# APPLICATION OF BIAS ADJUSTMENT TECHNIQUES TO THE ETA-CMAQ AIR QUALITY FORECAST

Daiwen Kang[$], Rohit Mathur[+], and S. Trivikrama Rao[+]

Atmospheric Sciences Modeling Division, Air Resources Laboratory,
National Oceanic and Atmospheric Administration, RTP, NC 27711
[+]On assignment to National Exposure Research Laboratory,
U.S. Environmental Protection Agency, Research Triangle Park, NC 27711
[$] On assignment from Science and Technology Corporation,
10 Basil Sawyer Drive, Hampton, VA 23666-1393
e-mail: kang.daiwen@epa.gov
Voice (919) 541-2460 Fax (919) 541-1379

## 1. INTRODUCTION

Advancement of computing technology and the need for providing accurate ozone health advisories are making it possible to apply numerical models to produce regional air quality forecasts. Numerical simulations depend much on such components as computation schemes, initial and boundary conditions. Imperfect settings of simulation processes (assumptions, approximations, and incompleteness of simulated physical processes) introduce errors into the simulation results. While limited in space and time, observations are the only way for us to know the real conditions of the environment, and, hence, simulation results should be evaluated against observations. Since numerical simulations are based on the solution of time differential equations, i.e, changes of quantities over time, a sound numerical simulation may not reproduce observations, but it should be able to simulate the changes from a given initial state reasonably well. If past and current observations exist, forecast accuracy may be improved by post-processing numerical simulation results with bias adjustment.

The air quality forecast system (AQFS) (Otte et al., 2005), developed by the National Oceanic and Atmospheric Administration (NOAA) in partnership with the US Environmental Protection Agency (EPA), entails coupling the Eta meteorological model (Black 1994; Rogers et al., 1996) with the Community Multiscale Air Quality (CMAQ) (Byun and Schere, 2006) model. The modeling system has been used to provide forecasts of ozone ($O_3$) concentrations since 2004. Comparison of the Eta-CMAQ forecast output and observations over time for $O_3$ revealed that the model has consistently over-predicted, but still simulates the day-to-day variability quite well. This suggests that the forecast results could be improved by combining observations with forecast biases.

Model post-processing techniques were first used in weather forecast, especially for precipitation forecast (Glahn and Lowry 1972; http://www.nws.noaa.gov/mdl/synop/products.shtml). In this study, two bias adjustment techniques (post processing model forecasts) are applied to the Eta-CMAQ $O_3$ and $PM_{25}$ forecasts during the period of July to September, 2005.

In this study, KF is applied to both $O_3$ and $PM_{25}$ forecast. The $O_3$ forecast covers the whole continental US with more than 1000 measurement sites and for the forecast period of 1 July to 30 September 2005, while the $PM_{25}$ forecast covers the eastern US domain with over 300 measurement sites for the 2005 annual forecast.

## 2. THE ETA-CMAQ FORECAST SYSTEM

The Eta model provides the meteorological fields for input to CMAQ (Otte et al., 2005). The processing of the emission data for various pollutant sources has been adapted from the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system (Houyoux et al., 2000) using input from the U.S. EPA national emission inventory. The Carbon Bond chemical mechanism (version 4.2) is used to represent the photochemical reactions. Detailed information on transport and cloud processes in

[*]*Corresponding author:* Daiwen Kang, US EPA, Mail Drop E243-03, NERL/AMD, RTP, NC 27711; e-mail: kang.daiwen@epa.gov

the CMAQ is described in Byun and Schere (2006). For this application, $O_3$ concentrations are forecast over the continental U.S. using 12-km horizontal grid spacing on the Lambert Conformal map projection. There are 22 layers in the vertical, which are set on a sigma coordinate extending from the surface to 100 hPa. The chemical fields for CMAQ are initialized using the previous forecast cycle. The primary Eta-CMAQ model forecast for next-day surface-layer $O_3$ is based on the current day's 12 UTC cycle.

Hourly, near real-time, $O_3$ (ppb) and $PM_{25}$ data obtained from EPA's AIRNow program are used in this study (http://www.epa.gov/airnow).

## 3. BIAS ADJUSTMENT TECHNIQUES

The first technique, namely, the Hybrid Forecast, is simply to combine the current observations with the difference between the forecasts made from current time step and the forecasts made from last time step.

The second bias adjustment technique is called Kalman filter predictor (KFP) (Kalman, 1960) which is a recursive, adaptive method that takes into account the time variation of forecast error at a specific location (Monache et al., 2006). It has been mainly used in data assimilation schemes to improve the accuracy of the initial conditions for both numerical weather prediction (Houtekamer et al., 2005) and air quality forecasts (Segers et al., 2006). It has been used for model forecasts as a predictor bias adjustment method during post-processing of short-term weather forecasts (Roeger et al., 2003) and Monache et al. (2006) extended the application to air quality forecast for $O_3$ forecast at five measurement sites in Canada for a 7-day period of 9-15 August 2004.

### 3.1 Hybrid Forecast (HF)

$$HF_{t+\Delta t} = O_t + (M_{t+\Delta t} - M_t) \qquad (1)$$

Where $O_t$ are observations at time t, $M_{t+\Delta t}$ and $M_t$ are ozone forecasts at time t+Δt and t, respectively.

### 3.2 Kalman Filter Predictor Forecast (KF)

A brief description of KF is given below. The detailed KF implementation was given by Monache et al. (2006).

In the KF forecast, there are two steps. First, the filter estimates the systematic component of the forecast bias using the recent past forecast and observations. Kalman (1960) showed that the optimal recursive predictor of $x_t$ (derived by minimizing the expected mean square error) can be expressed as:

$$x_{t+\Delta t|t} = x_{t|t-\Delta t} + \beta_{t|t-\Delta t}(y_t - x_{t|t-\Delta t}) \qquad (2)$$

Where Δt is a time lag, and t|t – Δt means that the value of the variable at time t depends on values at time t – Δt. β is the weighting factor, called Kalman gain, which is recursively computed using errors associated with forecast bias. Second, after bias $x_{t+\Delta t}$ is calculated, it is combined with the new model forecast to form the KF forecast as:

$$KF_{t+\Delta t} = y_{t+\Delta t} - x_{t+\Delta t|t} \qquad (3)$$

Where $y_{t+\Delta t}$ is the model forecast for next time step.

During this study, a uniform ratio (0.06) of white noise variance to random error variance is assumed for all the sites studied which may vary for location to location, even though this is the optimal value found by Homleid (1995). Further research is needed to apply KF bias adjustment to different locations to find an optimal ratio value for each location.

## 4. BIAS ADJUSTMENT FOR $O_3$ FORECAST

When KF is applied to the daily maximum 8-hr $O_3$ forecast, there are two ways to calculate the maximum 8-hr $O_3$ concentrations. One way is to calculate maximum 8-hr $O_3$ concentrations using the original model forecast time series, then apply KF to the maximum 8-hr $O_3$. In this case, there is only one value at each site per day. The other way is to apply KF to the hourly $O_3$ time series to get a KF adjusted hourly $O_3$ time series, then compute the daily maximum 8-hr $O_3$ from the adjusted hourly time series to do analysis. As shown in Figure 1, there is no significant difference between these two calculation schemes. So, the former scheme is adopted to apply KF filter.

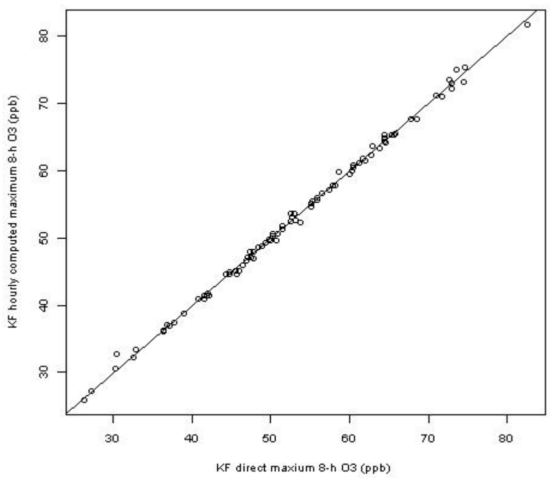


Fig.1. Maximum 8-h $O_3$ calculated from two ways of applying KF adjustment to the model forecasts.

### *4.1 Monthly Mean Errors*

Figure 2 displays the monthly boxplots of root mean square error (RMSE) for the forecast model, HF, KF, and persistence forecast. Persistence is used as a reference in that current observed concentrations are used as future forecast at next time step, in this case, it is next day.

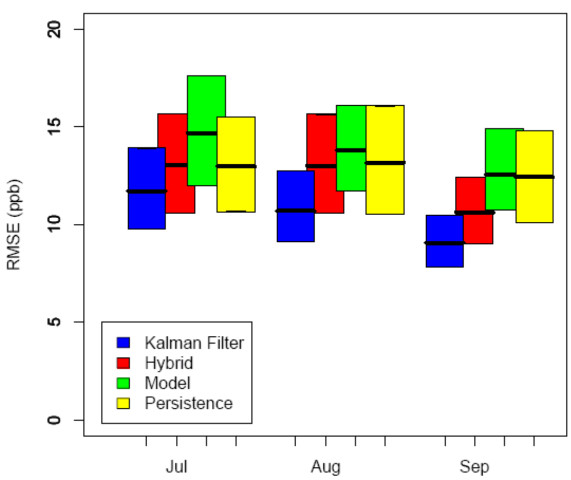


Fig. 2. Boxplots of monthly maximum 8-h $O_3$ of all the sites within the continental US domain: the boxplots show 25 (bottom) and 75 (top) percentiles and the median (cross line).

As shown in Figure 2, KF displays the smallest RMSE followed by HF implying that KF has the largest improvement followed by HF in the forecast results. The original model forecasts present the largest RMSE values, while persistence seems to have some improvement over model forecast, especially in July. But remember that this is the average value and persistence forecasts usually compensate over prediction with under-prediction. Persistence forecast can not respond to dramatic weather changes which often cause big fluctuations in $O_3$ concentrations, while the other forecasts including the original model forecast have a better skill.

### *4.2 Time Series*

Figure 2 displays the time series of observed, model forecast, HF forecast, and KF forecast maximum 8-h $O_3$ for the time period of 1 July – 30 September 2005 at Bushy Fork monitoring site at Raleigh, NC. The model (red dashed line) tends to overpredict almost all the days compared with observations (black solid line). The HF adjusted forecast (green dashed line) tends to bring the forecast closer to observation than the model forecasts, but sometimes it overreacted when dramatic changes occurred (the green dashed over shooting peaks and troughs). The KF adjusted forecasts (blue solid line) are the best among the three forecasts to match observations, even though some over reactions with smaller magnitude still exist with HF if the model forecasts and the observations have dramatic changes over the previous days. Similar trends are found in other monitoring sites.

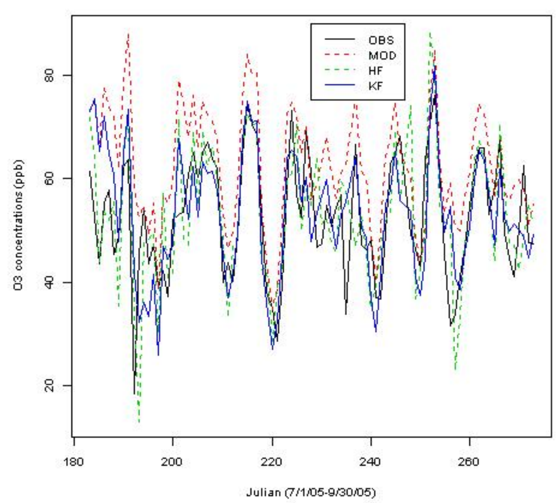

Fig. 3. Time series of observation, model forecast, HF adjusted, and KF adjusted forecast at Busy Fork monitoring site at Raleigh, NC.

### *4.3 Impact on Exceedence Events*

Figure 4, which displays the Hit rate (H) and False alarm ratio (FAR) of the model forecast and adjusted forecasts (KF and HF) for the study period within the continental US domain (Kang et al., 2006), indicates that for the continental US domain, both KF and HF show larger hit rate values than those of the original model forecasts; HF is better than KF to capture the exceedences. FAR values are largest in the original model forecast and smallest in the KF forecasts. However, the trends may vary from region to region if we examine the categorical metrics for sub regions. Generally speaking, the improvement of bias adjusted forecasts over original model forecasts for exceedence events is not as significant as overall performance when evaluated by the discrete metrics.

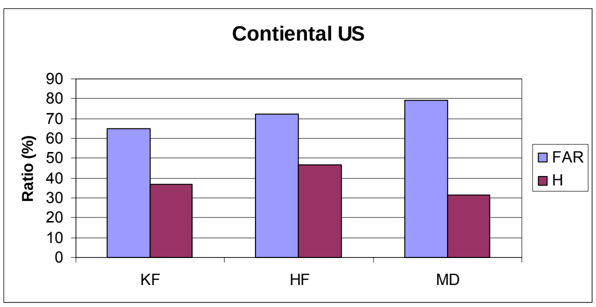


Fig. 4. Hit rate (H) and False alarm Ratio (FAR) of the model, Kalman Filter (KF), Hybrid Forecasts (HF), and model forecasts (MD) for the period of 1 July to 30 September 2005 within the continental US forecast domain.

## 5. BIAS ADJUSTMENT FOR $PM_{25}$ FORECAST

KF and HF bias adjustment are also applied to $PM_{25}$ forecasts. $PM_{25}$ forecast data from the Eta-CMAQ AQF system are available for eastern US domain during the whole year of 2005.

### *5.1 Monthly Mean Errors*

As the boxplots (Fig. 5) indicates, throughout the year, KF forecast always has the smallest RMSE values except in December when persistence has the smallest value. The improvement is more significant during summer months when the model over-predicted $PM_{25}$ concentrations. The original model forecasts or the persistence forecasts have the largest RMSE for majority of the months. Hybrid (HF) forecasts also show improvement over model forecasts for majority of the months, but less significant than the KF forecasts.

### *5.2 Time Series*

Figure 6 displays the time series of the observed (black solid line), model forecast (red dotted line), HF (green dotted line), and KF (blue solid line) forecast daily mean $PM_{25}$ concentrations during the year 2005 at Bryson, NC. The original model under-predicted $PM_{25}$ concentrations at this site for almost all the times with only a few exceptions during the summer. The forecast accuracy was significantly improved with HF and KF bias adjustment. Again, the HF bias adjusted forecasts have more over shooting of the peaks and troughs than the KF adjusted forecasts when dramatic changes occurred over the preceding days.

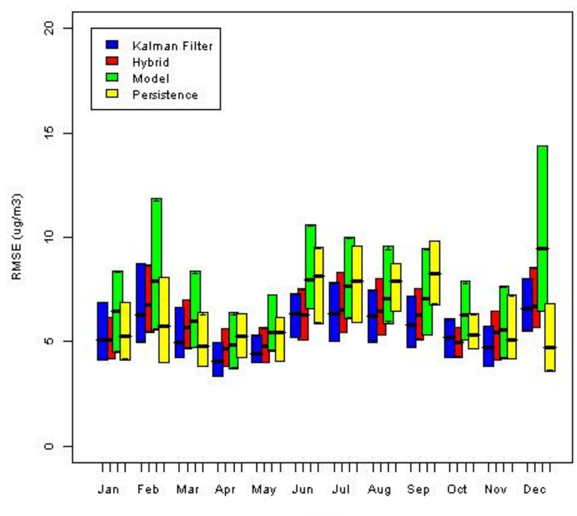


Fig. 5. Monthly RMSE boxplots for $PM_{25}$ for the year 2005 over the eastern US Domain: the boxplots show 25 (bottom) and 75 (top) percentiles and the mean (cross line).

## 6. SUMMARY

The bias adjustment techniques (Kalman filter predictor and Hybrid forecast) have demonstrated their ability to improve forecast accuracy for both $O_3$ and $PM_{25}$ by significantly reducing the root mean square errors. Kalman filter predictor bias adjustment is more powerful than Hybrid forecast, but Hybrid forecast is simpler and easier to implement. The bias adjustment techniques have also displayed some improvement over extreme (exceedence) event forecast for $O_3$, but it varies from region to region.

## 7. Acknowledgements

The authors are grateful to Luca Delle Monache and Roland Stull for providing their original source codes of Kalman Filter calculation. The research presented here was performed under the Memorandum of Understanding between the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) and under agreement number DW13921548. This work constitutes a contribution to the NOAA Air Quality Program. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views.

## 8. REFERENCE

Black, T. 1994: The new NMC meso-scale Eta Model: Description and forecast examples. *Wea. Forecasting*, **9**, 265-278.

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.

Chang, J., et al., 1990: The regional acid deposition model and engineering model, in *National Acid Precipitation Assessment Program, Acidic Deposition, State of Science and Technology*, vol. 1, NAPAP SOS/T Rep. 4, Natl. Acid Precip. Assess. Program, Washington, D.C.

Homleid, M., 1995: Diurnal corrections of short-term surface temperature forecasts using Kalman filter, Weather Forecasting 10, 707-989.

Houtekamer, P.L., H.L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation, Mon. Weather Rev. 133, 604-620.

Houyoux, M.R., J.M. Vukovich, C.J. Coats Jr., N.M. Wheeler, and P.S. Kasibhatla, 2000: Emission inventory development and processing for the seasonal model for regional air quality (SMRAQ) project. *J. Geophys. Res.,* **105**, 9079-9090.

Kalman, R.E., 1960: A new approach to linear filtering and prediction problems, J. of Basic Engineering 82, 35-45.

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, Journal of the Air & Waste Management Association 55, 1782-1796.

Kang, D., R. Mathur, S. Yu, K. Schere, 2006: Performance characteristics of Eta-CMAQ O3 forecast over different regions of the continental United States and application of new categorical evaluation metrics, Preprints, 28th NATO/CCMS International Technical Meeting on Air Pollution and its Application, May 15-19, Leipzig, Germany, pp. 324-321.

Monache, D.L., T. Nipen, X. Deng, Y. Zhou, and R. Stull, 2006: Ozone ensemble forecasts: 2. A Kalman-filter predictor bias correction, Journal of Geophysical Research 111, D05308, doi:10,1029/2005JD006311.

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. Davison, 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, *Wea. Forecasting*, **6**, 367-384.

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

Segers, A.J., H.J. Eskes, R.J. van der A, R.F. van Oss, and P.F.J. van Velthoven, 2005: Assimilation of GOME ozone profiles and a global chemistry-transport model using a Kalman filter with anisotropic covariance, Q.J.R. Meteorol. Soc. 131, 477-502.

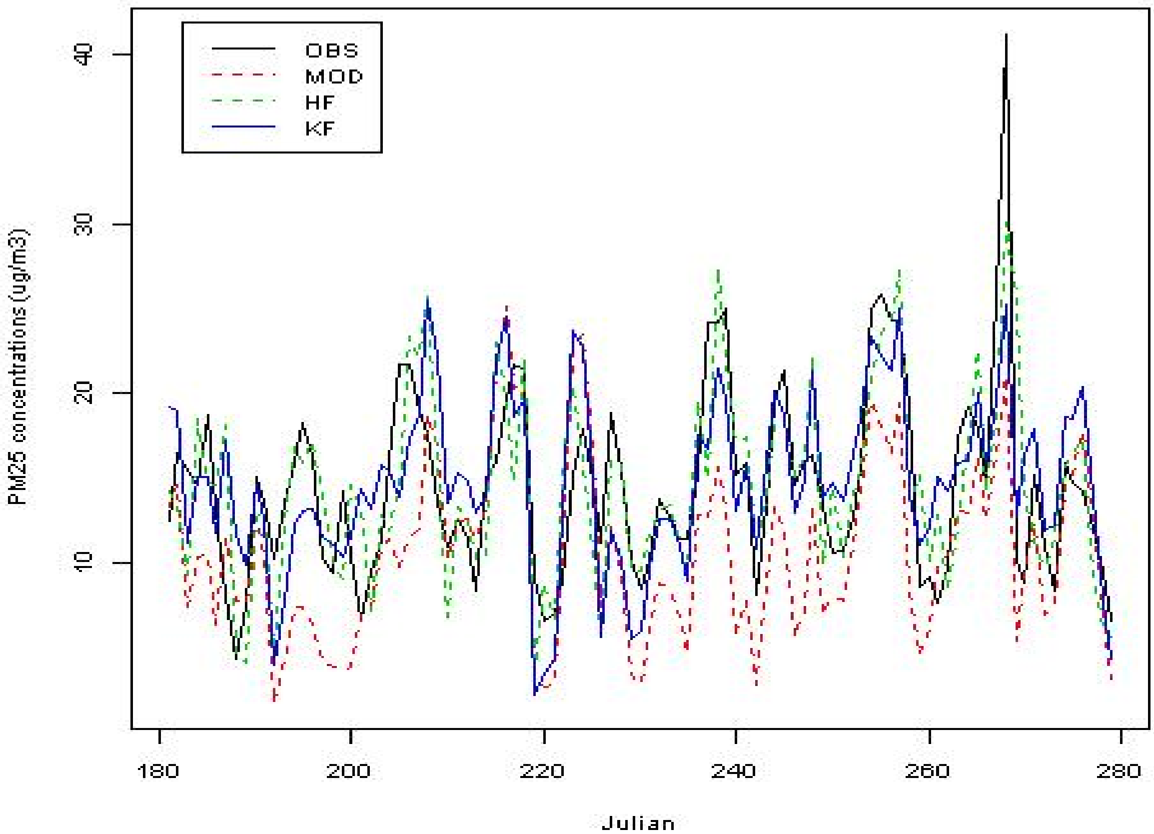


Fig. 6. Time series of observed, model forecast, HF, and KF forecast daily mean $PM_{25}$ concentrations during the year 2005 at Bryson, NC.