U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
OFFICE OF SCIENCE AND TECHNOLOGY
METEOROLOGICAL DEVELOPMENT LABORATORY

MDL OFFICE NOTE 06-01

PERFORMANCE OF THREE FORECAST SYSTEMS AT SELECTED MOS AND NON-MOS STATIONS

Michael Schenk

July 2006

PERFORMANCE OF THREE FORECAST SYSTEMS AT SELECTED MOS AND NON-MOS STATIONS

Michael Schenk

1. INTRODUCTION

This study examined the relative performance of the National Digital Forecast Database (NDFD; Glahn and Ruth, 2003), Gridded Model Output Statistics (GMOS; Dallavalle and Glahn, 2005), and Hydrometeorological Prediction Center (HPC; Glahn and Ruth, 2003) forecast systems for three weather elements at two sets of verification sites. The elements considered were daytime maximum temperature, nighttime minimum temperature, and dewpoint. The Interactive Forecast Preparation System (IFPS) Science Steering Team (ISST) has stated that the “HPC grids . . . should be thoroughly tested and evaluated to examine their impacts on . . . forecast accuracy” (ISST 2004). The ISST has also stated that verification should include a “. . . set of point observations . . . [at] lower order sites such as available in the RAWS network” (ISST 2004). In response, MDL undertook a study to examine the medium range (days 4-7) performance of the NDFD, GMOS, and HPC systems, both at traditional METAR verification sites (MOS sites) and at "lower-order" sites (e.g., RAWS network) which were not included in GMOS development (non-MOS sites), as illustrated in Fig. 1. The primary purpose of the study was to test the performance of the grids at these non-MOS sites in order to estimate the accuracy at any point on the grid. Also, the correlation between distance from verification site to NDFD grid point and accuracy was examined. Mean Absolute Error (MAE) was selected as the measure of accuracy.

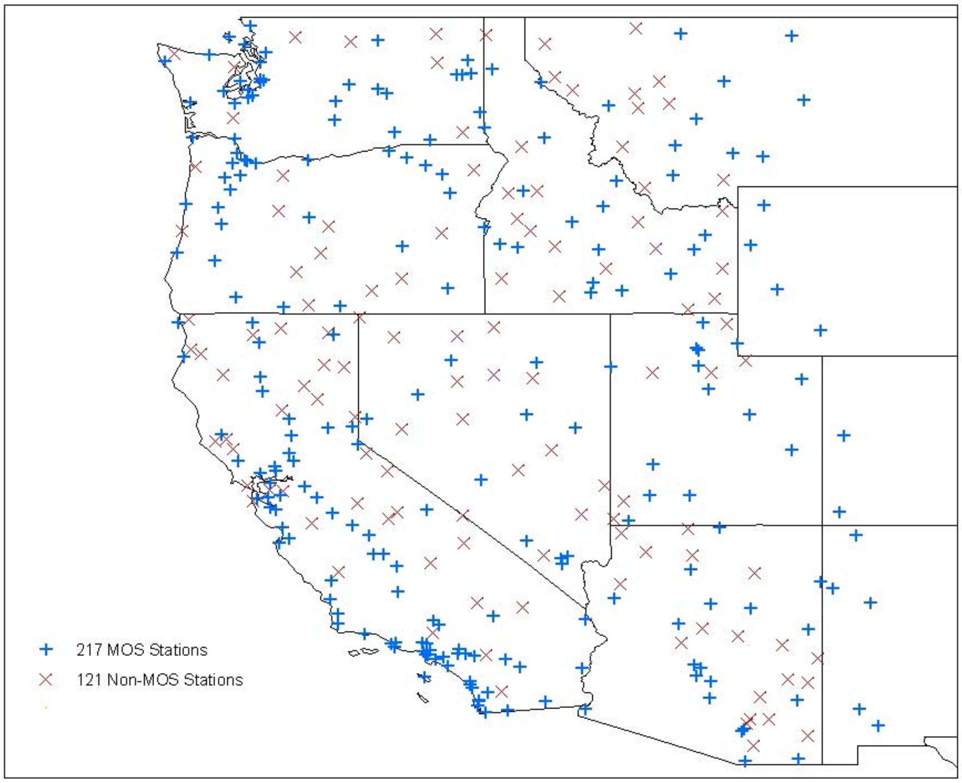

Figure 1. Study region and location of verification sites.

2. METHODS

A. Scope of Data

A special retrospective rerun of GMOS over the western third of the conterminous United States (CONUS) was performed. This was necessary to exclude from the gridding of the forecasts the non-MOS sites. It is important to note, when comparing forecasts, that these GMOS forecasts were not available to users in real time.

Availability of forecasts and of observation data of sufficient quality limited the scope of the spatial domain, the forecast projections, and the choice of weather elements. The Western CONUS was chosen because it was the area for which GMOS was available. This choice of region also allowed testing in complex terrain. NDFD and GMOS forecasts corresponding to HPC's days 4 to 7 forecasts were selected. Weather elements tested were dew point, daytime maximum temperature, and nighttime minimum temperature, chosen for common availability of forecasts and observation data. Because daytime max and nighttime min are not observed, they were inferred from hourly temprature data.

The observation data used in the study consisted of two weather elements, dew point and hourly temperature. Data for the non-MOS sites were obtained from a set of stations recommended by Western Region Weather Forecast Offices (WFO). Observation data for the MOS sites were obtained from METAR.

Forecasts were matched to observations spatially by using the forecast at the nearest-neighbor gridpoint, and temporally. The sequence of issuance of the five forecasts studied were: 00Z GMOS issued at 05 UTC (labeled 05ZGMOS); 12Z GMOS issued at 17 UTC (labeled 17ZGMOS); preliminary HPC issued at 15 UTC (labeled 15ZHPC); 18Z NDFD (labeled 18ZNDFD); and the 00Z NDFD issued the day following the 00Z GMOS (labeled 00ZNDFD). Thus, 17ZGMOS, 15ZHPC, and 18ZNDFD are available to users at roughly the same time.

We obtained observation data from the Global Systems Division (formerly Forecast Systems Laboratory) Meteorological Assimilation Data Ingest System (MADIS) website (http://www-sdd.fsl.noaa.gov/MADIS/). The metadata for the stations submitted by Western Region WFO's were compared to external sources for consistency and accuracy. A list of 167 stations was identified for potential use in the study.

We further reduced the list to a set of 121 stations. Many stations were removed because they did not report dew point or temperature data for the study period, or had very sparse data. Stations with low data frequency were eliminated if: a) the station reported less than 25% of the possible data for the study period, or b) the station did not show improvement with time in the percentage of possible observations. Two stations were removed from the study because of characteristics of the forecast grids: there was a slight difference in the lower left corner of the NDFD grid versus the HPC and GMOS grids, which resulted in different nearest-neighbor grid points being chosen for those stations. Some stations were removed because they fell outside the study area bounded by the GMOS grid. Finally, more stations were eliminated because all four surrounding gridpoints were either 500 feet lower or higher than the site or the gridpoint was located over water.

B. Data Conversion and Computation

Forecast and observation data were converted to TDLPACK (MDL's internal format), as needed for processing. We were careful to ensure that the data sets were comparable regarding unit conversion and rounding.

Hourly dew point and temperature observations were precise to whole degrees Fahrenheit (F); therefore, no rounding of data or unit conversion was necessary. Because daytime maximum temperature and nighttime minimum temperature are not reported at METAR and mesonet sites, we inferred those values using hourly temperature values. We used the highest hourly temperature during daylight hours (7:00 AM – 7:00 PM LST) and the lowest hourly temperature for the hours of 7:00 PM to 8:00 AM LST as the daily maxima and minima of temperature. These inferred values potentially underestimated the extremes of the actual daytime maxima and nighttime minima.

MOS2000 software was used to obtain the nearest-neighbor gridpoint forecasts from each system, to round the data to the nearest whole degree (consistent with observation precision), match the forecasts, and to generate MAE and bias scores for each forecast projection, both for the overall set of stations and for individual stations. The scores were computed for each month and for the overall 10-month study period, and a weighted mean of all projections combined was computed.

To examine the correlation between accuracy and distance to NDFD grid point, we regressed MAE on distance from non-MOS site to nearest NDFD grid point. In a few cases, the distance was greater than 5 kilometers because closer gridpoints were not used due to difference in elevation or location over water.

3. RESULTS

Fig. 2 shows the MAE by element at both MOS and non-MOS observation sites. The MAE of all forecasts is approximately 1 to 1.5 degrees F lower at MOS sites than at non-MOS sites in all five issuances, for all three weather elements. The GMOS MAE is generally lower than NDFD and HPC at both MOS and non-MOS sites. An exception is with minimum temperature for non-MOS stations, where HPC is as good as or better than GMOS for all projections combined. For non-MOS stations, there is generally less than one degree Fahrenheit spread in the MAE's among the five forecast issuances.

Fig. 3 illustrates the biases. The means of the bias for all projections (shown at the far right of the plot) are generally greater in magnitude at non-MOS stations than at MOS stations. The mean bias for all projections tended to be positive for maximum temperature and negative for minimum temperature, possibly due to the method of estimating daytime maxima and nighttime minima from hourly temperatures. In terms of mean bias, GMOS is closer to zero than NDFD for MOS stations, but NDFD is closer to zero for non-MOS sites.

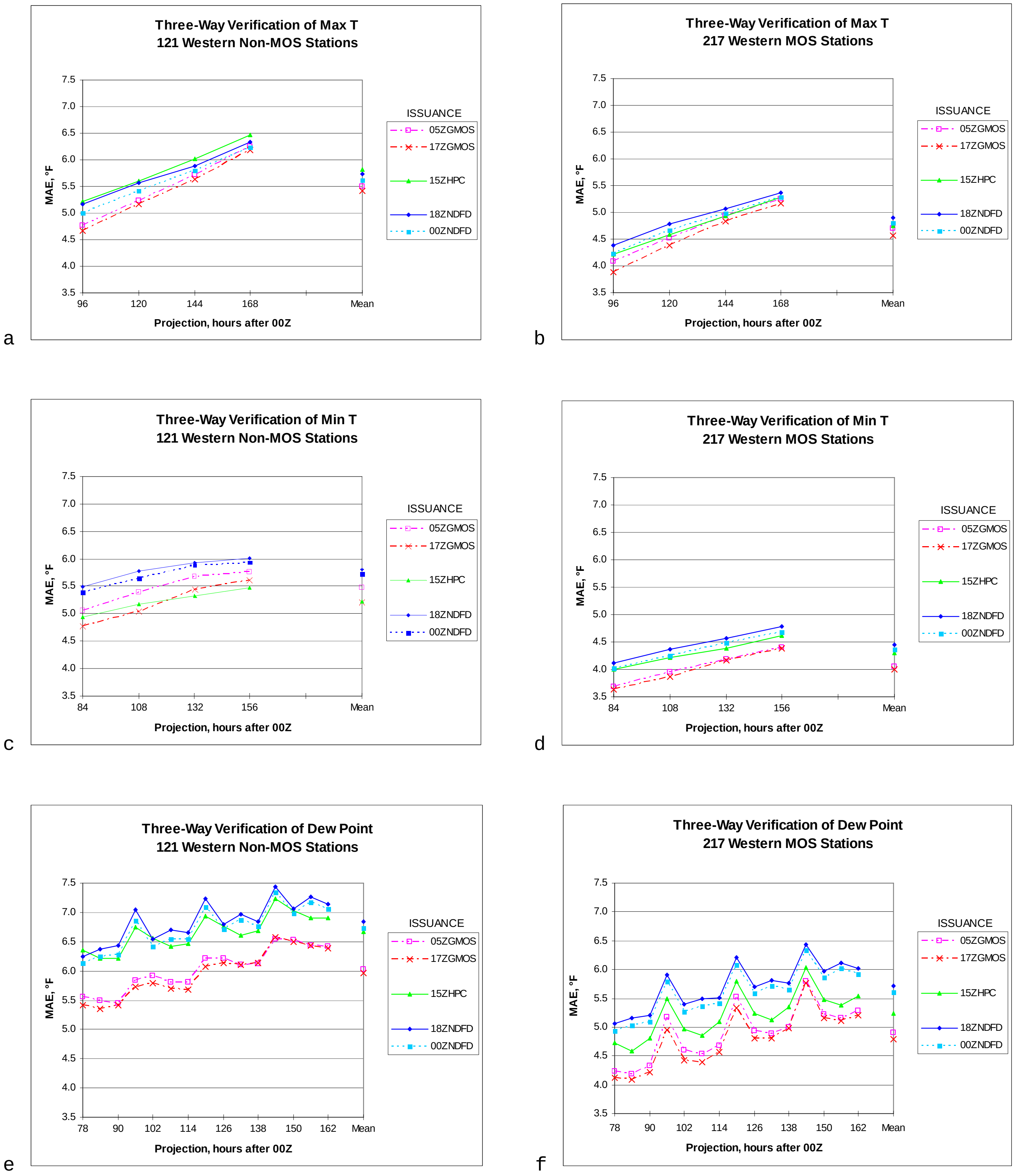

Figure 2. MAE of maximum temperature (a,b), minimum temperature (c,d), and dew point (e,f), at non-MOS and MOS observation sites. The mean of all projections is shown at far right.

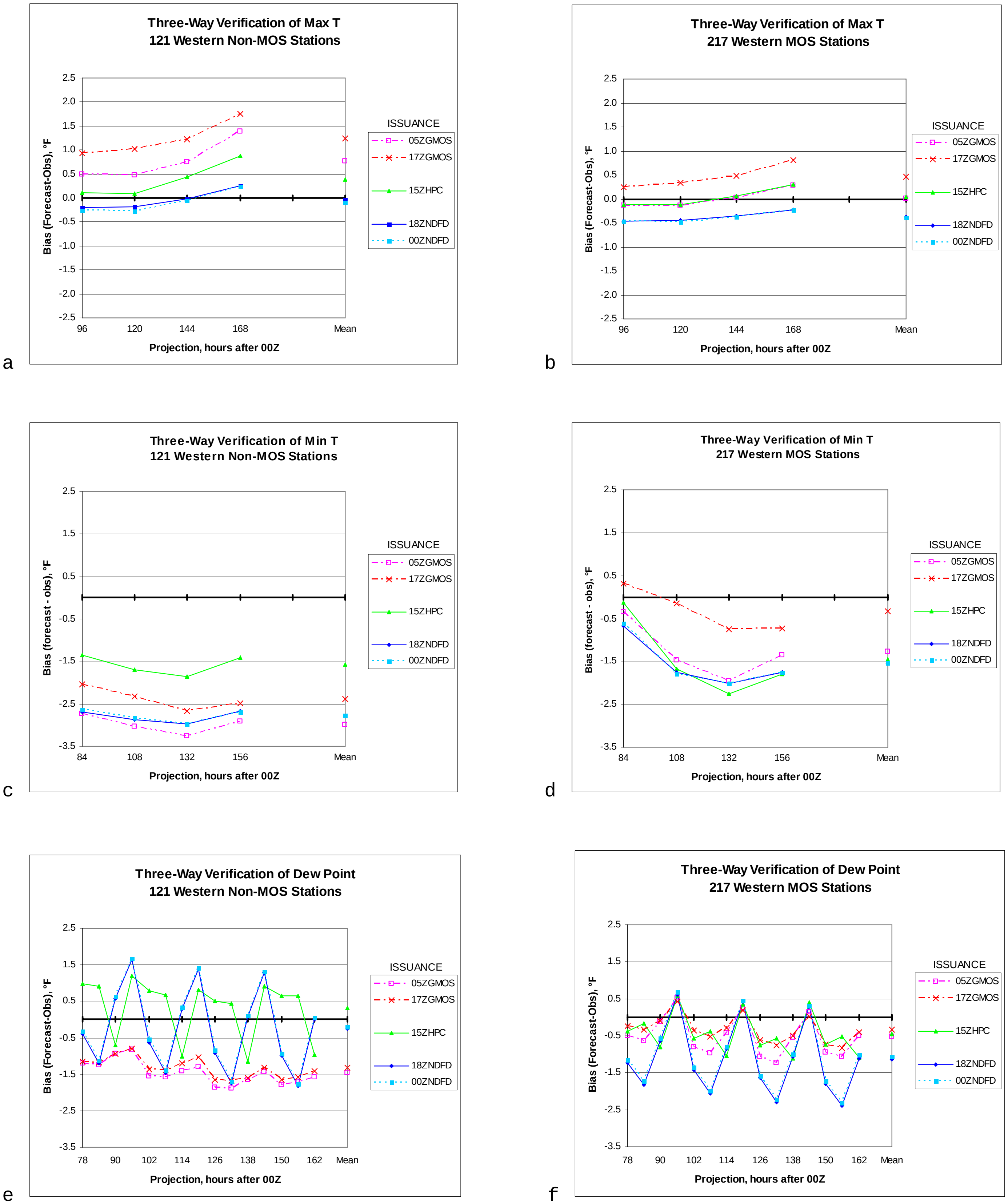

Figure 3. Bias of maximum temperature (a,b), minimum temperature (c,d), and dew point (e,f), at non-MOS and MOS observation sites. The mean of all projections is shown at far right.

Although there is an improvement for all forecasts at MOS sites with respect to non-MOS sites, the rankings of the forecasts do not change, and the difference between best and worst MAE is similar in magnitude between MOS sites and non-MOS sites.

There may also be a seasonal signal in the MAE. The first few months of the study period show less agreement of trend among the five issuances, and a wider dispersal of scores. Fig. 4 illustrates the MAE of maximum temperature, minimum temperature, and dew point, plotted by month and averaged for all projections.

There was a very low correlation between distance to NDFD grid point and MAE for non-MOS sites, as shown in Fig. 5.

4. DISCUSSION

MAE was greater at non-MOS sites than at MOS sites for all forecast systems. This difference in accuracy could be due to the characteristics of the observational data as well as to the difficulty of generating forecasts away from MOS sites. Although we found no quantitative evidence that the loss of accuracy is due to the non-MOS observation data, there are clear differences in the nature of the observation stations. Many non-MOS sites are remotely-located fire-weather stations, for example, while many MOS sites are situated at accessible locations such as airports. There is a possibility of more frequent maintenance and calibration at airport sites than remote sites. In addition to potential differences of accuracy due to observation data, the forecasts may be more accurate at MOS sites than at non-MOS sites due to local forecast procedures and the inherent nature of MOS, which makes the forecast specific for points used in its development. In general, the biases are better at MOS sites than at non-MOS sites.

There may be a seasonal effect on the magnitude of the MAE scores and the ranking of forecast issuances by accuracy. There is a much wider spread of scores in December and January than there is in May and June. For dewpoint, the MAE for NDFD is less than for HPC in the months June through September; however the opposite is true in the months January through April. An explanation may be that WFOs may pay closer attention to dewpoint forecasts during seasons of the year with higher fire danger (D. Ruth, 2006, personal communication). However, the 10-month, single year span of this study is probably insufficient to draw firm conclusions about seasonal effects.

The regression plots of MAE on distance in Fig. 5 indicate an extremely weak correlation between distance from observation to NDFD grid point and MAE. Individual correlations were examined for each projection and each of the five issuances. The means for all projections, all issuances do not differ greatly from any of the individual correlations. A tangential finding of this study is that proximity to an NDFD grid point, within a grid cell, does not give a verification site a strong advantage, and therefore, there is no penalty associated with verifying at 5 kilometer resolution rather than a finer resolution.

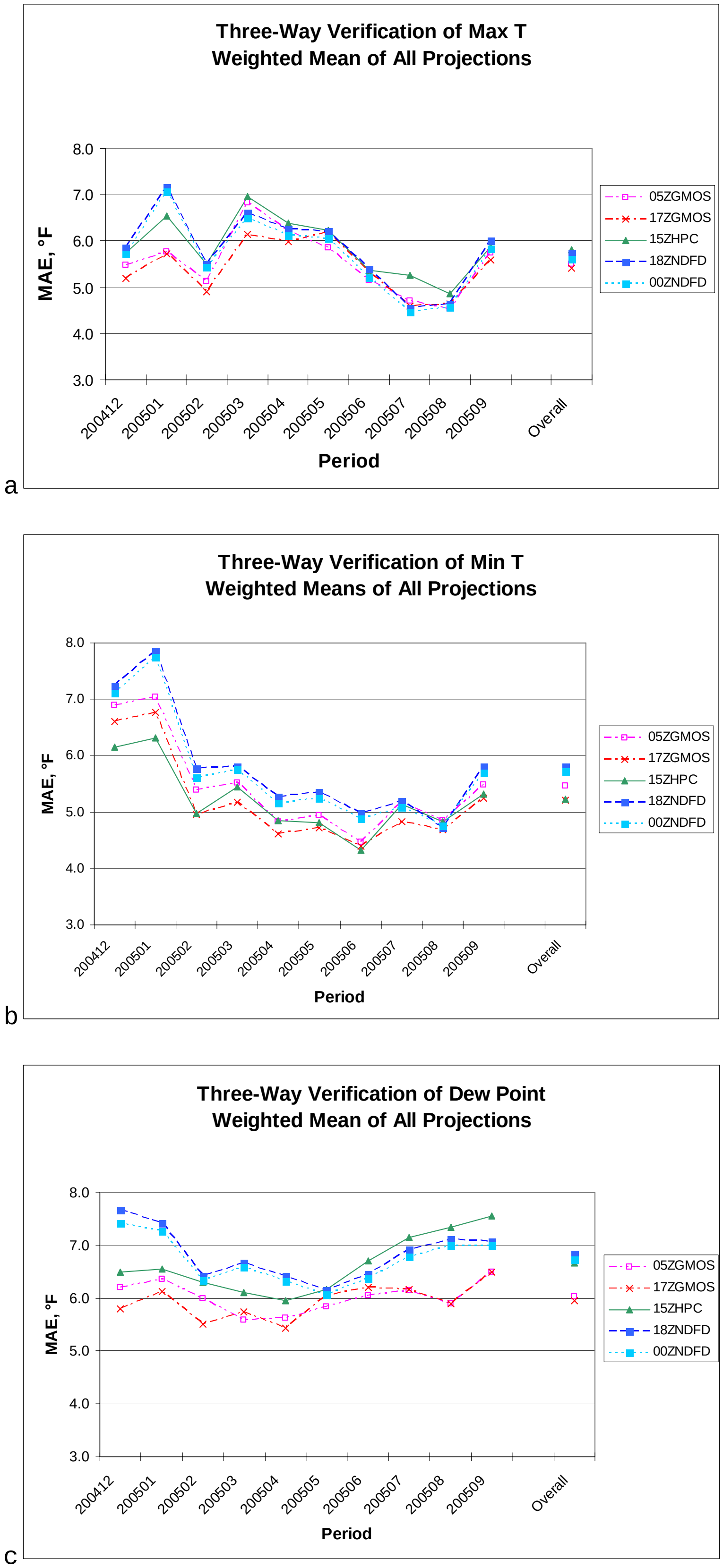

Figure 4. Apparent seasonal effect on MAE scores for maximum temperature (a), minimum temperature (b), and dew point (c), taken for non-MOS sites. The mean for all months is at the far right.

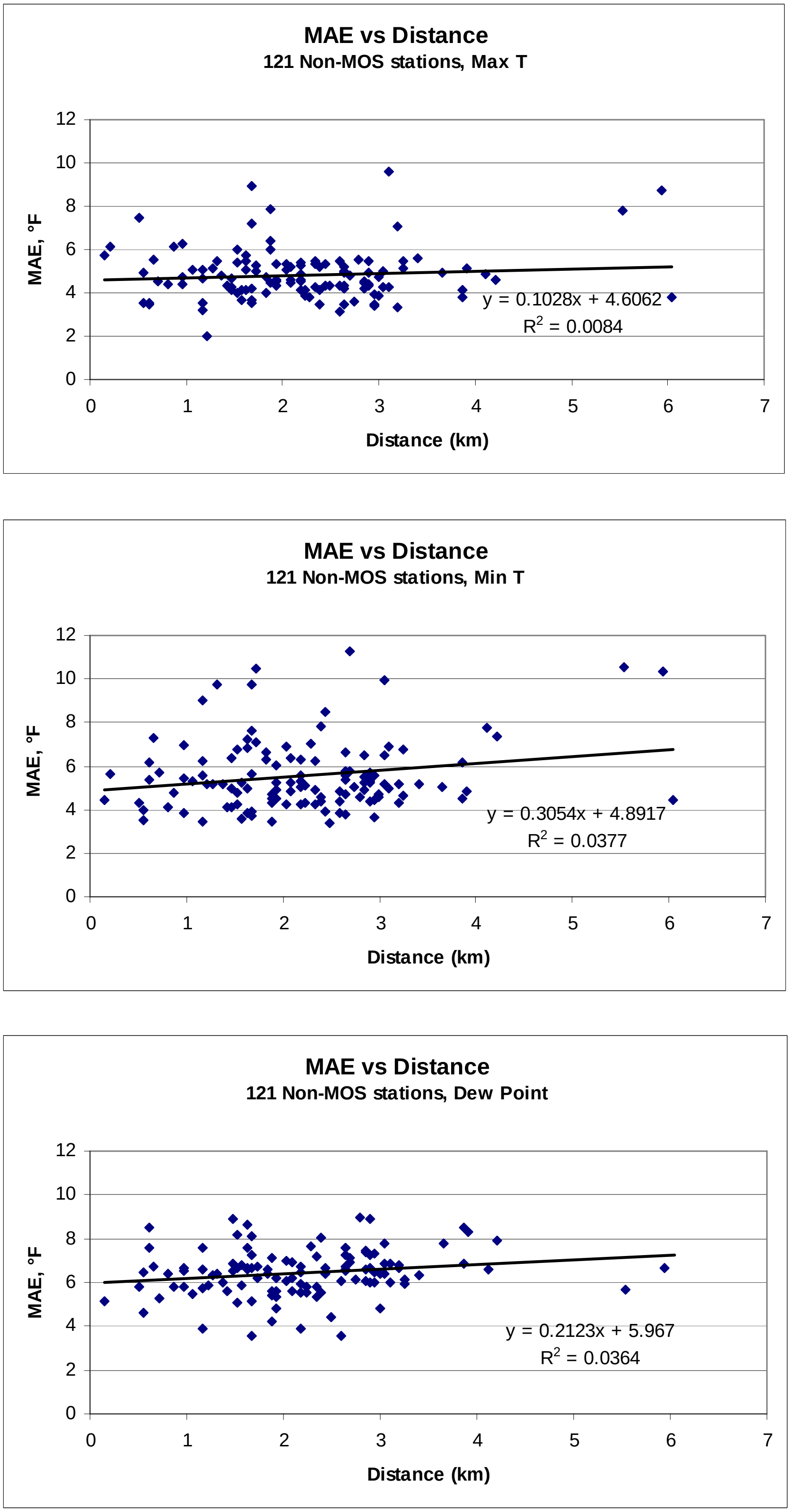

Figure 5. Regression of MAE on distance to nearest NDFD grid point. MAE is average of all projections, all issuances, for non-MOS sites.

ACKNOWLEDGEMENTS

I thank Valery Dagostaro for her assistance with MOS2000 code, and constant helpful comments and reviews; Statistical Modeling Branch of the Meteorological Development Laboratory for close cooperation and assistance with observation data and GMOS forecasts; Bob Glahn, Wilson Shaffer, Paul Dallavalle, and David Ruth for guidance and comments; and Arthur Taylor for assistance with forecast data.

REFERENCES

Dallavalle, J. P., and B. Glahn, 2005: Toward a gridded MOS system. Preprints, *21st Conference on Weather Analysis and Forecasting*, Washington, D.C., Amer. Meteor. Soc., 13B.2.

Glahn, H. R. and D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.*, **84**, 195-201.

IFPS Science Steering Team, 2004: HPC 4-7 Day Grid Generation and Field Use, www.nws.noaa.gov/ost/ifps_sst/pdf_docs/HPC_4-7_Day_ISST_Position_Paper.pdf