The Performance of MOS in the Digital Age

DAVID P. RUTH, BOB GLAHN, VALERY DAGOSTARO, AND KATHRYN GILBERT

Meteorological Development Laboratory, Office of Science and Technology, NOAA/National Weather Service, Silver Spring, Maryland

(Manuscript received 8 May 2008, in final form 25 August 2008)

ABSTRACT

Model output statistics (MOS) guidance forecasts have been produced for over three decades. Until recently, MOS guidance was prepared for observing stations and formatted in text bulletins while official National Weather Service (NWS) forecasts for stations and zones were prepared by forecasters typing text. The flagship product of today's NWS is the National Digital Forecast Database (NDFD). In support of NDFD, MOS is now also produced on grids.

This paper compares MOS and gridded MOS (GMOS) to the forecaster-produced NDFD at approximately 1200 station locations in the conterminous United States. Results indicate that GMOS should provide good guidance for preparing the NDFD. In those areas of the country where station observations well represent the grid, GMOS features accuracy comparable to that of NDFD. In areas of complex terrain not well represented by station observations, GMOS appears similar to NDFD in its depiction. A new score is introduced to measure convergence from a long-range forecast to the final short-range forecast. This shows good GMOS forecast continuity when compared to station MOS and NDFD.

1. Introduction

Model output statistics (MOS) guidance forecasts have been produced and provided to National Weather Service (NWS) forecasters and private entities for over three decades (Glahn and Lowry 1972; Carter et al. 1989). Until recently, MOS guidance was prepared for observing stations and formatted in text bulletins while official NWS forecasts for stations and zones were prepared by forecasters typing text. The flagship product of today's NWS is the National Digital Forecast Database (NDFD; Glahn and Ruth 2003). Legacy text products are automatically produced from digital forecasts and, more importantly, the database itself is made available to all customers and partners—public and private—so that those customers and partners can create a wide range of text, graphic, and image products of their own (information online at www.weather.gov/ndfd). In support of NDFD, MOS is also now produced on grids (Dallavalle and Glahn 2005), which are broadcast to NWS Weather Forecast Offices (WFOs) and posted for download via NDFD's companion—the National Digital Guidance Database (NDGD).

Figure 1 shows an example NDFD day 3 maximum temperature forecast with the corresponding NDGD gridded MOS forecast for the conterminous United States (CONUS). On this day in April, the NDFD has warm spring air extending farther into the northern plains than the NDGD. However, as on most days, overall patterns appear quite similar. Figure 2 shows a zoomed view of the same forecast for Utah and western Colorado, where similarities of detailed terrain features can be seen. Here, NDFD is generally a few degrees cooler than GMOS both in the mountains and in the valleys. Current NDFD forecast images can be viewed online (www.weather.gov/forecasts/graphical), as can current GMOS forecasts (www.weather.gov/mdl/synop/gridded/sectors). Side-by-side comparisons of the current GMOS and NDFD forecasts for the CONUS can also be viewed online (www.weather.gov/mdl/synop/gridded/sectors/conusCompare.php). The Meteorological Development Laboratory (MDL) routinely computes differences between GMOS and NDFD grids for several forecast elements to identify potential problems. While day-to-day differences can be large, average differences between NDFD and GMOS tend to be small.

Corresponding author address: David P. Ruth, Meteorological Development Laboratory, 1325 East–West Highway, Silver Spring, MD 20910.
E-mail: david.ruth@noaa.gov

DOI: 10.1175/2008WAF2222158.1

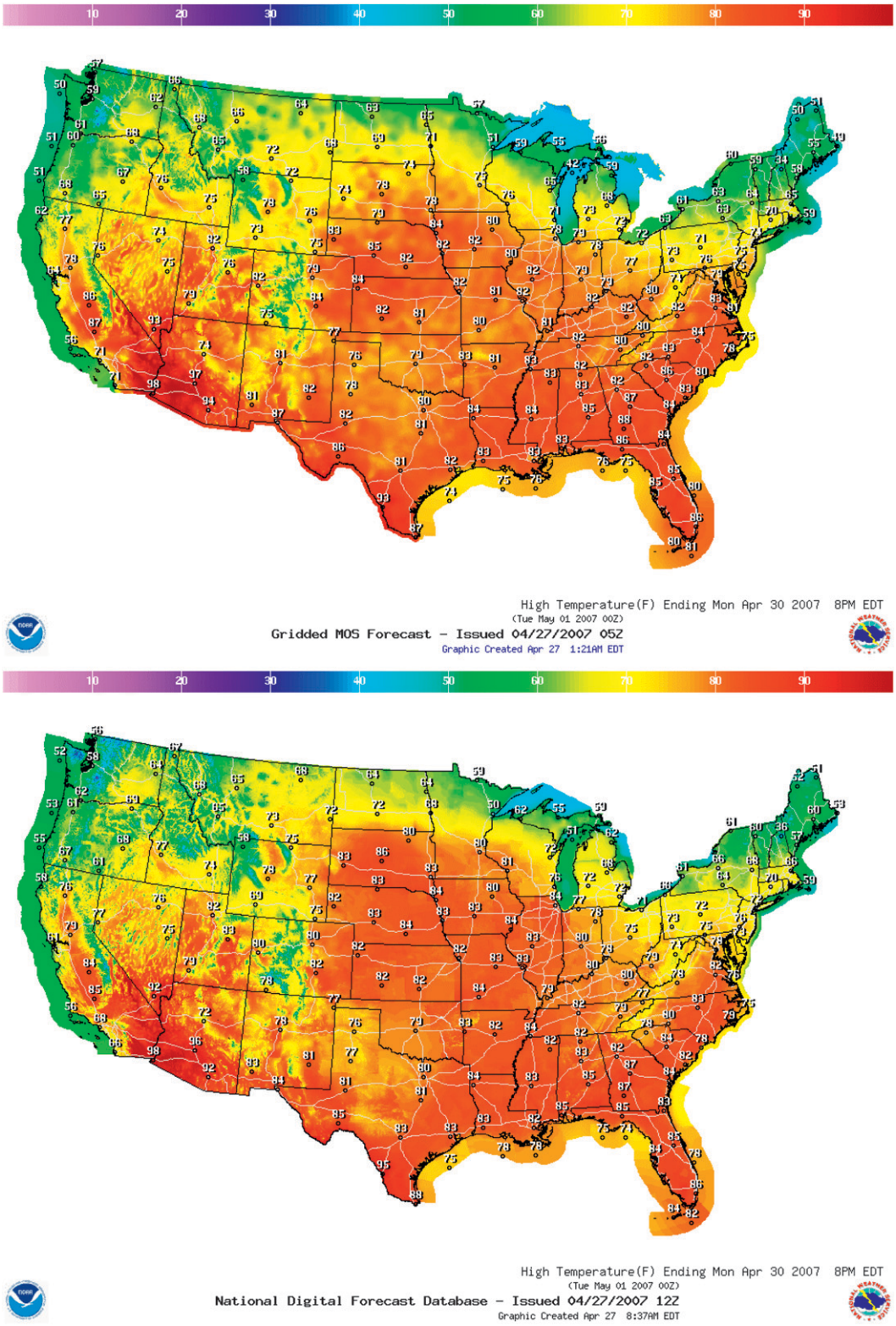

FIG. 1. Day 3 maximum temperature forecasts from (top) GMOS and (bottom) NDFD for 30 Apr 2007. The images and plotted values in the images depict gridpoint forecasts.

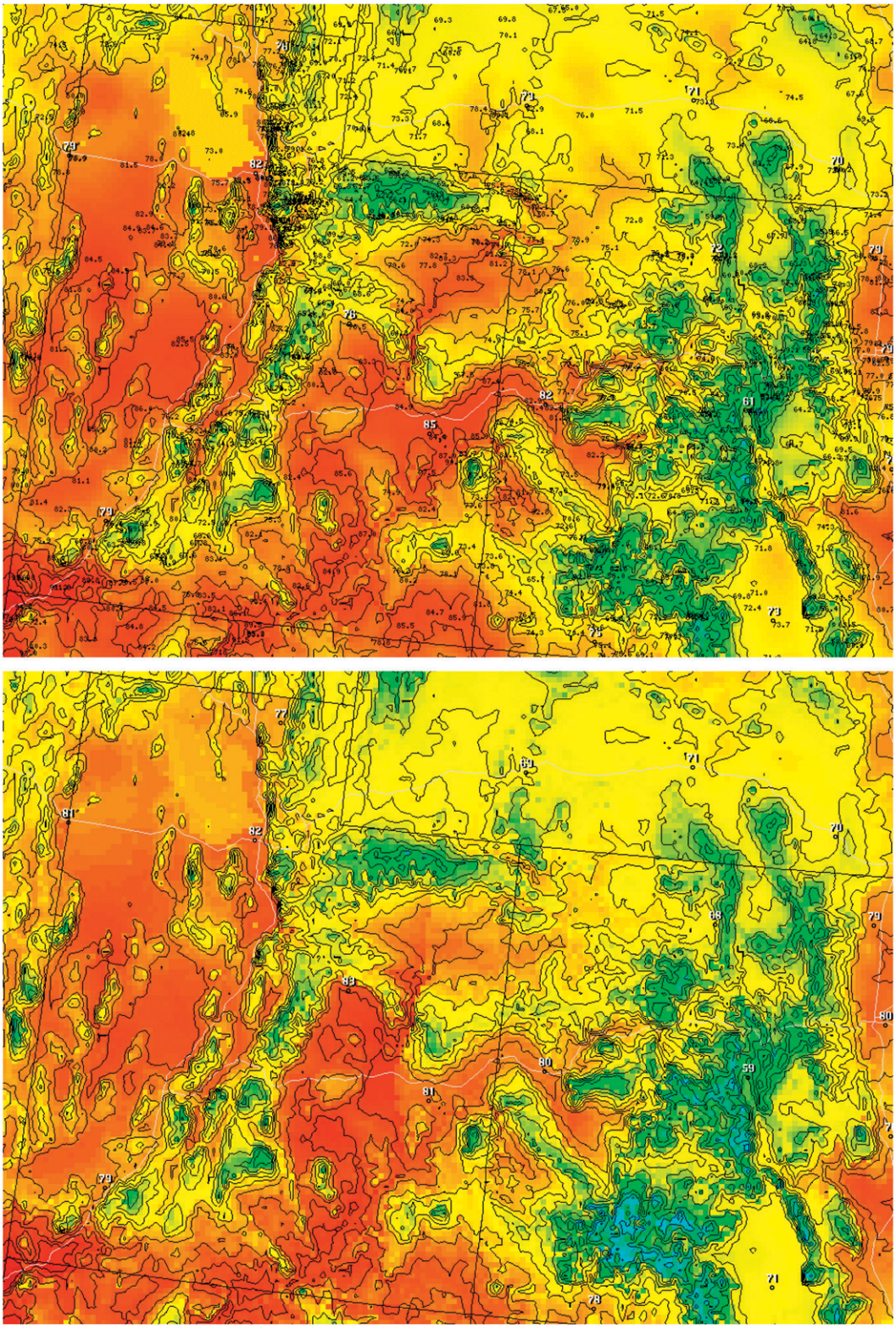

FIG. 2. Same as in Fig. 1, but with a zoomed view showing terrain detail in UT and western CO. Contours depict elevation. Temperature values (°F) plotted in black (top) show MOS station locations.

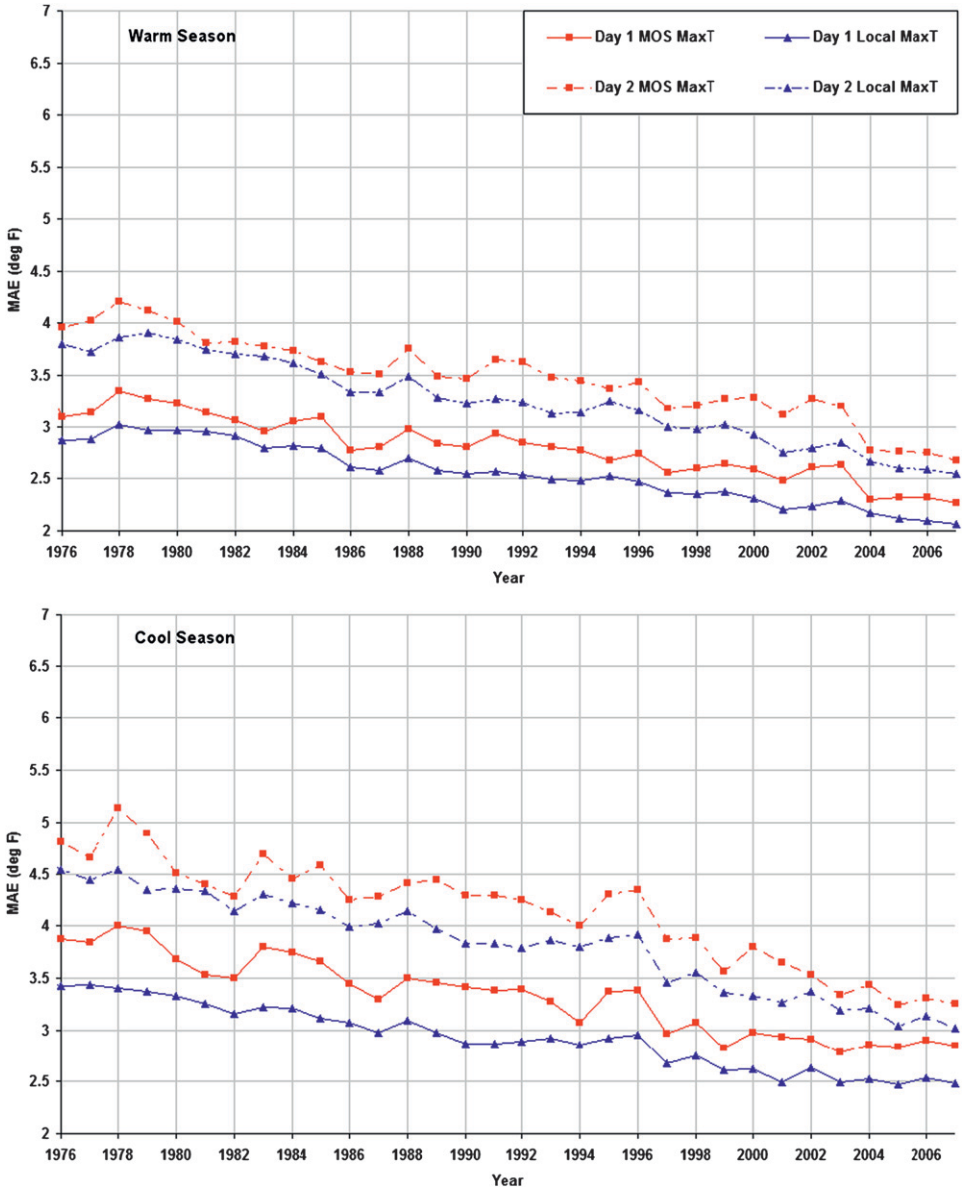

FIG. 3. Day 1 and day 2 maximum temperature MAEs for (top) warm (April–September) and (bottom) cool (October–March) seasons based on 0000 UTC model cycle guidance. Years show when season began.

A goal of MDL is to provide gridded MOS guidance for as many NDFD elements as possible. The guidance should be accurate, reflect high-resolution terrain, and provide good forecast continuity from when the forecast is first issued 7 days in advance, until it is last updated some hours before forecast valid time.

2. The long-term view

Dallavalle and Dagostaro (2004) documented the improvement in guidance products that objectively interpreted the output of numerical weather prediction models from 1966 through 2003. Figures 3–5 show verifications of MOS guidance compared to official forecasts prepared at local NWS offices for daytime maximum temperature (MaxT), nighttime minimum temperature (MinT), and 12-h probability of precipitation (PoP12) over the past three decades. Mean absolute errors (MAEs) are provided for temperatures, and Brier scores (Brier 1950) are provided for PoP12. Local forecasts are compared to MOS guidance that is available several hours prior to local forecast issuance. Specifically, local maximum temperature forecasts issued at approximately 0400 local time (LT) for the next two days are

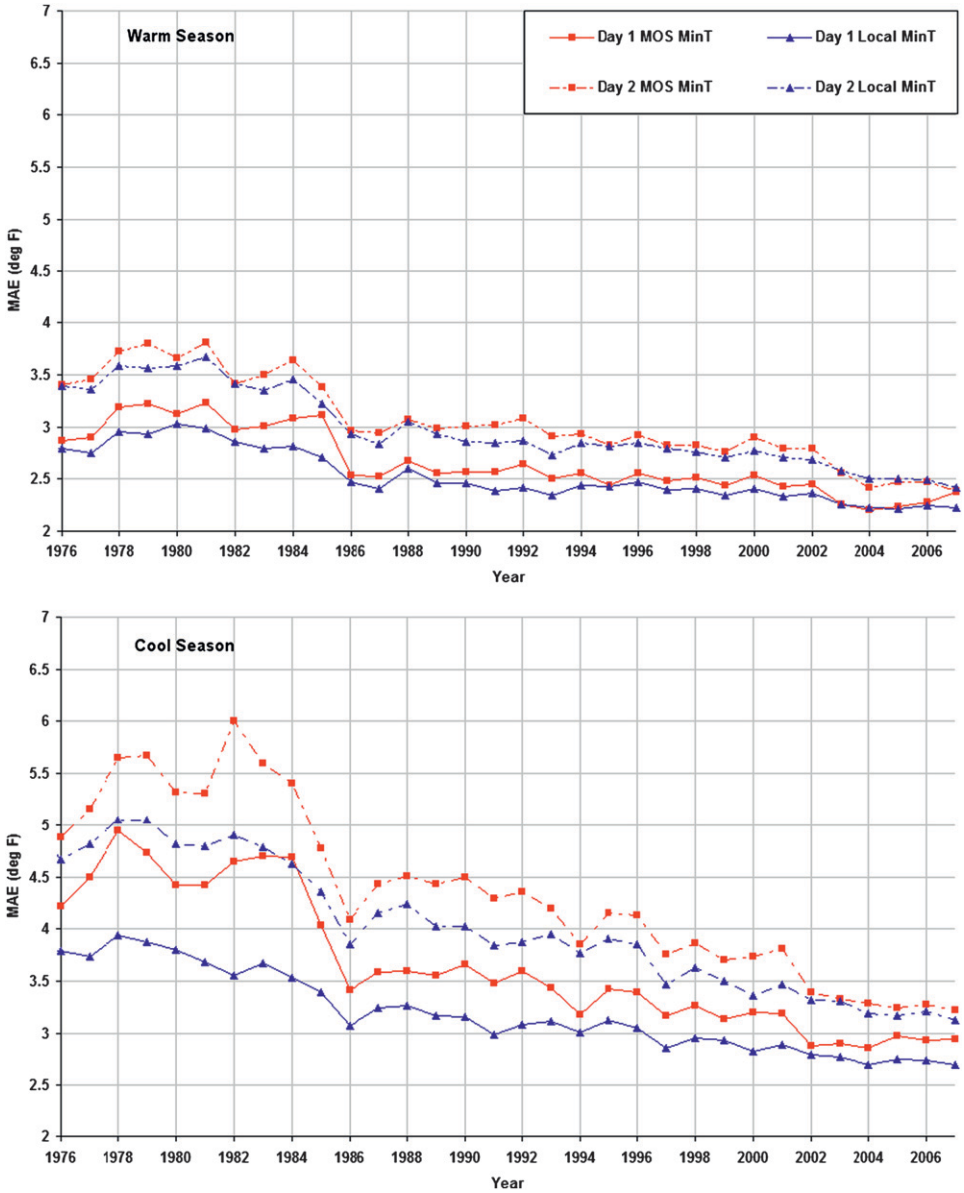

FIG. 4. Same as in Fig. 3, but for nighttime minimum temperature based on 1200 UTC model cycle guidance.

compared to MOS based on the 0000 UTC model cycle. Local minimum temperature forecasts issued at approximately 1600 LT for the next two nights are compared to MOS based on the 1200 UTC model cycle. The probabilities of precipitation forecasts issued at 0400 LT for the next two daytime periods, and at 1600 LT for the next two nighttime periods, are compared to MOS based on the 0000 and 1200 UTC model cycles, respectively. The Dallavalle and Dagostaro (2004) charts are here supplemented by scores from the 2004 warm season through the 2007 cool season for 79 stations available out of the original 80.

Year-to-year improvements in MOS and local forecasts are correlated, and can often be tied to the implementation of new or improved numerical and statistical models. For example, an increase in Global Forecast System (GFS) model resolution in October 2002 seems to have had a positive effect on MOS scores for subsequent seasons. Decreases in performance can sometimes be attributed to problems with models as well. Relatively large MOS errors for the day 1 nighttime minimum during the 2007 warm season were likely caused by recognized problems with the GFS surface sensible heat flux that year. In this case, locally prepared forecasts did not

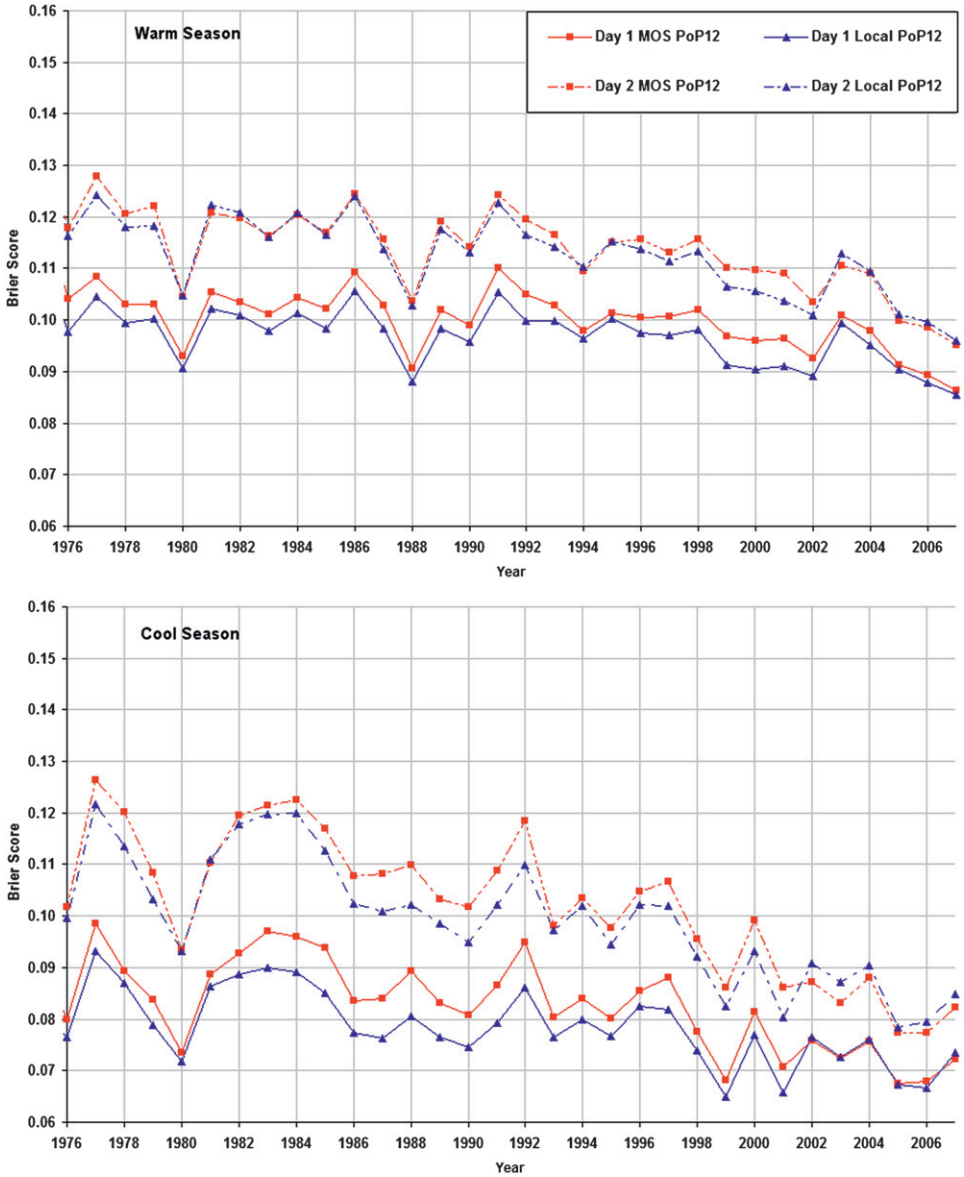

FIG. 5. Same as in Fig. 4, but for the PoP12 Brier score for the 0000 and 1200 UTC model cycle guidance combined.

appear to be negatively impacted by the poorer model performance.

Overall, two trends are evident in these figures. Forecasts are continually improving; MOS guidance for day 2 is now about as accurate as day 1 guidance was 10–15 yr ago. Second, the accuracy of MOS on day 1 and day 2 for these three elements is now growing close to that of the local forecast.

3. Forecast verification in the digital age

Until recently, NWS forecasters prepared and disseminated official forecasts primarily by writing narrative text for zones and stations. Our ability to verify these forecasts was limited to the few elements, forecast projections, and stations that were available in coded products (e.g., terminal aerodrome forecasts, coded cities forecasts), or that were required to be entered into special tables by hand (Ruth and Alex 1987). With the nationwide implementation of the Interactive Forecast Preparation System (IFPS; Ruth 2002) at NWS WFOs, local digital forecasts for many forecast elements became available at high temporal and spatial resolutions from forecast days 1–7. By the end of 2003, MDL had developed and implemented a prototype verification system for examining local WFO forecasts in NDFD and

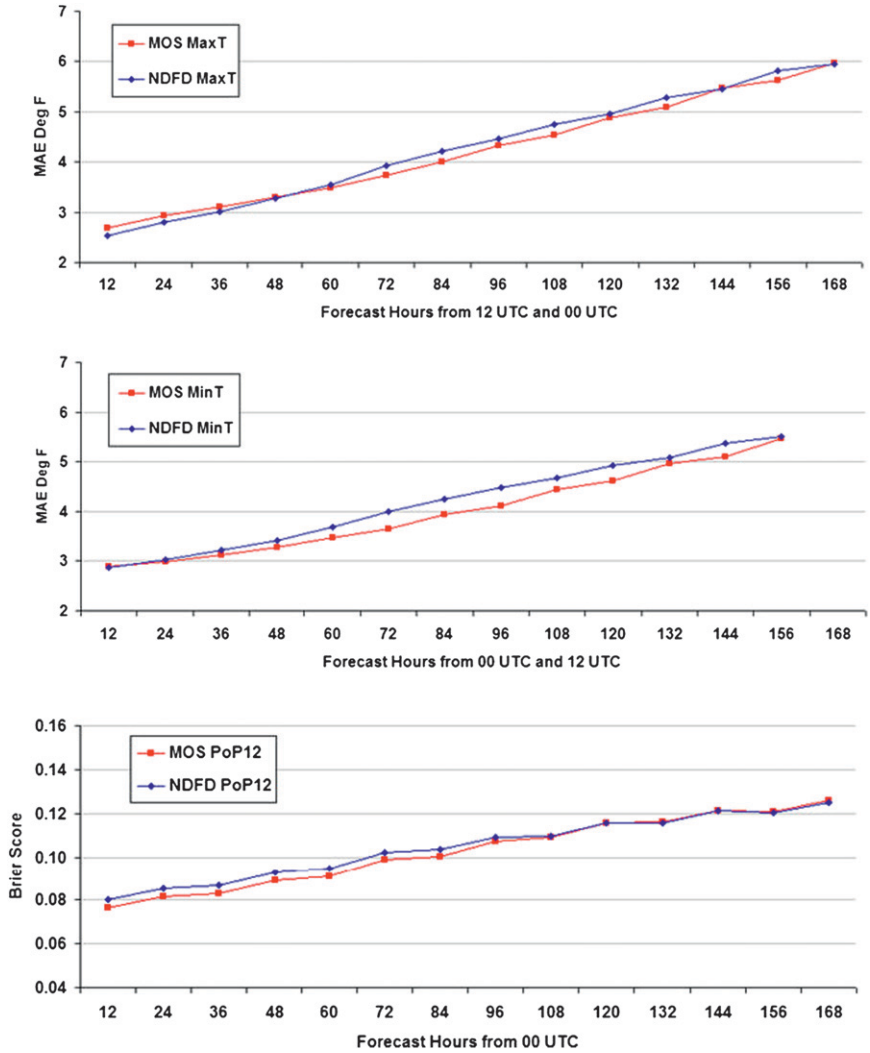

FIG. 6. Comparative accuracy of MOS and NDFD at about 1200 CONUS stations from October 2005 to September 2007 for (top) daytime maximum temperature, (middle) nighttime minimum temperature, and (bottom) PoP12. Forecast hours are in reference to NDFD issuance times.

corresponding MOS guidance forecasts at about 1200 locations in the CONUS on a monthly basis (Dagostaro et al. 2004). For the first time, the NWS could routinely verify many forecast elements at a wide range of projection times, and for a large number of locations.

Figures 6 and 7 compare the accuracy of GFS MOS (MOS developed on the Global Forecast System) at about 1200 CONUS stations to the NDFD forecast at the nearest 5-km grid point for daytime maximum temperature, nighttime minimum temperature, 12-h probability of precipitation, hourly temperature (T), and hourly dewpoint temperature (T_d) over a recent 2-yr period. Verification on matched case samples is performed for NDFD forecasts issued at 0000 and 1200 UTC with MOS guidance from the prior 1200 or 0000 UTC model cycle. For daytime maximum and nighttime minimum temperatures, scores for forecasts issued at 0000 and 1200 UTC are shown together. Only scores matched to the 0000 UTC NDFD forecast issuance are shown for other elements so as to preserve diurnal patterns. Other than this, scores matched to the 1200 UTC NDFD forecast issuance look very similar.

Forecasters can update NDFD at any hour of the day. Update habits vary by region and individual WFO. Typically, GFS MOS guidance becomes available to forecasters 4–5 h after model cycle time. WFOs routinely update NDFD forecasts 1–4 h prior to the 0000 and 1200 UTC issuances being verified. This provides a minimum of 3 h for NWS forecasters to consider forecast changes to NDFD based on the model guidance to

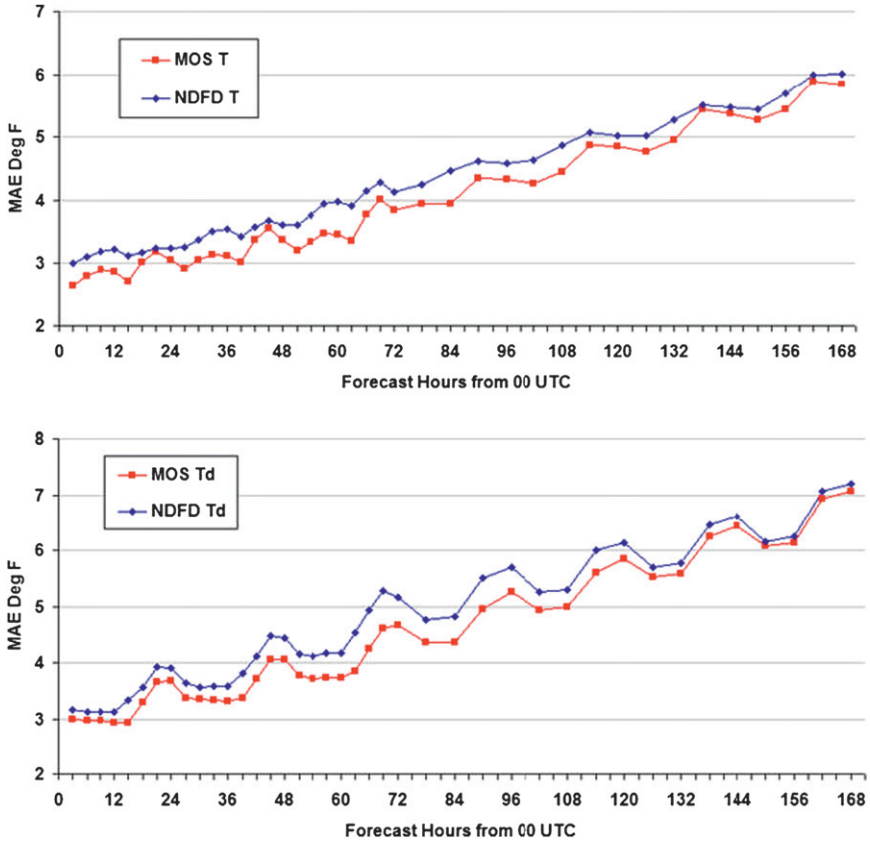

FIG. 7. Same as in Fig. 6, but for (top) hourly temperature and (bottom) hourly dewpoint temperature.

which NDFD is compared. Further updates to NDFD forecasts based on more recent model guidance or later observations can be included until shortly before the 0000 and 1200 UTC NDFD issuances. This is 7–8 h after the corresponding MOS guidance becomes available.

At some WFOs, updates to forecasts for days 4–7 are only made once per day based on forecast guidance from the NWS Hydrometeorological Prediction Center. As a result, NDFD day 4–7 forecasts issued at 0000 UTC are generally better when compared to MOS than are NDFD day 4–7 forecasts issued at 1200 UTC. This effect is evident as a slight wobble in NDFD MAEs beyond 96 h for daytime maximum and nighttime minimum temperatures (Fig. 6). A reverse phase wobble in MOS MAEs indicates that the improvement in MOS guidance from the 0000 UTC model cycle to the subsequent 1200 UTC model cycle is less than the improvement from the 1200 UTC model cycle to the subsequent 0000 UTC model cycle. This may be due to a 1-yr difference in the MOS development sample for these two model cycles.

Dallavalle and Dagostaro (2004) showed that the accuracy of MOS is approaching that of the local forecast for the early forecast periods. The charts presented here indicate that MOS can be a good source of guidance all the way out to day 7. The relatively inferior performance of NDFD for hourly temperature and dewpoint is likely the result of local tools used by WFOs to produce hourly values that are consistent with forecaster-edited maximum and minimum temperatures. When one element is adjusted, MOS guidance for the other can no longer be used directly. And while WFO forecasters almost certainly look at MOS maximum and minimum temperature guidance when they first submit a new forecast to NDFD 7 days in advance, the results here show that forecasters could benefit from a closer look at MOS guidance for these elements on days 3 and 4 as well.

Although one can get an idea for the relative strengths and weaknesses of MOS station guidance at various forecast hours from these charts, it would be incorrect to conclude that NDFD gridpoint forecasts are inferior to MOS. MOS forecasts are specific to the observing location, while NDFD forecasts represent conditions on a 5-km grid. This distinction may provide a significant advantage to MOS in regions of complex terrain in the western CONUS when the verification is based on station values.

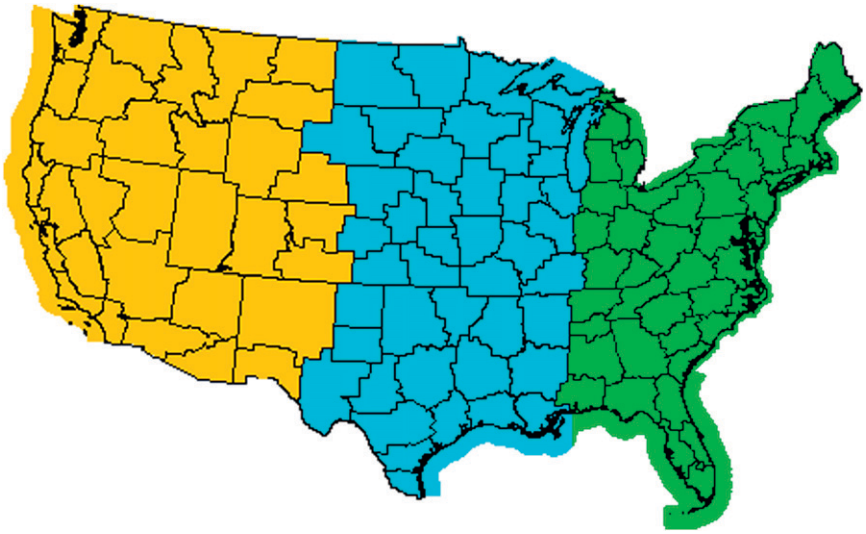

FIG. 8. Three CONUS regions grouped by WFO: West, Central, and East.

To see the distribution of improvements on MOS, scores from October 2006 to September 2007 were grouped by forecast days 1–3 (all projections) and forecast days 4–7 (all projections) according to three CONUS regions by WFO (see Fig. 8). Figures 9 and 10 show the percentage of WFOs and the number of months during the year that NDFD improved on MOS in each of these three regions. With the exception of PoP12, improvement on MOS is most common in the eastern and central CONUS on days 1–3, and in the central CONUS on days 4–7. For PoP12, the day 1–3 improvement is evenly distributed among WFOs nationwide while day 4–7 improvements are more common in the eastern CONUS. Improvement on MOS in the western CONUS is consistently less than that in the central and eastern CONUS except for day 1–3 PoP12.

An analysis of the magnitude of improvements for the same year (not shown) shows the most improvement (up to 26%) on day 1–3 daytime MOS maximum temperatures at WFOs located in the central plains. For nighttime minimum temperature, the largest improvements (up to 8%) for days 1–3 appear at WFOs in the southern Appalachians. For daytime maximum and nighttime minimum temperatures, the least improvement (down to −31%) on the day 1–3 MOS appears at WFOs in the mountainous West. The largest improvements on MOS PoP12 Brier scores (up to 17%) are on days 4–7 in the southern Appalachian Mountains. NDFD improvement on MOS PoP12 tends to be worse on days 1–3 due to relatively poor NDFD Brier scores during the three winter months in the northeast CONUS.

In light of recognized problems with point verification, MDL also provides NWS forecasters with NDFD verification based on real-time mesoscale analysis (RTMA) grids (De Pondeca et al. 2007). These gridded scores are meant to complement NDFD scores at observation points. However, the quality of these data does not yet appear to be sufficient to support comparative verification studies. In the future, the NWS plans to create an official analysis of record with additional quality control (Horel and Colman 2005). With gridded verification, doubts are directed at the representativeness of the analysis rather than the representativeness of values at observation points. For both methods, questions concerning representativeness will certainly persist.

4. Gridded MOS for NDFD

With the advent of NDFD, WFOs needed guidance that contained the high-resolution terrain features that forecasters want depicted in grids for their local area. In the fall of 2006, MDL began producing gridded MOS (GMOS) guidance for NDFD elements and forecast projections on the NDFD grid. For most elements, gridded MOS is created by analyzing all available MOS station forecasts (Glahn et al. 2009). For daytime maximum and nighttime minimum temperatures, MOS forecasts for about 8000 observation sites are available including METAR (aviation routine weather report, translated roughly from French), mesonet, and cooperative observation locations. For hourly temperatures and dewpoints, about 3000 stations are used including METAR and mesonet observations. For other NDFD

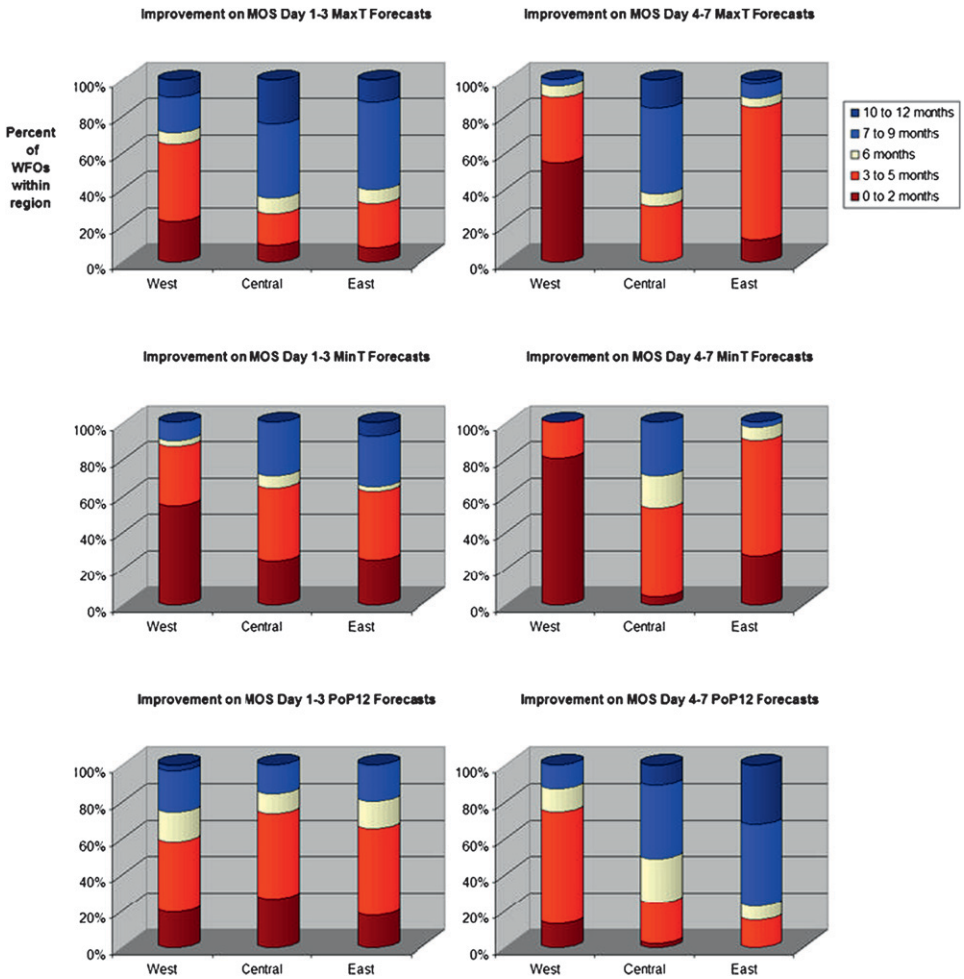

FIG. 9. The percent of WFOs and the number of months from October 2006 to September 2007 that NDFD improved (lower MAE or Brier score) on MOS (top) daytime maximum temperature, (middle) nighttime minimum temperature, and (bottom) PoP12 forecasts in three CONUS regions shown in Fig. 8: (left) 1–3 day and (right) 4–7 day forecasts.

elements, about 1640 METAR sites are used. It has been shown that MOS provided reasonable guidance at MOS points, but how good was the new guidance relative to the local NDFD forecast in areas away from MOS stations?

To determine this, MDL conducted a 10-month special study that examined the relative performance of NDFD and GMOS for daytime maximum, nighttime minimum, and dewpoint temperatures for days 4–7. These days were chosen because of interest by NWS management in having forecasters rely more on guidance for these projections. The study used 217 MOS stations and 121 non-MOS stations in remote areas of the western CONUS not normally included in national verification. Stations were recommended for use in the study by forecasters at local WFOs. The western CONUS was selected because of the complex terrain there. MOS station guidance was not available to either local forecasters or the GMOS analysis for the 121 non-MOS observing sites. GMOS was produced retrospectively for this study, and also was not available to local forecasters. For all stations, NDFD and GMOS forecasts for the grid point nearest to the observing site were verified against reported observations.

Figure 11 shows a comparison of NDFD and GMOS scores at the MOS and non-MOS stations. The results are described in detail by Schenk (2006). In summary, the study found that 1) errors both for GMOS and NDFD were 1°–1.5°F greater for the non-MOS sites than the MOS sites, 2) GMOS forecasts had smaller errors than NDFD at both MOS sites and non-MOS sites, and 3) the differences between NDFD and MOS errors at non-MOS sites were nearly the same as at MOS sites.

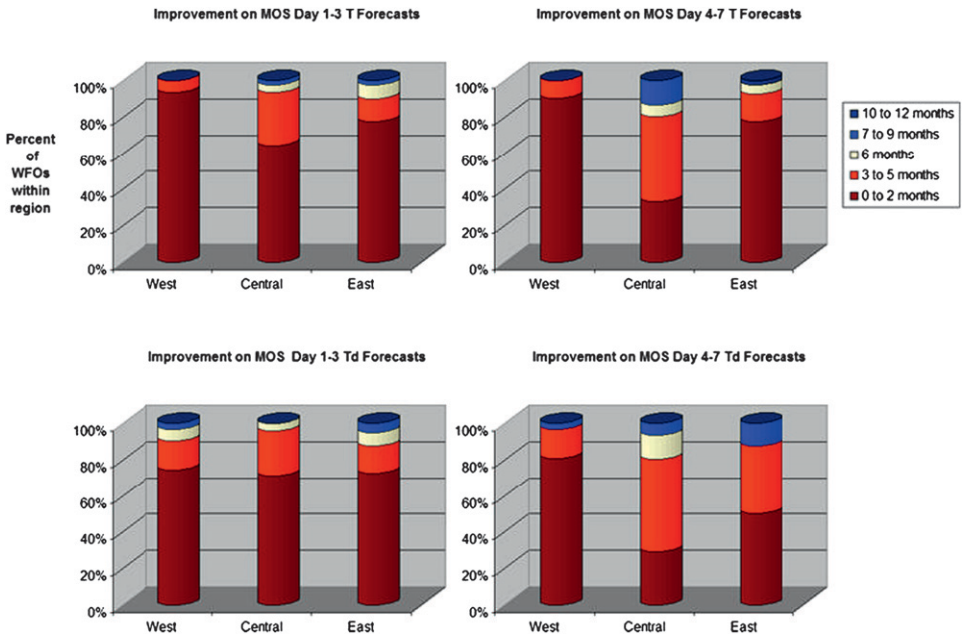

FIG. 10. Same as in Fig. 9, but for hourly temperature and hourly dewpoint temperature.

Since the fall of 2006, MDL has included GMOS as part of our routine monthly NDFD verification (Dagostaro et al. 2004). Over the past year, GMOS and MOS scores for most elements have been nearly identical. In contrast, GMOS scores for daytime maximum and nighttime minimum temperatures (based on approximately 8000 MOS stations) were about the same as station MOS in areas of smooth terrain, but were worse than scores for station MOS in areas of complex terrain. This result is consistent with our earlier finding that the approximately 1200 stations routinely used for NDFD verification do not adequately represent the spatial detail that NDFD, and now GMOS, provide in areas of complex terrain.

Figure 12 shows stations used for monthly NDFD verification compared to the total number of MOS stations included in the GMOS analysis for daytime maximum temperature. Poorer scores for GMOS can result when multiple nearby MOS stations influence the verification site's gridpoint forecast. In the cases of maximum and minimum temperatures, MOS forecasts for cooperative observing sites are likely inferior to forecasts for the METAR sites that are used for forecast verification. Performance can suffer from less reliable observations available for MOS development, as well as ambiguities concerning whether the maximum or minimum temperature occurred during the appropriate daytime or nighttime period at sites that only observe once daily. In areas of complex terrain, apparent errors in GMOS often show a warm or cool bias based on the elevation of the verification site compared to nearby sites. In areas where terrain is not an overriding issue, the use of neighboring MOS stations in the GMOS analysis generally makes the forecast a degree or two

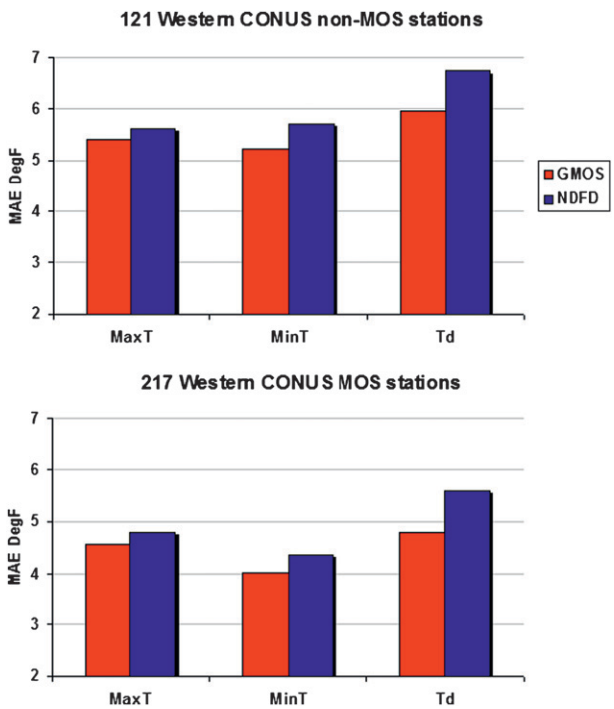

FIG. 11. MAE for day 4–7 forecasts from December 2004 through September 2005 at (top) non-MOS and (bottom) MOS stations in the western CONUS.

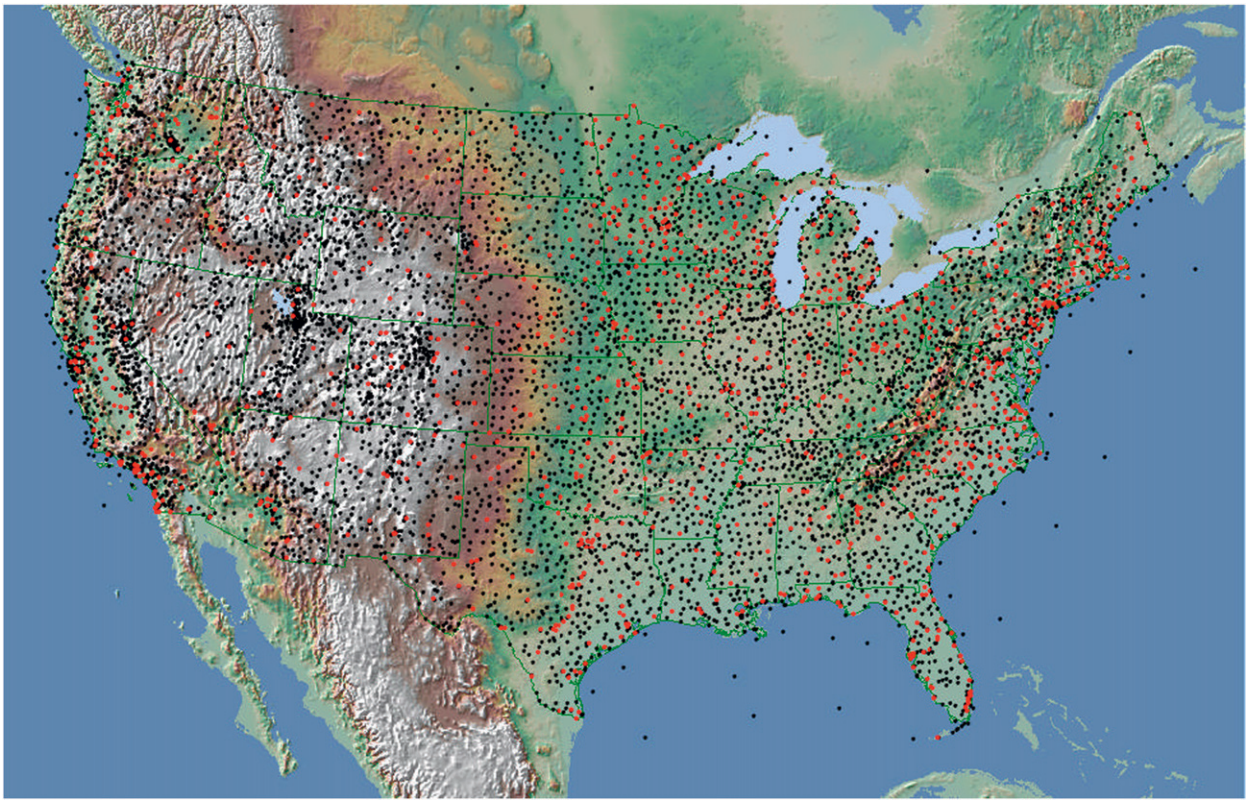

FIG. 12. Locations of approximately 1200 MOS stations used for NDFD monthly verification (red dots) compared to approximately 8000 MOS stations used by GMOS analysis for daytime maximum and nighttime minimum temperatures (red and black dots).

cooler than MOS. This is because the approximately 1200 sites used for NDFD verification tend to be located in less remote areas. In months that MOS has a warm bias, this makes the GMOS forecast appear better. In months that MOS has a cool bias, it makes the GMOS forecast appear worse. MDL does not routinely verify at all MOS sites because of delays in obtaining verifying observations, and forecaster concerns about observation quality. In fact, at the request of field forecasters, many stations that are available have been dropped from our monthly verification.

Figures 13 and 14 show the NDFD improvement on GMOS for the same year shown for the NDFD improvement on MOS in Figs. 9 and 10. NDFD improvements on 12-h probability of precipitation, hourly temperature, and hourly dewpoint temperature are nearly the same for both MOS and GMOS. On the other hand, the improvement on GMOS scores for daytime maximum and nighttime minimum temperatures is better—significantly so in the West.

If all MOS stations used in the creation of GMOS were verified, or if the current GMOS analysis were produced and verified on a finer-resolution grid, one would expect the results for station MOS and gridded MOS to be closer. However, the results would likely not be the same unless the area of influence an individual station can have in the GMOS analysis were significantly reduced.

5. Forecast continuity

In addition to producing accurate guidance that reflects high-resolution terrain, forecasts should provide good continuity from when the guidance is first issued 7 days in advance, until it is last updated on forecast day 1. Forecasters at NWS offices almost always take into consideration previously issued forecasts (Lashley et al. 2008). Model forecasts from adjacent cycles, however, are not tied to previous forecasts (beyond an initial first-guess analysis), and have been known to ''flip-flop'' as a result. To quantify this, MDL developed an index that measures the number of significant swings made over a series of forecast cycles for forecasts valid at the same time. This index is known as the Ruth–Glahn forecast convergence score (FCS).

When considering n forecasts made over a number of days for subsequent forecast cycles that decrease in forecast projection until the valid time of the forecast is reached, the FCS is defined as follows:

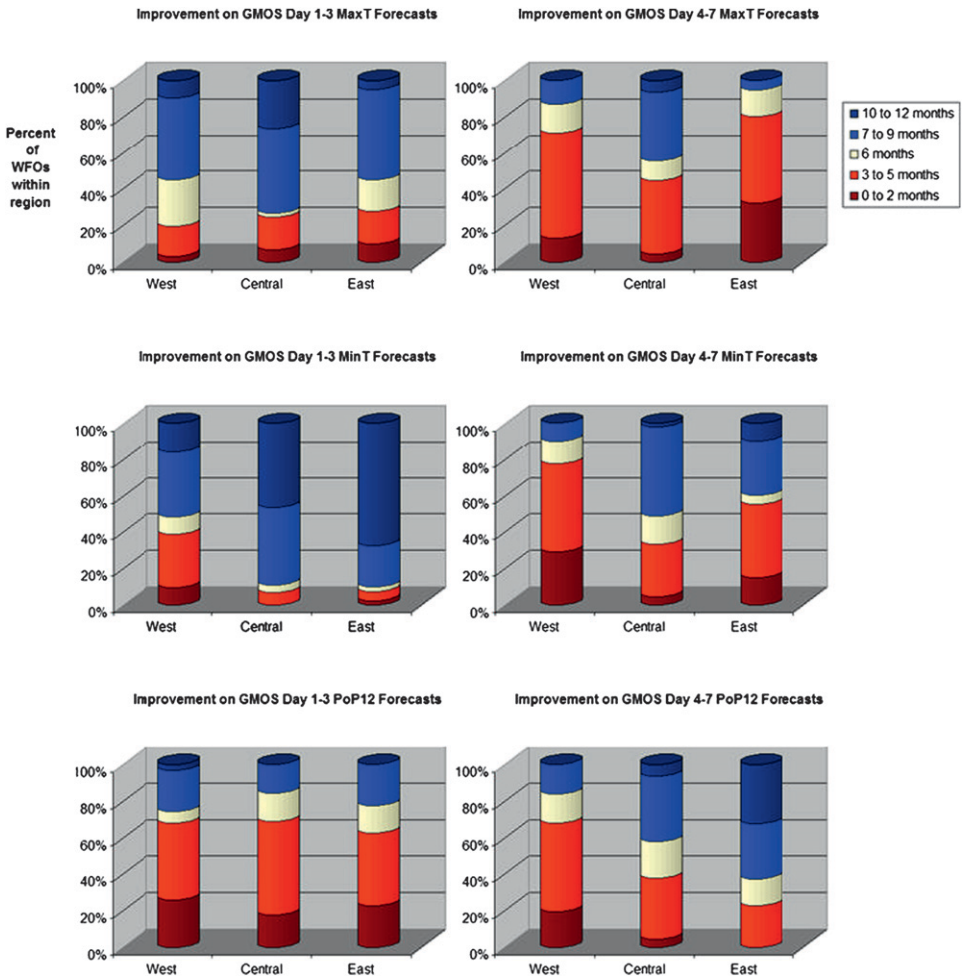

FIG. 13. Same as in Fig. 9, but for improvement relative to GMOS rather than MOS.

$$\mathrm{FCS} = (T_1 + T_2) / (T_3 + T_4).$$

The first term (T_1) is the number of forecasts that changed insignificantly (less than a threshold) from the previous forecast F_{i-1} *or* that moved closer to the next forecast F_{i+1}, where i varies from 2 to n. When $i = n$, the observation is used as the next forecast F_{i+1}:

$$T_1 = \sum_{i=2}^{n} \begin{cases} 1 \text{ if } |F_i - F_{i-1}| < \text{ threshold} \\ 1 \text{ if } |F_i - F_{i+1}| < |F_i - F_{i-1}|. \\ 0 \text{ if neither of above} \end{cases}$$

The second term (T_2) is the difference between the first and last forecasts scaled by the significance threshold:

$$T_2 = |F_n - F_1| \,/\, \text{threshold}.$$

The third term (T_3) is the number of possible forecast changes:

$$T_3 = n - 1.$$

The fourth term (T_4) is the sum of the forecast changes scaled by the significance threshold:

$$T_4 = \sum_{i=2}^{n} |F_i - F_{i-1}| \,/\, \text{threshold}.$$

The T_1 and T_3 terms account for the actual and possible numbers of swings, respectively. The T_2 and T_4 terms account for the magnitudes of the swings. The significance threshold specifies the minimum change necessary to count as a swing.

The FCS ranges from near 0 (many large swings away from the next forecast) to 1.0 (no swings). Repeatedly forecasting the same value (e.g., a climatic normal) will yield a perfect FCS of 1.0. The score does not measure forecast accuracy; accuracy is measured by other scores, such as MAE.

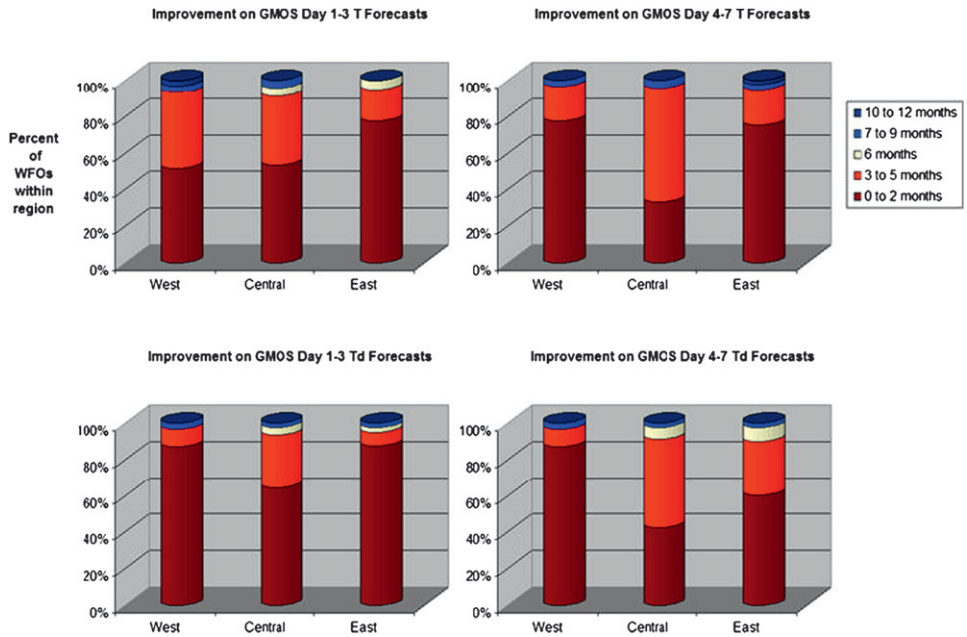

FIG. 14. Same as in Fig. 10, but for improvement relative to GMOS rather than MOS.

Table 1 provides an example FCS calculation in which we compare 14 forecasts issued at 12-h intervals over a period of 7 days. The case was chosen to include the forecast shown in Figs. 1 and 2 for a station in each of the regions in Fig. 8. These results demonstrate that MOS suffers from large swings away from the next forecast while NDFD and GMOS are much more consistent from issuance to issuance. GMOS continuity that is as good (or better) than NDFD continuity is achieved by averaging the most recent 0000 and 1200 UTC cycles

TABLE 1. FCS calculation for daytime maximum temperature forecasts valid 30 Apr 2007 at Salt Lake City, UT (KSLC); Chicago, IL (KORD); and Boston, MA (KBOS). Forecasts (F_i) for fourteen sequential issuances ($n = 14$) and the observed condition at the forecast valid time (Obs) are shown. A significance threshold of 3°F was used. Forecasts that failed the T_1 test are shown in boldface.

		KSLC			KORD			KBOS		
NDFD issuance	Forecast	NDFD	MOS	GMOS	NDFD	MOS	GMOS	NDFD	MOS	GMOS
0000 UTC 24 Apr 2007	F_1	79	74	76	67	72	69	66	62	62
1200 UTC 24 Apr 2007	F_2	76	72	75	67	73	72	66	64	64
0000 UTC 25 Apr 2007	F_3	76	77	77	70	71	72	70	**58**	62
1200 UTC 25 Apr 2007	F_4	74	77	79	74	77	75	70	64	64
0000 UTC 26 Apr 2007	F_5	80	**83**	81	74	78	77	65	**70**	70
1200 UTC 26 Apr 2007	F_6	80	79	80	78	78	78	65	64	68
0000 UTC 27 Apr 2007	F_7	81	79	81	78	**82**	80	67	66	67
1200 UTC 27 Apr 2007	F_8	81	81	82	84	78	78	67	**62**	65
0000 UTC 28 Apr 2007	F_9	82	78	81	82	**74**	75	64	64	65
1200 UTC 28 Apr 2007	F_{10}	81	78	80	80	76	75	67	67	66
0000 UTC 29 Apr 2007	F_{11}	78	78	78	79	**72**	74	67	61	64
1200 UTC 29 Apr 2007	F_{12}	78	80	79	76	**76**	75	67	61	63
0000 UTC 30 Apr 2007	F_{13}	79	82	81	76	72	74	68	63	64
1200 UTC 30 Apr 2007	F_{14}	80	81	81	72	73	73	68	65	64
	Obs	82	82	82	67	67	67	71	71	71
	T_1	13.00	12.00	13.00	13.00	9.00	13.00	13.00	10.00	13.00
	T_2	0.33	2.33	1.67	1.67	0.33	1.33	0.67	1.00	0.67
	T_3	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00
	T_4	6.33	9.00	5.67	9.67	12.33	6.67	6.00	15.67	7.33
	FCS	0.69	0.65	0.79	0.65	0.37	0.73	0.72	0.38	0.67

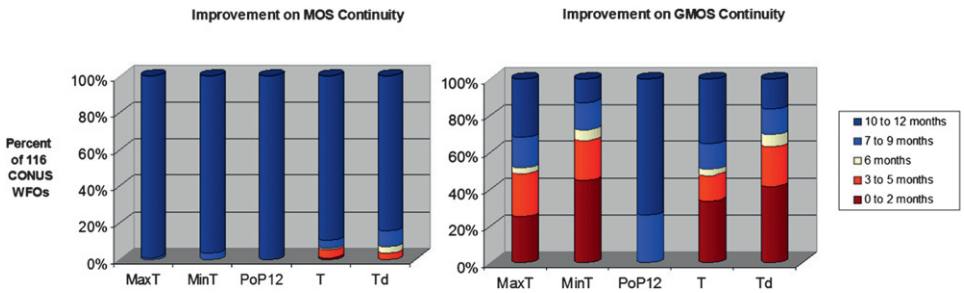

FIG. 15. The percent of NWS WFOs and the number of months from October 2006 to September 2007 that NDFD showed better forecast continuity than (left) MOS and (right) GMOS in the CONUS. Significance thresholds were 3°F for temperatures and 40% for PoP12.

of station MOS in the production of GMOS. Tests show that cycle averaging makes little difference in the accuracy of the GMOS forecast.

Figure 15 shows the percentage of WFOs and number of months during the year that NDFD improved on the MOS FCS and the GMOS FCS. The score for each element was computed by using forecasts issued at 12-h intervals (0000 and 1200 UTC) over a period of 7 days. GMOS forecast continuity over the past year appears to be worse for PoP12 because MDL only introduced the use of two cycles of station MOS for this element in June 2007. Cycle averaging for other GMOS elements was performed from the start.

Figure 16 shows monthly forecast convergence scores for the daytime maximum temperature. If the same significance threshold is used throughout the year, temperature scores will be slightly higher in summer than in winter. Similarly, if the same significance threshold is used throughout the CONUS, scores will generally be higher in the South than in the North. Although adjustments to significance thresholds will change the values of the scores, our tests show that the relative performance levels of NDFD, MOS, and GMOS remain very much the same. The increase in forecast continuity scores for GMOS daytime maximum temperature seen in June 2007 (Fig. 16) resulted from the implementation of equal weights for the 0000 and 1200 UTC MOS guidance cycles. In the initial implementation of GMOS, the previous cycle was only given half the weight of the current cycle.

6. Conclusions

As the NWS has moved from the preparation of forecasts as text to the creation and dissemination of digital data, MDL has adapted MOS guidance to meet the changing needs for NWS forecasters at WFOs. As best as we can determine using a limited set of stations for verification, the new GMOS provides guidance that is comparable in accuracy to official NDFD forecasts in those areas of the country where station observations well represent the grid. GMOS also appears similar to NDFD in the depiction of complex terrain and provides forecast continuity that is as good as NDFD from day 7 through day 1.

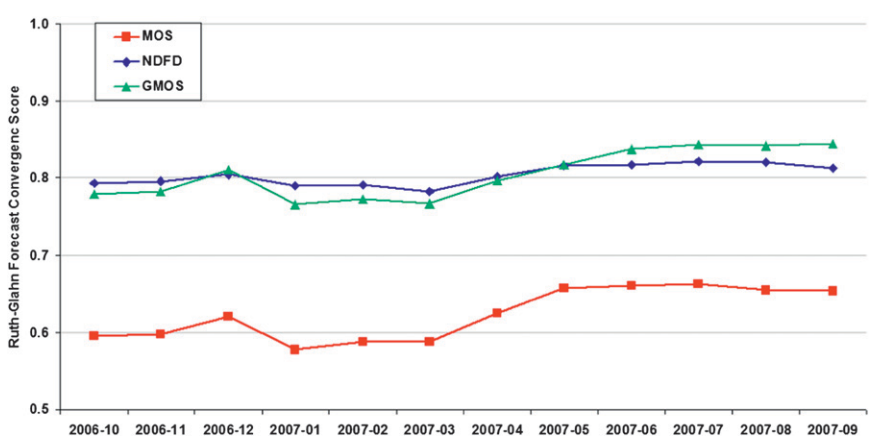

FIG. 16. Ruth–Glahn FCSs for daytime maximum temperature in the CONUS calculated with a significance threshold of 3°F.

Acknowledgments. NDFD verification depends on the dedicated efforts of the persons in the Evaluation Branch of MDL. Station scores used to update Figs. 3–5 were extracted from an interactive database maintained by the Performance Branch of the NWS Office of Climate, Water, and Weather Services. Tim Kempisty downloaded monthly verification scores from the NDFD central server; Kelly Malone provided valuable assistance in handling several very large spreadsheets of MOS, NDFD, and GMOS verification scores; and Kari Sheets furnished the station map for Fig. 12. This paper is the responsibility of the authors and does not necessarily represent a position of the National Weather Service or any other governmental agency.

REFERENCES

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

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.

Dagostaro, V. J., W. A. Shaffer, M. J. Schenk, J. L. Gorline, and A. A. Taylor, 2004: A prototype verification system for examining NDFD forecasts. Preprints, *17th Conf. on Probability and Statistics in the Atmospherics Sciences,* Seattle, WA, Amer. Meteor. Soc., 2.7. [Available online at http://ams.confex.com/ams/pdfpapers/73589.pdf.]

Dallavalle, J. P., and V. J. Dagostaro, 2004: Objective interpretation of numerical prediction model output—A perspective based on verification of temperature and precipitation guidance. *Symp. 50th Anniversary of Operational Numerical Weather Prediction,* College Park, MD, NCEP, 5.8. [Available online at http://www.nws.noaa.gov/mdl/pubs/Documents/Papers/JNWPUpaper_200404.pdf.]

——, and B. Glahn, 2005: Toward a gridded MOS system. Preprints, *21st Conf. on Weather Analysis and Forecasting/17th Conf. on Numerical Weather Prediction,* Washington, DC, Amer. Meteor. Soc., 13B.2. [Available online at http://ams.confex.com/ams/pdfpapers/94998.pdf.]

De Pondeca, M. S. F. V., and Coauthors, 2007: The development of the real time mesoscale analysis system at NCEP. Preprints, *23rd Conf. on Interactive Information Processing Systems,* San Antonio, TX, Amer. Meteor. Soc., P1.10. [Available online at http://ams.confex.com/ams/87ANNUAL/techprogram/paper_118813.htm.]

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.

——, and D. P. Ruth, 2003: The New Digital Forecast Database of the National Weather Service. *Bull. Amer. Meteor. Soc.,* **84,** 195–201.

——, K. Gilbert, R. Cosgrove, D. P. Ruth, and K. Sheets, 2009: The gridding of MOS. *Wea. Forecasting,* **24,** 520–529.

Horel, J., and B. Colman, 2005: Real-time and retrospective mesoscale objective analyses. *Bull. Amer. Meteor. Soc.,* **86,** 1477–1480.

Lashley, S. L., L. Fisher, B. J. Simpson, J. Taylor, S. Weisser, J. A. Logsdon, and A. M. Lammers, 2008: Observing verification trends and applying a methodology to probabilistic precipitation forecasts at a National Weather Service forecast office. Preprints, *19th Conf. on Probability and Statistics,* New Orleans, LA, Amer. Meteor. Soc., 9.4. [Available online at http://ams.confex.com/ams/pdfpapers/134204.pdf.]

Ruth, D. P., 2002: Interactive forecast preparation—The future has come. Preprints, *Interactive Symp. on the Advanced Weather Interactive Processing System (AWIPS)*, Orlando, FL, Amer. Meteor. Soc., 9.4. [Available online at http://ams.confex.com/ams/pdfpapers/4204.pdf.]

——, and C. L. Alex, 1987: AFOS–ERA forecast verification. NOAA Techniques Development Laboratory Computer Program NWS TDL CP 87-2, 50 pp. [Available from NWS Meteorological Development Laboratory, 1325 East–West Highway, Silver Spring, MD 20910-3283.]

Schenk, M., 2006: Performance of three forecast systems at selected MOS and non-MOS stations. MDL Office Note 06-01, NOAA/NWS, 9 pp. [Available online at http://www.nws.noaa.gov/mdl/pubs/Documents/OfficeNotes/Schenk2006Performance.pdf.]