

Improving NOAA NAQFC PM_{2.5} Predictions with a Bias Correction Approach

JIANPING HUANG,^{a,b} JEFFERY MCQUEEN,^b JAMES WILCZAK,^c IRINA DJALALOVA,^{c,d}
 IVANKA STAJNER,^c PERRY SHAFRAN,^{a,b} DAVE ALLURED,^{c,d} PIUS LEE,^f LI PAN,^{f,g}
 DANIEL TONG,^{f,g} HO-CHUN HUANG,^{a,b} GEOFFREY DIMEGO,^b SIKCHYA UPADHAYAY,^{e,h}
 AND LUCA DELLE MONACHEⁱ

^a *I.M. Systems Group, Inc., Rockville, Maryland*

^b *National Oceanic and Atmospheric Administration/National Centers for Environmental Prediction/Environmental Modeling Center, College Park, Maryland*

^c *NOAA/Earth System Research Laboratory, Boulder, Colorado*

^d *Cooperative Institute for Research in the Environmental Sciences, University of Colorado Boulder, Boulder, Colorado*

^e *NOAA/National Weather Service/Office of Science and Technology Integration, Silver Spring, Maryland*

^f *NOAA/Air Resources Laboratory, College Park, Maryland*

^g *Cooperative Institute for Climate and Satellite, University of Maryland, College Park, College Park, Maryland*

^h *Syneren Technologies Corporation, Arlington, Virginia*

ⁱ *National Center for Atmospheric Research/Research Applications Laboratory, Boulder, Colorado*

(Manuscript received 14 July 2016, in final form 24 October 2016)

ABSTRACT

Particulate matter with an aerodynamic diameter less than or equal to 2.5 μm (PM_{2.5}) is a critical air pollutant with important impacts on human health. It is essential to provide accurate air quality forecasts to alert people to avoid or reduce exposure to high ambient levels of PM_{2.5}. The NOAA National Air Quality Forecasting Capability (NAQFC) provides numerical forecast guidance of surface PM_{2.5} for the United States. However, the NAQFC forecast guidance for PM_{2.5} has exhibited substantial seasonal biases, with overpredictions in winter and underpredictions in summer. To reduce these biases, an analog ensemble bias correction approach is being integrated into the NAQFC to improve experimental PM_{2.5} predictions over the contiguous United States. Bias correction configurations with varying lengths of training periods (i.e., the time period over which searches for weather or air quality scenario analogs are made) and differing ensemble member size are evaluated for July, August, September, and November 2015. The analog bias correction approach yields substantial improvement in hourly time series and diurnal variation patterns of PM_{2.5} predictions as well as forecast skill scores. However, two prominent issues appear when the analog ensemble bias correction is applied to the NAQFC for operational forecast guidance. First, day-to-day variability is reduced after using bias correction. Second, the analog bias correction method can be limited in improving PM_{2.5} predictions for extreme events such as Fourth of July Independence Day firework emissions and wildfire smoke events. The use of additional predictors and longer training periods for analog searches is recommended for future studies.

1. Introduction

Particulate matter with aerodynamic diameter less than or equal to 2.5 μm (PM_{2.5}) and ground ozone (O₃) are the two major air pollutants in the United States. Exposure to high levels of ambient PM_{2.5} may pose significant health risks for people with heart or lung disease, older adults, and children (Brook et al. 2004; Nel 2005). For example, there are about 130 000 cases of

premature mortality attributable to PM_{2.5} pollution each year in the United States (Fann et al. 2012). To protect human health, the U.S. Environmental Prediction Agency (EPA) established the National Ambient Air Quality Standards (NAAQS) for PM_{2.5} in 1997 and lowered the NAAQS in 2006 and 2012, respectively. The current NAAQS for 24-h averaged PM_{2.5} concentration is 35 $\mu\text{g m}^{-3}$ while for annually averaged PM_{2.5} the concentration is 12 $\mu\text{g m}^{-3}$. According to the monitoring reports, many counties in the United States violated the new NAAQS of PM_{2.5} (EPA 2015). Thus, it is important to provide numerical forecast guidance as a

Corresponding author e-mail: Jianping Huang, jianping.huang@noaa.gov

basis for alerting the public to avoid or reduce exposure to unhealthy levels of $\text{PM}_{2.5}$.

The goal of the National Oceanic and Atmospheric Administration (NOAA) National Air Quality Forecasting Capability (NAQFC) is to provide timely and accurate operational numerical guidance for surface O_3 and $\text{PM}_{2.5}$ concentrations. The NAQFC was established by NOAA in partnership with the EPA to provide ozone and particulate matter pollutant forecasts. The capability was initially deployed in 2004 to provide surface ozone operational forecast guidance for the northeastern United States (Otte et al. 2005). The capability for providing surface ozone operational forecasts was expanded to the conterminous United States (CONUS) in 2007, Hawaii in 2009, and Alaska in 2010 (Stajner et al. 2012). Nationwide real-time developmental $\text{PM}_{2.5}$ forecast guidance has been provided from the operational NAQFC system since January 2015. This guidance exhibits substantial seasonal biases: $\text{PM}_{2.5}$ is usually underpredicted in summer and overpredicted in winter as compared with AirNow observational data (Stajner et al. 2012; Lee et al. 2017). Uncertainties in emission inventories, meteorological inputs, and air quality models may contribute to the biases in model predictions of airborne chemical species and particulate matter. Improving NAQFC $\text{PM}_{2.5}$ forecast skill is imperative to ensuring its readiness for operational use.

While many research efforts have been devoted to improving the core chemical, meteorological, and emissions model components, postprocessing approaches such as bias correction provide a complementary pathway to refine forecast products. Bias correction approaches range from complex statistical regression techniques to subjective corrections. Bias correction methods have been widely used in numerical weather forecasting (e.g., Glahn and Lowry 1972; Hamill and Whitaker 2006; Delle Monache et al. 2011, 2013; Cui et al. 2012; Durai and Bhradwaj 2014; Glahn 2014; Jo and Ahn 2015; Zhu and Luo 2015) and air quality forecasting (e.g., Delle Monache et al. 2006, 2008, 2011; Kang et al. 2008, 2010; Wilczak et al. 2006; Djalalova et al. 2010). These studies demonstrate that both numerical weather and air quality forecasts are improved substantially compared with model raw forecasts. Recently, Djalalova et al. (2015) evaluated bias correction methods for improving the NAQFC $\text{PM}_{2.5}$ predictions over the CONUS. They tested several postprocessing techniques, which include a 7-day running mean bias correction, a Kalman filter (KF) applied to standard time series data, an analog ensemble, KF applied to the series of ordered analog forecasts (KFAS), and KF applied to analog time series (KFAN).

All of these bias correction approaches show strong improvement compared with the NAQFC raw forecasts.

In this study, an analog ensemble bias correction approach is integrated into the operational NAQFC real-time system for improving the Community Multiscale Air Quality (CMAQ) model predictions of $\text{PM}_{2.5}$. In the study by Djalalova et al. (2015), a full year's worth of historical model predictions was used to identify 10 analog cases for each forecast, which were then used to determine $\text{PM}_{2.5}$ forecast biases. Given the time limitations of real-time forecast product delivery, the configuration used by Djalalova et al. (2015) for bias corrections needs to be optimized without degrading performance. The goals of this study are to integrate the bias correction approach into the NAQFC system and to identify a practical configuration for bias correction using the most recent version of NAQFC. The bias correction results are compared with the model raw forecasts and evaluated with EPA AirNow observational data (<http://www.airnow.gov>). The performance of the analog ensemble bias correction approach is evaluated during different seasonal months in 2015. Furthermore, the performance during several high $\text{PM}_{2.5}$ concentration events such as wildfire episodes is discussed to demonstrate the challenges encountered when using analog ensemble bias correction during rare high-impact events.

2. Methods

a. NAQFC and configurations

The NAQFC is an offline meteorology–chemistry coupling forecasting system. The NOAA North American Model Forecast System (NAM) Nonhydrostatic Multiscale Model with Arakawa B grid staggering (NMMB; Janjić and Gall 2012) is linked with the EPA's CMAQ model (Byun and Schere 2006) to provide predictions of spatially and temporally varying concentrations of gaseous and aerosol air pollutants for the United States. Currently, the NAQFC provides twice-daily 48-h forecasts at 0600 and 1200 UTC for the CONUS, Alaska, and Hawaii.

As illustrated in Fig. 1, NMMB provides hourly meteorological inputs to drive CMAQ. The NMMB provides 84-h operational weather forecast guidance for the United States at a horizontal resolution of 12 km. The EPA's CMAQ V4.6 with the Carbon Bond–2005 (CB05) gas-phase chemical mechanism and aerosol module version 4 (AERO-IV) has been modified to provide updated NAQFC operational ozone and experimental $\text{PM}_{2.5}$ predictions since January 2015. The NAQFC produces ozone and $\text{PM}_{2.5}$ predictions for the CONUS, Alaska, and Hawaii domains at 12-km

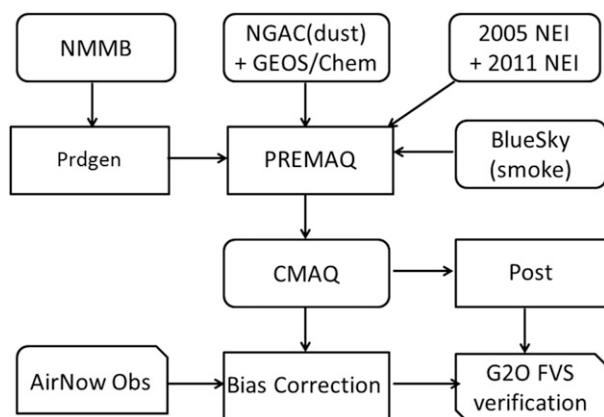


FIG. 1. Flowchart of the NAQFC in the United States.

horizontal grid spacing. The Prdgen and PREMAQ (preprocessor of CMAQ) customized interface processors handle horizontal map projection transformation from the NMMB B grid to the CMAQ C grid and vertical level coupling from the NMMB's hybrid sigma-pressure layers (i.e., sigma layers in the bottom and pressure layers in the top) to the CMAQ's sigma layers, respectively (Otte et al. 2005). The PREMAQ processor has been modified from the Meteorology-Chemistry Interface Processor (MCIP) of the CMAQ modeling system (Otte and Pleim 2010) by adding several new features. In particular, PREMAQ recalculates several important meteorological input fields such as the planetary boundary layer (PBL) height, eddy diffusivity, and cloud parameters from the NMMB outputs. It also computes deposition velocity, photolysis rate, and emission rates for CMAQ.

The NOAA Environmental Modeling System (NEMS) Global Forecast System (GFS) Aerosol Component (NGAC) provides dynamic lateral boundary conditions of dust-related aerosol species to the CMAQ runs. The simulations from the Goddard Earth Observing System (GEOS) with the Chemistry Component (GEOS/Chem) modeling system are used to generate lateral boundary conditions of gas-phase and other aerosol-phase chemical species to the CMAQ.

The emission inputs for NAQFC are processed in two different ways, depending on the nature of the emission sources and their sensitivity to meteorology (Pan et al. 2014; Tong et al. 2015). Anthropogenic sources including area, mobile, and point sources are obtained from North American environmental agencies. The U.S. emission sources are based on a mixture of the EPA National Emission Inventories (NEI) for 2005 and 2011. Most sectors in NEI 2011 are used in this study except for mobile sources and a few area sources (e.g., ocean-going ship emissions) that are associated with high

uncertainties or require inline emission modeling capability, which is not used in this version of CMAQ. Anthropogenic sources for the Canadian part of the domain are based on the 2006 Emission Inventories compiled by Environment Canada, and sources for the Mexican part of the domain come from the 2012 Mexico National Emissions Inventories. These inventory data are processed using the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system (Houyoux et al. 2000) to represent monthly, weekly, diurnal, and holiday/nonholiday variations that are specific for each year. Both wind-blown dust and wildfire emissions are included in the 2015 operational NAQFC system to account for their contributions to $PM_{2.5}$ predictions (Lee et al. 2017). For the wildfire smoke emissions, fire points and smoke plume locations are identified by the NOAA/National Environmental Satellite, Data, and Information Service (NESDIS) Hazard Mapping System (HMS) from satellite retrievals and human analysis (Ruminski et al. 2006). The HMS fire smoke products are processed by the U.S. Forest Service BlueSky framework modeling system (O'Neill et al. 2009; Larkin et al. 2009) to produce near-real-time wildfire smoke emissions for the CMAQ.

This study is focused on the NAQFC CONUS domain, which covers the CONUS, as well as parts of southern Canada and northern Mexico. There are 35 σ vertical levels that extend from the surface to 100 hPa, with the first 14 layers within the lowest 2 km of the atmosphere. The first layer of CMAQ is defined at the height of 39 m above ground level (AGL). The photolysis rate of organic nitrate (NTR) is increased by 10-fold within the CB05 gas-phase chemical mechanism to accelerate NTR removal (Saylor and Stein 2012; Carty et al. 2015). The modification typically shortens the predicted life of NTR in CMAQ from about 1 week to approximately 1 day (Pan et al. 2014). This reduces the overprediction of surface O_3 and has a minor impact on $PM_{2.5}$ prediction. A minimum PBL height of 50 m is employed to avoid excessive suppression of vertical diffusive mixing. More details about modifications to CMAQ and updates to emission inventories were given by Lee et al. (2017).

b. Analog ensemble bias correction

The analog ensemble approach, originally developed for improving numerical weather predictions, is integrated into the NAQFC system for $PM_{2.5}$ forecast bias correction. The analog ensemble method is based on the assumption that, if the climate is relatively stable, model forecast errors in past similar weather scenarios (or analogs) can be used to statistically correct current numerical forecasts (Hamill and Whitaker 2006). The key

to this approach is in determining a suitable metric for identifying analogs from the historical dataset. The metric used here follows [Delle Monache et al. \(2011\)](#):

$$\|A_{t'}, F_t\| = \sum_{i=1}^{N_v} \frac{w_i}{\sigma_{f_i}} \sqrt{\sum_{j=-\tilde{t}}^{\tilde{t}} (F_{i,t+j} - A_{i,t'+j})^2}, \quad (1)$$

where F_t is the forecast at the future time t ; $A_{t'}$ is an analog forecast at the past time t' ; N_v is the number of variables that are used for the analog search ($N_v = 4$ in this study); w_i and σ_{f_i} represent the weight and standard deviation of the i th variable, respectively; \tilde{t} is half of the time window over which the metric is computed ($\tilde{t} = 1$ h in this study); and $A_{i,t'+j}$ and $F_{i,t+j}$ represent the analog and forecast for the i th variable at time $t' + j$ and $t + j$, respectively. Following the study of [Djalalova et al. \(2015\)](#), PM_{2.5}, 2-m temperature, 10-m wind speed, and 10-m wind direction with the same weight are used in the calculation of the metric with Eq. (1).

Analog ensemble bias correction is accomplished through a multiple-step process. First, the NAM model's meteorological variables (e.g., temperature, wind speed/direction) and the CMAQ model's air quality variables (i.e., PM_{2.5}) are interpolated to the AirNow observational sites to form the set of analog predictors. Second, analog members are identified from past forecast time series based on the metric calculated with Eq. (1) and, then, are ranked according to their similarity with the current forecast. Third, forecast biases are computed between the analog ensemble mean at the AirNow observational sites and then spread to the entire CMAQ grids. The spreading technique is based on an eight-pass Barnes-type iterative objective analysis scheme, which is described in detail by [Djalalova et al. \(2015\)](#). The last step is to correct the future CMAQ raw forecasts with the historical analogs' forecast biases across the entire CMAQ grid.

It is noted that the length of the training period and the number of analog ensemble members are the two factors with substantial impacts on the bias correction results. This study evaluates the practical training period and the number of analog ensemble members for the bias-corrected NAQFC PM_{2.5} prediction.

c. Evaluation protocol

The NCEP Verification System (NVS) was originally developed for evaluating numerical weather prediction (NWP) model performance and modified for evaluation of the NAQFC operational predictions of surface ozone and experimental predictions of surface PM_{2.5}. The NVS comprises four parts: editbufr, prepfits, gridobs, and the Forecasting Verification System (FVS) (see [Fig. 2](#)). Among these parts, editbufr reads and retains the

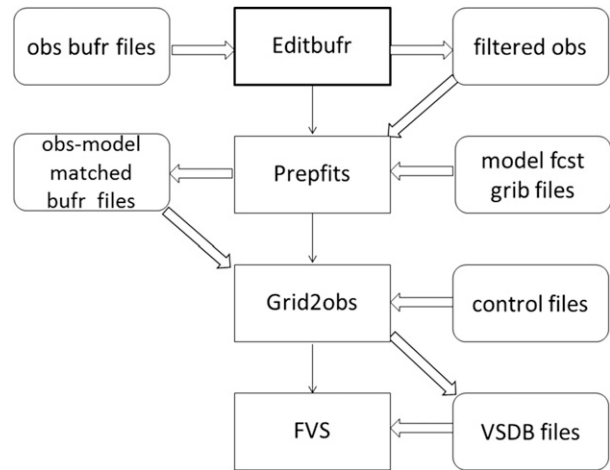


FIG. 2. Flowchart of the NCEP Grib2obs Verification System used for NAQFC performance evaluation.

observations from prepbufr files that contain point observations and quality control information, prepfits interpolates model forecast data to the AirNow observational sites, and grid2obs generates a series of Verification Statistics Data Base (VSDB) files, which include partial sums for the calculation of various statistics.

The FVS is used to compute traditional statistics including root-mean-square error (RMSE), bias, and correlation coefficients, and forecast skill scores like critical success index (CSI), hit rate, probability of detection (POD), and false alarm rate (FAR). The forecast skill scores are defined as follow ([Wilks 1995](#), 260–265):

$$\text{CSI} = \frac{a}{a + b + c}, \quad (2)$$

$$\text{hit rate} = \frac{a}{a + b}, \quad (3)$$

$$\text{POD} = \frac{a}{a + c}, \quad \text{and} \quad (4)$$

$$\text{FAR} = \frac{b}{a + b}, \quad (5)$$

where a denotes the number of occurrences when both the forecast and the observed are above a given threshold (i.e., both are “yes”), b represents the number of occurrences when the forecast is above but the observed below the given threshold (i.e., forecast is “yes” but observed is “no”), c denotes the number of occurrences of the forecast being below but the observed above the given threshold (i.e., forecast is “no” but observed is “yes”), and d denotes occurrences where both the forecast and observed values are below the given threshold (i.e., both are “no”). In this study, the AirNow hourly mean surface PM_{2.5} observational data

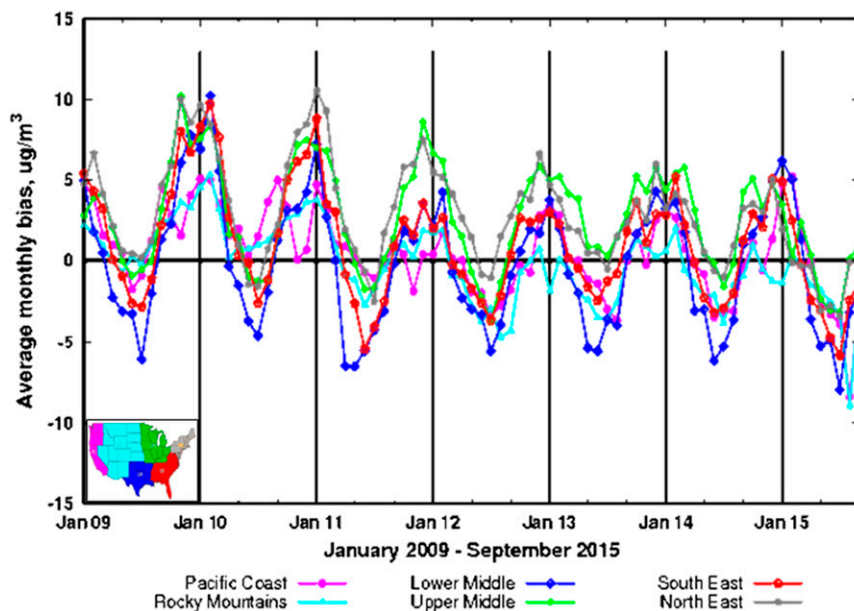


FIG. 3. Monthly mean forecast biases of the NAQFC in different subregions over the CONUS from January 2009 through September 2015.

at 551 sites are used to evaluate the NAQFC performance on $PM_{2.5}$ predictions. The evaluated forecast parameters include hourly mean, 24-h average, and daily maximum 1-h average $PM_{2.5}$ values. Eight thresholds are employed in the calculations of skill scores for $PM_{2.5}$. They include 5, 10, 12, 15, 20, 25, 30, and $35 \mu g m^{-3}$.

3. Evaluation of the NAQFC $PM_{2.5}$ predictions

The NAQFC monthly mean $PM_{2.5}$ forecast biases for six different subregions of the CONUS domain from January 2009 to September 2015 are shown in Fig. 3. The 48-h forecasts at the 0600 and 1200 UTC cycles each day are included in the calculation. The subregions include the Pacific Coast, the Rocky Mountains, the Lower Middle, the Upper Middle, the Southeast, and the Northeast. The subregions are indicated by different colors in the map included in Fig. 3. Substantial seasonal forecast biases persisted over the past several years. The $PM_{2.5}$ results were overpredicted in late autumn (e.g., November) and winter (i.e., December–February) but underpredicted in summer (i.e., June–August). The monthly mean forecast biases ranged from about $-9 \mu g m^{-3}$ in summer to about $10 \mu g m^{-3}$ in winter.

There are multiple likely reasons for the NAQFC $PM_{2.5}$ underpredictions in summer. The major plausible reasons causing underpredictions of $PM_{2.5}$ include 1) the underestimate of primary $PM_{2.5}$ emissions, 2) outdated mobile emission inventories, 3) incorrect representation of secondary organic aerosols (SOAs), 4) constant

climatological lateral boundary profiles except for dust-related aerosol species, 5) uncertainty of meteorological inputs related to meteorology–chemistry coupling (e.g., overpredicting the planetary boundary layer height and eddy diffusivity), 6) exclusion of transboundary-transported wildfire/smokes from Canada or Mexico, and 7) the outdated BlueSky fire emission processing system and the wildfire emissions not being used properly in the CMAQ.

The reasons causing the forecast biases vary from one region to another. Our analyses indicate that SOAs were not well simulated over the southeast United States (Carlton et al. 2010); organic carbon (OC) and elementary carbon (EC) were underestimated over the western United States; ammonium was underestimated over the Rocky Mountains, the Lower Middle, and the Upper Middle; and that wildfire/smoke emissions were still underestimated over the Northwest regions by the NAQFC during summer. The dust-storm-related emissions were not treated appropriately in the NAQFC over the Southwest region, such as in Arizona and Nevada, during spring. Furthermore, the fugitive dust emissions were significantly overestimated during winter. Contributions from source groups and regions to the ambient levels of primary and secondary $PM_{2.5}$ can be evaluated further through model analysis tools such as the source apportionment method (Kwok et al. 2013).

In the current version of NAQFC, only NGAC-predicted dust-related species are used to generate the dynamically varying lateral boundary conditions for the

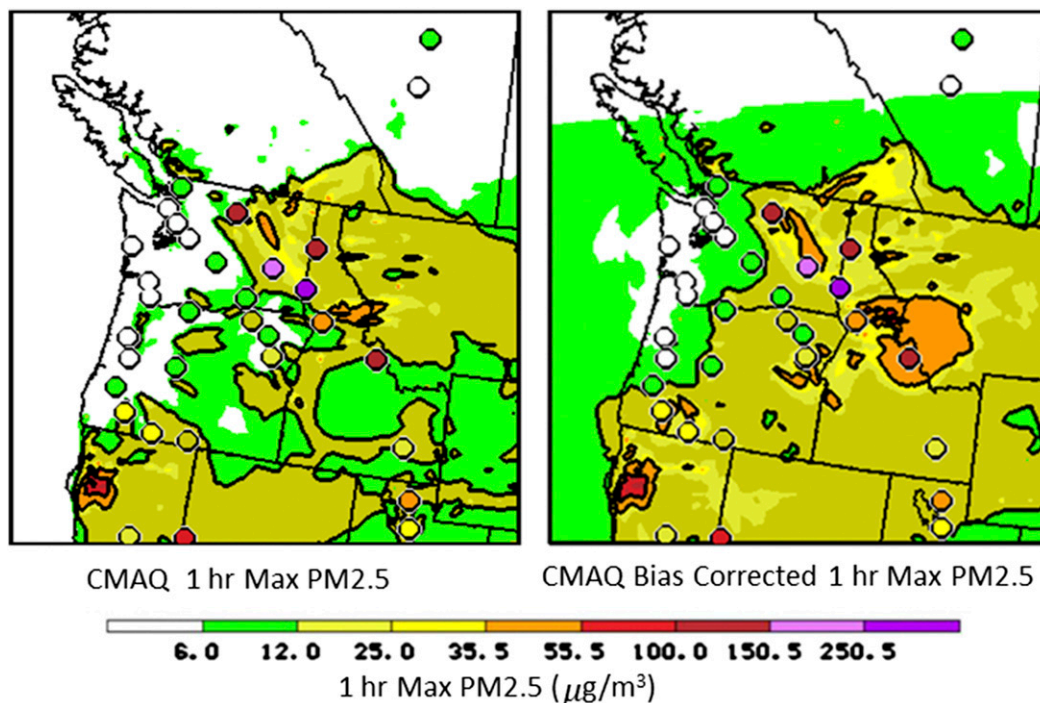


FIG. 4. Overlaid plots of day 1 maximum 1-h average $\text{PM}_{2.5}$ ($\mu\text{g m}^{-3}$) over the northwestern United States on 22 Aug 2015 for the 0600 UTC cycle run; shown are base runs (left) without and (right) with bias correction (background colors, forecasts; dotted points, AirNow observations).

NAQFC. Transboundary transport of wildfire and smoke from Canada or biomass burning from Mexico are not included. This could be another important reason causing the $\text{PM}_{2.5}$ underpredictions during the wildfire/smoke active season. The full aerosols predicted by the recently upgraded NGAC including wildfire smoke and dust will be used to generate lateral boundary conditions for the NAQFC predictions during the future implementation.

Uncertainty of meteorological inputs (e.g., PBL height) is another important factor in the winter overpredictions. We note that positive forecast biases in the winter months had in general decreased over the past several years. For example, the forecast bias decreased from about $10.0 \mu\text{g m}^{-3}$ in January 2009 to around $5.0 \mu\text{g m}^{-3}$ in January 2015. The improvement of $\text{PM}_{2.5}$ predictions in winter was related to advancements of the meteorological model (i.e., NMMB) and better estimates of anthropogenic emissions that were made over the past several years. The major changes and improvements of NMMB were described by Janjić and Gall (2012). Further details about emissions updates were given by Lee et al. (2017) and Tong et al. (2015).

It is noted that the underpredictions were worse in summer 2015 than in the preceding years. Several factors were responsible for this larger forecast bias. The

first possible factor was that more and larger wildfires occurred over the CONUS, especially in the northwestern United States and Canada in 2015 (see the total burned areas online: https://www.nifc.gov/fireInfo/fireInfo_stats_totalFires.html). The fire emissions were still largely underestimated in the NAQFC although the BlueSky fire emission modeling system with near-real-time satellite-based fire information was implemented in 2015. An extreme example is shown in Fig. 4. The observed $\text{PM}_{2.5}$ in eastern Washington was larger than $250 \mu\text{g m}^{-3}$ (indicated by a dark purple circle) whereas the predicted $\text{PM}_{2.5}$ was less than $35 \mu\text{g m}^{-3}$. In addition, contributions of wildfire smoke from outside the CMAQ domain were not considered.

The wildfire emissions used in the NAQFC were provided by the U.S. Forest Service BlueSky fire emissions modeling system (Larkin et al. 2009). The BlueSky operational system used the previous day's NOAA/NESDIS HMS fire information such as fire locations and durations for the emission calculation during the 0600 UTC cycle run. The second factor could be related to the incomplete inclusion of fire emission sources. For instance, prescribed biomass burning such as debris clearing and agricultural fire emissions were removed from the emission inventories to avoid double counting with the implementation of dynamically projecting these

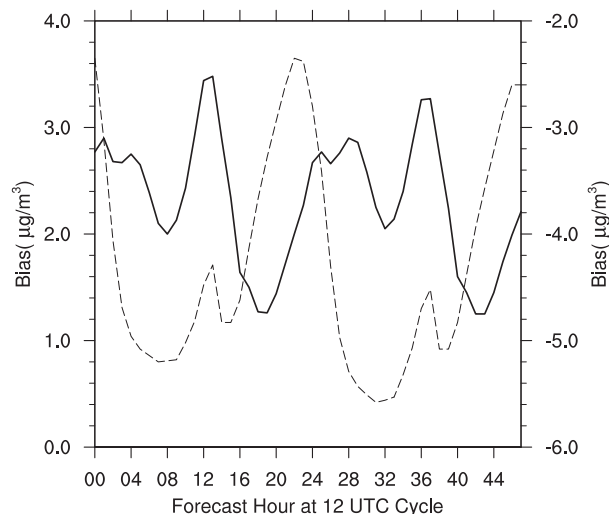


FIG. 5. Domain-averaged forecast bias diurnal variation of NAQFC over the CONUS during January and July 2015 (solid line, January; dashed line, July).

emissions using the HMS-BlueSky algorithm. Moreover, several other factors may cause uncertainties in the emissions results, which include the plume rise calculation algorithm, meteorological inputs, and the detection of wildfire smoke under cloudy conditions. The primary goal of this study is to evaluate whether bias correction approaches can improve CMAQ $PM_{2.5}$ predictions given the uncertainties in the emissions and meteorology.

Another feature worth noting is that the NAQFC forecast biases showed different diurnal variation patterns between winter and summer. Figure 5 indicates the average diurnal forecast biases over the CONUS for the 1200 UTC cycle CMAQ runs in January and July 2015. In January the monthly mean forecast bias ranged from 1.3 to $3.5 \mu\text{g m}^{-3}$ with the maximum forecast bias at forecast hour 14, or 0200 UTC [i.e., 2100 eastern standard time (EST)]. The maximum overpredictions during the nighttime were usually linked with underpredictions of the PBL heights or the setting of typical minimum PBL heights in the simulations. On the other hand, the NAQFC showed negative $PM_{2.5}$ forecast biases from approximately -5.5 to $-2.5 \mu\text{g m}^{-3}$ during July 2015. The worst forecast bias occurred during the daytime (i.e., the forecast hour 8, or 2000 UTC, or 1500 EST). Underestimation of wildfire smoke emissions could be one of the main reasons causing the underpredictions. However, such underpredictions during the nighttime could be compensated for by other factors such as meteorological inputs. Thus, further investigations are needed in the future to identify the specific factors, including emissions, chemistry, and meteorological inputs,

TABLE 1. Summary of the NAQFC parallel run and three cases of analog ensemble bias correction. Base case represents the CMAQ raw forecast. BC5E6M denotes the bias correction case using five ensemble members and a 6-month training period; BC5E12M and BC10E12M represent the bias correction cases with 12-month training periods and with 5 and 10 ensemble members, respectively.

Expt	No. of analog ensemble members	Length of training period (months)
Base case	None	None
BC5E6M	5	6
BC5E12M	5	12
BC10E12M	10	12

and to quantify their relative contributions to the forecast biases.

4. Testing of analog ensemble bias correction with different configurations

The length of the training period and the number of analog ensemble members are the two variable parameters for the analog ensemble bias correction approach (Djalalova et al. 2015). As summarized in Table 1, three sensitivity experiments were conducted to assess the impact of the training period length and the number of analog ensemble members on bias correction performance, and to identify a practical configuration for real-time operational applications.

Different bias correction configurations were evaluated for each of the following four months: July, August, September, and November in 2015. These months were chosen to evaluate the analog ensemble bias correction approach under different air quality scenarios. Among them, July was the month in which $PM_{2.5}$ was significantly underestimated. August was the month during which wildfires were very active, especially in the northwestern United States. September was the transition period when the NAQFC $PM_{2.5}$ predictions began to show positive biases. Finally, November was the month during which the NAQFC started to show significant $PM_{2.5}$ overprediction, especially over the eastern United States (EUS), which comprises the Lower Middle, Upper Middle, Southeast, and Northeast (shown in Fig. 3).

a. $PM_{2.5}$ forecast guidance and bias correction in July 2015

Comparisons of $PM_{2.5}$ among CMAQ raw and bias-corrected forecast guidance for different analog ensemble bias correction configurations over the western United States (WUS, consisting of the Pacific Coast and Rocky Mountain regions; Fig. 3) and EUS during July 2015 are shown in Fig. 6. Both WUS and EUS are

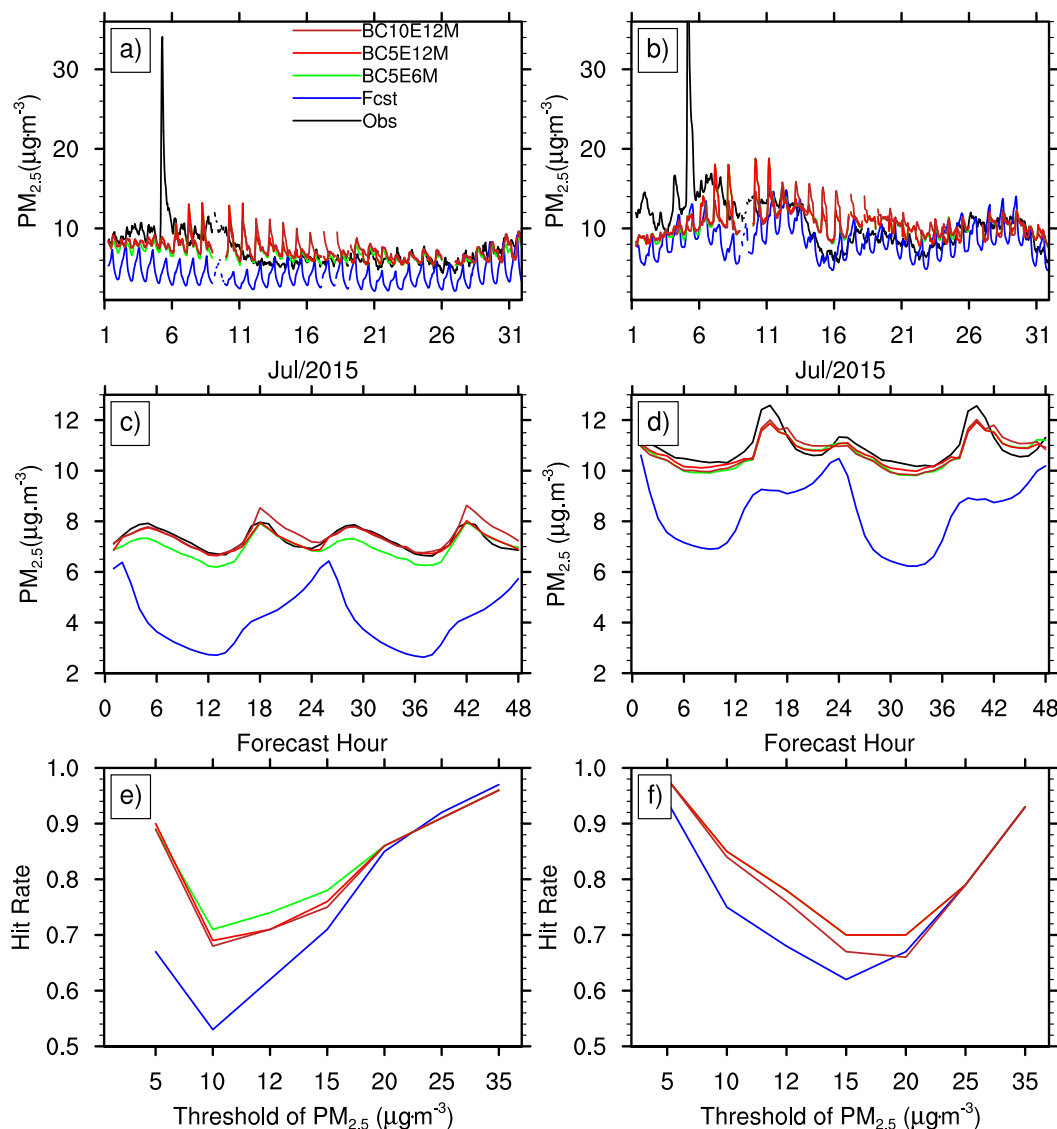


FIG. 6. Comparisons among bias-corrected, raw forecast, and observed $PM_{2.5}$ results for (a) time series of the 24th forecast hour during the 1200 UTC cycle run over the WUS, (b) hourly time series over the EUS, (c) mean diurnal variation over the WUS, (d) mean diurnal variation over the EUS, (e) hit rates for daily maximum 1-h $PM_{2.5}$ concentration over the WUS, and (f) hit rates for daily maximum 1-h $PM_{2.5}$ concentration over the EUS during July 2015 (BC5E6M, bias corrected with five members and a 6-month training period; BC5E12M, bias corrected with five members and a 12-month training period; BC10E12M, bias corrected with 10 members and a 12-month training period; Fcst, forecast; and Obs, observation).

verified separately as a result of the large discrepancy in $PM_{2.5}$ emission sources. Here, the $PM_{2.5}$ was substantially underpredicted by the NAQFC over both regions, but more strongly over the WUS. A large spike in $PM_{2.5}$ was observed over both EUS and WUS on 5 July due to Independence Day fireworks. The hourly averaged $PM_{2.5}$ rose sharply to approximately $47 \mu\text{g m}^{-3}$ over the EUS and approximately $35 \mu\text{g m}^{-3}$ over the WUS during the evening of 4 July and returned to the normal levels late in the day on 5 July. The NAQFC guidance failed to predict the event because the

firework emissions were not included in the current emission inventory. The approach was not able to capture the event even though July data from the previous year were included in the analog search. This is because the ensemble members were dominated by analogs from other days instead of 4 July. As illustrated in Eq. (1), the metric calculation relied on three meteorological factors (i.e., 2-m temperature, and 10-m wind direction and wind speed), but the meteorological conditions on 4 and 5 July in the previous year may not be similar to those in 2015.

TABLE 2. A comparison of standard statistical evaluation parameters and forecast skill scores with a $\text{PM}_{2.5}$ threshold above $12 \mu\text{g m}^{-3}$ among the three bias correction scenarios and the base case over the EUS and WUS during August 2015, where C_r is the correlation coefficient.

	CMAQPARA		BC5E6M		BC5E12M		BC10E12M	
	WUS	EUS	WUS	EUS	WUS	EUS	WUS	EUS
C_r	0.43	0.36	0.49	0.48	0.47	0.47	0.47	0.47
Bias	-7.14	0.42	-2.63	0.31	-1.31	0.33	-3.4	0.37
RMSE	16.46	5.98	14.58	5.07	14.67	5.01	14.87	5.01
CSI	0.20	0.24	0.49	0.28	0.56	0.31	0.42	0.23
Hit rate	0.27	0.74	0.28	0.75	0.31	0.76	0.29	0.75
POD	0.21	0.23	0.58	0.33	0.69	0.38	0.48	0.27
FAR	0.00	0.04	0.06	0.02	0.07	0.02	0.03	0.01

Overestimates of $\text{PM}_{2.5}$ were seen in the bias-corrected guidance for several days following 5 July. The magnitude of the overestimate was reduced when using a larger number of analog ensemble members, but the duration of the overprediction was longer with an increasing number of analog ensemble members (see Figs. 6a,b). Most likely 5 July was selected as one of the analog ensemble members in the following several days and the magnitude of the overprediction was reduced when more analog ensemble members were used or when a longer training dataset was available.

Monthly mean diurnal variations of $\text{PM}_{2.5}$ over the WUS and EUS in July 2015 are presented in Figs. 6c and 6d, respectively. The raw NAQFC guidance failed to simulate $\text{PM}_{2.5}$ diurnal variation patterns in terms of the magnitude and temporal phase for both subregions. The worst underpredictions by raw forecasts were found during the daytime, whereas the fewest underpredictions appeared at night. All configurations of the bias correction forecast guidance show excellent agreement with observations for both magnitude and phase.

Comparison of hit rates among three scenarios and the base case at different thresholds are presented in Figs. 6e and 6f. A large increase of hit rate appeared at thresholds below $20.0 \mu\text{g m}^{-3}$ over the WUS whereas a relatively small increase occurred at the thresholds below $15 \mu\text{g m}^{-3}$ over the EUS. Quantitative comparisons of hit rate and other statistical evaluation parameters among the base case and the three bias correction experiments are shown in Table 2. All of the three bias correction experiments show larger improvements over the WUS than over the EUS. This is because the model raw prediction biases over the WUS are significantly larger than those over the EUS during wildfire/smoke events. FAR is small and shows less change while both the hit rate and POD are improved. In addition, it is noticed that the reduction in the RMSE is much less than that of the forecast biases.

Overall, the monthly mean diurnal variations and forecast skill scores for the threshold of $15.0 \mu\text{g m}^{-3}$ and lower are improved substantially (see Figs. 6e,f). However, it is still a challenge for extreme events like the 4 July fireworks case and for the thresholds around $35.0 \mu\text{g m}^{-3}$ or above in July.

b. $\text{PM}_{2.5}$ forecast guidance and bias correction in August 2015

August was an active time for wildfires across the WUS and $\text{PM}_{2.5}$ air quality model predictions can be challenging given the uncertainties in wildfire smoke emissions. Several wildfire events were observed over the northwestern United States, with the largest events occurring on 22–25 August. The observed hourly averaged $\text{PM}_{2.5}$ results reached up to approximately $40 \mu\text{g m}^{-3}$ over the WUS on 24 August (Fig. 7a). However, the NAQFC raw $\text{PM}_{2.5}$ predictions were around $8 \mu\text{g m}^{-3}$ or less over the WUS. The $\text{PM}_{2.5}$ values increased to about $16 \mu\text{g m}^{-3}$ with bias correction using five members and longer training periods (i.e., 6 or 12 month), but did not reach the observed $\text{PM}_{2.5}$ levels.

Three causes for the underpredictions of $\text{PM}_{2.5}$ -associated wildfires by the analog ensemble bias correction approach are discussed. First, the variables (i.e., $\text{PM}_{2.5}$, 2-m temperature, and 10-m wind speed and wind direction) used to determine the analogs may not represent the most important indicators for wildfire episodes while some important fire-related indicators may not be included in the analog search. For example, high concentrations of OC are usually associated with biomass burning, and the ammonium sulfate $[(\text{NH}_4)_2\text{SO}_4]$ mainly comes from anthropogenic sources (Hand et al. 2011). Thus, the ratio of OC to ammonium sulfate is a good indicator for distinguishing wildfire sources of $\text{PM}_{2.5}$ from anthropogenic emissions. However, this parameter was not included in the analog search. Another important factor is that more and larger wildfires occurred over the CONUS, as discussed above. If the training period is very different from the period for

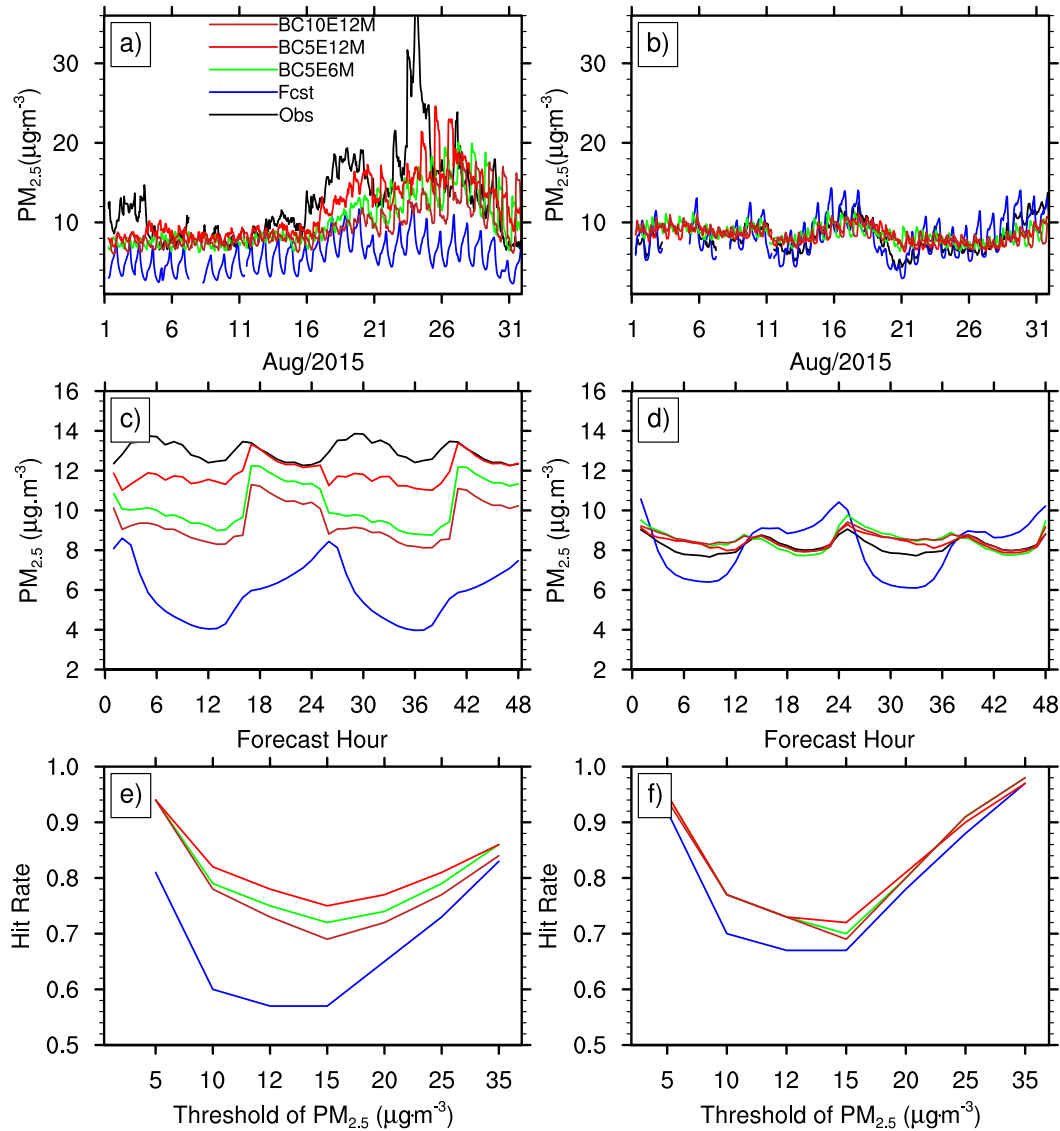


FIG. 7. As in Fig. 6, but during August 2015.

which analogs are searched, it becomes more challenging to find good matches. Finally, when a fire first erupted and affected the measured PM_{2.5} at an AirNow observation site, there may be no historical forecasts for that site that include fire-associated forecast biases, and therefore none of the selected analogs were used to correct the model for the presence of fire, indicating a need of a longer training period for analog searches or for the use of other indicators. Thus, in Fig. 7a it was only after 26 August, when the high fire-related PM_{2.5} was present for nearly a week, that the bias correction scheme finally was able to accurately increase the forecast PM_{2.5} to match the observed values.

In contrast, the NAQFC predictions showed much larger variability than the observations over the EUS in

August (Fig. 7d). Here, underpredictions and overpredictions were observed during daytime and nighttime, respectively. In contrast to the WUS, the influence of wildfire smoke was much smaller in the EUS during this period. The analog ensemble bias correction approach did not capture some of the day-to-day variability as well as the raw forecast guidance did over the EUS (Fig. 7b).

The forecast skill quantified by hit rates was clearly increased over the WUS (Fig. 7e). Among the results, the BC5E12M scenario (solid red, with five ensemble members and a 12-month training period) showed the best performance over the WUS. However, both the NAQFC forecast guidance and the bias correction approach require further study for the predictions during wildfire-smoke-driven PM_{2.5} events.

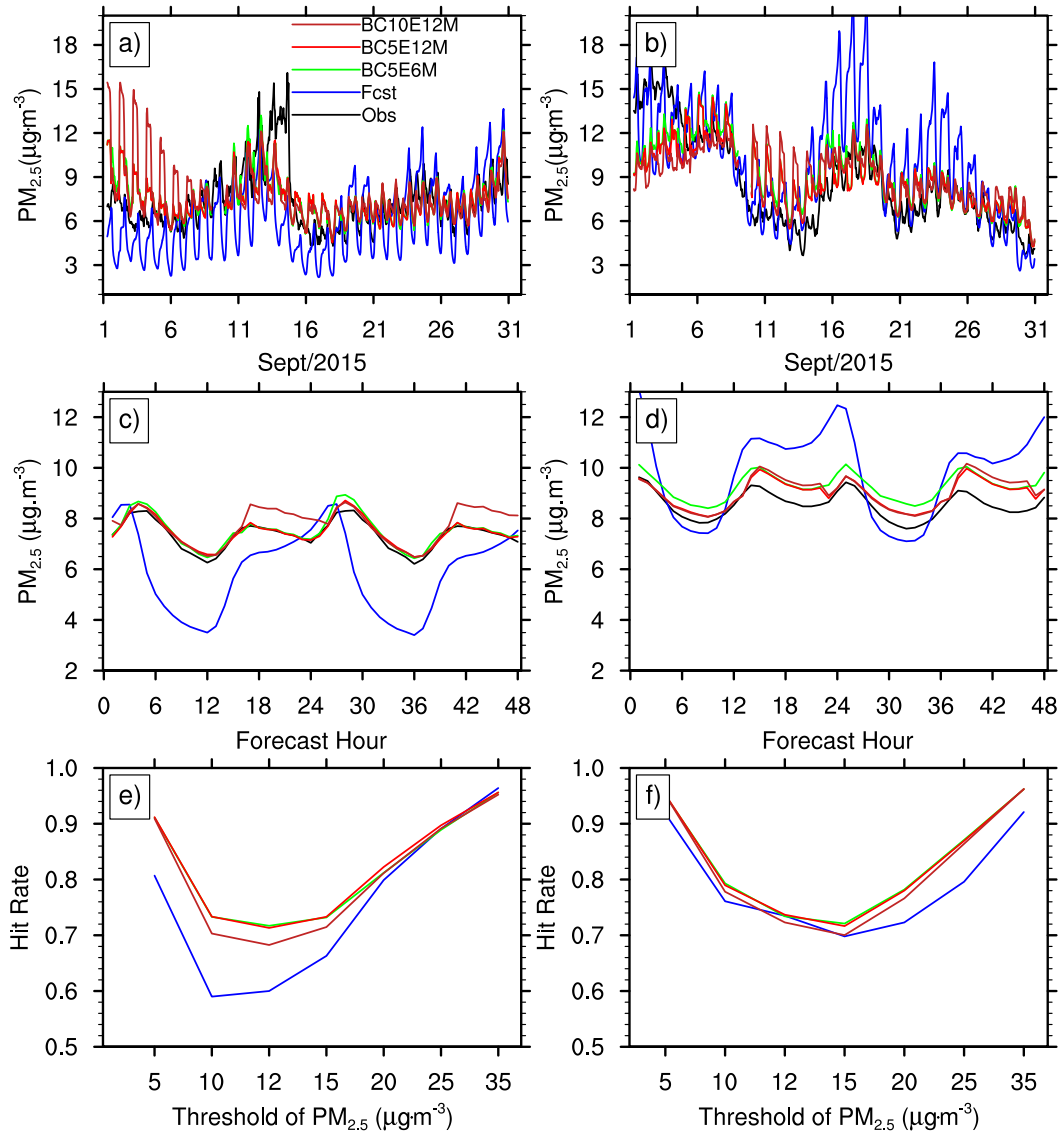


FIG. 8. As in Fig. 6, but during September 2015.

c. PM_{2.5} forecast guidance and bias correction in September 2015

For September, a typical transition month, the NAQFC predictions showed opposite patterns of behavior for the WUS and EUS (Fig. 8). The PM_{2.5} was underpredicted by the NAQFC predictions across the WUS, but overpredicted over the EUS. The BC5E12M runs (red lines) showed excellent agreement with the observed hourly variations (Figs. 8a,b) over the WUS, although a wildfire event on 13–14 September was still underpredicted by both the raw and bias-corrected forecasts. Overcorrections by the bias correction experiments were seen in the hourly time series during the

first week of September (see Fig. 8a), especially for the case of BC10E12M (using 10 members; green line). This was because the BC10E12M case had the most members and therefore was likely choosing more recent fire days for its analogs that were not ideal matches.

All bias correction experiments showed a substantial increase in the hit rate for thresholds below 15 $\mu\text{g m}^{-3}$ in the WUS (Fig. 8e) and moderate change for thresholds above 15 $\mu\text{g m}^{-3}$ in the EUS (Fig. 8f). The configurations with 5 members (i.e., BC5E6M and BC5E12M) showed higher hit rates than the configurations with 10 analog ensemble members. The readers are reminded that the performance of the configurations with 10 analog ensemble members could be improved further if the

training period were to be extended (to, say, 2–3 yr) but this is a considerable burden when creating real-time operational forecasts since CPU time is limited, and forecast models are generally updated on at least an annual basis. Little change in CSI skill score was seen when the daily 1-h maximum $\text{PM}_{2.5}$ was lower than $15 \mu\text{g m}^{-3}$, but the bias-corrected CSI was lower than that of the model raw forecast for thresholds of $15 \mu\text{g m}^{-3}$ and above, especially for the configuration with 10 ensemble members and a 1-yr training period (figure not shown). A slight degradation in CSI was found over the EUS (not shown). This was because the PODs decreased for the higher-threshold events after the bias correction was applied. Similar to July and August, significant reductions in the bias were seen over both the EUS and WUS in September but this was not the case for RMSE.

d. $\text{PM}_{2.5}$ forecast guidance and bias correction in November 2015

Wildfires episodes were less frequent in November. As a result, the NAQFC predictions showed better agreement with the observations over the WUS (see Fig. 9a). As seen in Figs. 9c and 9d, the diurnal variations improved substantially with bias correction in the EUS, but had only a slight impact in the WUS, where it most notably helped correct the hourly timing of the minimum and maximum $\text{PM}_{2.5}$ values.

A large increase in the hit rate was seen for all the thresholds over the WUS and for thresholds above $12.0 \mu\text{g m}^{-3}$ over the EUS. Over the EUS there was a small degradation in the bias-corrected CSI values for larger CSI thresholds of $12 \mu\text{g m}^{-3}$ and above (Table 3). Overall, the performance of bias correction in November was similar to that in July and September. Therefore, the combination of $\text{PM}_{2.5}$, temperature, and wind speed and wind direction was adequate to identify analogs for bias correction except for infrequent, but important, high- $\text{PM}_{2.5}$ events such as wildfires.

5. Discussion and future direction

Substantial improvement in the skill of $\text{PM}_{2.5}$ predictions is demonstrated with the analog ensemble bias correction approach. However, this method has limitations for handling extremely high concentration events such as the Independence Day fireworks, wildfires, and wind-blown dust episodes. The rarer the event, the longer the training dataset needs to be to find good analogs, or more effective methods are needed to determine analogs. Currently, $\text{PM}_{2.5}$ is combined with three meteorological variables (i.e., 2-m temperature, and 10-m wind speed and wind direction) for identifying

appropriate analog ensemble members; however, other parameters could be considered (e.g., model-predicted organic carbon to determine wildfire-smoke-influenced episodes). Moreover, as Junk et al. (2015) have shown, optimal weighting of the analog predictors (computed independently for every location and possibly forecast lead time) may help improve considerably the analog ensemble performance for $\text{PM}_{2.5}$ predictions. The latter is left to future investigations.

The ratio of OC to ammonium sulfate can also be a potential indicator for distinguishing wildfire emissions from anthropogenic emissions. Inclusion of such a parameter in the analog metric calculation could be helpful for determining analog members used in the $\text{PM}_{2.5}$ bias correction algorithm. In addition, inclusion of fire impacts on weather forecasts could provide more reasonable meteorological fields for finding the best analogs from the historical data.

For wind-blown dust events, soil moisture and surface friction velocity are critical parameters for calculating dust emissions. Inclusion of those dust-sensitive parameters may allow analog ensemble bias correction approaches to better correct raw forecasts for dust events.

The Independence Day firework event is a human activity and $\text{PM}_{2.5}$ concentrations do not have a strong dependence on weather conditions. The analog ensemble approach does not help unless the Fourth of July weather from the previous year happens to be similar to the current forecast. Inclusion of firework emissions into the emission inventory would improve $\text{PM}_{2.5}$ predictions on 4 July. An alternative would be to force the analog scheme to only use the previous (one or more) 4 July cases as analogs, or delete 4 July from the training data (in which case the forecast will have a low bias on 4 July).

Results also show that day-to-day or week-to-week variabilities are reduced in some instances after applying the analog ensemble bias correction approach. The problem becomes more evident when the number of analog ensemble members is increased and the training period is short. This is to be expected when using an ensemble mean approach as a bias correction. By definition, the ensemble mean reduces the variability of the ensemble members given its smoother estimate. Thus, more tests on appropriate ensemble member numbers and longer training periods (e.g., 2–3 yr) are needed for reinstating the day-to-day or week-to-week variabilities of the bias-corrected predictions.

Computational time is critical for real-time operational forecasts. The analog ensemble approach is demonstrated as the first step for improving real-time

KFAN will be tested for real-time forecast applications when a parallelized version of this code becomes available.

6. Summary and conclusions

In this study, a summary of the performance of the NOAA National Air Quality Forecast Capability (NAQFC) PM_{2.5} predictions with and without bias correction is presented. A persistent seasonal bias is noted over the past several years. Various efforts have been made at NOAA to improve the NAQFC predictions of surface PM_{2.5}, resulting in improved winter PM_{2.5} predictions. However, underprediction in summer has not improved and was even worse in 2015 than previous years. Out-of-date emission inventories could be one of the major reasons, in addition to intense wildfire activity during the summer of 2015.

To address these identified biases, an analog ensemble bias correction is integrated into the NOAA NAQFC. Tests of the analog ensemble approach with different configurations have been completed to assess the impact of training period length and number of analog members on the analog ensemble bias correction's performance during July, August, September, and November 2015. Results show that the diurnal variation patterns are improved greatly with all the configurations. The sensitivity run BC5E12M (using five analog ensemble members and a 12-month training period) provides the best performance overall. This configuration has been selected for the analog ensemble bias correction approach for the 2016 NAQFC operational implementation at NOAA/NCEP.

This study also highlights the limitation of the analog ensemble bias correction approach on improving PM_{2.5} predictions during infrequent, but extremely high concentration, PM_{2.5} episodes, such as the Fourth of July Independence Day fireworks and wildfire events. A more robust way of identifying analogs is critical to improving the analog ensemble bias correction approach. For example, including the ratio of organic carbon to ammonium sulfate might improve the search for good analogs during wildfire emission-type events. Soil moisture and surface friction velocity could be included for identifying better analogs for dust-storm cases. Overall, this study highlights the strengths and weaknesses of the analog ensemble approach, and provides direction for our next steps as well as future research.

Acknowledgments. This project is supported by the NOAA National Air Quality Forecast Capability Program and by the U.S. Weather Research Program within the NOAA/OAR Office of Weather and Air Quality. The authors are thankful to Jerry Gorline (NOAA/NWS/MDL) for the help in plotting Fig. 3.

REFERENCES

- Brook, R. D., and Coauthors, 2004: Air pollution and cardiovascular disease: A statement for healthcare professionals from the expert panel on population and prevention science of the American Heart Association. *Circulation*, **109**, 2655–2671, doi:10.1161/01.CIR.0000128587.30041.C8.
- Byun, D., and K. L. Schere, 2006: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Appl. Mech. Rev.*, **59**, 51–77, doi:10.1115/1.2128636.
- Canty, T. P., and Coauthors, 2015: Ozone and NO_x chemistry in the eastern US: Evaluation of CMAQ/CB05 with satellite (OMI) data. *Atmos. Chem. Phys.*, **15**, 10965–10982, doi:10.5194/acp-15-10965-2015.
- Carlton, A. G., P. Bhave, S. L. Napelenok, E. O. Edney, G. Sarwar, R. W. Pinder, G. A. Pouliot, and M. Houyoux, 2010: Model representation of secondary organic aerosol in CMAQ v4.7. *Environ. Sci. Technol.*, **44**, 8553–8560, doi:10.1021/es100636q.
- Cui, B., Z. Toth, Y. Zhu, and D. Hou, 2012: Bias correction for global ensemble forecast. *Weather Forecasting*, **27**, 396–410, doi:10.1175/WAF-D-11-00011.1.
- Delle Monache, L., T. Nipen, X. Deng, Y. Zhou, and R. Stull, 2006: Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction. *J. Geophys. Res.*, **111**, D05308, doi:10.1029/2005JD006311.
- , and Coauthors, 2008: A Kalman-filter bias correction of ozone deterministic, ensemble-averaged, and probabilistic forecasts. *Tellus*, **60B**, 238–249, doi:10.1111/j.1600-0889.2007.00332.x.
- , T. Nipen, Y. Liu, G. Roux, and R. Stull, 2011: Kalman filter and analog schemes to postprocess numerical weather predictions. *Mon. Wea. Rev.*, **139**, 3554–3570, doi:10.1175/2011MWR3653.1.
- , T. Eckel, D. Rife, and B. Nagarajan, 2013: Probabilistic weather prediction with an analog ensemble. *Mon. Wea. Rev.*, **141**, 3498–3516, doi:10.1175/MWR-D-12-00281.1.
- Djalalova, I., and Coauthors, 2010: Ensemble and bias-correction techniques for probabilistic forecast of surface O₃ and PM_{2.5} during the TEXAQS-II experiment of 2006. *Atmos. Environ.*, **44**, 455–467, doi:10.1016/j.atmosenv.2009.11.007.
- , L. D. Monache, and J. Wilczak, 2015: PM_{2.5} analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model. *Atmos. Environ.*, **119**, 431–442, doi:10.1016/j.atmosenv.2015.05.057.
- Durai, V. R., and R. Bhradwaj, 2014: Evaluation of statistical bias correction methods for numerical weather prediction model forecasts of maximum and minimum temperatures. *Nat. Hazards*, **73**, 1229–1254, doi:10.1007/s11069-014-1136-1.
- EPA, 2015: Areas designations for the 2012 the 2012 annual fine particle (PM_{2.5}) standard. [Available online at <http://www3.epa.gov/pmdesignations/2012standards/state.htm>.]
- Fann, N., A. D. Lamson, S. C. Anenberg, K. Wesson, D. Risley, and B. J. Hubbell, 2012: Estimating the national public health burden associated with exposure to ambient PM_{2.5} and ozone. *Risk Anal.*, **32**, 81–95, doi:10.1111/j.1539-6924.2011.01630.x.
- Glahn, B., 2014: Determining an optimal decay factor for bias-correcting MOS temperature and dewpoint forecasts. *Weather Forecasting*, **29**, 1076–1090, doi:10.1175/WAF-D-13-00123.1.
- Glahn, H., and D. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211, doi:10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2.

- Hamill, T. M., and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analog: Theory and application. *Mon. Wea. Rev.*, **134**, 3209–3229, doi:10.1175/MWR3237.1.
- Hand, J. L., and Coauthors, 2011: IMPROVE—Interagency Monitoring of Protected Visual Environments: Spatial and seasonal patterns and temporal variability of haze and its constituents in the United States. CIRA Rep. V, Cooperative Institute for Research in the Atmosphere, Colorado State University, 48 pp. [Available online at http://vista.cira.colostate.edu/Improve/wp-content/uploads/2016/04/Cover_TOC.pdf.]
- Houyoux, M. R., J. M. Vukovich, C. J. Coats Jr., N. J. M. Wheeler, and P. S. Kasibhatla, 2000: Emission inventory development and processing for the Seasonal Model for Regional Air Quality (SMRAQ) project. *J. Geophys. Res.*, **105**, 9079–9090, doi:10.1029/1999JD900975.
- Janjić, Z., and R. Gall, 2012: Scientific documentation of the NCEP Nonhydrostatic Multiscale Model on B grid (NMMB). Part 1 Dynamics. NCAR Tech. Note NCAR/TN-489+STR, 80 pp. [Available online at <https://openky.ucar.edu/islandora/object/technotes%3A502>.]
- Jo, S., and J. Ahn, 2015: Improvement of CGCM prediction for wet season precipitation over Maritime Continent using a bias correction method. *Int. J. Climatol.*, **35**, 3721–3732, doi:10.1002/joc.4232.
- Junk, C., L. Delle Monache, S. Alessandrini, L. von Bremen, and G. Cervone, 2015: Predictor-weighting strategies for probabilistic wind power forecasting with an analog ensemble. *Meteor. Z.*, **24**, 361–379, doi:10.1127/metz/2015/0659.
- Kang, D., R. Mathur, S. T. Rao, and S. Yu, 2008: Bias adjustment techniques for improving ozone air quality forecasts. *J. Geophys. Res.*, **113**, D23308, doi:10.1029/2008JD010151.
- , —, and —, 2010: Real-time bias-adjusted O₃ and PM_{2.5} air quality index forecasts and their performance evaluations over the continental United States. *Atmos. Environ.*, **44**, 2203–2212, doi:10.1016/j.atmosenv.2010.03.017.
- Kwok, R. H. F., S. L. Napelenok, and K. R. Baker, 2013: Implementation and evaluation of PM_{2.5} source contribution analysis in a photochemical model. *Atmos. Environ.*, **80**, 398–407, doi:10.1016/j.atmosenv.2013.08.017.
- Larkin, N. K., and Coauthors, 2009: The BlueSky smoke modeling framework. *Int. J. Wildland Fire*, **18**, 906–920, doi:10.1071/WF07086.
- Lee, P., and Coauthors, 2017: NAQFC developmental forecast guidance for fine particulate matter (PM_{2.5}). *Wea. Forecasting*, doi:10.1175/WAF-D-15-0163.1, in press.
- Nel, A., 2005: Air pollution-related illness: Effects of particles. *Science*, **308**, 804–806, doi:10.1126/science.1108752.
- O'Neill, S. M., and Coauthors, 2009: Regional real-time smoke prediction systems. *Wildland Fires and Air Pollution: Developments in Environmental Science*, A. Bytnerowicz et al., Eds., Developments in Environmental Science, Vol. 8, Elsevier, 499–534.
- Otte, T. L., and J. E. Pleim, 2010: The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system: Updates through MCIPv3.4.1. *Geosci. Model Dev.*, **3**, 243–256, doi:10.5194/gmd-3-243-2010.
- , 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, doi:10.1175/WAF855.1.
- Pan, L., D. Tong, P. Lee, H.-C. Kim, and T. Chai, 2014: Assessment of NO_x and O₃ forecasting performances in the U.S. National Air Quality Forecasting Capability before and after the 2012 major emissions updates. *Atmos. Environ.*, **95**, 610–619, doi:10.1016/j.atmosenv.2014.06.020.
- Ruminski, M., S. Kondragunta, R. Draxler, and J. Zeng, 2006: Recent change to the Hazard Mapping System. Preprints, *15th Int. Emission Inventory Conf.: Reinventing Inventories—New Ideas in New Orleans*, New Orleans, LA, EPA. [Available online at <http://www.epa.gov/ttn/chief/conference/ei15/session10/ruminski.pdf>.]
- Saylor, R. D., and A. F. Stein, 2012: Identifying the causes of differences in ozone production from CB05 and CBMIV chemical mechanisms. *Geosci. Model Dev.*, **5**, 257–268, doi:10.5194/gmd-5-257-2012.
- Stajner, I., P. Davidson, D. Byun, J. McQueen, R. Draxler, P. Dickerson, and J. Meagher, 2012: US National Air Quality Forecast Capability: Expanding coverage to include particulate matter. *Air Pollution Modeling and Its Application XXI*, D. G. Steyn and S. T. Castelli, Eds., NATO Science for Peace and Security Series C: Environmental Security, Springer, 379–384, doi:10.1007/978-94-007-1359-8_64.
- Tong, D. Q., and Coauthors, 2015: Long-term NO_x trends over large cities in the United States during the great recession: Comparison of satellite retrievals, ground observations, and emission inventories. *Atmos. Environ.*, **107**, 70–84, doi:10.1016/j.atmosenv.2015.01.035.
- Wilczak, J., and Coauthors, 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.
- Wilks, D., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- Zhu, Y., and Y. Luo, 2015: Precipitation calibration based on the frequency-matching method. *Wea. Forecasting*, **30**, 1109–1124, doi:10.1175/WAF-D-13-00049.1.