Evaluating the Impact of RAFS Changes on the NGM-Based MOS Guidance

MARY C. ERICKSON, J. BRENT BOWER, VALERY J. DAGOSTARO, J. PAUL DALLAVALLE,
ELI JACKS, JOHN S. JENSENIUS JR., AND JAMES C. SU

Techniques Development Laboratory, Office of Systems Development NWS/NOAA Silver Spring, Maryland

5 October 1990 and 5 November 1990

ABSTRACT

In the spring of 1990, the National Meteorological Center (NMC) tested several modifications to the Regional Analysis and Forecast System (RAFS). In order to compare the proposed version of the RAFS to the current operational RAFS, NMC ran both systems in parallel for a 4-week period. Simultaneously, the Techniques Development Laboratory (TDL) applied the operational RAFS-based Model Output Statistics (MOS) equations to output from both the operational and proposed (parallel) versions of the Nested Grid Model (NGM) to generate two sets of MOS forecasts. Our goal was to determine the impact of RAFS modifications on the NGM MOS forecasts. At the end of the 4-week test period, we verified both the operational and parallel NGM MOS forecasts. Virtually no differences in accuracy or skill existed between the operational and parallel MOS forecasts of max/min temperature, probability of precipitation, and surface wind. The skill of the cloud amount forecasts, however, deteriorated significantly. The NGM 300-mb relative humidity field changed markedly as a result of the RAFS modifications, and this change affected the cloud forecasts. Since the cloud cover forecasts were the only NGM MOS products adversely impacted by the new parallel RAFS, we rederived the cloud equations without the 300-mb relative humidity. These equations were implemented operationally in September 1990. When the new RAFS is implemented, we expect that the impact on the current NGM MOS guidance will be minimal.

1. Introduction

The National Meteorological Center's (NMC) Regional Analysis and Forecast System (RAFS) is currently the primary source of short-range (1 to 2 day) model forecasts for the United States (Hoke et al. 1989). At the Techniques Development Laboratory (TDL), we apply the Model Output Statistics (MOS) technique (Glahn and Lowry 1972) to output from the Nested Grid Model (NGM, the forecast component of the RAFS) to produce statistical forecasts of several surface weather elements (Jacks et al. 1990). NGM MOS forecasts of daytime maximum/nighttime minimum (max/min) temperature, probability of precipitation (PoP), surface wind speed and direction, and opaque cloud amount are generated twice daily and sent to users. In early 1990, NMC proposed implementing numerous changes to the first guess, analysis, and forecast steps of the RAFS (Petersen et al. 1991). These changes included using a regional, rather than global, first guess over the C grid (the innermost computational grid of the NGM); increasing the accuracy of the horizontal finite-differencing in the NGM from second to fourth order; and improving the NGM's lower boundary conditions, including an enhanced terrain field. Since the MOS technique accounts for some of the systematic errors found in a dynamical model, changes that alter these systematic errors can lead to less skillful MOS forecasts. Therefore, we were concerned that the proposed changes could have an adverse impact on the NGM MOS.

2. Evaluation procedure

In order to evaluate the new RAFS and its potential impact on NGM MOS, NMC conducted a "parallel test." NMC ran both operational and proposed (parallel) versions of the RAFS for a 4-week test period from 6 April 1990, to 4 May 1990. At the same time, TDL applied the operational NGM MOS equations to output from both versions of the NGM. Note that these equations were developed from a 2-year sample of the operational NGM. Once the parallel test was complete, we had 27 test cases for 0000 UTC and 26 test cases for 1200 UTC. NGM MOS forecasts were produced for stations in the contiguous United States (Fig. 1), and verified for all stations combined, as well as by region. The regional results showed little deviation from the national scores so, in the interest of brevity, the regional scores are not shown. By combining the forecasts at all stations, we had roughly 5000 cases for each weather element, cycle, and projection. In addition to

---

Corresponding author address: Mary C. Erickson, Techniques Development Laboratory, Office of Systems Development, National Weather Service/NOAA, 5200 Auth Road, Room 706, Camp Springs, MD 20746.

FIG. 1. The 204 stations used in the evaluation of parallel and operational NGM MOS.

the various verification scores, we used paired t-tests to evaluate the significance of the differences in verification scores between the operational and parallel NGM MOS guidance. In Section 3 (Results), we show verifications of the parallel NGM MOS, operational NGM MOS, and, for comparison, the operational MOS forecasts (Carter et al. 1989) based on output from the Limited-area Fine-mesh Model (LFM) (Gerrity 1977; Newell and Deaven 1981). Note that no significance testing was done between the NGM and LFM MOS guidance.

We also evaluated the direct model output, interpolated to the 204 stations, for various fields related to the forecasting of max/min temperature, PoP, surface wind, and cloud amount. For each model output variable of interest, we combined the data for all stations and days and determined the average forecast value for each projection. We also examined the correlation between weather observations and model predictors. As in the MOS verifications, we include the scores from the LFM as a basis for comparison. Since the results for 0000 UTC and 1200 UTC were similar, except for fluctuations due to the diurnal cycle, we show only the 0000 UTC results.

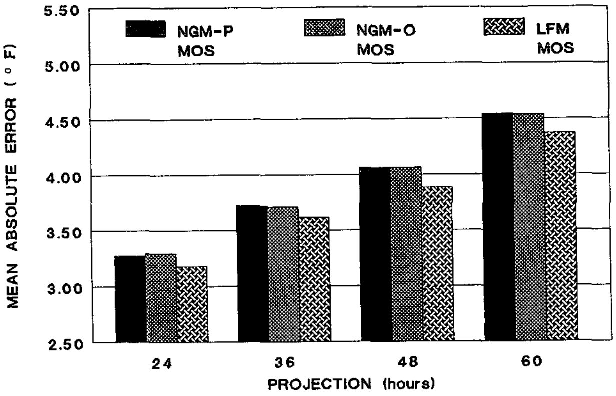

FIG. 2. Verification in terms of mean absolute error (°F) of parallel NGM MOS (NGM-P), operational NGM MOS (NGM-O), and LFM MOS max/min temperature forecasts for the 0000 and 1200 UTC cycles combined.

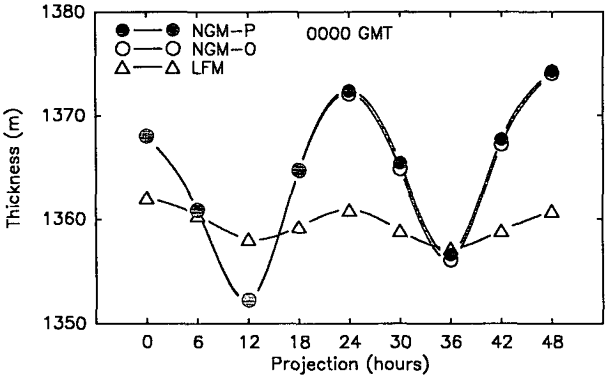

FIG. 3. A comparison of the parallel NGM (NGM-P), operational NGM (NGM-O), and LFM average 1000–850-mb thickness forecasts.

3. Results

a. Daytime max and nighttime min temperature

Using mean absolute error as a measure of accuracy, we verified max/min temperature forecasts valid approximately 24, 36, 48, and 60 h after both 0000 and 1200 UTC. At all eight projections, virtually no difference existed between the operational and parallel NGM MOS guidance. The significance tests confirmed that none of the small differences was statistically significant. The average mean absolute errors for both cycles combined are shown in Fig. 2. Although the LFM MOS provided the most accurate forecasts at every projection in this short sample, 6-month seasonal verifications show that the NGM MOS and LFM MOS guidance are about equally accurate.

One of the most important predictors used to forecast the max/min temperature is the 1000–850-mb thickness. Fig. 3 shows the average thickness for the parallel and operational versions of the NGM, as well as for the LFM. The mean values for both versions of the NGM were nearly identical, while the LFM thicknesses had a much weaker diurnal cycle than the NGM forecasts. The correlations between the observed surface temperature and the forecast 1000–850-mb thickness (not shown) were also essentially equal for the parallel and operational NGM fields.

b. PoP

The MOS PoP forecasts predict the probability of $\geqslant 0.01$ inches (0.25 mm) of liquid equivalent precipitation for 12-h periods ending at 24, 36, 48, and 60 h after 0000 and 1200 UTC. To evaluate the NGM MOS PoP forecasts, we used the improvement over climate in the Brier score (National Weather Service 1982). In Fig. 4, we've combined the results from the 0000 and 1200 UTC cycles for each of the four projections.

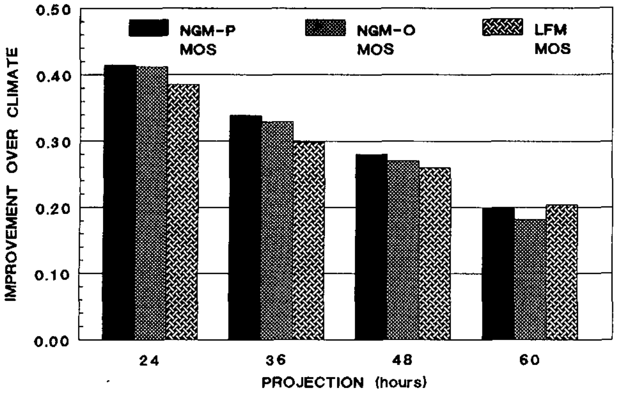

FIG. 4. Same as Fig. 2 except for percent improvement over climate in the Brier score for the PoP forecasts.

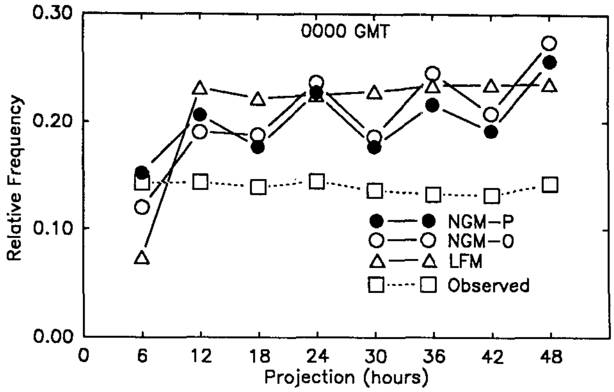

FIG. 6. Same as Fig. 3 except for the observed and forecast relative frequency of ≥0.01 in. (0.25 mm) of precipitation.

No significant differences existed between the operational and parallel NGM MOS guidance. However, both systems were generally more skillful than the LFM MOS forecasts.

In forecasting PoP, both the mean relative humidity (surface to approximately 500 mb) and the model forecast of ≥0.01 (0.25 mm) inches of precipitation are important predictors. Fig. 5 shows the average value of the mean relative humidity for each model during the test period. The values for the parallel and operational versions of the NGM were very similar. Note, however, that the average values for the LFM did not increase with increasing projection as did the NGM values. Fig. 6 shows the forecast relative frequency of precipitation ≥ 0.01 inches (0.25 mm) for each model, as well as the observed relative frequency during the test period. Although the mean values for the parallel and operational versions of the NGM followed the same trend, there were some differences in the values. The relative frequency in the parallel NGM was higher than the frequency in the operational NGM at the 6- and 12-h projections, and lower than the frequency in the operational NGM at the 18-h projection and beyond. The higher relative frequency in the first two 6-h periods of the parallel NGM was consistent with the reduced "spin-up" expected from the new analysis. The lower relative frequency of the parallel NGM at 18-h and beyond was closer to the observed relative frequency than the operational NGM.

c. Surface wind

The NGM MOS wind speed and direction forecasts were verified for eight projections at 6-h intervals from 6 to 48 h after 0000 and 1200 UTC. The skill of the MOS wind speed forecasts was evaluated by separating the forecasts into seven categories of wind speed, and then using the Heidke skill score (National Weather Service 1982). The wind speed categories were: ≤7 kts, 8–12 kts, 13–17 kts, 18–22 kts, 23–27 kts, 28–32 kts, and ≥33 kts. Fig. 7 shows the average skill scores for 0000 and 1200 UTC cycles combined. The results were mixed, but no significant differences between the skill scores of the parallel and operational NGM MOS wind speed forecasts were found. The LFM MOS wind speed scores were comparable to, but generally less skillful

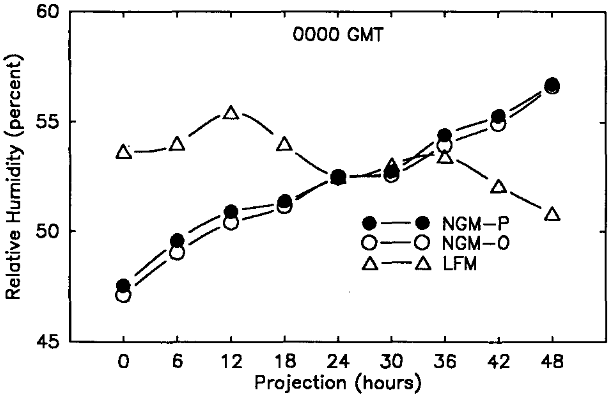

FIG. 5. Same as Fig. 3 except for the mean relative humidity (surface to approximately 500 mb).

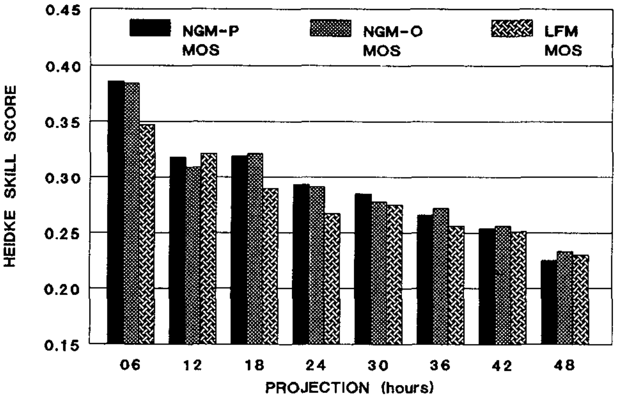

FIG. 7. Same as Fig. 2 except for Heidke skill scores for the wind speed forecasts.

than, the NGM-based MOS scores. The scores for the wind direction forecasts, in terms of the mean absolute error, are shown in Fig. 8. The parallel and operational NGM guidance performed similarly, and no significant differences were noted. In almost all cases, the parallel and operational NGM MOS forecasts were more accurate than the LFM MOS guidance. Note that the numerical model forecasts of wind speed and direction at various atmospheric levels are essential predictors in forecasting the surface wind speed and direction. In our tests, the average wind speeds in the parallel and operational NGM were virtually identical at all levels (not shown).

d. Cloud amount

The MOS cloud forecasts predict the probability of clear (0 tenths), scattered (1–5 tenths), broken (6–9 tenths), and overcast (10 tenths) opaque cloud amount for projections at 6-h intervals from 6 to 48 h after 0000 and 1200 UTC. In addition to these forecasts, we provide categorical cloud amount forecasts by comparing the probabilities to threshold probabilities (Jacks et al. 1990). We evaluated these best category cloud amount forecasts in terms of the Heidke skill score. In Fig. 9, we see that the parallel NGM MOS was consistently less skillful than the operational NGM MOS (and significantly so, in terms of the paired t-test) at the 0000 UTC cycle for the 12- to 36-h projections. At 1200 UTC (Fig. 10), the parallel NGM MOS was less skillful than the operational NGM MOS at almost all projections to 48 h. These differences were statistically significant at the 12-, 18-, and 24-h projections. Differences between the operational NGM MOS and LFM MOS guidance were comparatively small.

4. Discussion of cloud results

These results indicate that implementation of the parallel RAFS would have adversely affected the NGM MOS cloud guidance. But why? The mean relative humidity (RH) is the most important predictor used to forecast cloud amount, yet the differences between the mean relative humidity forecasts from the parallel and operational NGM (Fig. 4) were not large enough to cause this much deterioration in skill. After examining the operational NGM MOS cloud equations, we found that the 300-mb relative humidity was common to most equations. We then suspected that the parallel and operational NGM forecasts of this field were different. Fig. 11 shows that, indeed, the parallel NGM forecasts of the 300-mb RH were more moist than the corresponding operational NGM forecasts. The difference was nearly 20% at the initial time, and decreased to roughly 5% at 48 h. While the mean values in Fig. 11 were only valid for the 4-week test period, we suspect that the 300-mb RH forecasts from the parallel NGM also differed greatly from the NGM 300-mb RH forecasts used in developing the operational NGM MOS equations. Applying these equations to data with characteristics that were quite different from those in the

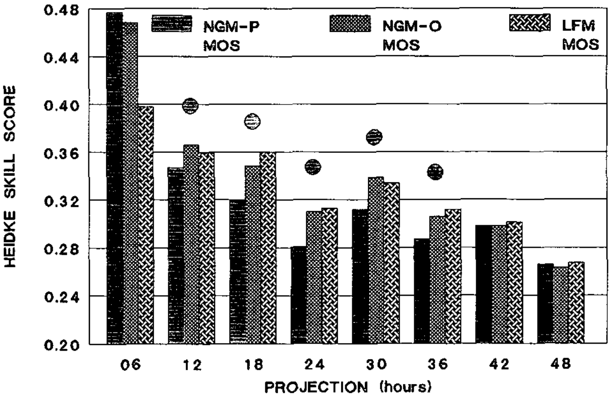

FIG. 9. Heidke skill scores for the parallel NGM MOS (NGM-P), operational NGM MOS (NGM-O), and LFM MOS 0000 UTC best category cloud cover forecasts. The dots indicate statistically significant differences in scores between the operational and parallel NGM MOS forecasts.

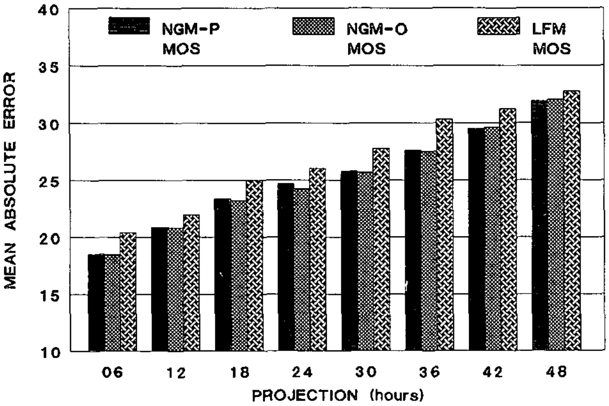

FIG. 8. Same as Fig. 2 except for mean absolute errors (degrees) of wind direction forecasts.

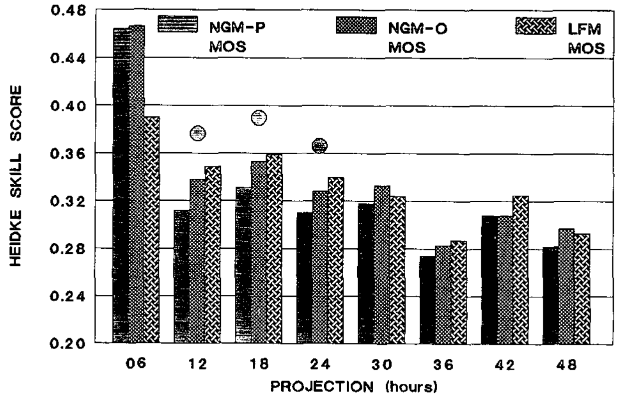

FIG. 10. Same as Fig. 9, except for 1200 UTC.

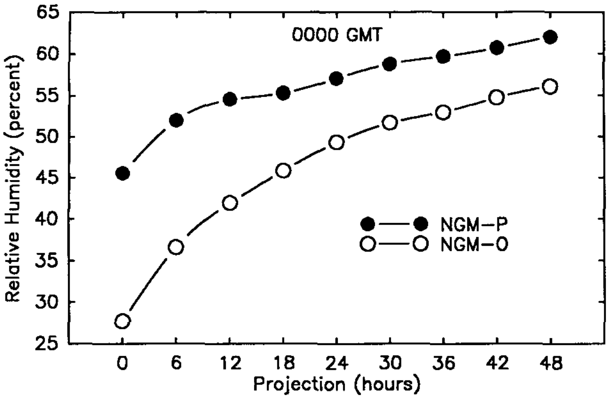

FIG. 11. Same as Fig. 3 except for the 300-mb relative humidity forecasts. LFM forecasts of the 300-mb relative humidity were not evaluated.

developmental sample was probably the cause of the parallel cloud forecasts being less skillful.

The large differences in the 300-mb RH are due to a change in the procedure used to vertically extrapolate moisture to the upper levels of the NGM. (DiMego and Mitchell 1990). In the operational model, the first guess (from the global model) and analysis are used to assign initial moisture values to the tropospheric sigma layers of the model atmosphere. Then a vertical extrapolation procedure is used to obtain the moisture for the remaining model layers above. Because this extrapolation procedure caused an overly rapid decrease in moisture with height in the upper layers of the model, a new extrapolation procedure was used in the parallel NGM (Petersen et al. 1991). The improved extrapolation procedure produces a more gradual decrease in the model moisture above approximately 350 mb, and therefore causes the 300-mb RH to be higher.

Since the cloud cover forecasts were the only NGM MOS products adversely impacted by the parallel RAFS, we redeveloped the cloud equations without the 300-mb RH as a predictor. The redevelopment was based on the same developmental sample as the operational NGM MOS equations. Test results shown in Fig. 12 for the 18-, 24-, and 30-h projections from 0000 UTC indicate that the new equations applied to the parallel NGM were comparable in skill to the operational equations applied to the operational NGM. The new cloud equations were implemented in September 1990 in anticipation of the RAFS changes.

5. Conclusions and comments

On the basis of this one-month sample, the implementation of the new RAFS should have minimal impact on the NGM MOS max/min temperature, PoP, and surface wind forecasts. Without any changes, the cloud forecasts would be significantly less skillful when the new RAFS is implemented. However, since we implemented cloud equations which do not use the 300-mb RH as a predictor, the cloud forecasts should not be noticeably less skillful.

This evaluation highlights an important aspect of the MOS system: equations generated by the MOS technique are strongly dependent on the characteristics of the model data used in the developmental sample. Subsequent changes to the numerical model can have one of several effects on the skill of the MOS forecasts (Glahn 1985). If the overall skill of the new model is substantially higher, the skill of the MOS forecasts would probably improve. If the overall skill of the new model is only slightly higher, however, the MOS forecast skill would probably decrease slightly. Finally, if the old and new models have roughly equal skill, but different systematic error characteristics, the MOS forecast skill would undoubtedly decrease. This last situation applies to the impact of the RAFS modifications on the cloud guidance. NMC made an effort to minimize the changes to the systematic characteristics of the forecast fields, but a procedural change significantly affected the bias of the 300-mb RH.

Although it appears that the current NGM MOS will not be degraded by the implementation of the new RAFS, a word of caution is prudent. We should not forget that this evaluation was based on one month of springtime data. This is a relatively short sample which is not representative of other seasons. In addition, since we have not yet finished developing equations for elements such as thunderstorms or precipitation type (which may be more sensitive to model changes), we are not able to assess the impact that the RAFS changes may have on these future products. Likewise, we suspect that changes to the NGM topography in Alaska may significantly impact the statistical characteristics of the model forecasts and the development of NGM MOS guidance for that area. Consequently, we've decided to forego development of NGM MOS equations

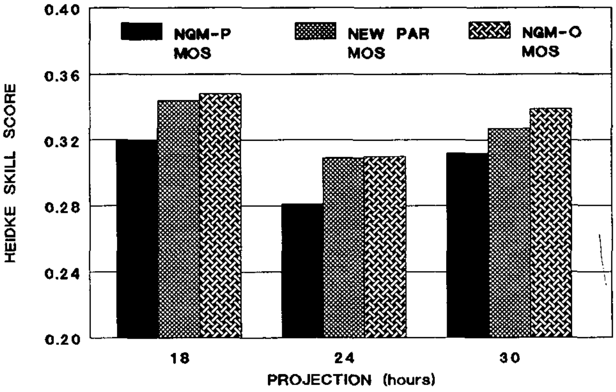

FIG. 12. Heidke skill scores for the original parallel NGM MOS (NGM-P), the revised parallel NGM MOS (NEW PAR), and operational NGM MOS (NGM-O) best category cloud cover forecasts. Forecasts for only 3 projections from 0000 UTC were tested.

for Alaska. Instead, we plan to use the modified perfect prog approach (Erickson 1988) to develop a forecast system for Alaska. The perfect prog technique does *not* account for the systematic biases of the numerical model or the deterioration of model forecasts with increasing projection, but the perfect prog method is less sensitive to changes in the statistical characteristics of the model forecasts.

Acknowledgments. The authors would like to thank Dr. Ralph Petersen and his staff in the Regional and Mesoscale Modeling Branch of NMC's Development Division for their hard work and cooperation in making the parallel test and TDL's evaluation possible.

REFERENCES

Carter, G. M., J. P. Dallavalle and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting,* **4,** 401–412.

DiMego, G. J., and K. E. Mitchell 1990: Personal communication. Regional and Mesoscale Modeling Branch, Development Division, NMC. 6 September 1990.

Erickson, M. C., 1988: Development and evaluation of perfect prog guidance based on output from the Nested Grid Model. Preprints: *Eighth Conference on Numerical Weather Prediction,* Baltimore, Amer. Meteor. Soc., 565–571.

Gerrity, J. P., Jr., 1977: The LFM model—1976: A documentation. NOAA Tech. Memo. NWS NMC-60, 68 pp. (NTIS PB-279-419).

Glahn, H. R., 1985: Statistical weather forecasting. *Probability, Statistics, and Decision Making in the Atmospheric Sciences,* A. H. Murphy and R. W. Katz, Eds., Westview Press, 289–335.

——, and D. A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1203–1211.

Hoke, J. E., N. A. Phillips, G. J. DiMego, J. J. Tuccillo and J. G. Sela, 1989: The regional analysis and forecast system of the National Meteorological Center. *Wea. Forecasting,* **4,** 323–334.

Jacks, E., J. B. Bower, V. J. Dagostaro, J. P. Dallavalle, M. C. Erickson and J. C. Su, 1990: New NGM-based MOS guidance for maximum/minimum temperature, probability of precipitation, cloud amount, and surface wind. *Wea. Forecasting,* **5,** 128–138.

National Weather Service, 1982: *National Verification Plan.* National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 81 pp. (Limited quantities available from the Office of Meteorology, NWS, 1325 East-West Highway, Silver Spring, Maryland, 20910).

Newell, J. E., and D. G. Deaven, 1981: The LFM-II model—1980. NOAA Tech. Memo. NWS NMC-66, 20 pp. (NTIS PB-82-156845).

Petersen, R. A., G. J. Dimego, J. E. Hoke, K. E. Mitchell, J. P. Gerrity, R. L. Wobus, H. M. Juang and M. J. Pecnick, 1991: Changes to NMC's Regional Analysis and Forecast System. *Wea. Forecasting,* **6,** 133–141.