Eastern Region Technical Attachment
No. 2004-01
January 2004

IMPROVING TEMPERATURE VERIFICATION RESULTS WITHIN THE IFPS/GFE FRAMEWORK

George J. Maglaras
NOAA/National Weather Service
Albany, New York

1. INTRODUCTION

Since September, 2002, verification scores for 12-hour maximum/minimum (MAX/MIN) temperature forecasts at the National Oceanic and Atmospheric Administration (NOAA)'s National Weather Service (NWS) Weather Forecast Office (WFO) at Albany, NY (ALY) have shown a dramatic improvement when compared to Model Output Statistics (MOS) guidance forecasts (Glahn and Lowry 1972). This improvement occurred over a very short period of time, and was the result of well defined changes to forecast methodologies and operational procedures. This paper will discuss the steps taken to achieve the improvement in the verification scores, and how they occurred within the framework of the transition to an entirely new forecast process.

2. BACKGROUND

During the past three years, NWS WFOs in the Eastern Region (ER) have made the transition from manually creating and editing text forecast products, to maintaining a constantly updated digital forecast data base (Glahn and Ruth 2003) from which a variety of forecast products are generated. Before the transition began, the primary software tools used to manually create and edit text forecast products were relatively simple word-processing software packages. During the transition, forecasters learned to use the Grid Forecast Editor (GFE) and the Interactive Guidance Revisor (IGR) of the Interactive Forecast Processing System (IFPS) to grid edit individual weather element forecast fields, edit data matrices, and then generate automated text and graphical forecast products. For WFO ALY forecasters, this transition has meant a strong focus on learning to use these new and complex software packages, while still producing all of the previous forecast products. At times during this three year transition period, the accuracy of the forecast values used to populate the gridded fields was compromised as forecasters learned to create and modify the grid fields, and produce various forecasts using the new software. As a result, temperature verification results at WFO Albany remained stagnant, despite improved model and MOS forecasts being made available to the forecasters.

Starting in December 2000, WFO Albany 12-hour MAX/MIN temperature forecasts were compared to the Global Forecast System (GFS) Model (Kanamitsu 1989) based MOS (MAV) guidance (National Weather Service

2000) for Albany (ALB), for the first time. In addition, WFO Albany forecasts were compared to the Nested-Grid Model (NGM; Hoke et al. 1989) based MOS (FWC) guidance (National Weather Service 1992), as had been done since the early 1990s. The WFO Albany temperature (WFO) forecasts are the 12-hour MAX/MIN forecasts for the first five periods from the coded cities forecast (CCF) product. Periods one through five are defined as 12-24, 24-36, 36-48, 48-60, and 60-72 hours after 0000 or 1200 UTC, respectively. For each period, the mean absolute error (MAE) score was calculated for WFO ALY, MAV and FWC forecasts. Based on the MAE, the performance of WFO ALY forecasts were determined by calculating the percent improvement over the guidance. Initially, for the first few months after December 2000, the results for ALB indicated percent improvements over the FWC guidance by WFO ALY forecasts that were similar to previous years. However, WFO ALY forecasts for ALB were considerably worse than the MAV guidance, especially for fifth period forecasts.

Also beginning in December 2000, WFO ALY forecasts were compared to MAV and FWC forecasts for Glens Falls (GFL), NY, and for Poughkeepsie (POU), NY, for the first time. Similarly, WFO ALY forecasts were considerably worse than the MAV forecasts for GFL and POU. In addition, the results showed little or no improvement over the FWC guidance by WFO ALY forecasts at these two stations.

Over time, WFO ALY forecasts were expected to improve in comparison to the MAV and FWC forecasts for all three stations. Maglaras (1999) showed that when MOS forecasts from a new and improved model first become available, WFO ALY forecasters initially have difficulty improving upon the new MOS forecasts, but within two years, forecasters should be doing better than the new guidance on a regular basis. For this study, the MAV MOS guidance represents the MOS forecasts from a new and improved model. In addition, once forecasters began doing better than the MAV MOS guidance, their improvement over the FWC guidance was also expected to increase since the MAV forecasts were routinely better than the FWC forecasts for ALB, GFL and POU. However, after almost two years, WFO forecast improvement over the MAV and FWC guidance remained virtually unchanged. Table 1 shows the percent improvements of WFO ALY forecasts over MAV and FWC guidance for ALB, GFL and POU, for the 21-month period from December 2000, through August 2002. Also shown (in parenthesis) are the percent improvements for July and August 2002. The verification results for July and August 2002, were similar, if not worse than the results for the entire 21-month sample. In addition, the results for July and August 2002, were similar to the verification results (not shown) for all other two-month samples during this 21-month period. As a result, after almost two years, there was no indication of an improvement trend in WFO ALY forecasts.

There are several possible reasons WFO ALY forecasts did not improve, when compared to MOS guidance, as they would have in the past. The causes are probably a combination of human considerations, technological limitations, and changes to operational procedures to create forecast products. First, the introduction of new technology (multiple versions of IFPS/GFE/IGR software), and constantly changing procedures to create and disseminate forecast products, including many new forecast products, caused many forecasters to spend considerable time learning how to use the new technology and new procedures. As a result, verification became secondary in importance.

Second, even though forecasters had the MAV guidance available to them during the 21-month period, the GFE software did not support direct ingest of these data until the spring of 2002. Even when the MAV guidance became available in GFE, the default MOS station-to-grid mapping method available used a linear interpolation approach, which was not suitable to populate a grid field within an area of varied terrain. As a result, forecasters were unwilling to use this grid field as it existed much of the time.

Third, and probably the most important reason, for IFPS/GFE operations, the prevailing philosophy had been to use the official/current temperature forecast grid field as a starting point in the forecast process, and to make modifications only when significant changes were needed. This philosophy stemmed from the labor intensive nature of looking at and modifying hundreds of grid fields, as well as forecaster inexperience with using the IFPS/GFE system. In addition, the early versions of IFPS featured relatively immature software, with few Smart Tools (GFE software code that performs operations on a grid-by-grid basis) and procedures available to make populating and modifying the grid fields more efficient. All these factors resulted in a much longer forecast preparation time. Although this philosophy was necessary to help forecasters manage the much longer IFPS/GFE forecast preparation time, this philosophy also resulted in temperature forecasts for longer-term periods transitioning into shorter-term periods with little if any modification. This was especially noticeable with forecasts for the sixth period (72-84 hours after 0000 or 1200 UTC) or beyond, transitioning into the fifth period. The fifth period MAV forecasts were far better than any method that could have been used to populate sixth period forecasts from 12-hours earlier, and these sixth period forecasts were frequently transitioning into the fifth period with little if any modification. As a result, as shown in Table 1, WFO ALY forecasts for the fifth period averaged 24.3 percent worse than the MAV guidance.

Even for periods one through four, using the official/current temperature forecast grid fields as your starting point, and not populating with the latest MAV guidance, can have a negative impact on verification scores. Based on MAE scores for ALB from December 2000 through April 2003, on average, MAV MOS guidance forecasts for a particular forecast period were 6.4% worse than for the earlier forecast period. For example, third period MAV forecasts averaged 6.4% worse than second period MAV forecasts, and 12.8% worse than first period MAV forecasts. Thus, over a long period of time, if MAV MOS guidance from a particular model run is used to populate third period temperature forecasts, but on a regular basis these forecasts are allowed to transition into second (first) period forecasts with little or no modification, then WFO ALY forecasts could be as much as 6.4% (12.8%) worse when compared to the second (first) period MAV forecasts from later model runs.

The final reason why WFO ALY forecasts may not have improved when compared to MOS guidance, may have to do with the preparation of the Coded Cities Forecasts (CCF) product. For the forecaster, preparing the CCF product is the last step in the forecast preparation process in terms of verification. The temperature forecasts entered into the CCF for ALB, GFL and POU are also used for local verification. Unfortunately, the early versions of IFPS did not select the temperature forecast value for a specific CCF station from the appropriate grid point. Instead, the average temperature value from a zone group was used for all CCF stations within that zone group. Forecasters could manually modify the CCF forecasts if they wished, but during much of this transition process, this was not considered a priority. In addition, during the transition to IFPS/GFE, the number of CCF stations that WFO Albany forecasters needed to prepare forecasts for, increased from one station with five forecast periods, to 13 stations with 14 forecast periods. Thus, manually editing CCF forecasts also became a labor intensive task.

3. STEPS TAKEN TO IMPROVE TEMPERATURE FORECAST VERIFICATION

Based on the verification statistics in Table 1, the MAV MOS MAX/MIN temperature forecasts would have been the best statistical guidance available for populating GFE temperature grid fields near combined CCF/MAV MOS sites. However, as already stated, for the 21-month period used in Table 1, these data were either not available within GFE, or the grid field produced was not suitable for an area with varied terrain.

In order to provide an improved MAV MOS temperature grid field that would be of greater utility to the forecasters, the labor intensive task of manually configuring a new MOS station-to-grid mapping method was begun late in the spring of 2002, and completed in August of 2002. The new MOS station-to-grid mapping method was developed by manually determining on a grid point-by-grid point basis, which MAV MOS forecasts would be used and the percentage each of these MAV MOS forecasts would contribute to the calculation of the temperature forecast for a particular grid point. The mapping was done for other weather elements and for other types of MOS guidance as well to facilitate initial grid population. At the beginning of September, 2002, the forecast staff at WFO Albany was informed that the MAV MOS was now available for populating GFE temperature grid fields in a useable form. In addition, the forecast staff was advised that the recommended procedure for populating GFE MAX/MIN temperature grid fields had been changed. The new procedure was to repopulate the temperature grids with MAV MOS for the first five periods whenever new guidance became available, and to use the repopulated MAV MOS grid fields as the starting point for further modification. Previously, it was recommended that forecasters use the official/previous grid forecast as their starting point when preparing a new forecast package.

These configuration/procedural changes resulted in a rapid and dramatic increase in percent improvement over guidance scores by WFO ALY forecasts. Table 2 shows the percent improvement over MAV and FWC guidance by WFO ALY forecasts, for ALB, GFL and POU, for the two-month period after September 1, 2002, and for the two-month period before September 1, 2002, in parenthesis. Table 2 indicates that for the two months after September 1, 2002, WFO ALY forecasts were now better than the MAV guidance for ALB and GFL. At POU, WFO ALY forecasts were still worse than the MAV guidance, but they had improved considerably. The most dramatic change occurred for fifth period forecasts. Table 2 shows that for the two months before September 1, 2002, WFO ALY forecasts for the fifth period averaged around 30% worse than the MAV guidance for all three stations combined. For the two months after September 1, 2002, WFO ALY forecasts averaged only about 3% worse. In addition, there was a substantial increase in WFO forecast percent improvement over FWC guidance in the two months after September 1.

4. LATEST RESULTS

Table 3 shows the latest verification results for September 1, 2002, through May 4, 2003, as well as the results for the 21-month period before September 1, 2002, in parenthesis. In general, the results in Table 3 indicate that WFO ALY forecasts continue to increase their percent improvement over MOS guidance. For the eight months after September 1, 2002, WFO ALY forecasts were 2.0% and 2.6% better than MAV guidance at ALB and GFL, respectively, for all five periods combined. For GFL, this represents a nearly 15% increase compared to the previous 21 months. At POU, WFO ALY forecasts remain poorer than the MAV guidance, but the percent improvement over MAV guidance scores have increased by around 7% compared to the previous 21 months.

Table 3 shows that the increase in WFO forecast percent improvement over FWC guidance has been even greater. For all four periods combined, the percent improvement over FWC guidance was 18.9%, 14.4% and 9.9% for ALB, GFL and POU, respectively, for the eight month sample. These percent improvements over guidance represent an increase of 11% to 15% compared to the 21-month sample. Note, in particular, that the percent improvement over guidance for ALB is now approaching 20%. WFO ALY forecasts have been compared to FWC guidance for more than a decade. Previously, any improvement over guidance that exceeded 10% for a six-month cool/warm season would have been considered exceptional.

5. DISCUSSION AND CURRENT OPERATIONS

WFO ALY forecasts have improved dramatically during the past year, when compared to MOS guidance. This improvement has been accomplished within the framework of preparing gridded and text forecast products using IFPS/GFE. When WFO ALY began the transition to IFPS/GFE, initial verification results showed that WFO ALY forecasts performed poorly when compared to MOS guidance, with no improvement trend noted over the subsequent 21 months. However, once a new MAV MOS station-to-grid point mapping method was implemented around September 1, 2002, and the office recommended procedure for populating MAX/MIN temperature forecasts was changed, the performance of WFO ALY forecasts when compared to MOS guidance improved dramatically over a short period of time.

The new MAV MOS station-to-grid mapping method implemented around September 1, 2002, produced GFE temperature grid fields which showed a smoothed representation of the major terrain features within WFO ALY's forecast area. Features such as the Adirondack, Catskill, Southern Green, Taconic and Berkshire Mountains, and the Hudson and Mohawk Valleys, were clearly evident. The smoothed representation of the major terrain features was acceptable at first. However, as the transition to gridded forecast products continued (current grid resolution is 5.0km), more terrain enhancement was needed.

In order to produce a temperature grid field that could represent individual mountain peaks or ridges, along with their associated valleys, WFO ALY forecasters used a locally developed lapse rate smart tool. This smart tool used lapse rate calculations and the local terrain map built into the GFE grid field, to create temperature forecast grid fields that were much more terrain specific. This method for populating GFE temperature grid fields required a two-step approach. First, forecasters would populate the grid field with the smoothed MAV MOS forecasts for each of the first five periods. Then, forecasters would decide what lapse rate they would use for each period, and run the lapse rate smart tool to make further terrain enhancements.

In the spring of 2003, the method for populating the GFE temperature grid fields was changed again. (The MOS station-to-grid mapping method that was manually developed was no longer used.) The latest method, and the one currently in use, populates the temperature grid fields through a combination of direct model gridded output, the corresponding MOS forecasts for that model, and the local terrain map in GFE. Using the GFS Model and MAV MOS forecasts as an example of how this method works, the GFS Model two-meter temperature forecasts are loaded into the GFE-temperature grid field, and terrain adjusted according to the local terrain map in GFE. The "MatchMOS" (Barker 2003) Smart Tool is then used to add MAV MOS forecasts to the grid field, and make additional adjustments to the grid field based on the forecasts from the MAV MOS forecast sites. The degree of MAV MOS adjustment is dependent on the number of MOS sites, the distance from each MOS site, and elevation change from the MOS site. The final grid field produced depends on the density of MAV MOS stations. The GFE temperature grid forecast in areas with a high density of MAV MOS stations will be determined primarily by the MOS forecasts. The forecasts for areas of the grid field with a low density of MAV MOS stations will be determined primarily by GFS Model gridded output and elevation adjustments based on the local GFE terrain map. This method can be employed for any combination of gridded model output and corresponding MOS forecasts, and for any weather element for which there is a MOS forecast.

6. CONCLUSION

Presently at WFO ALY, the recommended procedure for populating MAX/MIN grid fields is to use the method discussed in the last paragraph of section 5, with GFS Model gridded output and MAV MOS forecasts, for the first five periods. This grid field adjustment method allows forecasters to include both MOS and model gridded output as the starting point for their gridded temperature forecasts. Areas near MAV MOS stations will have the best temperature forecasts available, based on MAV MOS forecasts. Additionally, remote terrain areas will also have the best temperature forecasts available, based on model gridded output which include specific local elevation adjustments.

This method for populating GFE temperature grid fields has been in use at WFO Albany since the end of April, 2003. Preliminary verification results (not shown) indicate that the change to the current method for populating MAX/MIN temperature grid fields has had little effect on WFO Albany forecaster performance against MOS guidance, and the overall dramatic improvement in WFO forecast performance that has occurred since September 1, 2002, continues.

Future efforts to improve temperature verification results will include examining the utility of using mesoscale models (such as the Workstation-ETA) to populate the MAX/MIN grid fields, and determining meteorological situations where MOS guidance does not perform well.

ACKNOWLEDGMENTS

The author would like to thank David Zaff, Information and Technology Officer, at NWS WFO ALY, for all his dedicated work installing/modifying office GFE smart tools to fit our current procedures.

REFERENCES

Barker, T., 2003: MatchMOS Version 1.0. NOAA/NWS IFPS Smart Tool Repository. [Available online at: http://140.90.90.253/~applications/STR/index.php3]

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

Glahn, H. R., D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. S*oc., **84**, 195-202.

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.

Kanamitsu, M., 1989: Description of the National Meteorological Center global data assimilation and forecast system. *Wea. Forecasting,* **4**, 335-342.

Maglaras, G. J., 1999: Temperature and precipitation forecast verification trends at the Albany forecast office: forecasters continue to show improvement on MOS guidance - Part II. *Natl. Wea. Dig.,* **23:1,2**, 3-12.

National Weather Service, 1992: NGM-based MOS guidance - the FOUS14/FWC message. *NWS Technical Procedures Bulletin No. 408,* NOAA, U.S. Department of Commerce, 8 pp.

National Weather Service, 2000: Avn-based MOS guidance - the 0000/1200 UTC alphanumeric messages. *NWS Technical Procedures Bulletin No. 482,* NOAA, U.S. Department of Commerce, 12 pp.

Table 1. The percent improvement of WFO ALY forecasts over MAV and FWC guidance for ALB, GFL, and POU, for the 21-month period from December, 2000, through August, 2002. Also shown (in parenthesis)are the percent improvements for July and August, 2002.

ALB

PERIOD	MAV	FWC
1	5.2 (-5.6)	6.8 (10.0)
2	-0.1 (4.8)	5.5 (10.6)
3	-1.5 (2.3)	5.4 (6.0)
4	-7.5 (-16.2)	8.4 (11.4)
5	-22.8 (-33.2)	
ALL	-6.4 (-10.0)	6.6 (9.4)

GFL

PERIOD	MAV	FWC
1	-5.1 (0.3)	2.5 (-3.5)
2	-8.0 (5.8)	-2.7 (-2.0)
3	-10.8 (-1.2)	0.4 (-4.7)
4	-11.3 (-7.5)	0.9 (0.6)
5	-22.9 (-19.2)	
ALL	-12.0 (-4.7)	0.2 (-2.2)

POU

PERIOD	MAV	FWC
1	-2.0 (-0.4)	0.3 (-2.7)
2	1.2 (6.1)	-0.3 (5.7)
3	-3.9 (-6.4)	-1.9 (4.3)
4	-5.8 (-10.3)	-2.2 (-5.2)
5	-27.3 (-36.9)	
ALL	-8.2 (-9.6)	-1.1 (0.8)

Table 2. The percent improvement of WFO ALY forecasts over MAV and FWC guidance for ALB, GFL, and POU, for the two-month period after September 1, 2002,and (in parenthesis) for the two-month period before.

ALB

PERIOD	MAV	FWC
1	12.3 (-5.6)	9.2 (10.0)
2	3.0 (4.8)	16.2 (10.6)
3	5.3 (2.3)	9.7 (6.0)
4	-7.5 (-16.2)	12.3 (11.4)
5	-1.3 (-33.2)	
ALL	2.1 (-10.0)	11.9 (9.4)

GFL

PERIOD	MAV	FWC
1	5.7 (0.3)	-4.4 (-3.5)
2	5.9 (5.8)	-0.8 (-2.0)
3	4.0 (-1.2)	12.6 (-4.7)
4	-9.6 (-7.5)	8.2 (0.6)
5	0.8 (-19.2)	
ALL	1.1 (-4.7)	4.7 (-2.2)

POU

PERIOD	MAV	FWC
1	4.6 (-0.4)	-4.6 (-2.7)
2	1.1 (6.1)	-5.1 (5.7)
3	0.0 (-6.4)	13.1 (4.3)
4	-15.5 (-10.3)	13.6 (-5.2)
5	-8.7 (-36.9)	
ALL	-3.8 (-9.6)	6.0 (0.8)

Table 3. The percent improvement of WFO ALY forecasts over MAV and FWC guidance for ALB, GFL, and POU, for the eight-month period from September 1, 2002, through May 4, 2003. Also shown (in parenthesis) the results for the 21-month period before September 1, 2002 (repeated from Table 1).

ALB

PERIOD	MAV	FWC
1	10.2 (5.2)	17.7 (6.8)
2	3.4 (-0.1)	19.1 (5.5)
3	4.3 (-1.5)	18.0 (5.4)
4	-0.5 (-7.5)	21.0 (8.4)
5	-5.0 (-22.8)	
ALL	2.0 (-6.4)	18.9 (6.6)

GFL

PERIOD	MAV	FWC
1	6.4 (-5.1)	12.6 (2.5)
2	6.9 (-8.0)	13.0 (-2.7)
3	3.1 (-10.8)	16.7 (0.4)
4	0.2 (-11.3)	15.0 (0.9)
5	-1.2 (-22.9)	
ALL	2.6 (-12.0)	14.4 (0.2)

POU

PERIOD	MAV	FWC
1	1.3 (-2.0)	7.8 (0.3)
2	0.3 (1.2)	3.0 (-0.3)
3	-1.1 (-3.9)	10.1 (-1.9)
4	1.3 (-5.8)	16.5 (-2.2)
5	-5.8 (-27.3)	
ALL	-1.1 (-8.2)	9.9 (-1.1)