

Verification Statistics for the NCEP WRF Pre-Implementation Test. Part 1: Deterministic Verification of Ensemble Members

Ligia R. Bernardet^{1&*}, Louisa Nance², Hui-Ya Chuang³, Andrew Loughe¹, Steven Koch¹

¹NOAA Research – Forecast Systems Laboratory, Boulder, CO

²NCAR – Developmental Testbed Center, Boulder, CO

³National Centers for Environmental Prediction, Camp Springs, MD

I. Introduction

To facilitate a rapid and direct transfer of new numerical weather forecasting research results into the operational process of the National Weather Service (NWS), a joint Developmental Testbed Center (DTC) was established by the National Oceanic and Atmospheric Administration (NOAA) and the National Center for Atmospheric Research (NCAR). The first project of this new institution was designed to support the National Centers for Environmental Prediction's (NCEP's) goal of implementing the Weather Research and Forecasting (WRF) model into operations as an ensemble mesoscale forecast system. With the collaboration of the Air Force Weather Agency (AFWA), a series of tests were conducted using eight different configurations of the WRF model to identify potential members of this ensemble system.

This paper evaluates the high-resolution forecasts computed using two physics packages (NCAR and NCEP) and two dynamical cores: Eulerian Mass Core Non-hydrostatic Mesoscale Model core. The two WRF cores will hereafter be referred to as EM and NMM and the four setups are referred to as EM-NCAR, EM-NCEP, NMM-NCAR and NMM-NCEP). For the setups discussed in this paper, the EM was initialized with the Rapid Update Cycle (RUC) model and the NMM with the ETA model. Details of the model configurations, model domains and the seasons for which the retrospective tests were conducted are described in Seaman et al. (2004).

To evaluate the WRF forecasts, the WRF Post-processor and the WRF Verification System (Chuang et al., 2004) were used. The former interpolates the forecast fields to isobaric surfaces and computes derived fields, such as geopotential height. The latter interpolates the forecast fields to observation location and compares them with surface and upper air observations. The NCEP Precipitation Verification System

(<http://wwwt.emc.ncep.noaa.gov/mmb/ylin/pcpverif/>)

scores/docs/verfdoc.htm) was used to compute the Bias and Equitable Threat Scores of the forecasts.

II. Results

The temperature bias and root mean square error (RMSE) for the Eastern Domain in the fall are shown in Figure 1. The curves in these plots represent averages over both forecast hour (17 forecast times - 0 to 48 h, every 3 h) and forecast cycle (1 every 24 h) for a month long time period. The temperature biases for the various domains and seasons considered in this study exhibit a number of common features. The EM bias is always to the left of the NMM bias. The forecasts generally exhibit cold biases at low levels and warm biases aloft. At upper levels (e.g., 150 hPa), the EM curves cluster together, while the NMM curves cluster at a different value, indicating that dynamical core and/or initial condition (DC/IC) differences are the main factor influencing the solution. On the other hand, the grouping of verification statistics suggests the physics packages have a strong influence on the solution at lower levels. For instance, the NCAR physics configurations have a relative maximum at 700 hPa, while the NCEP physics configurations display a decrease with respect to the values at 500 hPa. The ETA model, which has a physics package very similar to the NMM, has a relative minimum at 700 hPa. The grouping of temperature bias by physics packages, which occurs in all seasons and domains, is possibly due to the behavior of the different planetary boundary layer schemes used in each physics package.

The average temperature RMSE also exhibited features common to all seasons and domain. The ETA model generally exhibited the lowest RMSE. All models had the largest errors in the upper troposphere and at lower levels, with a minimum at midlevels (500 or 400 hPa).

*Contract with Systems Research Group, Inc. (SRG), Colorado Springs, Colorado

& Corresponding author address: Dr. , NOAA FSL, 325 Broadway R/FS1, Boulder, CO, 80305, email: ligia.bernardet@noaa.gov.

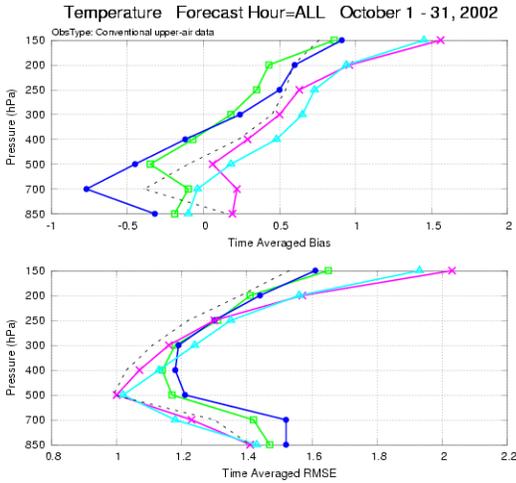


Figure 1. Vertical profile of temperature a) bias and b) RMSE averaged over all forecast times for the Eastern Domain in fall. EM- NCAR (square); EM-NCEP (circle); NMM-NCEP (triangle); NMM-NCAR (X); ETA (dashed).

The average relative humidity (RH) bias and RMSE for the Western Domain in summer are shown in Figure 2. In contrast to the temperature errors, the RH errors vary significantly by season and domain. In the example shown, the WRF models have biases near zero at 850 hPa, whereas the ETA model has a bias of -2.6% . While the bias for the EM models is positive at all levels, the bias for the NMM models reaches -1.7% at 500 hPa. All models have positive humidity biases at upper levels. The errors are very similar among all WRF models at 850 hPa, whereas larger differences appear aloft. At 700 and 500 hPa, the solutions cluster by DC/IC (the EM curves are close together and the NMM curves are close together). In contrast, the solutions cluster by physics package at upper levels (300 hPa): the NCAR physics package produces biases around 2.1%, while the NCEP physics package yields biases around 5.0%. Clustering by physics package at upper levels is a result common to most seasons and domains. It is possible the different cumulus parameterizations in the physics packages dictate the vertical transport of moisture and influence the moisture errors. The different radiation parameterizations in the packages may also lead to this result through different interactions with the cloud systems.

The average bias and RMSE for the vector winds for the Central Domain in the Spring are shown in Figures 3a and 3b, respectively. Except for the 150-hPa level, where the bias is positive, all models underpredicted the winds. The largest negative bias

occurred at 500 hPa for all models. The largest RMSE occurred at jet level, and no obvious grouping of the solutions is observed. This result is representative of all domains and seasons, except for winter (not shown), when there is a clear grouping of solutions by DC/IC, with the NMM producing biases close to zero and the EM underpredicting the winds by about 0.9 ms^{-1} on average.

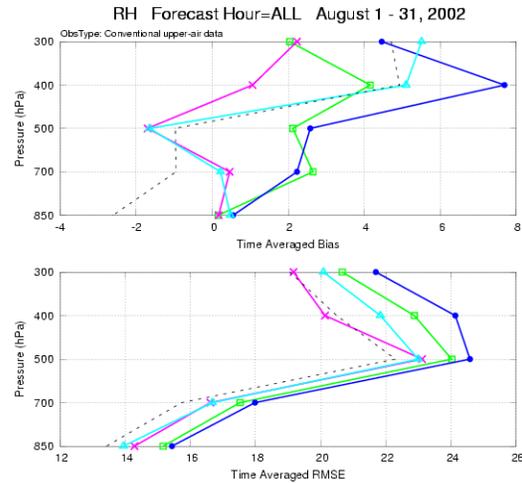


Figure 2. Vertical profile of relative humidity a) bias and b) RMSE averaged over all forecast times for the Western Domain in summer. EM- NCAR (square); EM-NCEP (circle); NMM-NCEP (triangle); NMM-NCAR (X); ETA (dashed).

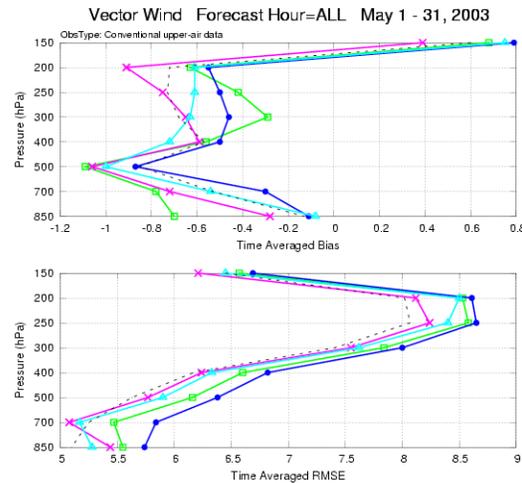


Figure 3. Vertical profile of vector winds a) bias and b) RMSE averaged over all forecast times for the Central Domain in spring. EM- NCAR (square); EM-NCEP (circle); NMM-NCEP (triangle); NMM-NCAR (X); ETA (dashed).

The sea level pressure (SLP) bias for the Central Domain in Spring is shown in Figure 4a. Most models underpredicted the SLP in general. In addition, the SLP bias has a diurnal cycle with the largest underpredictions at the time of maximum temperature and biases close to zero or positive at the time of minimum temperature. The ETA model has the flattest diurnal cycle, with biases closer to zero, while the EM model has the most negative biases, especially in the first 24 hours of integration.

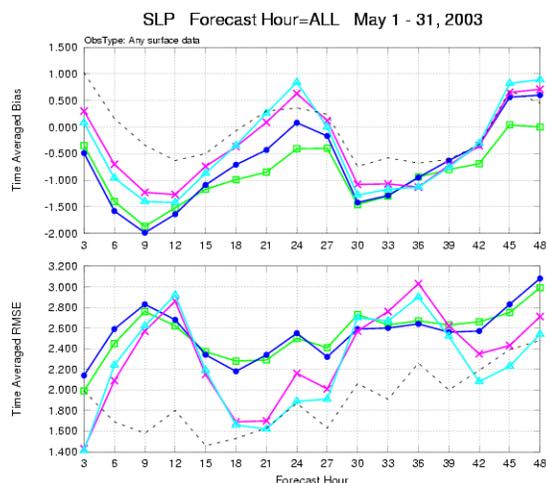


Figure 4. Sea level pressure a) bias and b) RMSE from forecast hour 3 through 48 for the Central Domain in spring. EM- NCAR (square); EM-NCEP (circle); NMM-NCEP (triangle); NMM-NCAR (X); ETA (dashed).

For the SLP forecast, the solutions group by DC/ICs in the first day of forecast, and have no clear grouping in the second day. The SLP RMSE is shown in Figure 4b. The ETA model has the smallest errors and, like the EM model, no evidence of diurnal cycle. The NMM model has the largest errors at sunset. These results are representative of the Central and Western domains for all seasons

The Eastern Domain, however, exhibits a distinct pattern. Figure 5 shows the SLP bias (5a) and RMSE (5b) for the fall in the Eastern Domain. The diurnal cycle is much less evident than for the other domains, and the clustering of solutions by DC/ICs is remarkable. The larger diurnal effect seen in the Western Domain is understandable for the warmer months, when sensible heating over elevated terrain is appreciable. Figure 5a indicates that, while most models still underpredict SLP at all times, the EM has an almost constant bias of about -2.0 hPa. This

negative bias leads to the relatively high RMSE values in Figure 4b.

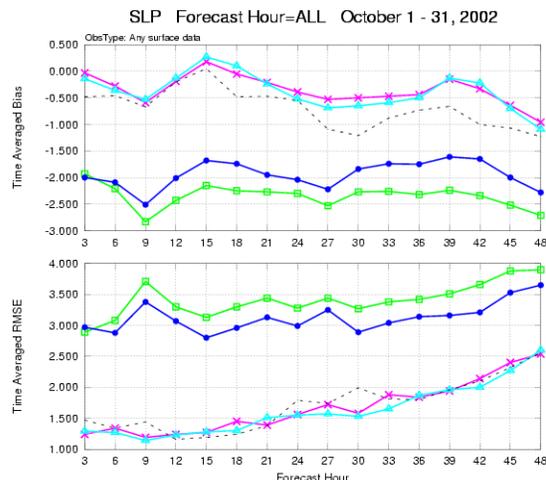


Figure 5. Sea level pressure a) bias and b) RMSE from forecast hour 3 through 48 for the Eastern Domain in fall. EM- NCAR (square); EM-NCEP (circle); NMM-NCEP (triangle); NMM-NCAR (X); ETA (dashed).

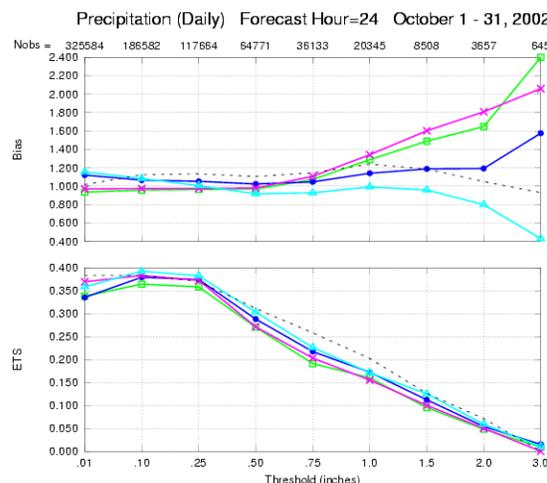


Figure 6. A) bias and b) Equitable Threat Score for 24-hour accumulation for the Eastern Domain in fall. EM- NCAR (square); EM-NCEP (circle); NMM-NCEP (triangle); NMM-NCAR (X); ETA (dashed).

The verification of the 24-hour accumulated precipitation valid at 12 UTC (corresponding to the 42-hr forecast) is shown in Figure 6 for the fall in the Eastern Domain. The bias (Figure 6a) shows that all models produce roughly the correct area of precipitation for the lower thresholds (up to 0.75 in); however, above that threshold, all models, except the

ETA and the NMM-NCEP, overestimate the precipitation area. The largest overestimations occur with the NCAR physics package: the EM-NCAR reaches a bias of 2.4 at the 3.0-in threshold and the EM-NCEP reaches 1.6 for that same threshold. The underestimation of the ETA model is 0.9 at the 3.0-in threshold, while the NMM-NCEP reaches 0.4 for that same threshold. Since both physics packages use the same microphysics parameterization, the difference can be attributed to the convective parameterization. The Equitable Threat Score (Figure 6b) is very similar among all models, with values around 0.37 for the 0.01, 0.10 and 0.25 inch thresholds and lower values for higher thresholds.

III. Discussion and conclusions

A sample of the verification statistics for the retrospective runs performed in support of the NCEP WRF Pre-Implementation Test Plan has been presented. It was shown that the pattern of the errors varies substantially depending on the variable, and is sometimes also sensitive to location or season. For example, while the profile of temperature RMSE errors displays a minimum at midlevels, the RH RMSE is maximum at those levels.

The goal of the Test Plan was not to establish which model performed the best, but to ascertain whether all models were performing sufficiently well that they could be considered qualified for the composition of the ensemble system. The results show that the errors are of the same order of magnitude for all models. Furthermore, the results show that the introduction of different DC/ICs and physics packages effectively created variability among the potential ensemble members. For some variables, seasons, domains and levels the errors cluster by DC/IC, indicating that the main factor that differentiates the solutions is DC/IC. Examples of that clustering are the temperature bias at upper levels and the RH bias at midlevels. However, in other cases, the same DC/IC produced dissimilar results; for example, RH bias at upper levels and precipitation bias at high thresholds. This conclusion, also reached by Stensrud et. al (2000), indicates that the use of multiple physics packages is a valid way of introducing diversity in the ensemble system. A companion paper (DiMego et al., 2004) discusses aspects of the NCEP WRF ensemble composition using Talagrand diagrams and deterministic verification of the ensemble mean.

Acknowledgements. The results shown here are the product of the work of a large DTC Team. The Test Plan was designed and supervised by NOAA-NWS's Nelson Seaman, in cooperation with Steve Koch, Robert Gall, Geoff DiMego, and Jordan Powers. Robert Gall is the new Director of the DTC. At NCEP, Zavisla Janjic, Thomas Black and Matthew Pyle provided and supported the NMM code, Keith Brill and Ying Lin developed the verification code and Geoff DiMego supervised the group. At NCAR, Jordan Powers, Jimmy Dudhia, David Gill, William Skamarock and Chris Davis provided and supported the EM code. At NOAA FSL, Brent Shaw supported the WRF initialization and Jacques Middlecoff and Christopher Harrop assisted with code porting and automation. Finally, part of the retrospective runs were performed at the Naval Oceanographic Office computers by AFWA's Daniel Lohaus, Frank Olson and Mark Noe under Jerry Wegiel's supervision.

IV. References

- Chuang, H., G. DiMego, M. Baldwin, and WRF DTC team, 2004. The NCEP WRF Post-Processing Software. *WRF/MM5 Joint Workshop, Boulder, CO.*
- DiMego, G., M. Tsidulko, K. Brill, H. Chuang, L. Bernardet, D. Lohaus, L. Nance, C. Davis, S. Gopalakrishnan and DTC Team, 2004. Verification Statistics for the NCEP WRF Pre-Implementation Test. Part 2: Ensemble Results. *WRF/MM5 Joint Workshop, Boulder, CO.*
- Seaman, N., R. Gall, L. Nance, S. Koch, , G. DiMego, Z. Janjic, J. Powers, W. Skamarock, and F. Olson. 2004. The WRF Process: Streamlining the Transition of New Science from Research into Operations. *WRF/MM5 Joint Workshop, Boulder, CO.*
- Stensrud, D. J., J. Bao and T. Warner, 2000. Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077-2107.