7.4

# VERIFICATION OF THE FSL ENSEMBLE OF MESOSCALE MODELS USED FOR A WINTER WEATHER APPLICATION

Paul Schultz[1]
NOAA Research - Forecast Systems Laboratory
Boulder, Colorado

## 1. INTRODUCTION

The LAPS group at FSL has built an ensemble of mesoscale models that runs in real time in support of field projects and demonstrations. One of these projects is sponsored by the Federal Highways Administration (FHWA) and is focused on winter weather. A companion paper (Schultz 2004) discusses the ensemble design as it was deployed during this demonstration last winter.

The FHWA Maintenance Decision Support System (MDSS, Mahoney 2001) is an effort to tailor weather forecasts for the purposes of winter road maintenance. FSL generates the mesoscale model forecasts (Schultz 2002) and transmits them to the NCAR/Research Applications Program, where they are used to make point forecasts along roadways. These point forecasts feed pavement temperature and chemical dilution algorithms (developed by the Cold Regions Research and Engineering Laboratory), which are used along with codified rules of practice (developed by MIT/Lincoln Labs) to automatically recommend timing and location for snow plowing and chemical applications.

Last winter, the MDSS ensemble consisted of six members: three mesoscale models (MM5, RAMS, and WRF) with two larger-scale models (NCEP's Eta and AVN) providing lateral boundary conditions. The models were run out to 27 h to provide a 24-h forecast service. The grid configuration, centered on the state of Iowa (Fig. 1), is the same for all models. For this test the grid increment was 12 km, with no convective parameterizations because of the focus on winter weather. The execution schedule was driven by the update frequency of the NCEP models; thus, all six members were run four times per day, upon receipt of the NCEP model grids. All model runs were initialized with the same LAPS "hot start" diabatic initialization grids (Shaw et al. 2001; Schultz and Albers 2001).

The MDSS model forecasts are evaluated by comparisons against observations of surface temperature, wind, dewpoint, and precipitation. Although the observations are taken hourly, for the 2003 demo the model outputs were provided in 3-h increments. Hourly precipitation observations were binned for comparison to 3-h model precipitation accumulations.

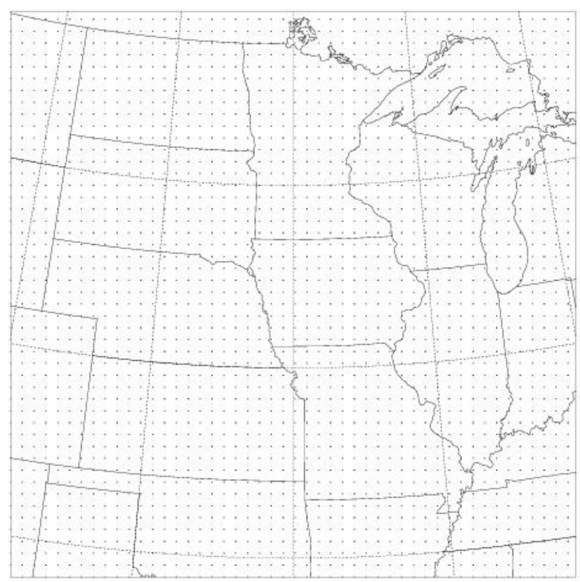

**Figure 1. MDSS modeling domain. Every third grid point is shown. The grid dimensions are 144×144.**

The object of verifying these forecasts is to determine if predictors (model forecasts) are adding quality to the forecast service. The concept behind ensemble modeling is that a properly combined group of forecasts can provide a better forecast than any single predictor in the ensemble. However, for

[1] *Corresponding author address:* Paul Schultz, NOAA/Forecast Systems Laboratory, R/FS1, 325 Broadway, Boulder, CO 80305; *e-mail* paul.j.schultz@noaa.gov

this advantage to be consistently realized, each model forecast in the ensemble should be equally likely to be the most accurate predictor for a given forecast event. Furthermore, the ensemble members should be as different as possible, so that errors among the different models are uncorrelated. In other words, if the models are all making the same kinds of errors for the same reasons, the ensemble forecast will not be optimum. Thus, verification efforts are also aimed at determining progress toward the goal of good *dispersion* among the ensemble members.

## 2. STATE VARIABLES

6-h forecast verification statistics (rms and bias) for the state variables temperature, wind speed, and dewpoint are given in the following table:

| | Temperature (K) | | Wind speed (m/s) | | Dewpoint (K) | |
|---|---|---|---|---|---|---|
| MM5-AVN | 3.1 | -0.7 | 2.5 | +0.8 | 5.6 | +1.5 |
| MM5-Eta | 3.0 | -0.5 | 2.5 | +0.8 | 5.5 | +1.6 |
| RAMS-AVN | 5.8 | -1.1 | 2.6 | +1.6 | 6.5 | -0.9 |
| RAMS-Eta | 5.9 | -1.1 | 2.6 | +1.7 | 6.9 | -1.0 |
| WRF-AVN | 3.1 | -0.4 | 2.4 | +1.1 | 5.7 | +1.4 |
| WRF-Eta | 3.1 | -0.4 | 2.4 | +1.0 | 5.7 | +1.3 |

Several trends are apparent. All the model configurations produced forecasts with a positive wind speed bias and very similar rms errors. The RAMS model configurations have significantly larger errors in temperature and, to a lesser extent, dew point temperatures. All the models made too-cold 6-hr forecasts.

The deviations between configurations that are different only in the lateral boundary models are very small compared to deviations among configurations with different mesoscale models. Further graphical evidence of this will be shown at the conference. This conclusion is somewhat disappointing from the perspective of building a robust ensemble, because it is likely that such model pairs are making the same errors and thus add little dispersion to the ensemble.

Direct, formal statistical comparisons with the NCEP models was not conducted for the 2003 MDSS demo (although this is planned for the 2004 demo). However, some comparisons have been performed, and graphics from these will be presented at the conference. Generally, the NCEP models are quite accurate for temperature forecasts: about 2K rms vs about 3K rms for the MDSS models. The NCEP models are also more accurate than the mesoscale models in humidity forecasts (but not wind). This comes as no surprise; the surface fields were not expected to be an area in which the regional models would add much value. This is because the NCEP models have surface flux formulations that are closely tied and tuned to their respective Land Surface Models (LSMs), which characterize the earth surface by vegetation type, roughness, albedo, moisture content, land use, etc. By contrast, the regional models inherit surface information from the NCEP models and use different surface flux formulations. We point this out because the LSM coupling with the WRF model is seen by the WRF modeling community as an opportunity for significant progress in forecasting surface state variables in regional weather modeling in the next few years.

## 3. PRECIPITATION

The following figures show 3-hr precipitation verification. The statistics are nearly identical for model configuration pairs that differ only in their lateral bounds models and are not shown separately.

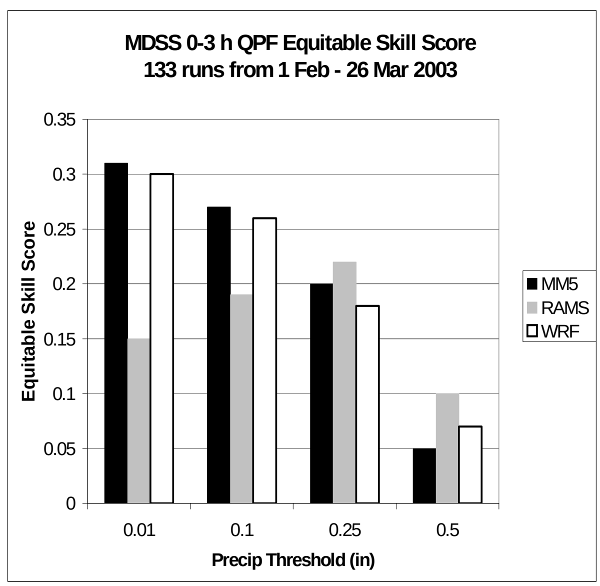


**Figure 2. Equitable skill scores for precipitation forecasts from the MDSS ensemble members.**

In Figure 2 it is clear that the MM5 and WRF configurations make similarly skillful precipitation forecasts, and much better than the RAMS forecasts for light amounts, whereas the RAMS forecasts appear to be somewhat better for larger amounts. The RAMS model has a very large overforecasting bias (too much precipitation predicted; Fig. 3).

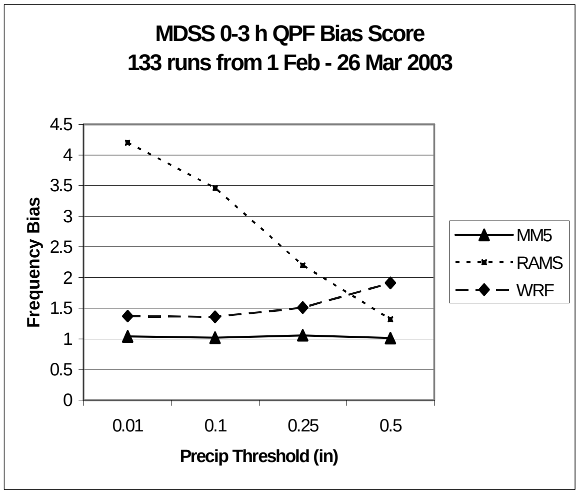


**Figure 3. Bias statistics for 3-h precipitation forecasts from the MDSS ensemble members.**

Regular visitors to the model output display web site found this obvious in virtually every model run. While the WRF model has a moderate bias problem, the MM5 bias values are nearly perfect (1.0 at all levels). We have reason to believe that microphysics parameterizations in the WRF and RAMS models can be optimized to improve their bias problems.

## 4. CONCLUSION

Following statistical evaluation of the models' performance during the 2003 Demo, we have begun experiments with alternate configurations of the ensemble modeling system. The pertinent lessons learned were 1) using two different models (AVN and Eta) for lateral boundary conditions did not provide much diversity, 2) the models did not provide much added value beyond 18 hours, 3) the RAMS model routinely had large errors in precipitation and temperature, and 4) the WRF model generated too much precipitation.

In light of this experience, we have developed an alternative strategy (Schultz 2004) to take better advantage of what these models do best, which is exploit more of the available observations (particularly radar and satellite) to improve precipitation forecasts in the range of 1-12 hours. This configuration consists of running MM5 and an improved version of WRF every hour, and using "time-lagged" ensembling techniques (e.g., Brundage et al. 2001). For example, a 6-hour ensemble forecast uses the current 6-hr forecast, the previous 7-hour forecast, and the 8-hour forecast from the cycle before that, all forecasts valid at the same time. It is expected that such practice will reduce the cycle-to-cycle "shock" in the MDSS forecast services that was sometimes caused during the 2003 demo when the models updated.

Other changes to the modeling system include extending the verification system to allow for direct comparisons between the NCEP model services and those provided by the ensemble.

For the 2004 MDSS demo, which runs from January through March, the ensemble model runs are posted in real time at:

`http://laps.fsl.noaa.gov/mdss`

*Acknowledgement.* The precipitation forecasts were verified by the Real-Time Verification System (Mahoney et al. 2002). Steve Weygandt reviewed and improved this paper. This work is sponsored by the Federal Highways Administration's Office of Transportation Operations under the Road Weather Management Program.

## 5. REFERENCES

Brundage, K.J., S.G. Benjamin, and M.N. Schwartz, 2001: Wind energy forecasts and ensemble uncertainty from the RUC. Preprints, *9th Conf. on Mesoscale Processes,* Fort Lauderdale, Amer. Meteor. Soc., 343-345.

Mahoney, J. L., J. K. Henderson, B. G. Brown, J. E. Hart, A. Loughe, C. Fischer, and B. Sigren, 2002: The Real-Time Verification System (RTVS) and its application to aviation weather forecasts. *10th Conference on Aviation, Range, and Aerospace Meteorology*, Portland, OR., Amer. Meteor. Soc., 323-326.

Mahoney, W., 2001: An advanced winter road maintenance decision support system. Preprints, *11th Annual ITS America Meeting*, Miami, Amer. Meteor. Soc.

Schultz, P., 2004: The MDSS ensemble of mesoscale models: Lessons learned from the 2003 Demo and changes for the 2004 Demo. Preprints, *20th Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Seattle, Amer. Meteor. Soc.

Schultz, P., 2002: Using ensembles of numerical weather forecasts for road weather prediction. Preprints, *18th Int'l Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Orlando, Amer. Meteor. Soc., 32-34.

Schultz, P. and S. Albers, 2001: The use of three-dimensional analysis of cloud attributes for diabatic initialization of mesoscale models. Preprints, *14th Conf. On Numerical Weather Prediction*, Ft. Lauderdale, Amer. Meteor. Soc., J122-J124.

Shaw, B.L., J.A. McGinley, and P. Schultz, 2001: Explicit initialization of clouds and precipitation in mesoscale forecast models. Preprints, *14th Conf. On Numerical Weather Prediction*, Ft. Lauderdale, Amer. Meteor. Soc., J87-J91.