Performance of NOAA-EPA Air Quality Predictions, 2007 – 2009

Jerry Gorline[1] and Pius Lee[2]
[1]Meteorological Development Laboratory, National Weather Service, NOAA, Silver Spring, MD, USA
[2]Air Resources Laboratory, Office of Oceanic and Atmospheric Research, NOAA, Silver Spring, MD, USA

1. INTRODUCTION

In response to Congressional direction to National Oceanic and Atmospheric Administration (NOAA) to build an operational air quality forecast capability, NOAA has been developing, testing and implementing phased expansions of a National Air Quality Forecast Capability (NAQFC) since 2003. The capability is being built in partnership with the Environmental Protection Agency (EPA). The initial operational capability was implemented at the National Weather Service (NWS) in September 2004 (Otte et al. 2005), producing twice-daily forecasts of ground-level ozone across the northeastern United States. In the initial capability, the NWS/National Centers for Environmental Prediction (NCEP) North American Mesoscale (NAM) model was used to drive the Community Multi-scale Air Quality (CMAQ) model to produce next-day ozone predictions at 12-km grid resolution. The NAQFC has been expanded via a program of phased development and testing with implementations of ozone predictions over the entire eastern U.S. in 2005, and to the conterminous United States (CONUS) in 2007. Further goals for the NAQFC include providing quantitative predictions of fine Particulate Matter (PM2.5), which together with ozone is associated with almost all of the poor air quality episodes in the U.S. As a step toward building particulate matter prediction capabilities, NOAA has been testing a version of the CMAQ model that includes an aerosol prediction module that incorporates contributions to PM2.5 from the EPA's National Emissions Inventory.

Surface ozone concentrations in the CONUS begin to increase in May, are highest June to August, and begin to decrease in September. Ozone activity is relatively low from November through March over the CONUS, with occasional outbreaks in October and May. In contrast, surface concentrations of fine aerosols tend to be active throughout the year, especially during summer and winter. Sulfate particles tend to accumulate in the eastern U.S. with higher concentrations in the summer. Nitrates are found in more abundance in the western U.S. and concentrations peak in the winter, (EPA 2003). Figure 1 shows the number of daily observed values above a threshold for 1-h aerosols in blue and for 8-h ozone in red, over the CONUS, June 2008 to September 2009. The thresholds are 35 ug/m^3 for aerosols and 76 parts per billion (ppb) for ozone. This plot shows that ozone is active in the summer months and is relatively quiet in the winter, while aerosols stay active throughout the year.

The NWS Meteorological Development Laboratory (MDL) compared the performance of CMAQ experimental ozone predictions and developmental fine aerosol predictions, over the CONUS and six geographic regions in the CONUS. The six regions are the Pacific Coast (PC), Rocky Mountains (RM), Lower Midwest (LM), Upper Midwest (UM), South East (SE), and North East (NE). A map of the six regions can be found in Gorline and Lee (2008). On some occasions developmental testing of the aerosol predictions was interrupted. Higher test priority for experimental ozone predictions resulted in fewer interruptions of daily predictions. Further information regarding differences in developmental and experimental test configurations is provided in McQueen et al. (2005). We produced verification scores for developmental aerosol predictions to provide feedback for possible model configuration changes. MDL also produced verification for experimental ozone to assist in the validation of performance evaluation metrics provided by NCEP.

MDL provided a performance evaluation of predicted surface ozone concentrations and fine aerosol concentrations against observations compiled by the EPA. Our verification metrics included categorical analyses for Fraction Correct (FC), Threat Score (TS), Probability of Detection (POD), and False Alarm Rate (FAR). For a more detailed discussion about two-by-two contingency table analyses, see Wilks (1995). We also calculated monthly and seasonal Mean Absolute Error (MAE) and bias, where bias is forecast minus observation. We compared categorical performance of next-day maximum 8-h average ozone predictions for June to September, 2007, 2008, and 2009, based on daily tests driven by the 1200 UTC NAM cycle. For developmental aerosol predictions, we compared

*Corresponding author: Jerry L. Gorline, Meteorological Development Laboratory, NWS, 1325 East West Highway, Silver Spring, MD. 20910; fax: 301-713-9316; phone: 301-713-1768; e-mail: jerry.gorline@noaa.gov

categorical performance of next-day maximum 1-h average 0600 UTC cycle predictions for January 2008 to September 2009.

2. PERFORMANCE OF 8-H OZONE PREDICTIONS

During the summer of 2007, 2008, and 2009, MDL generated categorical verification metrics for experimental ozone predictions. We concentrated our analyses on the daily maxima of 8-h averaged predictions above the 76 ppb threshold. The EPA provided ozone observations for 1,160 sites within CONUS and 90 sites in Canada. For this paper, all 8-h predictions or observations that were equal to or greater than the threshold during a predefined 24-h period will be called E-76 events. The 24-h window for counting E-76 events began hour 24 of the 1200 UTC CMAQ 48-h forecast period. For more information about the timing of our ozone performance metrics, see Gorline et al. (2006).

Figure 2 shows a plot of the mean 8-h experimental ozone predictions and observations for August 2009, overlaid with the mean bias for August 2009, 2008, and 2007. The mean bias was about three ppb higher in 2008 than in 2007. The mean bias in August 2009 was two ppb higher than in 2008 and five ppb higher than in 2007. The summer of 2009 was unusually cool in the eastern and northern U.S. There were significantly fewer observed 8-h E-76 events in four out of the six regions, RM, UM, SE, and NE, compared to 2007 and 2008. Figure 3 shows the number of observed 8-h ozone E-76 events by region, for the summer of 2007, 2008, and 2009. We can see that 2007 was the most active year, except for 2008 in the PC region where there were 1,841 observed 8-h E-76 events. In 2009, there were fewer than 100 observed E-76 events in the East and upper Midwest. The unusually cool summer may have played a role in the slightly higher model bias in 2009, compared to 2008. The experimental ozone predictions were produced with the CBIV chemical mechanism in 2007 and CB05 was used in 2008 and 2009. Therefore, differences in performance between 2007 and later years also reflect the difference in chemical mechanism. Comparing predictions with CBIV and CB05 for the same year show that CB05 systematically increases over-prediction (not shown). For more information about CBIV and CB05, see Yarwood et al. (2005).

Figure 4 shows the TS by region, of 8-h ozone predictions for the summers of 2007, 2008, and 2009. The TS was calculated for June 15 to August 31. Figure 4 also shows the 95% Confidence Interval (CI), a measure of uncertainty for the TS performance metric that was calculated using the non-parametric bootstrap method described by Jolliffe (2007). We see in Fig. 4 that the TS was much lower in the UM, SE, and NE regions in 2009 compared to 2007 and 2008. These lower 2009 TS values were most likely caused by the lower number of observed E-76 events. Among all regions, PC and LM have the highest number of observed E-76 events and the highest TS during 2009. For PC and LM regions the TS values are similar from 2007 to 2009. Daily performance of the model improves on days with 50 or more observed E-76 events over the CONUS, compared to days with fewer than 50 E-76 events (not shown). The results shown in Figs. 3 and 4 suggest that this assessment is also true for seasonal comparisons.

3. PERFORMANCE OF 1-H AEROSOL PREDICTIONS

During 2008 and 2009, MDL generated categorical verification metrics for aerosols over the CONUS developmental domain. For this paper, all daily maxima of 1-h average aerosol predictions or observations that were equal to or greater than a threshold of 35 ug/m^3 during a predefined 24-h period, will be called E-35 events. The 24-h window for counting E-35 events began hour 22 of the 0600 UTC CMAQ 48-h forecast period. The EPA provided fine aerosol observations for 620 sites within CONUS and 41 sites in Canada. These observations were temperature corrected and quality-control edited by the EPA. If an observation or model prediction for a station was missing, we excluded that station from our calculations. For more details about the CMAQ aerosol module, see Gorline and Lee (2008).

Figure 5 shows a monthly average of the FC of 1-h aerosol predictions for the six geographic regions, January 2008 to August 2009. The FC is often higher in the summer months than in the winter. Under-prediction in the summer reduces the number of false alarms compared to the winter, resulting in the higher FC values. The decrease in the FC was not as large in the winter of 2009 compared to the winter of 2008, for five out of six of the regions. The PC region had a lower FC in January 2009 compared to January 2008. But the PC region also had a more active winter in 2009 and may be the reason for the lower FC. In June 2008, the chemical mechanism was updated to CB05 and the aerosol module to AERO-4. The updates may have reduced the over-prediction in these regions during the winter of 2009, resulting in the higher FC values.

Figure 6 shows the monthly average bias of 1-h aerosol predictions for the six geographic regions, January 2008 to August 2009. The bias values were higher in the winter months, October to March, compared to the summer, April to September. The bias values were not as high in January 2009 compared to January 2008 in all regions except the PC region. There were more over-predictions in the winter, when the bias was positive. Comparing Figs. 5 and 6, we see that the FC decreased with the higher bias during the winter months and the FC increased with the lower bias in the summer months.

4. CONCLUSIONS

Comparisons of experimental ozone predictions over 2007 to 2009 are helpful for assessing impacts of several different model configurations. Experimental predictions for ozone in 2007 over CONUS were based on the CBIV chemical mechanism. Beginning in 2008, experimental predictions were based on the newer CB05 mechanism. The average bias over the CONUS for experimental ozone predictions was about 3 ppb higher in 2008, than in 2007. Small seasonal increases in bias for ozone predictions in 2007 and 2008 were noted. Biases ranged from slight under-prediction in June to slight over-prediction by July. This bias change is consistent with NAM predictions that exhibited a cold/wet bias in the June surface temperature and dew point predictions. The slightly higher bias in 2009 compared to 2008, may have been caused by the unusually cool summer in the East in 2009. The lower TS values in the East in 2009 are associated with the lower number of observed 8-h E-76 events, and related to the cooler weather, compared to 2007 and 2008.

For developmental aerosol predictions, there were strong seasonal bias changes, from under-prediction in the warm season, April to September, to over-prediction in the cool season, October to March. While these biases are consistent with missing source contributions (e.g. wildfires) in the summer months, additional complexity of the aerosol test predictions are contributing to large prediction errors, and are the subject of ongoing investigation. The CB05 chemical mechanism and AERO-4 aerosol module introduced in June 2008 may have reduced over-prediction in five out of the six regions for January 2009, compared to January 2008. Work is continuing on determining the causes for the seasonal changes in bias of the aerosol predictions.

5. ACKNOWLEDGMENTS

This paper was prepared in cooperation with development activities carried out under the auspices of NWS's National Air Quality Forecast Capability. We would like to thank NCEP for providing the forecasts of surface concentration of ozone and fine aerosols and the EPA for providing the associated verifying observations.

6. REFERENCES

Environmental Protection Agency, 2003: Guidelines for Developing an Air Quality (Ozone and PM2.5) Forecasting Program, U.S. EPA, Office of Air Quality Planning and Standards.

Gorline, J. L., M. Schenk, W. Shaffer and A. Taylor, 2006: Verification of the NOAA-EPA Air Quality Forecasts For Summer 2005, *5th Annual Community Modeling and Analysis System Conference*, October 16-18, University of North Carolina, Chapel Hill.

Gorline, J. L. and P. Lee, 2008: Performance Evaluation of NOAA-EPA Developmental Aerosol Forecasts, *Journal Environmental Fluid Mechanics*, DOI:10.1007/s10652-008-9090-7, Springer Netherlands.

Jolliffe, I. T., 2007: Uncertainty and Inference for Verification Measures, *Wea. Forecasting*, **22**, 637-650.

McQueen, J. T., P. C. Lee, M. Tsidulko, G. DiMego, T. Otte, J. Pleim, G. Pouliout, J. Young, D. Kang, P. M. Davidson, and N. Seaman, 2005: Update to and Recent Performance of the NAM-CMAQ Air Quality Forecast Model at NCEP operations, *17th Conference on Numerical Weather Prediction*, Amer. Meteor. Soc., Washington, D.C., 12A.2.

Otte, T. L., G. E. Pouliot, J. E. Pleim, J. O. Young, K. L. Schere, D. C. Wong, P. C. S. Lee, M. Tsidulko, J. T. McQueen, P. Davidson, R. Mathur, R. H. Chuang, G. DiMego, and N. L. Seaman, 2005: Linking the Eta Model with the Community Multiscale Air Quality (CMAQ) Modeling System to Build a National Air Quality Forecasting System, *Wea. Forecasting*, **20**, 367–384.

Wilks, D. S., 1995: *Statistical Methods in the Atmosphic Sciences: An Introduction.* Academic Press, 238 – 241.

Yarwood, G., S. Rao, M. Yocke, and G. Whitten, 2005: Updates to the Carbon Bond Chemical Mechanism: CB05, U.S. EPA, RT-0400675.

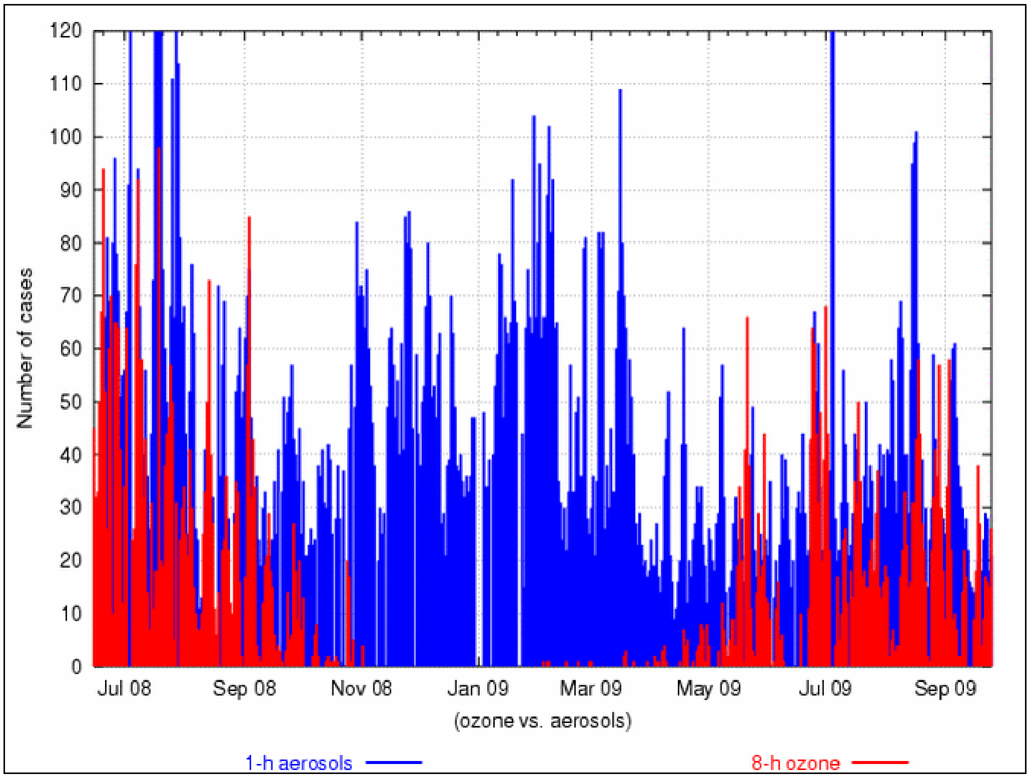

Fig. 1. Number of observed values higher than threshold, 8-h avg ozone vs. 1-h avg aerosols, June 2008 to September 2009, threshold: 76 ppb (ozone), 35 ug/m^3 (aerosols).

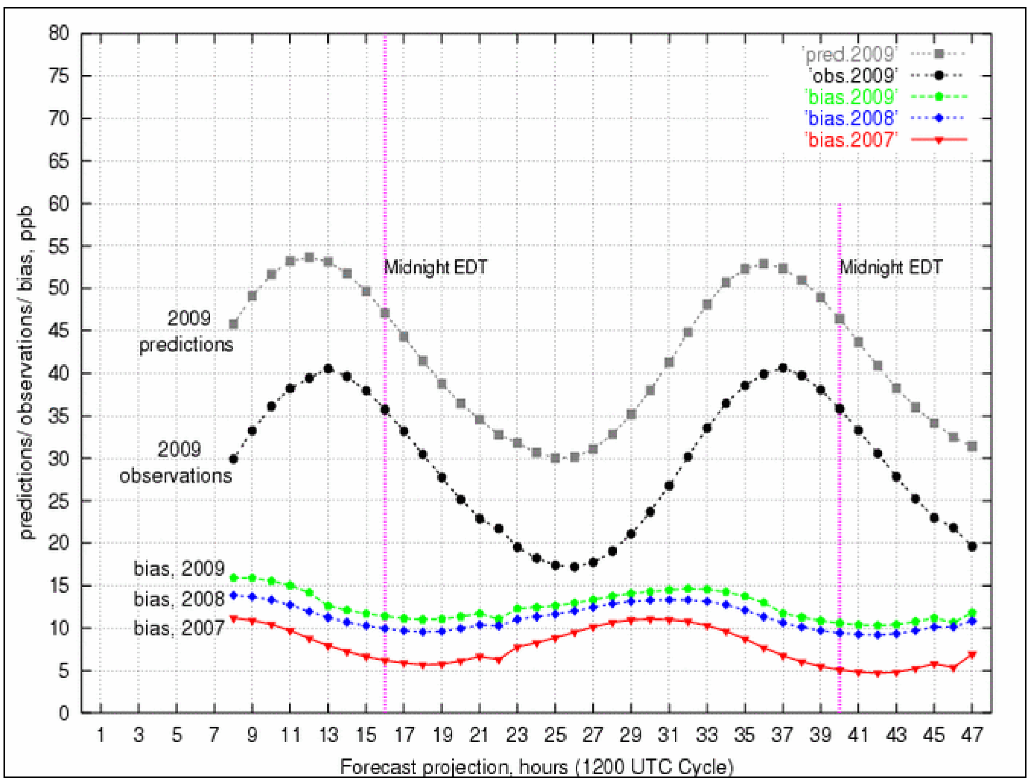

Fig. 2. Mean bias, three-year comparison, 8-h ozone predictions vs. EPA observations, August 2007, 2008, and 2009, 1,250 stations.

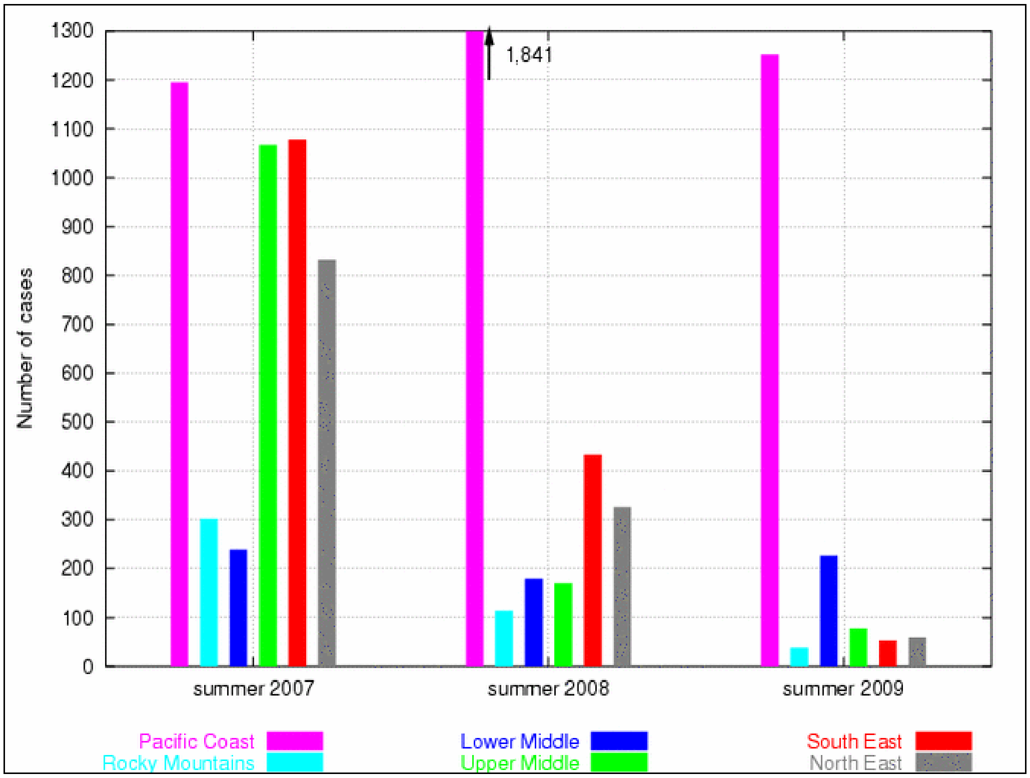

Fig. 3. Number of observed values higher than 76 ppb, all regions, three-year comparison, 8-h average ozone, summer (June 15 to August 31) 2007, 2008, and 2009, 1,250 stations.

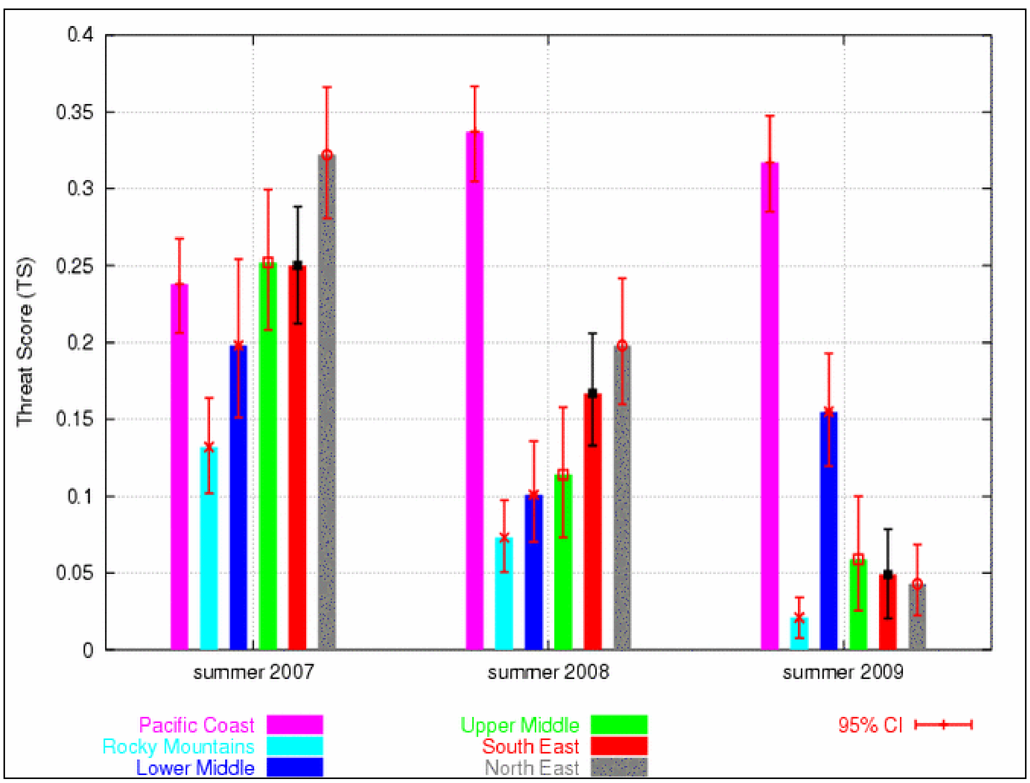

Fig. 4. Threat Score, all regions, 8-h ozone predictions vs. EPA observations, summer 2007, 2008, and 2009, 1,250 stations.

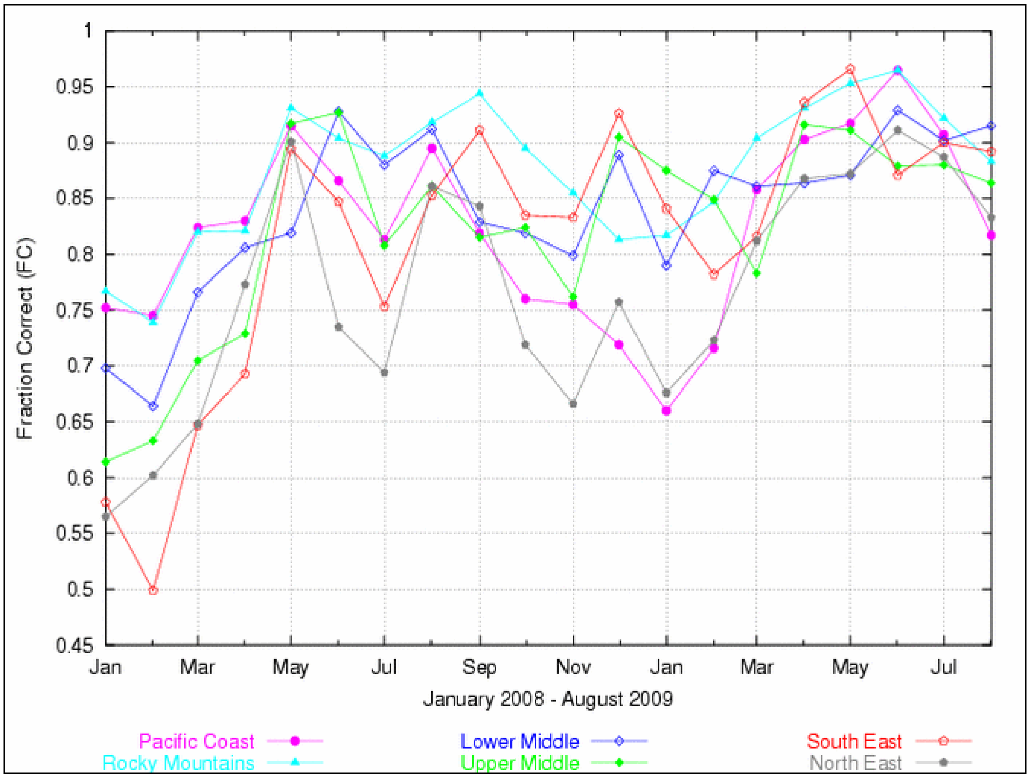

Fig. 5. Fraction Correct, all regions, 1-h average aerosol predictions vs. EPA observations, monthly average, January 2008 to August 2009, 661 stations.

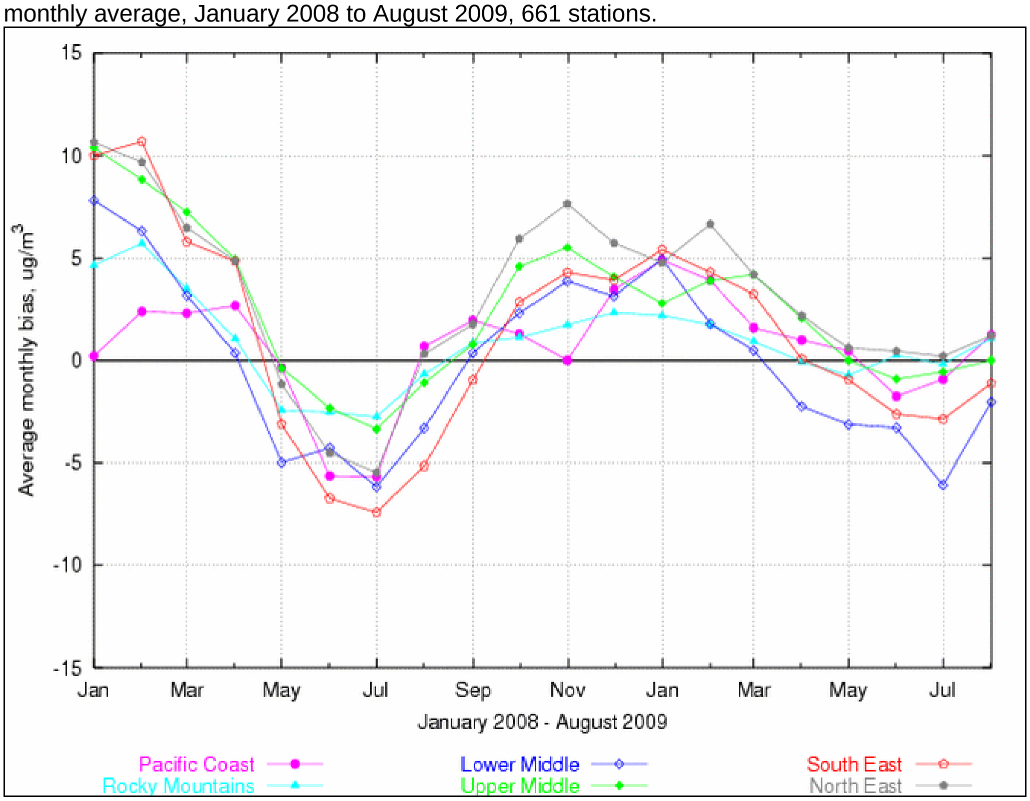

Fig. 6. Average monthly bias, all regions, 1-h average aerosol predictions vs. EPA observations, January 2008 to August 2009, 661 stations.