Preliminary statistical evaluation of Rapid Refresh cycled forecasts initialized with GSI

Steve Weygandt[1], Dezso Devenyi[1,2], Ming Hu[1,2],
Bill Moninger[1], Barry Schwartz[1], Stan Benjamin[1]

[1]Global System Division, Earth System Research Laboratory, NOAA
[2]Cooperative Institute for Research in Environmental Sciences, University of Colorado

March 14, 2008

DRAFT

PLEASE CHECK BACK TO:

http://ruc.noaa.gov/RR_preliminary_verification_report.pdf

FOR AN UPDATED VERSION

1. Introduction.

This report provides a very preliminary summary of the ongoing statistical evaluation of the current real-time experimental Rapid Refresh (RR) cycle forecast. The Rapid Refresh cycle is an hourly mesoscale meteorological analysis and prediction system currently under development at GSD, and includes collaboration with NCAR and NCEP EMC. The RR will replace the current Rapid Update Cycle (RUC) system, run operationally at the National Centers for Environmental Prediction (NCEP). The RR utilizes the WRF ARW model and the Gridpoint Statistical Interploation (GSI) analysis package. The GSI was developed at NCEP and includes full satellite data assimilation capabilities. For this particular application of GSI to the Rapid Refresh, a generalized cloud analysis procedure using satellite cloud top pressure, METAR cloud information, and radar reflectivity mosaic data has been added to the GSI. A recent inclusion in our Rapid Refresh application is the use of full satellite radiance data with the NCEP cycled satellite bias prediction package.

Development of the fully functioning 1-h cycled RR system, including GSI with the generalized cloud analysis, full use of the satellite data, and use of expanded hourly observation data file (transferred hourly from NCEP GSD), on the GSD supercomputer has been slowed by computer difficulties related to application of the GSI on a computer architecture different from the one it was developed. These difficulties have been surmounted, however, and we now have in a place a relatively stable 1-h RR cycle with all major components. This hourly cycle complements previously implemented RR cold start capabilities and a 6-h RR cycle. Thus, while our work efforts up to now have focused almost entirely on assembling the RR system on the GSD wjet supercomputer, we are now at a point where we can turn our attention to optimizing the various components of the RR system to maximize forecast skill. In this document we present some extremely preliminary verification statistics. **Please check back to the link listed on the front page, as this document will be updated on a periodic basis as further verification and model and analysis refinement work is completed.**

2. Rapid Refresh verification modules

Toward the goal of maximizing RR forecast skill, we have recently implemented or begun to implement automated verification packages needed to evaluate the RR forecast performance relative to the current operation RUC model and other standards. These packages include: 1) an upper-air verification package, 2) a precipitation verification package, and 3) a surface verification package. Details on the verification packages are summarized below.

2a) Upper-air verification package

The upper air verification package is run at GSD, utilizes a mySQL database and has been improved based on TAMDAR evaluation work. Specific details of the improvement are as follows:

Under the previous verification system,

- comparisons are made only at mandatory sounding levels (850, 700, 500, 400, 300, 250, 200, and 150 mb)
- Model data are taken from pressure-based files, which are interpolated in the horizontal and vertical from the models native coordinate system
- RAOB data that fail quality control checks made with the *operational* RUC cycle are not used.

Under the new verification system,

- Full RAOB soundings, interpolated to every 10 mb, are compared with model soundings
- Model soundings, interpolated to every 10 mb, are generated directly from model native files
- Comparisons are made every 10 mb up from the surface.
- No RAOB data are automatically eliminated. (About a dozen obviously erroneous RAOBs have been eliminated by hand since 23 February 2006).

By interpolating from full RAOB soundings (including the significant level data) to every 10 mb, the new verification system yields verification soundings with much higher effective vertical resolution compared to the old verification system (which was effectively substantially sub-sampling the vertical profile of information provided by the raw RAOB data). This issue was found during the TAMDAR moisture evaluation, because with the old system vertical variations in relative humidity that were in both the raw RAOB data and the TAMDAR profiles were lost when only the mandatory levels were used. Also, during the TAMDAR evaluation, it was found that excluding RAOB data based on differences from operational RUC was occasionally removing good sounding data. These changes have yielded a new upper-air verification system that improves upon previous version, especially in its ability to validate detailed vertical structures. Upper-air skill scores can also be computed for a variety of different regions and partitioned by 00z and 12z RAOB.

2b) Precipitation verification package

The precipitation verification package was developed within GSD, but utilizes components from NCEP EMC including the “IPOLATES” interpolation routine. Verification data come from the quality controlled 24-h “stage 4” analysis provided by NCEP and shorter duration radar estimated accumulations. For the 24-h precipitation verifications, two accumulation methods are used: 1) summing the 12-h accumulation from 2 successive 12-h forecasts (00z, 12z) and 2) summing the 3-h accumulations from 8 successive 3-h forecasts. Additional work is ongoing to sum 6-h accumulations from 4

successive 6-h accumulations. Each one of these provides different information and allows the duration of the impact from various assimilation procedures (especially radar reflectivity assimilation) to be evaluated. Precipitation verification statistics can also be computed for shorter duration periods, allowing for aspects such as the diurnal cycle of precipitation prediction skill (especially important during the warm season) to be evaluated. Precipitation skill scores can also be computed over various subsections of the CONUS domain.

2c) Surface verification package

The surface verification package was developed within GSD to evaluate forecast skill for surface temperature, dew point, wind, and also for ceiling and visibility. METAR observations provide the input data to the verification database. Model-based surface fields are provided based on optimized forward models in the various model post-processing program (HYBPOST for RUC and WRFPOST for RR). Standard scores (RMS, bias, etc.) are computed for surface T, Td and wind) and standard categorical scores (POD, CSI, bias, etc.) are computed for detection of specific ceiling and visibility thresholds.

3. Rapid Refresh verification results

We are currently running three different RR configurations (all over a domain covering all of North American including Alaska and covering much of the Caribbean including Puerto Rico): 1) a "cold start" 24-h forecast initialized from the GFS and using GFS lateral boundary conditions 2 times per day (00z, 12z), 2) a 6-hourly cycled RR forecast with 12-h forecasts using GFS boundary conditions 4 times per day (00z, 06z, 12z, 18z), and 3) a 1-hourly cycled RR forecast with forecasts of up to 12-h using GFS boundary conditions. Besides the inclusion of the GSI data assimilation module in the cycled runs, an important difference is that the cold start runs are utilizing the diabatic digital filter initialization (DDFI – coded and fully operational for ARW), but it has not yet been introduced into the RR cycle version of the ARW. This will be introduced in the next few weeks with the upgrade to WRF version 3.

As noted in the introduction, while the major RR-specific components are in place for the 1-h RR cycle (the 6-h RR cycle is not using the satellite radiance assimilation module), we have done virtually no calibration or optimization of the various GSI aspects. In particular, we are using background error covariance and observation statistics from the 2006 NAM (will be updated soon), have not updated to the "Q-option2" moisture option (currently being implemented for ARW by Dezso Devenyi) and are not using SSMI data or GPS precipitable water data. So these statistics provide a look at the RR forecasts with the components assembled "out of the box" and provide a baseline from which we can begin to calibrate and optimize the forecast performance. It should be noted that we have only examined statistics for the CONUS region; however the upper-air package is currently producing statistics over the entire North American domain of the RR.

We have examined the performance cold start runs (statistics not shown) and found surprisingly good results, especially for upper-level winds. This reflects the superior balance in the GFS model forecasts and indicates it may be desirable to use GFS lateral boundary conditions for the RR cycled runs.

In this document we will restrict our assessment to a very preliminary look at the first upper-air statistics from a comparison of the RR 1-h cycle, with a 6-h RR cycle, and the NCEP operational RUC. The results are from an extremely short period, and we have just now begun to accumulate precipitation statistics and are still getting the surface verification applied to the RR grids, so as noted in section 1, **please check back to the link listed on the front page, as this document will be updated on a periodic basis as further verification and model and analysis refinement work is completed.**

Despite the severe limitations of our statistical analysis to date, the initial upper-air statistics suggest that 1) the current non-optimized RR system is competitive with current systems and more importantly 2) that the 1-h cycle is beating the 6-h cycle for upper-level wind and temperature forecasts.

Fig. 1 shows the vertical profile of RMS errors for the analysis (fit to observations) and 6-h upper-level temperature forecasts from the operational RUC, 6-h cycled RR and 1-h cycled RR. As can be seen, both of the RR cycles provide analyses that fit the observations less closely that the RUC. This is not a problem, as overly fitting the observations (which contain errors) can produce unbalanced, less skillful forecasts. It is, however, something we will be examining as we evaluate the observation and background error specification in the RR. Consistent with its use as a situational awareness model, the RUC has traditionally provided a very close fit to the observations. As shown in the lower panel of Fig. 2, 6-h forecast upper level temperature forecasts from the 1-h RR cycle are not quite as good as the operational RUC, but better than the 6-h RR cycle. Given the non-optimized state of the RR, this is an encouraging result.

A consistent, but more pronounced, pattern of improvement from the 6-h cycle to the 1-h cycle is seen in the wind RMS errors (Fig. 2). In particular, for the crucial 6-h forecast, the upper level wind errors for the 1-h cycle show a definite improvement compared to the 6-h cycle and are fairly close to the errors for the operational RUC. Again, we caution that these results are from an extremely short period, and we will be continuing to monitor the statistics and updating this report.

Finally, the upper-level relative humidity RMS errors are shown in Fig. 3. Here, the 1-h cycle has slightly larger errors than the 6-h cycle at most levels for both the analysis and the 6-h forecast, and both RR cycles are worse than the operational RUC. Several noted deficiencies in the RR (including lack of Q-option2 and no assimilation of SSMI or GPS precipitable water) likely play a role here, and further evaluation of the relative humidity results is deferred until these deficiencies have been remedied.

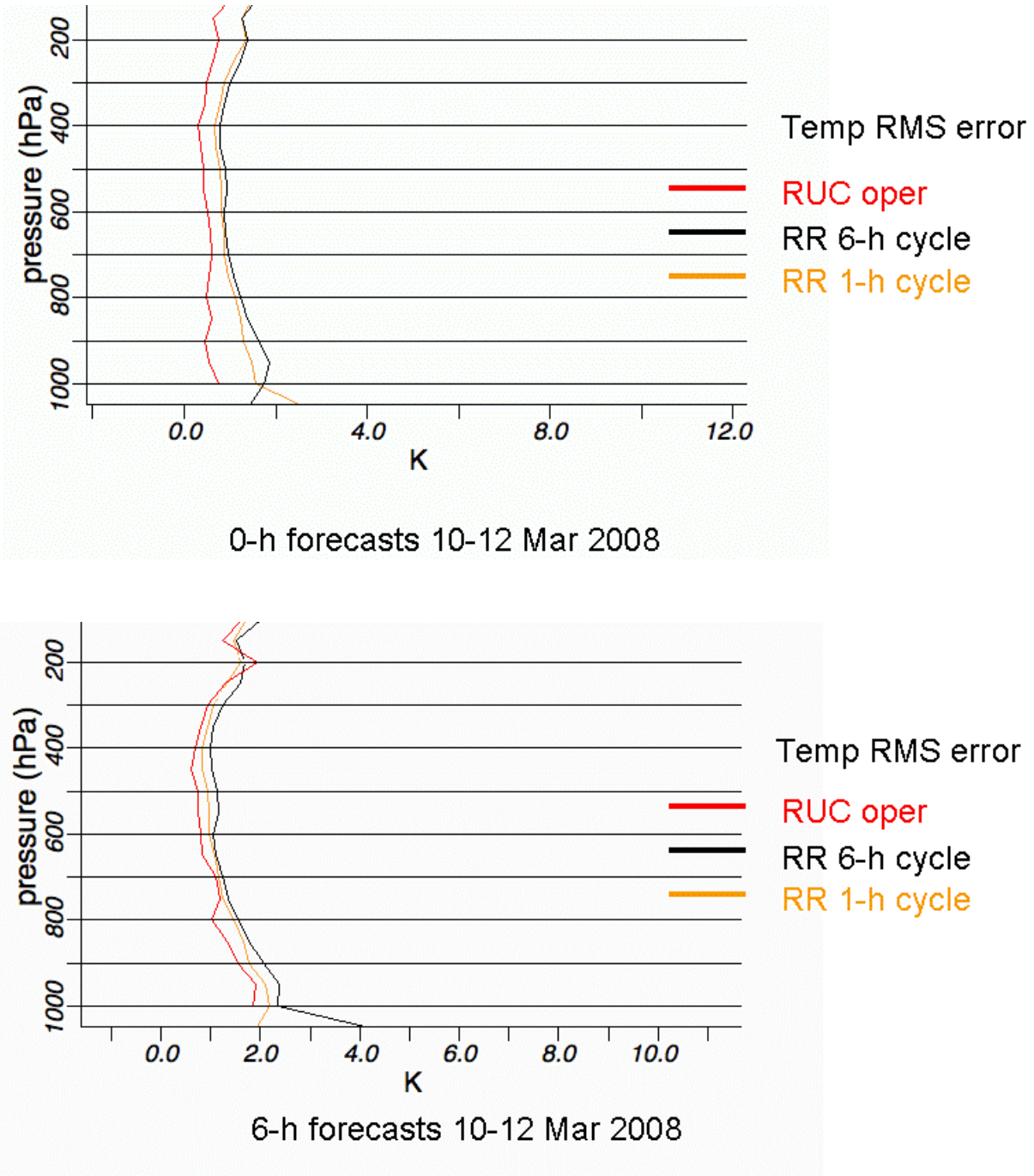

Fig. 1 Vertical profile of temperature RMS errors (computed from US rawinsonde network) for period 10-12 March 2008 for (top panel) analysis and (bottom panel) 6-h forecast. The red curve is for the NCEP operational RUC, black curve is for the RR 6-h cycle and arrange curve is for the 1-h RR cycle.

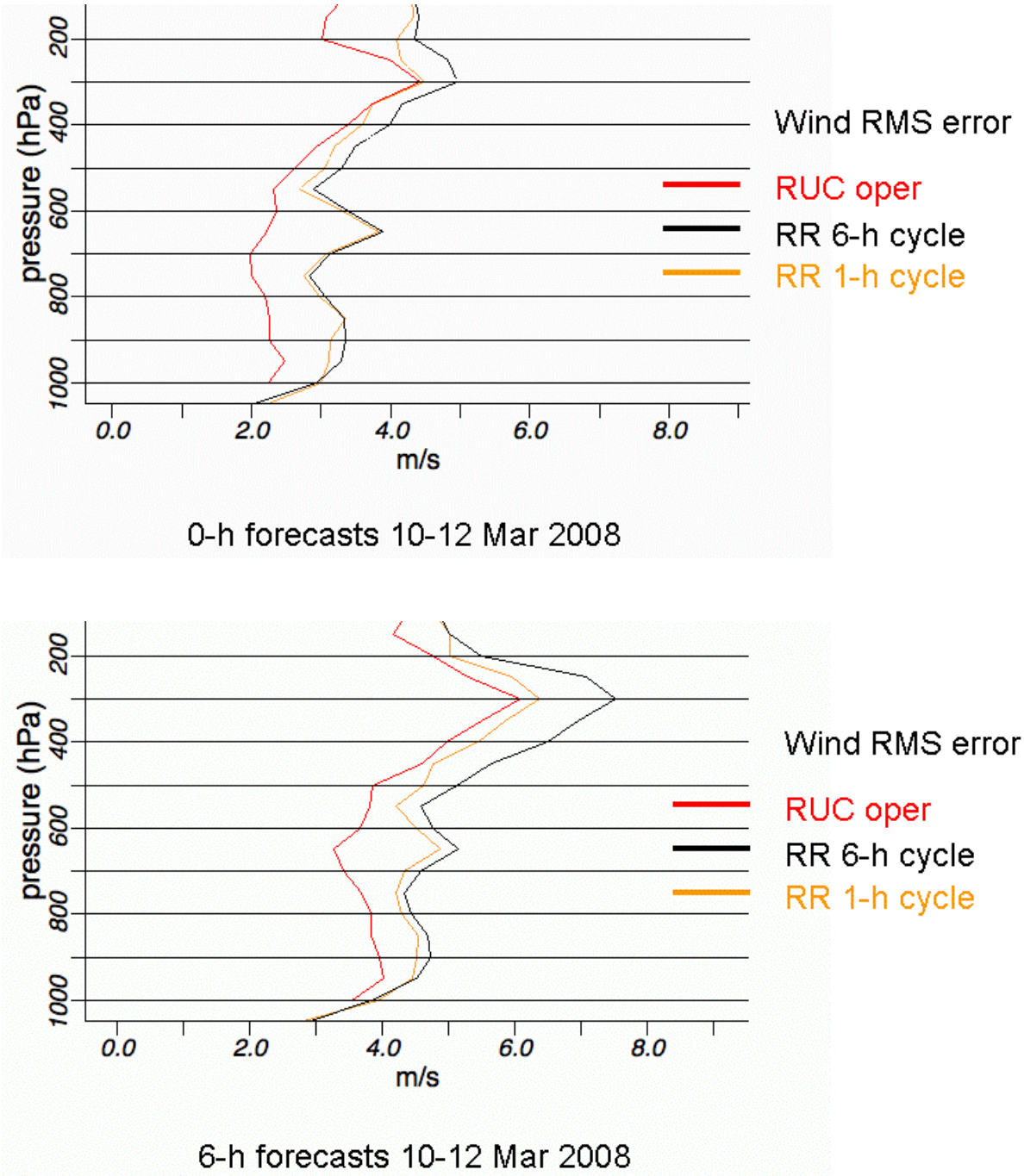

Fig. 2 Vertical profile of wind RMS errors (computed from US rawinsonde network) for period 10-12 March 2008 for (top panel) analysis and (bottom panel) 6-h forecast. The red curve is for the NCEP operational RUC, black curve is for the RR 6-h cycle and arrange curve is for the 1-h RR cycle.

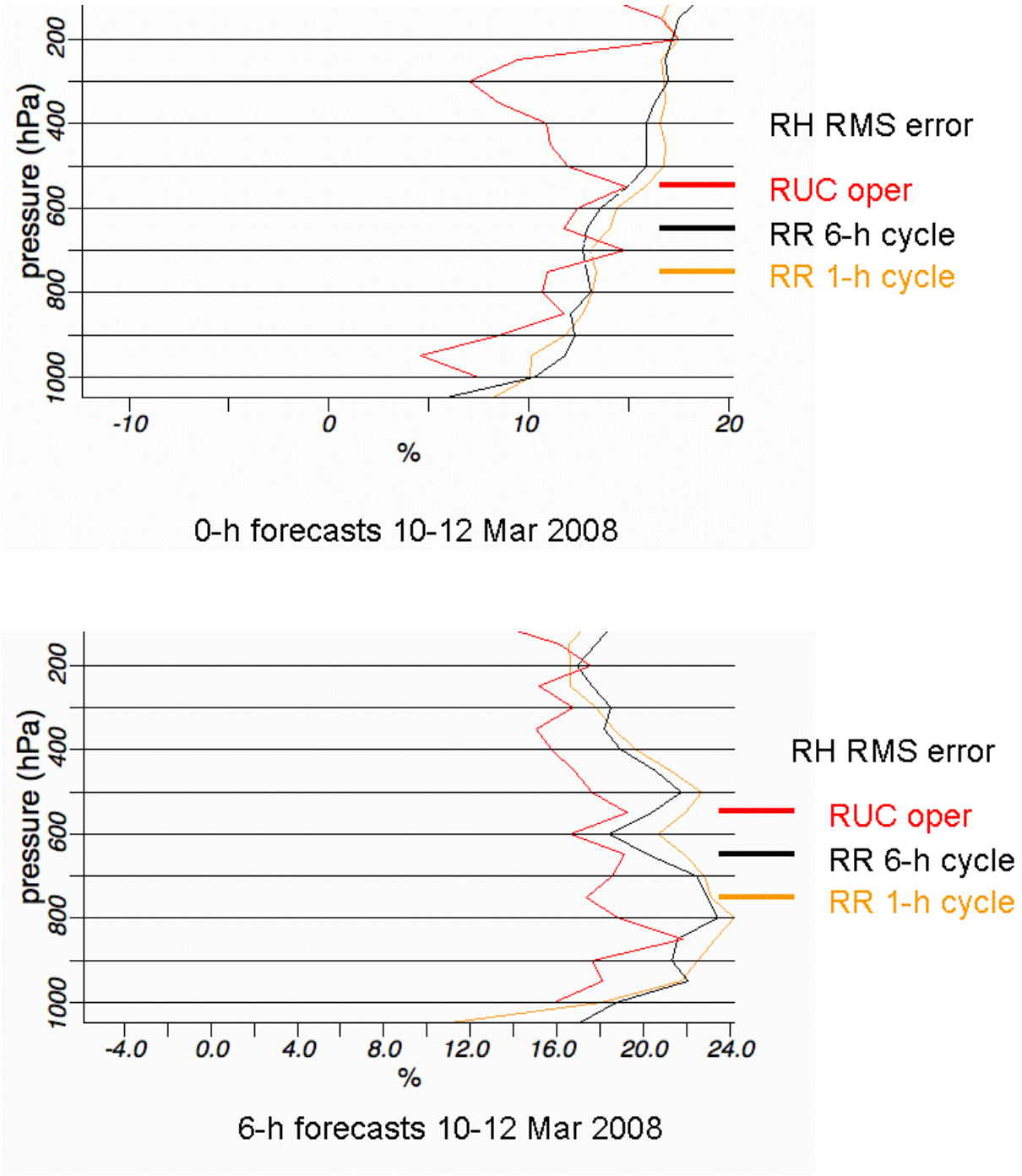

Fig. 3 Vertical profile of relative humidity RMS errors (computed from US rawinsonde network) for period 10-12 March 2008 for (top panel) analysis and (bottom panel) 6-h forecast. The red curve is for the NCEP operational RUC, black curve is for the RR 6-h cycle and arrange curve is for the 1-h RR cycle.

5. Summary

This document provides a first look at some extremely preliminary upper-air verification statistics from the Rapid Refresh cycles. Based on this first look, we see two encouraging results: 1) the current (non-optimized) RR system provides a credible forecast, with errors that are reasonable close to the operational RUC system that it will replace, and 2) the 1-h cycle has smaller 6-h forecasts errors than the 6-h cycle. This second result is especially encouraging and suggests that a high frequency update cycle can continue to yield short range forecast improvement for the RR as it has for the RUC.

Again, we emphasize that these results are extremely preliminary. **Please check back to the link listed on the front page, as this document will be updated on a periodic basis as further verification and model and analysis refinement work is completed.**

5. Acknowledgements.

Thanks to Tanya Smirnova for running the cold start RR runs; Chris Harrop for creating and assisting with application of the workflow manager software; Jacques Middlecoff, Leslie Hart, and others of the JET management team for assistance with issues related to running GSI on JET; and Russ Treadon, Wan Shu Wu, Dave Parrish, Manual Pondeca, and James Purser of NCEP EMC with GSI assistance, and other NCEP personnel who have contributed to verification software.