Rapid Refresh Pre-Implementation Performance Assessment
15 March 2010

1. Introduction

The Rapid Refresh (RR) mesoscale short-range forecast weather prediction system (http://rapidrefresh.noaa.gov) is in final testing for a planned operational implementation at the National Centers for Environmental Prediction (NCEP) Environmental Modeling Center (EMC) in late 2010. In this EMC implementation, the RR will replace the Rapid Update Cycle (RUC), which currently runs operationally at NCEP. The EMC RUC has an hourly cycled configuration generating a fresh 18-h forecast each hour over a domain covering the Continental U.S. A pre-implementation version of the Rapid Refresh system is currently running in real-time at ESRL, in a partial hourly cycling configuration (see sect. 2 for details).

In this report, we summarize recent statistical verification comparisons between this pre-implementation version of the RR and current operational RUC. Analysis metrics covered include: 1) upper-air verification (wind, temperature, moisture), 2) surface verification (wind, temperature, dewpoint), and 3) precipitation verification. Future skill assessments will also include ceiling verification and precipitable water verification. Unless otherwise indicated, the error statistics have been aggregated over a period from 30 Jan 2010 through 11 March 2010 (a 40-day verification period).

2. Upper-air verification

In this section, we present upper-air forecast verification statistics computed against the standard radiosonde network within the RUC domain. Note that scores are computed from native level data as opposed to pressure level data, which can reveal more detail in the vertical structure (Moninger et al. 2010, Section 4.1), but generally leads to slight larger error values. Fig. 1 shows comparisons of the RUC vs. RR **vector wind** RMS errors for 3-h and 12-h lead time. As can be seen, the RR skill exceeds that of the RUC (i.e., smaller forecast error) at all levels for both verification periods. Note the changes in scale for the errors between the 3-h and 12-h forecast comparisons (smaller range shown for the 12-h forecast comparison). Fig. 2 shows a similar comparison for the **temperature** RMS errors. Again, note the change in scale between the 3-h and 12-h forecasts comparisons (much smaller range shown for the 12-h forecast comparison. For the 3-h forecast comparison, the RR temperature errors are smaller at virtually all level, with a significant improvement at lower levels. At 12 h, the errors are similar, with a small edge for the RUC, except at low-levels where the RR is again better. Fig. 3 shows the same 3-h and 12-h RR vs. RUC comparison for **relative humidity**. For the 3-h forecast time, the RR is more accurate than the RUC at for levels up to around 650 hPa, then slightly worse from 650 – 500 hPa. At 12-h, the RUC is somewhat better at most levels, except around 800 hPa. We are currently examining a number of possible factors in this difference for the 12-h forecast. Examination of relative humidity bias statistics (not shown) indicates that much of this RMS difference is related to a positive bias in the RR moisture that exceeds that in the RUC.

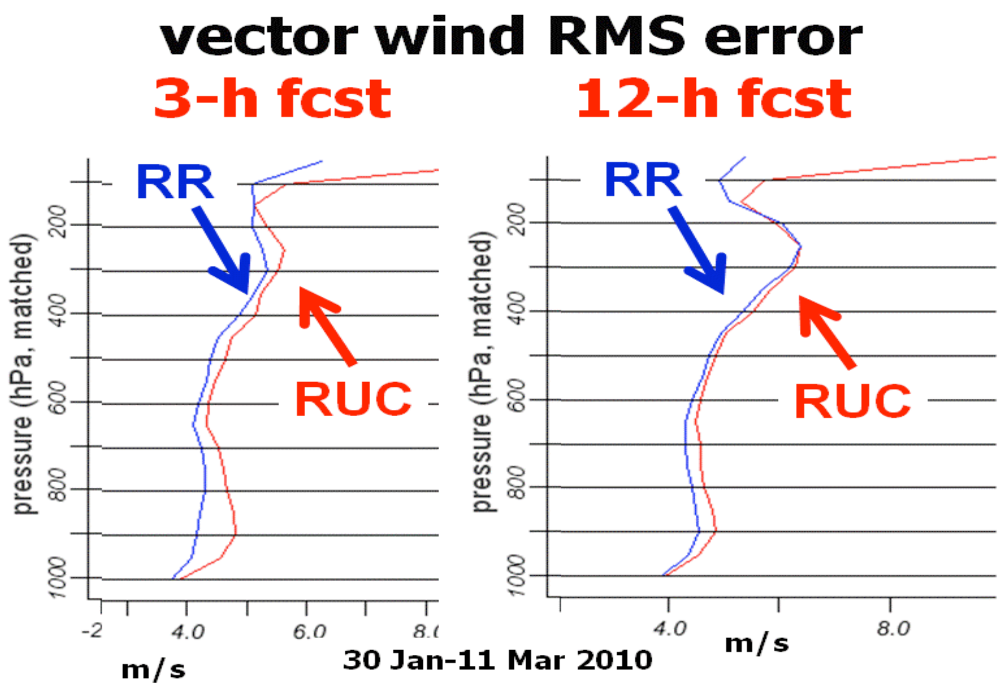

Fig. 1 *3-h and 12-h vector wind RMS errors (m/s) for Rapid Refresh vs. RUC for period 30 Jan 2010 through 11 March 2010.*

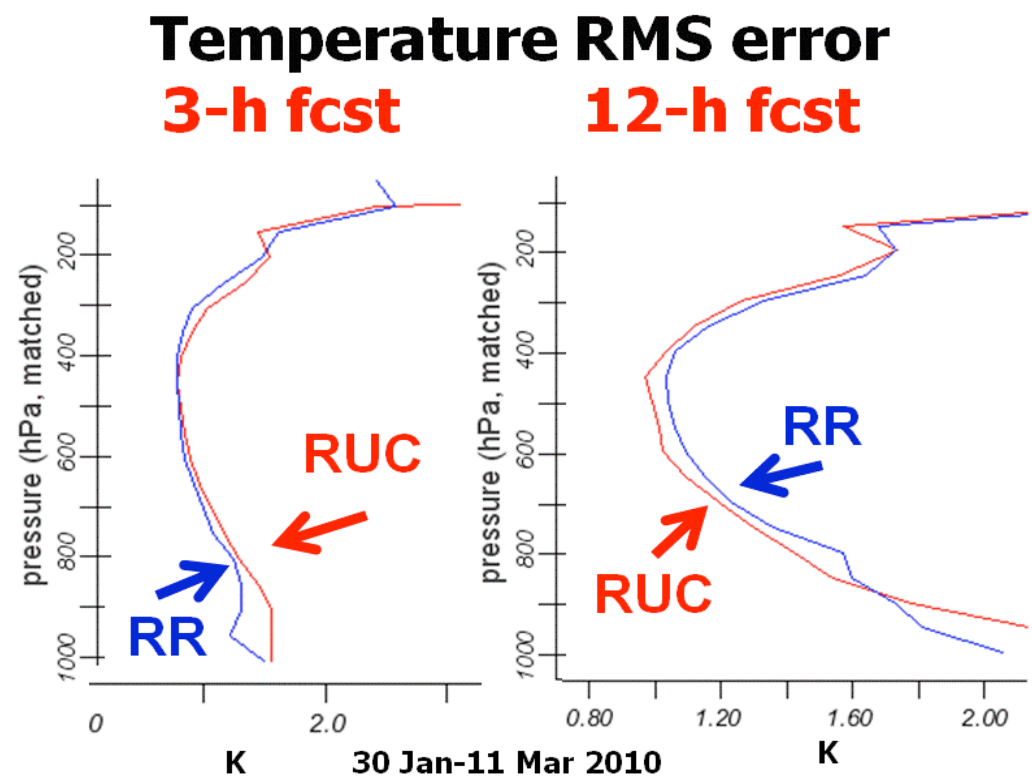

Fig. 2 *3-h and 12-h temperature RMS errors (K) for Rapid Refresh vs. RUC for period 30 Jan 2010 trough 11 March 2010.*

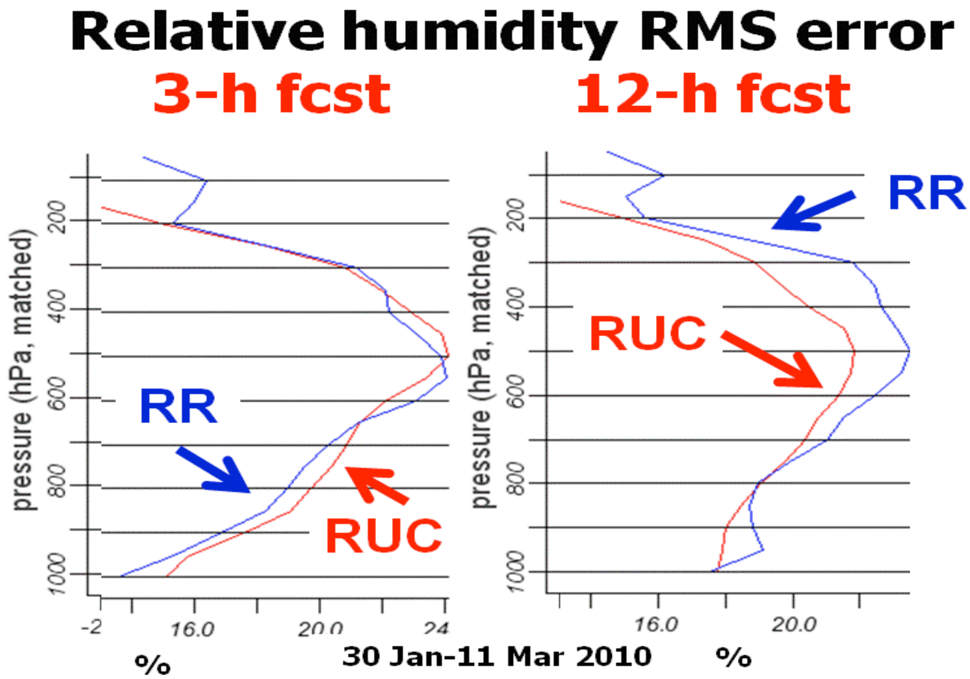

Fig. 3 *3-h and 12-h relative humidity RMS errors (%) for Rapid Refresh vs. RUC for period 30 Jan 2010 through 11 March 2010.*

These statistics are influenced by the partial cycling configuration, in which GFS forecast fields are introduced two times per day (9z and 21z) with a 6-h spin-up at each time. At these two times, these additional 6-h-long "spin-up" RR cycles are initiated from GFS global fields starting with GFS 3-h fields valid at 03z and 15z respectively, but retaining the ongoing hourly cycled RR land surface fields. Then, a series of cycled 1-h RR forecasts are run through the 8z and 20z 1-h forecasts, respectively. These 1-h forecasts from 8z and 20z, respectively, are then used as background for the 9z and 21z RR analyses. The procedure prevents the cycled RR from drifting away from the GFS, since the global data assimilation gives more skill than regional assimilation for larger-scale waves that affect RR accuracy, but allows for small-scale surface and precipitation features to be spun up. It also maintains the important cycling of the detailed land surface model fields (multi-layer soil temperature/moisture, snow water equivalent, multi-layer snow temperature) at 13km resolution. There are also minor differences between the cloud analysis implementation in the RUC vs. RR.

3. Surface verification

In this section, we present surface forecast verification statistics computed against the standard METAR network within the RUC domain. Results shown here are for the portion of the U.S. east of 100 deg. W longitude. Fig. 4 showed the temperature bias for RR and RUC 6-h forecasts. The RR bias is significantly better than the RUC bias, especially during the daytime, when the RUC is too cool. Fig. 5 shows a similar plot for the surface dewpoint. Here, both the RUC and the RR are too moist. The high bias is most pronounced in the RR during the daytime. The cause for the especially high RR dewpoints during the day is under active investigation. Possible causes include too much moisture in RR soil and a post-processing issue in deriving the temperature.

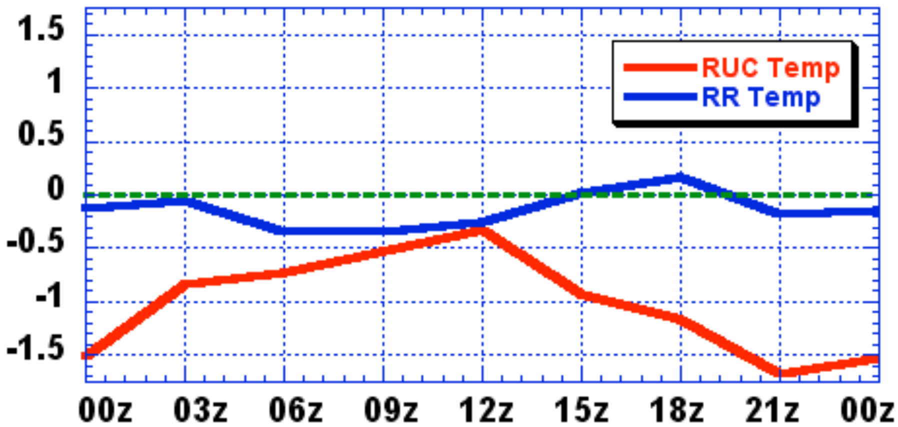

Fig. 4 *6-h surface 2-m temperature bias (K) for Rapid Refresh vs. RUC verified against METAR observations over the RUC domain for period 30 Jan 2010 through 11 March 2010.*

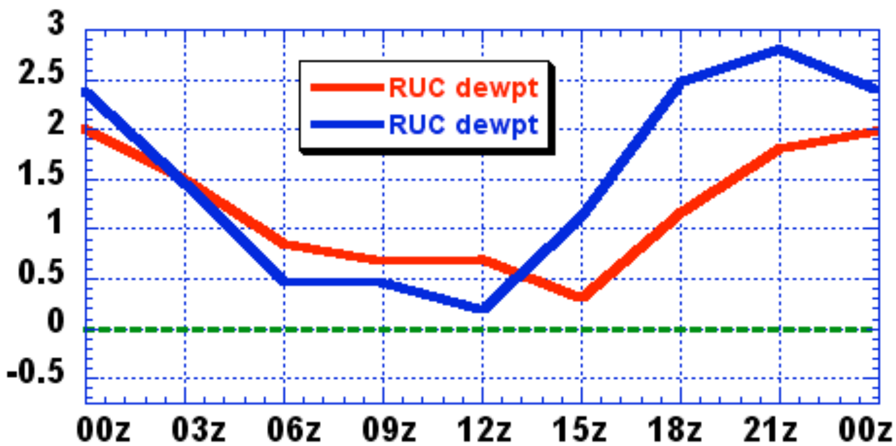

Fig. 5 *6-h surface 2-m dew point bias (K) for Rapid Refresh vs. RUC verified against METAR observations for period 30 Jan 2010 through 11 March 2010.*

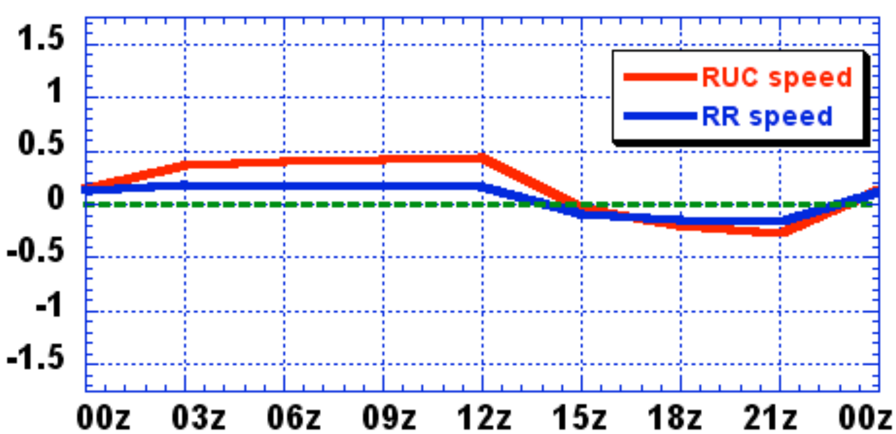

Fig. 6 *6-h surface windspeed bias (m/s) for Rapid Refresh vs. RUC verified against METAR observations for period 30 Jan 2010 through 11 March 2010.*

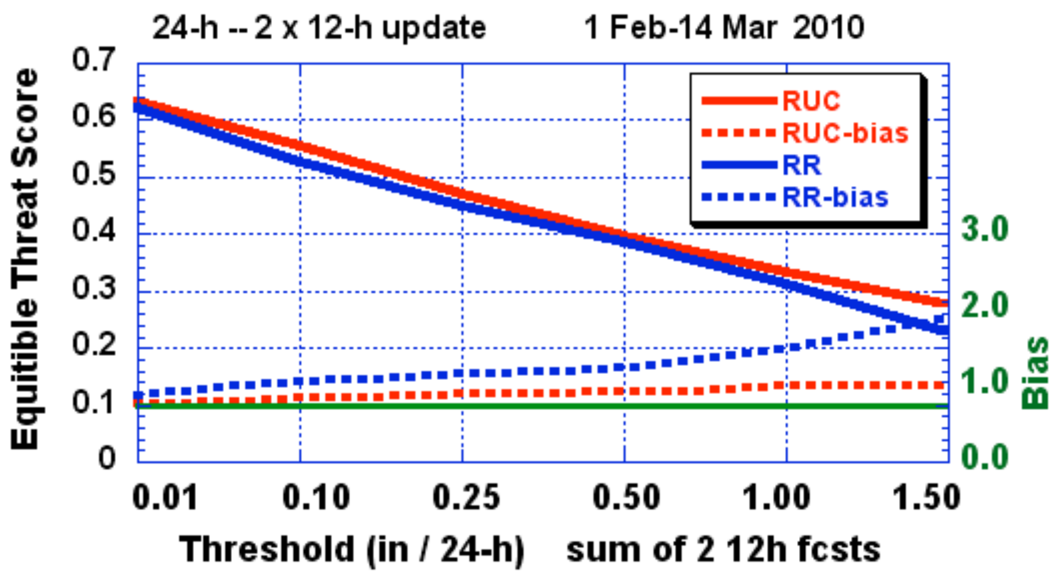

***Fig.** 7 Comparison of 24-h (2 x 12-h forecast) precipitation verification for Rapid Refresh vs. RUC for period 1 Feb 2010 through 14 March 2010. Equitable threat scores are shown with solid lines and bias is shown with a dashed line.*

4. Precipitation verification

In this section, we present precipitation verification statistics computed against the CPC 24-h accumulated rainfall dataset for a winter period (1 Feb – 14 March 2010). 24-h RUC and RR forecasts are obtained by aggregating two successive 12-h runs onto a 40-km common domain. Fig. 7 shows equitable threat scores (EQTS) and biases for several thresholds. Overall, the equitable threat scores indicate similar performance for both models, with a slide edge to the RUC for most thresholds. The RR bias falls off a bit at the highest precipitation thresholds (1.0-1.5 inches / 24h), consistent with the high bias shown for those thresholds. One possible cause for this high bias, is the fact that the radar-DFI in the RR is still being used with an assumed time-scale that longer than that of the RUC.

5. Summary

Taken as a whole, these statistics indicate very good overall performance for the RR, comparable to or better than the RUC for most measures. A few RR deficiencies have been uncovered and are currently being investigated.