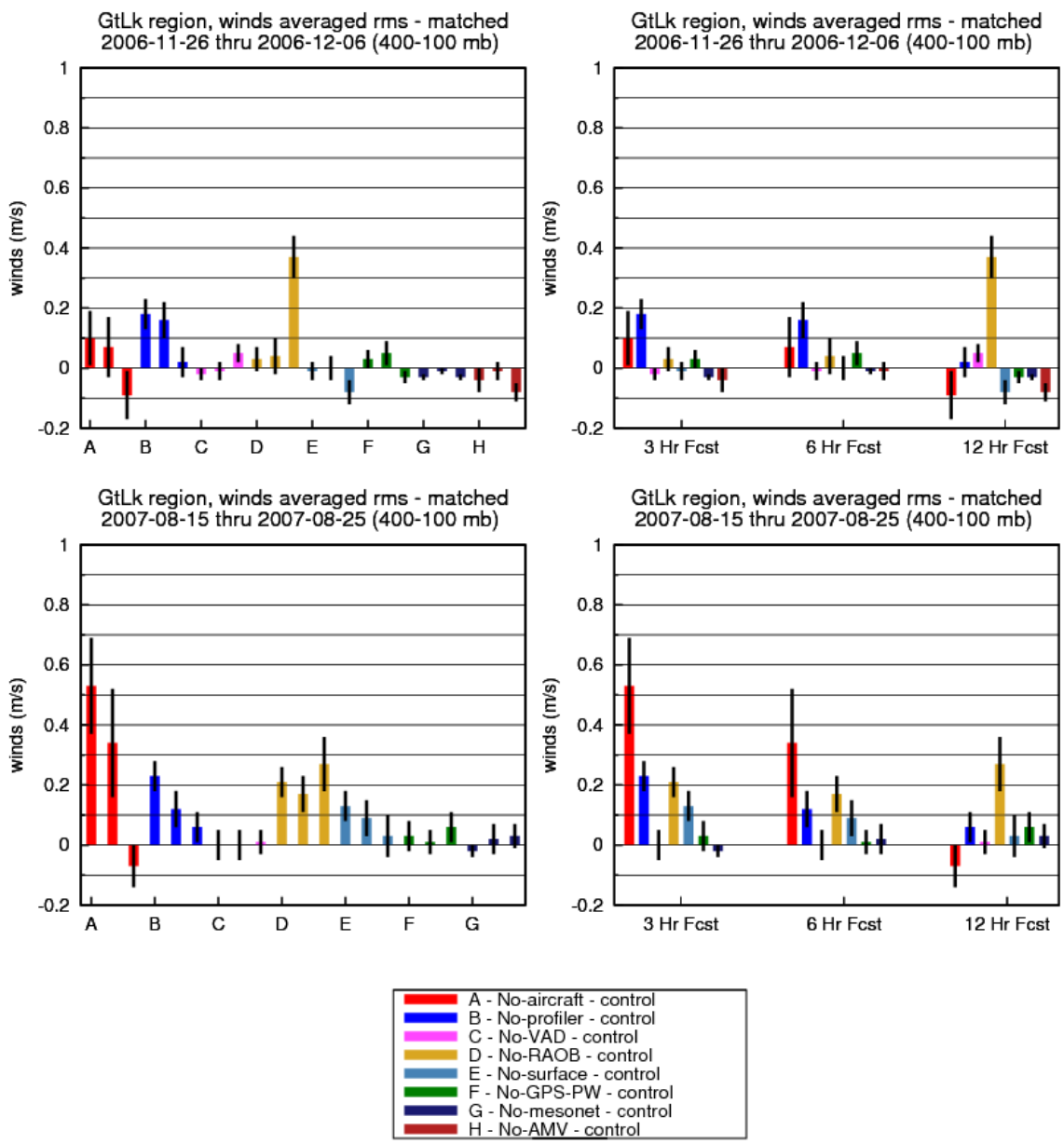

Fig. 18.

*Same as Fig 15 (**Wind, Midwest** domain), but for **400-100 hPa** layer.*

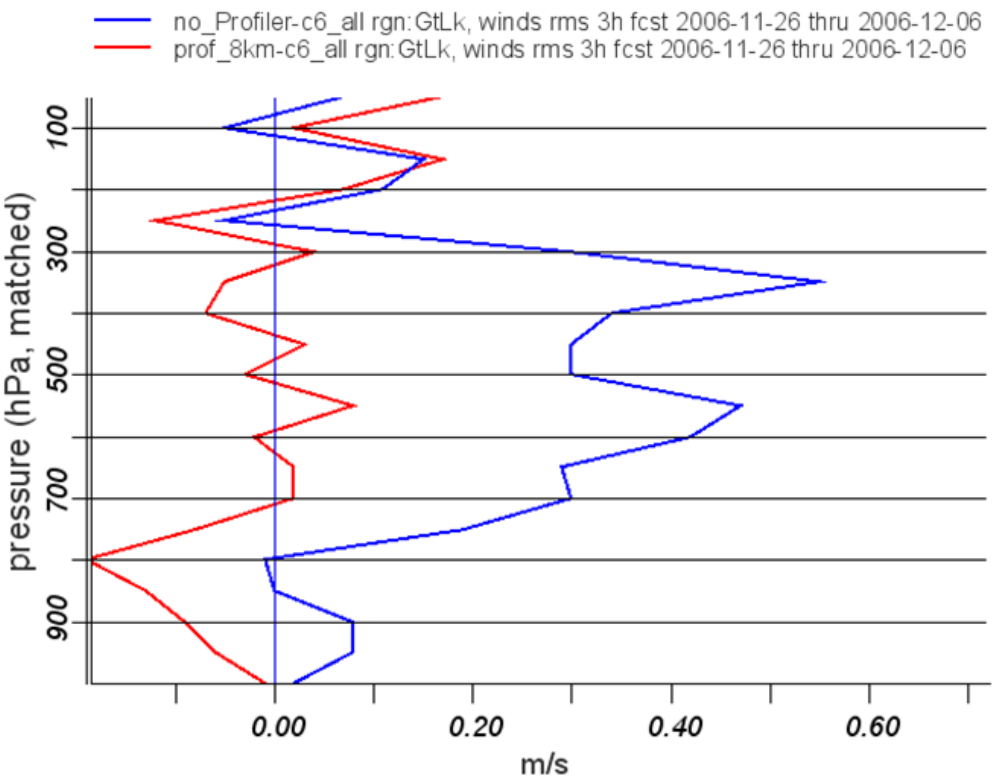

Fig. 19. *Differences in 3-h vector wind errors between No-profiler-minus-Control (N-C, blue line) and 8-km (quarter-scale) minus Control (8-C, red line) over the Midwest region.*

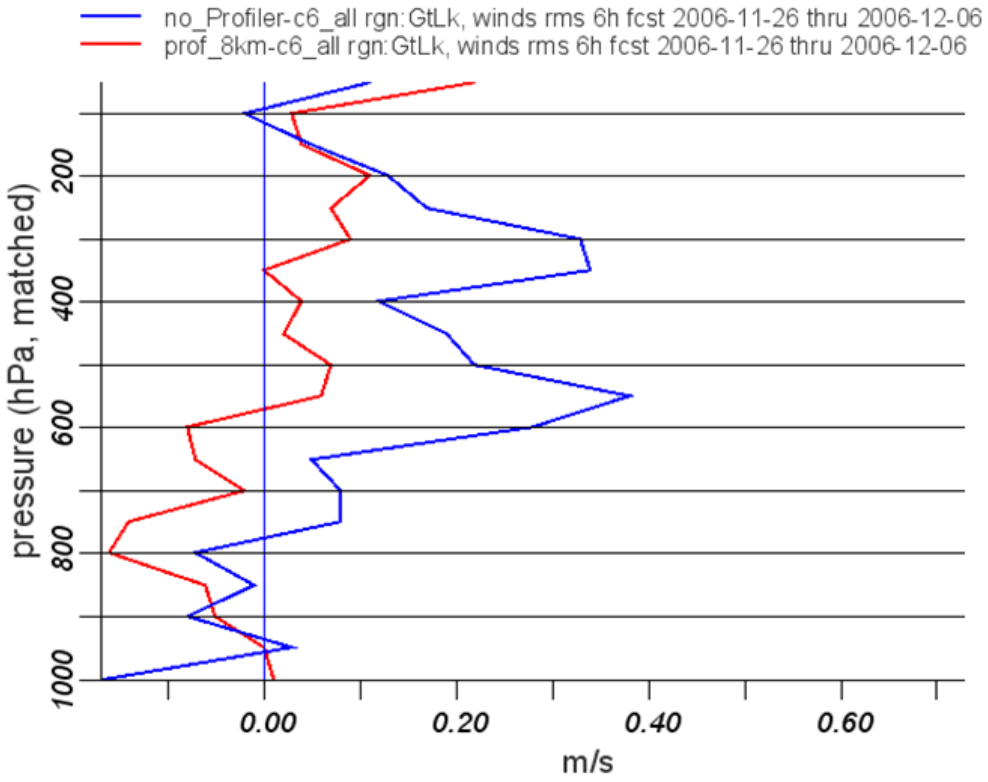

Fig. 20. *Same as Fig. 19 but for **6-h** forecasts instead of 3-h forecasts.*

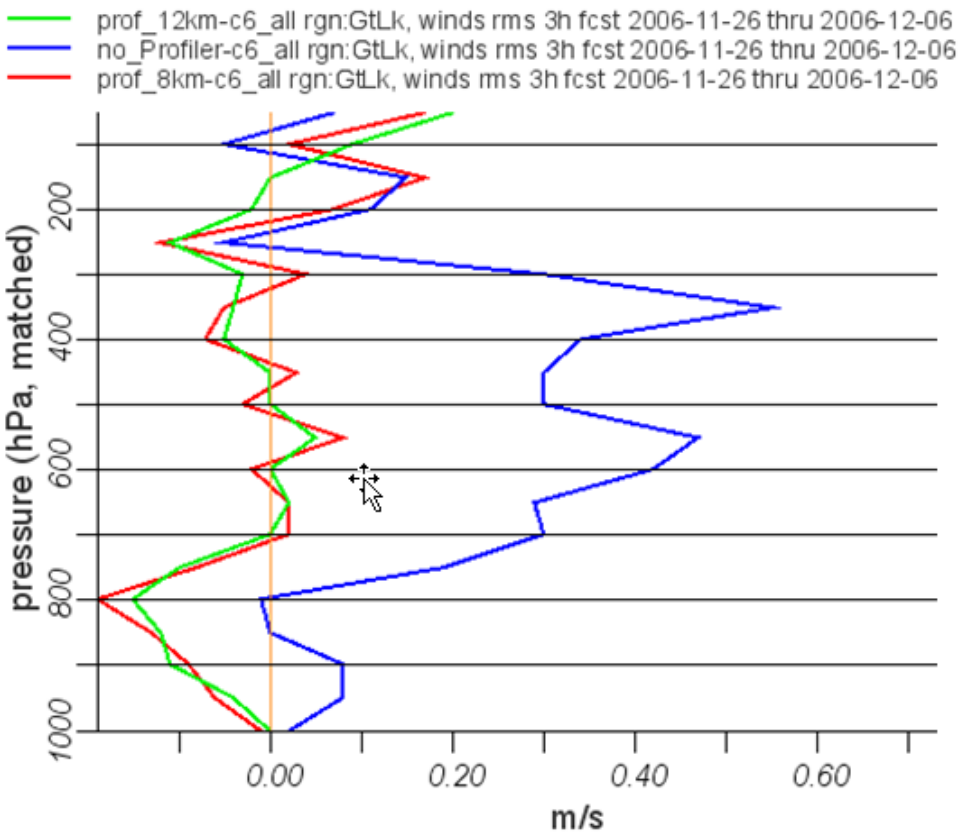

Fig. 21. *Similar to Fig. 20 for 3-h wind forecast impact in Midwest domain, but with* ***results added for 12-km profilers****. Green line (12-C) is for difference between experiments with 12-km profilers (12) vs. Control (C).*

Observation data type	Variables measured	Frequency	Volume (approx)
Radiosonde	P,Z,T,V,RH	12h	80-85
NOAA profilers – 404 MHz	V (by Z)	1h	30
Boundary-layer profilers – 915 MHz, RASS	V (by Z), Tv (by Z)	1h 1h	25 14
VAD winds	V	1h	100-130
Aircraft (AMDAR, not TAMDAR)	V,T	1h	1400-7000
TAMDAR aircraft	V,T,RH	1h	0-800
GOES AMVs (cloud-drift winds)	V	1h	1000-2500
GOES cloud-top pressure, temp	P,T	1h	10km res
GOES precipitable water	PW	1h	10km res - clear areas
GPS precipitable water	PW	1h	250-300
Surface - METAR	P,T,V,Td	1h	1800-2000
Mesonet	P,T,Td,V	1h	7000

Table 1. *Observation types assimilated in the RUC for observation system experiments used in this study. P – pressure, Z – height, T – temperature, V – horizontal wind, RH – relative humidity, PW – precipitable water.*

Experiments	Variables denied	Carried out for Nov-Dec 2006	Carried out for Aug 2007
Control – all observations used		X	X
No radiosonde	Z,T,V,RH	X	X
No profiler winds (NOAA Network or CAP)	V	X	X
No VAD	V	X	X
No aircraft (AMDAR or TAMDAR)	V,T	X	X
No TAMDAR aircraft *(in Moninger et al 2009)*	V,T,RH	X	X
No GPS-PW	PW	X	X
No surface (METARs, buoy, or mesonet)	P,T,V,Td	X	X
No mesonet	P,T,V,Td	X	X
All obs but using 8-km NOAA Network profilers	V	X	
All obs but using 12-km NOAA Network profilers	V	X	
No GOES atmospheric motion vectors (AMVs) (from visible and IR channels, not water vapor)	V	X	

Table 2. *Observation impact experiments in this study. Those observational variables denied to the RUC are shown for each experiment. P – pressure, Z – height, T – temperature, V – horizontal wind, RH – relative humidity, PW – precipitable water.*

Experiment period	Beginning	End	Notes
Winter	26 November 2006	5 December 2006	Strong winter storm early, moderate winter weather later (in Midwestern US)
Summer	15 August 2007	25 August 2007	Active period for convective storms in Great Lakes region

Table 3. *Experiment periods.*

- **Nine experiments (control, 8 observation denial experiments)**
    - Control
    - No aircraft (meaning no AMDAR or TAMDAR, no aircraft of any type, listed as ‘ noAMDAR’ – automated aircraft reports)
    - No profiler (no profiler of any kind, NOAA or CAP)
    - No VAD winds
    - No raobs
    - No surface (no surface of any kind; METAR, mesonet or buoy)
    - No GPS precipitable water
    - No satellite AMVs (cloud-drift winds – listed as ‘CDW’)
    - No mesonet

- **Two Regions**
    - US National (data rich)
    - Midwest (very data rich)

- **Four layers**
    - 1000-100 hPa (full depth), 1000-400 hPa for RH only
    - 1000-800 hPa (near surface)
    - 800-400 hPa (mid-troposphere)
    - 400-100 hPa (upper troposphere, lower stratosphere)

- **Two seasons**
    - winter
    - summer
- **Forecast duration**
    - 3h, 6h, 12h

Table 4. *Verification stratifications for OSEs*