REAL-TIME BIAS-ADJUSTED O_3 AND $PM_{2.5}$ AIR QUALITY INDEX FORECASTS AND THEIR PERFORMANCE EVALUATIONS OVER THE CONTINENTAL UNITED STATES

Daiwen Kang[1], Rohit Mathur[2], and S. Trivikrama Rao [2]

[1] Computer Science Corporation, Research Triangle Park, 79 T.W. Alexander Drive, NC 27709, USA
[2]Atmospheric Modeling and Analysis Division, National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

Corresponding Author: Daiwen Kang

Computer Science Corporation
79 T.W. Alexander Drive
Building 4201 Suite 260
Research Triangle Park, NC 27709

(919) 558-8782 Ext. 207
kang.daiwen@epa.gov

December 2009

Revised: March 2010

ABSTRACT

The National Air Quality Forecast Capacity (NAQFC) system, which links NOAA's North American Mesoscale (NAM) meteorological model with EPA's Community Multiscale Air Quality (CMAQ) model, provided operational ozone (O_3) and experimental fine particular matter ($PM_{2.5}$) forecasts over the continental United States (CONUS) during 2008. This paper describes the implementation of a real-time Kalman Filter (KF) bias-adjustment technique to improve the accuracy of O_3 and $PM_{2.5}$ forecasts at discrete monitoring locations. The operational surface level O_3 and $PM_{2.5}$ forecasts from the NAQFC system were post-processed by the KF bias-adjusted technique using near real-time hourly O_3 and $PM_{2.5}$ observations obtained from EPA's AIRNow measurement network. The KF bias-adjusted forecasts were created daily, providing 24-hour hourly bias-adjusted forecasts for O_3 and $PM_{2.5}$ at all AIRNow monitoring sites within the CONUS domain. The bias-adjustment post-processing implemented in this study requires minimal computational cost; requiring less than 10 minutes of CPU on a single processor Linux machine to generate 24-hr hourly bias-adjusted forecasts over the entire CONUS domain.

The results show that the real-time KF bias-adjusted forecasts for both O_3 and $PM_{2.5}$ have performed as well as or even better than the previous studies when the same technique was applied to the historical O_3 and $PM_{2.5}$ time series from archived AQF in earlier years. Compared to the raw forecasts, the KF forecasts displayed significant improvement in the daily maximum 8-hr O_3 and daily mean $PM_{2.5}$ forecasts in terms of both discrete (i.e. reduced errors, increased correlation coefficients, and index of agreement) and categorical (increased hit rate and decreased false alarm ratio) evaluation metrics at almost all locations during the study period in 2008.

Keywords: Air quality index forecast; Bias-adjustment; O_3; $PM_{2.5}$; Kalman filter

1. INTRODUCTION

Ozone (O_3) and fine particulate matter ($PM_{2.5}$ – particles with aerodynamic diameters less than 2.5 μm) pollution is of concern due to their adverse effects on human and ecosystem health. Ambient levels of O_3 and $PM_{2.5}$ are the two primary components used in the calculation of the Air Quality Index (AQI), a standardized indicator of air quality degradation at a given location (Federal Register, 1999). The National Oceanic and Atmospheric Administration (NOAA), in partnership with the United States Environmental Protection Agency (US EPA), has been operationally implementing the National Air Quality Forecasting Capacity (NAQFC) system. This program, which couples NOAA's North American Mesoscale (NAM) weather prediction model with EPA's Community Multiscale Air Quality (CMAQ) model, has provided forecasts of ozone (O_3) mixing ratios since 2004 (Eder et al., 2006; Eder et al., 2010). Developmental $PM_{2.5}$ forecasts were initiated during the summer of 2004 (Mathur et al., 2008; Yu et al., 2008). The modeling domain for both the operational and developmental predictions currently covers the continental United States (CONUS).

Despite continuous refinement and improvement, all numerical models suffer from significant errors and uncertainties due to numerical solvers, emissions inventory, boundary conditions, as well as our incomplete understanding of the physical and chemical processes occurring in the atmosphere. Incorporating recent model forecasts with observations to adjust model forecasts, the bias-adjustment method has been proven to be an effective way to reduce the systematic errors in numerical model outputs (Kang et al., 2008). The implementation of bias-adjustment postprocessing for air quality forecasts relies on the availability of near real-time observations. The U.S EPA's AIRNow measurement network, which reports near real-time hourly O_3 and $PM_{2.5}$

observations nationwide, provides an ideal opportunity to perform bias-adjusted O_3 and $PM_{2.5}$ air quality forecasts for air quality forecast modeling systems.

Bias-adjustment techniques have been used to correct systematic biases in surface O_3 predictions (McKeen et al., 2005; Delle Monache et al., 2006; Wilczak et al., 2006; Delle Monache, et al., 2008; and Kang et al., 2008), and more recently have also been extended to $PM_{2.5}$ forecasts (Kang et al., 2009). Among these techniques, the Kalman Filter (KF) (Kalman, 1960) predictor forecast method has shown the most improvement in forecast skill. However, all previous research efforts on bias-adjustment predictions were performed on retrospective basis, i.e., the bias-adjusted predictions were formulated by using archived model predictions and observations. To test the applicability of the methods in the operational real-time setting during 2008, the KF bias-adjustment technique (Kang et al., 2008, Kang et al., 2009) was implemented, for the first time, in real-time along with the NAQFC system to provide daily bias-adjusted O_3 and $PM_{2.5}$ forecasts at all the locations where observations from EPA's AIRNOW network are available within the CONUS domain. The bias-adjusted O_3 forecasts were performed for the April to mid- September period covering the entire O_3 season while $PM_{2.5}$ bias-adjusted forecasts were conducted throughout the entire calendar year. This paper presents the implementation of the KF bias-adjusted forecasts and its performance evaluation for O_3 and $PM_{2.5}$ forecasts.

2. EXPERIMENTS AND METHODS

2.1 The NAM-CMAQ Air Quality Forecast System

The NAQFC system is based on the National Centers for Environmental Prediction's (NCEP's) NAM meteorological model (Black 1994; Rogers et al., 1996) and EPA's CMAQ (Byun and Schere, 2006) air quality modeling system. A brief summary of the linkage between the NAM and the CMAQ models, relevant to this study, is presented below. A more in-depth description of the NAM-CMAQ system can be found in Otte et al. (2005).

For this application, O_3 and $PM_{2.5}$ were forecast over the CONUS US at 12-km horizontal grids on the Lambert Conformal map projection. The vertical domain was discretized with 22 layers set on the sigma coordinate, extending from the surface to ~100 hPa. The Carbon Bond IV (CB-IV) chemical mechanism was used to represent the gas phase reaction pathways for O_3 forecasts and for the early part of $PM_{2.5}$ forecasts. The chemistry mechanism was updated to the CB05 (Yarwood et al., 2005; Sarwar et al., 2008) for the $PM_{2.5}$ forecasts on July 15, 2008. The AERO3 aerosol module was used with CB-IV model configuration; the module was updated to AERO4 when the chemistry mechanism was updated to CB05. Three-dimensional chemical fields were initialized from the previous forecast cycle. The primary NAM-CMAQ model forecast for next-day surface-layer O_3 was based on the current day's 12 UTC cycle, while for PM2.5 forecasts, the 06 UTC cycle was used. The target forecast period was local midnight through local midnight next day.

The processing of the emission data for various pollutant sources was adapted from the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system (Houyoux et al., 2000). Emission estimates were based on the 2005 U.S. EPA National Emission Inventory. NO_X and SO_2 emitted from elevated point sources were projected to 2008 using the 2006 Continuous Emission Monitoring (CEM) data in conjunction with projections derived from the Department of Energy's Annual Energy Outlook (Pouliot and Pierce, 2003).

2.2 Domain, Evaluation Regions, and Observational Data

As shown in Figure 1, the NAQFC domain covers the CONUS US. Due to large region-to-region differences in the atmospheric physical and chemical processes, the CONUS domain is divided into six subregions to facilitate the performance evaluations (see Figure 1). The four easternmost subregions, northeast (NE), southeast (SE), upper Midwest (UM), and lower Midwest (LM), are based on climatology that identified areas of homogeneous concentration variability using the Principal Component Analysis technique (Eder et al., 1993; Gogo et al., 2005). The Rocky Mountain (RM) subregion is characterized by high elevation (generally > 1000 m) and complex terrain. The Pacific Coast (PC) subregion contains the west coast states which are often under marine influence from the Pacific Ocean.

Hourly, near real-time, surface O_3 (ppb) and $PM_{2.5}$ ($\mu g/m^3$) data obtained from EPA's AIRNow program were used in the KF bias-adjustment forecasts and performance evaluations. Roughly 1000 O_3 (crosses) and 500 $PM_{2.5}$ (circles) routine measurement stations, mostly in urban areas, are available (Figure 1) for the study period. For O_3 forecasts, the daily maximum 8-hr concentrations were calculated at each station for each day over the study period. The running 8-hr average O_3 concentrations were computed using the concentration at the current and succeeding 7 hours; the daily maximum 8-hr O_3 is the maximum of the 8-hr average values over the day. For $PM_{2.5}$ forecasts, the 24-h daily mean at each site was used in the performance evaluations. To facilitate performance evaluations for $PM_{2.5}$, the study period is divided into a cool season (from January to April 20th and from September to December) and a warm season (from April 21st to August 31st).

2.3 Implementation of the KF bias-adjustment method

The KF predictor bias-adjustment algorithm (Kalman, 1960) is described in detail by Delle Monache et al. (2006). The adaptation and implementation of the technique for our applications has been presented by Kang et al., (2008). In that study, a key parameter in the KF approach which determines the relative weighting of observed and forecast values was investigated extensively with O_3 forecasts at over 1000 monitoring locations. Even though the optimal error ratios inherent in the KF algorithm implementation were found to vary across space, the impact of using the optimal values on the resultant bias-adjusted predictions was found to be insignificant when compared with using a reasonable single fixed value of this parameter across all locations within the modeling domain. We further tested the error ratio values in the range 0.01 to 0.10 for the entire domain, and found that the impact on the performance was relatively insignificant when the error ratios were in this range, consistent with results in Kang et al. (2008). In this study, the same single fixed error ratio value of 0.06 was used at all the locations for the real-time bias-adjusted O_3 and $PM_{2.5}$ forecasts.

The KF bias-adjustment technique was implemented for O_3 and $PM_{2.5}$ forecasts separately. First, the KF was initialized with the initial estimates of KF parameters as outlined in Kang et al. (2008) and with two days of hourly observations and raw model predictions. It then generated the third day's bias-adjusted forecasts by combining the third day's raw forecasts with the updated KF parameters. All updated KF parameters at each site for each hour were saved into a file for use in the next KF run. The KF runs then continued by reading the previous day's KF parameters and two preceding days' observations and raw model predictions to continuously generate the next day's bias-adjusted forecasts through combining with the next day's raw forecasts. The KF simulations run daily when the preceding days' observations and the raw forecasts for next day (issued on current day) were available. In our implementation, if data at two consecutive days were missing at a site, the method would automatically drop this site from future bias-adjustment forecasts; however, if a new site with two consecutive days' data appeared in the observation data set, the KF would initialize the site with initial values of KF parameters and generate bias-adjusted forecasts further on. This implementation is very adaptable to the variable nature of monitoring stations reporting hourly observations to the AIRNow network, and can be easily combined with the operational AQF system to provide bias-adjusted forecasts operationally. The bias-adjusted forecasts were initialized on January 4 and April 3 for $PM_{2.5}$ and O_3 forecasts, respectively, and the programs were run daily on a Linux system; it took less than 10 minutes of computation to create a bias adjusted forecast.

2.4 Verification statistics

To assess the performance of the KF bias-adjusted forecasts, model verification statistics commonly used by the air quality modeling community (Kang et al., 2005; Eder et al., 2006; Kang et al., 2008) are used in this study and include Root Mean Square Error (RMSE), Normalize Mean Error (NME), Mean Bias (MB), Normalized Mean Bias (NMB), and correlation coefficient (r). In addition, the index of agreement (IOA) (Willmott, 1981; Kang et al., 2008) is also calculated to specify the degree to which the observed deviations about the mean observed value agree, both in magnitude and sign, to the predicted deviations about the mean observed value. For a forecast product, another set of verification statistics is the categorical metrics (Kang et al., 2005); among those the False Alarm Ratio (FAR) and Hit Rate (H) are used in the current study.

3. RESULTS

3.1 Overall Performance

Table 1 presents a summary of domain (Dom) and sub-regional mean discrete statistics for the raw model and the KF forecast daily maximum 8-h O_3 mixing ratios during the study period. Table 2 presents similar model performance statistics for the daily average $PM_{2.5}$ concentrations for warm and cool seasons. As seen in Table 1, for the daily maximum 8-h O_3 raw forecasts, RMSE values ranged from 10.4 to 16.0 ppb. The application of the KF bias-adjustment reduced the RMSE to the range of 8.5 to 10.5 ppb, reflecting more than a 25% improvement. Similar improvement was reflected in NME. More noticeable improvement by the KF forecasts over the raw model is seen in the MB and NMB; the MB values were reduced from several ppb to about 1 ppb across all the regions, and NMB from as high as 17% to less than 2%. The correlation coefficients (r) also increased systematically from 0.5 to 0.7 range for the raw forecasts to 0.7 - 0.84 range in the KF forecasts. Similar forecast skill improvement in $PM_{2.5}$ forecasts by the KF forecasts over raw forecasts is shown in Table 2. Compared to O_3 forecasts, the overall statistics for $PM_{2.5}$ forecasts still need to be improved due to the difficulty in simulating the complexity of $PM_{2.5}$ formation and distribution by the NAM-CMAQ system.

Figures 2 and 3 present scatter plots of selected forecast and observed percentiles for the daily maximum 8-h O_3 and daily mean $PM_{2.5}$, respectively. In these figures, both measured and forecast time series were examined at each site and percentiles of the concentration distributions over the study period were computed for both observations and forecast values following Mathur et al. (2008). Scatter plots of specific percentiles of the concentration distributions of the modeled and observed time series are then examined to assess the ability of the model to capture the spatial variability in frequency distributions of the species of interest across the sites. As shown in Figures 2 and 3, the KF forecasts displayed a much improved match with the observed distributions as reflected by the reduced and even scatter about the 1:1 line (perfect prediction) when compared to the raw forecasts. The r^2 associated with the forecast and observed percentile distributions increased from 0.80 to 0.98 for the daily maximum 8-h O_3 forecasts, and increased from 0.42 to 0.90 for the daily average $PM_{2.5}$ forecasts, When the KF bias-adjustment procedure was implemented.

The improvement in the performance of the KF bias-adjusted forecasts over the raw forecasts is also evident in the index of agreement (IOA) comparisons (Figures 4 and 5). As seen in Figure 4, for the daily maximum 8-h O_3 forecasts over the entire domain and across all the subregions, the IOA values associated with the KF forecasts increased significantly when compared with those of the raw forecasts. The median IOA values for the raw forecasts were generally less than 0.80, while the median IOA values for the KF forecasts were generally greater than 0.80. For the daily average $PM_{2.5}$ forecasts (Figures 5a and 5b), the KF forecasts again resulted in larger IOA values compared to the raw forecasts for both the warm and cool seasons as well as across all subregions. Comparison of Figure 4 with Figure 5 indicates that the IOA values for both the raw forecasts and KF forecasts for O_3 were larger than those for $PM_{2.5}$; for the raw forecasts, the difference in IOA was about 20%, while for the KF forecasts, the difference was reduced to about 10%. Another important feature is that for O_3 forecasts, both the raw forecasts and the KF forecasts performed better in the eastern portions of the domain (NE, SE, UM, and LM) than in the western regions (RM and PC). However for $PM_{2.5}$ forecasts, the raw forecasts did not perform well in SE for both seasons, though the IOA values significantly increased by the KF forecasts; both the raw forecasts and the KF forecasts displayed lower IOA values for LM and RM during both seasons than for the rest of the regions, while they performed better in PC during the cool season than during the warm season. Both raw forecasts and KF bias-adjusted forecasts displayed the largest

IOA values in NE for both O_3 and $PM_{2.5}$ among all the regions except that the IOA values in UM were larger than those in NE for the KF $PM_{2.5}$ forecasts for the cool season. It should also be pointed out that the performance of KF bias-adjusted forecasts is always dependent on the performance of raw forecasts, i.e., if the IOA values associated with the raw forecasts were lower at a region than at other regions, then the IOA values associated with the KF bias-adjusted forecasts at this region will also be lower than at other regions.

3.2 Temporal and Spatial Performance

Figure 6 presents time-series comparisons of mean daily maximum 8-h O_3 mixing ratios forecast by the raw and bias-adjusted models with corresponding measurements. Figure 7 presents a similar comparison for the daily-average $PM_{2.5}$ forecasts. As seen in Figure 6, the NAM-CMAQ system underestimated the daily maximum 8-h O_3 concentrations at the beginning of the study period, then transitioned to overestimation with time; significant overestimation occurred towards the end of the study period. However, the KF forecasts were able to correct for both overestimation and underestimation and track the observed time series quite well. As noted in Figure 7, the raw model overestimated daily mean $PM_{2.5}$ concentrations during cool season and underestimated during warm season. Again the KF time series tracked the observed time series very well throughout the entire year and was able to reduce the systematic seasonal biases considerably.

To further investigate the temporal and spatial performance, boxplots of monthly RMSE values for daily maximum 8-h O_3 and daily mean $PM_{2.5}$ for each of the subregions are displayed in Figures 8 and 9, respectively. As illustrated in Figure 8, the RMSE values associated with the raw model daily maximum 8-h O_3 forecasts for the SE, UM, and RM subdomains exhibited a tendency to increase as the O_3 season progressed;

similar trends in the NAM-CMAQ O_3 forecast error were also noted for prior years (Eder et al., 2009). These trends were found to be partially related to trends in temperature forecast error in the meteorological model (NAM); the subsequent impacts on modeled emissions and chemistry are currently under investigation. Despite the systematic trend for RMSE values associated with the raw forecasts, the KF bias-adjusted forecasts were able to effectively adjust the errors and produced comparable distributions of RMSE values during the entire period.

For the year-long daily mean $PM_{2.5}$ forecasts (Figure 9), the monthly RMSE values associated with the raw forecasts started higher at the beginning months (January, February, and March). The significant decrease of RMSE values associated with the raw forecasts during the October to December period compared to the January to March period is attributable to the switch in the chemical mechanism from CB-IV to CB05 and the corresponding aerosol module from AERO3 to AERO4 on July 15, 2008. The significantly higher RMSE values during June and July in the PC region can be attributed to missing emissions from wide spread wild fires in California during these two months which resulted in elevated observed $PM_{2.5}$ concentrations which were not simulated by the raw model. Nevertheless, the KF bias-adjusted forecasts were able to produce significantly smaller RMSE values compared to the raw forecasts for all the regions during each of the months.

The ability of the KF technique to improve O_3 and $PM_{2.5}$ forecast across the continental U.S. is further illustrated in Figures 10 and 11 which compare maps of mean bias in the raw and bias-adjusted forecasts. As seen in Figure 10a, the NAM-CMAQ system generally overestimated in the eastern part of the domain, especially in the northeast and southeast with MB values greater than 5 ppb. In California, the MB values indicated mixed results with both overestimation (MB $\geq$ 5 ppb) and underestimation (MB

< -5 ppb) coexisting in the same area. However, with the application of KF bias-adjustment (Figure 10b), the MB values at almost all the locations were reduced to be within ±2 ppb, demonstrating the robustness of the KF bias-adjustment technique for O_3 forecasts across all locations.

Similar effects are also demonstrated for $PM_{2.5}$ forecasts over the CONUS domain (Figure 11). During warm season, underestimation of the daily average $PM_{2.5}$ concentrations by the raw forecasts dominated the entire domain (orange and purple squares in Figure 11a). During the cool season (Figure 11c), the raw forecasts generally overestimated in the east, and displayed mixed results in the west. However, during both warm and cool seasons, the KF forecasts were able to adjust either the overestimation or underestimation concentrations very effectively with mean bias of ±2 $\mu g/m^3$ at majority of the sites (Figures 11b and 11d). Even at the sites where MB values were greater than 2 $\mu g/m^3$ or less than -2 $\mu g/m^3$, the magnitude of the MB values was significantly reduced in the KF forecasts compared with those in the raw forecasts.

3.3 Performance over observation concentration bins

To examine the performance of the KF bias-adjustment technique over different concentration ranges, RMSE and MB for both O_3 and $PM_{2.5}$ forecasts were examined as a function of observed ambient levels. As seen in Figure 12a, the RMSE values for daily maximum 8-h O_3 forecasts were larger at both lower (<30 ppb) and higher (≥85 ppb) O_3 levels than those in the middle. For the $PM_{2.5}$ forecasts, the raw forecasts displayed lower RMSE values at lower observation bins and higher RMSE values at higher observation bins for both the warm (Figure 13a) and cool seasons (Figure 13b). Compared to the O_3 forecasts (Figure 12b), the distribution of MB values for $PM_{2.5}$ forecasts (Figures 13 c and d) over concentration bins displayed very different features; during the warm season, the distribution of the MB values associated with the raw forecasts showed very little variations, while when the observed concentrations were greater than 10 μg/m^3, the raw model displayed increased underestimation with the increasing concentrations. In contrast, during the cool season (Figure 13d), the $PM_{2.5}$ MB values associated with the raw model showed very little variation, even though the distributions became increasingly wider at higher observation bins. The KF bias-adjustment technique is able to effectively reduce the errors and biases across all concentration ranges and for both the warm and cool seasons.

3.4 Categorical Performance

It is important for an air quality forecast product to be able to accurately predict exceedance and non-exceedance events (categorical predictions). Figure 14 presents the categorical evaluations for the raw forecasts and KF bias-adjusted forecasts for daily maximum 8-hr O_3 and daily mean $PM_{2.5}$ concentrations, respectively. The statistical measures presented include the FAR and H (Kang et al., 2005). Exceedance threshold of both the 85 ppb 8-hr maximum O_3 and the revised NAAQS of 75 ppb are examined; the corresponding metrics are denoted as FAR85, FAR75, H85, and H75. The threshold value for daily mean $PM_{2.5}$ exceedance is 35 μg/m^3 and the corresponding metrics are denoted as FAR35 and H35.

As shown in Figure 14, the KF bias-adjusted forecasts were able to significantly reduce FAR values and increase H values for both daily maximum 8-h O_3 forecasts and daily average $PM_{2.5}$ forecasts. Comparison of the categorical metrics for the two threshold values for O_3 forecasts indicates that for the new standard of 75 ppb, both the raw forecasts and the KF forecasts provide better categorical forecasts relative to those with the old standard. The KF forecasts produced an H value of 50% based on the new exceedance standard and the FAR was slightly higher than 50%, further illustrating the robustness of the KF bias-adjustment technique. For the $PM_{2.5}$ forecasts, the FAR reduced from 93% to 76% and H increased from 24% to 38% through the application of the KF bias-adjustment technique.

3.5 Performance for Air Quality Index

The air quality index (AQI) is frequently used to report daily air quality conditions. The index is an indicator of how clean or polluted the air is and what the associated health effects might be for sensitive populations. The breakpoints for converting from O_3 mixing ratio (ppb) or $PM_{2.5}$ concentrations (μg/m^3) to AQI values are presented in Table 3. As seen in Table 3, AQI values range from 0-500, with higher values representing greater level of air pollution and a greater associated health concern; an AQI value of 100 generally corresponds to the National Ambient Air Quality Standard (NAAQS) for the pollutant. The AQI is divided into six color-coded categories; values of 0-50 (code green) represent good air quality conditions, 51-100 (code yellow) represent moderate pollution, 101-150 (code orange) represent air pollution levels unhealthy for sensitive groups, 151-200 (code red) represent unhealthy conditions, while 201-300 (code purple) and 301-500 (code maroon) represent very unhealthy and hazardous air quality conditions, respectively. The ability of the bias correction technique to improve the AQI forecasts for O_3 and $PM_{2.5}$ at each of the monitoring locations was examined. Figure 15 presents comparisons of the category hit rate for each AQI category across all monitoring locations for O_3 forecasts with the raw model and the KF bias-adjusted forecasts. The Category Hit Rate (cH) (Eder et al., 2009b) is defined as: $cH_i = \frac{N_f^i}{N_{obs}^i}$, where i is the AQI index (1, 2, 3, 4, 5), and N_f^i is the number of correctly forecast instances in the ith category and N_{obs}^i is the number of observed instances in the ith category. Figure 16 presents similar comparisons for surface-level $PM_{2.5}$ forecasts. For forecasts of both pollutants, a systematic improvement in the predictions of the different AQI categories is evident when the bias-adjustment technique was applied. The improvements in the accuracy of the AQI forecasts for the moderate to unhealthy categories, further demonstrate the applicability of the methodology and suggest the need to adopt bias-adjustment operationally for improving the reliability of model-based air quality forecasts.

4. SUMMARY

The near real-time Kalman filter bias-adjustment technique was applied to NAM-CMAQ derived O_3 and $PM_{2.5}$ air quality forecasts for the continental United States. These bias-adjustment forecasts were implemented to run daily for improving the next-day forecast. Bias-adjustment on operational basis adds minimal computational burden; on a daily-basis, it required less than 10 minutes of CPU on a single processor Linux machine. Hourly O_3 and $PM_{2.5}$ bias-adjusted forecasts have been generated for all the locations where the observations are available from the AIRNow network. The performance evaluation of the KF forecasts for both O_3 and $PM_{2.5}$ has shown significant improvement over the raw forecasts for a variety of statistical measures. Specifically, systematic errors or biases have been reduced, correlation coefficients increased, false alarm ratios reduced, while hit rates have gone up. The robustness of this bias-adjustment technique is evident for various concentration ranges over CONUS. The forecast skill has improved at all the locations within the domain during all seasons. Though the bias-adjustment technique was only applied at discrete points in this study, the bias-adjusted spatial maps of O_3 and $PM_{2.5}$ forecasts could be readily developed by using appropriate statistical methods (e.g., Hogrefe et al., 2009; Denby et al., 2009; Garcia et al., 2010). Comparison of model forecasts skills for $PM_{2.5}$ and O_3 have clearly indicated that more work needs to be done to improve the accuracy of $PM_{2.5}$ forecasts. Improvements in the representation of fine particulate matter emissions as well as the physical and chemical processes regulating sources and sinks in atmospheric models are expected as a result of on-going research over the next several years. Nevertheless, our analysis indicates that despite the current limitations in the representation of atmospheric processes dictating the distribution of ambient $PM_{2.5}$, bias-adjustment techniques can be used to help improve the accuracy and reliability of short-term $PM_{2.5}$ forecasts and AQI from such models.

5. ACKNOWLEDGEMENTS

The authors thank Drs. Luca Delle Monache and Roland B. Stull for providing their original Kalman filter codes. 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. Although this work has been reviewed by EPA and approved for publication, it does not necessarily reflect agency policy or views.

6. REFERENCES

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

Byun, D.W., Schere, K.L., 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.

Delle Monache, L., Nipen, T., Deng, X., Zhou, Y., Stull, R., 2006. Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction, *J. Geophys. Res., 111,* D05308, doi:10.1029/2005JD006311.

Delle Monache, L., Wilczak, J., Mckeen, S., Grell, G., Pagowski, M., Peckham, S., Stull, R., Mchenry, J., McQueen, J., 2008. A Kalman-filter bias correction method applied to deterministic, ensemble averaged, and probabilistic forecasts of surface ozone, *Tellus, Ser. B, 60*, 238-249, doi:10.111/j.1600-0889.2007.00332.x.

Denby, B., Garcia, V., Holland, D., Hogrefe, C., 2009. Integrating of air quality modeling and monitoring data for enhanced health exposure assessment, EM, October, 46-49.

Eder, B., Davis, J., Bloomfield, P., 1993. A characterization of the spatiotemporal variation of non-urban ozone in the eastern United States, *Atmos. Environ. 27A*: 2645-2668.

Eder, B., Kang, D., Mathur, R., Yu, S., Schere, K., 2006. An operational evaluation of the Eta-CMAQ air quality forecast model, *Atmospheric Environment*, *40*, 4894-4905.

Eder, B., Kang, D., Mathur, R., Pleim, J., Yu, S., Otte, T., Pouliot, G., 2009. A performance evaluation of the national air quality forecast capability for the summer of 2007, *Atmos. Environ.*, doi:10.1016/j.atmosenv.2009.01.033.

Eder, B., Kang, D., Rao, S.T., Mathur, R., Yu, S., Otte, T., Schere, K., Wayland, R., Jackson, S., Davidson, P., McQueen, J., 2010. A demonstration of the use of national air quality forecasts for developing local air quality index forecasts, *Bull. Amer. Meteorol. Soc.*, in press.

Federal Register, 1999. Air quality reporting, Final Rule, 40 CFR Part 58, August 4, 1999, *Fed. Regist*. 64.

Garcia, V. C., Foley, K. M., Gego, E., Holland, D. M., Rao, S. T., 2009. A comparison of statistical techniques for combining modeled and observed concentrations to create high-resolution ozone air quality surfaces, *J. Air Waste Manage. Assoc.*, in review.

Gego, E. L., Porter, P. S., Irwin, J. S., Hogrefe, C., Rao, S. T., 2005. Assessing the comparability of ammonium, nitrate and sulfate concentrations measured by three air quality monitoring networks, Pure Appl. Geophys., 162, 1919-1939, doi 10.1007/s00024-005-2698-3.

Houyoux, M.R., Vukovich, J.M., Coats Jr., C.J., Wheeler, N.M., Kasibhatla, P.S., 2000. Emission inventory development and processing for the Seasonal Model for Regional Air Quality (SMRAQ) project. *J. Geophys. Res., 105*, 9079-9090.

Hogrefe, C., Lynn, B., Goldbery, R., Rosenzweig, C., Zalewsky, E., Hao, W., Doraiswamy, P., Civerolo, K., Ku, J. Y., Sistla, G., Kinney, P. L., 2009. A combined model-observation approach to estimate historic gridded fields of PM2.5 mass and species concentrations, *Atmospheric Environment*, 43, 2561-2570, doi:10.1016/j.atmosenv.2009.02.031.

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

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

Kang, D., Mathur, R., Rao, S.T., Yu, S., 2008. Bias adjustment techniques for improving ozone air quality forecasts, *J. Geophys. Res., 113*, D23308, doi: 10.1029/2008JD010151.

Kang, D., Mathur, R., Rao, S.T., 2009. Assessment on bias-adjusted $PM_{2.5}$ air quality forecasts over the continental United States during 2007, *Geosci. Model Dev. Discuss.,* 2, 1375-1406.

Mathur, R., Yu, S., Kang, D., Schere, K.L., 2008. Assessment of the wintertime performance of developmental particulate matter forecasts with the Eta-Community Multiscale Air Quality modeling system, *J. Geophys. Res., 113*, D02303, doi:10.1029/2007JD008580.

McKeen, S., and coauthors, 2005. Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004, *J. Geophys. Res., 110*, D21307, doi:10.1029/2005JD005858.

Otte, T.L., and coauthors, 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.

Pouliot, G., Pierce, T. C., 2003. Emission processing for an air quality forecasting model, Preprints, *12th Int. Emission Inventory Conf. on Emission Inventories – Applying New Technologies*, Atlanta, GA, U.S. Environmental Protection Agency, J85-86.

Rogers, E., Black, T. L., Deaven, D. G., DiMego, G. J., Zhao, O., Baldwin, M., Junker, N. W., Lin, Y.,1996. Changes to the operational "early" Eta Analysis/Forecast System at the National Centers for Environmental Prediction. *Wea. Forecasting*, *11*, 391-413.

Sarwar, G., Luecken, D., 2008. Impact of an updated carbon bond mechanism on predictions from the CMAQ modeling system: preliminary assessment, *J. Appl. Meteorol. Climatol.*, 47, 3-14, doi: 10.1175/2007JAMC1393.1.

Wilczak, J., McKeen, S., Djalalova, I., Grell, G., Peckham, S., Gong, W., Bouchet, V., Moffet, R., McHenry, J., Lee, P., Tang, Y., Carmichael, G. R., 2006. Bias-corrected ensemble and probabilistic forecasts of surface ozone over eastern North America during the summer of 2004, *J. Geophys. Res.*, *111*, D23S28, doi:10.1029/2006JD007598.

Willmott, C. J., 1981. On the validation of models, *Phys. Geogr.*, 2. 184-194.

Yarwood, G., Rao, S., Yocke, M., Whitten, G., 2005. Updates to the carbon bond chemical mechanism: CB05. Final report to the U.S. EPA, RT-0400675. [Available online at http://www.camx.com.]

Yu, S., Mathur, R., Schere, K., Kang, D., Pleim, J., Young, J., Tong, D., Pouliot, G., Mckeen, S., Rao, S.T., 2008. Evaluation of real-time PM2.5 forecasts and process analysis for PM2.5 formation over the eastern United States using Eta-CMAQ forecast model during the 2004 ICARTT study, *J. Geophys. Res., 113*, D06204, doi:10.1029/2007JD009226.

List of Tables

Table 1. Regional summary of discrete statistics for raw model and KF bias-adjusted daily maximum 8-hr O_3 forecasts during 2008 summer season. RMSE: Root Mean Square Error, NME: Normalized Mean Error, MB: Mean Bias, NMB: Normalized Mean Bias, and r: Correlation Coefficient

Table 2. Regional summary of discrete statistics for raw model and KF bias-adjusted daily mean $PM_{2.5}$ forecasts during 2008 warm/cool season: In each cell, the value on the left of slash (/) is for warm season and the value on the right of the slash is for the cool season. The values in the rows of each table with white background marked with "–mod" represent statistics associated with raw forecasts, while those in the rows with shaded background and with the extension "–kf" represent the statistics associated with the KF bias-adjusted forecasts.

Table 3. Air quality index categories with their O_3 and $PM_{2.5}$ concentrations breakpoints

List of Figures

Figure 1. Model domain, evaluation regions, and observational O3 sites (crosses) and PM2.5 sites (cirles): The regions are Northeast (NE), Southeast (SE), Upper Midwest (UM), Lower Midwest (LM), Rocky Mountains (RM), and Pacific Coast (PC)

Figure 2. Scatterplots between forecasts and observations for selected percentiles for the daily maximum 8-h O_3 mixing ratios (ppb)

Figure 3. Scatterplots between forecasts and observations for selected percentiles for daily mean $PM_{2.5}$ concentrations ($\mu g/m^3$)

Figure 4. Box plots of index of agreement (IOA) of daily maximum 8-h O_3 (ppb) for the raw model (MOD) forecasts and KF bias-adjusted forecasts over the domain (DM) and across all subregions

Figure 5. Box plots of index of agreement (IOA) of daily mean $PM_{2.5}$ ($\mu g/m^3$) for the raw model (MOD) forecasts and KF bias-adjusted forecasts over the domain (DM) and across all subregions during (a) warm season and (b) cool season

Figure 6. Time series of observed, raw model forecast, and KF bias-adjusted forecast mean daily maximum 8-h O_3 (ppb) over the domain

Figure 7. Time series of observed, raw model forecast, and KF bias-adjusted forecast average daily 24-h mean $PM_{2.5}$ ($\mu g/m^3$) over the domain

Figure 8. Monthly box plots (only 25th and 75th percentiles and median values are shown) of RMSE values of the daily maximum 8-h O_3 (ppb) for the raw model and KF bias-adjusted forecasts for all sub-regions

Figure 9. Monthly box plots (only 25th and 75th percentiles and median values are shown) of RMSE values of the daily mean $PM_{2.5}$ ($\mu g/m^3$) for the raw model and KF bias-adjusted forecasts for all sub-regions

Figure 10. Mean Bias (MB, ppb) for daily maximum 8-h O_3 forecasts at each location within the continental U.S. domain: (a) raw model, (b) KF bias-adjustment

Figure 11. Mean Bias (MB, $\mu g/m^3$) for daily mean $PM_{2.5}$ forecasts at each location within the continental U.S. domain: (a) raw model during warm season, (b) KF bias-adjustment during warm season, (c) raw model during cool season, (d) KF bias-adjustment during cool season

Figure 12. RMSE and MB values over observed daily maximum 8-h O_3 mixing ratio bins for the raw forecasts and Kalman filter bias-adjusted forecasts over the domain: (a) RMSE and (b) MB

Figure 13. RMSE and MB values over observed daily mean $PM_{2.5}$ concentration bins for the raw forecasts and Kalman filter bias-adjusted forecasts over the domain: (a) RMSE during warm season, (b) RMSE during cool season, (c) MB during warm season, and (d) MB during cool season

Figure 14. FAR and H values for both raw model and KF forecasts: a. daily maximum 8-h O_3 (ppb), and b. daily mean $PM_{2.5}$ ($\mu g/m^3$)

Figure 15. Categorical Hit Rates for each AQI Category for daily maximum 8-hr O_3, the AQI indexes (1, 2, 3, 4) are corresponding to the categories (Good, Moderate, Unhealthy for sensitive groups, and Unhealthy). Note that indexes greater than 4 (the very Unhealthy and Hazardous categories) didn't occur in the data.

Figure 16. Categorical Hit Rates for each AQI Category for daily mean $PM_{2.5}$

Table 1. Regional summary of discrete statistics for raw model and KF bias-adjusted daily maximum 8-hr O_3 forecasts during 2008 summer season. RMSE: Root Mean Square Error, NME: Normalized Mean Error, MB: Mean Bias, NMB: Normalized Mean Bias, and r: Correlation Coefficient

TYPE	RMSE (ppb)	NME (%)	MB (ppb)	NMB (%)	r
Dom-mod	12.5	20.1	3.2	6.8	0.65
Dom-kf	9.1	14.5	0.6	1.3	0.81
NE-mod	10.6	16.9	2.7	5.6	0.70
NE-kf	8.9	13.8	0.7	1.4	0.78
SE-mod	12.2	20.1	5.8	12.2	0.70
SEkf	9.1	14.7	0.5	1.1	0.80
UM-mod	10.4	17.5	2.5	5.4	0.59
UM-kf	8.5	13.7	0.7	1.5	0.72
LM-mod	13.6	27.0	7.0	16.9	0.64
LM-kf	9.8	17.7	0.8	1.9	0.77
RM-mod	11.4	16.4	2.7	5.1	0.50
RM-kf	8.9	12.8	0.7	1.3	0.70
PC-mod	16.0	21.9	-3.2	-5.9	0.60
PC-kf	10.5	14.5	0.2	0.3	0.84

Table 2. Regional summary of discrete statistics for raw model and KF bias-adjusted daily mean $PM_{2.5}$ forecasts during 2008 warm/cool season: In each cell, the value on the left of slash (/) is for warm season and the value on the right of the slash is for the cool season. The values in the rows of each table with white background marked with "–mod" represent statistics associated with raw forecasts, while those in the rows with shaded background and with the extension "–kf" represent the statistics associated with the KF bias-adjusted forecasts.

TYPE	RMSE ($\mu g/m^3$)	NME (%)	MB ($\mu g/m^3$)	NMB (%)	r
Dom-mod	9.6/10.5	47.5/70.5	-2.3/4.5	-19.6/45.1	0.33/0.53
Dom-kf	6.6/6.4	32.9/42.5	-0.1/1.7	-0.4/16.5	0.71/0.68
NE-mod	7.5/12.3	39.5/76.1	-2.4/6.6	-17.8/59.9	0.56/0.63
NE-kf	5.5/7.3	29.1/44.7	-0.4/2.4	-2.7/22.1	0.76/0.72
SE-mod	7.8/9.1	41.5/62.1	-3.9/4.6	-27.5/43.8	0.40/0.47
SE-kf	5.3/5.4	27.1/37.2	-0.4/1.3	-2.7/12.8	0.63/0.58
UM-mod	6.0/10.7	36.6/68.3	-0.7/6.5	-6.0/57.4	0.58/0.62
UM-kf	5.0/6.1	30.7/37.3	-0.2/1.7	-1.7/15.2	0.69/0.73
LM-mod	8.7/9.4	52.4/67.7	-4.0/3.6	-32.9/36.8	0.17/0.32
LM-kf	5.8/5.9	34.9/42.5	-0.2/1.2	-1.5/12.2	0.37/0.49
RM-mod	6.4/9.3	50.5/75.7	-1.5/3.5	-17.2/43.1	0.18/0.37
RM-kf	4.6/5.6	33.5/44.4	0.0/1.3	0.2/16.2	0.57/0.62
PC-mod	15.3/10.2	57.9/60.2	-3.4/1.8	-30.6/15.8	0.23/0.53
PC-kf	10.5/7.0	39.0/40.9	0.2/1.2	1.9/10.4	0.73/0.72

Table 3. Air quality index categories with their O_3 and $PM_{2.5}$ concentrations breakpoints

AQI Category	Index values	Daily maximum 8-hr O_3 (ppb)	24-hr $PM_{2.5}$ ($\mu g/m^3$)
Good	0 - 50	0 - 59	0 - 15
Moderate	51 - 100	60 - 75	16 - 35
Unhealthy for Sensitive Groups	101 - 150	76 - 95	36 - 55
Unhealthy	151 - 200	96 - 115	56 - 140
Very Unhealthy	201 - 300	116 - 375	141 - 210

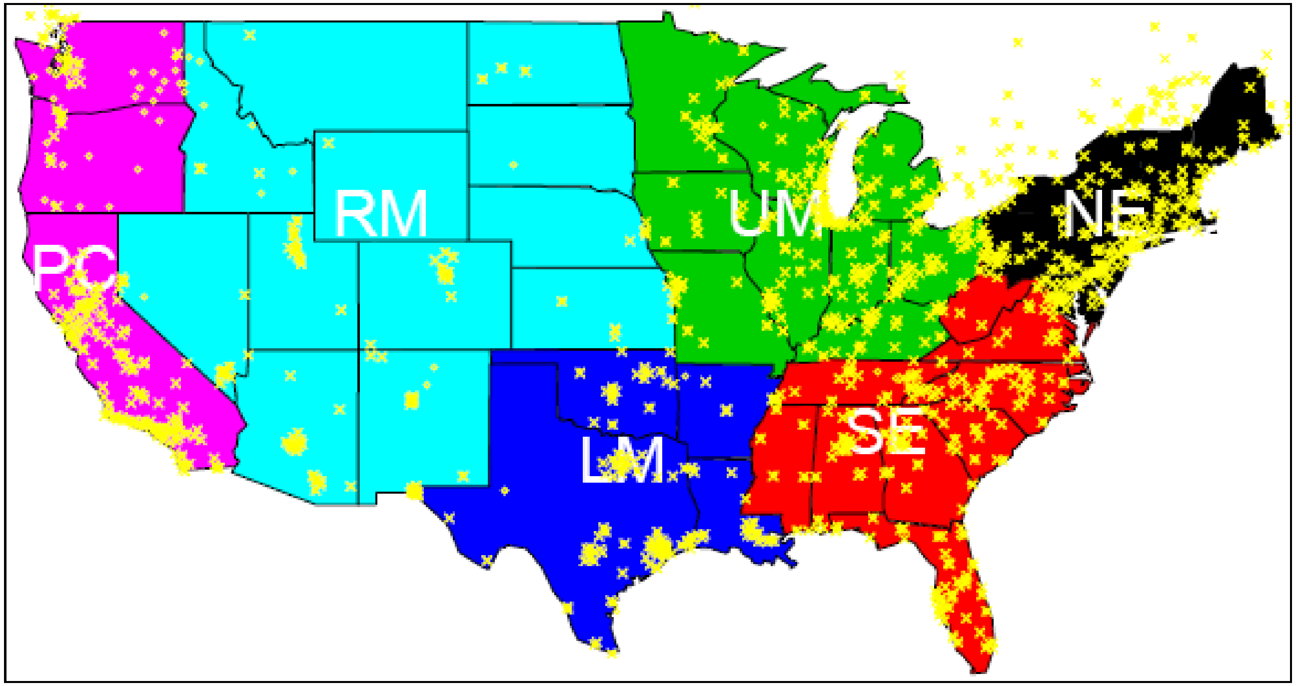

Figure 1. Model domain, evaluation regions, and observational O_3 sites (crosses) and $PM_{2.5}$ sites (cirles): The regions are Northeast (NE), Southeast (SE), Upper Midwest (UM), Lower Midwest (LM), Rocky Mountains (RM), and Pacific Coast (PC)

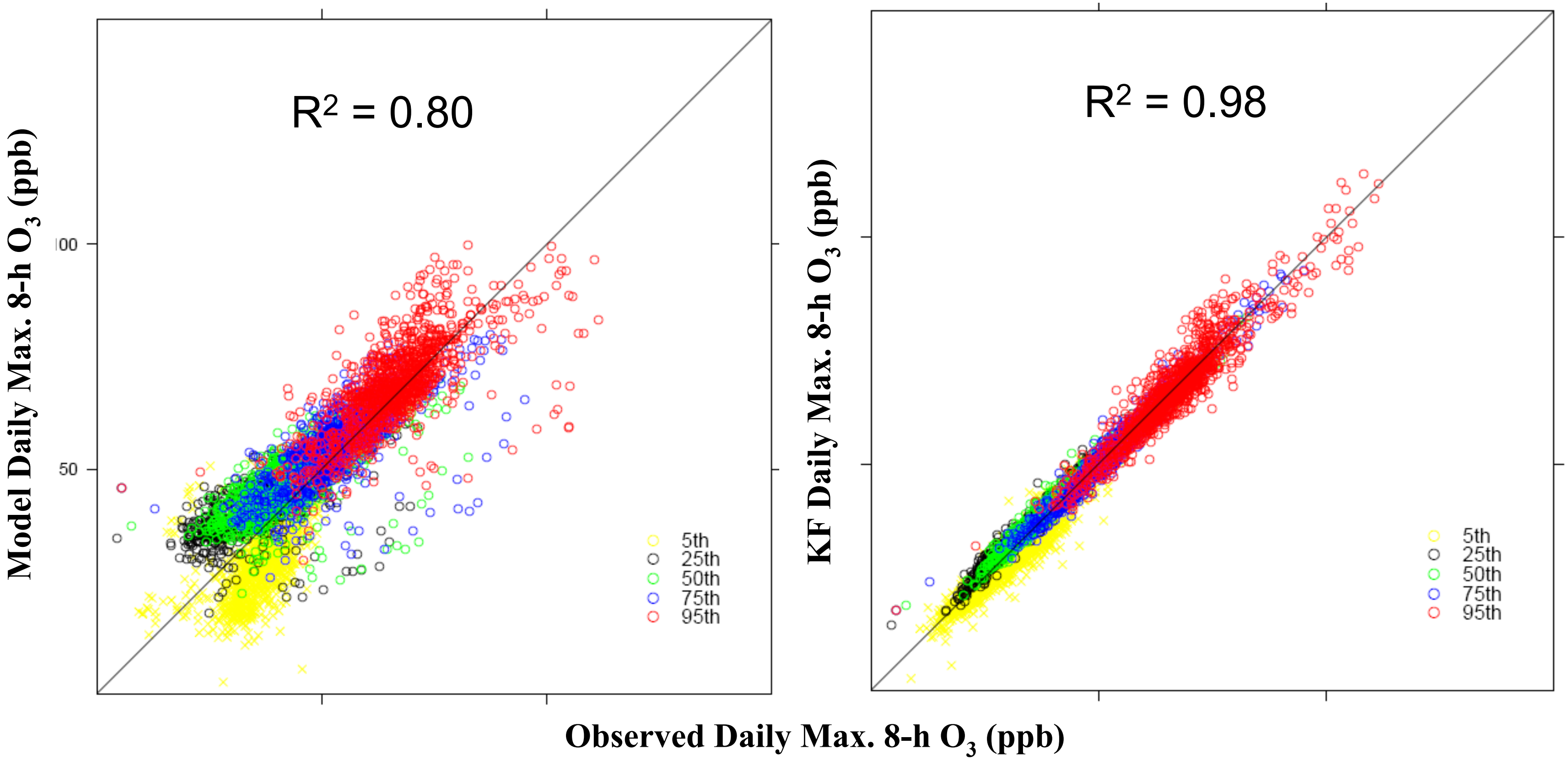

Figure 2. Scatterplots between forecasts and observations for selected percentiles for daily maximum 8-h O_3 mixing ratios (ppb).

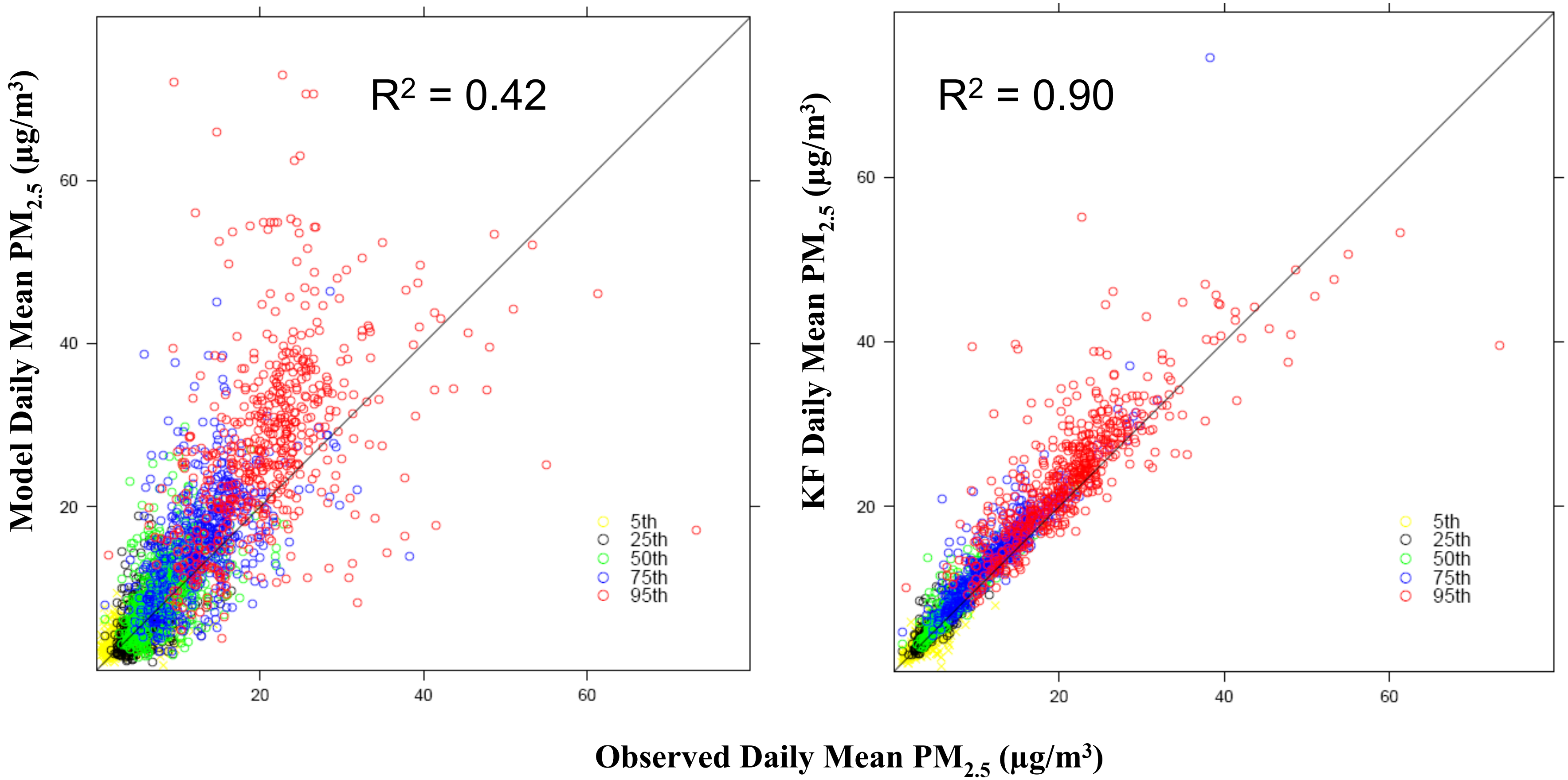

Figure 3. Scatterplots between forecasts and observations for selected percentiles for daily mean $PM_{2.5}$ concentrations (μg/m^3).

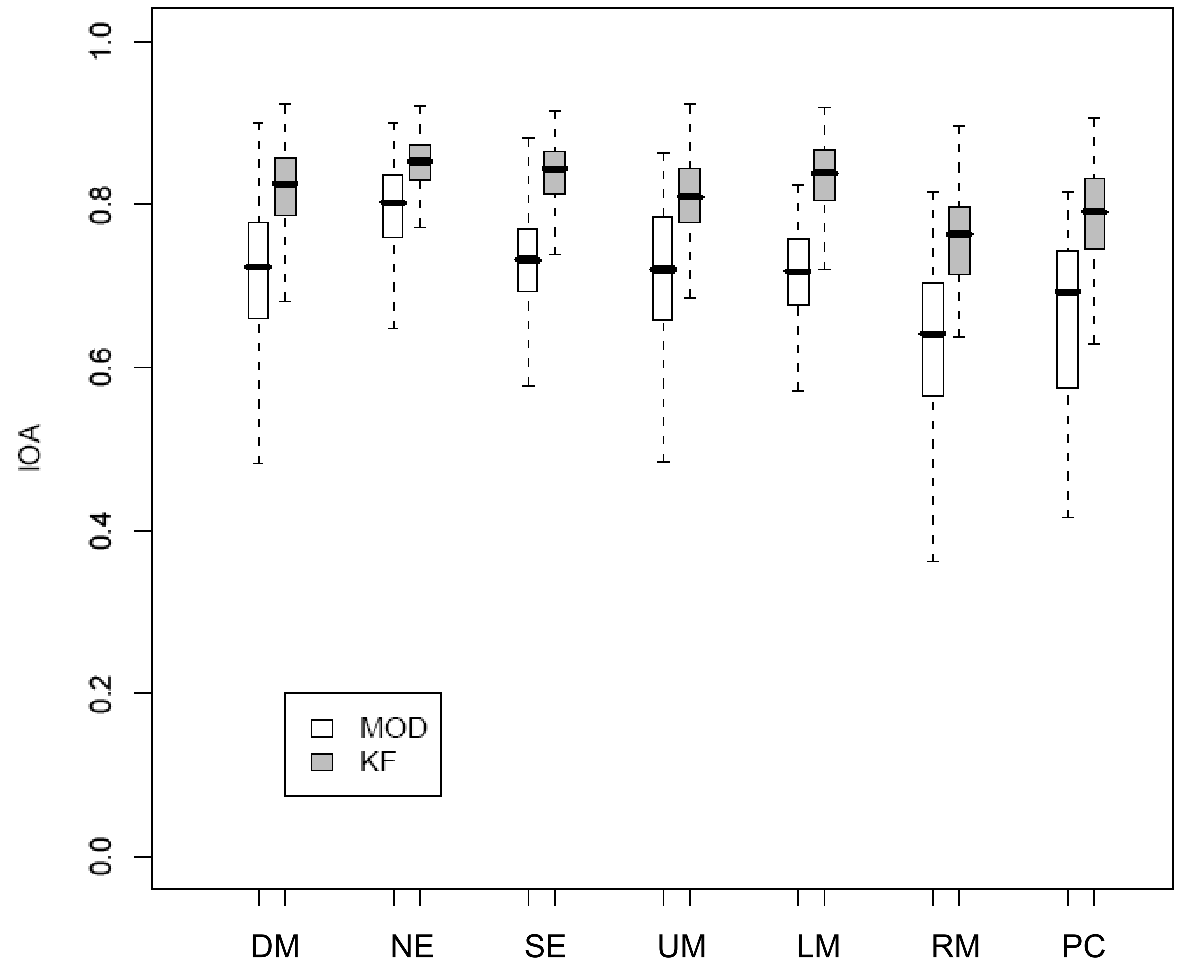

Figure 4. Box plots of index of agreement (IOA) of daily maximum 8-h O_3 (ppb) for the raw model (MOD) forecasts and KF bias-adjusted forecasts over the domain (DM) and across all subregions

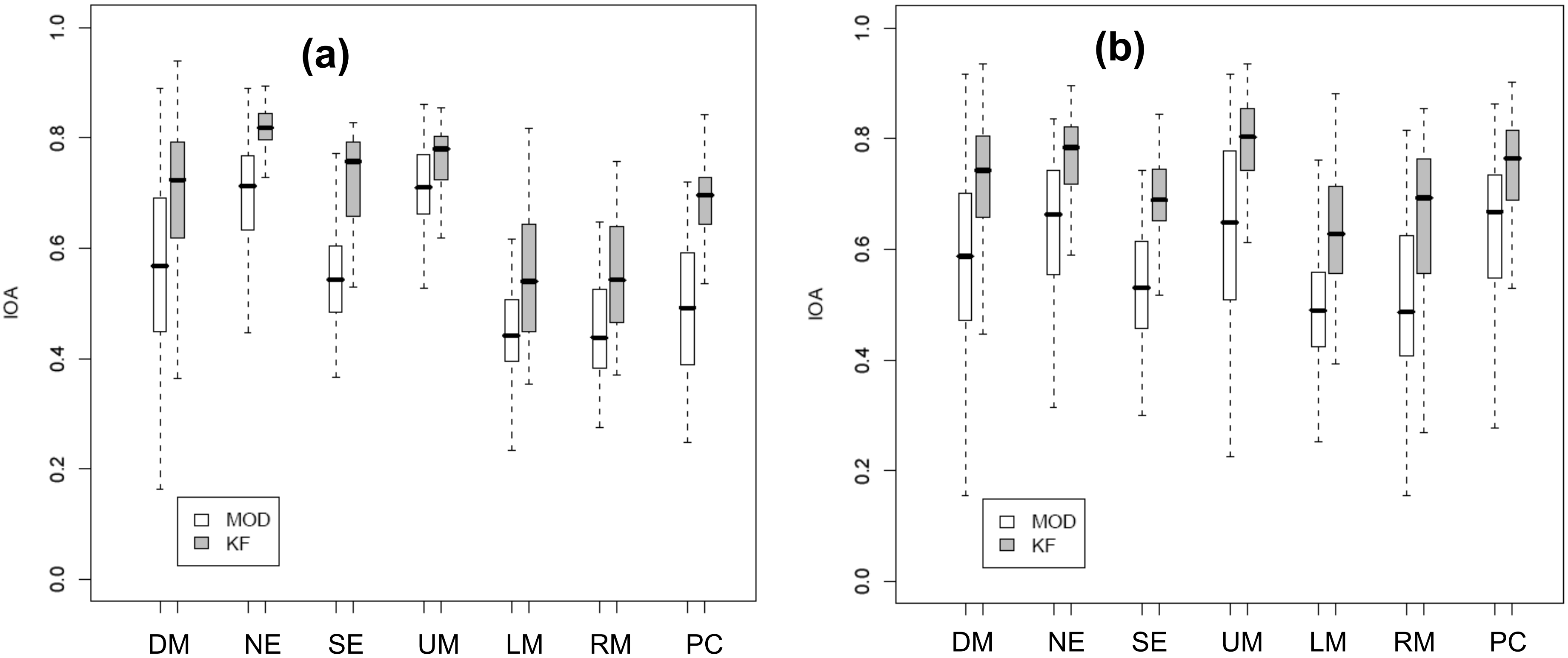

Figure 5. Box plots of index of agreement (IOA) of daily mean $PM_{2.5}$ ($\mu g/m^3$) for the raw model (MOD) forecasts and KF bias-adjusted forecasts over the domain (DM) and across all subregions during (a) warm season and (b) cool season

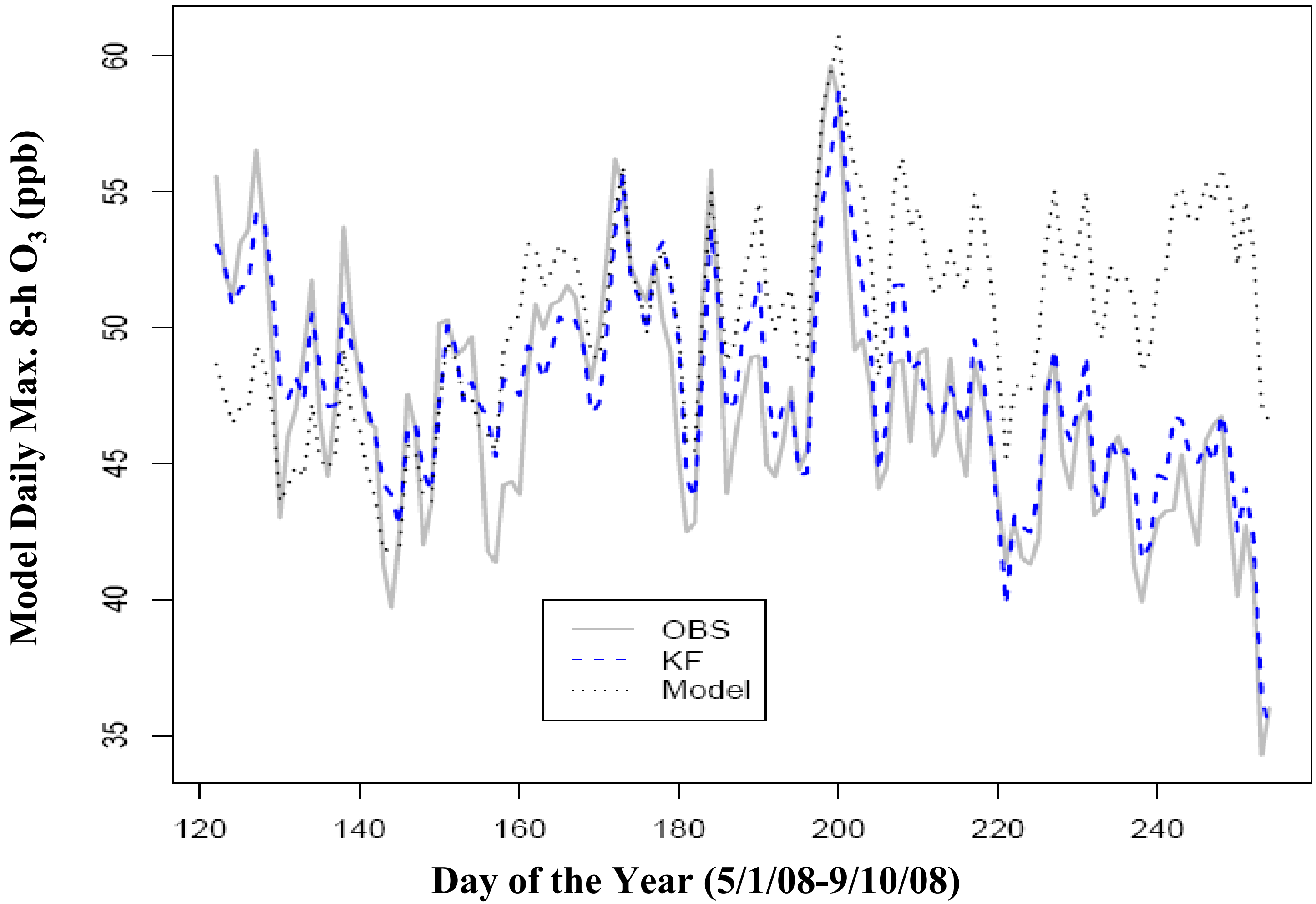

Figure 6. Time series of observed, raw model forecast, and KF bias-adjusted forecast mean daily maximum 8-h O_3 (ppb) over the domain

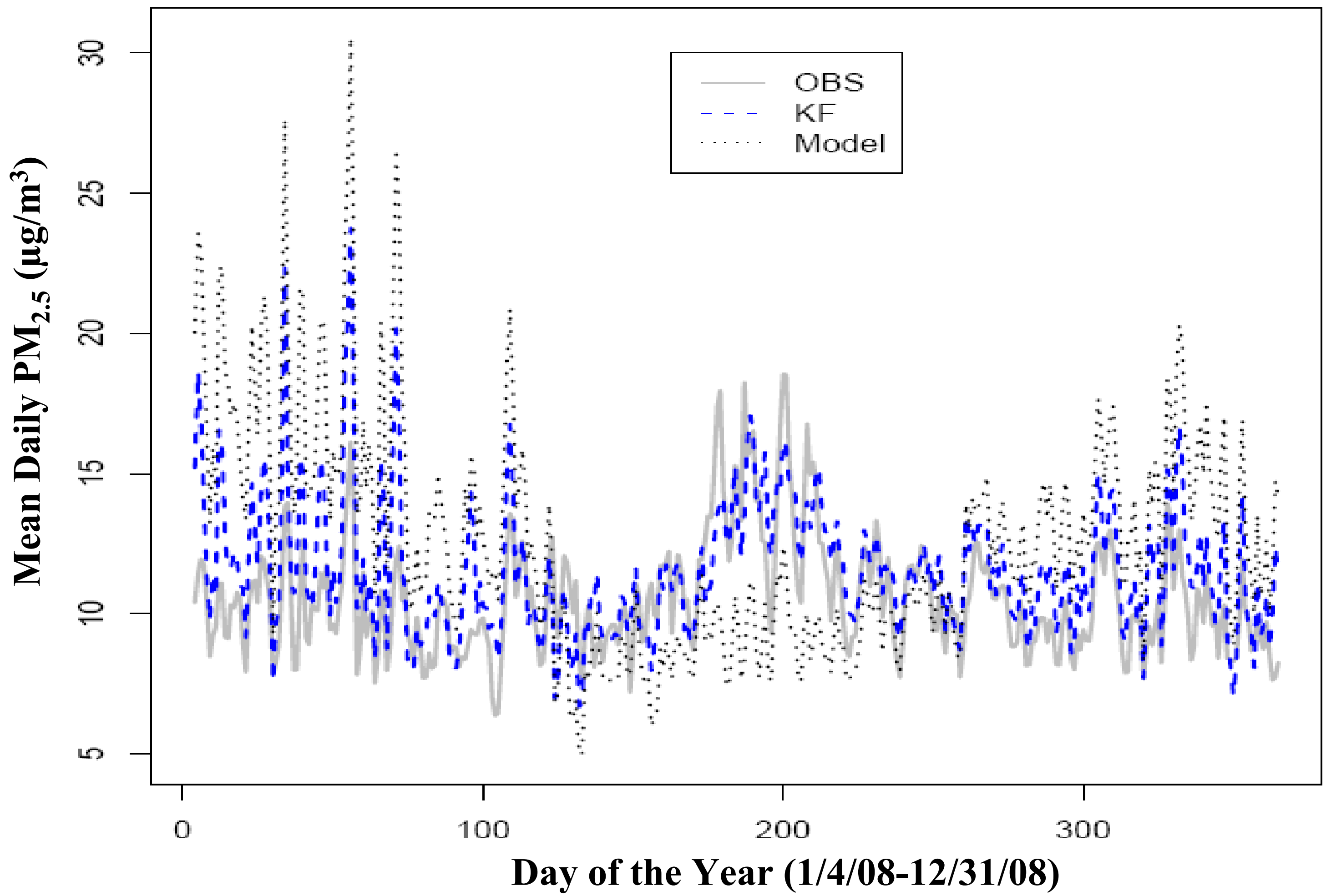

Figure 7. Time series of observed, raw model forecast, and KF bias-adjusted forecast mean daily $PM_{2.5}$ ($\mu g/m^3$) over the domain

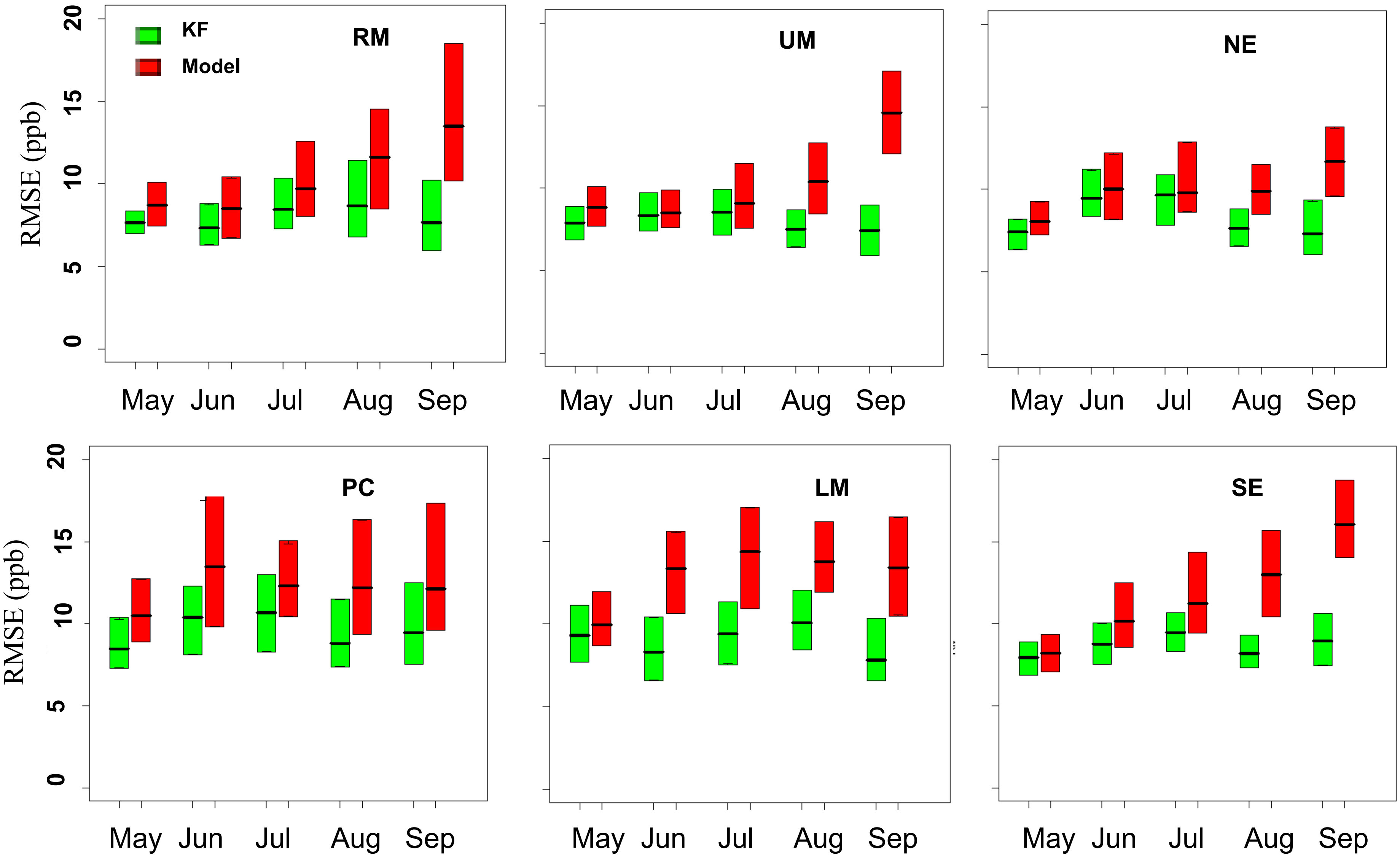

Figure 8. Monthly box plots (only 25th and 75th percentiles and median values are shown) of RMSE values of the daily maximum 8-h O_3 (ppb) for the raw model and KF bias-adjusted forecasts for all sub-regions

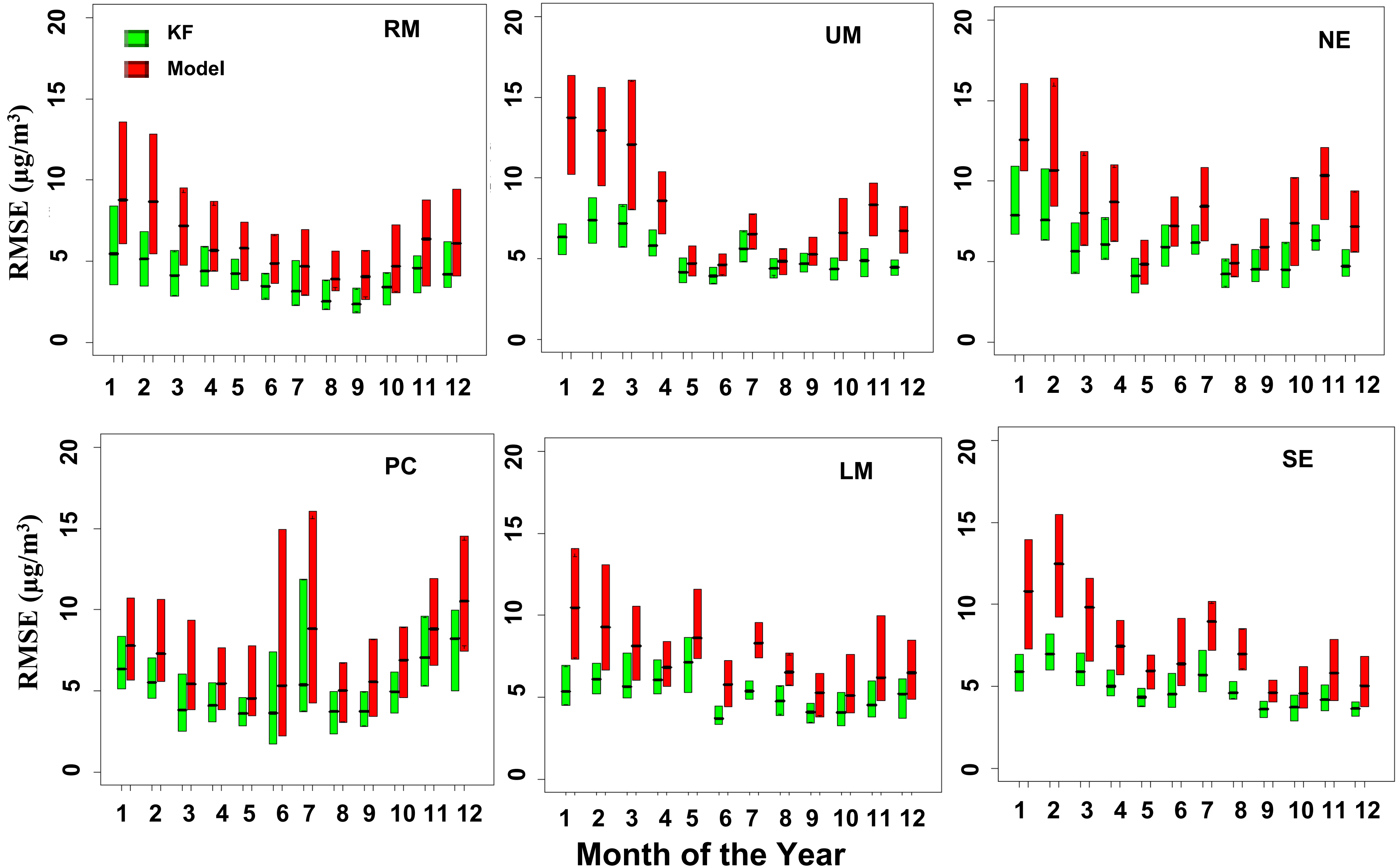

Figure 9. Monthly box plots (only 25th and 75th percentiles and median values are shown) of RMSE values of the daily mean $PM_{2.5}$ ($\mu g/m^3$) for the raw model and KF bias-adjusted forecasts for all sub-regions

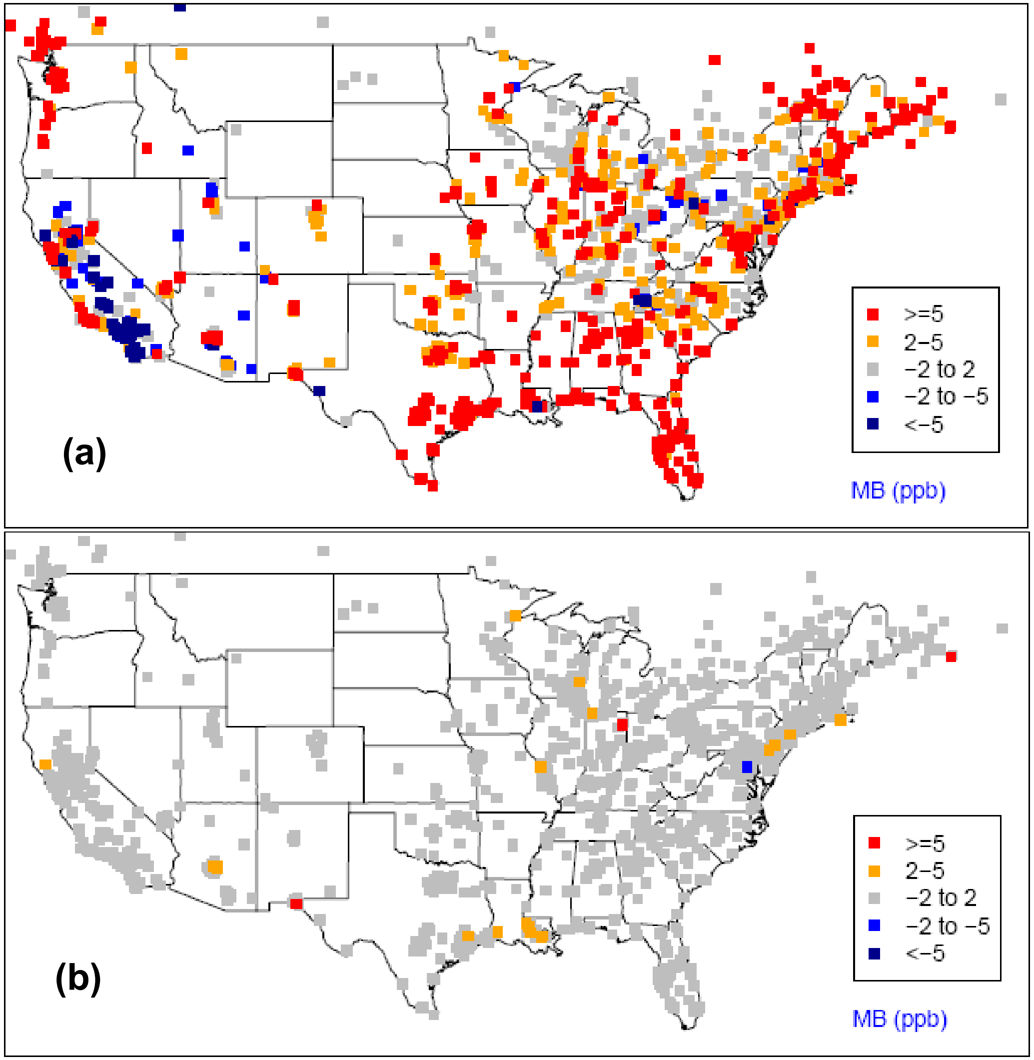

Figure 10. Mean Bias (MB, ppb) for daily maximum 8-h O_3 forecasts at each location within the continental U.S. domain: (a) raw model, (b) KF bias-adjustment

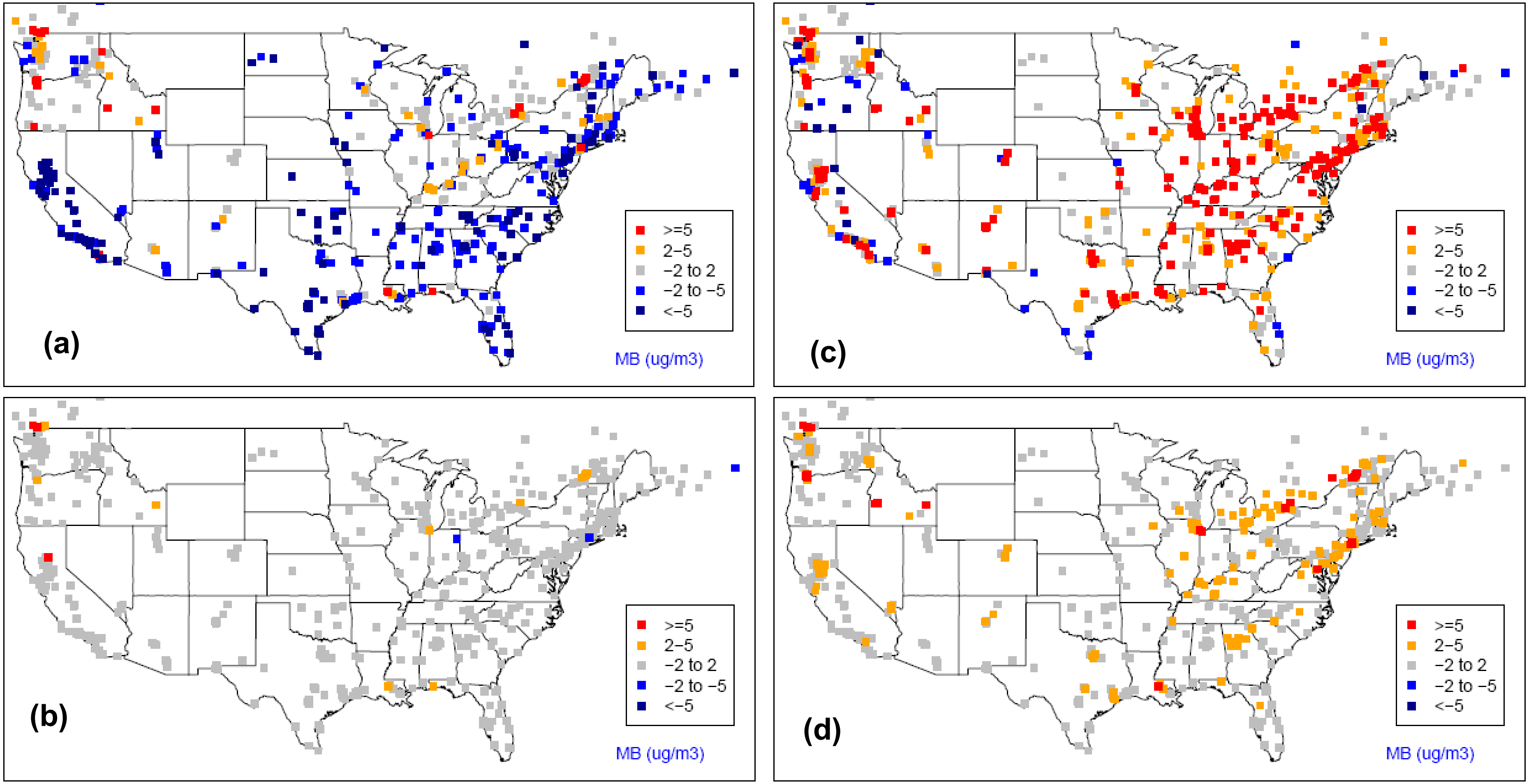

Figure 11. Mean Bias (MB, $\mu g/m^3$) for daily mean $PM_{2.5}$ forecasts at each location within the continental U.S. domain: (a) raw model during warm season, (b) KF bias-adjustment during warm season, (c) raw model during cold season, (d) KF bias-adjustment during cold season

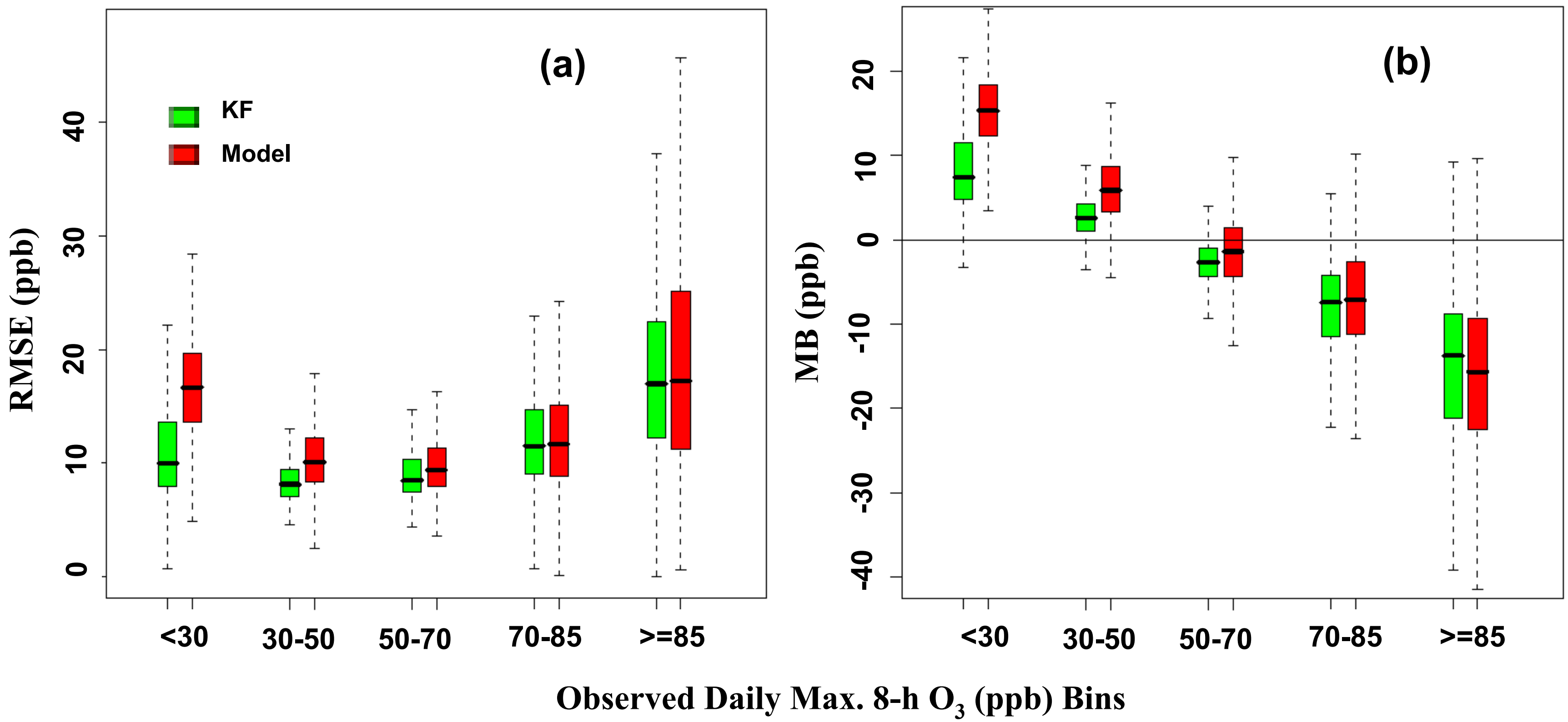

Figure 12. RMSE and MB values over observed daily maximum 8-h O_3 mixing ratio bins for the raw model forecasts and Kalman filter bias-adjusted forecasts over the domain: (a) RMSE and (b) MB

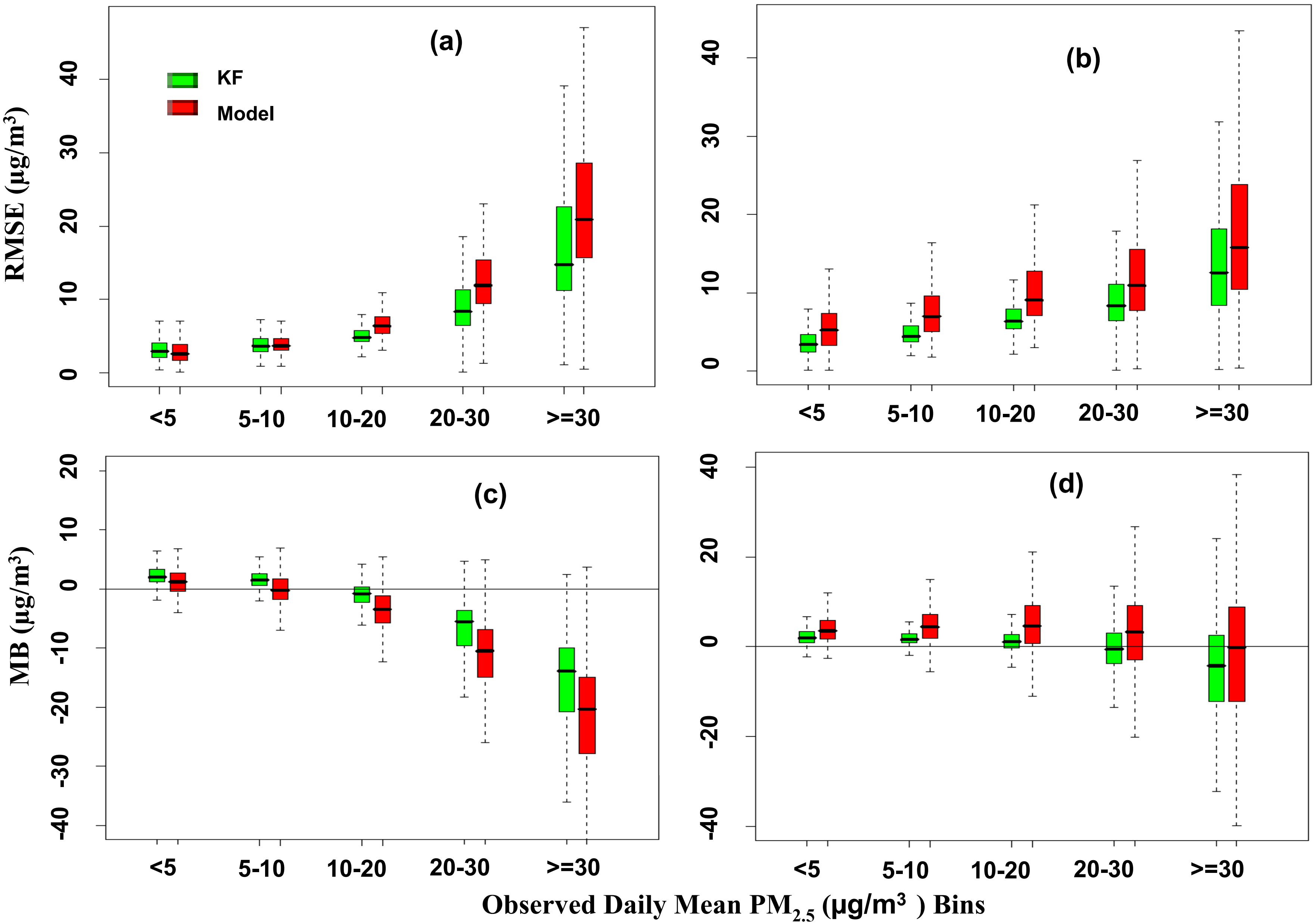

Figure 13. RMSE and MB values over observed daily mean $PM_{2.5}$ concentration bins for the raw model forecasts and Kalman filter bias-adjusted forecasts over the domain: (a) RMSE during warm season, (b) RMSE during cool season, (c) MB during warm season, and (d) MB during cool season

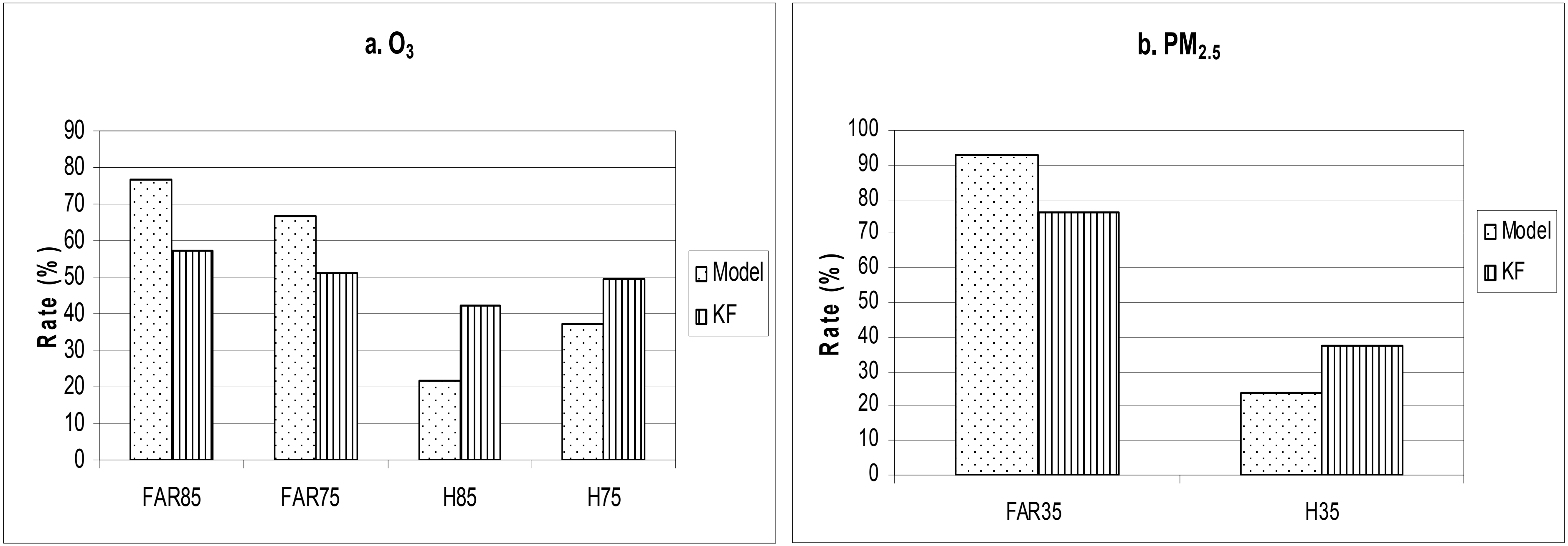

Figure 14. FAR and H values for both raw model and KF forecasts: a. daily maximum 8-h O_3 (ppb), and b. daily mean $PM_{2.5}$ ($\mu g/m^3$)

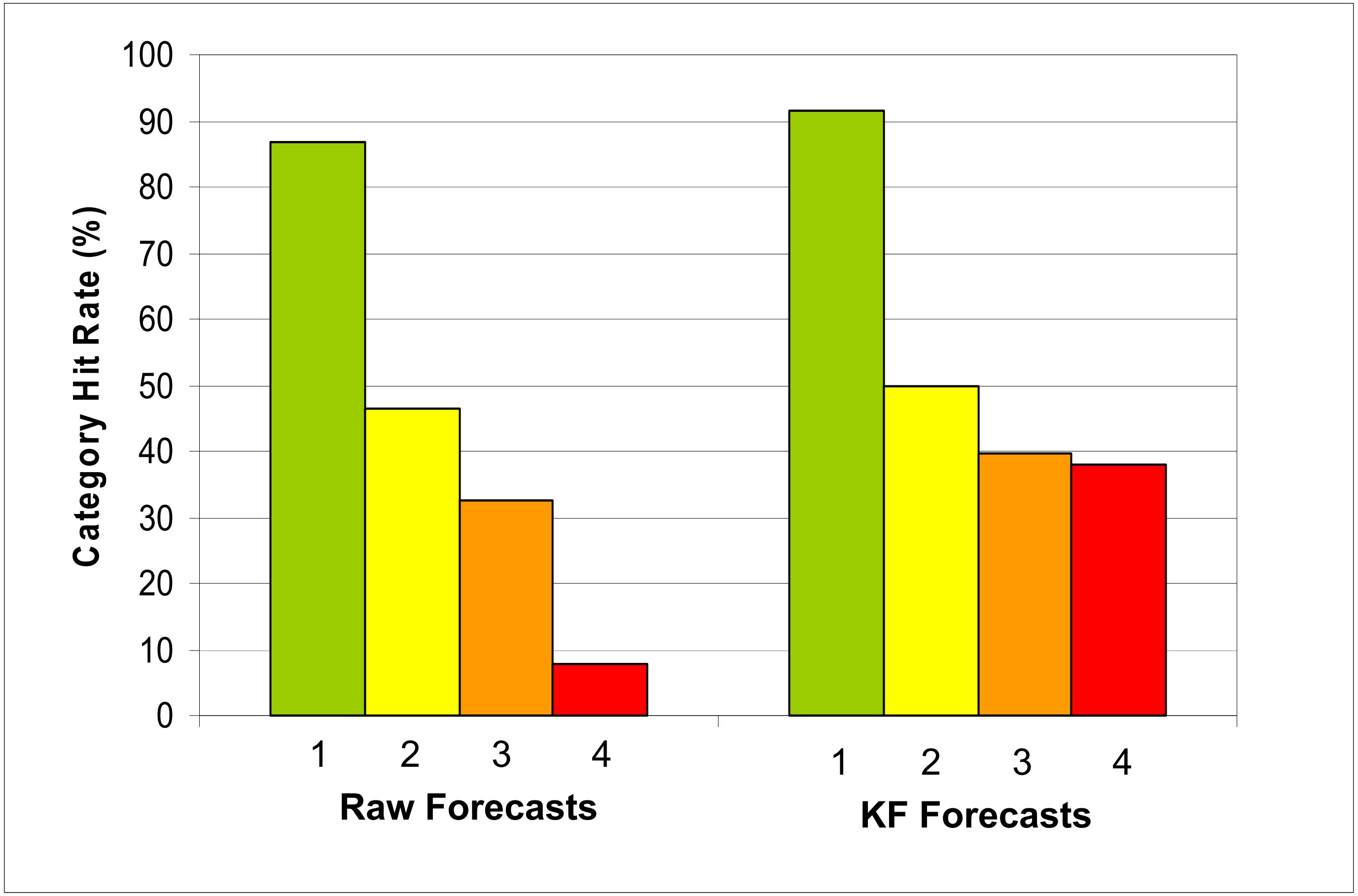

Figure 15. Categorical Hit Rates for each AQI Category for daily maximum 8-hr O_3

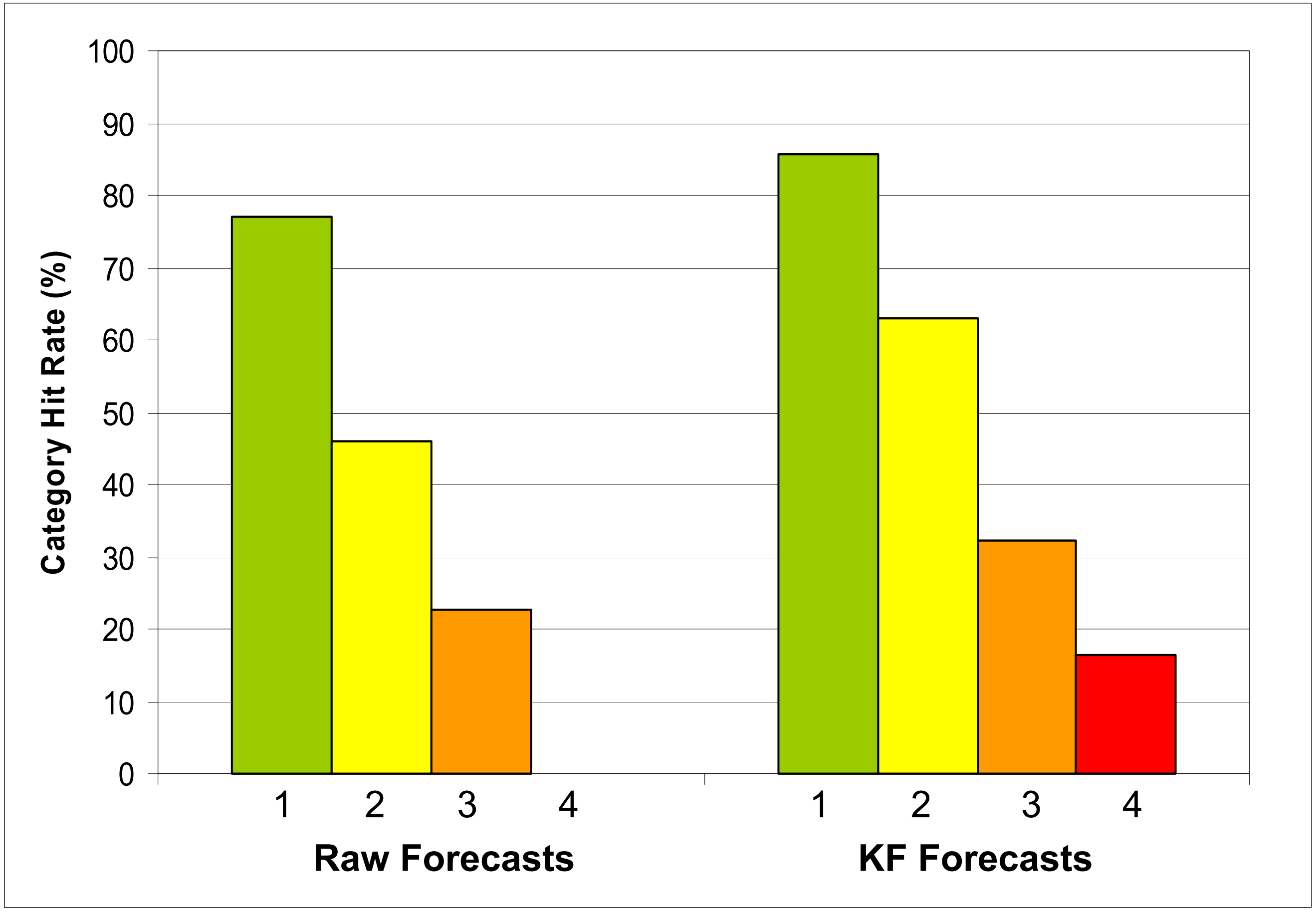

Figure 16. Categorical Hit Rates for each AQI Category for daily mean $PM_{2.5}$