J4.1 A FIRST LOOK AT THE METEOROLOGICAL DEVELOPMENT LABORATORY'S EXPERIMENTAL ECMWF MOS SYSTEM

David E. Rudack*, David P. Ruth, Kathryn K. Gilbert, and Tamarah Curtis
Meteorological Development Laboratory
Office of Science and Technology
National Weather Service, NOAA

1. INTRODUCTION

The skillful performance of the European Centre for Medium-Range Weather Forecasts (ECMWF) model (White 2002) is well documented and widely recognized in the meteorological community (Hamill 2012). While quite skillful, ECWMF model output does contain systematic bias. Further, the model does not directly provide forecasts for weather elements such as probability of precipitation. In order to enhance the usefulness of the ECMWF model to NWS forecasters, the Meteorological Development Laboratory (MDL) has applied a Model Output Statistics (MOS) approach to ECMWF model output. The MOS approach, originally developed by Glahn and Lowry (Glahn and Lowry 1972), has been successfully employed by MDL to post-process numerical model output for several decades.

A short- and extended-range experimental MOS product has been developed for stations from 0000 UTC ECWMF model output. The statistical guidance contains many of the weather elements found in MDL's station-based Global Forecast System (GFS) MOS for both the short- and extended-range (Dallavalle et al. 2004, Dallavalle and Cosgrove 2004, Maloney et al. 2010, Erickson and Carroll 1999). In this paper, we focus on the most recent performance of ECMWF MOS guidance for 2-m temperature and dewpoint, local daytime maximum 2-m temperature (MaxT), local nighttime minimum 2-m temperature (MinT), and 12-h probability of precipitation (PoP12) (probability of measurable liquid-equivalent precipitation over a 12-h period). These weather elements have been selected for evaluation because of their impact on public and agricultural interests. As with any type of verification study, a baseline must be established to measure the relative performance of a new product. For this reason, we included verification of operational GFS MOS guidance for the same independent test sample.

In this extended abstract, Section 2 discusses some of the ECMWF model characteristics relevant to this ECMWF MOS development. Section 3 explains basic procedures for developing equations for the weather elements verified here. In Section 4, we present verification results of ECMWF MOS guidance in comparison to GFS MOS guidance. A summary of the results can be found in Section 5, and a short description of future plans is presented in Section 6.

2. ECMWF MODEL DATA

The parent model used in this MOS development is the deterministic portion of the Integrated Forecast System also commonly referred to as the "ECMWF" or "European" Model (Simmons et al. 1989). This global spectral model is run twice daily for the 0000 and 1200 UTC cycles and generates forecasts for a variety of weather elements for Days 1 through 10. The native vertical resolution of the model (91 vertical levels, approximately every 11 mb) remained constant during the sample period used for ECMWF MOS development (see Section 3.1). However, the horizontal resolution did increase from ~25 km to ~12 km beginning in January 2010. Unfortunately, we were unable to take advantage of these higher horizontal resolution forecasts because at the time, the National Centers for Environmental Prediction (NCEP) only received ECMWF model forecasts on a global 1-degree latitude/1-degree longitude grid. Since MDL provides forecasts for a subset of this global domain, only a portion of the ECMWF model grid was saved (Fig. 1). The temporal resolution of the available ECMWF model data remained constant throughout the archive period. Forecast fields were available beginning with the analysis and extending through 240 hours at 6-h increments.

* *Corresponding author address:* David E. Rudack
1325 East-West Hwy, Silver Spring, MD 20910.
E-mail: David.Rudack@noaa.gov

3. ECMWF MOS DEVELOPMENT

Procedurally, developing station-based ECMWF MOS (henceforth referred to as ECMWF MOS) was no different than developing any other MDL station-based MOS product. Tasks included interpolating ECMWF gridded model data to station locations for predictor use, investigating sample size availability, stratifying data by season and station, and selecting a reasonable pool of possible predictors for equation development. We now briefly discuss each of these aspects pertaining to the development of ECMWF MOS equations.

3.1 *Development Sample Period*

As with any MOS development, consistency in the availability of model forecasts dictates the developmental period. In our case, since a consistent set of model fields in the archived 0000 UTC ECMWF data were unavailable prior to April 2008 the model data used in this development began in April 2008 and ended in September 2011. Following the practice of other MOS developments, we stratified the data sample into two seasons, cool (October-March) and warm (April-September). Cool season equations were developed by using data from 2008-2009, 2009-2010, and 2010-2011 while warm season equations were derived from the 2008, 2009, 2010, and 2011 sample periods. The developmental data are stratified by season to better capture the relationships between predictor(s) and predictand(s). However, one negative effect of stratifying the data in this manner is that the sample size for a specific set of equations is cut in half. This can limit the predictability of rare events. Despite this drawback, a sufficient sample size of model and observational data was present to develop stable ECMWF MOS seasonal equations.

3.2 *ECMWF MOS 2-m Temperature, 2-m Dewpoint, MaxT, and MinT*

Single-station development is the preferred method for developing stable regression equations when a sufficient data sample is available. In this instance, the single-station equation can be tuned to the local weather observed at a particular site. This approach was used for the weather elements of temperature, dewpoint, daytime MaxT, and nighttime MinT. Single-station regression equations were developed for a total of 1279 METAR stations inside the contiguous United States (CONUS) and outside the contiguous United States (OCONUS). Unique regression equations valid for a specific projection or period were generated for each station. Equations for 2-m temperature and dewpoint were generated for Days' 1 through 8, at 3-h intervals beginning at the 6-h projection and ending at the 192-h projection. Single-station MaxT and MinT regression equations were generated for Days 1 through 8 and Days 2 through 7, respectively.

Many of the predictors used in the regression analysis (Table 1) were similar, if not identical, to those used in the development of the current operational GFS MOS equations. As one might expect, the ECMWF model 2-m temperature and dewpoint were generally the top two predictors chosen.

To reduce the chances of generating inconsistent meteorological forecasts between weather elements (e.g., temperature less than dewpoint), equations for 2-m temperature, dewpoint, MaxT, and MinT were developed simultaneously. In other words, we forced the regression analysis to use the same set of predictors for certain sets of predictands. Thus, each prediction equation in a set possessed the same set of predictors, but with different coefficients tuned to the predictand. Additional consistency checks are applied to the final MOS forecast guidance.

3.3 *ECMWF MOS PoP12*

As noted above, a sufficient sample must exist to generate stable regression equations for a particular weather event. Since the occurrence of measurable precipitation is relatively infrequent, single-station development is generally avoided. So, how does one increase the sample size for situations such as this? The method is to regionalize the equations. Regionalizing is the process of collating data within a geographical region that experiences the same general type of weather phenomenon. Pooling the data in this manner supplies the regression analysis with sufficient cases for developing forecast equations for less frequent weather events. An added benefit of developing regional equations is that forecasts can be made for those observational sites within a particular region that do not contribute to the regression analysis. Note that regionalization does not guarantee that a sufficient number of cases will be present in the sample to develop equations for rare events such as rainfall greater than two inches.

The stations and regional boundaries used in the development of ECMWF MOS PoP12 guidance equations were primarily the same as those used in the GFS MOS PoP12 development. The sample period was identical to that noted in Section 3.1. In all, 873 stations with reliable observations were available for equation development. These stations (along with the remaining 1415 stations not sampled) were divided into 20 regions for the cool season and 19 regions for the warm season (Fig. 2). For the short-range forecast projections (18 through 84 hours), PoP12 equations valid at 6-h increments were generated. For extended-range projections (96 through 192 hours), PoP12 equations valid every 12 hours were generated. The projection assigned to the PoP12 forecast corresponds to the ending 12-h period for which the forecast is valid. For example, a MOS 48-h PoP12 is the probability that measurable precipitation will fall between the 36-h and 48-h projections.

Table 2 shows a list of model predictors offered to the regression. As one might expect, the model accumulated precipitation amount was the most frequently chosen predictor for both the short and extended ranges. Although the usefulness of this predictor in the PoP12 regression analysis diminished in lockstep with the declining skill of the model, it was still chosen even beyond the 156-h projection.

4. ECMWF MOS AND GFS MOS VERIFICATION RESULTS

All MOS forecasts were verified over a matched, independent sample covering the 2012-2013 cool and 2013 warm season. A reliable set of 334 uniformly distributed stations across the CONUS and OCONUS was used to evaluate the overall performance of experimental ECMWF MOS and operational GFS MOS guidance (Fig. 3). Because the CONUS and Alaskan verification results were very similar, we have elected in this paper to limit the verification discussion to the overall results. Note that verifications for Hawaii and Puerto Rico MOS guidance were also included in the overall results, but did not resemble the overall scores. Separate MOS verification results for the CONUS, Alaska, Hawaii, and Puerto Rico regions are available upon request.

Because of time zone differences across NWS Weather Forecast Offices and the latency of the arrival of 0000 UTC ECMWF model data, ECMWF MOS is not always available in the nightly guidance suite used to prepare NWS official forecasts. In these instances, the 1200 UTC GFS MOS forecasts along with the previous night's 0000 UTC ECMWF MOS, would likely be used in tandem when preparing the subsequent day's morning or afternoon forecast package. With this in mind, we have also included the subsequent day's 1200 UTC GFS MOS forecasts in this verification analysis. Because the 0000 UTC ECMWF MOS is not available to the NWS forecaster in a timely manner, we have verified the ECMWF MOS product in comparison with both the 0000 UTC GFS MOS guidance, as well as the 1200 UTC GFS MOS guidance available 12 hours later. The latter scenario more closely matches what the forecaster has available in the operational environment. These 1200 UTC GFS MOS forecasts were matched to the same valid time as the 0000 UTC ECMWF and 0000 UTC GFS MOS forecasts.

4.1 *2-m Temperature, 2-m Dewpoint, MaxT, and MinT*

The mean absolute error (MAE) is used to measure the accuracy of continuous weather elements such as temperature. Lower MAE values represent more accurate forecasts. Figure 4a and 4b show the MAE for MOS 2-m temperature forecasts for both the cool and warm seasons beginning at the 18-h projection and extending to the 192-h projection at 6-h intervals, respectively. For both cool and warm seasons, the ECMWF MOS guidance was more accurate than the corresponding 0000 UTC GFS MOS at all projections. The improvement varied by projection but was consistent throughout. Although the MAE improvement is only a few tenths of a degree, it has been our experience that this small amount translates into a meaningful forecast improvement.

Somewhat surprisingly, ECMWF MOS guidance was also more accurate than most of the subsequent 1200 UTC GFS MOS guidance valid at the same projection. The only notable exception was in the warm season prior to the 24-h projection. For these very early projections, the ECMWF MOS guidance was less accurate because the 1200 UTC GFS MOS guidance had the advantage of the most recent observations.

For the weather element of 2-m dewpoint, the same overall verification picture emerged with a couple of notable exceptions (Figure 4c and 4d). For the cool season, the subsequent 1200 UTC GFS MOS was almost as accurate as the ECMWF MOS for the 126-h projection and beyond. Also, the cool season GFS MOS 2-m dewpoint guidance in the projections prior to 24 hours was more accurate than the ECMWF MOS.

Figure 5 shows the cool and warm season MAE scores for MaxT and MinT. The verifications for each season display the same overall behavior. ECMWF MOS guidance was consistently more accurate than the corresponding 0000 and 1200 UTC GFS MOS guidance at all projections with the exception of the 1200 UTC GFS MOS Day–8 MaxT and Day-7 MinT. These results suggest that NWS forecasters can still add value to their morning updates and afternoon forecast products by considering the 0000 UTC ECMWF MOS.

4.2 *PoP12*

The reliability diagram (Wilks 2006) is one method to visually assess the bias behavior of a set of probabilistic forecasts like PoP12. Probability forecasts are generally partitioned into evenly spaced bins (0-10%, 10-20%,…, 90-100%) and compared to the observed relative frequency of the event within that bin. Probabilistic forecasts are deemed reliable when the average probability forecast and average observed frequency of the event are about the same for all or a majority of bins. On a reliability diagram, perfect reliability is denoted by a diagonal line beginning at the origin and extending to the upper right corner of the graph. Reliability values above this line at a particular probability threshold represent underforecasting the event at that threshold. Reliability values below the diagonal line imply over-forecasting of the event.

Figure 6 shows the ECMWF and GFS MOS reliability performance of the 72-h, 120-h, and 168-h PoP12 guidance for both the cool and warm seasons. Overall, the ECMWF MOS POP12 guidance is more reliable than the GFS MOS guidance. The ECMWF MOS reliability line primarily lies close to the diagonal with only a slight tendency to underforecast PoP12 in the middle probability ranges. In contrast, GFS MOS forecasts above the 20% bin consistently underforecast PoP12. The only exceptions are those in the high probability bins for the warm season for the 120-h and 168-h projections. However, these reliability values are less meaningful because of the small number of cases in these bins.

Another metric used in determining the efficacy of probabilistic forecasts is the property of sharpness. Forecast sharpness refers to the ability of probabilistic forecasts to differ from the event's climatological relative frequency (approximately 10-20% in this case). In Fig. 6, we see this is indeed the case for the ECMWF MOS guidance where a higher proportion of cases differed from the climatological normal of approximately 20%. This behavior was especially prevalent in the 72- and 120-h projections and to a lesser extent at the 168-h projection for both cool and warm seasons. While GFS MOS PoP12 guidance also differed from climatology, an appreciable proportion of the GFS MOS guidance was clustered in the climatological range relative to ECMWF MOS forecasts.

The overall reliability and sharpness of a set of probabilistic forecasts can be quantified by the Brier score (Brier 1950, Wilks 2006). The Brier score can be understood as the mean squared error for probabilistic forecasts. Smaller Brier scores imply more accurate forecasts.

Since the Brier scores of the ECMWF and GFS MOS systems were small, we used the Brier score percent improvement (Brier skill score) to compare ECMWF MOS and GFS MOS PoP12 guidance. As with any skill score, a reference system must be chosen. Here, we have used the 0000 UTC GFS MOS PoP12 guidance. Figure 7 shows the Brier skill scores of the 0000 UTC ECMWF MOS and 1200 UTC GFS MOS guidance relative to the 0000 UTC GFS MOS guidance. For both the cool and warm seasons, the ECMWF MOS demonstrated more skill than the GFS MOS for both cycles across all projections. This was was especially evident in the cool season, short-range projections where an average 12% improvement over the 0000 UTC GFS MOS guidance could be seen in the Brier skill score. This overall improvement can likely be attributed to the ECMWF's ability to better resolve the evolution of longwave and shortwave weather patterns across the globe. The Brier score percent improvement was not as pronounced in the warm season. This is likely due in part to the difficulty associated with predicting convection during this season. Although no statistical significance testing was done, the PoP12 verifications shown here, as measured by reliability, sharpness, and Brier score improvement, strongly suggest that ECMWF MOS PoP12 forecasts are more skillful than GFS MOS PoP12 forecasts.

5. SUMMARY AND CONCLUSION

MDL has developed an experimental 0000 UTC ECWMF MOS short- and extended-range station-based product. The guidance contains many of the weather elements that are found in MDL's operational, station-based GFS MOS system. In this paper, we have evaluated the performance of ECMWF MOS and operational GFS MOS guidance for the weather elements of 2-m temperature and dewpoint, MaxT, MinT, and PoP12 for the 2012-2013 cool season and 2013 warm season. The verification results shown here were generated from a matched set of independent forecasts made on 334 CONUS and OCONUS METAR stations.

We have shown that 0000 UTC ECMWF MOS guidance outperformed 0000 UTC GFS MOS guidance for all weather elements examined here. This was true across all projections and seasons. These results (at least for MaxT, MinT, and PoP12) can be interpreted as the accuracy of the 0000 UTC GFS MOS having about the same accuracy as the 0000 UTC ECMWF MOS issued 1.0 to 1.5 days earlier. We have also shown that the subsequent 1200 UTC GFS was generally less accurate than the 0000 UTC ECMWF MOS. This of course does not necessarily imply that on a case-by-case basis the ECMWF MOS always provides the more accurate forecast. In fact, real-time verification of these two systems shows quite a few instances when GFS MOS forecasts verify better.

One might argue that the dominance of the ECMWF MOS guidance is primarily due to the somewhat dated GFS MOS equations. In fact, the last major update to the GFS MOS system was implemented in 2010. With the planned redevelopment of the GFS MOS, the gap will likely be narrowed. However, we believe that as long as the ECMWF model remains more skillful than GFS model, ECMWF MOS forecasts will also generally remain more skillful than GFS MOS forecasts.

6. FUTURE WORK

MDL is currently in the process of developing 0000 and 1200 UTC ECMWF MOS guidance that will be implemented operationally.[1] This ECMWF MOS product will replace the current experimental ECMWF MOS. The upgraded ECMWF MOS package will include all the current weather elements that are in MDL's operational station-based GFS MOS package. The development will include both a larger and more recent sample of ECMWF model data, along with guidance for an additional several hundred METAR stations. Given adequate resources, MDL is planning to implement this package by the end of 2014 or early 2015.

[1] ECMWF data is provided to the National Weather Service for internal use only. ECMWF model data is considered proprietary and confidential. The ECMWF MOS guidance derived from ECMWF data is restricted for internal use only.

7. ACKNOWLEDGEMENTS

We would like to express our appreciation to Geoff Wagner of MDL who generated some of the figures for this paper. We would also like to thank Paul Dallavalle, Bob Glahn, and Mark Oberfield for reviewing this manuscript.

8. REFERENCES

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

Dallavalle, J. P. and Cosgrove, R. L., 2004: GFS-Based MOS Guidance – The Short-Range Alphanumeric Messages from the 0000/1200 UTC Forecast Cycles, *MDL Technical Procedures Bulletin* No. 05-03, NOAA, U.S. Dept. of Commerce, 13 pp.

Dallavalle, J. P., M. C. Erickson, and J. C. Maloney, 2004: Model Output Statistics (MOS) Guidance for Short-Range Projections. *Preprints 20th Conference on Weather Analysis and Forecasting, Seattle, Amer. Meteor. Soc.*

Erickson, M. C. and K. L. Carroll, 1999: Updated MRF-based guidance: Another step in the evolution of objective medium-range forecasts. *Preprints, 17th Conference on Weather Analysis and Forecasting, Denver, Amer. Meteor. Soc.*

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, 2012: Verification of TIGGE Multi-model and ECMWF Reforecast-Calibrated Probabilistic Precipitation Forecasts over the contiguous United States, *Mon. Wea. Rev*., **140**, 2232-2252.

Maloney, J. C., K. K. Gilbert, M. N. Baker, and P. E. Shafer, 2010: GFS-Based MOS Guidance – The Extended-Range Alphanumeric Messages from the 0000/1200 UTC Forecast Cycles, *MDL Technical Procedures Bulletin* No. 2010-01, NOAA, U.S. Dept. of Commerce, 12 pp.

Simmons, A. J., Burridge, D. M., Jarraud, M., Girard, M. and Wergen, W. 1989: The ECMWF medium-range prediction models development of the numerical formulations and the impact of increased resolution. *Meteorol. Atmos. Phys.*, **40**, 28–60

White, P.W.: IFS Documentation, ECMWF, Reading, 2002.

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

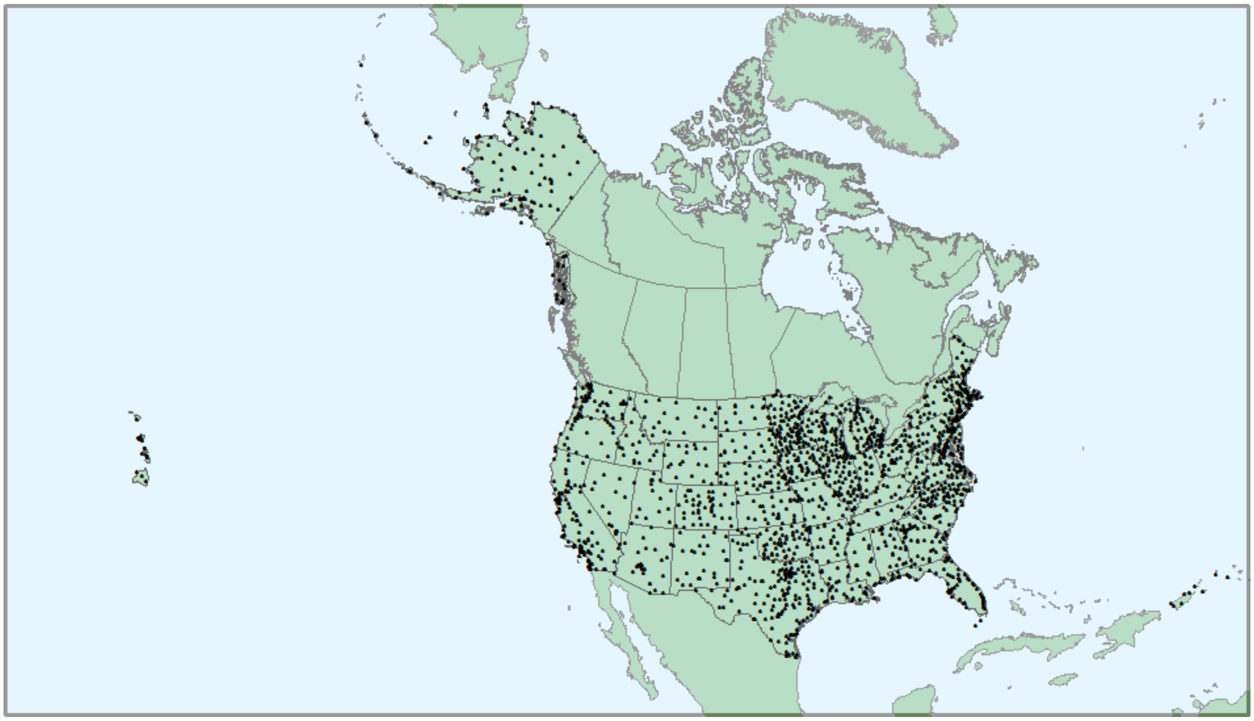

Figure 1. Geographical extent of MDL's archive of ECMWF model data and locations of stations used to develop experimental ECMWF MOS.

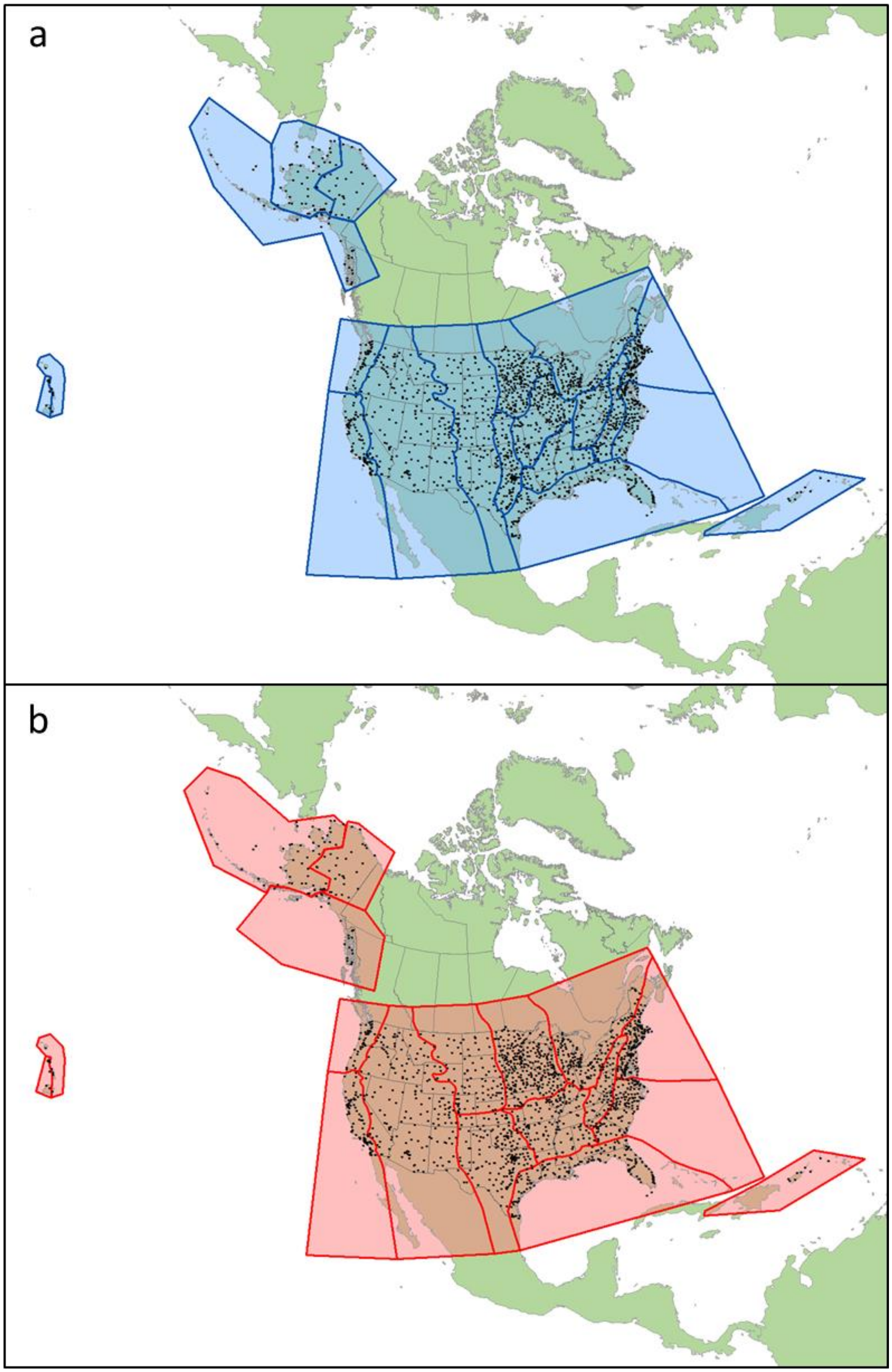

Figure 2. Regions used for developing 0000 UTC ECMWF MOS PoP12 (a) cool and (b) warm season equations.

Table 1. List of 0000 UTC ECMWF model fields offered as predictors for the development of 2-m temperature, 2-m dewpoint, Daytime MaxT, and Nighttime MinT 0000 UTC ECMWF MOS equations.

ELEMENT	**LEVEL**
12-H TOTAL PRECIPITATION	SURFACE
6-H TOTAL PRECIPITATION	SURFACE
DEWPOINT	700 MB
DEWPOINT	850 MB
DEWPOINT	925 MB
DEWPOINT	1000 MB
DEWPOINT	2-M
LAPSE RATE	1000-850 MB
RELATIVE HUMIDITY	700 MB
RELATIVE HUMIDITY	850 MB
RELATIVE HUMIDITY	925 MB
RELATIVE HUMIDITY	1000 MB
TEMPERATURE	700 MB
TEMPERATURE	850 MB
TEMPERATURE	925 MB
TEMPERATURE	1000 MB
TEMPERATURE	2-M
THICKNESS	850-1000 MB
THICKNESS	500-1000 MB
THICKNESS	500-700 MB
TOTAL CLOUD COVER	
VERTICAL VELOCITY	700 MB
WIND SPEED	700 MB
WIND SPEED	850 MB
WIND SPEED	925 MB
WIND SPEED	10-M
TOTAL COLUMN OF WATER	

Table 2. Same as Table 1, except for PoP12.

ELEMENT	**LEVEL**
EARTH U WIND	850 MB
EARTH U WIND	500 MB
EARTH U WIND	300 MB
EARTH V WIND	850 MB
EARTH V WIND	500 MB
K-INDEX	
MEAN RELATIVE HUMIDITY	1000-850 MB
MEAN RELATIVE HUMIDITY	1000-700 MB
MEAN RELATIVE HUMIDITY	1000-500 MB
MEAN RELATIVE HUMIDITY	850-500 MB
MOISTURE DIVERGENCE	850 MB
MOISTURE DIVERGENCE	500 MB
RELATIVE VORTICITY	300 MB
RELATIVE VORTICITY	500 MB
SEA LEVEL PRESURE	SURFACE
LAPSE RATE	1000-925 MB
LAPSE RATE	925-850 MB
U-WIND	10-M
V-WIND	10-M
MASS DIVERGENCE	300 MB
MASS DIVERGENCE	500 MB
MASS DIVERGENCE	700 MB
MASS DIVERGENCE	850 MB
MASS DIVERGENCE	925 MB
RELATIVE VORTICITY	700 MB
RELATIVE HUMIDITY * VERTICAL VELOCITY	700 MB

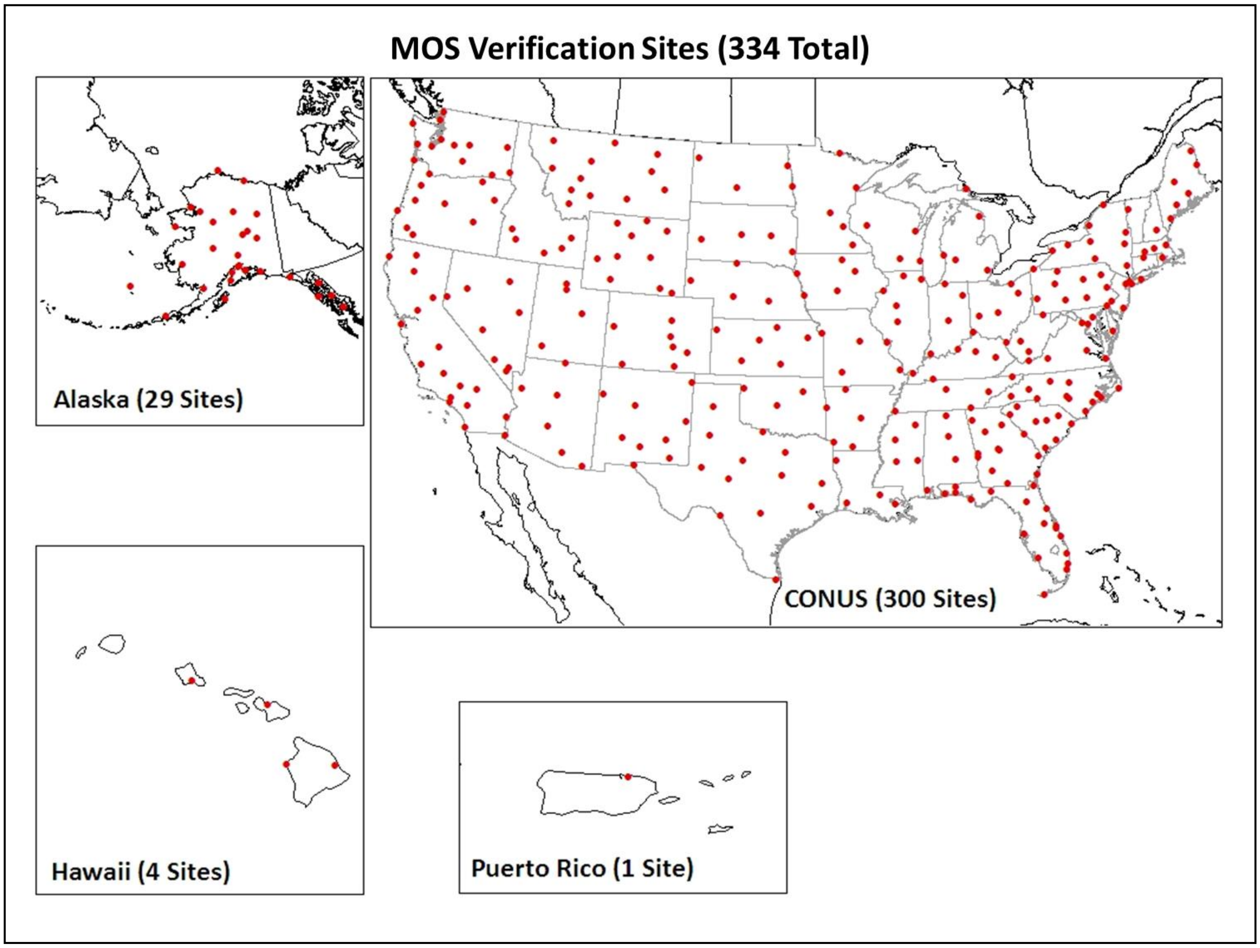

Figure 3. Locations of the 334 METAR stations used to verify cool and warm season experimental 0000 UTC ECMWF MOS and operational 0000 and 1200 UTC GFS MOS forecasts.

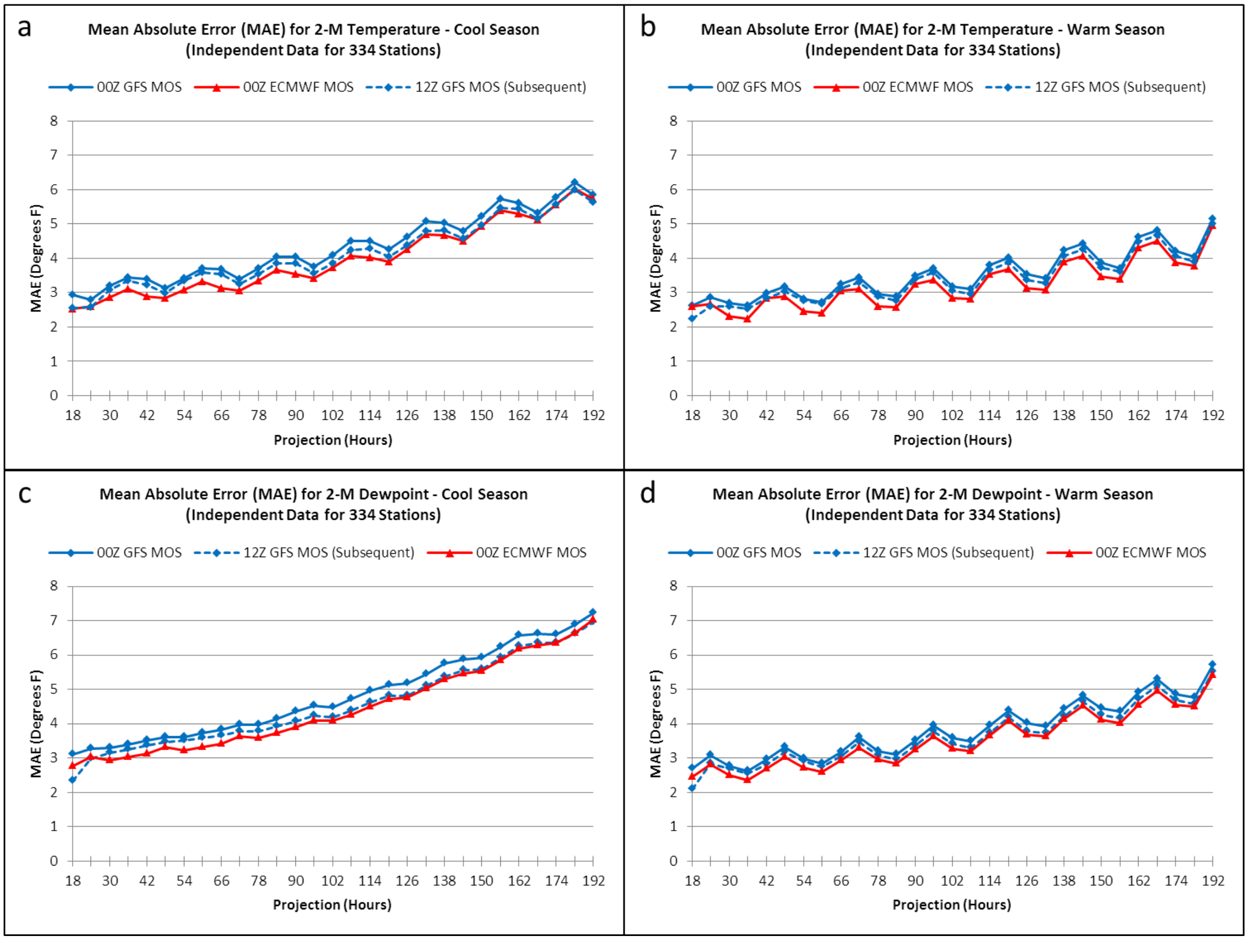

Figure 4. ECMWF MOS and GFS MOS mean absolute error (MAE) scores for 2-m temperature (a) 2012-2013 cool and (b) 2013 warm season; (c) and (d) Same as (a) and (b) except for 2-m dewpoint forecasts.

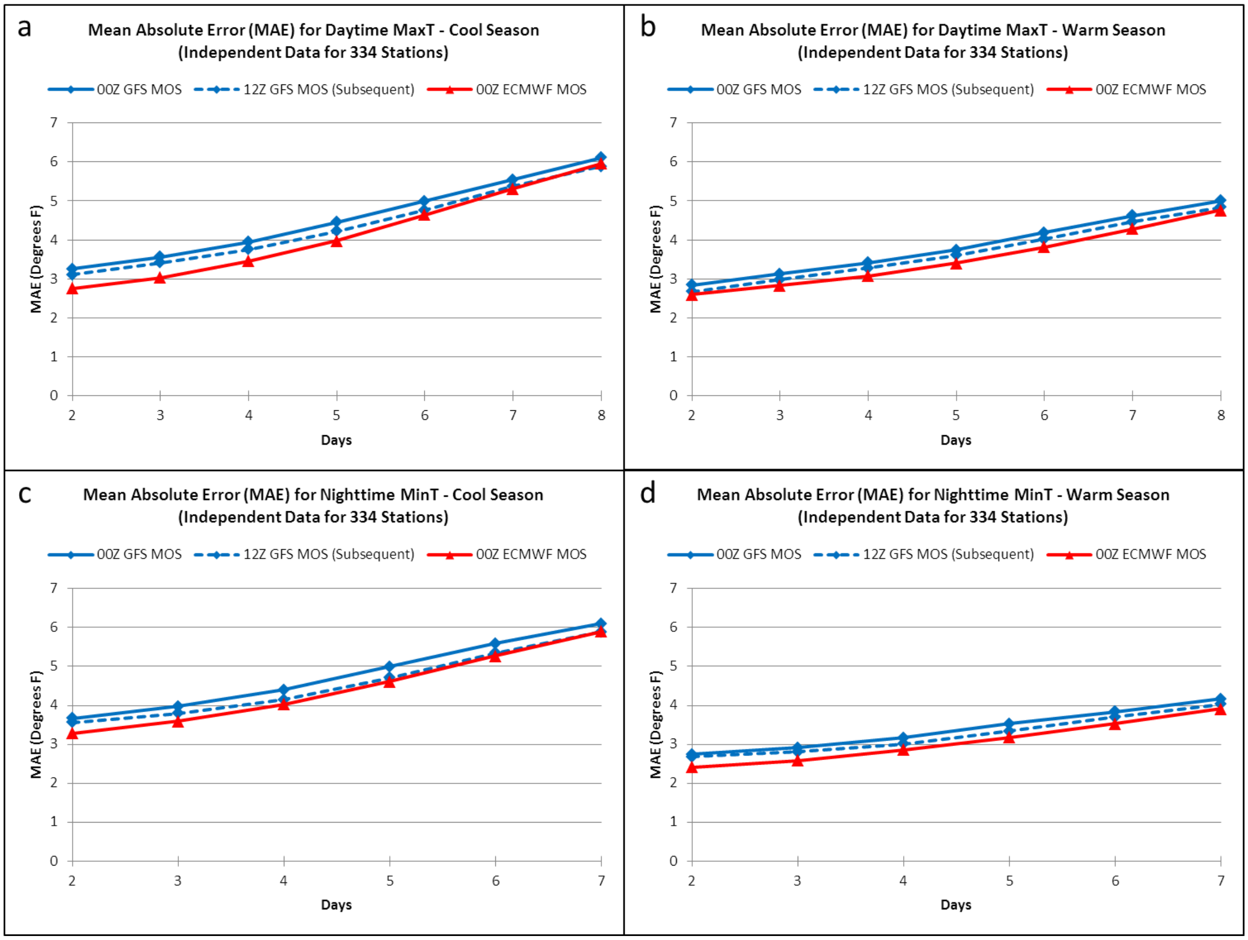

Figure 5. ECMWF MOS and GFS MOS mean absolute error (MAE) scores for daytime MaxT (a) 2012-2013 cool and (b) 2013 warm season; (c) and (d) Same as (a) and (b) except for nighttime MinT forecasts.

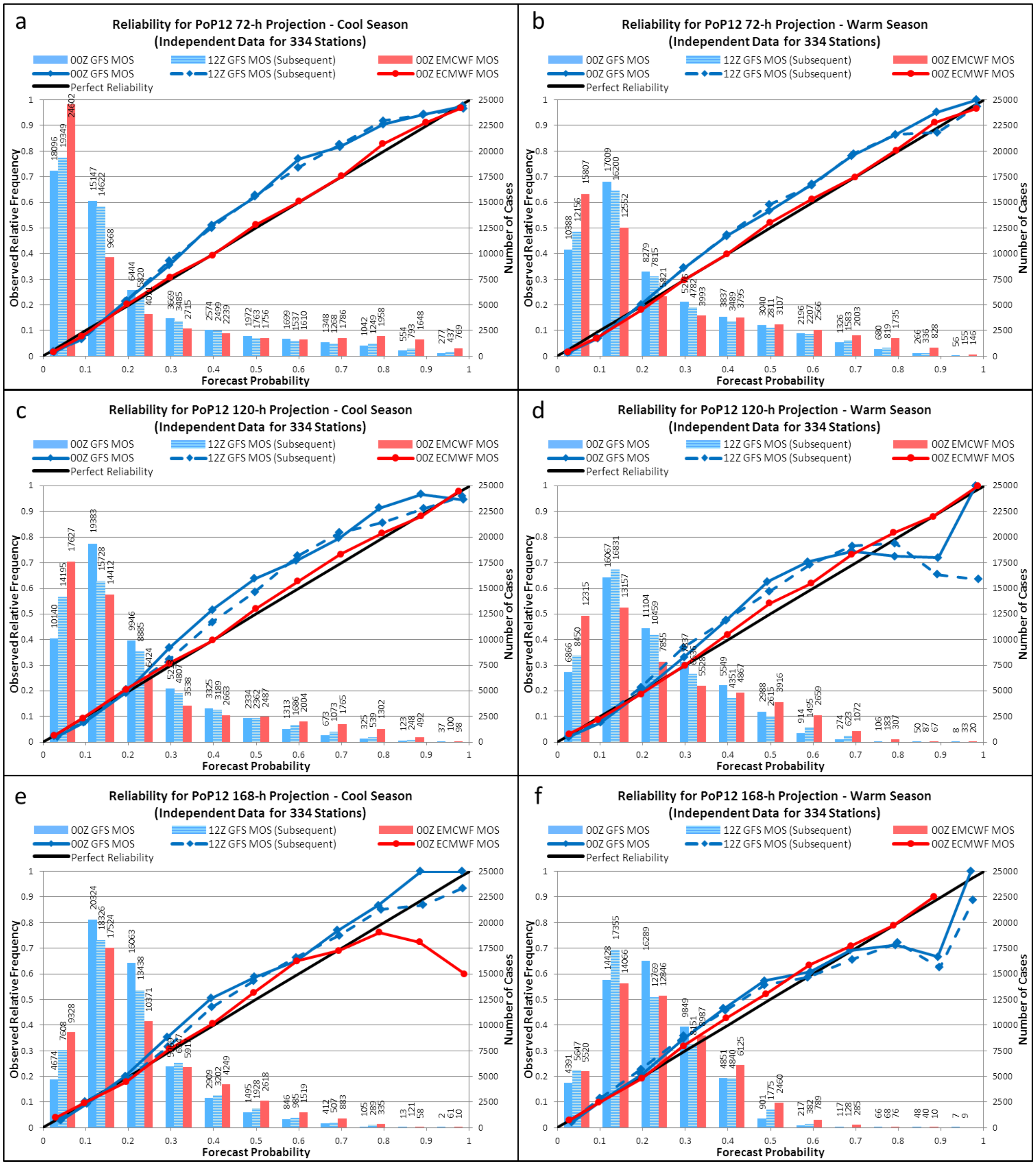

Figure 6. Reliability lines and corresponding histograms for 72-h (a) and (b), 120-h (c) and (d), and 168-h (e) and (f) ECMWF MOS and GFS MOS PoP12 2012-2013 cool and 2013 warm season forecasts. Note the overall better reliability and sharpness of ECMWF MOS forecasts. Because the number of cases in the two highest bins is rather low, the reliability may not be meaningful.

Figure 7. Brier score percent improvement of 0000 UTC ECMWF MOS and subsequent 1200 UTC GFS MOS forecasts over 0000 UTC GFS MOS forecasts for the (a) 2012-2013 cool and (b) 2013 warm season. Note that ECMWF MOS PoP12 forecasts show skill over GFS MOS forecasts at all projections, especially in the short-range.