Geosci. Model Dev. Discuss., 2, 1375–1406, 2009
www.geosci-model-dev-discuss.net/2/1375/2009/

Geoscientific Model Development Discussions

This discussion paper is/has been under review for the journal Geoscientific Model Development (GMD). Please refer to the corresponding final paper in GMD if available.

Assessment of bias-adjusted $PM_{2.5}$ air quality forecasts over the continental United States during 2007

D. Kang[1], **R. 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, US Environmental Protection Agency, Research Triangle Park, NC, USA

Received: 29 October 2009 – Accepted: 16 November 2009 – Published: 8 December 2009

Correspondence to: D. Kang (kang.daiwen@epa.gov)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

To develop fine particular matter ($PM_{2.5}$) air quality forecasts, a National Air Quality Forecast Capability (NAQFC) system, which linked NOAA's North American Mesoscale (NAM) meteorological model with EPA's Community Multiscale Air Quality (CMAQ) model, was deployed in the developmental mode over the continental United States during 2007. This study investigates the operational use of a bias-adjustment technique called the Kalman Filter Predictor approach for improving the accuracy of the $PM_{2.5}$ forecasts at monitoring locations. The Kalman Filter Predictor bias-adjustment technique is a recursive algorithm designed to optimally estimate bias-adjustment terms using the information extracted from previous measurements and forecasts.

The bias-adjustment technique is found to improve $PM_{2.5}$ forecasts (i.e. reduced errors and increased correlation coefficients) for the entire year at almost all locations. The NAQFC tends to overestimate $PM_{2.5}$ during the cool season and underestimate during the warm season in the eastern part of the continental US domain, but the opposite is true for the pacific coast. In the Rocky Mountain region, the NAQFC system overestimates $PM_{2.5}$ for the whole year. The bias-adjustment forecasts can quickly (after 2–3 days' lag) adjust to reflect the transition from one regime to the other. The modest computational requirements and systematical improvements in forecast results across all seasons suggest that this technique can be easily adapted to perform bias-adjustment for real-time $PM_{2.5}$ air quality forecasts.

1 Introduction

Ozone (O_3) and fine particulate matter ($PM_{2.5}$; particles with aerodynamic diameters less than 2.5 μm) in the atmosphere have been a major concern because of their adverse effects on human and ecosystem health. O_3 and $PM_{2.5}$ are the two pollutants used to compute the Air Quality Index (AQI), a standardized indicator of air quality conditions at a given location (http://www.airnow.gov). Thus, to develop accurate AQI based health advisories, it is desirable that air quality forecast systems at least be capable of forecasting these two species well. Real-time O_3 forecasts using air quality models have been publicly available for several years over different domains (McHenry et al., 2004; McKeen et al., 2005; Otte et al., 2005; Eder et al., 2006), while real-time $PM_{2.5}$ forecasts are mainly in the developmental stage. The NAQFC (Otte et al., 2005), developed by the National Oceanic and Atmospheric Administration (NOAA) and the US Environmental Protection Agency (EPA) couples NOAA's operational North American Mesoscale (NAM) weather prediction model (Black 1994; Rogers et al., 1996) with EPA's Community Multiscale Air Quality (CMAQ) model (Byun and Schere, 2006). It has the capability to provide real-time forecasts for both O_3 and $PM_{2.5}$. The developmental mode model predictions are available for the year of 2007 over the continental US domain, representing a consistent and integral data set to perform comprehensive model performance evaluations and bias-adjustment studies.

Adverse health effects in humans have been shown to be associated with exposure to elevated ambient $PM_{2.5}$ levels (e.g., NRC, 1998). While it is recognized that $PM_{2.5}$ pollution results from both primary emissions and secondary formation through complex photochemical and heterogeneous chemical pathways, significant scientific and technical challenges surround the characterization of ambient $PM_{2.5}$ distributions both through modeling and measurements (e.g., McMurry, 2000; Donahue et al., 2009). The emissions, physical, chemical, and removal processes controlling the day-to-day levels of ambient $PM_{2.5}$ and precursor concentrations also exhibit seasonal variability resulting in significant spatial and seasonal variability in ambient $PM_{2.5}$ mass and its chemical composition. Existing uncertainties in these individual components poses enormous challenges for developing accurate short-term $PM_{2.5}$ forecasts (Mathur et al., 2008; Yu et al., 2008). Nevertheless, a need exists for local air quality agencies to accurately forecast $PM_{2.5}$ concentrations to alert the sensitive population of the onset and duration of unhealthy air associated with elevated $PM_{2.5}$ levels. To address this need, the utility of $PM_{2.5}$ forecast guidance obtained from comprehensive atmospheric models can, in the short-term, be improved through the application of post-process bias adjustment methods; this serves as the primary motivation for the analysis presented in this study. It should be noted that post processing bias-adjustment techniques are routinely used in conjunction with numerical weather prediction models, despite decades of research to improve the formulations of the models, to develop more accurate forecast products. Given the relatively early state of $PM_{2.5}$ forecast model and the current uncertainties in process representations, the exploration of bias adjustment techniques to improve the usefulness of $PM_{2.5}$ forecasts is warranted.

Different bias-adjustment (also referred to as bias-correction) techniques have been used for improving surface O_3 predictions in recent years (McKeen et al., 2005; Delle Monache et al., 2006; Wilczak et al., 2006; Delle Monache et al., 2008; Kang et al., 2008). Among these techniques, the Kalman Filter (KF) predictor (hereafter refer to KF bias-adjustment or simply KF) forecast method yielded the most forecast skill improvement. Kang et al. (2008) presented a comprehensive study on the application of KF technique to O_3 forecasts over the continental US domain for a three-month period from July to September 2005. While the techniques were found to improve the forecast skill for O_3, it is not clear if they are readily applicable for PM forecasts and whether they would yield similar improvements in PM forecast skill. This is primarily due to the fact that unlike O_3, elevated $PM_{2.5}$ concentrations are encountered throughout the year and that significant seasonal biases exist in current models both in the representation of total $PM_{2.5}$ mass as well as its composition (cf. Mathur et al., 2008; McKeen et al., 2007; Appel et al., 2008). Additionally, the chemical constituent dominating the bias could also vary both spatially and seasonally. Thus, for improved PM forecasts, the bias adjustment techniques should be capable of correcting biases and errors that not only change with time but that also may have widely varying sources of origin.

In this study, the KF bias-adjustment technique is applied to $PM_{2.5}$ forecasts for the year of 2007 over the continental US domain; to our knowledge this is the first comprehensive assessment of this bias-adjustment technique for $PM_{2.5}$ forecasts. Within the continental US domain, there are about 500 AIRNow sites that report hourly $PM_{2.5}$ concentrations which are measured using the Tapered Element Oscillating Microbalance

(TEOM) method. The year-long forecast period over the continental US has provided a unique data set covering a wide range of atmospheric conditions and a broad $PM_{2.5}$ concentration range to test the performance of the bias-adjustment technique for $PM_{2.5}$ forecasts.

The objectives of this study include: (1) apply the KF post-processing technique to improve skills for real-time $PM_{2.5}$ forecasts, (2) investigate the spatial and temporal characteristics of this technique when applied to $PM_{2.5}$ forecasts, and (3) analyze the impact of bias-adjustment on forecast errors of different types (i.e., systematic versus unsystematic).

2 Experiments and methods

2.1 The NAQFC system

The NAQFC system consists of three primary components: (1) the National Weather Service's North American Mesoscale (NAM) model based on the Weather Research Forecast nonhydrostatic mesoscale model (WRF-NMM) which provides the meteorological and atmospheric dynamic conditions for the AQF; (2) the US EPA's Community Multiscale Air Quality (CMAQ) (Byun and Schere, 2006) model simulates the transport, chemical evolution, and deposition of atmospheric substance; and (3) an interface component (PREMAQ) that processes both the meteorological and emission inputs to conform with the CMAQ grid structure, coordinate system, and input format. For this application, $PM_{2.5}$ concentrations are forecast over the continental US (Fig. 1) using a 12-km horizontal grid spacings on the Lambert Conformal map projection and 22 layers of variable thickness in the vertical. Since the $PM_{2.5}$ forecasts were in the developmental stage, changes or modifications to the AQF components were allowable to accommodate new developments from evolving science. For instance, on 17 September 2007, the treatment for the PBL mixing height in CMAQ was changed from the Turbulence Kinetic Energy (TKE)-based method to the Asymmetric Convective Model-

2 (ACM2)-based method, which on average decreased the PBL depth, helping reduce forecast errors for both O_3 and $PM_{2.5}$ in the Pacific Coast region. However, this study does not deal with the impacts of the various changes or modifications to the forecast model, rather it focuses on how the bias-adjustment technique can improve the forecast results over the raw model forecasts. Since the bias-adjustment technique employed in this study is statistical, it does not involve any perturbations in the physical and chemical processes treated in the forecast model.

The emissions inventories used by the AQF system were updated to represent the 2007 forecast year based on the input from the US EPA national emission inventory. The Carbon Bond chemical mechanism (version 4.2) is used to represent the photochemical reactions and AERO3 aerosol module is used to represent aerosol formation and distribution. The chemical fields for CMAQ are initialized using the previous forecast cycle. The primary NAM-CMAQ model forecast for the next 48-h surface-layer $PM_{2.5}$ is based on the current day's 6 UTC cycle.

2.2 Observations

Hourly, near real-time, $PM_{2.5}$ measurements ($\mu g/m^3$) obtained from EPA's AIRNow program are used in this study (http://www.epa.gov/airnow). All measurements are made using TEOM instruments and concentrations are averaged over hourly intervals from the beginning of the one hour to the next. It should be recognized that TEOM measurements are somewhat uncertain and are believed to be lower limits to a "true" value because of volatilization of semivolatile material (ammonium nitrate and organic carbon) in the drying stages of the measurement (Eatough et al., 2003; Grover et al., 2005). Nevertheless, the TEOM measurements are the only real-time hourly $PM_{2.5}$ observation data available for use in the purpose of this study. About 500 $PM_{2.5}$ monitoring stations are available within the continental US domain (Fig. 1) for the year of 2007. For verification purposes and forecast products, the daily (24-h) mean $PM_{2.5}$ concentrations are often used.

2.3 Implementation of the KF bias-adjustment method

The KF predictor bias-adjustment algorithm (Kalman, 1960) was described in detail by Delle Monache et al. (2006) and a concise description of its implementation is provided in Kang et al. (2008). The specification of the error ratio, a key parameter in the KF approach which determines the relative weighting of observed and forecast values, was previously investigated extensively for O_3 forecasts. Even though the optimal error ratios were found to vary across space, the impact of using the optimal values on the resultant bias-adjusted predictions was insignificant when compared to using a reasonable single fixed value across all locations (Kang et al., 2008). Thus, in this study, we use the same single fixed error ratio value of 0.06 at all the locations for developing bias-adjusted $PM_{2.5}$ forecasts.

There are two steps to implement the KF bias-adjustment technique. First, the KF is initialized with the initial estimates of KF parameters as outlined in Kang et al. (2008) and hourly observations and raw model predictions for the prior 2 days. Then the updated parameters and the third day's raw model forecasts are used to create bias-adjusted forecasts for the 3rd day. All the updated KF parameters for each hour and at each site are saved into a file for use in the subsequent KF run. The KF runs then continue by reading the previous day's KF parameters and observations and raw model predictions from the prior 2 days to generate the next day's bias-adjusted forecasts through combining with the next day's raw model forecasts. Thus, in developing the daily KF forecasts, if 2 consecutive days' data are missing at a site, the KF will automatically drop this site from future bias-adjustment forecasts; however, if a new site with 2 consecutive days' data appears in the observation data set, the KF will initialize the site with initial values of KF parameters and generate bias-adjusted forecasts further on. This implementation is adaptable in real-time to the variable nature of monitoring stations which report hourly observations to the AIRNOW network and can be easily combined with AQF system to produce real-time bias-adjusted forecasts.

2.4 Verification statistics and spatial-temporal considerations

To assess the performance of the KF bias-adjusted forecasts, a variety of statistical metrics are used, including Root Mean Square Error (RMSE) and its systematic and unsystematic components, Normalize Mean Error (NME), Mean Bias (MB), Normalized Mean Bias (NMB), and correlation coefficient (r). For a forecast product, it is also important to evaluate its performance over categorical forecasts (Kang et al., 2005). The categorical metrics, including False Alarm Ratio (FAR) and Hit Rate (H), are used in this study.

Since the NAQFC domain covers the continental United States, and given large differences in physical and chemical processes from region to region, the continental US domain is divided into seven subregions to facilitate the performance evaluations (Fig. 1). The four easternmost subregions, northeast (NE), southeast (SE), upper Midwest (UM), and lower Midwest (LM), are based on an O_3 climatology that identified areas of homogeneous variability using principal component analysis (Eder et al., 1993).

Figure 2 shows comparisons of time series of the domain-wide daily average observed, raw model forecasts, and KF bias-adjustment forecasts of $PM_{2.5}$ concentrations during 2007. As shown in Fig. 2, compared to the observations the raw model tends to overpredict during cool season (before mid-April and after August) and underpredict during warm season (mid-April to end of August). To facilitate the temporal performance evaluations, the time series is divided into cool season (from January to 20 April and from September to December) and warm season (from 21 April to 31 August). Further more, the cool season is divided into the first cool season (from January to April) and second cool season (from September to December).

GMDD

2, 1375–1406, 2009

Assessment of bias-adjusted $PM_{2.5}$ air quality forecasts

D. Kang et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◂◂ | ▸▸
◂ | ▸
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

3 Results

3.1 General performance

As seen in Fig. 2, the raw model overestimated the $PM_{2.5}$ concentrations during the cool season, especially during the second cool season (from September to December). During the warm season, the raw model significantly underestimated, and the KF predictions were well above the raw model predictions and much closer to the observations. From late July to early September, the raw model underwent a transitional period from underestimation to overestimation. From early September until the end of the year, the overestimation of the raw model became larger. This is partially attributed to the change of the PBL mixing scheme for CMAQ as mentioned earlier on 17 September. Nevertheless, the KF bias-adjustment technique could quickly respond to the transitions from one regime to another and tracked the observations well in the time series. Since Fig. 2 presents the aggregate results for the entire domain, some important information may be hidden due to smoothing during the averaging process. Figure 3 displays same time series as Fig. 2 for two representative sub-regions: Southeast and Pacific Coast. The time series of the Southeast resembles that of the domain with raw model overestimation during cool season and underestimation during warm season. However, the under-prediction during the warm season was more pronounced for the Southeast than for the entire domain. The time series of the Pacific Coast presented a completely different story, in which the raw model generally over-predicted during the cool season and under-predicted during the warm season. The over-prediction was much stronger at the beginning of the year (January and early February) than that over the rest of cool season. The over-prediction for the second cool season was reduced, and during most times the raw model could reproduce the observations quite well. The performance change of the raw model during cool season is attributed to the adoption of the new PBL mixing height parameterization when the TKE-based PBL height was replaced by the ACM2-based PBL height. The ACM2-based PBL height generally leads to higher $PM_{2.5}$ concentrations than the TKE-based

PBL height since the ACM2-based PBL height is generally lower than the TKE-based PBL height. The under-prediction is thus reduced for the west region of the domain during the cool season, but the over-prediction is further aggravated for the eastern part of the domain during the same period. Nonetheless, the time series of the KF bias-adjusted predictions tracked the observed time series better than the raw model predictions.

To further investigate the performance of the KF bias-adjusted forecasts and compare with the raw model forecasts, Fig. 4 displays the scatterplots between the forecast and observed values across various percentiles for the daily mean $PM_{2.5}$ for all the stations within the continental US domain. Following Mathur et al. (2008), at each site the time series of both measured and model (or KF bias-adjusted model) daily mean $PM_{2.5}$ over the entire year was examined and percentiles of the distribution over the study period were computed for both the model and the measurements. Scatterplots of specific percentiles of the concentration distributions (e.g., median) of the model and observed time series are then examined to assess the ability of the model to capture the spatial variability in frequency distributions of $PM_{2.5}$ concentrations across the sites (Mathur et al., 2008). As shown in Fig. 4, compared with the raw model forecasts (left), the KF bias-adjusted forecasts displayed a much better match with the observed distributions as reflected by the reduced scatter about the 1:1 line, especially for the higher percentiles. The overall correlation between model forecasts and observations was greatly improved with the value of R^2 increasing from 0.43 for the raw model forecasts to 0.90 for the KF bias-adjusted forecasts. Similar improvements in O_3 forecasts after the application of the KF bias-adjustment were previously reported in Kang et al. (2008).

The ability of the KF bias-adjustment technique to improve the predicted $PM_{2.5}$ concentration distributions is further illustrated in Fig. 5 which displays the histograms of observed daily mean $PM_{2.5}$ concentrations along with the fitted probability density functions (PDFs) of daily mean $PM_{2.5}$ concentrations for the observations, raw model forecasts, and KF bias-adjusted forecasts. Figure 5a displays the overall distribution for the entire domain during 2007, while Fig. 5b presents the distribution for Lower Midwest during warm season and Fig. 5c for Pacific Coast during cool season to typify the sub-regional and seasonal signals. As seen in Fig. 5, the KF technique brings the PDFs of forecast values much closer to those of the observations. The improvements are more pronounced in the sub-regional and seasonal distribution comparisons illustrated in Fig. 5b and c. The distributions of the raw model forecasts for both cases were out of phase compared to those of the observations, especially for Lower Midwest during warm season. The KF bias-adjusted forecasts were able to reproduce the observations very well in both cases.

3.2 Regional performance

Tables 1 and 2 present the domain and sub-regional summary of discrete statistics for the raw model and the KF bias-adjusted daily mean $PM_{2.5}$ forecasts during cool and warm seasons, respectively. Examination of Table 1 reveals that during the cool season, the RMSE values range from 7.2 to 11.4 ($\mu g/m^3$) for the raw model forecasts, and from 5.2 to 7.6 ($\mu g/m^3$) for the KF bias-adjusted forecasts; this translates to about a 20% reduction in RMSE as a result of the application of the bias-adjustment. Similar reductions are also noted for the NME. The MB and NMB indicate that during the cool season, the raw model systematically over-predicted daily mean $PM_{2.5}$ across all the sub-regions except the Pacific Coast where it under-predicted. The KF bias-adjusted forecast reduced NMB values across all the sub-regions. Correlation coefficients also increased significantly across all the regions as a result of the bias-adjustment, with the largest increase in the LM and RM regions. The summary statistics during warm season (Table 2) indicate comparable improvement in the error statistics (RMSE and NME) for the KF bias-adjusted forecasts relative to the raw model. In contrast to the cool season, systematic under-predictions are noted in the warm season raw model $PM_{2.5}$ forecasts (Mathur et al., 2008). The application of the KF bias-adjustment helps reduce both the cool season high bias and the warm season low bias, and also results in consistently improved correlations with measurements across all seasons.

Figure 6 presents comparisons of the distribution of monthly RMSE values of daily mean $PM_{2.5}$ for the raw model and KF forecasts for the different sub-regions. In these plots, the lower and upper borders of the box represent the first and third quartiles, while the middle line represents the median value. As seen in Fig. 6, the RMSE values are consistently lower for the KF forecasts relative to those of the raw model across all sub-regions and months. In addition, the error distribution range (the size of the boxes) for the KF forecasts is also much smaller than the raw model forecasts. During October–December, the raw model forecasts exhibited large RMSE values for both the UM and LM sub-regions (partly attributed to a change in the PBL height parameterization discussed earlier). The KF bias-adjustment was able to reduce these large RMSEs significantly. In making comparisons across the regions, it should be noted that the relatively larger spread in RMSE for the RM and PC regions, especially for the raw model forecasts likely result from a combination of effects related to complex topography, land-sea breeze transitions in the PC region, greater spatial heterogeneity in emissions, and their impact on chemistry leading to $PM_{2.5}$ formation and distribution.

Figure 7 presents the spatial distribution of mean biases at each site within the modeling domain for both the cool and warm seasons. As illustrated in Fig. 7a, during warm season, the raw model predominantly under-predicted at most sites (orange and purple squares) in the eastern part of the domain, over-predicted in the northwest regions and exhibited both over- and under-predictions at sites in California. During cool season, the raw model generally over-predicted (Fig. 7c) in the east, but under-predictions dominated at sites in western portions of the domain. The application of the KF bias-adjustment was able to effectively rectify these biases at more than 90% of the sites (Fig. 7b and d) to less than 2 $\mu g/m^3$. Even at the sites where absolute bias was greater than 2 $\mu g/m^3$, the magnitude of the bias was significantly reduced.

The forecast skill improvement over space by the KF forecasts over the raw model forecasts is further demonstrated by the index of agreement (IOA) as shown in Fig. 8. The IOA increased on average from 8% (at NE and UM) to 30% at PC during the warm season (Fig. 8a) and from 15% (at NE and SE) to 28% at RM during the cool season

(Fig. 8b). The domain-wide average IOA values increased by 13% and 19% for the warm season and the cool season, respectively.

3.3 Systematic/unsystematic errors and performance over concentration bins

The RMSE can be further decomposed into its systematic and unsystematic components (Willmott, 1981) based on the least-square linear regression relationship between forecast values and observations (Kang et al., 2008). The boxplots in Fig. 9 show the distribution of the RMSE, and its systematic (RMSEs) and unsystematic (RMSEu) components of the predicted daily mean $PM_{2.5}$ for the raw model and KF forecasts across all the stations within the continental US domain. Shown in the boxplots are the first quartile (lower border of the box), the third quartile (upper border of the box), and the median (the central line) values of the distributions. The whiskers represent the 1.5 IQR (inter-quartile range). The decomposition of the RMSE displays different error characteristics for $PM_{2.5}$ relative to those noted previously for O_3 forecasts (Kang et al., 2008). First, for the raw model forecasts, while systematic errors were larger than the unsystematic components for O_3, the converse is noted for $PM_{2.5}$ forecasts. The larger contribution of unsystematic errors to the $PM_{2.5}$ RMSE not only reflect the bigger uncertainty in its emissions and in our understanding of the atmospheric processes regulating its measurement uncertainty, but also the local variability in the predominantly urban AIRNOW measurement network. The application of the KF bias-adjustment helps reduce both the unsystematic and systematic errors in $PM_{2.5}$ forecasts.

To further examine the performance of the KF bias-adjustment technique over different concentration ranges, Fig. 10 displays the forecast RMSE and MB values as a function of observed concentrations for both the warm and cool seasons. During warm season (Fig. 10a), when observed $PM_{2.5}$ concentrations were less than 10 $\mu g/m^3$, the KF bias-adjustment technique was unable to reduce RMSE values compared to the raw model forecasts, though the distributions were more condensed. This may in part be attributed to the fact that during warm season, the weather conditions tend to be more variable (more convective weather conditions) than those during cool season and lower concentrations are often associated with precipitation processes, and the raw model generally has difficulty to accurately simulate these weather conditions, resulting in larger unsystematic errors in the prediction of $PM_{2.5}$ concentrations. When the observed $PM_{2.5}$ concentrations were larger than 10 $\mu g/m^3$, the RMSE values associated with KF forecasts were much smaller in both the mean values and the distributions compared to the raw model forecasts. In contrast, during the cool season (Fig. 10b), the KF forecasts performed better than the raw model forecasts across all the concentration bins. Examination of the MB distributions over the observed concentration bins (Fig. 10c and d) reveals that the raw model over-predicted at lower concentrations and under-predicted at higher concentrations, which is similar to the raw model performance for O_3 forecasts (Kang et al., 2008). The under-prediction at higher concentration bins for $PM_{2.5}$ forecasts during the warm season was more severe than that during the cool season. In general, the KF forecasts were able to adjust the MB towards the zero line over all the concentration bins for both seasons.

3.4 Categorical performance

It is equally important to evaluate the performance of an air quality forecast system using the categorical metrics, because for the general public, it is more important to know if the NAQFC system could simulate the occurrences of an exceedance or non-exceedance. Categorical evaluations for O_3 forecasts have been extensively performed in the past (Kang et al., 2005; Eder et al., 2006, 2009), but similar assessments for $PM_{2.5}$ forecasts have been limited. Figure 11 displays the false alarm ratio (FAR) and hit rate (H) (see Kang et al., 2005) for the raw model and KF bias-adjusted daily mean $PM_{2.5}$ forecasts for each of the sub-regions during both the warm and cool seasons. A threshold value of 35 $\mu g/m^3$ for the 24-h mean $PM_{2.5}$, based on the National Ambient Air Quality Standard (NAAQS) for $PM_{2.5}$, is used. As seen in Fig. 11, the FAR values associated with the raw model forecasts were similar (~85%) for both seasons over the entire domain, but the H values varied from less than 10% during the warm season to greater than 30% during the cool season. For the KF forecasts, the FAR values were reduced by more than 20% during both seasons, and the H values were more than doubled during the warm season and were increased by about 20% during the cool season for the entire domain. Compared to the raw model forecasts, the KF forecasts reduced the FAR values across all the sub-regions with differing magnitude and increased the H values for all the sub-regions except for the LM and RM. In general, the H values were higher during the cool season than those during the warm season for both the raw model forecasts and the KF forecast, while the FAR values didn't differ significantly.

4 Summary

The Kalman filter bias-adjustment technique has been applied to post-process $PM_{2.5}$ air quality forecasts over the continental US domain during the year of 2007. Though the application and analysis were conducted on archived $PM_{2.5}$ model forecast output, the methodology is easily adopted for real-time applications. To facilitate performance evaluation, the entire domain was divided into six sub-regions and the year was split into a cool season and a warm season to examine spatial and seasonal characteristics of the performance of the method. The assessment of the raw model performance indicates that the daily mean $PM_{2.5}$ concentrations were generally over-predicted over the eastern part of the domain during the cool season and under-predicted during the warm season; while the opposite is true for the western part of the domain, i.e., the daily mean $PM_{2.5}$ concentrations were typically under-predicted along the Pacific Coast during the cool season and over-predicted during the warm season; the Rocky Mountain region is an exception where the daily mean $PM_{2.5}$ concentrations were over-predicted throughout the year.

The KF bias-adjustment technique significantly improved the $PM_{2.5}$ forecasts as revealed by reductions in errors and biases, and higher correlation coefficients throughout the year and across the entire model domain. The analysis also shows that the KF bias-adjustment can quickly respond to transitions from one regime to another during the transition of seasons.

Analysis of RMSE and MB as a function of observed concentrations suggests that the KF method significantly reduces the raw model error and bias across all concentration ranges except at lower concentration bins during the warm season. However, the significant reductions in error and bias at the moderate-high concentration ranges helps improve the ability to predict exceedances which is desirable for air quality forecasting. The effectiveness and benefits of bias-adjustment of $PM_{2.5}$ model forecasts is also reflected in the categorical evaluations; the KF bias-adjustment technique improved the categorical evaluation metrics significantly by reducing the false alarm ratio and increasing the hit rate for almost all the regions during both cool and warm seasons.

It should be pointed out that the performance of bias-adjusted forecasts is dependent on the performance of the raw model to which the bias-adjustment technique is applied. Because of the complexity in $PM_{2.5}$ composition, formation, and distribution, it is even more critical for the raw model to provide a stable and well-behaved basis to make bias-adjusted forecasts more reliable. This bias-adjusted forecast study was based on the total mass of $PM_{2.5}$. If the components of $PM_{2.5}$ could be bias-adjusted separately, the results may be further improved than those derived from the bias-adjustment for the total $PM_{2.5}$ mass performed in this study. However, the lack of real-time measurements of speciated $PM_{2.5}$ hampers the use of KF adjustments on individual species. Improvements in the representation of fine particulate matter emissions as well as 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 uncertainties in the representation of atmospheric processes dictating the distribution of ambient $PM_{2.5}$, bias-adjustment techniques can be used to improve the reliability of short term $PM_{2.5}$ forecasts from such models and consequently help in issuance of air quality degradation related health advisories.

Acknowledgements. The authors are grateful to Brian Eder for his constructive and insightful comments on initial drafts of this manuscript. We thank L. Delle Monache and R. B. Stull for providing their original Kalman filter codes. The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. It has been subjected to Agency's administrative review and approved for publication.

References

Black, T.: The new NMC mesoscale eta model: description and forecast examples, Weather Forecast., 9, 265–278, 1994.

Byun, D. W. and Schere, K. L.: 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, 2006.

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

Delle Monache, L., Wilczak, J., McKeen, S., Grell, G., Pagowski, M., Peckham, S., Stull, R., McHenry, J., and McQueen, J.: A Kalman-filter bias correction method applied to deterministic, ensemble averaged, and probabilistic forecasts of surface ozone, Tellus Ser. B, 60x, 238–249, doi:10.1111/j.1600-0889.2007.00332.x, 2008.

Donahue, N. M., Robinson, A. L., and Pandis, S. N.: Atmospheric particulate matter: from smoke to secondary organic aerosol, Atmos. Environ., 43, 94–106, doi:10.1016/j.atmosenv.2008.09.055, 2009.

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

Eder, B., Kang, D., Mathur, R., Yu, S., and Schere, K.: An operational evaluation of the Eta-CMAQ air quality forecast model, Atmos. Environ., 40, 4894–4905, 2006.

Eder, B., Kang, D., Mathur, R., Pleim, J., Yu, S., Otte, T., and Pouliot, G.: A performance evaluation of the national air quality forecast capability for the summer of 2007, Atmos. Environ., 43, 2312–2320, doi:10.1016/j.atmosenv.2009.01.033, 2009.

Eatough, D. J., Long, R. W., Modey, W. K., and Eatough, N. L.: Semi-volatile secondary organic

aerosol in urban atmospheres: meeting a measurement challenge, Atmos. Environ., 37, 1277–1292, doi:10.1016/S1352-2130(02)01020-8, 2003.
Grover, B. D., Kleinman, M., Eatough, N. L., Eatough, D. J., Hopke, P. K., Long, R. W., Wilson, W. E., Meyer, M. B., and Ambs, J. L.: Measurement of total $PM_{2.5}$ mass (nonvolatile plus semivolatile) with the Filter Dynamic Measurement System tapered element oscillating microbalance monitor, J. Geophys. Res., 110, D07S03, doi:10.1029/2004JD004995, 2005.
Kalman, R. E.: A new approach to linear filtering and prediction problems, J. Basic Eng., 82, 35–45, 1960.
Kang, D., Mathur, R., Rao, S. T., and Yu, S.: Bias adjustment techniques for improving ozone air quality forecasts, J. Geophys. Res., 113, D23308, doi:10.1029/2008JD010151, 2008.
Kang, D., Eder, B. K., Stein, A. F., Grell, G. A., Peckham, S. E., and McHenry, J.: The New England air quality forecasting pilot program: Development of an evaluation protocol and performance benchmark, J. Air Waste Manage., 55, 1782–1796, 2005.
Mathur, R., Yu, S., Kang, D., and Schere, K. L.: 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, 2008.
McHenry, J. N., Ryan, W. F., Seaman, N. L., Coats Jr., C. J., Pudykiewicz, J., Arunachalam, S., and Vukovich, J. M.: A real-time Eulerian photochemical model forecast system: Overview and initial ozone forecast performance in the northeast US corridor, B. Am. Meteorol. Soc., 85, 525–548, 2004.
Mckeen, S., J. Wilczak, G. Grell, et al.: 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, 2005.
McMurry, P. A.: A review of atmospheric aerosol measurements, Atmos. Environ., 34, 1959–1999, 2000.
National Research Council: Research Priorities for Particulate Matter 1: Immediate Priorities and a Long-Range Research Portfolio, Natl. Acad. Press, Washington, DC, 1998.
Otte, T. L., Pouliot, G., Pleim, J. E., et al.: Linking the Eta Model with the Community Multiscale Air Quality (CMAQ) modeling system to build a national air quality forecasting system, Weather Forecast., 20, 367–384, 2005.
Rogers, E., Black, T. L., Deaven, D. G., DiMego, G. J., Zhao, O., Baldwin, M., Junker, N. W., and Lin, Y.: Changes to the operational "early" Eta Analysis/Forecast System at the National Centers for Environmental Prediction, Weather Forecast., 11, 391–413, 1996.

Wilczak, J., McKeen, S., Djalalova, I., Grell, G., Peckham, S., Gong, W., Bouchet, V., Moffet, R., McHenry, J., Lee, P., Tang, Y., and Carmichael, G. R.: 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, 2006.

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

Yu, S., Mathur, R., Schere, K., Kang, D., Pleim, J., Young, J., Tong, D., Pouliot, G., Mckeen, S., and Rao, S. T.: Evaluation of real-time $PM_{2.5}$ forecasts and process analysis for $PM_{2.5}$ 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, 2008.

Table 1. Regional summary of discrete statistics for raw model and KF bias-adjusted daily mean $PM_{2.5}$ forecasts during 2007 cool season.

Type	RMSE ($\mu g/m^3$)	NME (%)	MB ($\mu g/m^3$)	NMB (%)	r
Dom-mod	9.6	59.3	2.5	24.2	0.52
Dom-kf	6.3	39.8	1.2	11.2	0.70
NE-mod	11.4	63.9	4.5	39.8	0.57
NE-kf	7.0	41.7	1.8	16.3	0.68
SE-mod	7.2	44.9	2.1	18.6	0.54
SEkf	5.2	33.4	0.9	8.1	0.66
UM-mod	9.9	56.0	4.2	34.8	0.56
UM-kf	6.3	37.2	1.3	11.2	0.69
LM-mod	9.4	65.2	2.9	28.8	0.39
LM-kf	5.6	40.2	1.3	12.5	0.59
RM-mod	9.4	70.5	2.7	31.1	0.41
RM-kf	5.9	43.8	1.1	12.1	0.68
PC-mod	10.1	52.7	–0.9	–7.4	0.58
PC-kf	7.6	38.9	0.5	4.2	0.75

GMDD

2, 1375–1406, 2009

Table 2. Regional summary of discrete statistics for raw model and KF bias-adjusted daily mean $PM_{2.5}$ forecasts during 2007 warm season.

Type	RMSE ($\mu g/m^3$)	NME (%)	MB ($\mu g/m^3$)	NMB (%)	r
Dom-mod	8.4	46.0	–0.7	–5.4	0.52
Dom-kf	6.3	34.1	0.5	3.8	0.72
NE-mod	8.7	41.1	–1.8	–12.1	0.61
NE-kf	7.1	35.0	0.2	1.5	0.74
SE-mod	9.6	36.9	–4.1	–23.8	0.49
SEkf	7.8	29.3	–0.0	–0.2	0.61
UM-mod	7.3	35.3	–1.8	–11.7	0.62
UM-kf	6.2	30.0	0.0	0.2	0.72
LM-mod	9.1	52.8	–1.1	–8.6	0.30
LM-kf	6.3	37.1	0.8	5.9	0.51
RM-mod	8.2	63.1	1.5	17.0	0.25
RM-kf	5.8	40.8	0.9	10.0	0.48
PC-mod	8.4	59.9	2.3	24.2	0.52
PC-kf	5.4	35.3	1.1	11.4	0.76

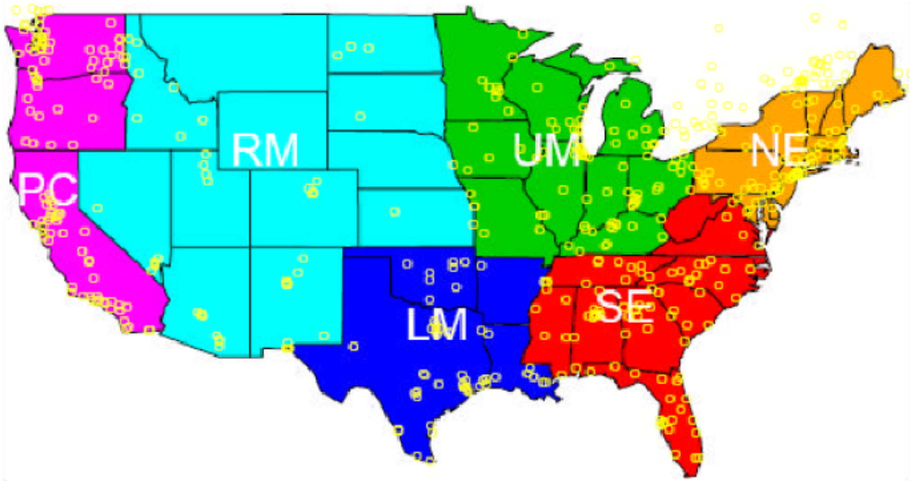

Fig. 1. Forecast domain, analysis sub-regions, and monitoring sites (AIRNOW network). NE: Northeast, SE: southeast, UM: Upper Midwest, LM: Lower Midwest, RM: Rocky Mountains, PC: Pacific Coast.

GMDD
2, 1375–1406, 2009

Assessment of bias-adjusted $PM_{2.5}$ air quality forecasts

D. Kang et al.

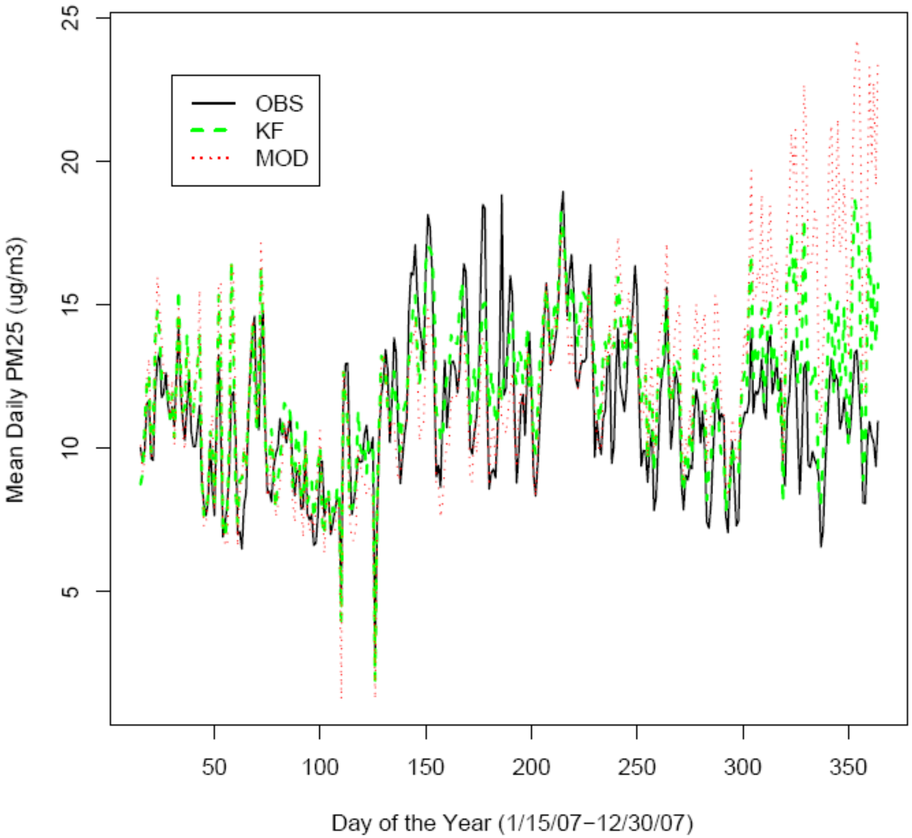

Fig. 2. Time series of observed, raw model forecast, and KF bias-adjusted forecast daily mean $PM_{2.5}$ ($\mu g/m^3$). OBS: observations, KF: Kalman filter bias-adjustment, MOD: raw model.

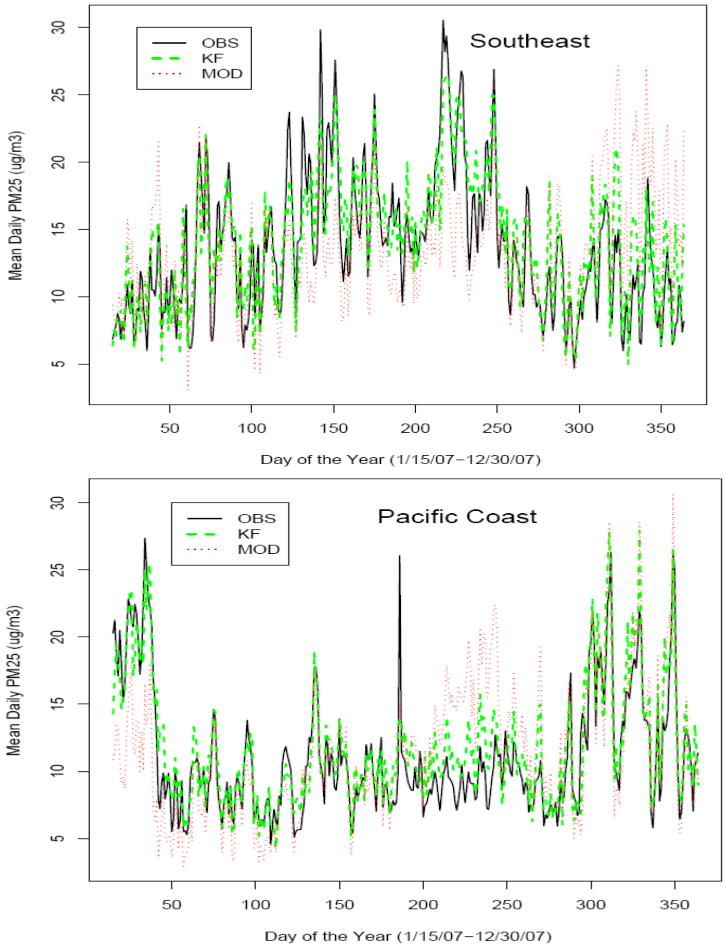

Fig. 3. Time series of observed, raw model forecast, and KF bias-adjusted forecast daily mean $PM_{2.5}$ ($\mu g/m^3$) at Southeast and Pacific Coast.

GMDD

2, 1375–1406, 2009

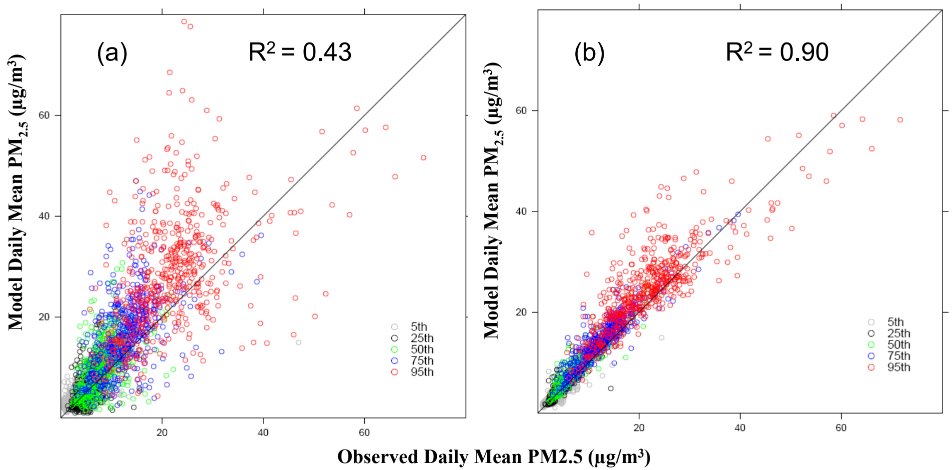

Fig. 4. Scatterplots between forecasts and observations for selected percentiles for the daily mean PM2.5 concentrations ($\mu g/m^3$): **(a)** raw model forecasts, **(b)** Kalman filter-adjusted forecasts.

GMDD
2, 1375–1406, 2009

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

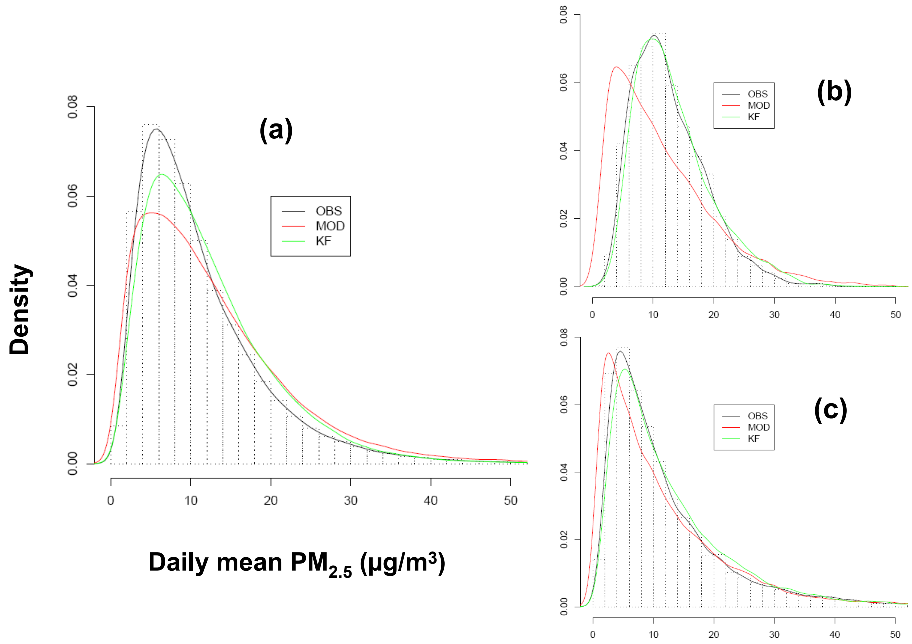

Fig. 5. The histogram of observed and the fitted Gaussian probability density function of observed, raw model forecast, and KF forecast daily mean $PM_{2.5}$ concentrations ($\mu g/m^3$): **(a)** Domain over entire year, **(b)** LM during warm season, and **(c)** PC during cool season.

GMDD
2, 1375–1406, 2009

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀◀ ▶▶
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

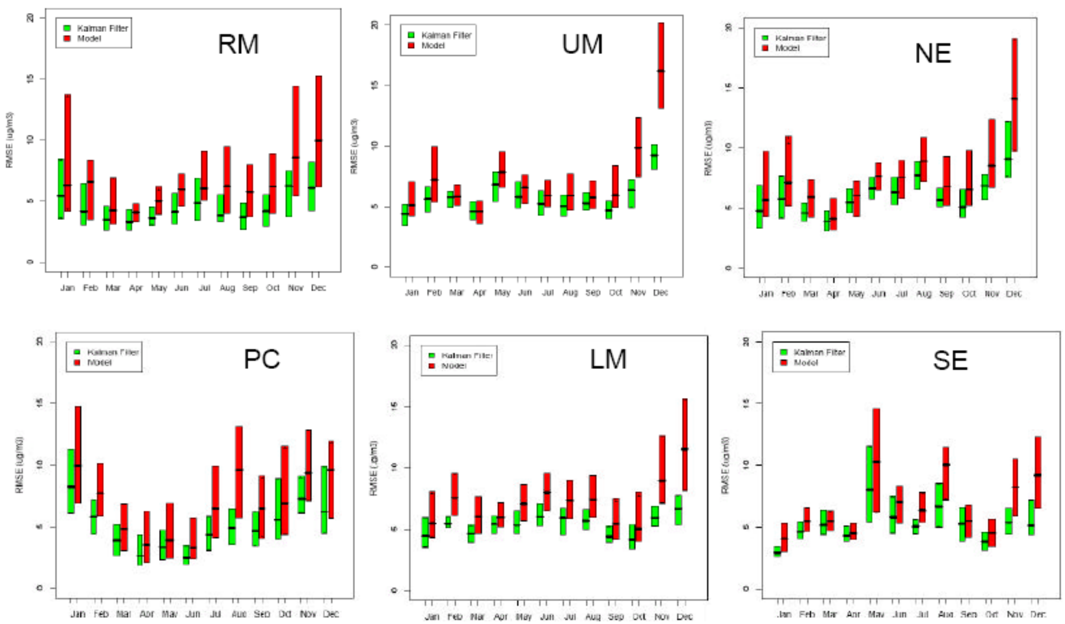

Fig. 6. Monthly box plots (only 25 and 75 percentiles and median values are shown) of RMSE values of the daily mean $PM_{2.5}$ concentrations ($\mu g/m^3$) for the raw model forecasts and KF bias-adjusted forecasts for all the sub-regions.

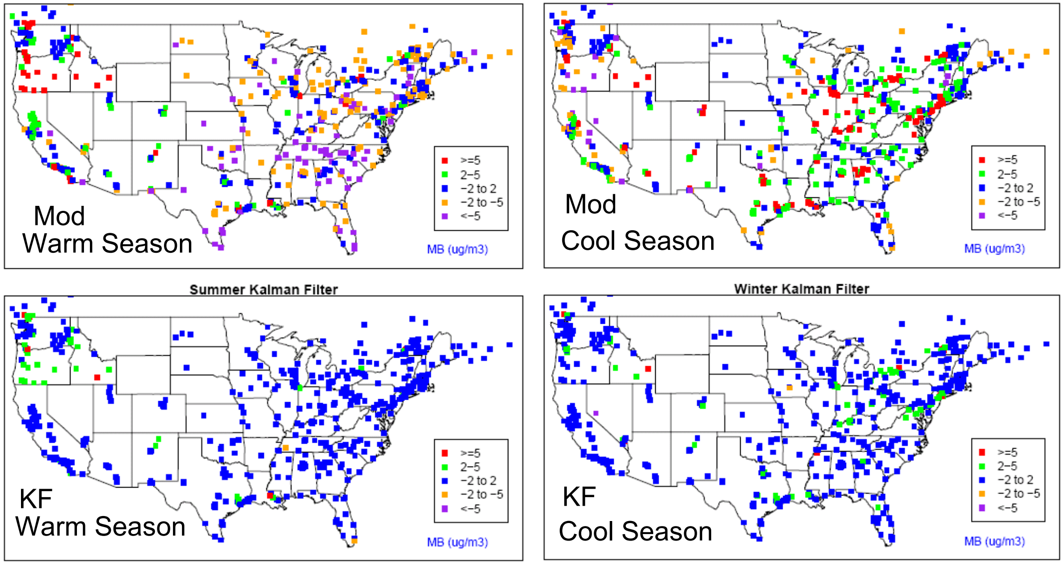

Fig. 7. Mean Bias (MB, $\mu g/m^3$) at each location within the continental US Domain: **(a)** raw model during warm season, **(b)** KF bias-adjustment during warm season, **(c)** raw model during cool season, and **(d)** KF bias-adjustment during cool season.

GMDD

2, 1375–1406, 2009

Assessment of bias-adjusted $PM_{2.5}$ air quality forecasts

D. Kang et al.

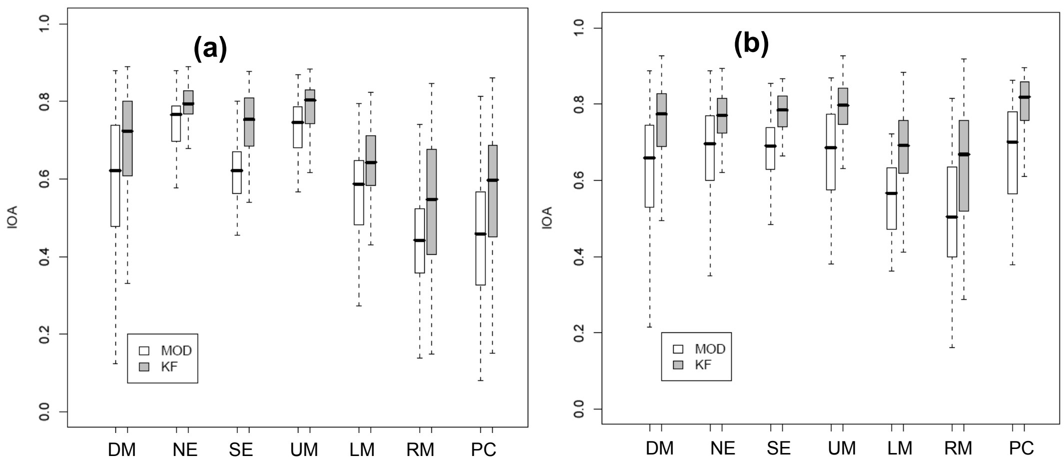

Fig. 8. Box plots of index of agreement (IOA) of daily mean $PM_{2.5}$ ($\mu g/m^3$) for the raw model (MOD) forecasts and KF bias-adjusted forecasts over the domain (DM) and across all subregions during **(a)** warm season and **(b)** cool season.

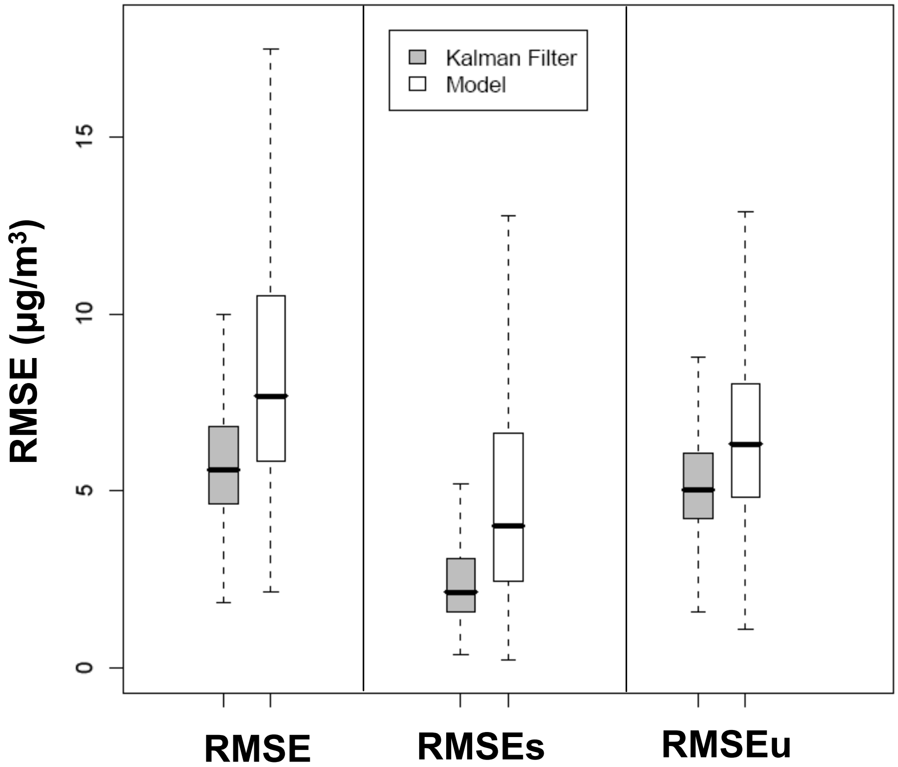

Fig. 9. Box plots of RMSE and decomposed RMSE (systematic, RMSEs; unsystematic, RMSEu) values of the daily mean $PM_{2.5}$ concentrations ($\mu g/m^3$) for the raw model forecasts and KF bias-adjusted forecasts.

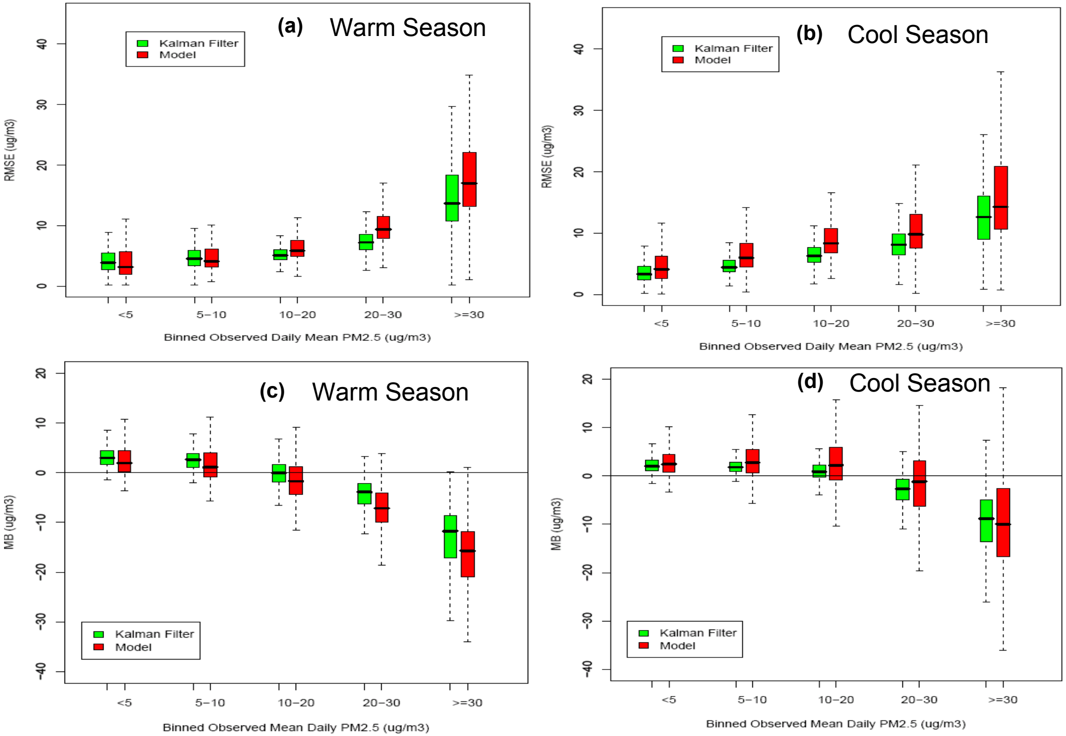

Fig. 10. (**a** and **b**) RMSE and (**c** and **d**) mean bias (MB) values over observed daily mean $PM_{2.5}$ concentration ($\mu g/m^3$) bins for the raw model forecasts and the KF bias-adjusted forecasts. Figure 9a and c for warm season, and Fig. 9b and d for cool season.

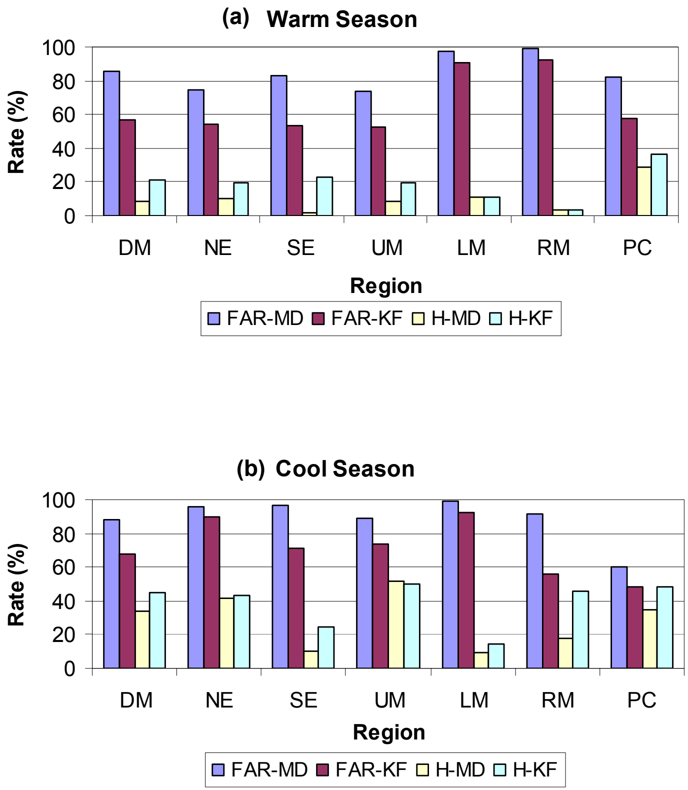

Fig. 11. False alarm ratio (FAR) and hit rate (H) for the daily mean $PM_{2.5}$ forecasts by the raw model and the KF bias-adjustment over the domain (DM) and all the sub-regions during **(a)** warm season and **(b)** cool season: FAR-MD, FAR associated with raw model forecasts; FAR-KF, FAR associated with KF forecasts; H-MD, H associated with raw model forecasts; and H-KF, H associated with KF forecasts.