EXPLORING APPROACHES TO INTEGRATE OBSERVATIONS AND CMAQ SIMULATIONS FOR IMPROVED AIR QUALITY FORECASTS

C. Hogrefe[1,2,*], W. Hao[1], K. Civerolo[1], J.-Y. Ku[1], and G. Sistla[1]

[1]New York State Department of Environmental Conservation, [2]Atmospheric Sciences Research Center, State University of New York at Albany

1. INTRODUCTION

A recent study compared real-time air quality simulations performed with the ETA/CMAQ modeling system over New York State against observations and routine expert-based air quality forecasts for O_3 and $PM_{2.5}$ (Hogrefe et al., 2006). Results indicated that in most regions of New York State, the routine air quality forecasts based on previous day's observed concentrations and expert judgment showed slightly better agreement with the observed distributions of Air Quality Index (AQI) categories than CMAQ simulations. However, CMAQ showed skill similar to these routine forecasts in terms of capturing the AQI tendency, i.e. in predicting changes in air quality conditions. Therefore, it was concluded that it might be beneficial to develop tools that combine CMAQ's predictive capability in terms of temporal trends with real-time observations of ambient pollutant levels to generate improved forecasts.

In this study, we present several potential approaches for implementing this concept. The utility of these approaches is evaluated by comparing the resulting forecast fields fields against observations and unadjusted CMAQ simulations. These analyses focus on daily maximum 8-hr ozone and 24-hr average $PM_{2.5}$ concentrations for the summer of 2005.

2. DATABASE AND METHODS OF ANALYSIS

2.1 Model Setup and Observational Database

In this study, we utilize observations and archived CMAQ forecasts that were generated by the New York State Department of Environmental Conservation (NYSDEC) for June 1 – September 30, 2005. The CMAQ air quality model is documented in Byun and Schere (2005). The specific setup of the CMAQ air quality forecasting system employed by NYSDEC is described in Hogrefe et al. (2006) and is based upon the forecasting system developed by the National Weather Service (NWS), the National Oceantic and Atmospheric Administration (NOAA), and the Environmental Protection Agency (EPA) (Davidson et al., 2004; Otte et al., 2005).

Observations of hourly ozone and total $PM_{2.5}$ concentrations for monitors in New York State were downloaded from the EPA AIRNOW system. Daily maximum 8-hr ozone concentrations and 24-hr average $PM_{2.5}$ concentrations were then determined from the hourly data and used in the subsequent analyses. It should be noted that the focus of the AIRNOW database is on providing access to monitoring information in near real-time. Therefore, while basic quality assurance is performed via automated checks on minimum/maximum values, rates of change, etc., these data are considered preliminary and are subject to more complete quality assurance prior to integration into the AQS database.

2.3 Postprocessing Approaches

In this study, we compare five methods to adjust CMAQ model simulations for improved air quality forecasts. In the equations below, the next-day forecasts resulting from these five methods are denoted as F1 – F5.

$$F1_{i+1} = CMAQ_{i+1} - (\overline{CMAQ} - \overline{Obs})$$

where *i* represents today, *i+1* represent tomorrow, and the overbar represents a temporal average. This can be viewed as a simple bias adjustment of the original CMAQ predictions.

$$F2_{i+1} = CMAQ_{i+1} - [\overline{CMAQ} - \overline{Obs}] \Big| CMAQ_{i+1}$$

where the vertical bar indicates that the average difference between observations and CMAQ predictions is computed for a number of model-predicted concentration bins and the bias adjustment on any given day corresponds to the bin in which the original CMAQ forecast falls.

$$F3_{i+1} = Obs_i + (CMAQ_{i+1} - CMAQ_i)$$

[*] *Corresponding author contact information: Christian Hogrefe, BAQAR, New York State Department of Environmental Conservation, 625 Broadway, Albany, NY 12233-3259, Phone (518) 402 8402, Fax (518) 402 9035, Email chogrefe@dec.state.ny.us*

This represents a forecasting approach that utilizes the CMAQ-predicted change between today and tomorrow and adds this tendency to today's observed value.

$$F4_{i+1} = Obs_i + (CMAQ_{i+1} - CMAQ_i) \times \frac{\sigma_{Obs}}{\sigma_{CMAQ}}$$

where σ_{obs} and σ_{CMAQ} represent the standard deviation of the observed and CMAQ simulated concentrations, respectively. This is similar to method 3, but adjusts the CMAQ-predicted changes to account for differences in the spread of observed and model-predicted distributions.

$$F5_{i+1} = Obs_i + (CMAQ_{i+1} - CMAQ_i) \times s$$

where *s* represents the slope of the least-squares linear regression fit between unadjusted CMAQ simulations and observations. This is an alternate approach to method 4 for utilizing CMAQ-predicted changes adjusted for differences in the spread of observed and model-predicted distributions.

It is important to point out that methods 1-2 and 4-5 all rely on incorporating observations not just for the current day but for an extended time period. In a routine forecast setting, this extended time period could be the past week, month, or season. In this study, we utilized the fixed time period from June 1 – September 30 also used for evaluating these methods, i.e. for any given day, both past and future observations were included. While this choice would not be feasible in a routine forecast setting, it is still suitable for demonstrating the characteristics of these adjustment methods since no special emphasis is placed on the actual observed next-day values relative to all other values from the 4-month time period that are included in approaches 1-2 and 4-5. Future analysis might consider the impact of the choice of the "calibration" or "learning" period over which the adjustment parameters in methods 1-2 and 4-5 are calculated on the performance of these approaches, but this is beyond the scope of the current study.

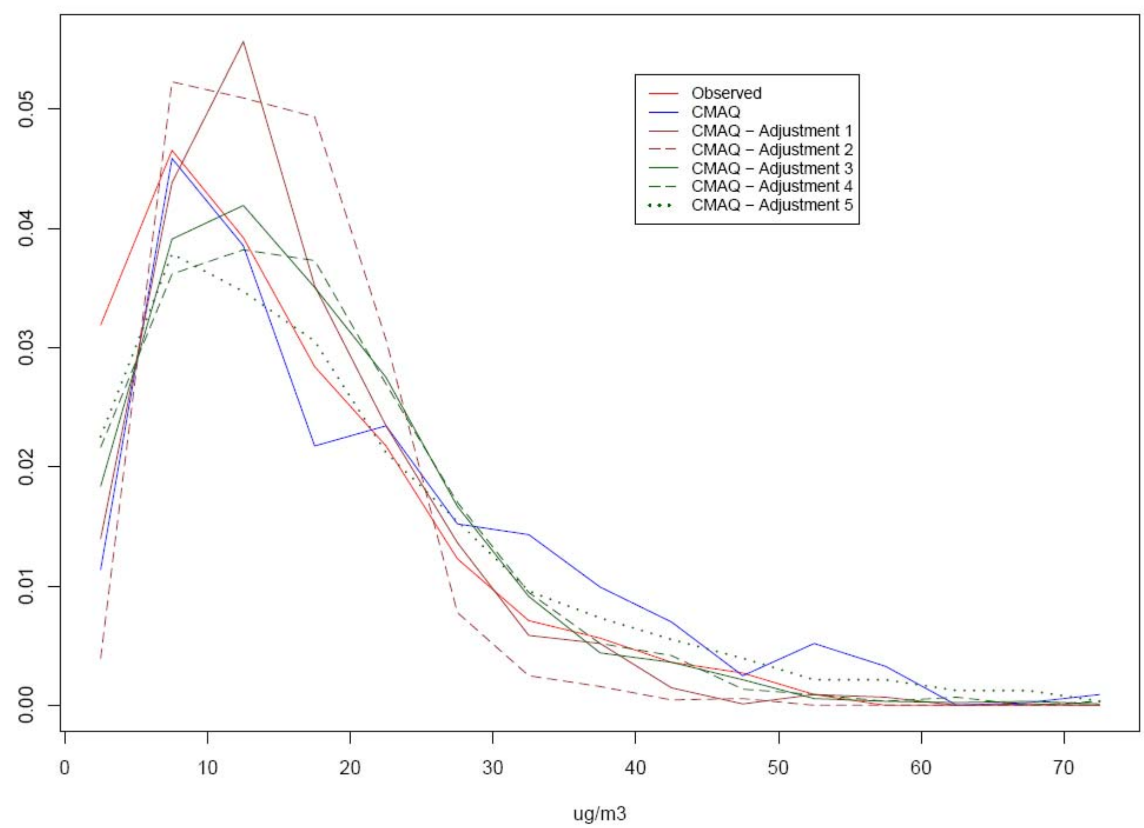

Figure 1: Distributions of 24-hr average $PM_{2.5}$ concentrations from observations, original CMAQ predictions, and adjustment methods 1 – 5. These distributions were constructed from all available data at the 21 $PM_{2.5}$ monitors considered in this analysis.

2.4 Evaluation Metrics

To compare the ability of the five model adjustment methods to provide improved forecasts, we employ discrete as well as categorical performance measures (Kang et al., 2005). The discrete forecast measures are the total, systematic, and unsystematic root mean square error (RMSE) as defined by Willmott (1982) and the categorical metrics are the False Alarm Ratio (FAR), Probability of Detection (POD), and Critical Success Index (CSI) as described by Kang et al. (2005). For the categorical metrics, we selected a threshold that corresponds to the transition from the “moderate” to the “unhealthy for sensitivty groups” range of the Air Quality Index (AQI). For ozone, this threshold corresponds to 84 ppb, while for $PM_{2.5}$, it corresponds to 40 µg/m^3 (U.S. EPA, 1999).

3. RESULTS AND DISCUSSION

As an illustration of the effect of adjustment methods, Figure 1 shows distributions of 24-hr average $PM_{2.5}$ concentrations from observations, original CMAQ predictions, and adjustment methods 1 – 5. These distributions were constructed from all available data at the 21 AIRNOW $PM_{2.5}$ monitors considered in this analysis. It can be seen that the distribution from the original CMAQ predictions has a heavier tail than the one constructed from observations, consistent with the overprediction of $PM_{2.5}$ at many monitors in New York State reported by Hogrefe et al. (2006). This figure illustrates that all adjustment methods show closer agreement with the observed distribution than the unadjusted CMAQ simulation. However, adjustment methods 1 and especially 2 now tend to underestimate high observed concentrations while methods 3 through 5 lead to distributions that are close to the observed distributions for concentrations above 25 µg/m^3.

While Figure 1 illustrates that the adjustment methods investigated generally lead to a closer agreement between observed and predicted distributions, this does not necessarily indicate that they improve the ability to capture the

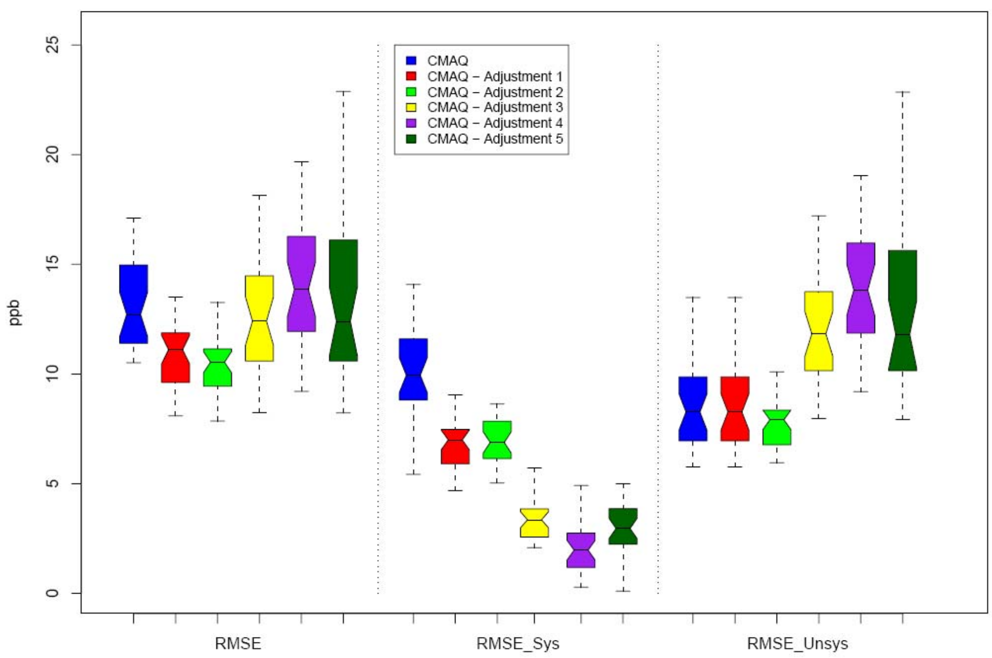

Figure 2: Total, systematic, and unsystematic RMSE of daily maximum 8-hr ozone concentrations predicted by the unadjusted CMAQ simulations and adjustment methods 1 – 5. Results were computed separately at each site, and the boxplots indicate the range of results across the 38 ozone monitors considered in this analysis. The center bar of each box represents the median, the lower and upper limits of the box represent the 25th and 75th percentiles, respectively, and the whiskers extend to the most extreme data point having a distance from the box that is no more than 1.5 times the height of the box. If the notches for different boxes do now overlap, it indicates that the medians associated with each box are significantly different at the 95% confidence level.

temporal fluctuations present in the observations. As discussed in Section 2, we measure this ability through the RMSE (total, systematic, and unsystematic), and the categorical metrics FAR, POD, and CSI. Figure 2 shows RMSE results for daily maximum 8-hr ozone. The boxplots illustrate the range of the various metrics across the 32 ozone monitors utilized here. It is evident that methods 1 and 2, which focus on reducing the overall model bias, reduce the median overall RMSE. This reduction is achieved primarily through a reduction of the systematic RMSE while the unsystematic RMSE remains largely unchanged. On the other hand, adjustment methods 3-5 that use today's observations as starting point and adjust it with unscaled or scaled model-predicted changes show an even stronger reduction in the systematic RMSE, but also an increase in the unsystematic RMSE, leading to little changes in the overall RMSE. The significant reduction of the systematic RMSE for methods 3-5 is expected since these methods aim at moving the best-fit line of model predictions vs. observations closer to a 1:1 line, the departure from which is measured by the systematic RMSE. A possible interpretation of the increase of the unsystematic RMSE might be that model predicted day-to-day changes, i.e. the first derivative of the CMAQ predicted time series, contain a higher level of random fluctuations than the original time series. These random fluctuations are then further amplified if the model-predicted changes are multiplied by the ratios of observed and predicted standard deviations. Results for $PM_{2.5}$ are qualitatively similar to those shown in Figure 2 for ozone.

Figure 3 shows categorical performance metrics for $PM_{2.5}$ calculated for a threshold of 40 $\mu g/m^3$ for unadjusted and adjusted CMAQ predictions. The numbers shown in this Figure were calculated as average over all $PM_{2.5}$ monitors considered in this study. Methods 1 and 2 reduce both the FAR and the POD, consistent with Figure 1 that showed a tightening of the distributions for these two methods. On the other hand, method 3 shows an increase in the FAR and a constant POD, while methods 4-5 shows an increase in both the FAR and the POD. The results for the CSI, which measures the overall ability of a prediction to capture exceedances while avoiding false alarms, indicate that methods 1-4 all provide better performance in this category

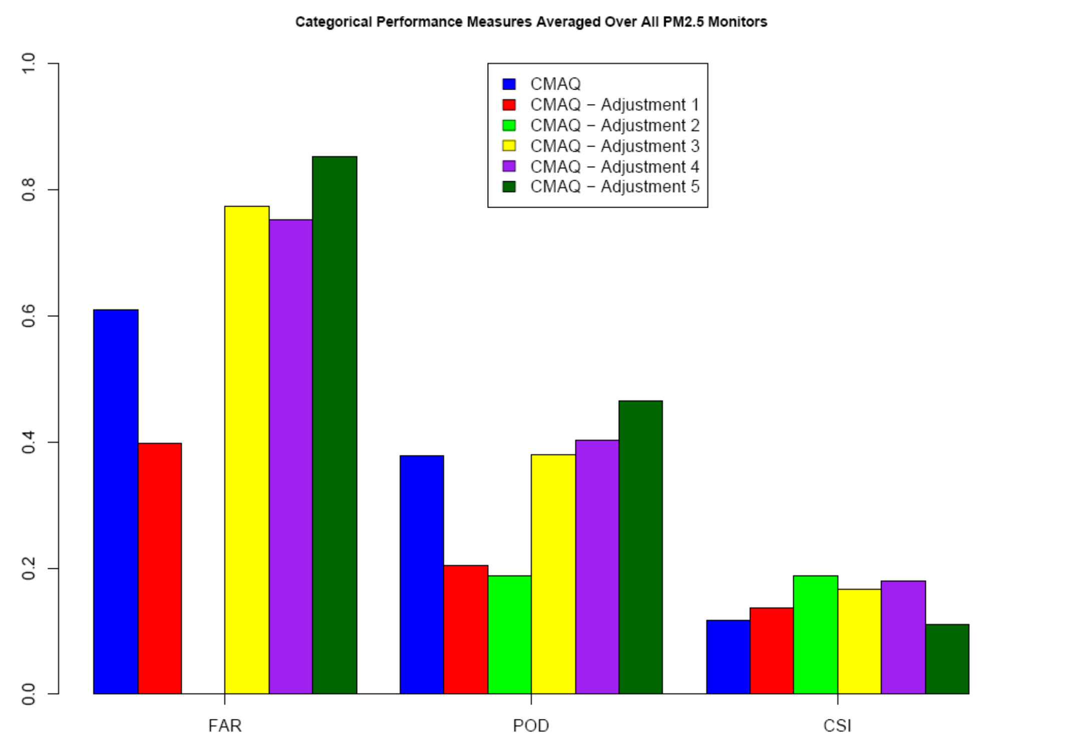

Figure 3: False Alarm Ratio (FAR), Probability of Detection (POD), and Critical Success Index (CSI) of daily average $PM_{2.5}$ concentrations predicted by the unadjusted CMAQ simulations and adjustment methods 1 – 5. Results were computed separately at each site, and the bars represent the average results across the 21 $PM_{2.5}$ monitors considered in this analysis.

compared to the unadjusted CMAQ results. The results for ozone (not shown here) indicate that methods 2 and 3 perform best for this metric. This is driven by only a slight increase in the FAR but a larger increase in the POD when method 3 is used to adjust CMAQ predictions.

While the results shown in figure 3 were averaged over all stations, we also determined the best adjustment method separately for each metric (lowest total RMSE or highest CSI) and pollutant (ozone and $PM_{2.5}$) at each site. Figure 4 shows the percentage of sites at which a given adjustment method performed best for a given metric and pollutant. If there was a tie for the best performance between different metrics at a given station, they were all included in the count for that station and weighted accordingly. This figure clearly illustrates that method 2, the binned bias correction, is the best approach to lower the overall RMSE at most monitors for both O3 and $PM_{2.5}$. However, the results are more varied when one considers the CSI. For both ozone and $PM_{2.5}$, there is a small number of stations at which the original CMAQ predictions yield a higher CSI than any of the adjustment methods explored in this study. At about half of the remaining stations where one of the postprocessing approaches yields a higher CSI than the unadjusted CMAQ results, method 3 is the method of choice for ozone, followed by methods 5, 4, 1, and 2. For $PM_{2.5}$, however, methods 2 – 4 appear to improve the forecasts at a roughly similar fraction of sites. If one draws a broad distinction between methods 1-2 (the "bias-correction" approaches) and methods 3-5 (the "CMAQ tendency" approaches), the latter appear to be superior to the former in the case of ozone at most sites, but this finding does not necessarily hold true for $PM_{2.5}$.

4. SUMMARY

This paper described the comparison of five approaches aimed at providing improved air quality forecasts based on both observations and CMAQ simulations. While the "bias-correction" approaches 1 or 2 work best for reducing the total RMSE at most sites, the approaches that combine today's observations with unadjusted or adjusted CMAQ-predicted temporal changes often work best for improving the CSI, especially for ozone. Moreover, it appears that the best adjustment

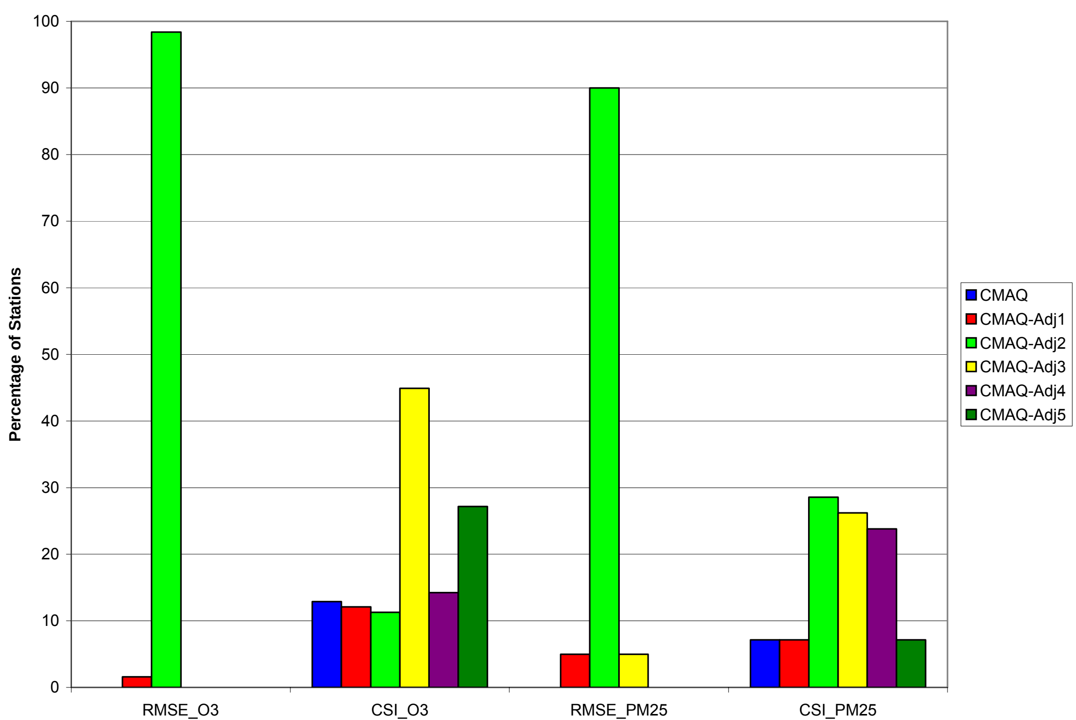

Figure 4: Percentage of sites at which a given adjustment method performed best for a given metric and pollutant. Grouped from left to right, the metrics and pollutants considered are: total RMSE for daily maximum 8-hr ozone; CSI for daily maximum 8-hr ozone using a threshold of 84 ppb; total RMSE for daily average $PM_{2.5}$; and CSI for daily average $PM_{2.5}$ using a threshold of 40 μg/m^3.

method to improve the CSI, which measures the quality of categorical forecasts, needs to be chosen on a pollutant-by-pollutant and station-by-station basis. It should be noted that the list of approaches investigated here is by no means complete. For example, Delle Monache et al. (2006) and Kang et al. (2006) describe the application of a Kalman filter bias correction to generate improved air quality forecasts and report good success. Additional methods might aim at including spatial correlation structures into the model adjustment algorithm rather than relying solely on temporal structures at individual monitors. Such analyses will be performed in the future.

5. ACKNOWLEDGMENTS

The work presented in this paper was performed by the New York State Department of Environmental Conservation with partial support from the U.S. EPA under cooperative agreement CR83228001. The views expressed in this paper do not necessarily reflect the views or policies of the New York State Department of Environmental Conservation or those of the U.S. EPA.

6. REFERENCES

Byun, D.W. and K.L. Schere, 2006: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. Appl. Mech. Rev., 59, 51-77.

Davidson, P.M., N. Seaman, K. Schere, R.A.Wayland, J.L. Hayes and K.F. Carey, 2004: National Air Quality Forecasting Capability: First Steps toward Implementation. Preprints, 6th Conference on Atmospheric Chemistry: Air Quality in Megacities. Seattle, WA, Jan 11-15, 2004.

Delle Monache, L., T. Nipen, X. Deng, Y. Zhou, Yongmei, and R. Stull, 2006: Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction, J. Geophys. Res., Vol. 111, No. D5, D05308 10.1029/2005JD006311

Hogrefe, C. W. Hao, K. Civerolo, J.-Y. Ku, R.S. Gaza, L. Sedefian, G. Sistla, K. Schere, A. Gilliland, and R. Mathur, 2006: Daily photochemical simulations of ozone and fine particulates over New York State: Findings and challenge, *Journal of Applied Meteorology*, in press

Kang, D., B. K. Eder, A. F. Stein, G. A. Grell, S. E. Peckham, and J. McHenry, 2005: The New England air quality forecasting pilot program: development of an evaluation protocol and performance benchmark. J. Air Waste Manag. Assoc., 55, 1782-1796.

Kang, D., R. Mathur, and S.T. Rao, 2006: Application of Bias Correction Techniques to the Eta-CMAQ Air Quality Forecast, presentation at the 5th Annual CMAS Conference, October 16-18, 2006, Chapel Hill, NC

Otte, T.L., G. 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, H.-Y. 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, Weather Forecasting, 20, 367-384

U.S. EPA, 1999: Guideline for reporting of daily air quality – air quality index (AQI), EPA-454/R-99-010, United States Environmental Protection Agency, Research Triangle Park, NC 27711, 29 pp.

Willmott, C.J., 1982. Some Comments on the Evaluation of Model Performance. *Bulletin of the American Meteorological Society*, **63**, 1309-1313.