6A.1

ON THE ABILITY TO DEVELOP MOS GUIDANCE WITH SHORT DEPENDENT SAMPLES FROM AN EVOLVING NUMERICAL MODEL

Mark S. Antolik * and Michael N. Baker
Meteorological Development Laboratory
Office of Science and Technology
National Weather Service, NOAA
Silver Spring, MD

1. INTRODUCTION

A properly-constructed statistical post-processing system, such as those routinely developed by using the Model Output Statistics technique (MOS; Glahn and Lowry 1972), can help to remove systematic biases in numerical weather prediction (NWP) model output. In addition, MOS systems can provide forecasts of quantities not explicitly predicted by the numerical model and can help account for local effects, effectively "downscaling" the NWP model forecasts to selected points of interest. It is not surprising, then, that MOS guidance has become widely accepted by field forecasters as providing a valuable "first guess" regarding the sensible weather to be expected given a specific set of model-forecast conditions.

Statistical forecast systems, including MOS, perform best when long, stable samples of dependent data are available for their development. However, changes to operational NWP models routinely occur as advances in computer power enable ongoing enhancements to model resolution and complexity. Hence, "ideal" samples of NWP model output for statistical post-processing are difficult to obtain within the constraints of the modern operational environment. In addition, sudden changes to the stochastic properties of NWP models can compromise previously-developed statistical systems. Nonetheless, forecasters still desire robust and reliable statistical guidance based on the operational NWP models, and this guidance is most useful if it can be made available within a reasonable length of time following major model changes.

Fortunately, experience at the National Weather Service's (NWS) Meteorological Development Laboratory (MDL) has shown that useful MOS guidance still can be produced under these conditions. A decade ago, experiments with output from an early version of the Eta-coordinate (Eta) model suggested that skillful MOS precipitation forecasts could be produced with as little as two years of dependent data collected while the model was actively undergoing development (Antolik 1998). Since that time, MDL successfully has implemented major operational MOS systems based on considerably longer samples of output from NWP models which were not strictly "frozen" to design changes (e.g., Dallavalle, et al. 2004).

In this paper, we discuss the specific challenges faced as MDL sought to update the operational short-range MOS guidance suite after the configuration of the North American Mesoscale (NAM) model, run at the National Centers for Environmental Prediction (NCEP), underwent significant changes in June 2006. We describe the effects of the change on existing MOS guidance, our strategy for mitigating these effects on NWS operations, and experimental work aimed at producing a new, updated NAM-based MOS package with the available model data. We also discuss some encouraging implications of the NAM MOS development with regard to the more general problem of producing MOS guidance based upon short samples of output from NWP models in today's operational environment. A companion paper (Maloney et al. 2009) describes the operational NAM MOS system implemented in December 2008 as a result of these studies.

2. MOTIVATION

2.1 *Eta Model Replacement*

On June 20, 2006, NCEP replaced components of the data initialization and assimilation system, analysis, and forecast model used in the NAM portion of its operational NWP suite. Specifically, NCEP replaced the operational version of the Eta model (Rogers, et al. 2005) with the newer, Nonhydrostatic Mesoscale Model (NMM; Janjic et al. 2001), part of the Weather Research Forecast (WRF) system. In addition, the Eta model's 3DVAR (Three-Dimensional Variational; Barker et al. 2003) analysis system was replaced by the Gridpoint Statistical Interpolation (Wu et al. 2002) analysis. A number of other changes to the NAM initialization and data assimilation process also were made at that time (NCEP 2009a).

One of the underlying principles of the MOS approach is that the operational forecast equations should be applied to the same (or nearly the same) analysis and prediction system from which they were developed. The June 2006 modifications to the components of the NAM were quite extensive; could the previously-developed Eta MOS equations still be used to produce meaningful forecast guidance when applied to NMM output?

* *Corresponding author address:*
Mark S. Antolik, 1325 East-West Highway, Rm. 11321, Silver Spring, MD, 20910-3283; e-mail: Mark.Antolik@noaa.gov

2.2 ***Effects of NAM Changes on Eta-based MOS***

Prior to NCEP implementation of the above NAM changes, MDL conducted a study to determine the extent to which the existing Eta-based MOS guidance might be affected. Test guidance was generated by applying the Eta-based MOS equations to output from the NMM running in a real-time parallel mode at NCEP over the period from March 1, 2006, through May 31, 2006. The performance of MOS guidance generated by this test method was compared to the performance of the operational Eta MOS forecasts for a number of weather elements over the same time period. Several of these comparisons showed a significant degradation in skill when this alternative "Eta-on-NMM" MOS approach (hereafter denoted Eta-NMM) was used.

The effects of using an Eta-NMM approach were not uniform for all weather elements, however. It appeared that the reduction in skill observed with the Eta-NMM MOS depended, at least in part, upon the particular weather element being forecast and, by extension, upon the particular type of equation employed. In general, MDL develops single-station MOS equations for commonly-observed elements such as surface (2-m) temperature and dewpoint. MOS guidance for other, less frequently-observed weather elements is developed using a regionalized-operator technique, since even a relatively long sample of NWP model output usually is not sufficient to obtain robust statistical relationships for individual stations. In this technique, data are pooled from a number of climatologically similar observing sites in order to increase the effective sample size. (For example, regionalized equations are used for prediction of precipitation probabilities and amounts as well as for most sky cover variables.) Because of this, regionalized equations are less finely "tuned" to individual forecast locations and thus tend to exhibit less sensitivity to local changes in NWP model characteristics.

This behavior generally was seen to hold true in the 2006 parallel verifications. Figure 1a shows the mean absolute error (MAE) of MOS 2-m temperature forecasts from the 0000 UTC model cycle time over the parallel evaluation period. Scores are aggregates calculated for a MDL standard set of 335 METAR (Meteorological Terminal Air Report) observing sites located over the conterminous U. S. (CONUS), Alaska, Hawaii, and Puerto Rico. These sites were chosen for reporting quality and consistency and their locations are plotted in Fig. 2. In Fig. 1a, note the rather significant MAE differences between the Eta-NMM MOS and the operational Eta-based MOS system at all forecast projections. Applying the existing Eta MOS equations to NMM output was seen to increase the overall system MAE by as much as 0.5 deg F during the 3-month parallel evaluation period, depending upon projection. This represents a reduction in forecast accuracy of up to 25% at the earliest lead times, where MOS generally performs best. The scores for the operational MOS guidance based on the NCEP Global Forecast System (GFS) are included (in red in Fig. 1) as a benchmark for comparison.[1]

On the other hand, Brier scores (Brier 1950) for 6-h PoP forecasts (Fig. 1b) showed considerably less degradation with the Eta-NMM approach. Brier Score differences between the operational Eta MOS and the Eta-NMM forecasts are noticeable only at the later forecast projections and are nowhere more than about 5%. Apparently, the (regionalized) PoP equations were less sensitive to the individual underlying NWP model and any possible changes in the stochastic properties of the model output. Differences in the skill scores between the operational Eta MOS and Eta-NMM forecasts for other regionalized elements were generally even smaller. More complete details of this parallel evaluation and verification scores for the full set of weather elements that were evaluated can be found online in MDL (2009a).

One particularly striking result was found in the parallel evaluation of 2-m dewpoint forecasts. Even though the single-station Eta MOS dewpoint forecast equations were developed simultaneously with those for 2-m temperature (i.e., were regressed together and contain a common set of model variables as predictors), their observed MAEs were not as severely impacted by the substitution of NMM output (Fig. 3a). Curiously, examination of the mean algebraic error (bias) of the forecasts over the test period (Fig. 3b) reveals a significantly *improved* overall bias resulting from the Eta-NMM substitution.

The precise reasons for the dewpoint behavior are unclear. It may be that the test period was somewhat drier overall than the norm (note that all statistical systems shown in Fig. 3 have an overall positive bias), and that the Eta MOS largely was unable to remove a resulting (positive/wet) bias in the underlying model output. If the NMM implementation improved the overall fidelity of the NAM-forecast moisture fields, then it is possible that this improvement could have been reflected in the bias statistics for the Eta-NMM MOS. A second possibility is that this result was simply fortuitous, and largely dependent upon the particular meteorological characteristics of the period chosen for the parallel tests. In any event, the stochastic properties of NMM output moisture variables (used as input to the MOS equations) appeared to be different from those of the corresponding Eta model output variables over the test period, despite the improved Eta-NMM bias performance. This behavior did not serve to increase our confidence that the existing Eta MOS dewpoint equations could be applied successfully to NMM output, but it did foreshadow some

[1] The GFS MOS was developed with a relatively long sample of data from a model which, while not strictly frozen, was not subject to drastic changes in configuration during collection of dependent data. Thus, GFS MOS scores could be considered a proxy for scores obtainable under conditions which are as close to optimal as is feasible operationally.

interesting results that were observed during our efforts to develop replacement dewpoint equations as part of a NAM MOS prototype (discussed below in Section 3).

2.3 *Mitigating the Impact on Operations*

Our 2006 parallel tests indicated that output from the NMM could not simply be substituted in the operational Eta-based MOS system without possible changes to the character of the forecasts and potential degradation of forecast skill. Therefore, at the time of WRF-NMM implementation, NCEP began running an interim 32-km version of the Eta model with NAM initial conditions at 0000 and 1200 UTC expressly to support production of the Eta MOS. However, no direct output from this interim version of the Eta model was distributed to the field, thus eliminating the ability of users to refer to the underlying model data when evaluating the Eta MOS guidance. This was not an ideal situation from the forecaster's perspective, and user responses to a NWS open public comment period underscored the need to devise a suitable replacement product from the NAM before removing both the Eta model and its associated MOS guidance. Furthermore, successful implementation of a new NAM-based MOS system would allow NCEP to terminate running of the Eta model in support of the Eta MOS, thus freeing up computational resources for other modeling tasks.

Consequently, in mid-2007 MDL set out to develop a new MOS system based on output from the NAM suite in a configuration which more closely resembled its current operational state. All Eta-based products were to be replaced by identical NAM-based MOS products generated by using either new equations redeveloped from samples of dependent data which specifically included NMM output, or by application of existing Eta-based equations to NMM output. The latter (Eta-NMM) approach would be used only where tests indicated that there would be no adverse effects resulting from the substitution of NMM output in the original equations, and only on an interim basis until sufficient samples of NMM data could be collected to ensure stable statistical relationships. In general, this meant deferring redevelopment of regionalized equations and/or equations for the most rarely-observed weather elements.

Initial emphasis was placed on developing NAM-based replacement equations for forecasting 2-m temperature and dewpoint (at 3-h intervals), daily local maximum/minimum (max/min) temperature, and wind speed and direction (also at 3-h intervals). This strategy was chosen because MOS guidance for these elements appeared to be the most affected by the June 2006 NCEP model changes. Given the importance of accurate precipitation guidance to field operations, development of replacement MOS equations for probability of precipitation (PoP) also was included in our initial efforts.

3. EXPERIMENTS AND RESULTS

3.1 *Available Model Data and Dependent Samples*

Prior MDL experience has indicated that a minimum of two, 6-month seasons of dependent data from a stable NWP model generally are required to produce robust statistical relationships when developing MOS equations[2]. This requirement was not exactly satisfied as we set out to develop our Eta MOS replacement. Not only were two complete warm and cool seasons of NMM output unavailable, but data collection efforts also were complicated by ongoing changes to the underlying model. In addition to the wholesale replacement of the Eta model by the NMM in mid-2006, an extensive set of changes to the operational configuration of the NMM was implemented on 19 December 2006 (NCEP 2009b). This involved a major revision of the model's convective parameterization and three distinct changes to the cloud microphysics, among other modifications. Therefore, since the NAM had been running in its latest WRF-NMM configuration for only a few months prior to system development, dependent samples for our experiments involve a mixture of data from the older Eta model and the first two operational configurations of the NMM.

3.2 *Initial Tests*

The authors conducted a limited, initial study to examine the feasibility of deriving updated NAM-based MOS guidance under the above conditions. Our preliminary tests focused on the cool season and the 0000 UTC model cycle time. We examined only 2-m temperature and dewpoint forecasts at a limited number (49 total) of the standard 335 MOS verification sites across the CONUS, Alaska, Hawaii, and Puerto Rico (see Fig. 2). We also confined our examination to early forecast lead times, specifically the 18-, 24-, and 30-h projections. A description of these experiments and subsequent results are given below.

The changes to the NMM on 19 December 2006 effectively split the 2006-07 six-month cool season into two approximately equal halves, thus creating a convenient division for our test data samples. The first "half" cool season refers to 1 October-19 December 2006 while the second "half" covers 20 December 2006-31 March 2007. We also employed Eta model data from the previous cool season (2005-06) in an effort to determine the effects of a longer mixed-sample development. Though we had initial concerns about the efficacy of combining the two half seasons of NMM, cursory investigations into the statistical properties of NMM output from the two sub-samples revealed no evidence that this would be detrimental. The test systems are described below, and Fig. 4 depicts a timeline of the

[2] MDL currently stratifies the dependent samples for development of MOS equations into warm (April-September) and cool (October-March) seasons for most weather elements.

various dependent samples used. In order to simplify the experiments, no post-processing was applied to the resulting forecasts. (Ordinarily, operational MOS forecasts of temperature, dewpoint, and max/min are post-processed to ensure meteorological consistency between related weather elements over a given forecast time period.) In addition, all development was for the cool season only.

3.2.1 *Half-season Tests*

Our first set of tests (Test 1) involved deriving MOS equations from various combinations of "half seasons" of NMM and Eta model data, and applying these to an independent sample comprising the second half of the 2006-07 cool season. Predictors were based upon those in the operational Eta MOS system. The test development systems included the following:

NMM1: NMM output from the first half of the 2006-07 cool season only.

NMM1ETA2: NMM output from the first half of the 2006-07 cool season plus Eta output from the second half of the 2005-06 cool season, effectively patching together a "full cool season" of a mixed NMM plus Eta data.

NMM1ETA12: NMM output from the first half of the 2006-07 cool season plus Eta output from the entire 2005-06 cool season (both halves), representing an even longer mixed sample than NMM1ETA2.

ETA-NMM: Baseline comparison system with operational Eta equations applied to the NMM.

Looking at the 2-m temperature and dewpoint mean absolute error (MAE; Fig. 5a, b), the most evident outcome of these tests is that a longer sample (i.e., more than just a half-season of NMM data alone) performed better, even compared to the baseline ETA-NMM system. This is similarly true for the bias (Fig. 5c, d). Most notable is the improvement of the 2-m dewpoint bias over ETA-NMM, seen even in the NMM1 result (Fig. 5d).

3.2.2 *Full-season Tests*

The second set of experiments (Test 2) attempted to expand the dependent sample to a full season of both NMM and Eta model data. The resulting equations were applied to an independent sample covering 1-19 December 2006, 1-15 March 2007, and 1-30 April 2007. (Data from December and March were withheld from test system development.) Again, predictors were based upon those used in the operational Eta MOS system. This set of tests included:

NMMCOOL: NMM output from the full 2006-07 cool season.

NMMETACOOL: NMM output from the full 2006-07 cool season plus Eta output from the full 2005-06 cool season.

ETA-NMM: Baseline comparison system with operational Eta equations applied to the NMM.

As in the previous set of tests, inclusion of Eta data in our sample (thus making it longer) results in an improvement in MAE over using a single season of NMM output only, and results in a system which compares favorably to the baseline ETA-NMM (Fig. 6a, b). This is true both for 2-m temperature and, especially, for the 2-m dewpoint, which clearly outperformed the ETA-NMM system. Temperature biases (Fig. 6c) were not as good as the baseline ETA-NMM, though the longer NMMETACOOL sample did better than the single season NMMCOOL. Dewpoint (Fig. 6d), on the other hand, again displayed a remarkable improvement in bias over the ETA-NMM.

3.3 *NAM MOS Prototype System*

Based upon the results from the preliminary tests discussed in the previous section, we then developed a full NAM MOS prototype system for the cool season (0000 UTC cycle) to predict 3-hourly temperature, dewpoint, and max/min out through the entire 84-h forecast extent of the NAM. This system incorporated post-processing of the MOS forecasts, as performed operationally, and the addition of an elevation-adjusted 2-m temperature as a predictor (found to be useful in other MOS temperature developments). The independent verification sample withheld for these tests included only time periods in the cool season, 1-19 December 2006 and 1-15 March 2007, and was conducted for the entire compliment of MOS verification sites (Fig. 2). Comparisons included:

PROTO: Our prototype NAM MOS system. This consisted of a predictor set based upon the operational Eta MOS equations (with slight modifications) and the addition of elevation-adjusted 2-m temperature. The dependent sample was the same as that for NMMETACOOL in Section 3.2.2 (see also the dependent sample timeline, Fig. 4).

ETA-NMM: Baseline forecasts from the operational Eta equations applied to NMM, with post-processing.

ETAMOS: The operational Eta MOS system, with post-processing.

Since all these systems are post-processed, this represents a true test comparison for our NAM MOS prototype in an operational configuration. The MAE of the prototype development through 84 hours clearly improves upon the ETA-NMM, and is close to or even better than the operational Eta, for both temperature and dewpoint (Fig. 7a, b). Improvement over ETA-NMM is most evident in the earlier projections, through about

36-h. Temperature biases for the PROTO (Fig. 7c) are also much improved over ETA-NMM (especially through about 36-h), and are in line with or better than the operational ETAMOS. Prototype dewpoint biases (Fig. 7d) are much superior to both ETA-NMM and the operational ETAMOS. This is consistent with each of our previous preliminary tests. Evidently, the addition of at least some NMM data to the dependent sample is able to help the prototype system adjust for most of the previously-existing bias in the Eta-based MOS.

Daytime maximum and nighttime minimum temperature from the PROTO system (not shown) also exhibited general improvement over the ETA-NMM and were similar to or better than the operational ETAMOS. Less variability in the bias with forecast projection was also noted for these variables.

In addition to the temperature prototype development, MDL also created similar prototypes for 10-m wind and probability of precipitation (PoP) forecasts (Fig. 8a, b). for both the warm and cool seasons. PROTO results for both wind speed and direction also improved notably upon the ETA-NMM and were at least similar to the ETAMOS. PoP results, on the other hand, were less clear-cut, with little difference noted between the PROTO, ETA-NMM, and ETAMOS systems. A more complete set of results for both the warm- and cool-season NAM prototype systems is available in MDL (2009b).

3.4 *Sample Size and Stability of Result*

The results of the previous section look promising in terms of our ability to develop NAM MOS based on short available samples. In the earlier tests, it appeared that lengthening the dependent sample by adding output from the final Eta model configuration would be beneficial to the overall performance of the prototype. However, one caveat is that the independent sample used for verification was rather small and potentially subject to representativeness issues. In order to further validate the design of our NAM MOS prototype, we compared it to an essentially identical system derived solely from a single-season sample of NMM model data (NMMONLY). Hence, the dependent sample for the NMMONLY development was the same as the NMMCOOL sample used above in Test 2 (see Fig. 4). At the time of this final experiment, data from the full 2007-08 cool season was available for use as an independent verification sample, which was substantially longer than the original verification sample used to evaluate the skill of the prototype.

Both wind and temperature developments were compared, for the PROTO and NMMONLY systems. Results for both these variables (Fig. 9a, b) indicated that the NAM MOS prototype and NMMONLY systems were very similar, though the NMMONLY did have a slightly larger error, especially beyond the earliest forecast projections. This is consistent with our experimental findings in Section 3.2, and provides additional evidence that a longer mixed sample tends to perform better than using a short sample of one model configuration. The relatively close agreement (MAE within 0.2°F) between PROTO and NMMONLY further indicates that our prototype NAM MOS, as derived, provided enough stability in the developmental data sample. The relative similarity in model configurations between the older Eta model and newer NMM likely facilitated this ability to combine them into a longer dependent sample.

4. DISCUSSION

Our experiments confirm earlier indications that the MOS approach may be used effectively even when the underlying NWP model undergoes changes and short data samples must be used for MOS development. Though conducted on an assortment of output samples from different configurations of the NAM, our results were consistent in that nearly all test variations suggested that longer data samples would produce a more robust and skillful MOS system, in spite of the variations in the underlying model. This result extends findings of the previous Antolik (1998) study concerning precipitation amount forecasts to MOS forecasts of temperature, dewpoint, and winds. In general, it seems beneficial to add data to MOS dependent samples even when doing so means that output from an earlier configuration of the NWP model must be included. This seems to hold true even though the added data likely does not have precisely identical statistical properties.

The effectiveness of mixing output from different NWP model configurations presumably is dependent upon the overall representativeness of the joint sample of model data and observed weather events. For a MOS system to perform effectively on independent data, this joint sample must accurately reflect the stochastic characteristics of the operational configuration of the NWP model as well as the statistics of the overall population of the predictand(s). Thus, two factors compete when attempting to develop MOS systems based on short dependent samples: The improvement in representativeness (with respect to the observed weather) that may be gained by lengthening the data collection period must be balanced against the potentially negative effects of including model output with differing statistical properties. While it is best if all dependent data are collected from exactly the same version of the NWP model, the beneficial effects of using longer dependent samples seem to dominate as long as the model versions remain “substantively similar” to each other, at least in a statistical sense. This appears to be the case with the final, Rogers et al. (2005) Eta model configuration and the two NMM configurations employed in our experiments. Our approach likely would not work if output were combined from two numerical models which are markedly different in design or numerics, such as the NMM and GFS for example.

While the effects of future NWP model configuration changes on operational MOS systems cannot be anticipated, the “substantively similar” condition also would seem relevant to the application of previously-developed MOS equations to output from a new version of the underlying model. The Eta MOS system was based on data collected from a much earlier version of the model; this version was probably quite dissimilar to its eventual, final state as well as to the NMM. Therefore, it is not surprising that attempts to apply the older, Eta MOS equations to the latest NMM configuration (i.e. our Eta-NMM system) yielded unsatisfactory results for some weather elements.

If the bias characteristics of the underlying NWP model are changed significantly and the biases of individual output variables are largely independent of each other (which probably is not the case), then adverse impacts on pre-existing MOS guidance might result. However, if model changes reduce random model errors, then MOS guidance based on a previous model version should perform satisfactorily after the model has been changed. Multivariate MOS equations probably help to promote robustness as well, since they contain several different output variables from the underlying NWP model, many of which are correlated with each other. Any change in the statistical properties of a single predictor variable is offset somewhat by the presence of the other variables in the equations.

Antolik (1998) hypothesized that if an NWP model were changed during dependent data collection, then the resultant MOS system might still perform better than one developed solely with data collected prior to the change. While not ideal, this mixed sample should be more representative of the post-change conditions than a sample containing no information at all from the final model state. Since changes to the underlying NWP models are designed to result in better model forecasts, these improvements should be reflected in the associated MOS guidance even if a mixed sample of output is used. In other words, the inclusion of at least some output from the latest configuration of the NWP model should tend to improve MOS performance even if the bias characteristics of the new model version are somewhat different. Our latest results seem to suggest this. The inclusion of even a small amount of output from the NMM significantly improved the bias performance of the NAM MOS dewpoint forecasts despite the previously observed bias differences when Eta MOS equations were applied to NMM output.

It follows that “updating” existing MOS equations by adding output from the newest model version to the previously collected dependent sample should work well in many cases where there have been model changes. This should be true even if only a short time has elapsed since the latest version of the NWP model was introduced. From the standpoint of representativeness, MOS updates using this “mixed sample” MOS approach would seem to have an advantage over simple model bias correction techniques which rely upon adaptive statistics calculated over a short training sample of, say, the previous 30-60 days.

Our results are likely to be somewhat dependent upon the particular weather element being forecast and upon the specific design changes being made to the underlying model. The results also are likely to vary somewhat given the meteorology of the sample chosen for development, and by season and region of the country. The MOS forecasts for weather elements predicted by single-station equations appear to be most affected by NWP model changes. Not only are single-station equations more sensitive to local changes in the error characteristics of the underlying model, but the MOS equations for these weather elements (e.g., temperature, dewpoint) also tend to rely more heavily upon predictor variables from the lowest levels of the model. Given that changes to modern NWP models most often involve boundary-layer physics and parameterizations, MOS forecasts for these elements would be most likely affected. By the same token, forecasts for these weather elements would be most improved by MOS system updates. Conversely, regionalized MOS equations (such as those used for PoP forecasts) may neither be as severely affected by model changes nor as improved by MOS system redevelopment. This seems to be the case in our study.

Hamill et al. (2006) have suggested the use of reforecast datasets to improve the MOS guidance. In our opinion, 20-30 years of reforecast output from an NWP model are not needed for effective development of MOS equations. However, it is critical to MOS applications that these reforecasts be done on the model in its operational configuration. Given our results with non-static NWP models, a reforecast sample of 2-3 years from the latest configuration of a numerical model probably is sufficient to develop robust MOS equations for most weather elements. The representativeness of the reforecast sample could be improved further by re-running the new version of the model every third or fourth day over a longer time period. Constructing a reforecast sample of this length is quite feasible given current data storage and computing capability. Perhaps reforecast datasets of this type could be included routinely as a component of future NWP model implementations.

When our results are considered together with the Maloney et al. (2009) work on the now-operational NAM MOS system, it is clear that the current MOS approach remains viable in today’s operational environment. While the use of output from a non-static NWP model is not completely novel to MOS development, the current NAM MOS system is notable in that it represents the first time that MDL has replaced an operational MOS package in response to model changes which adversely impacted the existing guidance, and has done so with a limited amount of data from the new version of the model. This bodes well for our continued ability to provide forecasters with timely, updated MOS guidance after future NWP model enhancements.

5. ACKNOWLEDGMENTS

The authors wish to acknowledge the efforts of all MDL Statistical Modeling Branch personnel involved in the 2006 Eta/NMM parallel evaluation described in Section 2 of this paper. Some of their results have been reproduced here. The authors also wish to acknowledge the work of the three MDL developers (Geoff Wagner, Eric Engle, and Joseph Maloney) who helped to develop a prototype system of NAM MOS forecast equations for temperature, wind, and precipitation elements, and who also provided some of the verifications included in this paper. We also thank J. Paul Dallavalle for a number of insightful thoughts on the possible effects of NWP model changes on MOS guidance.

6. REFERENCES

Antolik, M. S., 1998: Eta-based MOS quantitative precipitation forecasts in an era of rapidly changing numerical models: How well can we do? Preprints, *16th Conf. Wea. Analysis and Forecasting*, Phoenix, AZ, Amer. Meteor. Soc, 446-449.

Barker, D. M., W. Huang, Y.-R. Guo, and A. Bourgeois, 2003: A three-dimensional variational (3DVAR) data assimilation system for use with MM5. NCAR Tech Note, NCAR/TN-453+STR, 68 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO, 80307.]

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1-3.

Dallavalle, J. P., M. C. Erickson, and J. C. Maloney, 2004: Model output statistics (MOS) guidance for short-range projections. Preprints, *20th Conf. on Wea. Analysis and Forecasting*, Seattle, WA, Amer. Meteor. Soc., CD-ROM, 6.1.

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

Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: Reforecasts: An important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33–46.

Janjic, Z. I., J. P. Gerrity, Jr., and S. Nickovic, 2001: An alternative approach to nonhydrostatic modeling. *Mon. Wea. Rev.*, **129**, 1164-1178.

Maloney, J., E. Engle, P. Shafer, and G. Wagner, 2009: The NMM MOS replacement for the Eta MOS. Preprints, *23rd Conf. on Wea. Analysis and Forecasting/19th Conf. on Numerical Wea. Prediction*, Omaha, NE, Amer. Meteor. Soc., CD-ROM, 6A.3.

MDL, cited 2009a: WRF MOS evaluation. [Available online at: http://www.weather.gov/mdl/synop/wrfmoseval.htm]

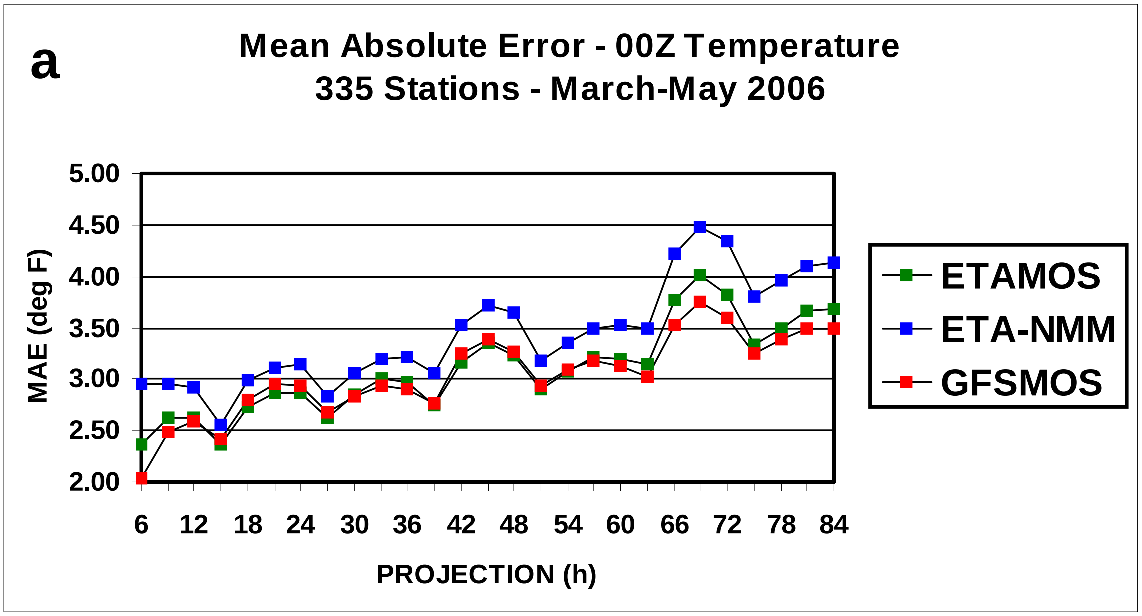
_____, cited 2009b: NAM-NMM MOS prototype evaluation. [Available online at: http://www.nws.noaa.gov/mdl/synop/nmmmoseval.php]

NCEP, cited 2009a: WRF-NMM and GSI analysis to replace Eta model and 3DVAR in NAM decision brief. [Available online at: http://www.emc.ncep.noaa.gov/WRFinNAM/Update_WRF-NMM_replacing_Eta_in_NAM2.pdf]

_____, cited 2009b: December 2006 NAM changes. [Available online at: http://www.emc.ncep.noaa.gov/mmb/namchanges_dec2006/nam_upgrades.nov2006.html]

Rogers, E., Y. Lin, K. Mitchell, W.-S. Wu, B. Ferrier, G. Gayno, M. Pondeca, M. Pyle, V. Wong, and M. Ek, 2005: The NCEP North American mesoscale modeling system: final eta model/analysis changes and preliminary experiments using the WRF-NMM. Preprints*, 21st Conf. on Wea. Analysis and Forecasting/17th Conf. on Numerical Wea. Prediction*, Washington, DC, Amer. Meteor. Soc., CD-ROM, 4B.5.

Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev*, **130**, 2905-2916.



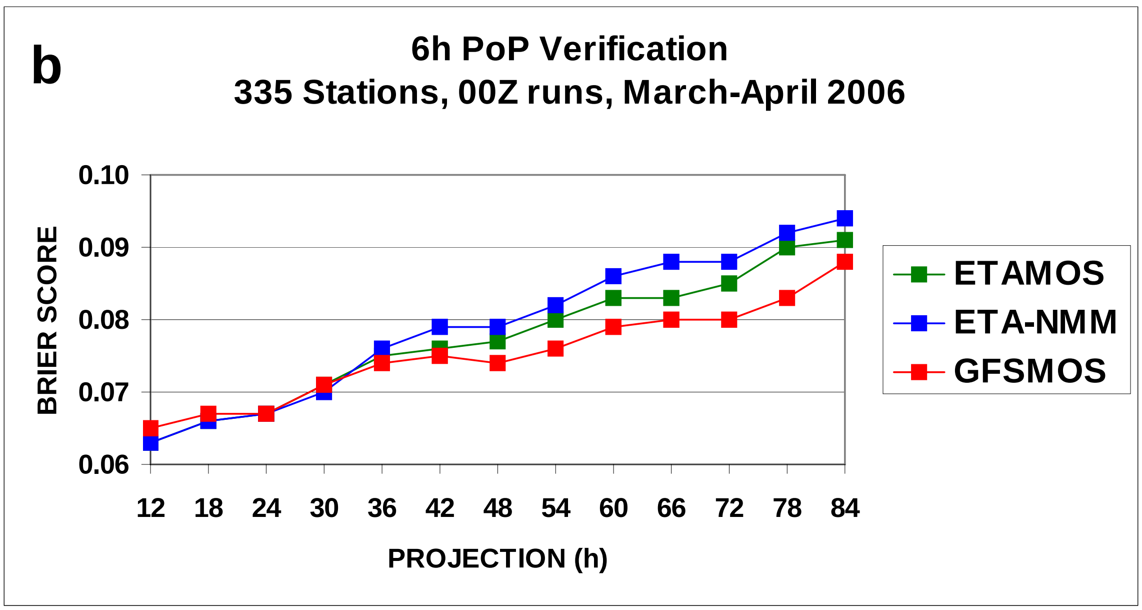


Figure 1. Parallel test results comparing the Eta MOS, Eta equations applied to the NMM (ETA-NMM), and the GFS MOS for (a) 2-m temperature MAE (°F) and (b) 6-h PoP Brier score. Adapted from MDL (2009a).

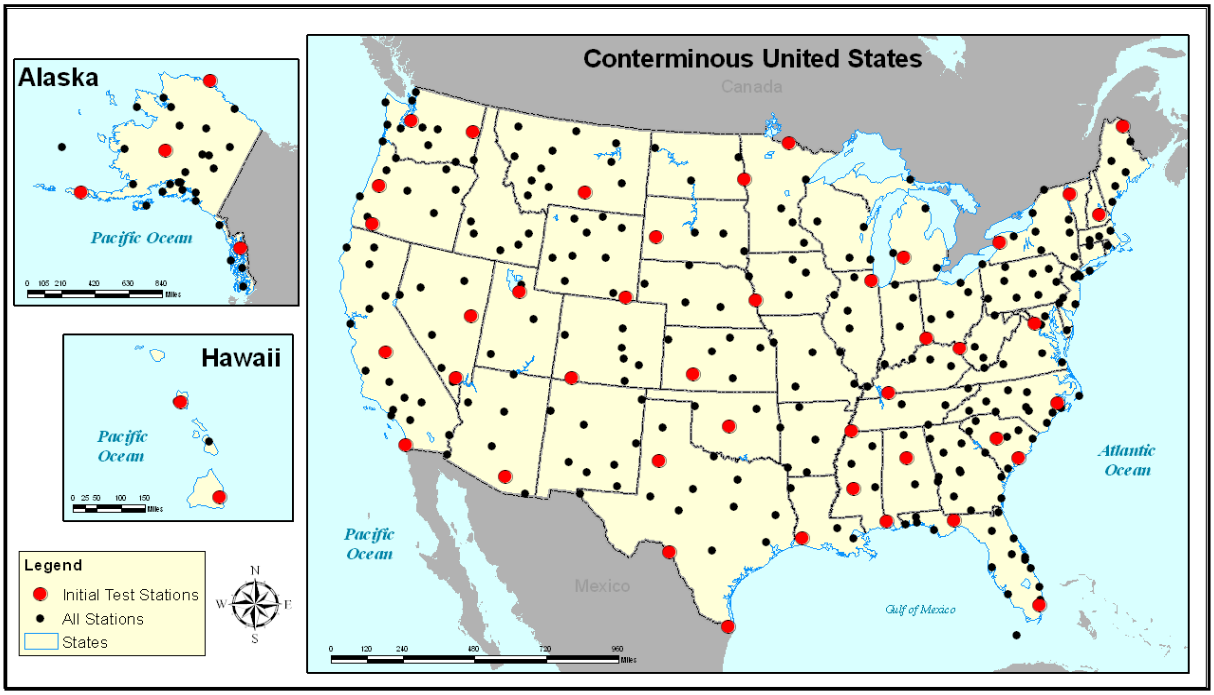
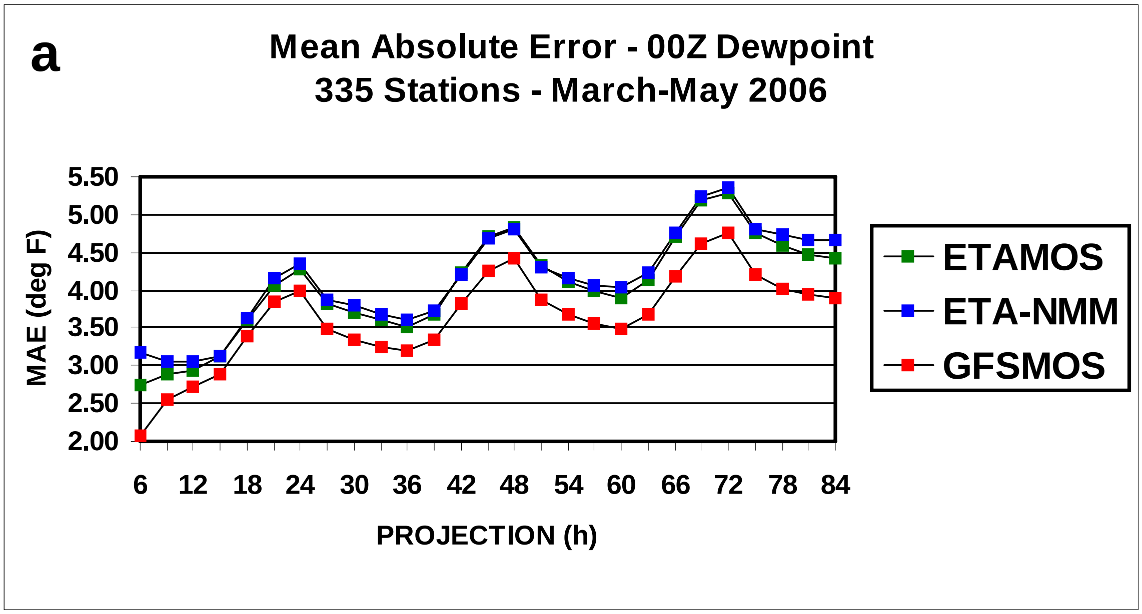


Figure 2. Locations of the 335 verification sites used in this study (all dots). Larger red dots depict the sub-sample of 49 sites used for the initial tests described in Section 3.2.



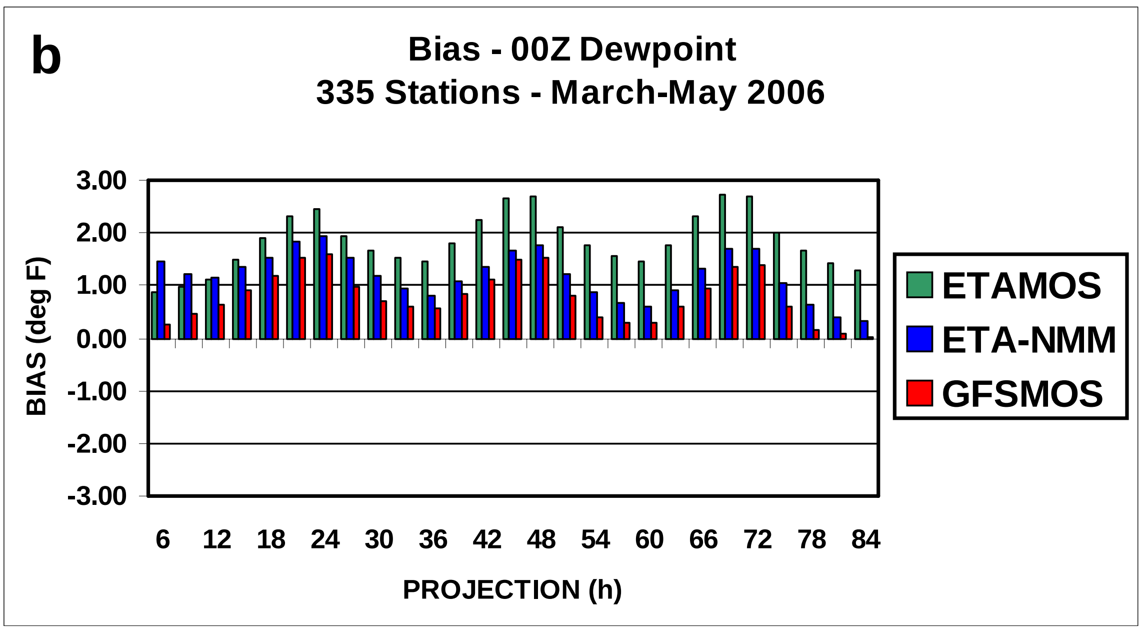


Figure 3. As in Fig. 1, but for (a) 2-m dewpoint MAE (°F) and (b) 2-m dewpoint bias (°F).

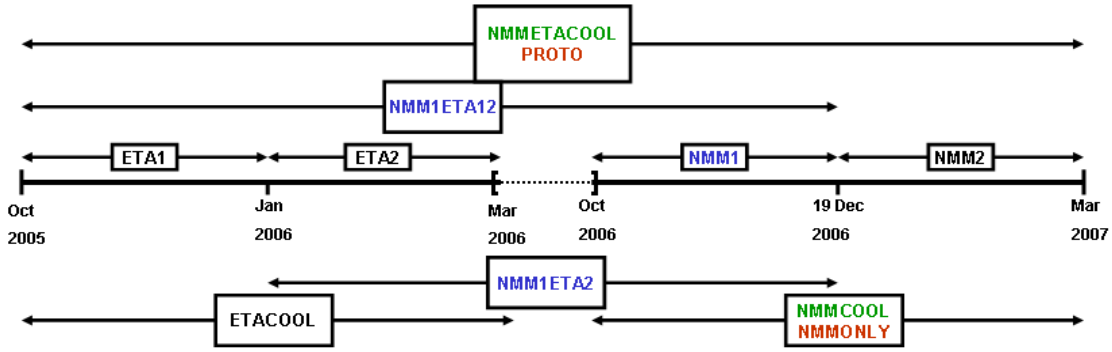


Figure 4. Timeline depicting the various dependent samples available for the test configurations described in Section 3. Samples used for Test 1 are shown in blue, Test 2 in green, NAM MOS prototype (PROTO) and NMMONLY in red. Note that some sample time periods overlap. See text for further discussion.

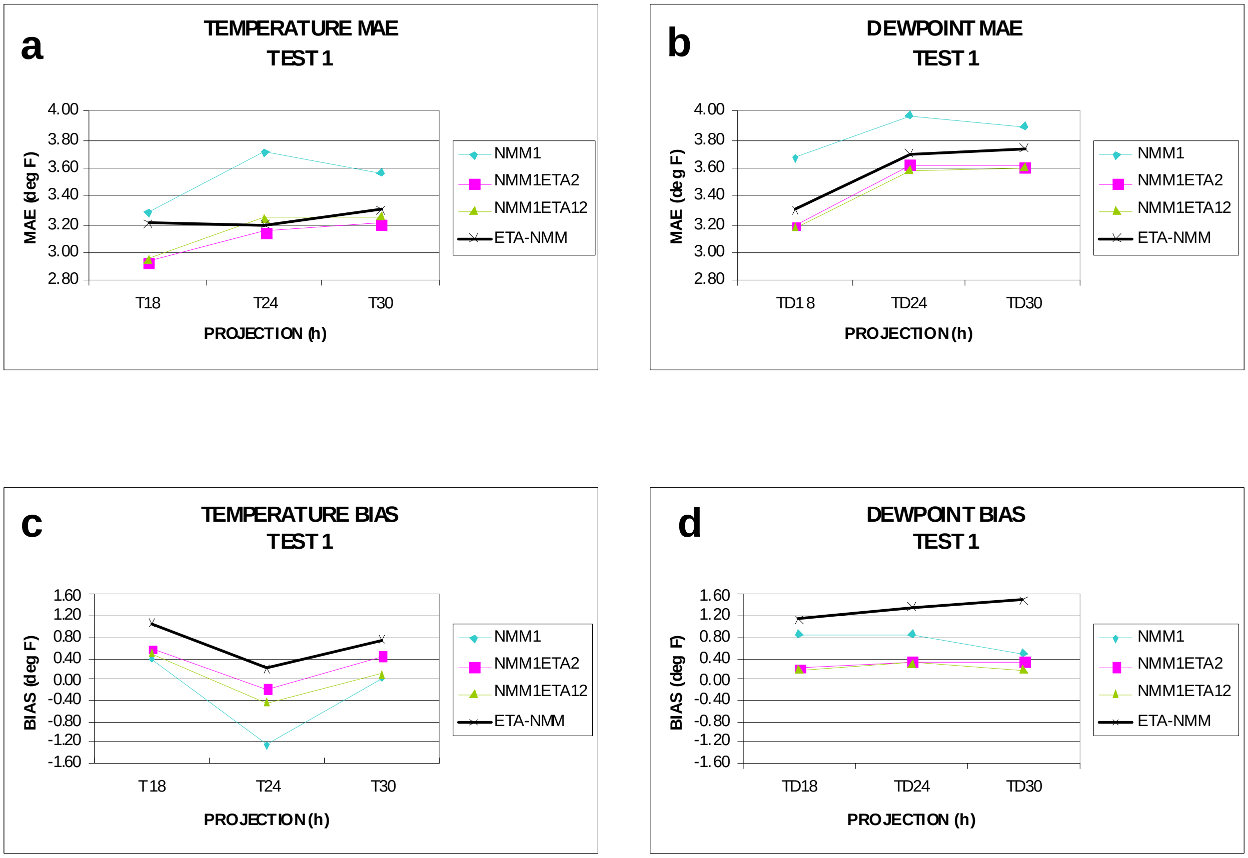


Figure 5. Test 1 MAE and bias (°F) results for projections 18-30-h: (a) 2-m temperature MAE, (b) 2-m dewpoint MAE, (c) 2-m temperature bias, and (d) 2-m dewpoint bias.

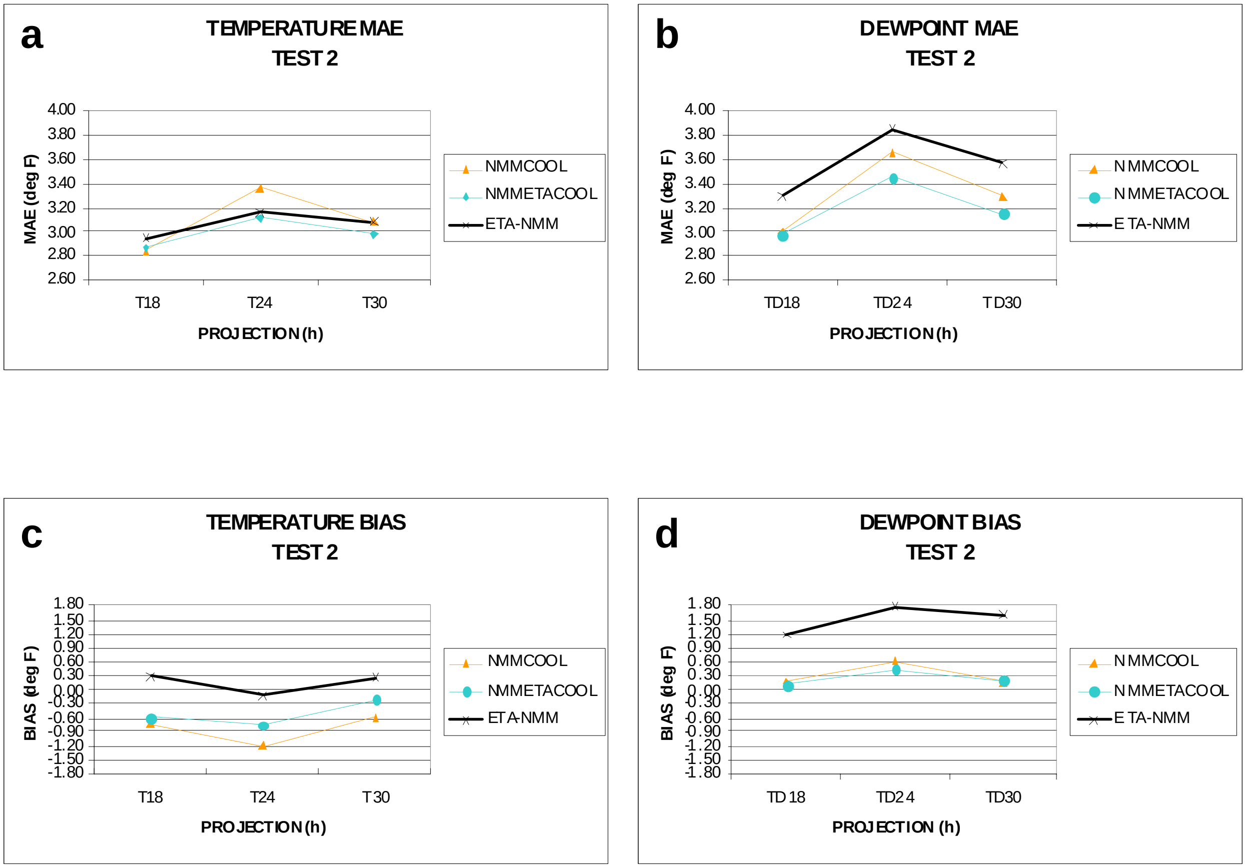


Figure 6. Same as in Fig. 5, but for Test 2.

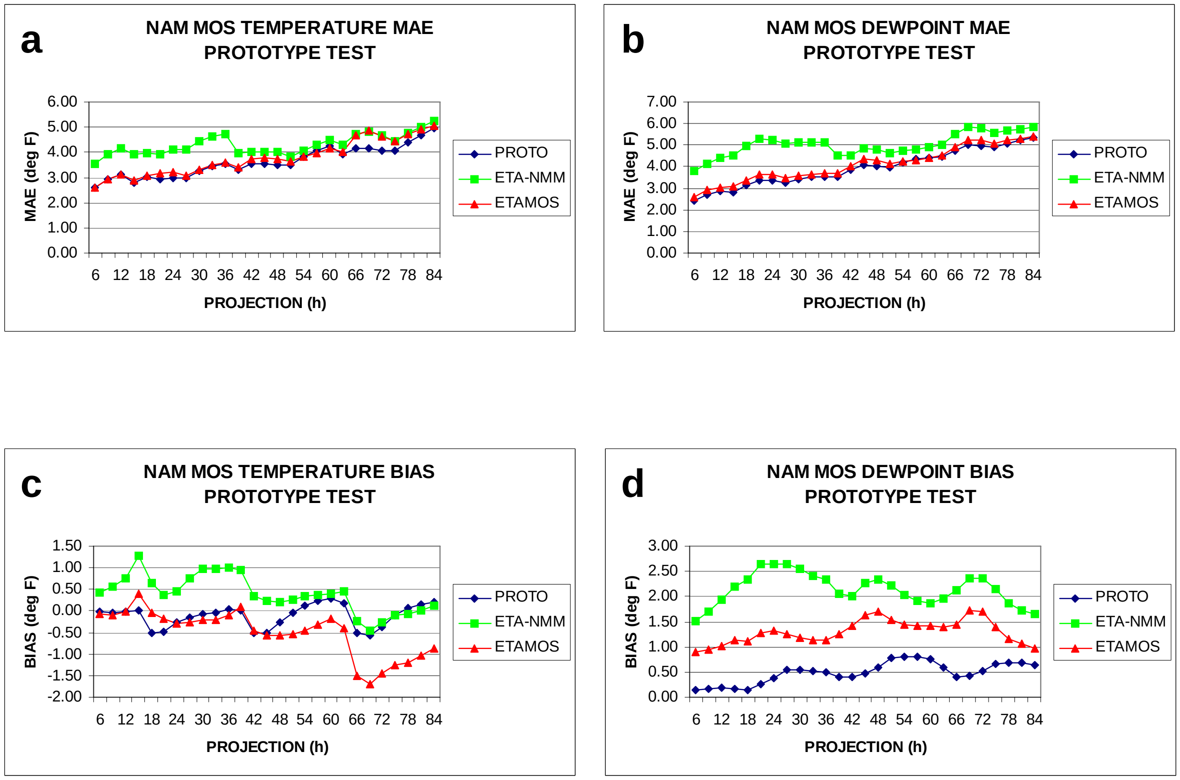


Figure 7. Cool-season MAE and bias comparison (°F) for NAM MOS prototype system (blue), operational Eta MOS (red), and Eta MOS equations applied to NMM output (green), as discussed in Section 3.3. Panels show results for: (a) 2-m temperature MAE, (b) 2-m dewpoint MAE, (c) 2-m temperature bias, and (d) 2-m dewpoint bias.

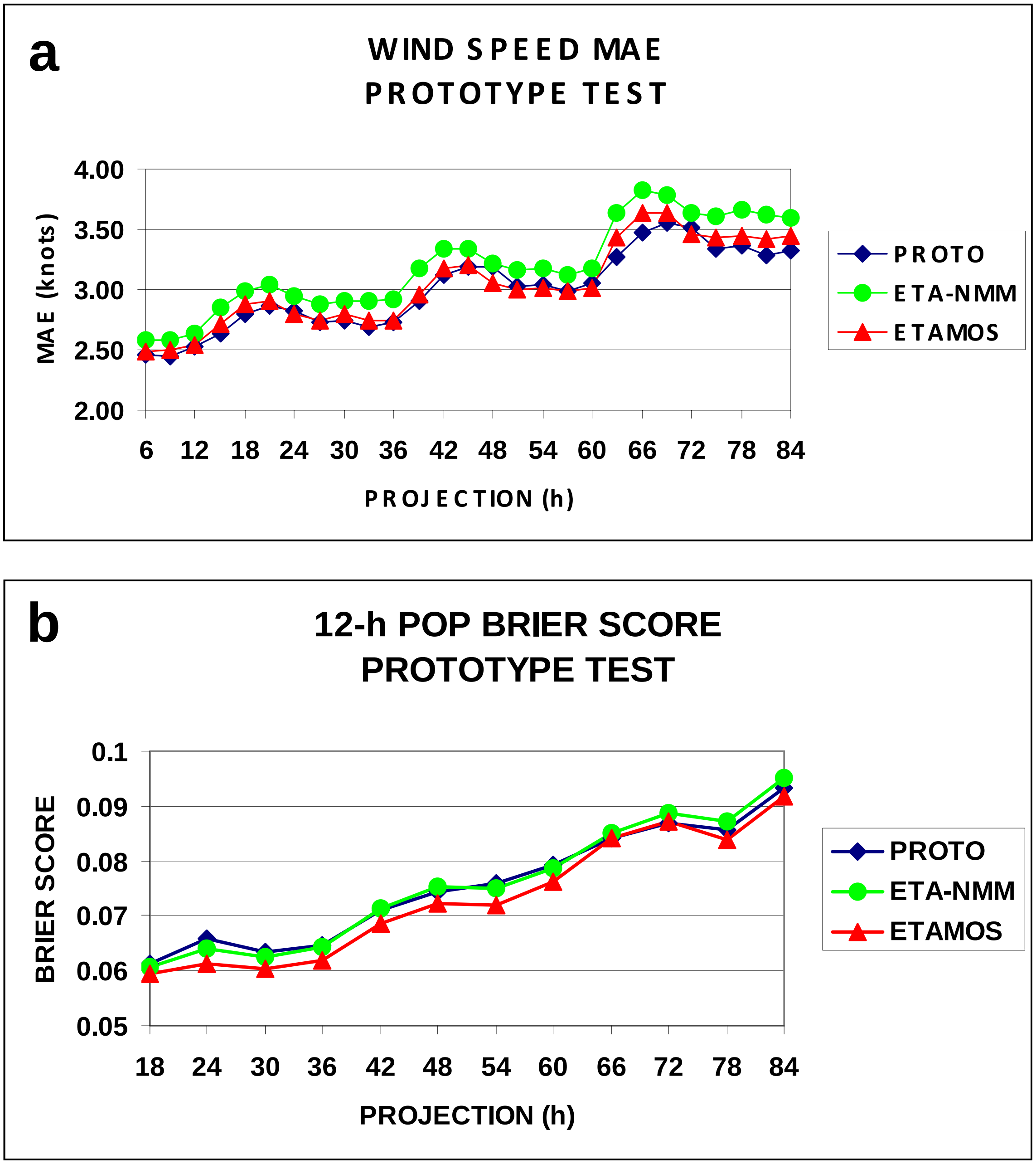
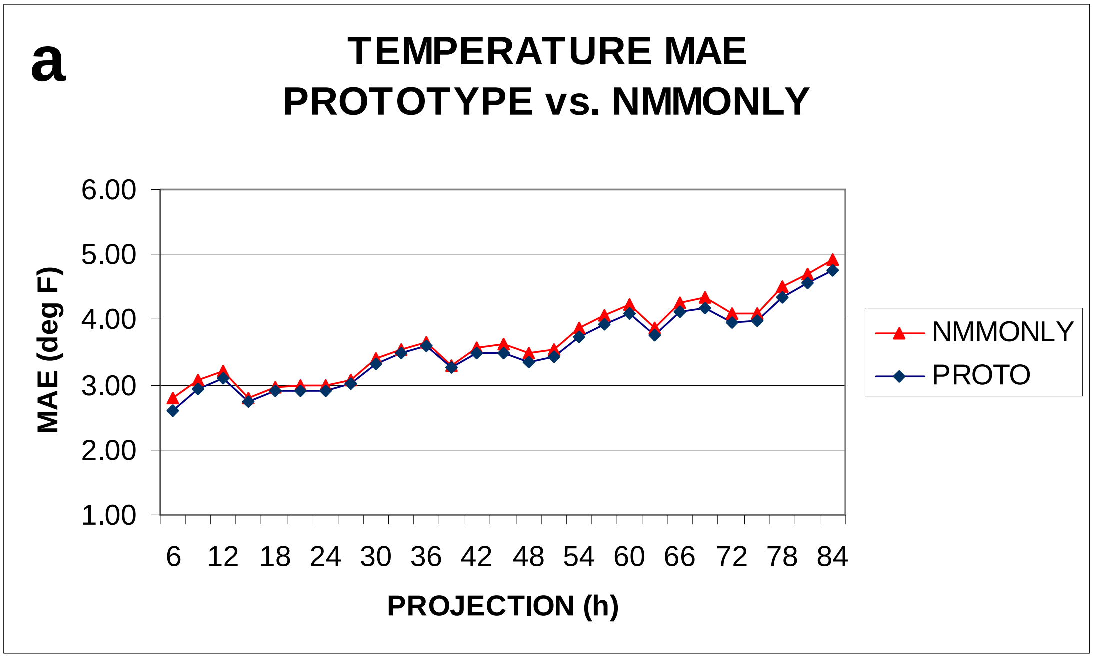


Figure 8. NAM MOS prototype results for (a) 10-m wind speed MAE (knots) and (b) 12-h PoP Brier score.



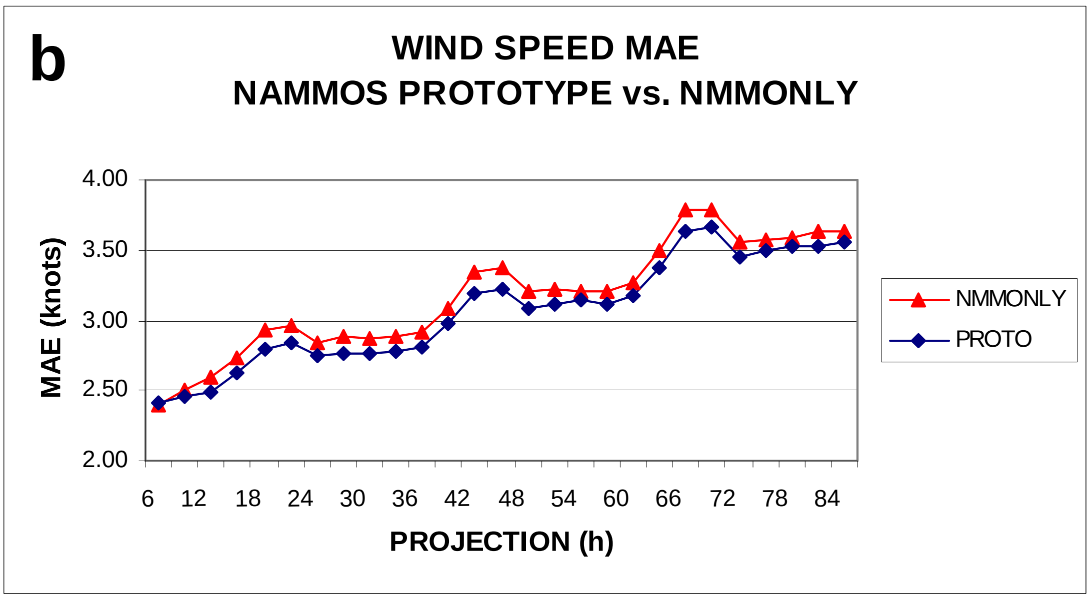


Figure 9. MAE (°F) of the NAM MOS prototype vs. NMM-only systems for (a) 2-m temperature and (b) 10-m wind speed. Scores are for the 2007-2008 cool season.