4.3 RESULTS OF SENSITIVITY TESTING OF MOS WIND SPEED AND DIRECTION GUIDANCE USING VARIOUS SAMPLE SIZES FROM THE GLOBAL ENSEMBLE FORECAST SYSTEM (GEFS) RE-FORECASTS

David E. Rudack*, Meteorological Development Laboratory
Office of Science and Technology
National Weather Service, NOAA

1. INTRODUCTION

A collaborative effort between NOAA's Oceanic and Atmospheric Research (OAR) and the Meteorological Development Laboratory (MDL) is underway to statistically calibrate model output as part of the National Blend of Global Models Project. MDL is employing Model Output Statistics (MOS) (Glahn and Lowry 1972) to statistical post-process GEFS model forecasts for input to the national blend. As with any MOS development, a representative training sample size is crucial for developing stable forecast equations to generate guidance. In the testing phase of this project, MDL has used NOAA/OAR's 30 year GEFS reforecast data set (Hamill et al. 2014) to determine the optimum training sample length to accurately predict both the common and extreme events for a variety of weather elements.

Due in part to time constraints and human resource availability, only 0000 UTC GEFS mean reforecasts were used. In particular, MDL evaluated the performance of MOS-based 2-m temperature, wind speed and direction, and precipitation forecasts. What follows is a summary of the wind speed and direction MOS development procedure and the verification results from various sample size testing of GEFS reforecasts. In this extended abstract, Section 2 explains the development and sampling method employed in this study. Section 3 summarizes the verification results from various sample-size sensitivity experiments for both GEFS MOS wind speed and wind direction forecasts. A short summary along with concluding remarks can be found in Section 4.

2. DEVELOPMENT

2.1 *Methodology*

To address the question of how many years of GEFS reforecast data is necessary to properly calibrate MOS station-based wind forecasts, the Blender Team at MDL decided to explore the effects of sample length sizes of 15, 10, 5, 3, 2, and 1 year(s). In this context, a "year" refers to a season so that 15 "years" of data refers to 15 seasons – either "cool" (October-March) or "warm" (April-September). Seasonal stratification generally works best at capturing the relationships between predictors and predictand. To illustrate how the equation development and cross-validation verification was done, consider the following example: For the period of 1985-1999, MOS equations were developed that predict wind speed, u-, and v- wind components for projections 6- through 192-h every six hours. From these equations, 2000 warm season wind speed and direction forecasts were generated and verified. Next, using the 15-year sample period of 1986-2000, a second set of equations for the same projections was developed. From these new equations, MOS forecasts of wind speed and direction were generated and verified for the 2001 warm season. In this manner, marching along from 1985 through 1999, 14 separate sets of equations and independent forecasts were developed from this 15-year sliding sample window. This methodology was used to generate all equations and independent forecasts for the remaining sample years. This technique allowed for the cross-comparison of independent forecasts generated by each of the sample lengths (Table 1).

2.2 *Model and Observational Data*

The full GEFS sample used in this study was comprised of 29 years of reforecast data (1985-2013) projected onto a global 1-degree latitude/1-degree longitude grid. Reforecasts from each of the 51 members and the ensemble mean were available. However, due in part to time constraints, only single-valued ensemble mean forecasts were used. Since the regression analysis was performed at stations, GEFS data was first interpolated to 334 reliable reporting stations in the CONUS and OCONUS (Fig. 1). The corresponding hourly observations for this period of wind speed, and u- and v- components were collated

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

for the regression and verification portions of this study.

2.3 *Model and Observational Data*

Single-station MOS equation development is MDL's preferred method for developing stable regression equations when a sufficient data sample (~180 cases) is available. In this way, single-station equations can be tuned to the local weather observed at a particular site. Meeting this case number threshold for this study was a non-issue except for the 1-year sample, where the tolerance was lowered to 150 cases.

A consistent pool of interpolated GEFS predictors valid at each particular projection was offered to the regression for each station at each forecast projection (Table 2)[1]. In this way, using step-wise, forward selection, equations for 10-m wind speed, and the u- and v- wind components. For any particular projection, each station's prediction set of equations contained the same set of predictors, but possessed different coefficients tuned to each of the three predictands, wind speed, u- and v-components. Not surprisingly, GEFS forecasts of 10-m wind speeds, u and v were chosen as the most important predictors by the regression for all sample periods, regardless of season. Other levels such as 1000 mb and 925 mb were also found to be useful.

2.4 *Generating MOS Wind Speed and Wind Direction Forecasts*

GEFS MOS forecasts (henceforth referred to as MOS) for wind speed, and u- and v wind components were generated beginning at the 6-h projection and ending at the 192-h projection. Because the regression equations for wind speed have a tendency to underestimate the wind speed forecasts above the mean of the predictand distribution, a technique (Carter and Schwartz 1985) referred to as "partial inflation" was used to artificially inflate wind speeds above the mean. This increased the frequency of higher wind speed forecasts but did not influence wind speed forecasts below the predictand mean. Although a negative effect of inflation is to increase the overall MAE scores, it is an effective method to introduce higher wind speeds into MOS forecasts. The MOS wind direction forecasts were computed from the u- and v- wind components. Once wind speeds and wind directions were calculated, post-processing checks were made to ensure that (1) MOS wind speeds were greater than or equal to zero and (2) all wind direction forecasts were set to calm when either the wind speed forecasts were less than 0.5 knots or when both the u- and v- wind components equaled 0.

3. WIND SPEED VERIFICATION RESULTS

3.1 *Contiguous Daily Sampling*

Verification results of MOS wind speed and wind direction forecasts for the 15-, 10-, 5-, 3-, 2-, and 1-year samples are shown in Figure 2. MAE scores were only generated for those cases where a 10 knot or greater wind was observed. The top panels 2(a) and 2(b) show the MAE scores for MOS wind speeds forecasts spanning the 2000/2001 - 2012/2013 cool seasons and 2000-2013 warm seasons, respectively. MAE scores show a general increasing trend during the 6- through 192-h projections for all six samples with the 15-year sample performing the best and the 1-year sample performing the worst. With the exception of the 1-year sample, all other sample-years have MAE scores that are somewhat clustered with each other. While the differences in MAE scores between the 2- and 15-year samples in the early projections are somewhat clustered, a more sizeable difference can be seen beyond the 120-h projection. It is in these extended range projections that the larger sample sizes appear to be paying dividends. One additional take-home message from these graphs is that the 5-year sample performs almost as well as the 15- and 10-year samples throughout the entire projection period. This is an important result because this suggests that the accuracy of recalibrated MOS reforecasts using a 5-year sample approaches the accuracy MOS forecasts calibrated from both a 15- and 10-year sample. Figures 2(c) and 2(d) display the MAE scores for wind direction where a 10 knot or greater wind was observed. While the clustering between different sample years is tighter than what is seen for wind speed, the overall pattern and comments noted above for wind speed are true for wind direction as well.

[1] Climatic predictors such as sine and cosine of the day were also offered to the regression to compensate for the decline in model performance in the extended range.

3.2 *Non-Contiguous Daily Sampling*

Presumably, the longer training sample such as 15 years should capture a larger number of high wind events than the shorter but yet competitive sample size of 5 years. That said, a suggestion was made to develop equations for the 15-year sample in the manner described in Section 2.4 but only include every third day. In this way, a 15-year sampling period is still maintained but the sample size is limited to 5-years. Figures 3(a) and 3(b) show the verification results of this 15-year, every third day sample alongside the contiguous 15- and 5-year samples for MOS wind speed forecasts. For the most part, especially in the extended projections, a slight but noticeable improvement can be seen in the 15-year, every third day sample over the contiguous 5-year sample. This suggests that given the choice of having two sample of the same duration, it is more desirable to have the sample that covers a longer historical period. The overall results for MOS wind direction forecasts (Figs. 3(c) and 3(d)) are similar to those of MOS wind speeds, with once again the greatest differences between the three samples appearing in the later projections.

3.3 *Sensitivity to Contiguous and Non-Contiguous Sampling of Stronger Winds*

To further investigate the sensitivity of sample size as it relates to the calibration of higher MOS wind speeds, MOS wind speed forecasts for the same three sample sizes noted in Section 3.2 were verified for a portion of the NWS wind speed thresholds identified in Table 3 (NWS Products and Services Reference Guidebook 2013). Specifically, threat scores were calculated for wind speeds $\geq$ 15 mph (13 knots), $\geq$ 20 mph (17 knots), $\geq$ 30 mph (26 knots), and $\geq$ 40 mph (34 knots). Figure 4 shows that for the first three thresholds, the threat scores for all three samples at any given projection are virtually identical (independent of season) showing no distinct advantage of one sample size over the other. For the fourth threshold of $\geq$ 40 mph, the threat scores do vary somewhat between the sample sizes. However, since the magnitude of threat scores at this threshold level is quite small, any cross comparisons between sample sizes is not meaningful.

4. SUMMARY AND CONCLUSION

This paper has explored the effects of calibrating MOS wind speed and direction forecasts with various sample sizes of GEFS mean, single-valued reforecasts. The results shown here suggest that while there is an MAE improvement in overall MOS wind speed forecasts with increasing sample size, especially in the extended range, the relative improvements wane for sample sizes of five years or greater. In fact, when verifying MOS wind speed forecasts at various NWS set thresholds of $\geq$ 15 mph, $\geq$ 20 mph, $\geq$ 30 mph, and $\geq$ 40 mph, all sample sizes of 5 years or greater perform almost equally. Furthermore, the MAE performance of MOS wind direction forecasts that were generated from sample sizes of five years or greater are generally insensitive to the sample length even in the extended range. Given these results, it appears that the added resources required for generating and processing samples longer than 5 contiguous years may not be justified.

5. ACKNOWLEDGEMENTS

The author would like to thank Tamarah Curtis for generating many of the figures in this paper and Bob Glahn for reviewing this manuscript.

6. REFERENCES

Carter, G. M., and B. E. Schwartz, 1985: The use of Model Output Statistics (MOS) for predicting surface wind. NWS Technical Procedures Bulletin No. 347, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 11 pp.

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., *et al.*, cited 2014: A Recommended Reforecast Configuration for the NCEP Global Ensemble Forecast System. [Available on line at http://www.esrl.noaa.gov/psd/people/tom.hamill/White-paper-reforecast-configuration.pdf.]

NWS Products and Services Reference Guidebook, cited 2013: NWS/OCWWS. [Available on line at http://www.nws.noaa.gov/os/guide/.]

Table 1. Sample sizes and dates catalogued by season (cool and warm) used to generate GEFS MOS wind speed and u- and v- wind component equations. The dates used to independently verify each sample can be found in the column labeled "Independent Verification Periods."

Sample Sizes	Training Periods (Cool Season)	Independent Verification Periods (Cool Season)	Training Periods (Warm Season)	Independent Verification Periods (Warm Season)
15 Years	1985/1986 – 1999/2000 1986/1987 – 2000/2001 . . 1997/1998 – 2011/2012	2000/2001 2001/2002 . . 2012/2013	1985 - 1999 1986 - 2000 . . 1998 - 2012	2000 2001 . . 2013
10 Years	1990/1991 – 1999/2000 1991/1992 – 2000/2001 . . 2002/2003 – 2011/2012	2000/2001 2001/2002 . . 2012/2013	1990 - 1999 1991 - 2000 . . 2003 - 2012	2000 2001 . . 2013
5 Years	1995/1996 – 1999/2000 1996/1997 – 2000/2001 . . 2007/2008 – 2011/2012	2000/2001 2001/2002 . . 2012 /2013	1995 - 1999 1996 - 2000 . . 2008 - 2012	2000 2001 . . 2013
3 Years	1997/1998 – 1999/2000 1998/1999 – 2000/2001 . . 2009/2010 – 2011/2012	2000/2001 2001/2002 . . 2012/2013	1997 - 1999 1998 - 2000 . . 2010 - 2012	2000 2001 . . 2013
2 Years	1998/1999 – 1999/2000 1999/2000 – 2000/2001 . . 2010/2011 – 2011/2012	2000/2001 2001/2002 . . 2012/2013	1998 - 1999 1999 - 2000 . . 2011 - 2012	2000 2001 . . 2013
1 Year	1999/2000 2000/2001 . . 2011/2012	2000/2001 2001/2002 . . 2012/2013	1999 2000 . . 2012	2000 2001 . . 2013

Table 2. List of GEFS mean predictors used in calibrating MOS wind speed and direction sample size sensitivity experiments.

Predictor	Vertical Level
Wind speed	10-m
u-wind component	10-m
v-wind component	10-m
Earth u-wind component	1000, 925, 850, 700, 500 (mb)
Earth v-wind component	1000, 925, 850, 700, 500 (mb)
Lapse Rate	1000-925, 1000-850, 1000-700, 925-850, 850-700 (mb)
Wind Speed	1000, 925, 850, 700, 500 (mb)
Relative Mass Divergence	925, 850, 700, 500 (mb)
Relative Vorticity	925, 850, 700, 500 (mb)
Temperature at [t-(t+12)]	850 (mb)
925 mb to 10-m wind speed Ratio	
K- Index	

Table 3. NWS definitions for various wind speed thresholds.

Wind Speed (mph)	Description
0 to 5 mph Light	Calm
5 to 20 mph	None used
15 to 25 mph	Breezy, Brisk, Blustery
20 to 30 mph	Windy
30 to 40 mph	Very Windy
40 to 73 mph	High, Strong, Damaging
74 mph or Greater	Hurricane Force

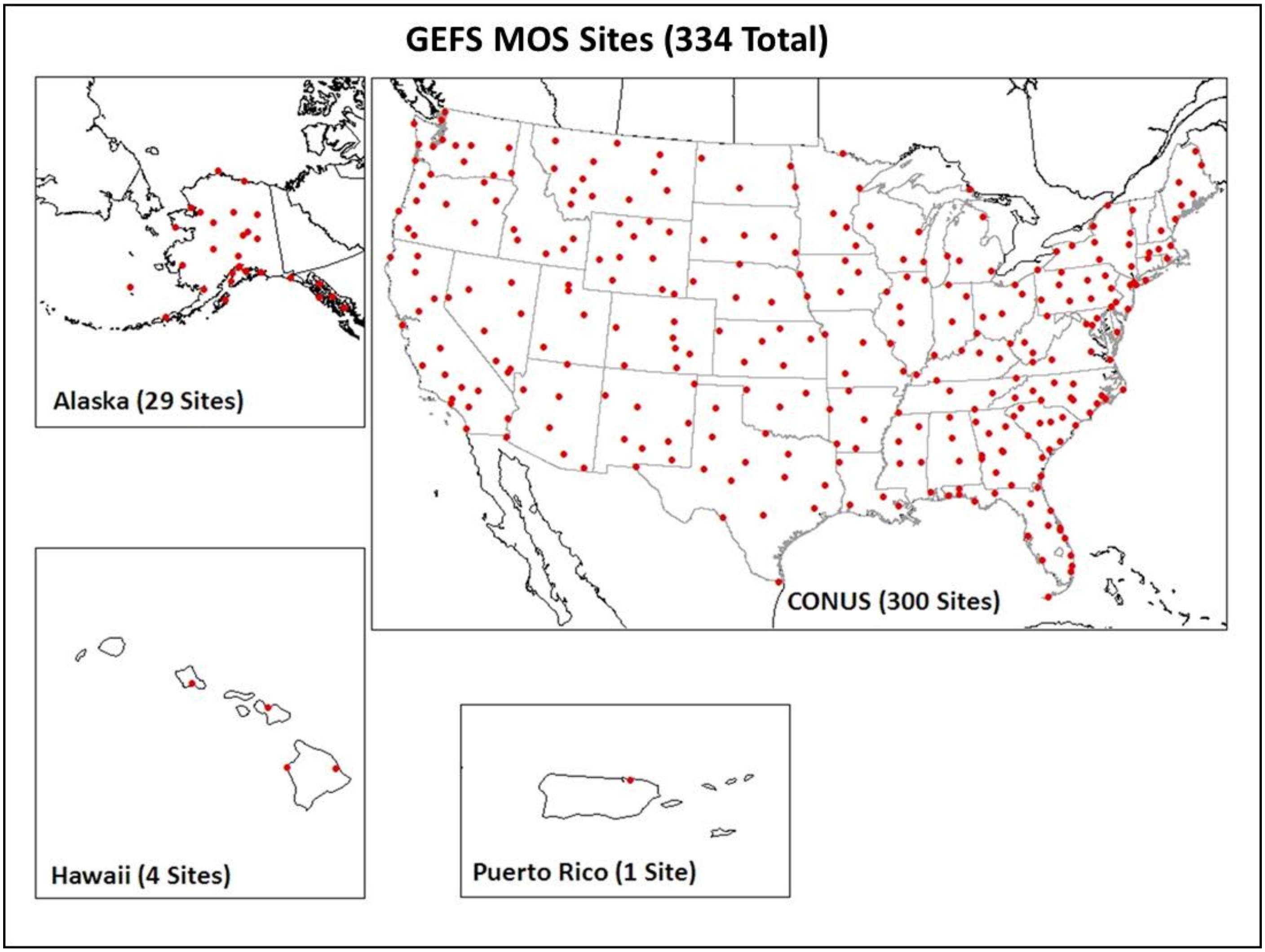

Figure 1. Locations of the 334 METAR stations used to generate and verify 0000 UTC GEFS MOS wind speed and direction forecasts.

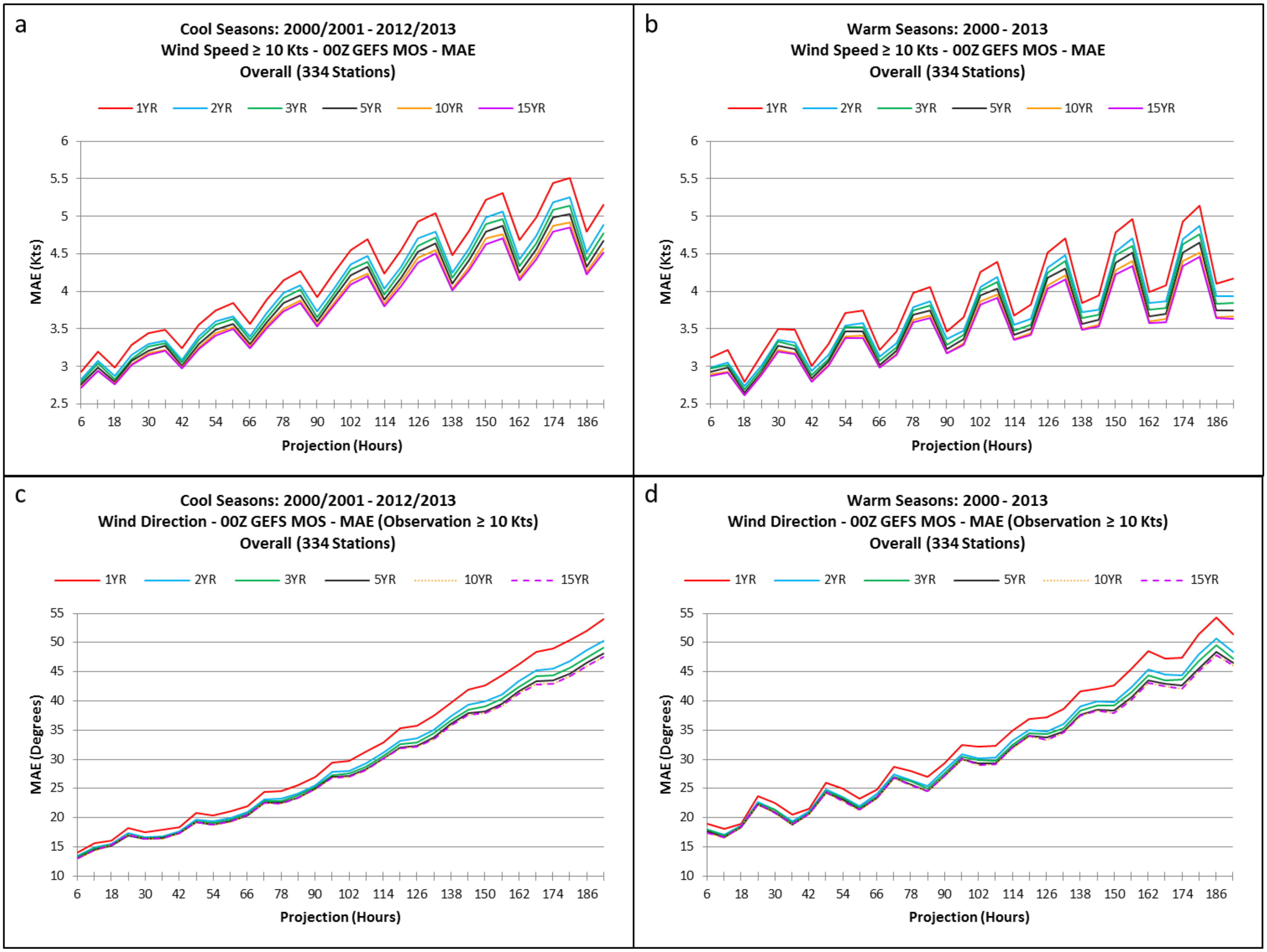

Figure 2. Cool and warm season mean absolute error (MAE) scores by projection for GEFS MOS wind speed and direction forecasts for the 15-, 10-, 5-, 3-, 2-, 1-year samples.

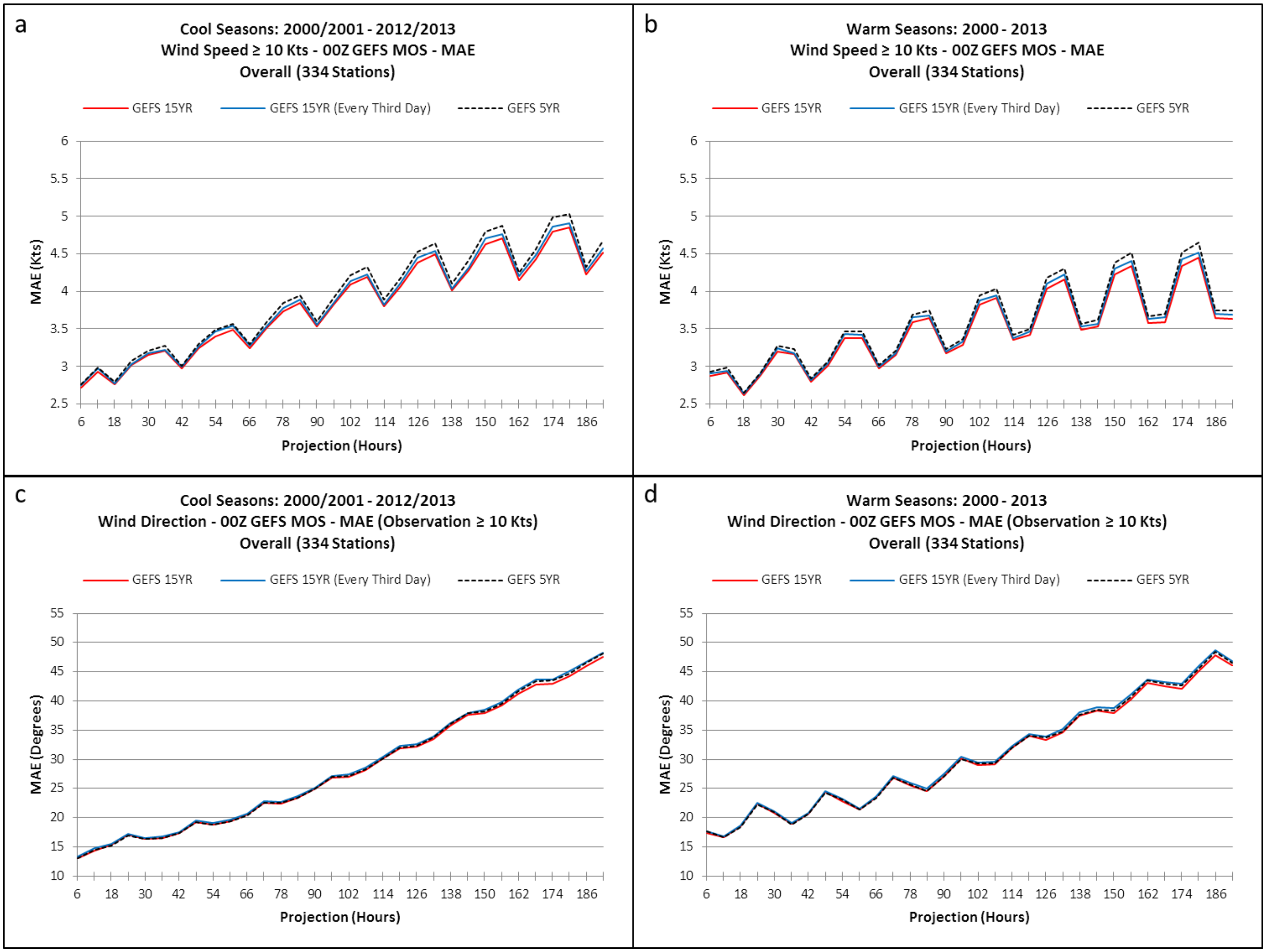

Figure 3. Cool and warm season mean absolute error (MAE) scores by projection for GEFS MOS wind speed and direction forecasts for the contiguous 15- and 5-year samples, and 15-year sample sampled every third day.

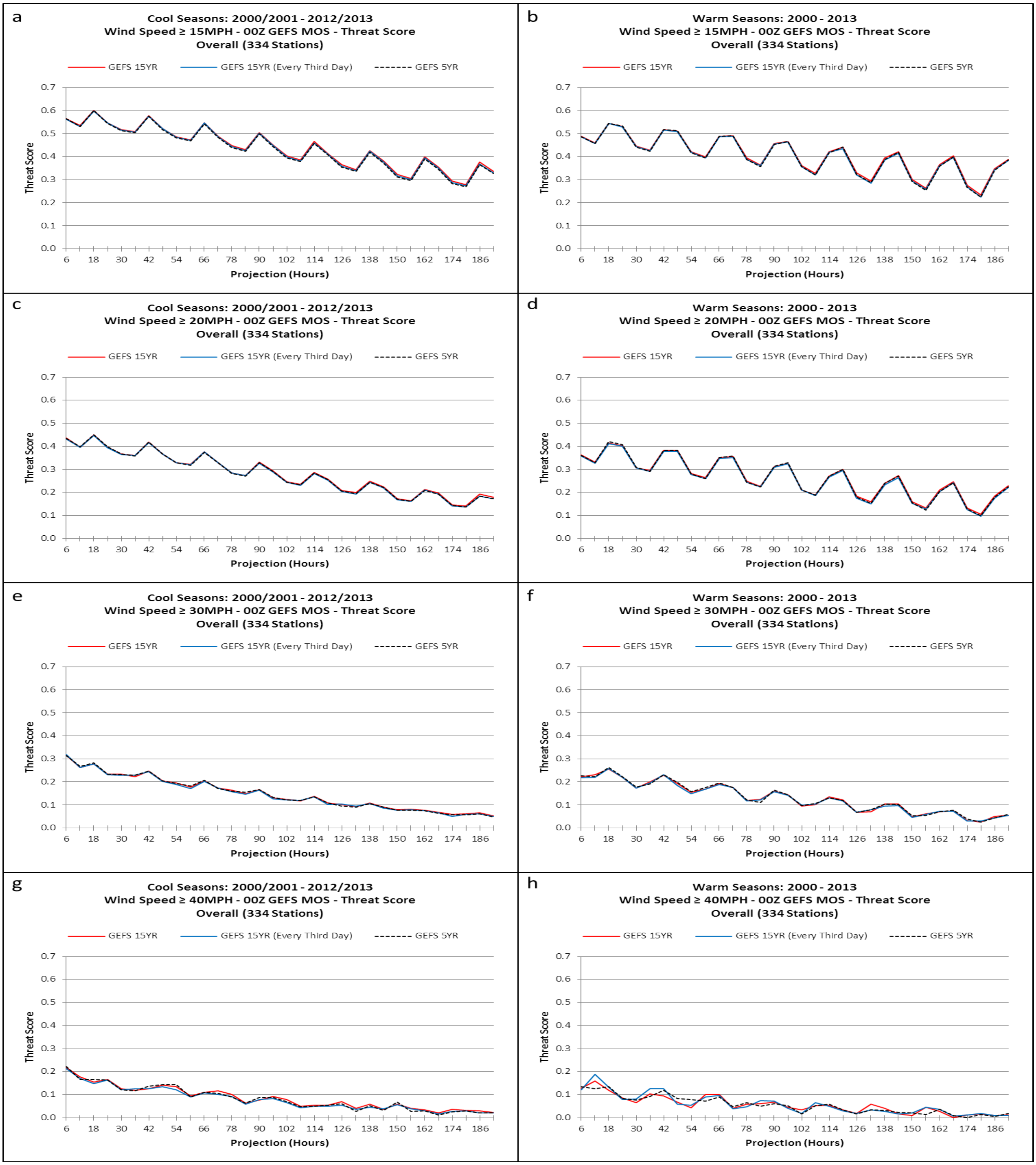

Figure 4. Threat scores of GEFS MOS wind speeds for the contiguous 15- and 5-year samples, and 15-year sample sampled every third day at thresholds of ≥15 mph, ≥20 mph, ≥30 mph, and ≥40 mph.