449

FORTY YEARS OF NWS FORECASTS: PAST PERFORMANCE AND FUTURE ADVANCES

Tabitha L. Huntemann*, David E. Rudack, David P. Ruth
Meteorological Development Laboratory
Office of Science and Technology
National Weather Service, NOAA
Silver Spring, MD

1. INTRODUCTION

The Meteorological Development Laboratory (MDL) of the National Weather Service (NWS) has issued model output statistics (MOS) forecast guidance for nearly four decades (Glahn and Lowry 1972). For many years, MOS guidance was generated for observing stations and formatted in text bulletins while official NWS forecasts for stations and zones were created by forecasters typing text. Today, the National Digital Forecast Database (NDFD) contains official NWS forecasts produced by forecasters at local Weather Forecast Offices (WFO) and National Centers on a fine-resolution grid (Glahn and Ruth 2003). MDL supports NDFD with both station-based and gridded MOS guidance (Glahn et al. 2009).

The NWS and MDL routinely evaluate official forecasts at stations and compare the skill of the human forecast to the MOS guidance for the same weather element (Dagostaro et al. 2004). Improvements in NWS public weather forecasts and in statistically post-processed numerical weather prediction (NWP) can be traced by the verification of the weather elements.

In this paper, we examine the performance at stations of four decades of official NWS maximum temperature (MaxT), minimum temperature (MinT), and 12-h probability of precipitation (PoP12) forecasts compared to MOS guidance for forecast periods to approximately 60 hours in advance. We investigate the performance of the last two years of NDFD, Global Forecast System-based MOS (GFS MOS), and European Centre for Medium-Range Weather Forecasts-based MOS (ECMWF MOS) forecasts for forecast periods out to seven days in advance.

** Corresponding author address:*
Tabitha L. Huntemann, Meteorological Development Laboratory, 1325 East-West Highway, Silver Spring, MD 20910; e-mail: Tabitha.Huntemann@noaa.gov

2. PAST PERFORMANCE

Dallavalle and Dagostaro (2004) and Ruth et al. (2009) documented the improvement in guidance products that objectively interpret the output of numerical weather prediction models. Figure 1 shows the 79 CONUS stations with suitable data availability over the last four decades for verification. Figures 2-4 show verifications for the last four decades of MOS guidance compared to official forecasts prepared at local NWS offices for MaxT, MinT, and PoP12. For temperatures, mean absolute errors (MAE) are provided. For PoP12, Brier scores are provided. The Brier score assesses the accuracy of probability forecasts (Wilks 2006). A perfect Brier score is zero. The results shown here are an update to the Ruth et al. (2009) charts through the 2014 warm season.

For information on how scores prior to 2007 cool season were obtained, see Dallavalle and Dagostaro (2004) and Ruth et al. (2009). Scores from the 2007 cool season through the 2011 cool season were extracted from Stats on Demand, an interactive database maintained by the Performance Branch of the NWS Office of Climate, Water, and Weather Services (NWS 2011). This database contains Public Forecast Matrices (PFM) used by each WFO twice a day at 0400 Local Time (LT) and 1600 LT, guidance from available MOS products, and observations from all METAR/SPECI reports issued for each location in the PFMs. Forecast data from the PFMs issued at 0400 LT (1600 LT) for the next two days are matched to the MOS guidance based on the 0000 UTC (1200 UTC) model cycle from the same date. The PFM verification provided by Stats on Demand was discontinued in May 2012.

Scores for the 2012 warm season through the 2014 warm season were obtained using MDL's NDFD point verification system (Dagostaro et al. 2004). Local forecasts are compared to MOS guidance that is available several hours prior to local forecast issuance. In this system, NDFD forecasts issued at 0000 UTC to MOS guidance based on the previous day's 1200 UTC model cycle. NDFD forecasts issued at 1200 UTC are matched to MOS guidance based on the same day's 0000 UTC model cycle. MDL's NDFD point verification system was run for the 2011 warm and cool seasons to compare to results obtained from Stats on Demand. MAEs and Brier scores for these two seasons were found to be very similar for MaxT, MinT, and PoP12 forecasts.

Improvements in NWS public weather forecasts and in statistically post-processed numerical weather prediction can be traced by the verification of the weather element guidance as shown in Figs. 2-4. For example:

- The transition from Perfect Prog guidance to the MOS approach in 1973 resulted in a clear improvement to MaxT guidance scores (Fig. 2).
- The change from MinT valid for a calendar day to a nighttime MinT in late 1984 reduced errors of both local forecasts and guidance (Fig. 3, Erickson and Dallavalle 1986).
- Forecasts are continually improving: Day 2 local forecasts (dark blue) are now as good as they were for day 1 (light blue) 10 to 20 years earlier.
- Problems with models contribute to decreases in performance: GFS model changes in 2010-2011 changed bias characteristics and negatively affected MOS. A refresh of the GFS MOS guidance was made in January 2015, coincident with the GFS model upgrade.

3. FUTURE ADVANCES

3.1 *ECMWF MOS*

While ECMWF model output is widely recognized in the meteorological community for its accuracy, it can contain systematic bias. Recently, MDL has applied the MOS approach to the ECMWF model to enhance its usefulness to NWS forecasters (Rudack et al. 2014). Application of the MOS technique to ECMWF model output has also produced guidance for weather elements found in MDL's GFS MOS including elements not available directly in ECMWF model output, such as probability of precipitation.

3.2 *Data and methods*

For this analysis, NDFD gridded forecasts of MaxT, MinT, PoP12, temperature (T), and dewpoint (Td) from 19 September 2012 to 30 September 2014 were collected for the 1200 UTC issuance time. NDFD forecasts at the 79 verification stations were extracted from the 5-km NDFD grid by using a nearest neighbor technique (Dagostaro et al. 2004). The NDFD forecasts were made at 5 km resolution for most of this sample. Ruth et al. (2009) concluded that NDFD forecast extraction from the 5 km grid gives an advantage to MOS, particularly in regions of complex terrain. Starting 19 August 2014, NDFD forecasts were made at 2.5 km resolution. This should reduce that advantage.

The NDFD forecasts were matched with the same day's 0000 UTC model cycle GFS MOS and ECMWF MOS station forecasts available to forecasters at the time the NDFD forecasts were prepared. T and Td forecasts were verified directly against METAR observations. MaxT, MinT, and PoP12 verifying observations were computed from hourly or 6-hourly information contained in METAR reports. As for the long-term verification, MAEs are provided for temperatures and Brier scores for PoP12 verification.

3.3 *Results*

Figures 5 and 6 compare the accuracy of GFS MOS, NDFD, and ECMWF MOS for projections spanning days 1 through 7. The ECMWF MOS MaxT guidance was nearly as accurate as the NDFD forecast. For MinT, T, and Td, the ECMWF guidance is consistently more accurate than the corresponding GFS MOS and NDFD, particularly at later projections. ECMWF MOS PoP12 is more accurate than GFS MOS or NDFD. GFS MOS PoP12 is less accurate than both ECMWF MOS and NDFD.

Reliability diagrams can be used to assess how closely the forecast probabilities of an event correspond to the actual chance of observing the event. Figure 7 shows reliability diagrams for four different forecast projections. In general, the ECMWF MOS is very reliable. NDFD and GFS MOS PoP12 both tend to underforecast higher precipitation categories for this sample. Forecasts exhibit "sharpness" if they are capable of predicting events with probabilities with extreme values relative to the observed event frequency (Wilks 2006). The ECMWF MOS produces more reliable high-probability forecasts than GFS MOS or NDFD, and thus ECMWF MOS exhibits more sharpness than GFS MOS and NDFD.

The ECMWF MOS guidance has not been implemented operationally, but these results suggest that NWS forecasters could add value to their morning updates and afternoon forecast products by considering the 0000 UTC ECMWF MOS.

4. SUMMARY & CONCLUSIONS

MDL has continually adapted MOS guidance to meet the needs of NWS forecasters at WFOs. Both local forecasts and guidance have clearly improved in quality over the last 40 years. Day 2 local forecasts are as good as they were for day 1 about 10 years ago. Guidance quality has been negatively impacted in recent years by model changes, but remains valuable.

Improvements to MOS guidance continue. A refresh of GFS MOS station guidance is ongoing in response to the planned GFS model upgrade in early 2015. 0000 UTC and 1200 UTC ECMWF MOS is planned to be made operationally available to forecasters by mid-2015. In addition, the ECMWF MOS will be used to leverage European models for the National Blend of Models project (Gilbert et al. 2015). We anticipate that NWS forecasters will continue to find MDL's MOS products valuable for many years to come.

5. ACKNOWLEDGMENTS

The authors wish to thank Valery Dagostaro for her insights on MDL's NDFD verification processes. Bob Glahn also provided helpful suggestions for this research.

6. REFERENCES

Dagostaro, V. J., W. A. Shaffer, M. J. Schenk, J. L. Gorline, 2004: A prototype verification system for examining NDFD forecasts. *17th Conf. on Probability and Statistics in the Atmospheric Sciences*, Seattle, WA, Amer. Meteor. Soc., 2.7. [Available online at http://www.nws.noaa.gov/mdl/pubs/Documents/Papers/Dagostaro2004APrototype.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.]

Erickson, M. C., and J. P. Dallavalle, 1986: Objectively forecasting the short-range maximum/minimum temperature – A new look. Preprints *11th Conf. on Wea. Forecasting and Analysis,* Kansas City, MO, Amer. Meteor. Soc., 33-38.

Gilbert, K. K., J. P. Craven, D. R. Novak, T. M. Hamill, J. Sieveking, D. P. Ruth, and S. J. Lord, 2015: An introduction to the National Blend of Global Models project. *Special Symp. on Model Postprocessing and Downscaling,* Phoenix, AZ, Amer. Meteor. Soc., 3.1. [Available online at https://ams.confex.com/ams/95Annual/webprogram/Paper267282.html.]

Glahn, B., K. Gilbert, R. Cosgrove, D. P. Ruth, and K. Sheets, 2009: The Gridding of MOS. *Wea. Forecasting*, **24**, 520–529. [Available online at http://journals.ametsoc.org/doi/abs/10.1175/2008WAF2007080.1.]

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.

Glahn, H. R., and D. P. Ruth, 2003: The New Digital Forecast Database of the National Weather Service. Bulletin of the American Meteorological Society [Available online at http://www.nws.noaa.gov/ndfd/resources/bamsarticle.pdf.]

NWS, 2011: Verification. National Weather Service Instruction 10-1601, 100 pp. [Available online at http://www.nws.noaa.gov/directives/sym/pd01016001curr.pdf.]

Rudack, D. E., D. P. Ruth, K. K. Gilbert, T. Curtis, 2014: A first look at the Meteorological Development Laboratory's experimental ECMWF MOS system. *26th Conf. on Wea. Analysis and Forecasting*, Atlanta, GA, Amer. Meteor. Soc., J4.1. [Available online at https://ams.confex.com/ams/94Annual/webprogram/Paper231722.html.]

Ruth, D. P., B. Glahn, V. Dagostaro, K. Gilbert, 2009: The performance of MOS in the digital age. *Wea. Forecasting*, **24**, 504-519. [Available online at http://www.nws.noaa.gov/mdl/pubs/Documents/Papers/Glahn2009TheGridding.pdf.]

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. Elsevier, 627 pp.

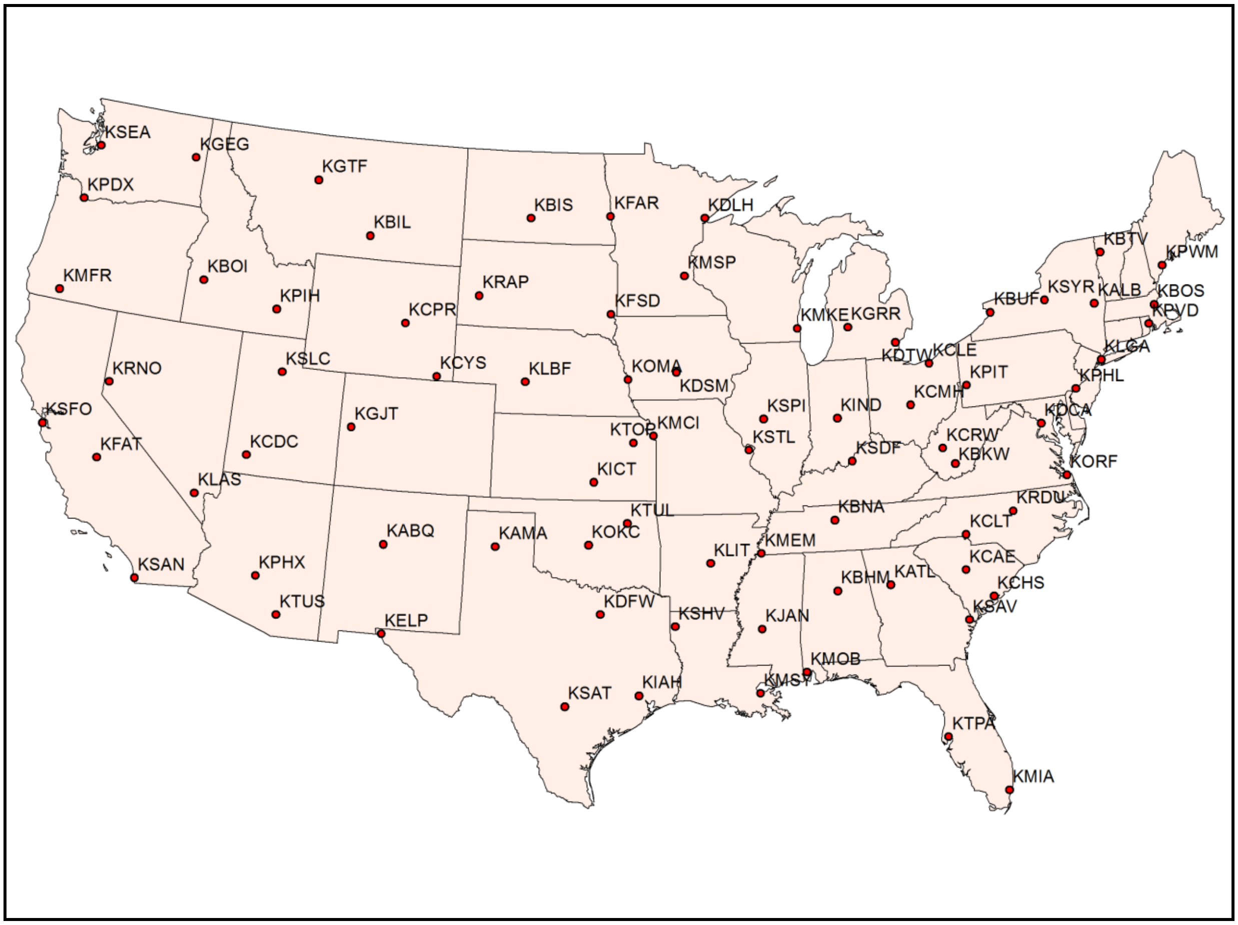

Figure 1. Locations of the 79 METAR stations used to verify local and guidance forecasts.

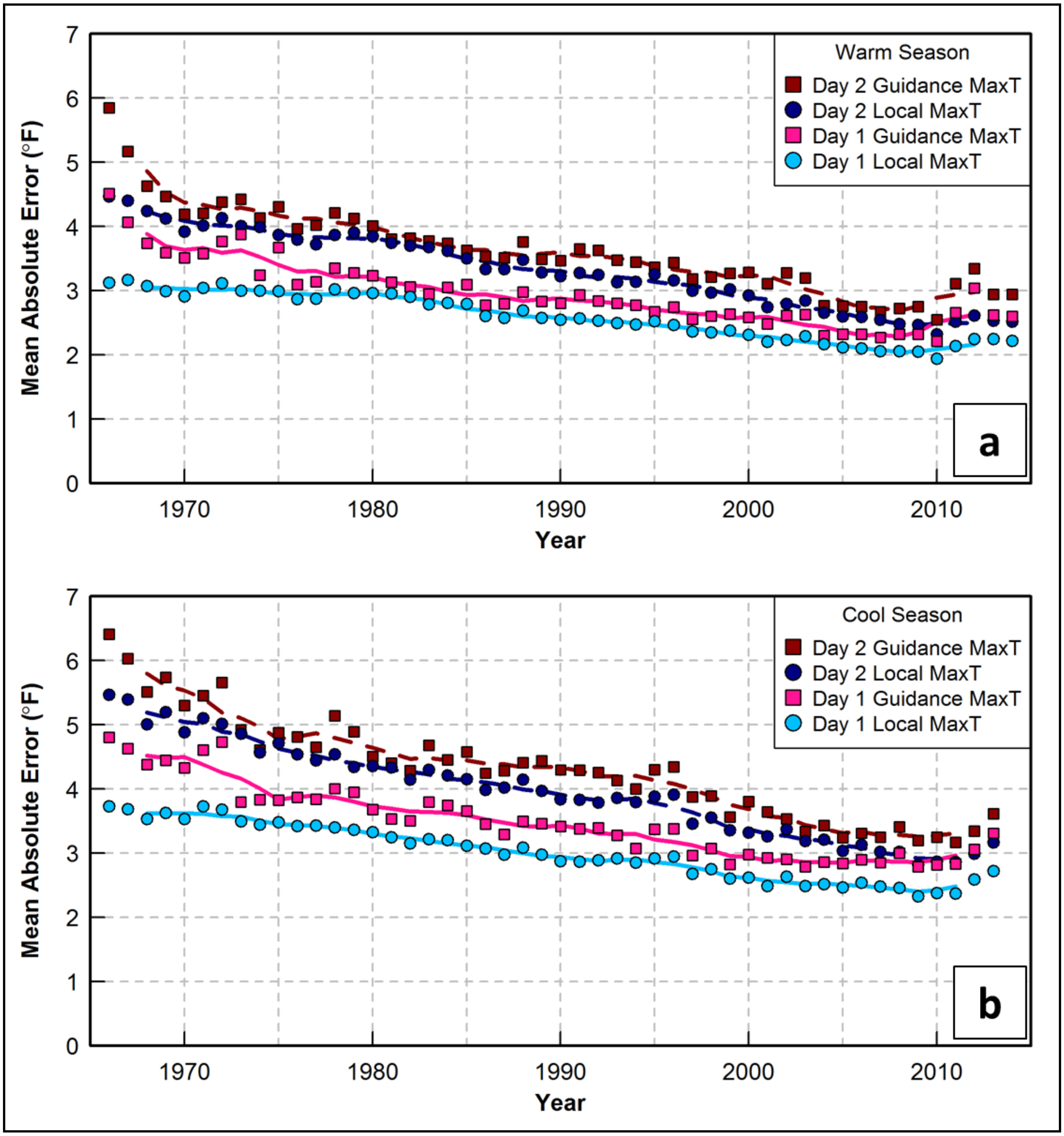

Figure 2. Day 1 and day 2 MaxT MAEs for (a) warm (April-September) and (b) cool (October-March) seasons based on 0000 UTC model cycle guidance. Years show when season began. Lines are 5-year moving averages.

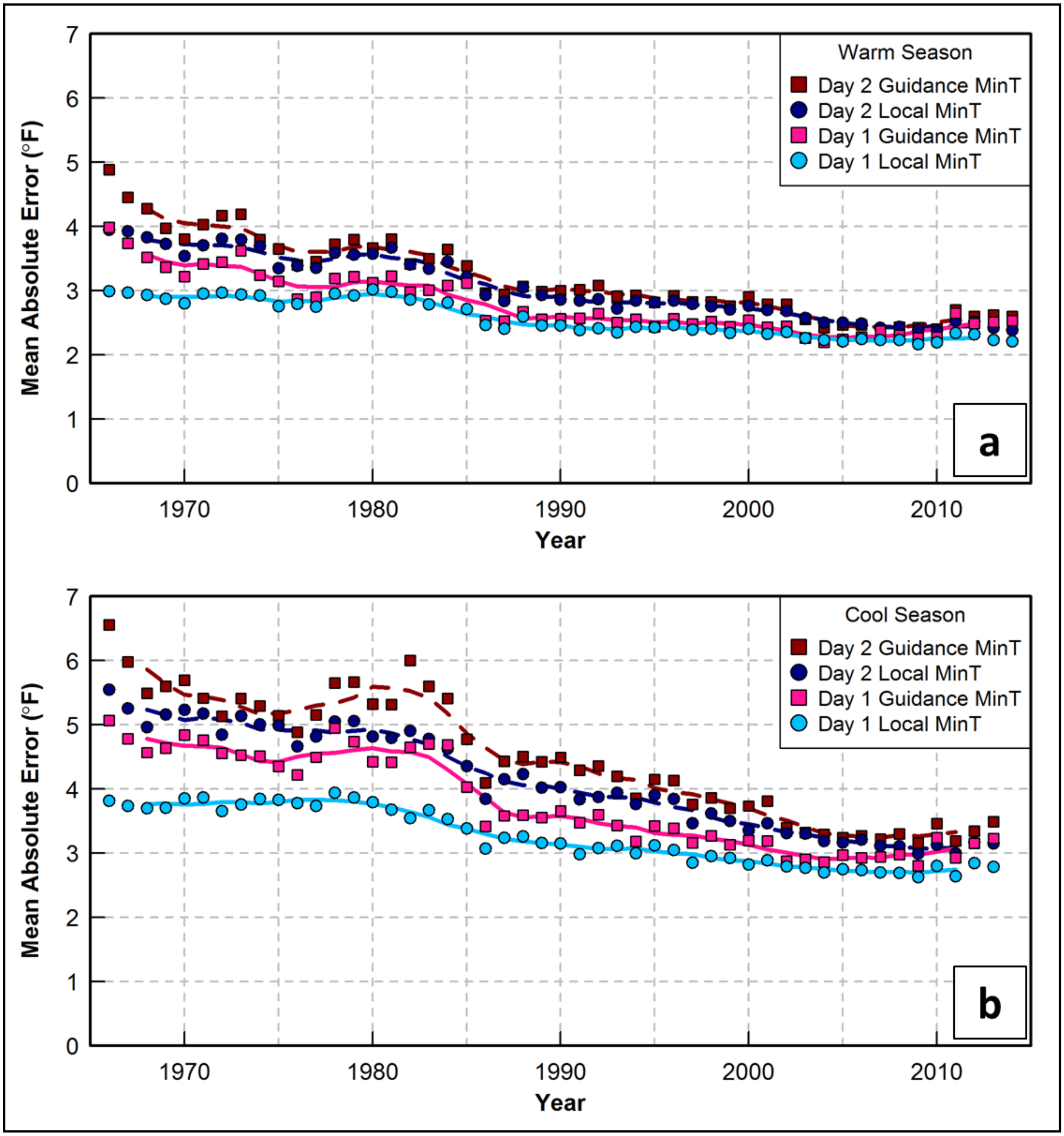

Figure 3. Day 1 and day 2 MinT MAEs for (a) warm (April-September) and (b) cool (October-March) seasons based on 1200 UTC model cycle guidance. Years show when season began. Lines are 5-year moving averages.

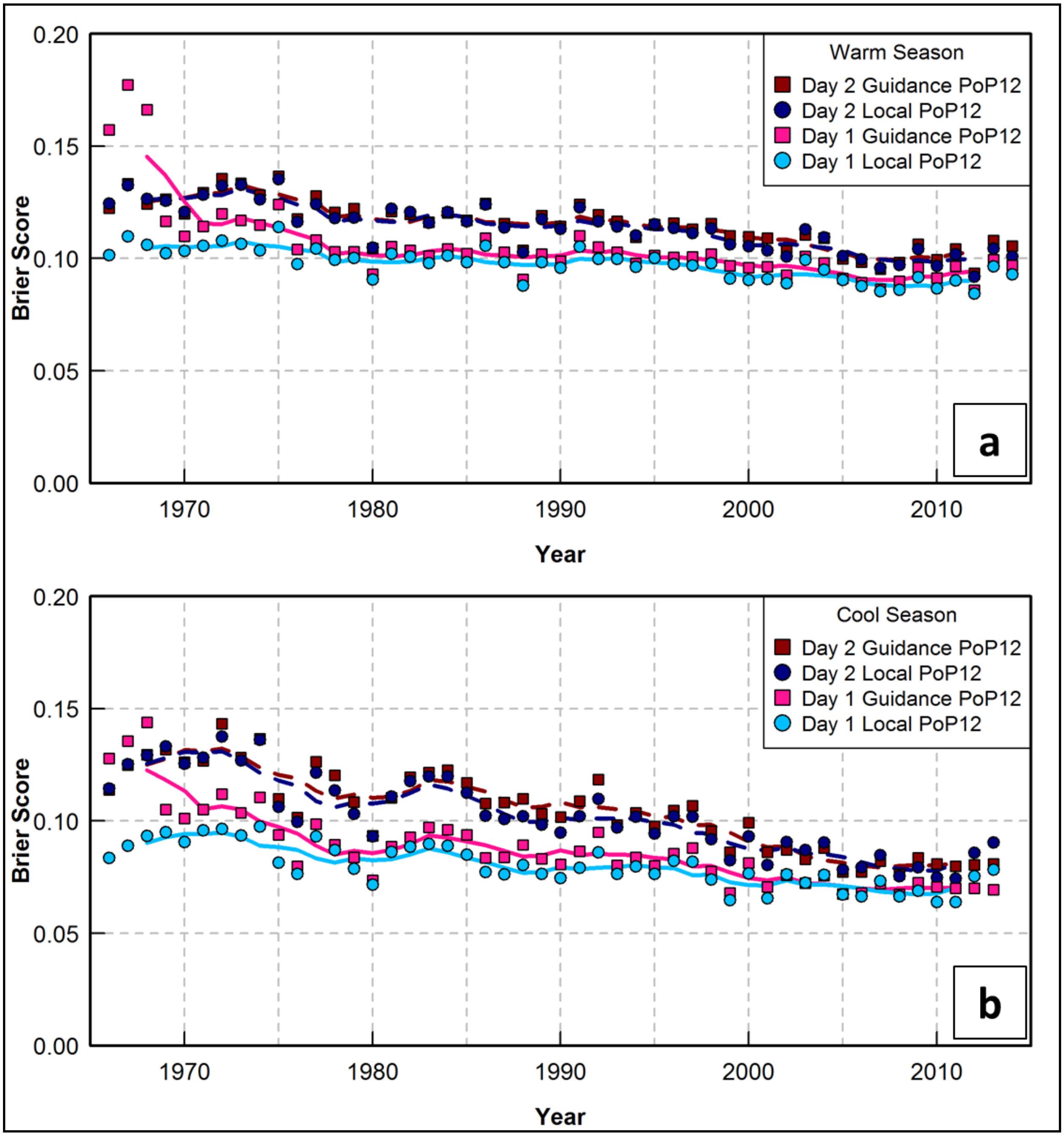

Figure 4. Day 1 and day 2 PoP12 Brier scores for (a) warm (April-September) and (b) cool (October-March) seasons based on the combined 0000 and 1200 UTC model cycle guidance. Years show when season began. Lines are 5-year moving averages.

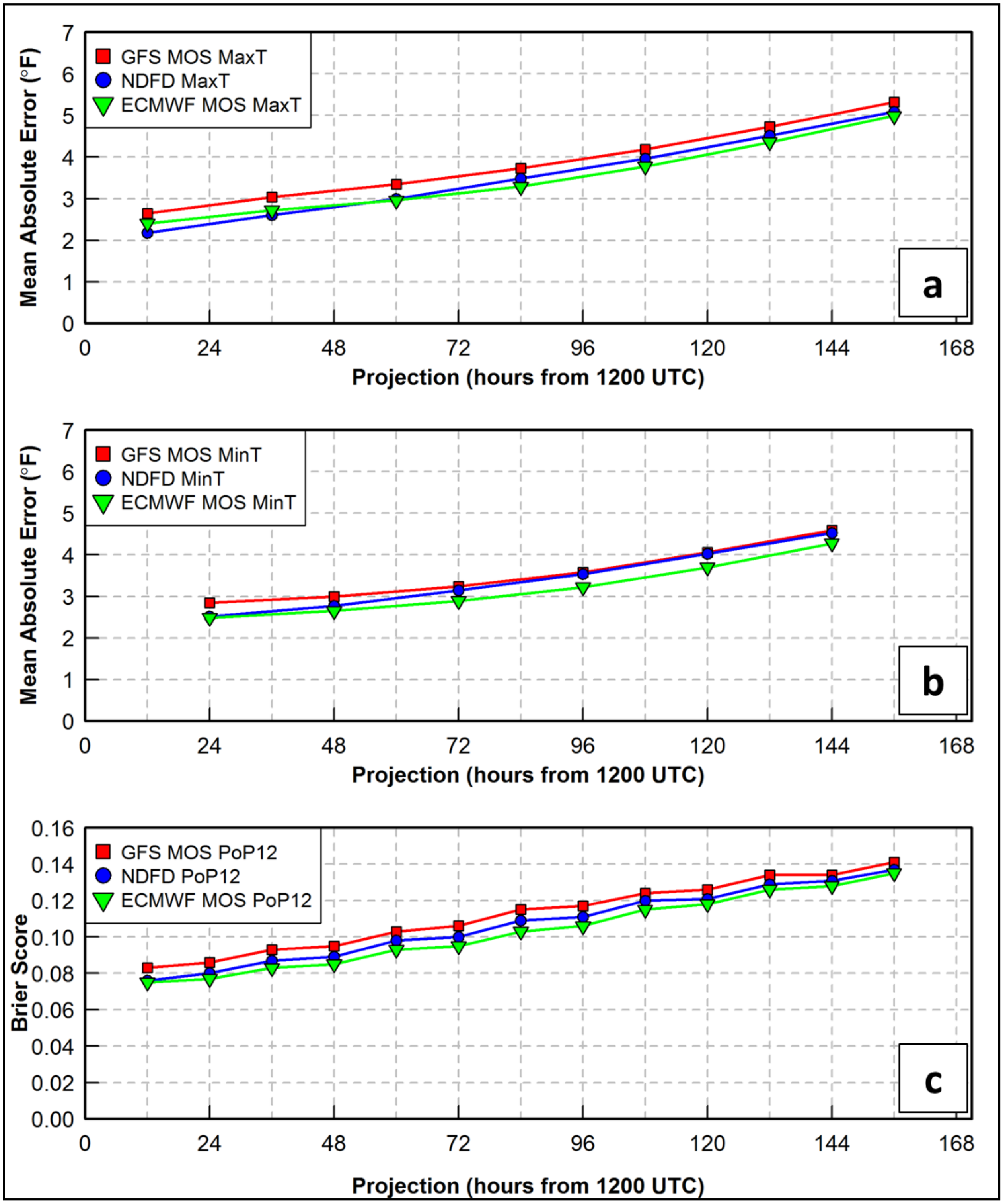

Figure 5. Comparative accuracy of GFS MOS, NDFD, and ECMWF MOS at 79 CONUS stations from September 2012 to September 2014 for (a) MaxT, (b) MinT, and (c) PoP12. Forecast hours relative to NDFD issuance times.

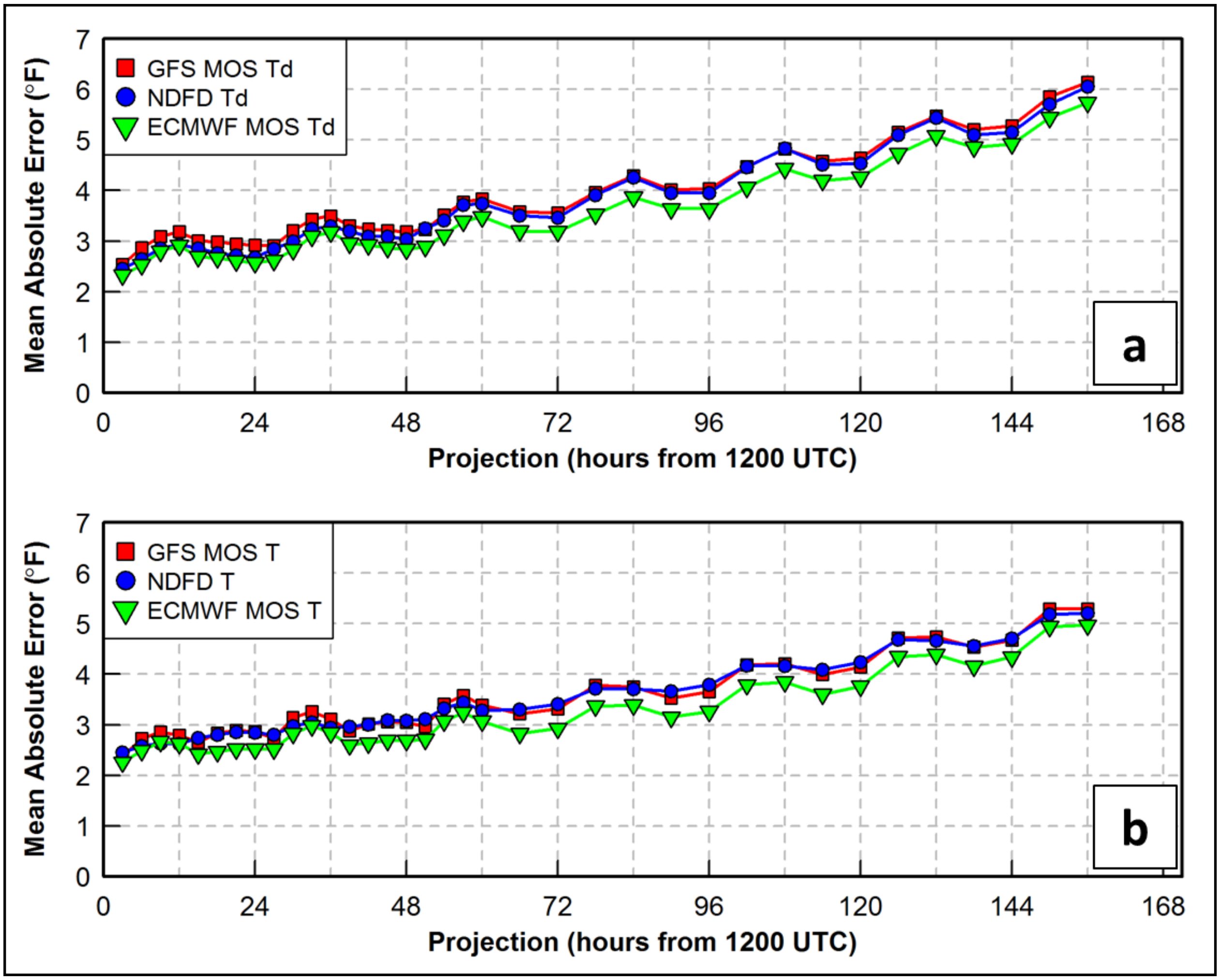

Figure 6. Comparative accuracy of GFS MOS, NDFD, and ECMWF MOS at 79 CONUS stations from September 2012 to September 2014 for (a) T and (b) Td. Forecast hours are relative to NDFD issuance times.

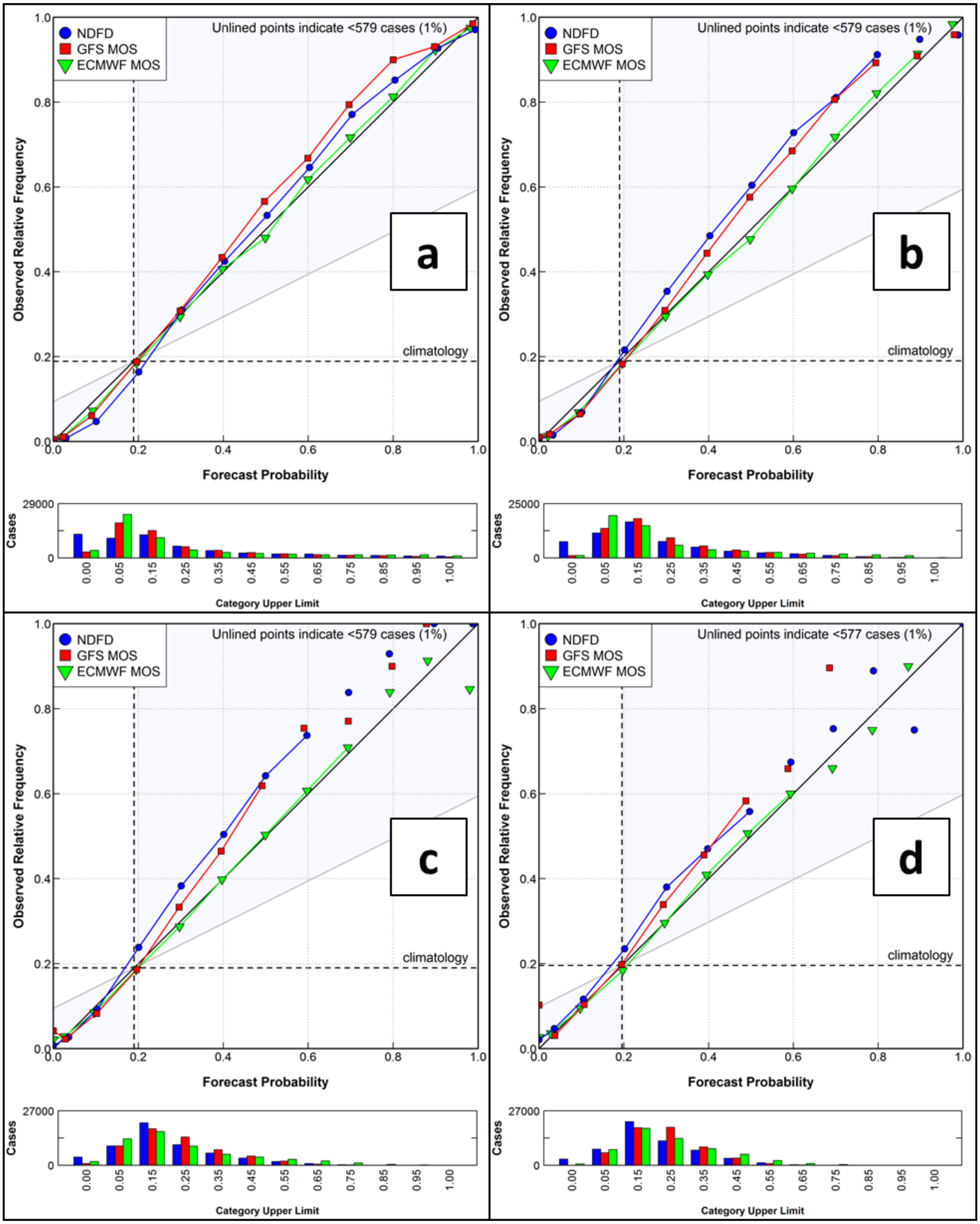

Figure 7. Reliability diagrams for (a) 24-h, (b) 72-h, (c) 120-h, and (d) 156-h PoP12. Forecast hours are relative to NDFD issuance times.