6.3

DETERMINING AN OPTIMAL DECAY FACTOR FOR BIAS-CORRECTING MOS TEMPERATURE AND DEWPOINT FORECASTS

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

1. Introduction

The National Weather Service (NWS) has continuously provided MOS dry bulb temperature and dewpoint temperature forecasts to customers and partners since their introduction in 1978 (TPB, 1978).[1] Initially, the NWP products on which MOS forecasts were based did not contain forecasts of "surface" weather elements, such as 2-m temperature and dewpoint and 10-m wind, and postprocessing was crucial in the use of those NWP products. Today, NWP *does* produce these weather elements directly, and their accuracy makes them prime predictors in the MOS regression equations. Even so, NWP forecasts can be improved by postprocessing, and in particular the NWP forecasts have biases. NWP variables have been bias corrected at the National Centers for Environmental Prediction (NCEP) since 2006 by a method called "decaying average" (Cui, et al. 2012).

MOS forecast relationships developed by the Meteorological Development Laboratory (MDL) have usually been based on so-called cool (October-March) and warm (April-September) season samples. This separation accounts to some degree for the annual biases and inaccuracies that would exist because of different relationships between predictors and predictands in the warm and cool months. Regression equations that produce MOS forecasts give unbiased estimates over the developmental sample period, but they may have bias over intervals within that sample and over other samples, including future forecasts. Biases in operational forecasts can exist because of NWP model changes, which render the developed relationships less than optimum; changing weather regimes that the equations do not handle adequately (e.g., blocking highs); or local environmental changes. A source of MOS error is the inability to "keep up" with the operational model changes. Changes in a model, without redeveloping MOS on an adequate sample, can create larger errors in MOS forecasts.[2] Short-sample bias correction is seen as a way of correcting such biases, and now that NWP models are much better than a few years ago, possibly bias-corrected raw NWP 2-m temperature is of comparable accuracy to MOS temperature. However, Cheng and Steenburgh (2007) found that in 2003 MOS forecasts were better than bias-corrected model forecasts of temperature except in periods of "quiescent large-scale patterns."

According to the usual definition of bias used in meteorology (Wilks 2011, p. 304; Jolliffe and Stephenson 2003, pp. 99-100; Murphy and Daan 1985, p. 385), if the average of a consecutive set of errors is sufficiently different from zero, the forecasts would be considered to be biased.[3] However, most of the authors discussing bias do not specifically state over what period of time the errors would have to be consistently above or below zero for them to consider the forecasts biased.

Routine verification has shown that the MOS temperature and dewpoint forecasts currently being made have some small but consistent bias (Glahn et al. 2009, Fig. 11). This paper shows the results of investigating the adaptability and effectiveness of applying the decaying average method to those forecasts.

2. Decaying Average Algorithm

Cui et al. (2012) call the decaying average implemented at NCEP in 2006 a "Kalman filter type algorithm" but do not state how it relates to Kalman's original work. The use of the Kalman filter (KF) as originally proposed (Kalman 1960) requires knowledge or estimates of several constants and parameters. As stated by Crochet (2008), "The design of a reliable KF procedure can prove to be difficult without any prior knowledge of the noise characteristics." Kalman's

---

[1] Although the Techniques Development Laboratory (now the Meteorological Development Laboratory) has been furnishing NWP postprocessed products for implementation since 1965 (TPB,1968), temperature forecasts at 3-h intervals did not start until 1978 (TPB 1978) and later dewpoint in 1980 (TPB 1980). Before then, statistical temperature guidance was of only maximum and minimum temperature.

* Corresponding author address: Bob Glahn, Meteorological Development Laboratory, National Weather Service Scientist Emeritus, 1325 East West Highway, Silver Spring, MD 20910; email: harry.glahn@noaa.gov

[2] A *major* improvement in the model would likely decrease the MOS errors, but such major improvements are rare; improvement usually comes in small steps.

[3] Some authors have called a single forecast error a "bias" (Cui et al. 2012, p. 398; Cheng and Steenburgh, 2007, p. 1317; Roeger and Stull, 2003, p. 1158), but that terminology can be confusing and does not follow the definition of bias.

original work (1960) has been described in several publications in the meteorological literature (e.g., Roeger et al. 2003; Homleid 1995; Galanis and Anadranistakis 2002; Cheng and Steenburgh 2007) and won't be repeated here. These and other authors have investigated various methods of reducing bias, success usually being measured as an average over a relatively long period (e.g., Yussouf and Stensrud 2007). NCEP's algorithm is attractive because it is very easy and cost effective to implement.

To implement the algorithm, one has only to carry forward a delta and apply it to the current forecast. Then to prepare for the next forecast cycle, the delta "d" would be updated by:

$$d_{t+1} = (1-\alpha)\, d_t + \alpha\, (F - O)_t$$

where d_{t+1} is the delta to apply at time t+1, d_t is the delta applied at time t, F is the forecast "verified" by the observation O at time t, and α is the weight to apply to the most recently calculated forecast error (F - O) at time t. There would optimally be a specific delta for each station and forecast projection. If (F - O) is occasionally missing, zero can be assumed (see the next section for more discussion of this situation).

3. Performance of the Algorithm Applied to MOS Forecasts

Before implementing bias correction into the MOS system, several questions have to be addressed, including not only whether the accuracy and bias of the forecasts are improved, but also what effect this might have on the customers and partners who use the "forecasts" directly or manipulate them by either automated or manual means to achieve a "final" forecast. In the NWS, this further processed guidance would be the "official" forecast.

a. Data and Processing

I used the operational Global Forecast System (GFS)-based MOS temperature and dewpoint forecasts for projections every 12 hours out to 264 hours (11 days) made at 0000 UTC over the period January 1, 2011, through May 31, 2012.[4] This provided a sufficient sample on which to investigate the bias correction method. A change was made to the GFS at NCEP during the summer of 2012 that had some major negative effects on MOS in some portions of the United States, but primarily to wind, not temperature.

Testing on temperature was done separately on the traditional MDL warm (March-September) and cool (October-March) seasons; however, for dewpoint, all 18 months of data were processed together. When comparing forecasts of different projections, one can either match verification times, the forecasts having been made previously at different times, or one can verify the forecasts based on when they were made, and the period over which observations were used will not exactly match for different projections. The latter is the way it is usually done and the way I carried out the verification. Evaluated over 6-month or longer seasons, the offset in verification times is not a significant factor in conclusions for projections up to 11 days. The comparison of MOS forecasts and bias corrected (BC) forecasts for a particular projection were, of course, matched samples. The verification was done with forecasts rounded to whole degrees Fahrenheit, the same precision as the observations. The deltas, however, were carried to three decimal places.

Errors in automated, data-driven processes occasionally occur. An extremely erroneous forecast or observation could cause a very large change in a delta. To address this possibility, a cap on the error can be imposed. The testing reported here was done with such a cap. The magnitude of the error allowed for a 24-h forecast was 20 degrees; for an 11-day forecast it was 40 degrees, and it varied linearly in between by projection. For instance, if an error of 90 degrees were calculated for an 11-day forecast, likely due to an incorrect observation, it would be capped at 40 degrees–still a sizeable error, but the effect would not be disastrous. In an operational setting, it is likely large errors would have already been culled out with quality control procedures.

b. Performance of different alphas

1) BIAS

NCEP and MDL have investigated different alphas up to at least 0.1; NCEP uses 0.02 for all projections for GFS raw model data.[5] Testing at MDL has indicated a higher value might be better for MOS. I tested four different values: 0.025, 0.05, 0.075, and 0.1. I used 1319 stations in the conterminous states (CONUS), the same set used in routine MDL verifications. Of course, a few observations, and even forecasts, may be missing in the sample. The adjustment algorithm can deal with occasional missing data; if there is no forecast or matching observation for a particular projection, no change in the delta is necessary, and on the next cycle for which there is a forecast, the delta last calculated

---

[4] Being operational forecasts, the dewpoint forecasts had been checked with temperature forecasts, and if a forecast was greater than the temperature, it was set to the temperature. If bias correction were implemented, the temperature/dewpoint check would logically come after the bias correction. Using the checked values of dewpoint for this study instead of unchecked ones is not seen as a problem, because the number and magnitude of the changes for consistency are rather small and do not materially affect the temporal characteristics of the forecasts.

[5] Bo Cui 2012 personal communication.

could be used. However, a station may miss reporting for an extended period or stop altogether. MOS forecasts will likely still be made because the model data are available. To perpetuate a particular alpha indefinitely would not be prudent, so a better alternative would be to use the difference as zero and let the delta decay gradually to zero. That is, given no recent history of the station's bias, a correction is not made. In either case, no elaborate backup software is necessary to accommodate one or more missing pairs of data. For the testing reported here, the decay toward zero was not used.

Figure 1 shows the bias for the original MOS temperature and the BC forecasts with all four alphas tested for all stations and projections for the 2011-2012 cool season. It can be seen that there was a significant cold bias in the MOS forecasts that varied each 12 hours and generally was more pronounced with increasing projection. Forecasts verifying at some times of the day have larger errors than forecasts verifying at other times; that is what causes the saw-tooth effect. It can also be seen that the BC forecasts were better in terms of bias, being positive but near zero at 24 h, but still having a cold bias approaching 0.4 degrees F at later projections. It is interesting that for the early projections, the projections with the largest negative bias for MOS had the largest positive bias for BC MOS. This mirror performance lasts until about 228 h, when it reverses and the two curves come into phase. The bias was improved substantially for all alphas. The largest improvement was for alpha = 0.1 and the smallest for 0.025 for all projections, but the differences are small, especially compared to the MOS bias.

The bias varies considerably over the CONUS. Figure 2 is the same as Fig. 1 but for the NWS Central Region, the region for which the MOS bias was greatest, and similarly, Fig. 3 shows biases for the Western Region, the region for which the MOS bias was least (note the different ordinate scales in these figures). Fig. 2 shows a remarkably negative bias of 3 deg. F at 264 h; the BC forecasts are for every projection better, but still drift systemically down to -0.6 degrees F at 264 hours for alpha = 0.05. On the other hand, the biases for the Western region range generally from -0.5 to +0.8 degrees, and the BC forecasts are better for most projections, especially for the larger alphas.

Figure 4 is the same as Fig. 1 except for the 2011 warm season, April through September. The pattern of MOS error by projection is dissimilar to the cool season, and the improvement is questionable except for projections $\leq$ 72 h. For projections of 84 hours and beyond, the MOS bias was quite small, and the correction was not, in general, helpful. As with the cool season, alpha = 0.1 was the best and 0.025 the least helpful.

Rather than separate the dewpoint data into seasons, the whole sample January 2011 through May 2012 was used. For each test done, the process was "cold-started" with a delta = 0 on January 1. Because of this, the verification period for dewpoint started on January 15, allowing a stabilizing period. Figure 5 shows the results. Again, MOS bias is improved, and all alphas produce a very small positive bias. Here, the small values of alpha were best, but the differences are minuscule. Probably one reason the biases of the corrected forecasts do not vary much and are so close to zero is because of the long averaging period over all 16.5 months (January 15, 2011-May 31, 2012).

2) MEAN ABSOLUTE ERROR

Figures 6 and 7 show mean absolute errors (MAE) for all stations for MOS and the four values of alpha tested for temperature for the cool and warm seasons, respectively. Only forecasts verifying at 0000 UTC are shown. Forecasts verifying at 1200 UTC show the same pattern, but the errors are considerably larger, and if both are shown on the same graph, the saw-tooth pattern would present a less clear picture. One might hope with the improvement in bias, that the MAEs would also improve (decrease). Indeed they do, but not with all alphas. For the lower values of 0.025 and 0.05, there is consistent improvement, but with the two higher values of 0.075 and 0.1 for projections > 144 h, there are larger MAEs in the cool season. Because the improvement with the smaller alphas is consistent for all projections, especially for the warm season, and `paired t-tests for each projection show very high significance,[6] some slight improvement can be expected in the future.

Figure 8 is similar to Figs. 6 and 7, except for dewpoint. The conclusion for dewpoint is the same as for temperature; the improvement with bias correction is consistent for all projections, on the order of 0.3 degrees F at shorter projections; improvement is small at longer projections. The improvement does not vary much with alpha, but the lower values are better at longer projections.

3) SMALL AND LARGE ERRORS

Figures 9 and 10 show the percentage of small errors, those < 5 degrees F, for temperature for the cool and warm seasons, respectively. These small errors are more frequent with all alphas, but the smaller alphas give slightly better results. The improvement is roughly equal to a 24-h improvement for the warm season. The same general conclusion is reached from Fig. 11, which is for dewpoint; the relative frequencies of small errors

---

[6] The paired t-tests were corrected for 1st order autoregression (Katz 1985, p. 275; Wilks 2011, p. 147). The errors of the MOS forecasts and the bias corrected ones are highly autocorrelated; the MAE pairs are also autocorrelated, but less so. Even allowing a reduction in degrees of freedom by a factor of 10 for spatial correlation, the paired t-tests show high significance.

are higher for BC forecasts than for MOS, and lower alphas are slightly better.

Figures 12 and 13 are similar to Figs. 9 and 10, except they are for large errors, those $\geq$ 15 degrees. There are very few such errors for the short projection times, and reach 1 percent at 11 days for the cool season. For that season, alpha = .025 is the only one that actually has fewer large errors than MOS, but 0.05 does not have more. For the warm season, alpha = 0.1 is the only one that does not improve on MOS for the longer projections, with 0.025 and 0.05 being of about equal value according to this score.

Figure 14 shows the percentage of large errors for dewpoint. Improvement over MOS holds for all projections, but only for alpha = 0.025 and 0.05 at longer projections.

4) CONSISTENCY OF FORECASTS OVER PROJECTIONS

Long-projection forecasts have more error than short-protection forecasts. As the forecasts for a particular verifying time are improved with time, they should be as consistent as possible, and not “bounce around” from forecast to forecast. The convergence score (Ruth et al. 2009) measures the tendency of the forecasts to march “consistently” from the longer range forecast toward the final short range forecast, a higher score being better with a possible maximum of 1.0. Figure 15 and 16 show this score for the four NWS regions and overall for temperature for the cool and warm seasons, respectively. For the cool season, the higher alphas give worse results than MOS for all regions except the Western; the two lower alphas have essentially the same or better scores than MOS. For the warm season, the highest alpha gives worse results than MOS except in the Western Region. The lower alphas are generally the best, the results for 0.025 and 0.05 being essentially indistinguishable.

Figure 17 shows the convergence scores for dewpoint for the warm and cool seasons combined. Only alpha = 0.025 is able to be about as good as MOS for all regions, but alpha = 0.05 is about the same overall. The two higher alphas are consistently worse than MOS.

5) BIAS BY AVERAGING TIME

The biases over the sample have been shown in previous sections. However, biases can be present over shorter periods, and may be negative for one period and positive for another period, and cancel out over the sample. In order to see what effect the bias correction has on biases of a shorter time period, running means of forecasts and of observations were computed over a 20-day period–a period long enough someone might consider consistently high (or low) forecasts to be “biased.” Then, the MAEs of these running averages were computed. This measures the biases, both positive and negative, over 20-day periods. Figure 18 shows the results for the warm season temperature. All alphas show improvement over MOS, with the larger values giving the smaller biases, consistent with the biases over the whole samples. The results, not shown, are similar for the cool season, as they are for dewpoint for the cool and warm seasons combined as shown in Fig. 19.

Figure 20 shows the 72-h forecast MAEs with running means of various lengths up to 30 days. The conclusion is the same, regardless of averaging time; short-term biases are less for BC forecasts than for MOS for all alphas. This means that if a user is interested in mean forecasts, as opposed to daily non-averaged forecasts, bias corrected forecasts are better than uncorrected, and larger alphas, at least up to 0.1, the maximum value tested, are better. The value at 1 day would be exactly the same as shown in Fig. 6 for 72 hours if the sample periods were the same; for averaged forecasts, the ends of the period are slightly different. It is emphasized that the values in Fig. 20 account for the bias being either positive or negative over the x-day period, and are not the same as the overall biases shown in other figures.

6) PERFORMANCE ON INDIVIDUAL DAYS

Figure 21 shows the MOS 48-h forecasts for Appleton, Minnesota, and the deltas for the four alphas for the cool season forecasts from February 1, 2011, through November 28, with the warm season forecasts April through September omitted. The alphas operative on March 31 were used on October 1. During March, the forecasts were generally too high. As this period started in late February, the deltas became negative, and for alpha = 0.1 reached a low of -5.0 degrees. For the first few days in October, the MOS forecasts were decidedly too low. The deltas rose rather quickly, and for alpha = 0.1 became as high as +2.5 degrees. But by the time this peak was reached, the period of consistently low forecasts was over, and then a march back toward zero delta began.

The delta for alpha = .025 was much more conservative, never reaching a negative value of more than about 2 degrees. It, too, rose starting October 1, but never peaked above zero. The other alphas performed at an intermediate manner as expected.

4. DISCUSSION OF RESULTS

It is clear from the data presented that the more volatile nature of larger alphas, while improving bias, hurt the overall quality of the MOS forecasts. Also, the deltas (corrections) to apply to the MOS forecasts can vary quite dramatically over a few days with large alphas. It is doubtful that users would welcome this volatility imposed by large alphas. Alpha = 0.025 is quite conservative and is competitive with alpha = 0.05, although 0.05 is slightly better in providing accuracy as judged by some scores for some projections. An alpha in that range, 0.025 to 0.05, seems to be best overall.

5. CONCLUSIONS

The decaying average algorithm has been tested with alphas ranging from 0.025 to 0.1. It has been found that all alphas improve the bias of MOS temperature and dewpoint forecasts, but only the alphas of 0.025 and 0.05 consistently improve the MAEs. These results hold for both warm and cool season temperature and for dewpoint over a period consisting of both warm and cool seasons. It is noted that the biases are not only reduced over a several-month period (where the pluses and minuses can cancel out), but also over periods of a few days up to a month (see Figs. 18-20) The lower alphas also provide more forecasts of less than 5 degrees F error and fewer forecasts with >15 degrees error. In addition, only the lower alphas improve the consistency of the MOS forecasts from the longer range to the shorter range, as shown by the convergence score. Plots of the actual performance in terms of the corrections and their volatility indicate that the larger alphas might not be acceptable to the users of MOS forecasts. All of these results lead to the strong indication that bias correction would improve the MOS temperature and dewpoint forecasts if implemented with an alpha in the range 0.025 to 0.05, the exact value within that range not being very important. Testing was done here on data from 2011 and 2012. It is likely that any medium- to long-sample MOS could be improved by this decaying average method.

While the decaying average method could logically be applied to any quasi-continuous variable, such as wind speed, it should be used with caution. This and other successful bias-correcting methods generally produce smoother forecasts than the forecasts to which they are applied. Smoother forecasts may not be desired even if the bias is decreased, because the forecasts will have less variance and the extremes will be reduced. For instance, the usefulness of wind speed forecasts depends critically on strong winds being reliably forecasted, and in some operational situations (e.g., orchard operations), calm or very weak winds. MOS wind speed forecasts, for instance, have been themselves postprocessed to increase their variance above the mean with partial inflation (Schwartz and Carter 1982; Jacks et al. 1990; Glahn and Allen 1966).[7]

Acknowledgments. I am indebted to Tamarah Curtis for furnishing me data and scripts, and for tutoring me in generating the graphs; to Valery Dagostaro for assistance in retrieving verification data from MDL's verification program; to Stephen Olson and Brian Zeigler for arranging for and setting up a script to access MDL's IBM computers from my PC; to Yun Fan, Eric Eagle, and David Rudack for their previous work and for intensifying my interest in this subject, and to Kathryn Gilbert for careful proofreading. This paper is the responsibility of the author and does not necessarily represent the views of the NWS or any other governmental agency.

6. REFERENCES

Cheng, W. Y. Y., and W. J. Steenburgh, 2007: Strengths and weaknesses of MOS, running mean bias removal, and Kalman filter techniques for improving model forecasts over the western United States. *Wea. Forecasting* **22**, 1304-1318.

Crochet, P., 2004: Adaptive Kalman filtering of 2-metre temperature and 10-metre wind-speed forecasts in Iceland. *Meteorol. Appl*. **11**, 173-187.

Cui, B., Z. Toth, Y. Zhu, and D. Hou, 2012: Bias correction for global ensemble forecast. *Wea Forecasting*, **27**, 396-410.

Galanis, G., and M. Anadranistakis, 2002: A one-dimensional Kalman filter for the correction of near surface temperature forecasts. *Meteorol. Appl.*, **9**, 437-441.

Glahn, H. R., and R. A. Allen, 1966: A note concerning the "inflation" or regression forecasts. *J. Appl. Meteor*., **5**, 124-126.

_____, M. Peroutka, J. Wiedenfeld, J. Wagner, G. Zylstra, B. Schuknecht, and B. Jackson, 2009: MOS uncertainty estimates in an ensemble framework. *Mon. Wea. Rev.*, **137**, 246-268.

Homleid, M., 1995: Diurnal corrections of short-term surface temperature forecasts using the Kalman filter. *Wea. Forecasting,* **10**, 689-707.

Jacks, E., B. Bower, V. J. Dagostaro, J. P. Dallavalle, M. C. Erickson, and J.. C. Su, 1990: New NGM-based MOS guidance for maximum/minimum temperature, probability of precipitation, cloud amount, and surface wind. *Wea. Forecasting*, **5**, 128-138.

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification, A Practitioner's Guide in Atmospheric Science*. Wiley, 240 pp.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME. J. Basic Eng*. **82**, 35-45.

Katz, R. W., 1985: Probabilistic models. *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, A. H. Murphy and R. W. Katz, eds., Westview Press, 261-288.

---

[7] Inflation was proposed by Isadore Enger and first applied by Klein et al. (1959). Inflated forecasts are obtained by subtracting from the regression estimate the developmental sample mean, dividing the difference by the (multiple) correlation coefficient, and adding the result to the sample mean. MDL found that this worked well for forecasts above the mean, but those below the mean were too weak, so the established practice is to "partially inflate" by using the procedure on only those regression estimates above the mean. Inflation, either full or partial, will increase the variance of the forecasts and increase the mean square error [see Glahn and Allen (1966) for details.]

Klein, W. H., B. M. Lewis, and I. Enger, 1959: Objective prediction of five-day mean temperatures during winter. *J. Meteor.*, **16**, 672-682.

Murphy, A. H., and H. Daan, 1985: Forecast evaluation. *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, A. H. Murphy and R. W. Katz, eds., Westview Press, 379-437.

Roeger, C., R. Stull, D. McClung, J. Hacker, X. Deng, and H. Modzelewski, 2003: Verification of mesoscale numerical weather forecasts in mountainous terrain for application to avalanche prediction. *Wea. Forecasting,* **18**, 1140-1160.

Ruth, D. P., B. Glahn, V. Dagostaro, and K. Gilbert, 2009: The performance of MOS in the digital age. *Wea. Forecasting*, **24**, 504-519.

TPB, 1968: Experimental computer forecasts of maximum and minimum surface temperature. *Tech. Procedures Bul*. **16**, Weather Bureau, ESSA, U.S. Department of Commerce, 9 pp.

_____ 1978: Automated maximum/minimum and 3-hourly surface temperature guidance. *Tech. Procedures Bul.* **238**, Weather Bureau, ESSA, U.S. Department of Commerce, 14 pp.

_____ 1980: Automated maximum/minimum, 3-hourly surface temperature, and 3-hourly surface dew point guidance. *Tech. Procedures Bul*. **285**, Weather Bureau, ESSA, U.S. Department of Commerce, 16 pp.

Schwartz, B. E., and G. M. Carter, 1982: An evaluation of a modified speed enhancement technique for objective surface wind forecasting. *TDL Office Note* **82-1**, 10 pp.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*, Third Edition. Elsevier, Inc., 676 pp.

Yussouf, N., and D. J. Stensrud, 2007: Bias-corrected short-range ensemble forecasts of near-surface variables during the 2005/06 cool season. *Wea. Forecasting*, **22**, 1274-1286.

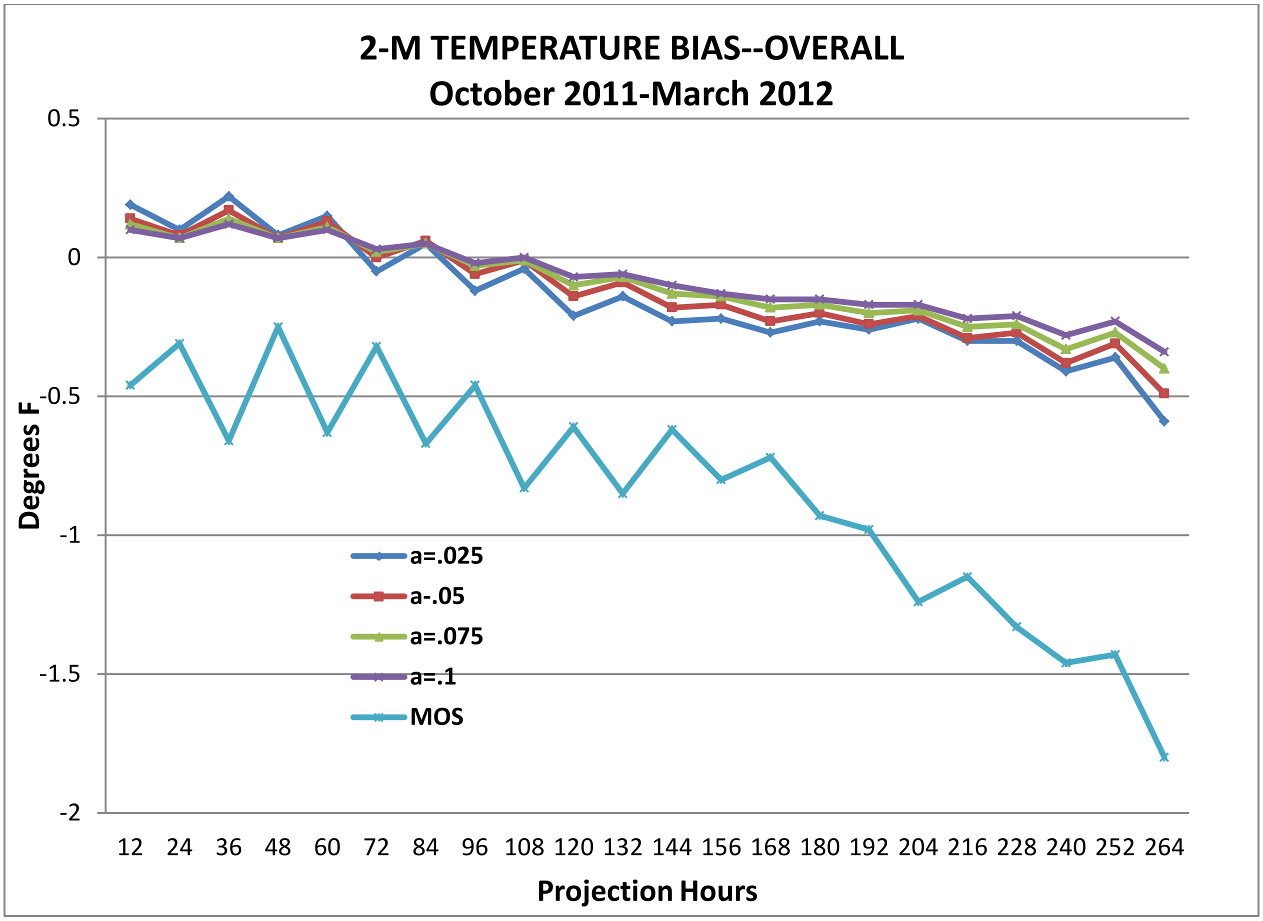


Figure 1. The bias over the 2011-2012 cool season for CONUS MOS temperature forecasts and the MOS bias corrected forecasts for the four alphas tested (alpha = a in the key).

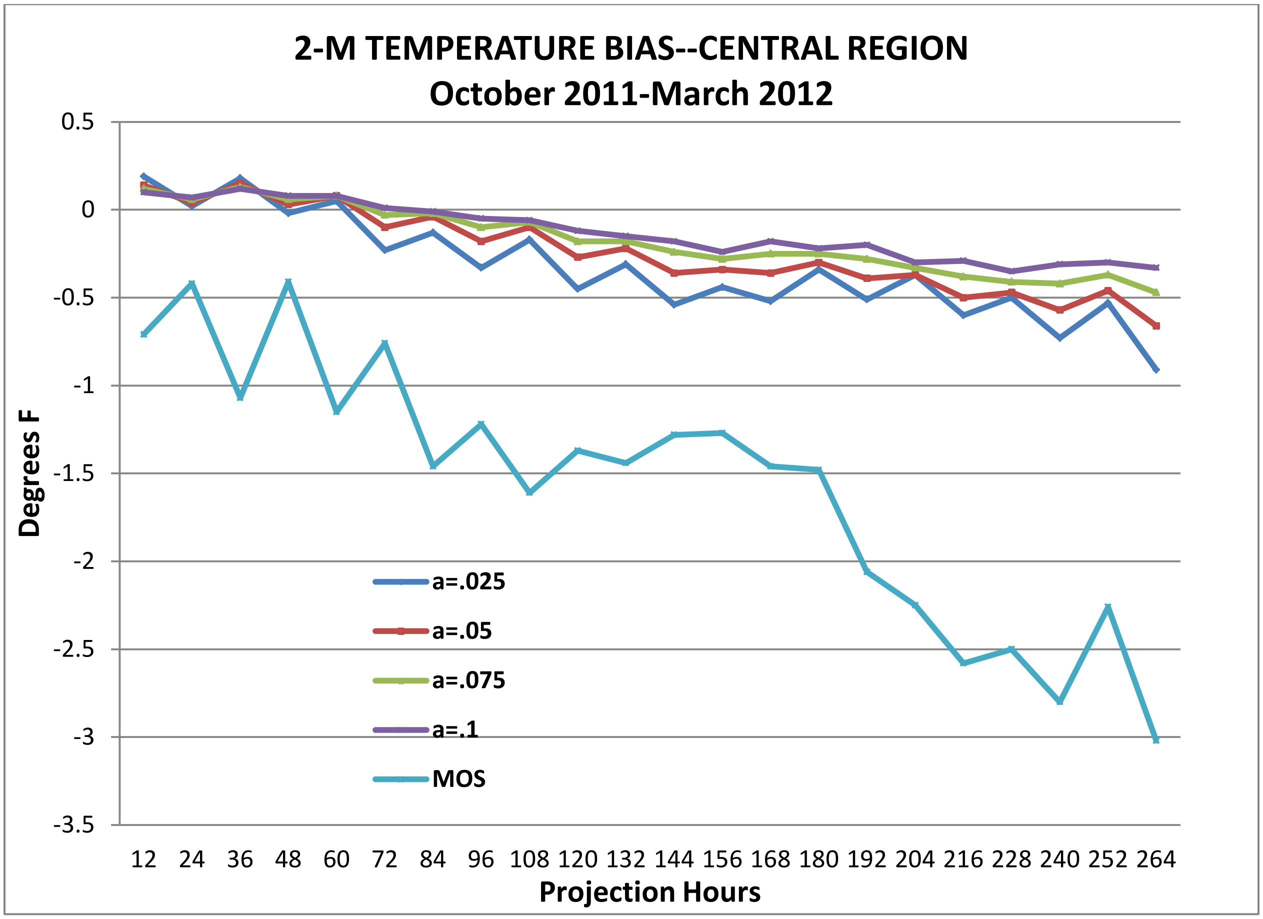


Figure 2. The same as Fig. 1, except for the NWS Central Region.

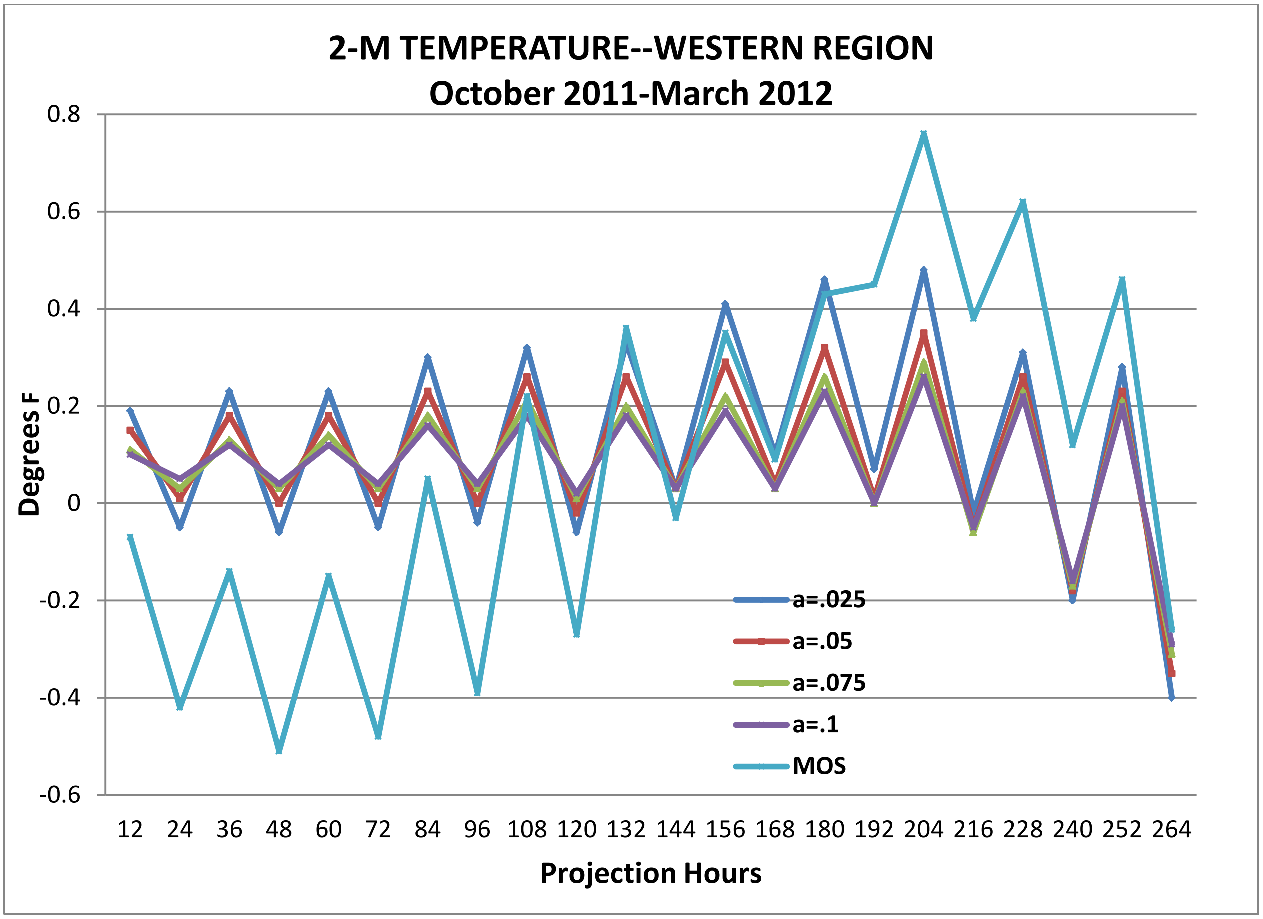


Figure 3. The same as Fig. 1, except for the NWS Western Region.

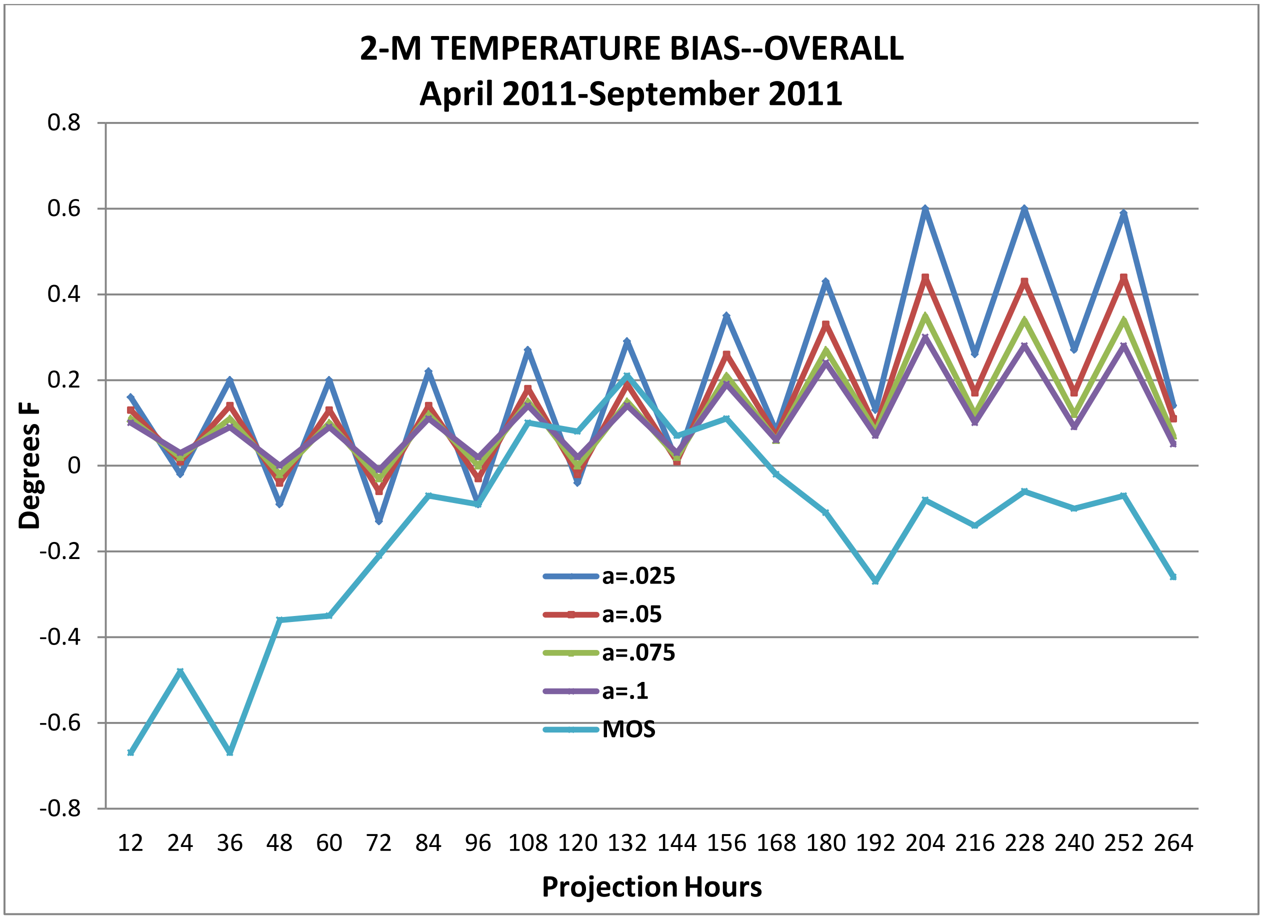


Figure 4. The same as Fig. 1, except for the 2011 warm season.

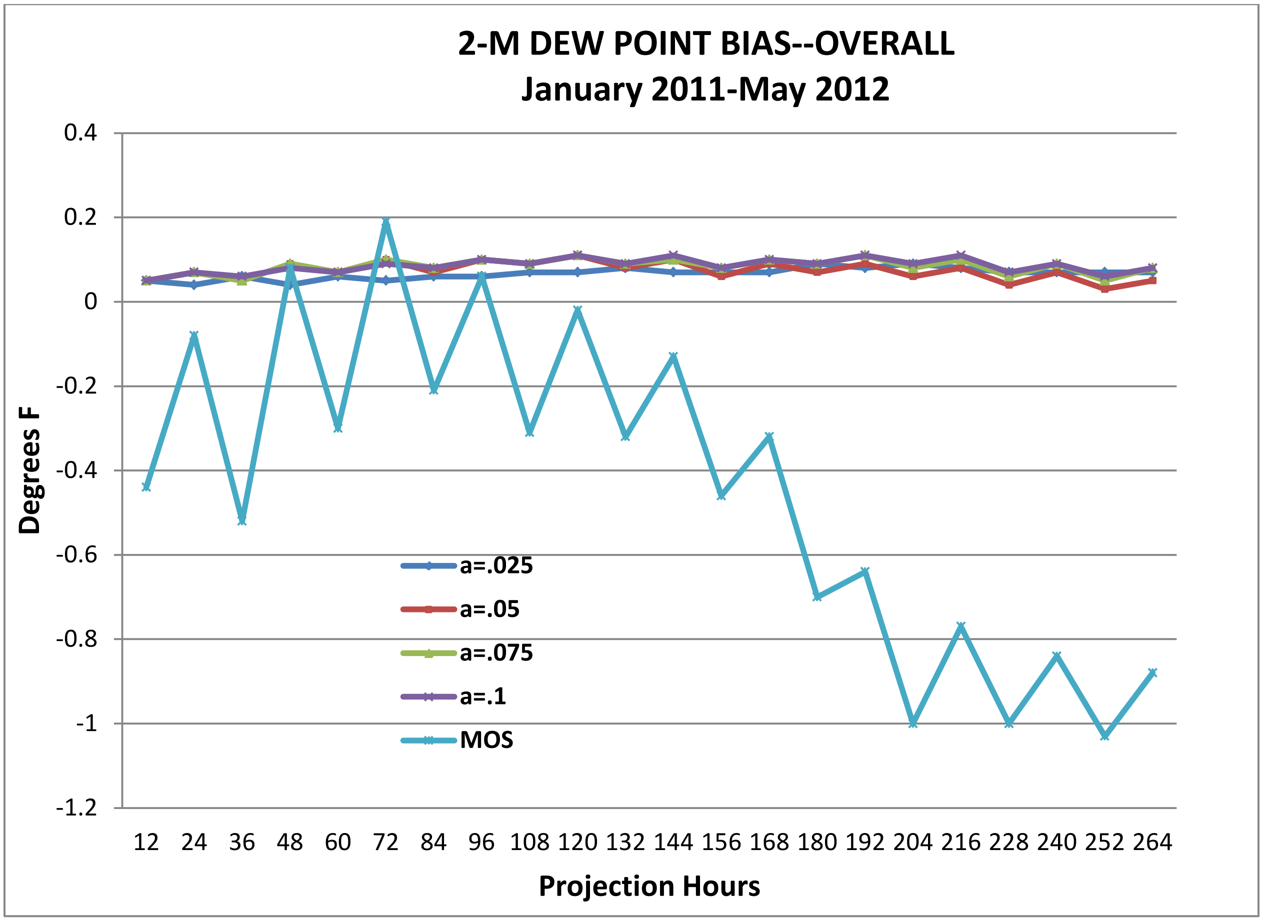


Figure 5. The same as Fig. 1, except for warm and cool seasons combined for dew point.

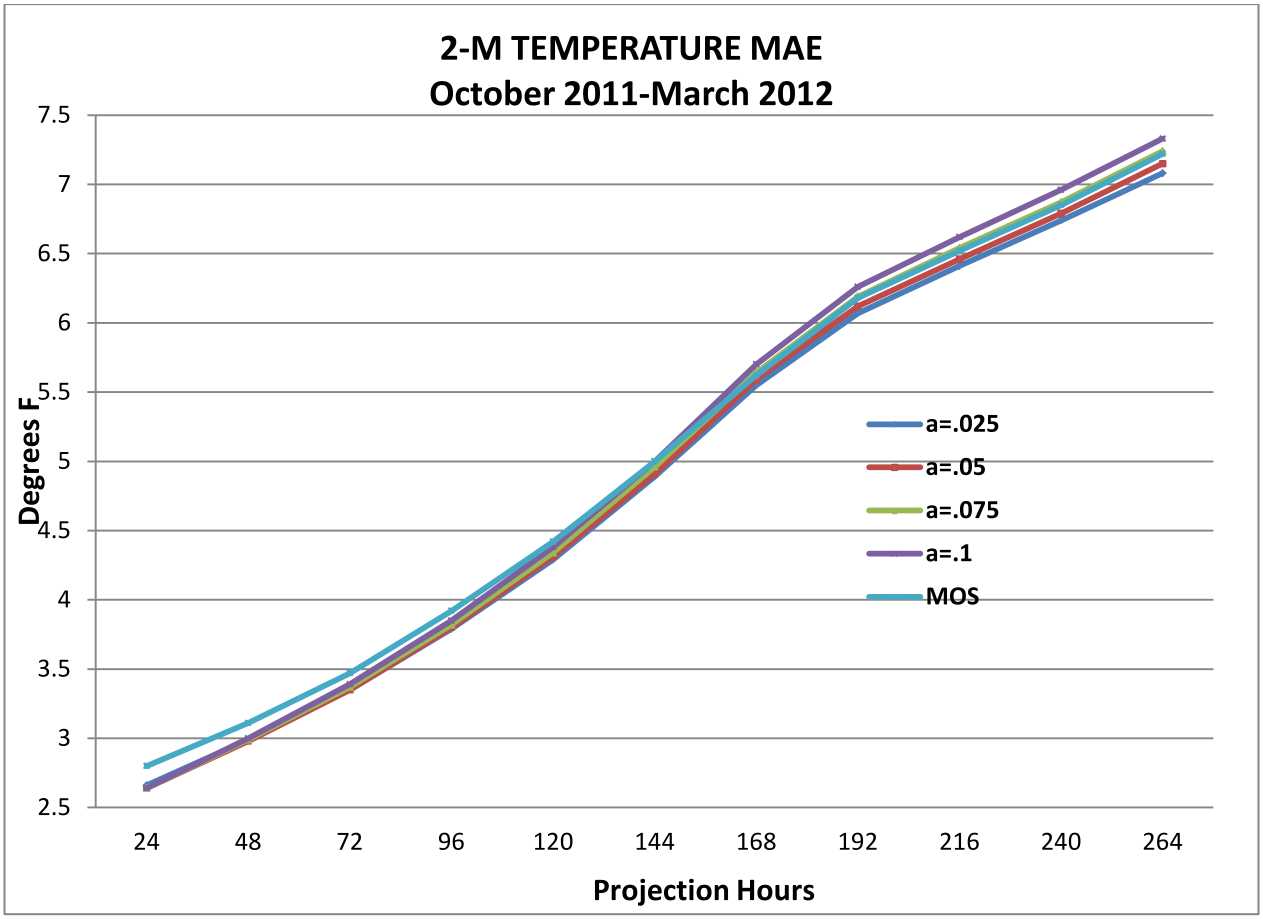


Figure 6. The MAEs over the 2011-2012 cool season for MOS temperature forecasts and the MOS bias corrected forecasts for the four alphas tested (alpha = a in the key). Only forecasts verifying at 0000 UTC are shown.

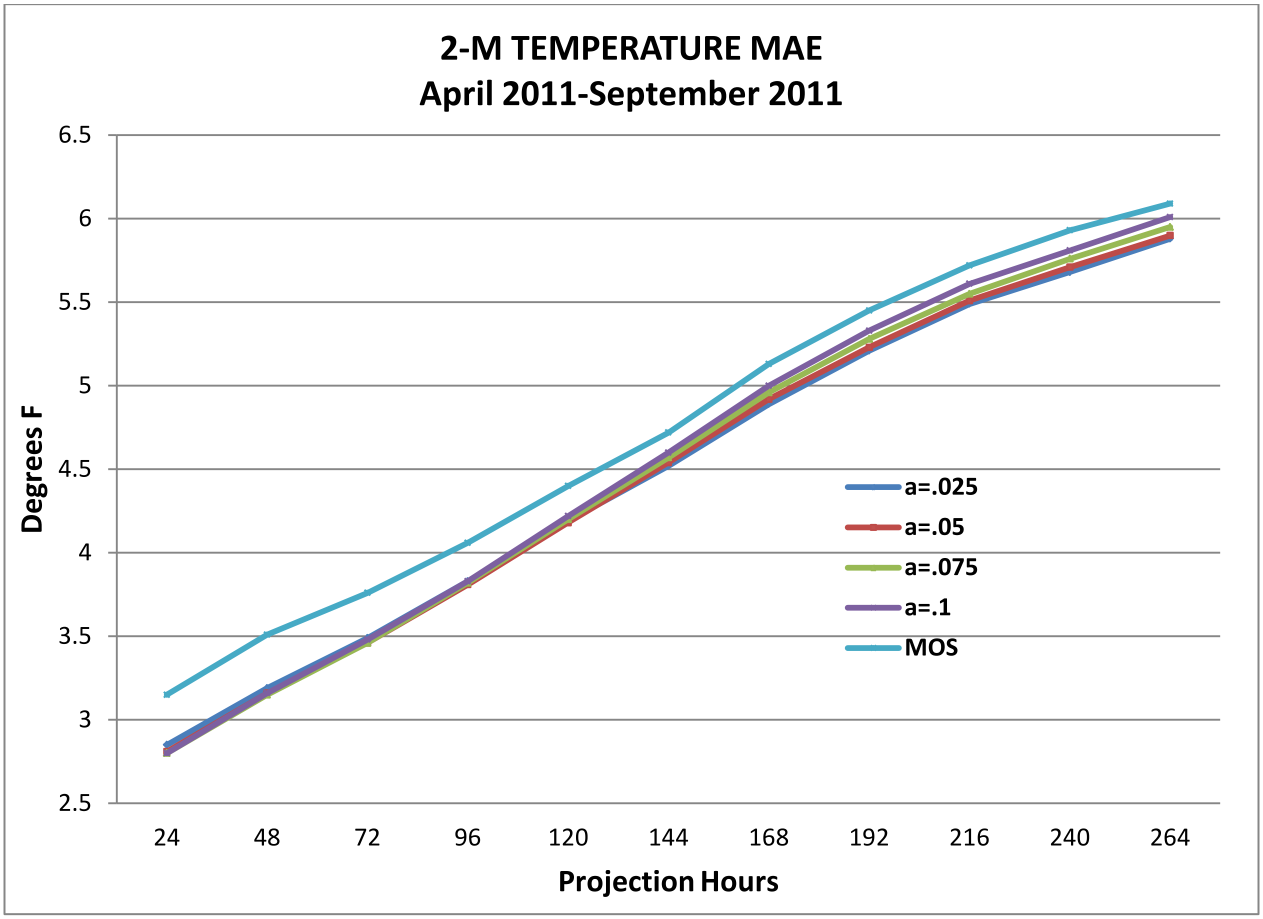


Figure 7. The same as Fig. 6, except for the 2011 warm season.

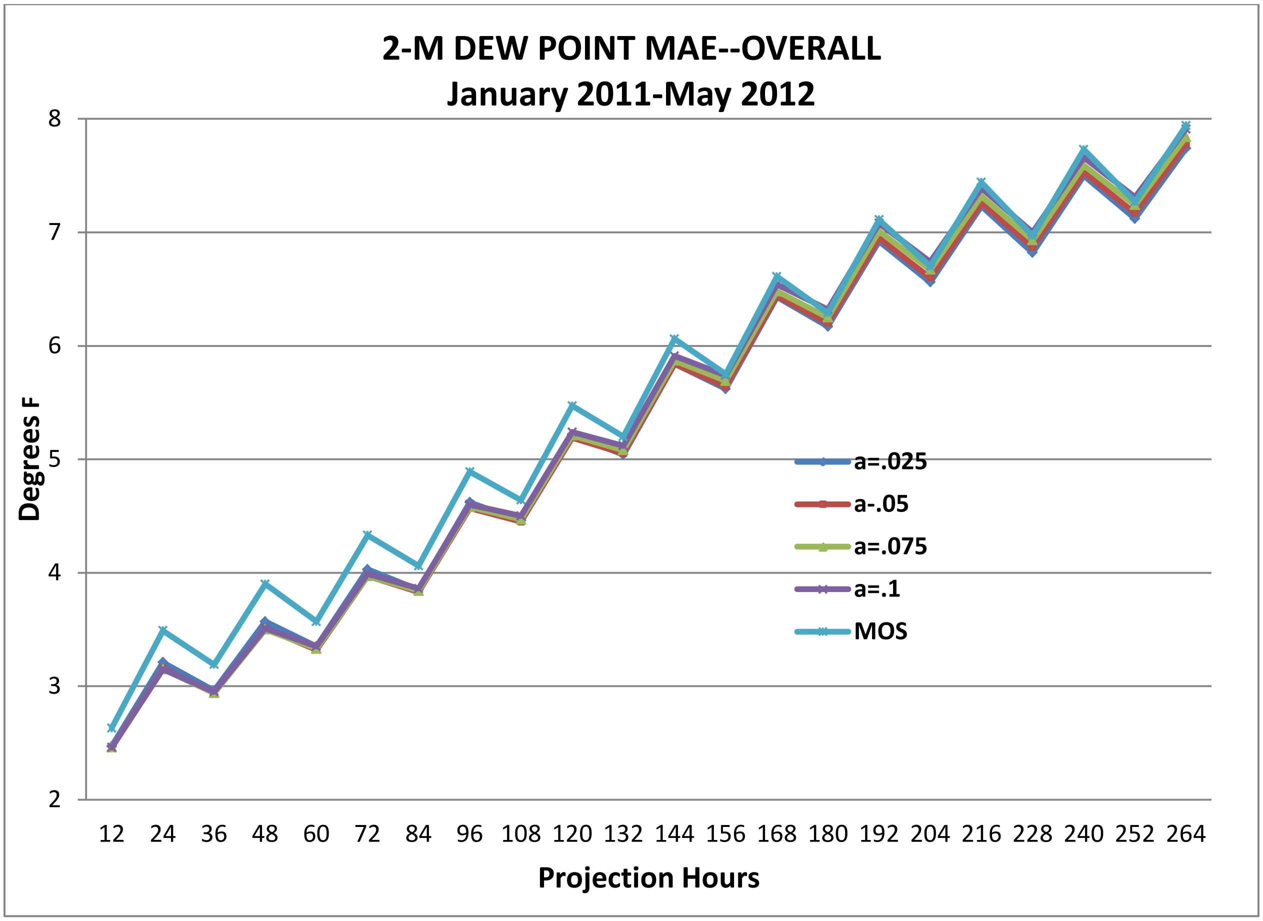


Figure 8. The MAEs over the warm and cool seasons for MOS dew point forecasts and the MOS bias corrected forecasts for the four alphas tested (alpha = a in the key).

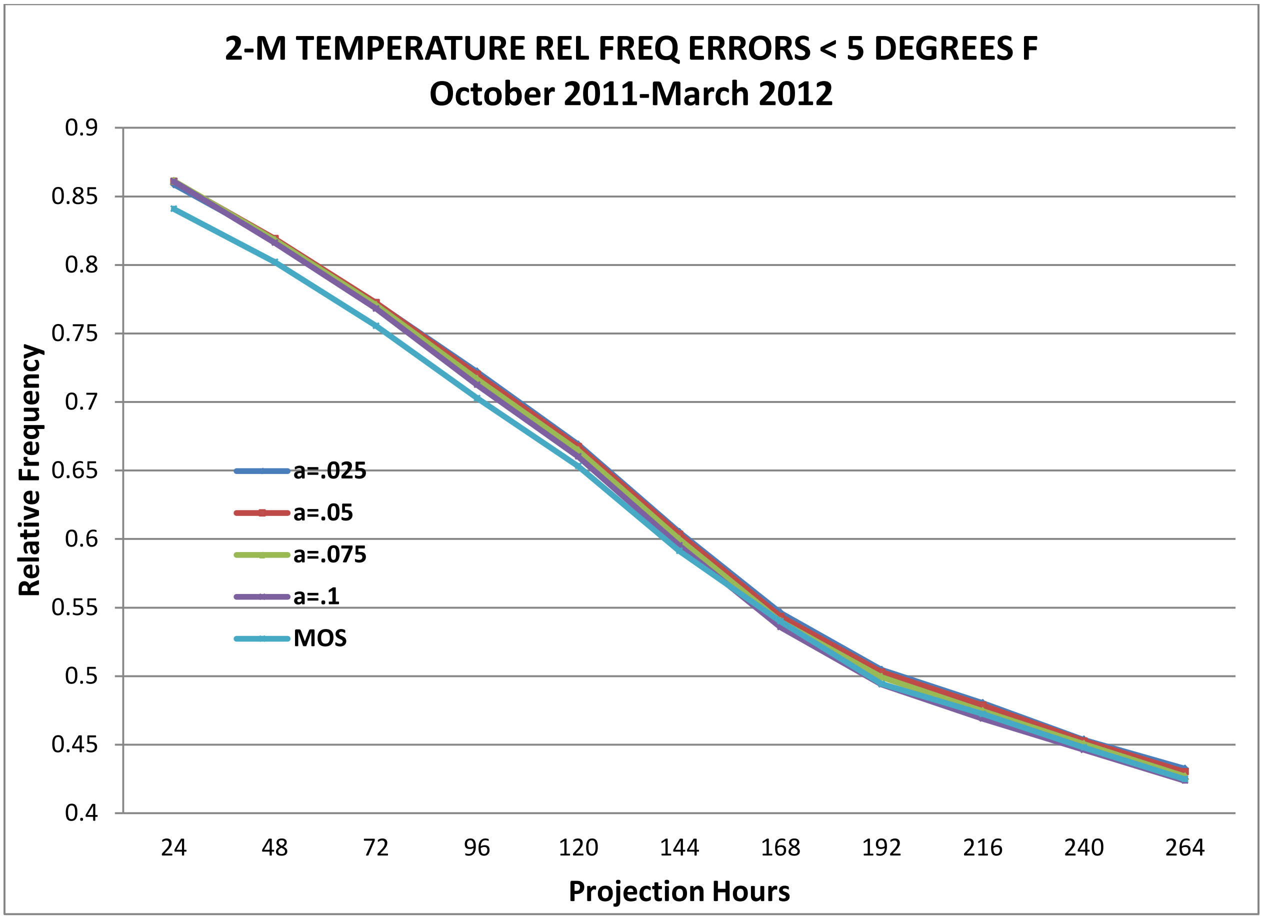


Figure 9. The percentage of temperature errors < 5 degrees F for the cool season.

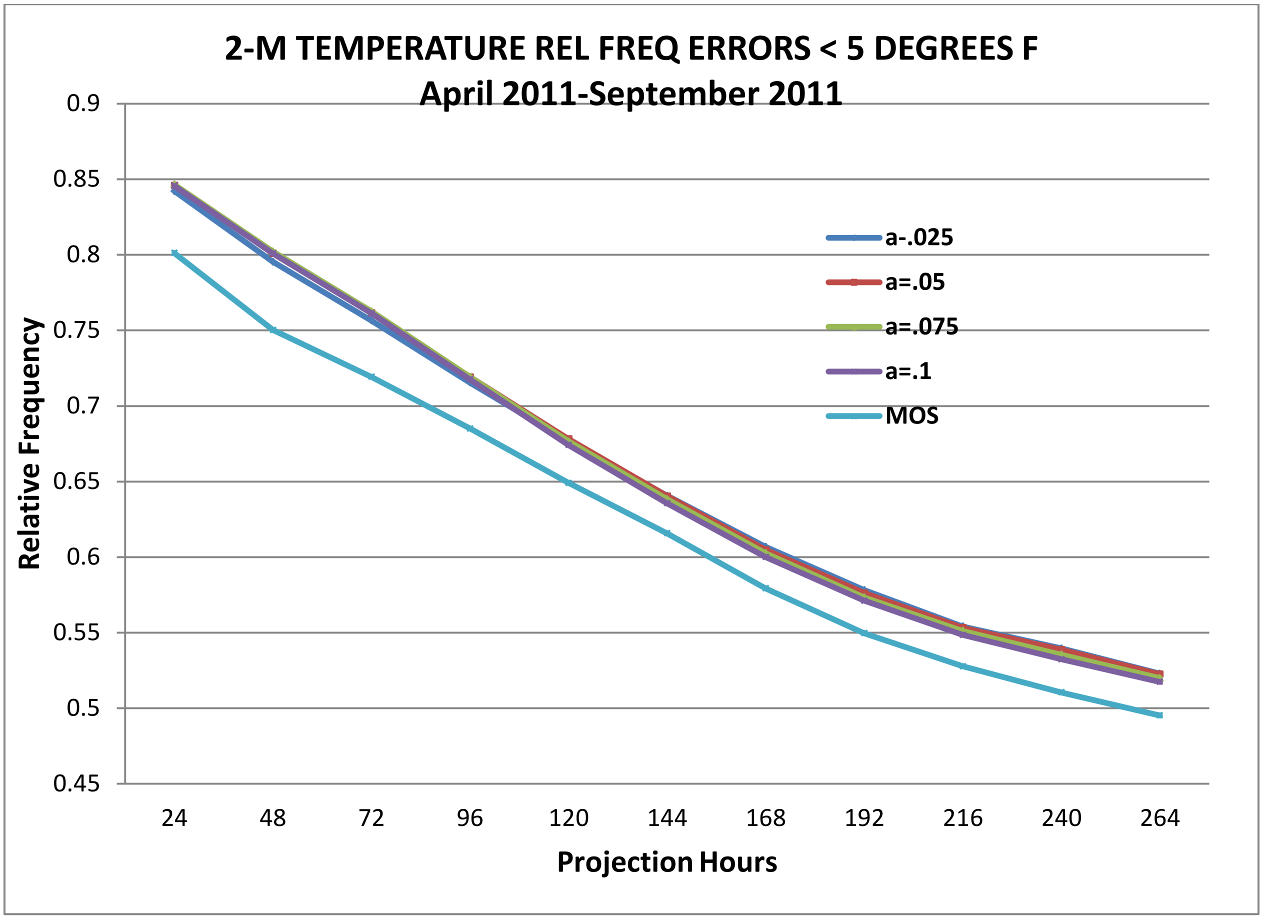
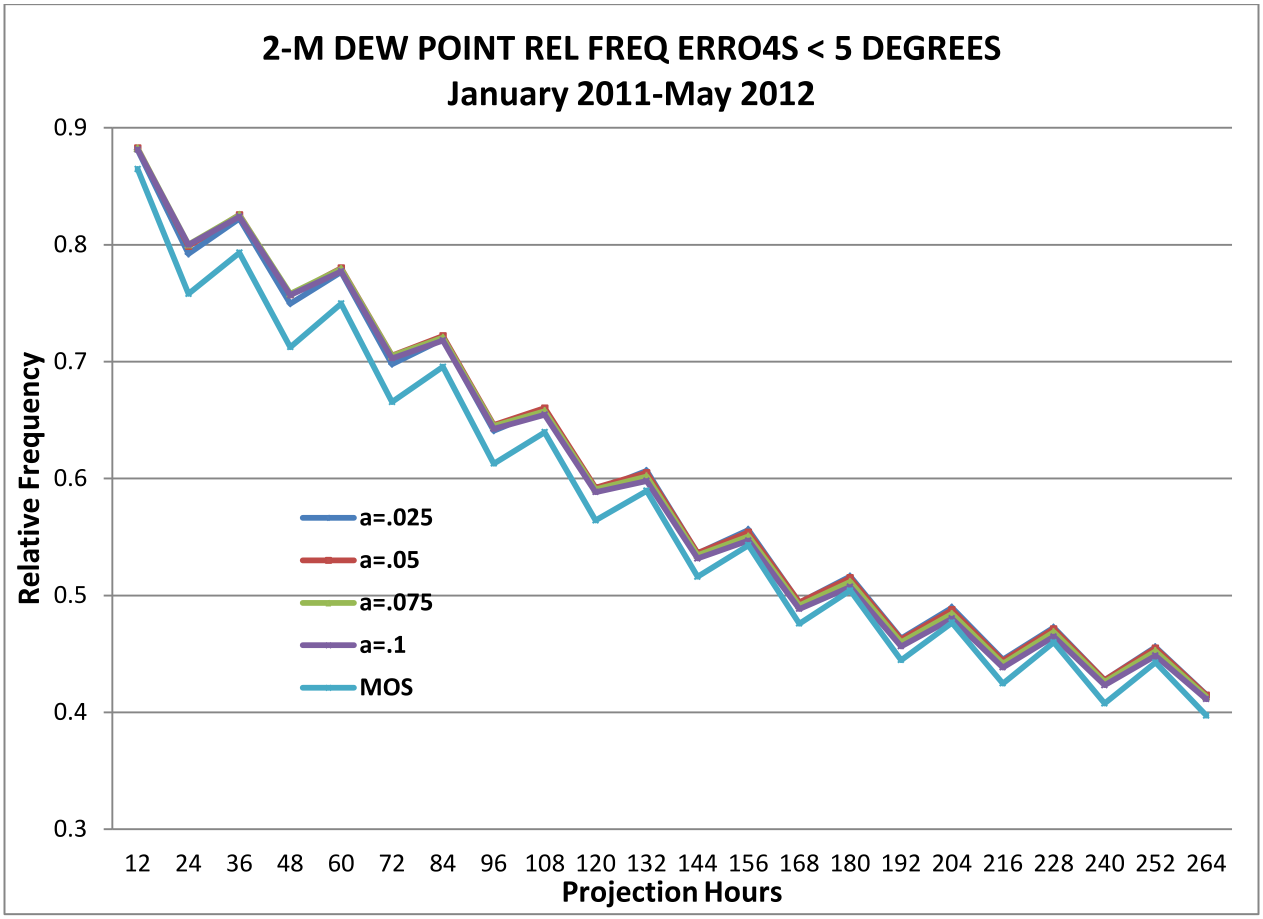


Figure 10. The same as Fig. 9, except for the 2012 warm season.



\Figure 11. The same as Fig. 9, except for the warm and cool seasons combined for dew point.

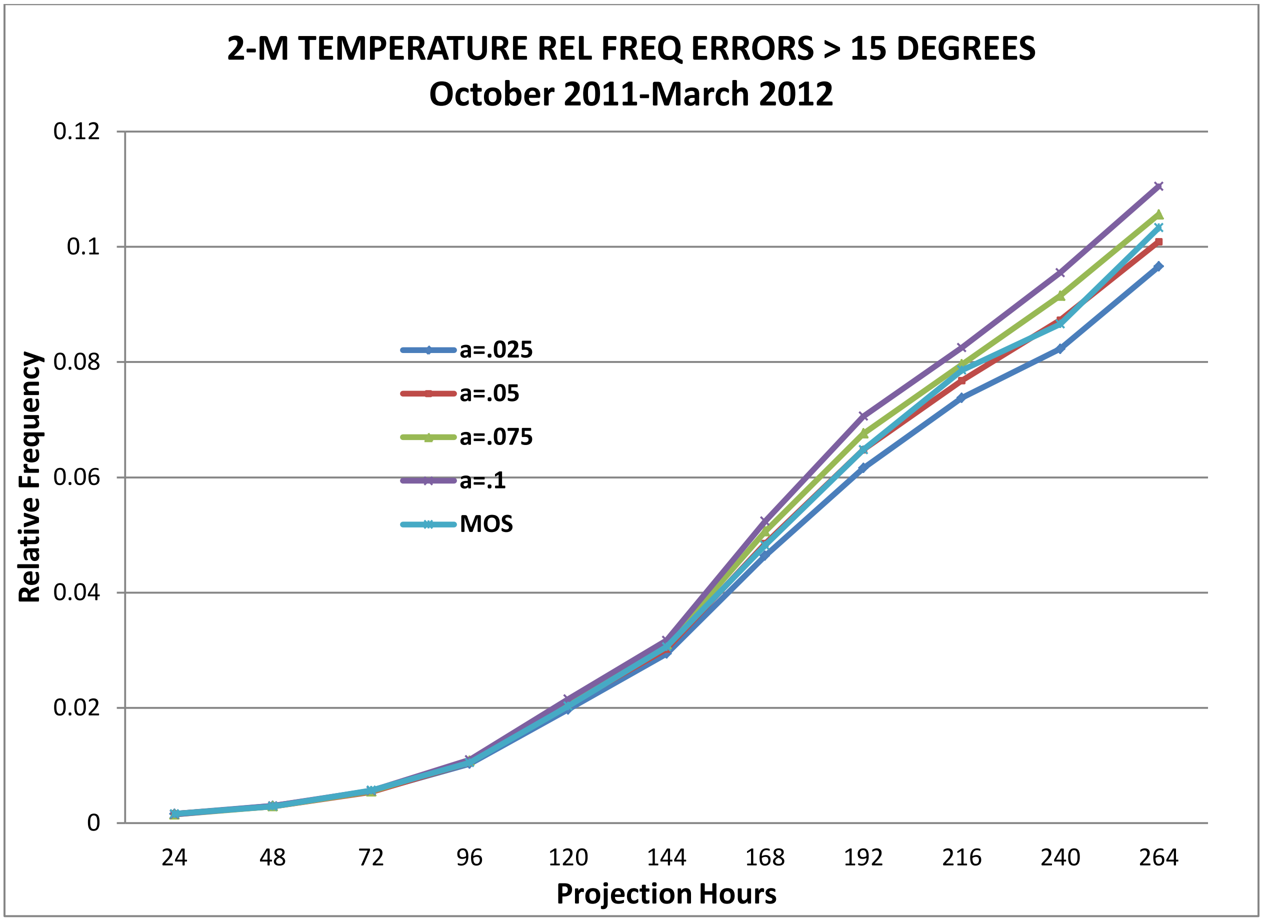


Figure 12. The percentage of temperature errors > 15 degrees F for the cool season.

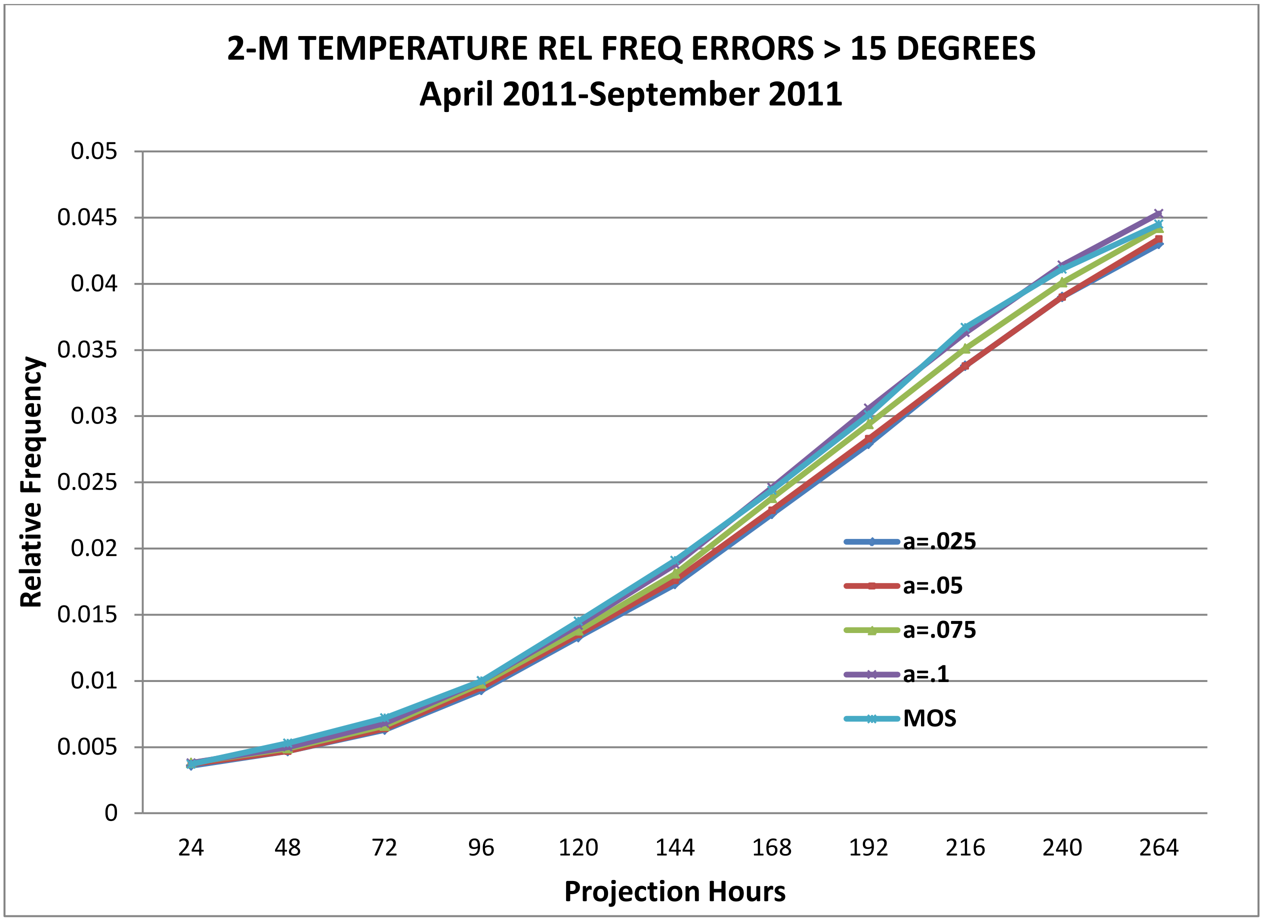


Figure 13. The same as Fig. 12, except for the 2012 warm season.

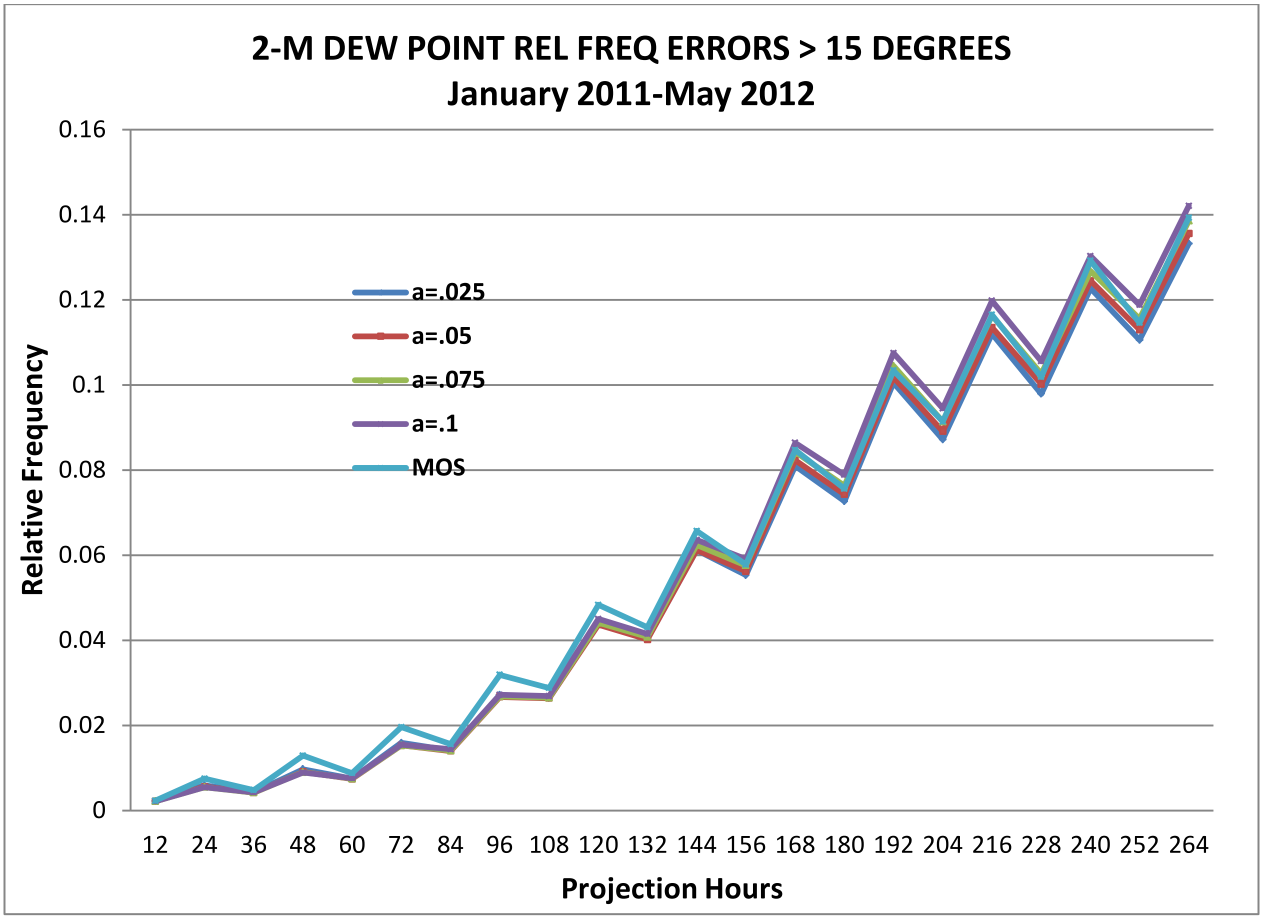


Figure 14. The same as Fig.12, except for the warm and cool seasons combined for dew point.

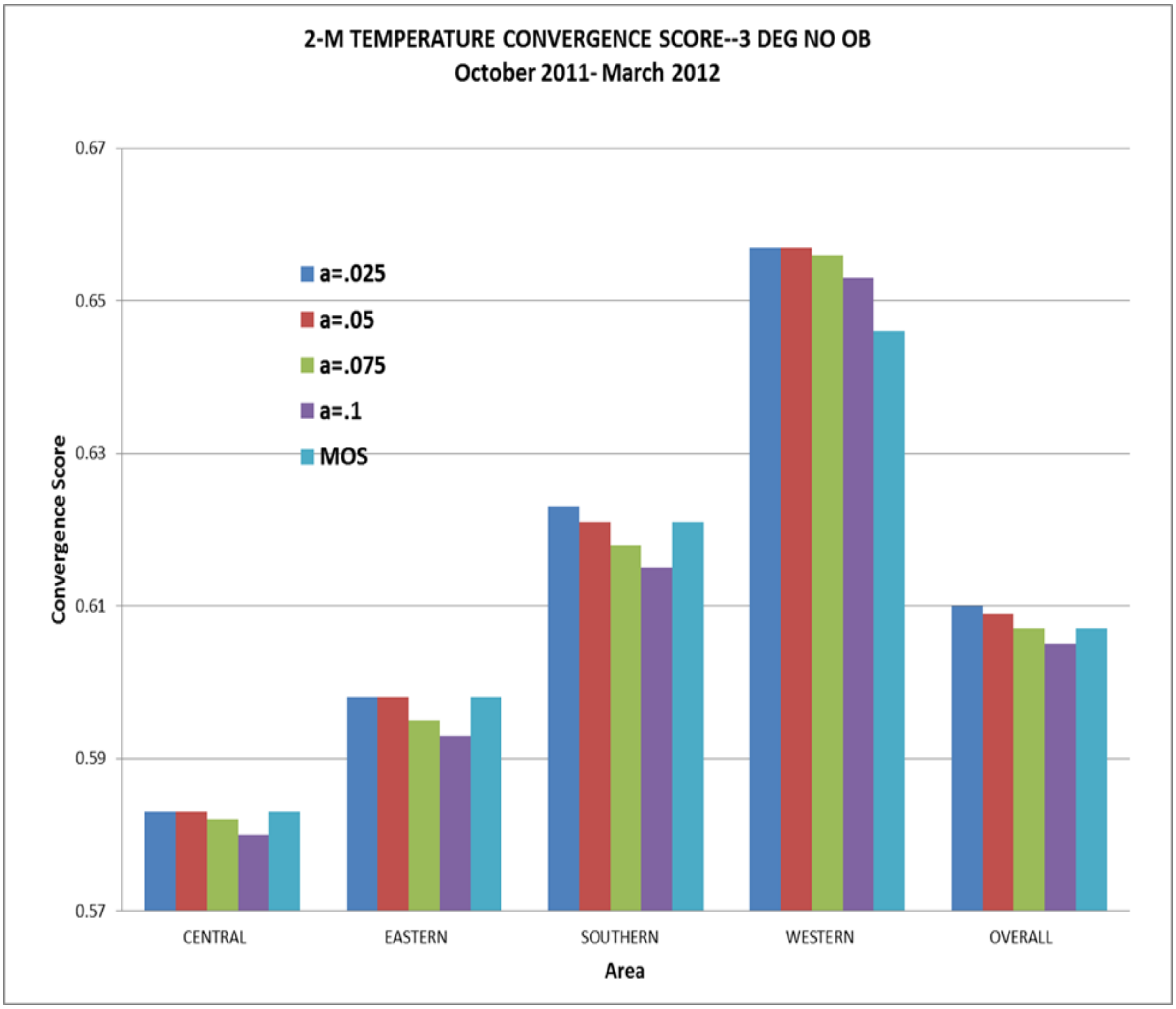


Figure 15. Convergence scores for the four NWS CONUS regions and overall for the cool season (alpha = a in the key). The “3 deg no ob’ refers to parameters for the score. Here, there is no penalty to the score if the change forecast is < 3 degrees F, and the observation to which the forecasts aspire is not used as a 0-h anchor point.

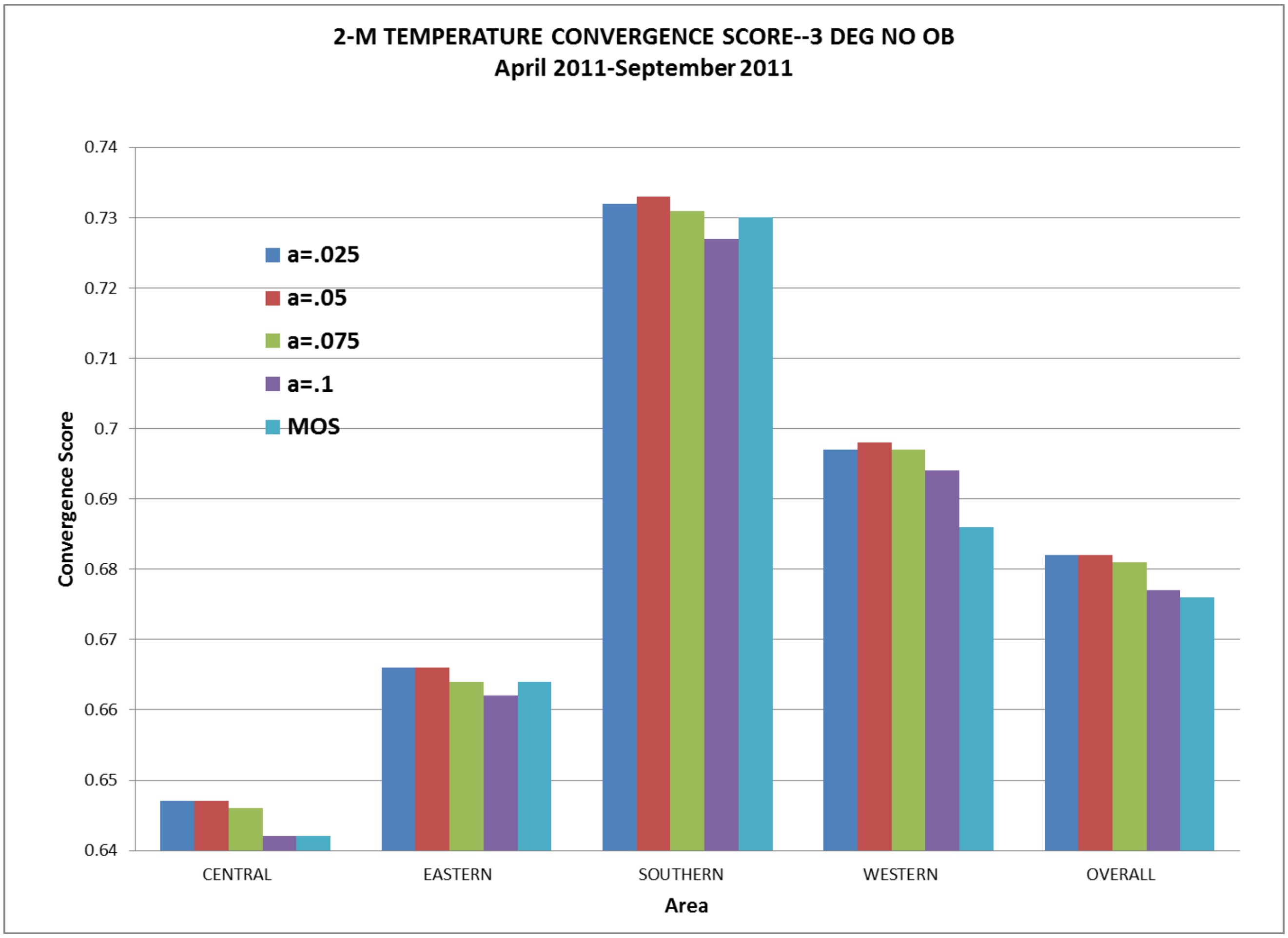


Figure 16. The same as Fig. 15, except for the warm season.

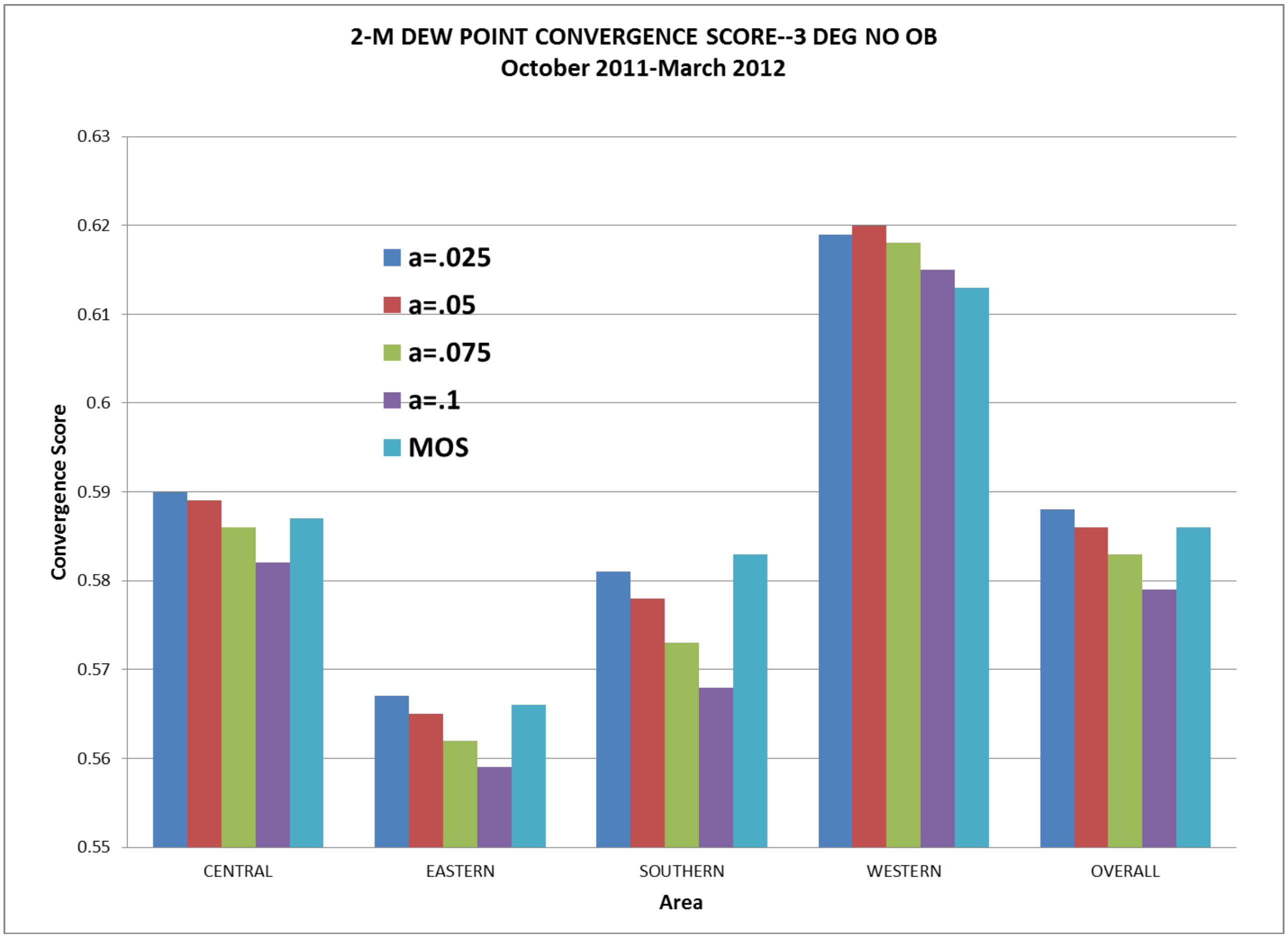


Figure 17. The same as Fig. 15, except for dew point for cool and warm seasons combined.

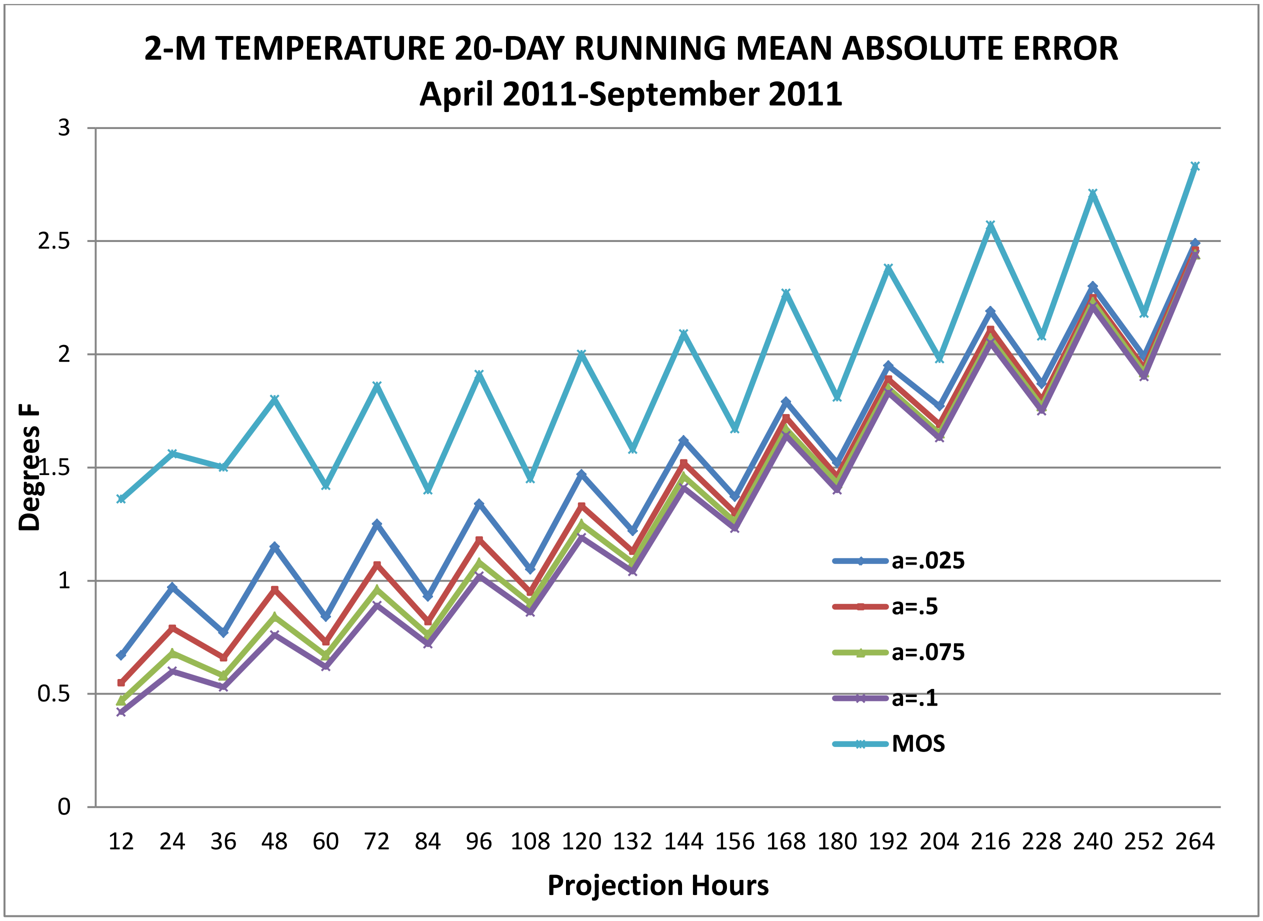


Figure 18. MAEs of forecasts over 20-day running periods for the warm season.

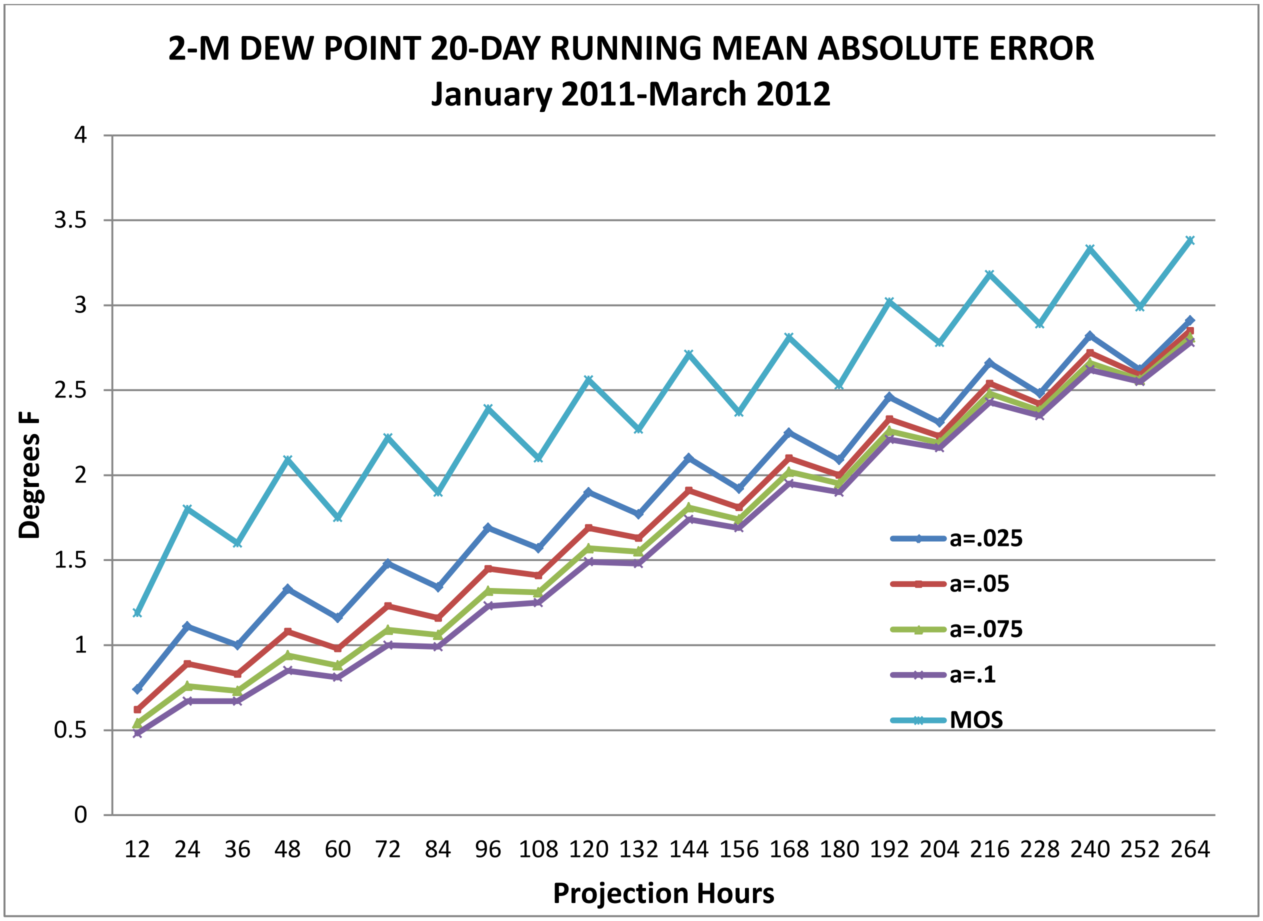


Figure 19. The same as Fig. 18, except for the cool and warm seasons combined for dew point.

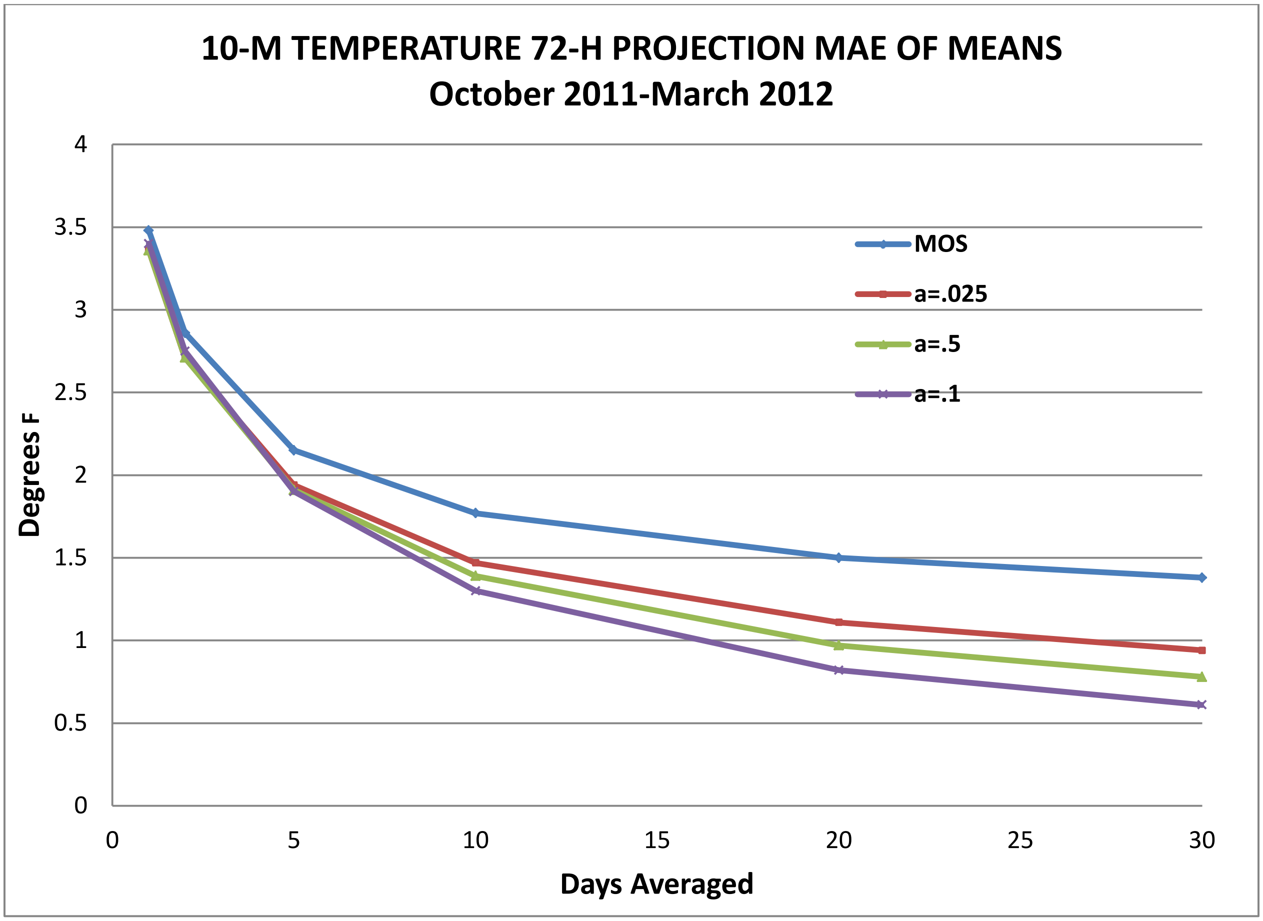


Figure 20. MAEs of 72-h forecasts as a function of averaging period (a = .075 is not shown).

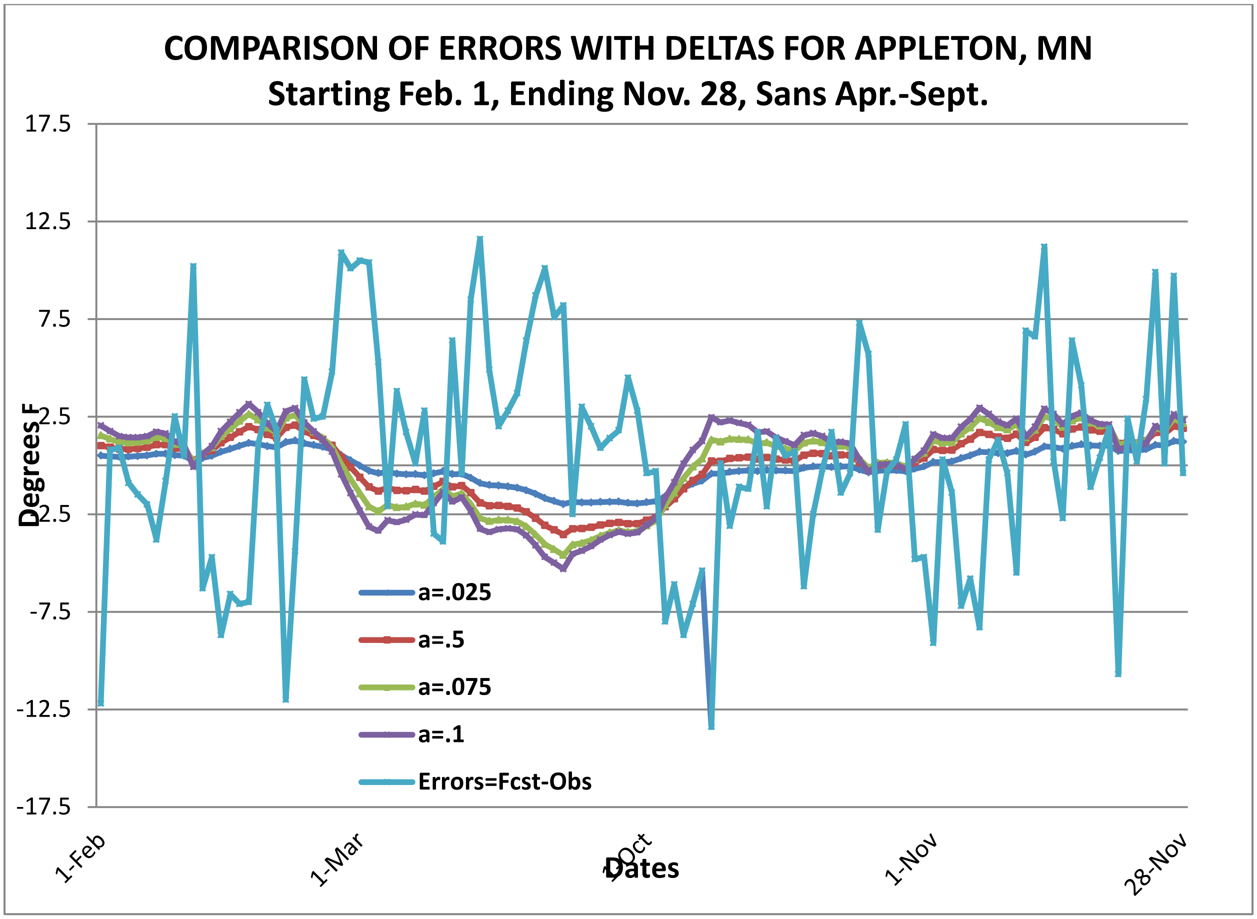


Figure 21. Errors in MOS temperature forecasts and the deltas associated with the four alphas tested (alpha = a in the key).