

# JGR Atmospheres

## RESEARCH ARTICLE

10.1029/2019JD032293

### Key Points:

- The chemical transport models (GEOS-Chem, WRF-Chem, and WRF-CMAQ) show systematically low bias of PM<sub>2.5</sub>
- Model output postprocessing with surface data and the ensemble Kalman filter technique improves the PM<sub>2.5</sub> forecast at both local and urban scale
- The Successive Correction Method extends the PM<sub>2.5</sub> forecast improvement from the local to regional scale

### Correspondence to:

H. Zhang and J. Wang,  
jun-wang-1@uiowa.edu;  
huanxin-zhang@uiowa.edu

### Citation:

Zhang, H., Wang, J., Garcia, L. C., Ge, C., Plessel, T., Szykman, J., et al. (2020). Improving surface PM<sub>2.5</sub> forecasts in the United States using an ensemble of chemical transport model outputs: 1. Bias correction with surface observations in nonrural areas. *Journal of Geophysical Research: Atmospheres*, 125, e2019JD032293. <https://doi.org/10.1029/2019JD032293>

Received 19 DEC 2019

Accepted 22 MAY 2020

Accepted article online 7 JUN 2020

### Author Contributions:

**Conceptualization:** Huanxin Zhang, Jun Wang

**Data curation:** Huanxin Zhang, Jun Wang, Lorena Castro Garcia, Todd Plessel, James Szykman, Benjamin Murphy, Tanya L. Spero

**Funding acquisition:** Jun Wang, James Szykman

**Investigation:** Huanxin Zhang, Jun Wang, Lorena Castro Garcia, James Szykman

**Methodology:** Huanxin Zhang, Jun Wang

**Project administration:** Jun Wang

**Resources:** Huanxin Zhang, Jun Wang

**Software:** Lorena Castro Garcia, Cui Ge

**Supervision:** Jun Wang

**Visualization:** Lorena Castro Garcia  
(continued)

©2020. American Geophysical Union.  
All Rights Reserved.

## Improving Surface PM<sub>2.5</sub> Forecasts in the United States Using an Ensemble of Chemical Transport Model Outputs: 1. Bias Correction With Surface Observations in Nonrural Areas

Huanxin Zhang<sup>1,2</sup> , Jun Wang<sup>1,2</sup> , Lorena Castro Garcia<sup>1,2</sup>, Cui Ge<sup>1,2</sup> , Todd Plessel<sup>3</sup>, James Szykman<sup>4</sup>, Benjamin Murphy<sup>4</sup>, and Tanya L. Spero<sup>4</sup> 

<sup>1</sup>Department of Chemical and Biochemical Engineering, The University of Iowa, Iowa City, IA, USA, <sup>2</sup>Center for Global and Regional Environmental Research, The University of Iowa, Iowa City, IA, USA, <sup>3</sup>General Dynamics Information Technology, RTP, NC, USA, <sup>4</sup>U.S. Environmental Protection Agency, RTP, NC, USA

**Abstract** This work is the first of a two-part study that aims to develop a computationally efficient bias correction framework to improve surface PM<sub>2.5</sub> forecasts in the United States. Here, an ensemble-based Kalman filter (KF) technique is developed primarily for nonrural areas with approximately 500 surface observation sites for PM<sub>2.5</sub> and applied to three (GEOS-Chem, WRF-Chem, and WRF-CMAQ) chemical transport model (CTM) hindcast outputs for June 2012. While all CTMs underestimate daily surface PM<sub>2.5</sub> mass concentration by 20–50%, KF correction is effective for improving each CTM forecast. Subsequently, two ensemble methods are formulated: (1) the arithmetic mean ensemble (AME) that equally weights each model and (2) the optimized ensemble (OPE) that calculates the individual model weights by minimizing the least-square errors. While the OPE shows superior performance than the AME, the combination of either the AME or the OPE with a KF performs better than the OPE alone, indicating the effectiveness of the KF technique. Overall, the combination of a KF with the OPE shows the best results. Lastly, the Successive Correction Method (SCM) was applied to spread the bias correction from model grids with surface PM<sub>2.5</sub> observations to the grids lacking ground observations by using a radius of influence of 125 km derived from surface observations, which further improves the forecast of surface PM<sub>2.5</sub> at the national scale. Our findings provide the foundation for the second part of this study that uses satellite-based aerosol optical depth (AOD) products to further improve the forecast of surface PM<sub>2.5</sub> in rural areas by performing statistical analysis of model output.

**Plain Language Summary** Air quality forecasting plays an important role in informing the general public and decision-makers on reducing exposure to air pollution. Air quality models simulating atmospheric constituents such as particulate matter with a diameter less than 2.5 μm (PM<sub>2.5</sub>) are often used to provide daily forecasts. However, these models are subject to large error and uncertainty as a result of the incomplete representation of the real atmosphere. Here, we develop a computationally efficient framework to improve model forecasts by performing bias correction on model outputs. We focus on nonrural areas in the continental United States and show that our technique improves model forecasts of surface PM<sub>2.5</sub>. In a companion paper, we focus on the application of satellite data to improve PM<sub>2.5</sub> forecasting in rural areas.

## 1. Introduction

Exposure to ambient PM<sub>2.5</sub> (fine particulate matter with aerodynamic diameter less than 2.5 μm) can cause adverse health issues such as cardiovascular and respiratory illness (Peng et al., 2005; Pope et al., 2009). The Global Burden of Disease Study ranked ambient particulate matter pollution as one of the top risk factors for global deaths (Forouzanfar et al., 2016) and a recent study by Cohen et al. (2017) estimated that exposure to ambient PM<sub>2.5</sub> caused more than 4 million deaths worldwide and 88,400 deaths in the United States in 2015. PM<sub>2.5</sub> is one of the criteria air pollutants regulated in the United States. The U.S. Environmental Protection Agency (EPA) established the national ambient air quality standards (NAAQS) for 24-hr-averaged PM<sub>2.5</sub> concentration as 65 μg m<sup>-3</sup> in 1997 and has made revisions in 2006 and 2012 to lower the concentration

**Writing - original draft:** Huanxin

Zhang, Jun Wang

**Writing - review & editing:** Huanxin

Zhang, Jun Wang

to  $35 \mu\text{g m}^{-3}$ . Air quality forecasting has played an important role in providing information to alert the general public and decision makers to take preventive cautions.

Air quality models are subject to large uncertainty associated with the parametrization of chemical and physical processes, inaccurate emission inventories, and errors in model initial and boundary conditions (Simon et al., 2012; Solazzo et al., 2012). As a result, different techniques have been developed to improve model forecasts such as postprocessing bias correction, chemical data assimilation, and ensemble forecasting methods (Bocquet et al., 2015; Wang, Nair, & Christopher, 2004; Zhang et al., 2012a, 2012b). Commonly used chemical data assimilation techniques are the variational method (3D-Var or 4D-Var) (Schwartz et al., 2012; Wang et al., 2012; Wang et al., 2014; Xu et al., 2013; Wang, Wang, et al., 2016; Zhang et al., 2008), optimal interpolation (OI) (Chai et al., 2017), and the Kalman filter (KF) or ensemble KF (Carmichael et al., 2008; Tang et al., 2011) to optimize initial and/or boundary conditions or to adjust emission factors. Due to the large computational cost in advanced data assimilation techniques, modelers have used simple bias correction techniques in postprocessing, for example, the so-called mean subtraction (McKeen et al., 2005; Wilczak et al., 2006) and the KF (De Ridder et al., 2012; Delle Monache, Nipen, et al., 2006; Delle Monache et al., 2008; Delle Monache et al., 2011; Djalalova et al., 2010; Kang et al., 2008; Kang et al., 2010a). All techniques resulted in improved forecasts, especially when using a Kalman filter, which can reduce systematic errors. Kang et al. (2010b) were the first to show the operational capability of the KF technique in predicting surface ozone ( $\text{O}_3$ ) and  $\text{PM}_{2.5}$  in real time over the continental United States.

In addition to the KF, the ensemble approach is widely used in numerical weather forecasts (Kalnay, 2003). In recent years, the ensemble approach has been applied to air quality forecasting. The ensemble technique has usually been performed by either using one model (such as WRF-Chem) with different emission inventory inputs and meteorological fields from various global models (Ge et al., 2017) or multimodels. Early work by Delle Monache and Stull (2003) found that the ensemble mean of all four photochemical model forecasts performed better than any of the individual models while studying an ozone event in Europe. The following work also showed improvements in  $\text{O}_3$  forecasts using an ensemble technique (Delle Monache, Deng, et al., 2006; McKeen et al., 2005; Pagowski et al., 2005). Similar to McKeen et al. (2005) but focusing on real-time  $\text{PM}_{2.5}$  forecasts, McKeen et al. (2007) again indicated that the ensemble average of seven air quality forecasting models calculated with equal weights for each model member resulted in the best performance when compared with the observational data. One possible reason for the improved performance is that ensemble modeling can reduce random errors. Additionally, some work has combined the KF technique with an ensemble technique. For instance, Delle Monache, Nipen, et al. (2006) demonstrated that the ensemble mean (EM) of 12 KF-corrected  $\text{O}_3$  air quality forecast models and KF-corrected EM showed the best forecasting skill. Djalalova et al. (2010) applied two bias correction techniques (a simple running mean average and a KF approach) separately to seven individual air quality forecasting models before generating the ensembles. Weighted model ensembles were further created using a linear regression method to minimize the bias. Overall, the combination of KF and weighted averaging provided the best performance.

Real-time data repositories such as the U.S. EPA's AirNow program (Dye et al., 2004) contribute significantly to the successful implementation of some of the aforementioned techniques. Currently, the AirNow program provides near-real-time hourly surface  $\text{PM}_{2.5}$  concentration measurements at about 600 ground sites across the United States and daily air quality forecasts for major U.S. cities. The air quality information is distributed to the general public and social media outlets in the form of an air quality index (AQI). The measured hourly surface  $\text{PM}_{2.5}$  is crucial for real-time forecasts as well as model evaluation. However, the spatial coverage is still limited and unevenly distributed with more monitoring sites along the west and east coast in urban areas and less in between such as the Great Plains and Rocky Mountain region. Because of this limitation, most of the previous bias correction conducted in the U.S. has only focused on places where ground observations were available. Recently, the work of Djalalova et al. (2015) propagated the bias information obtained from surface  $\text{PM}_{2.5}$  measurement sites to surrounding areas with no ground observations. This method is integrated into the operational  $\text{PM}_{2.5}$  forecast of the CMAQ modeling system provided by the U.S. National Air Quality Forecasting Capability (NAQFC) (Huang et al., 2017; Lee et al., 2017).

Satellite data could also fill in the gaps due to its global spatial coverage (Hoff & Christopher, 2009). The launch of the National Aeronautics and Space Administration (NASA) Earth Observing System (EOS) satellite Terra in 1999 started a new era of studying air quality from space (Hoff & Christopher, 2009;

Martin, 2008; Wang & Christopher, 2003). Aerosol optical depth (AOD) is one of the satellite-derived aerosol properties and has been correlated with surface measured  $PM_{2.5}$  (Chu et al., 2003; Liu et al., 2004; Van Donkelaar et al., 2006; Van Donkelaar et al., 2010; Wang & Christopher, 2003). Satellite-derived surface  $PM_{2.5}$  has been widely used in studying air quality and human health. Furthermore, a few studies (Al-Saadi et al., 2005; Szykman et al., 2004; Szykman et al., 2012) have demonstrated the potential of using satellite observations for forecasting air quality.

This is the first of a two-part study that aims to develop a bias correction framework to improve surface  $PM_{2.5}$  forecasts in the United States by integrating chemical transport model outputs, surface observations and satellite remote sensing data. Since hourly measured surface data is mostly located in nonrural areas, the present work will apply a multi-model ensemble approach along with the KF technique and a successive correction technique to adjust model forecast biases in these nonrural areas. The companion paper investigates the use of a multi-AOD ensemble approach for model forecast bias correction with a special focus on rural areas. As the first of a two-part series, this paper is motivated by the poor forecasting skills of  $PM_{2.5}$  (as discussed above), and hence, we continue investigating the robustness of the KF technique in  $PM_{2.5}$  forecasting in the continental United States, especially its combination with both equally weighted and nonequally weighted (optimized) ensemble techniques. Furthermore, we refine the radius of influence that propagates the model bias information from ground sites to nearby places that have no surface measurements available. More importantly, this paper sets the foundation for the two-part series of work that as a whole, differs from past work in that multiple model outputs and multiple satellite data products are used as ensemble members for Kalman filter processing, and the emphasis is on the synergy between surface observation networks and satellite observations for improving air quality forecasts in rural areas. This paper is organized as follows: In section 2, we describe the chemical transport models and surface measurements of  $PM_{2.5}$ . Section 3 presents the bias correction techniques. Results are summarized in section 4, and discussion and conclusions make up section 5.

## 2. Model Description and Observational Data

We used hindcast outputs from three chemical transport models here. The hindcast used the reanalysis meteorology data and therefore, it does not resemble exactly an operational air quality forecasting (because only the atmospheric chemistry is forecasted with no constraints from observations). Nevertheless, the hindcast has been commonly used in the research for developing new techniques that are valuable for improving real-time forecasts, and in section 3.4 we describe how the bias correction techniques were implemented as if they were in a forecasting mode. At the time of performing model simulations, the GEOS-Chem model did not carry the forecasting capability, but the recent development of WRF-GC (<http://wrf.geos-chem.org/>) would make the GEOS-Chem model applicable in real-time air quality forecasting. Furthermore, its chemistry has been implemented in NASA's Earth system models (Hu et al., 2018; Long et al., 2015), creating opportunities for experimental global air quality forecasting. Detailed model configurations are provided in this section, and a brief summary is also given in Table 1.

### 2.1. GEOS-Chem

We use the global 3-D chemical transport model (GEOS-Chem) (Bey et al., 2001) Version v11-01 (<http://geos-chem.org>) together with its nested version over the North American domain (10–70°N and 40–140°W) (Chen et al., 2009; Wang, McElroy, et al., 2004). Both global and nested simulations are driven by assimilated meteorological fields from the NASA Goddard Earth Observing System (GEOS5). The GEOS5 data have a native resolution of 0.5° latitude by 0.667° longitude and a temporal resolution of 3 hr (1 hr for surface variables and mixing depths). First, we conducted a global GEOS-Chem simulation at 2° latitude  $\times$  2.5° longitude horizontal resolution from 1 June 2011 to 30 June 2012 with the first 12 months as a spin up. Then, the results from the global simulation provided the hourly dynamic boundary conditions for the nested simulation with 0.5° latitude  $\times$  0.667° longitude horizontal resolution. The nested grid simulation ran from 1 May 2012 to 30 June 2012, where the first month was used for model initialization. Both global and nested simulations had the same vertical distribution of 47 pressure levels as well as the same model setup, which will be described below.

**Table 1**  
The Summary of Model Configurations

	GEOS-Chem	WRF-Chem	CMAQ
Version	v11-01	v3.8.1	v5.0.2
Horizontal spatial resolution	0.5° × 0.667°	12 km	12 km
Temporal resolution	Hourly	Hourly	Hourly
Meteorology	GEOS5	WRv3.8.1	WRv3.4
Chemistry boundary conditions	GEOS-Chem v11-01 (2° × 2.5°)	Model default <sup>a</sup>	GEOS-Chem v8-03-02 with GEOS5
Anthropogenic emissions	NEI 2011	NEI 2011	NEI 2011
Biomass burning emissions	FINN	FLAMBE	BlueSky
Biogenic emissions	MEGAN 2.1	Online MEGAN	BEIS
Chemical mechanism (gas phase)	Trop Chem	RADM2	CB05TUCL
Aerosol module	(NO <sub>x</sub> -O <sub>x</sub> -Aer-Br)	MADE/SORGAM	AERO6

<sup>a</sup>In WRF-Chem, the model default chemical boundary conditions represent a clean North America summer day. Only a limited number of chemical species are included and the majority of them are gas species.

The GEOS-Chem aerosol simulation includes the sulfate-nitrate-ammonium system (Park et al., 2004; Pye et al., 2009), primary (Park et al., 2003) and secondary (Liao et al., 2007) carbonaceous aerosols (SOAs), mineral dust (Fairlie et al., 2007; Fairlie et al., 2010), and sea salt in accumulation and coarse modes (Jaeglé et al., 2011). SOA is disabled in the current simulation. The aerosol chemistry was coupled to the gas-phase chemistry (HO<sub>x</sub>-NO<sub>x</sub>-VOC-O<sub>3</sub>-BrO<sub>x</sub>) through the gas-aerosol equilibrium partitioning of nitric acid and ammonia (Fountoukis & Nenes, 2007). Aerosol dry and wet deposition follow Zhang et al. (2001) and Liu et al. (2001), respectively. Global anthropogenic emission inventories are taken from Emissions Database for Global Atmospheric Research (EDGAR version v4.2) (<http://edgar.jrc.ec.europa.eu/>), overwritten by available regional emission inventories: the U.S. EPA 2011 National Emission Inventory (NEI) in the United States (Travis et al., 2016), the Critical Air Contaminants (CAC) in Canada, Big Bend Regional Aerosol and Visibility Observational Study Emission Inventory (BRAVO) (Kuhns et al., 2005) in Mexico, European Monitoring and Evaluation Programme (EMEP) in Europe, and MIX (Li et al., 2017) in Asia. Biomass burning emissions are from the Fire INventory from NCAR (FINN) (Wiedinmyer et al., 2011) with daily resolution. Biogenic emissions are from the MEGAN2.1 emission inventories and updates from Guenther et al. (2012). When calculating the mass concentration of surface PM<sub>2.5</sub>, the modeled relationship between each simulated aerosol mass and relative humidity for each aerosol type was accounted for (Van Donkelaar et al., 2010), in order to comply with surface measurement standards. For dust and sea salt aerosols, only the fraction of the total aerosol mass within the cut off diameter of 2.5 μm were included in the calculated PM<sub>2.5</sub> mass concentration.

## 2.2. WRF-Chem

The Weather Research and Forecasting (WRF) model coupled with chemistry, WRF-Chem Version v3.8.1 is used in this study (Fast et al., 2006; Grell et al., 2005). The 1° × 1° National Center for Environmental Prediction (NCEP) Final Analysis (FNL) data provide the meteorological initial and boundary condition (Kalnay et al., 1996; Kistler et al., 2001). The default chemical boundary conditions representing a North America summer day is used. This default boundary condition includes a limited number of species, of which the majority are gas-phase species and was originally developed for tropospheric ozone forecast. The model simulation was conducted from 22 May to 30 June 2012 with the first 10 days of the simulation for initialization. Model output from 1 June to 30 June was analyzed. The model was designed to have one domain over the United States centered around 39.52°N and 94.68°W with a horizontal resolution of 12 km and 28 vertical levels.

We use the Regional Acid Deposition Model, Version 2 (RADM2) for gas-phase chemistry (Stockwell et al., 1990). The aerosol modules are the Modal Aerosol Dynamics Model for Europe (MADE) (Ackermann et al., 1998) and the Secondary Organic Aerosol Model (SORGAM) (Schell et al., 2001). In the MADE/SORGAM model, aerosol species include sulfate, nitrate, ammonium, black carbon (BC), organic matter (OM), sea salt, mineral dust, and water. The MADE/SORGAM model uses the modal approach to represent the aerosol size distribution (Binkowski & Shankar, 1995). The modal approach employs three modes (the Aitken, accumulation and coarse mode) and each mode assumes a log-normal

distribution. In addition, aerosol species are assumed to be internally mixed in each mode so that all particles have the same chemical composition in each mode.

Anthropogenic emissions were taken from the U.S. EPA NEI 2011. Biomass burning emissions come from Fire Locating and Modeling of Burning Emissions Inventory (FLAMBE) (Reid et al., 2009). The ratio of organic carbon and black carbon in FLAMBE was set as 9 and more details of the FLAMBE implementation in WRF-Chem can be found in Wang et al. (2013) and Ge et al. (2014). The model calculates the biogenic emissions online using the U.S. Geological Survey (USGS) land-use classification, generated by the WRF Preprocessing System. No dust and sea salt emissions were included in our simulations. As a result, the  $PM_{2.5}$  mass concentration is calculated as the sum of sulfate, nitrate, ammonium, BC, and OM.

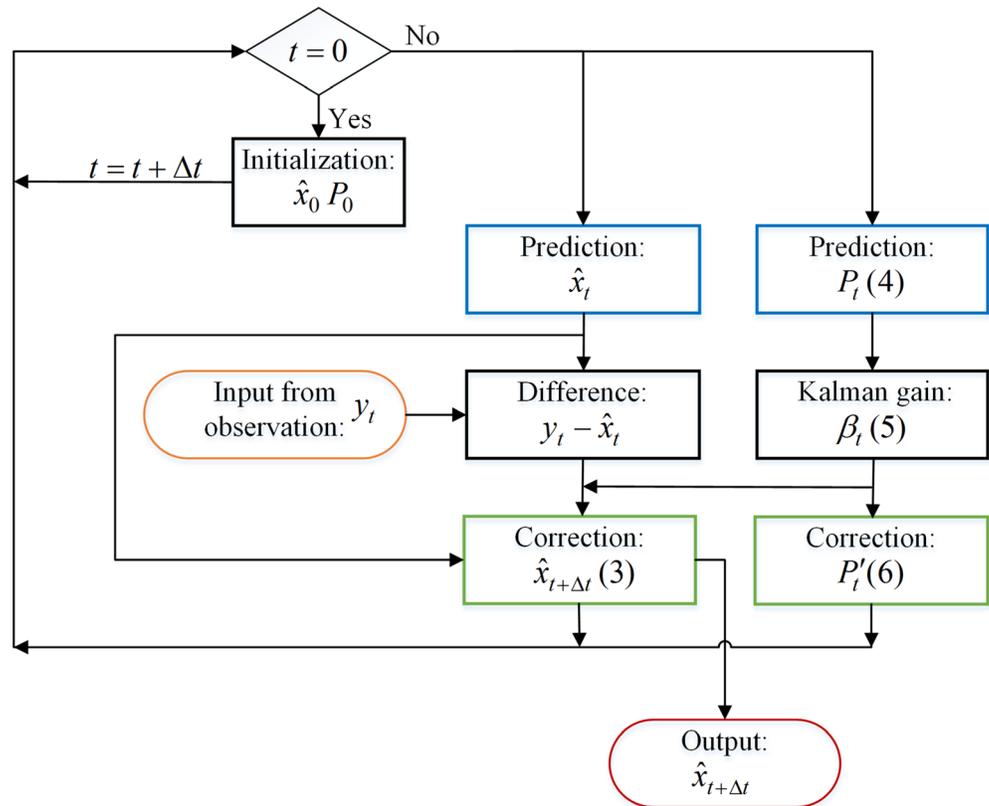
### 2.3. CMAQ

Output from the Community Multiscale Air Quality (CMAQ) (Byun & Schere, 2006) used in this study was downloaded through the Remote Sensing Information Gateway (RSIG) (<https://www.epa.gov/hesc/remote-sensing-information-gateway>). Here, we provide a brief summary of the model configuration that generated the model output. The CMAQ modeling system Version 5.0.2 was used. The model simulations were performed for the continental U.S. domain with a 12 km horizontal grid size and a Lambert Conformal projection. There are 35 vertical levels from the surface to the top of the free troposphere with a bottom layer thickness of 19 m. The meteorology inputs for the CMAQ model came from the WRF Version 3.4 simulations (Skamarock et al., 2008). The WRF simulations were conducted following Version 2 four-dimensional data assimilation (FDDA) with no nudging in the planetary boundary layer and using blended 3-hourly reanalysis fields (combination of 6-hr Meteorological Assimilation Data Ingest System (MADIS) Data and intermediate North American Mesoscale Model (NAM) 3-hr forecast). The WRF outputs were post processed with the Meteorology-Chemistry Input Processor (MCIP v4.1.3) (Otte & Pleim, 2010). Hourly data from global GEOS-Chem simulations (v8-03-02) with GEOS5 meteorology provide the chemistry boundary conditions (Bey et al., 2001).

The model used the CB05TUCL chemical mechanism for gas chemistry (Whitten et al., 2010; Yarwood et al., 2005) and AERO6 (Appel et al., 2013; Binkowski & Roselle, 2003) for aerosol module. The aerosol module simulates inorganic aerosols (sulfate, nitrate, and ammonium), water, secondary organic aerosols (Carlton et al., 2010), elemental carbon, anthropogenic primary organic carbon, and unspecified material of anthropogenic origin. Aerosols are represented in three log-normal modes: Aitken, accumulation, and coarse (Binkowski & Roselle, 2003). Inorganic aerosols in the Aitken and accumulation modes are assumed to be in equilibrium with the gas phase using ISORROPIA II (Fountoukis & Nenes, 2007). Anthropogenic emissions come from the U.S. EPA 2011 NEI projected to 2012. No dust emissions were included. Biogenic emissions were calculated from the Biogenic Emission Inventory System (BEIS). Fire emissions were developed using the U.S. Forest Service's BlueSky modeling framework (<https://airfire.org>). CMAQ outputs  $PM_{2.5}$  concentrations by calculating online the fraction of each log-normal particle mode that would be collected with an aerodynamic cutoff diameter of 2.5  $\mu\text{m}$ . All simulated pollutant species are included in the calculation of  $PM_{2.5}$ . Although aerosol water is included when calculating the size of the particles to determine the  $PM_{2.5}$  fraction, it is then excluded as a constituent of  $PM_{2.5}$  mass, consistent with measurement collection and sample preparation practices. Excluding dust emissions in the model could potentially affect surface  $PM_{2.5}$  concentration in the southwestern United States as previous work have shown the large number of dust events in the summer over this region (Tong et al., 2012) and that dust could contribute to approximately 20–30% of surface  $PM_{2.5}$  concentration in the region (Hand et al., 2017).

### 2.4. Ground-Based Measurements of $PM_{2.5}$

Hourly surface  $PM_{2.5}$  measurements over the United States for June 2012 were obtained from the U.S. EPA Air Quality System (AQS) (<https://www.epa.gov/aqs>). Even though the data from AQS have been quality controlled, we applied several additional quality control techniques to ensure the continuous application of Kalman filter: (1) For any site, if the difference between  $PM_{2.5}$  at the present hour and the previous hour/next hour is both greater than 50  $\mu\text{g m}^{-3}$ , the  $PM_{2.5}$  concentration for this current hour is rejected; (2) sites that have missing data on two consecutive days are excluded; (3) only sites that have at least 26 days of data were kept. As a result, a total of 508 ground sites with hourly  $PM_{2.5}$  data from 1 June through 30 June 2012 were used for this work (see Figure 5 for the spatial distribution).



**Figure 1.** The flow diagram of the KF technique used in this study and a more detailed description can be found in section 3.1. At time step  $t = 0$ , the KF is initialized. After the initialization, the KF will update the KF parameters on its own. First, the KF will predict the current estimates of the true bias  $\hat{x}_t$  and the error covariance  $P_t$  (shown in blue box). Then Kalman gain  $\beta$  can be calculated (5). Kalman gain is used to weight the difference of the forecast error  $y_t$  and the estimate of the true bias  $\hat{x}_t$  to generate a corrected bias  $\hat{x}_{t+\Delta t}$  and also update and correct the error covariance to  $P'_t$  (shown in green box). At each time step, the corrected bias  $\hat{x}_{t+\Delta t}$  will be outputted to correct model forecasts of surface  $\text{PM}_{2.5}$  mass concentration; (3)–(6) corresponds to Equations 3–6 in section 3.1.

### 3. Methodology

Categorically, the bias correction of model outputs includes two steps. First, we correct biases of model-simulated surface  $\text{PM}_{2.5}$  for the model grids collocated with EPA ground observational sites. Here, we apply the KF, the ensemble technique or the combined technique of the two to correct the model biases. Two ensemble techniques are used, one with the equal weights and the other one with optimized weights. Second, we spread the corrected model bias to surrounding model grids without ground observations. Experimentally, implementation of the bias correction for the hindcast data requires the elaboration of the time window for making the correction, and it is described in section 3.4.

#### 3.1. Bias Correction With the Kalman Filter Technique and its Implementations

Following Delle Monache, Nipen, et al. (2006) and Kang et al. (2008), we apply the Kalman filter in a post-processing predictor bias correction manner. The KF algorithm applied here is recursive, adaptive and optimal. It is recursive in a way that the KF coefficients at any time step depend on the previous time step. It is adaptive because it can update the KF coefficients at each time step based on all the available information. It is optimal in a least squares error sense (Delle Monache, Nipen, et al., 2006).

In the current KF framework, the state variable, in this case, the true bias  $x_t$ , is defined as the difference between model (i.e., WRF-Chem) forecast of  $\text{PM}_{2.5}$  and the true (unobserved) concentration of  $\text{PM}_{2.5}$  at each time step  $t$ . This true bias is related to the bias at the previous time step ( $t - \Delta t$ ) plus a white noise  $\eta$ :

$$x_t = x_{t-\Delta t} + \eta_{t-\Delta t} \quad (1)$$

where  $\eta$  is assumed to be uncorrelated in time and normally distributed with zero mean and variance  $\sigma_\eta^2$ .  $\Delta t$  is the time interval of KF cycling (24 hr in this study) for updating the bias (for a given hour of a day) of the model outputs of  $PM_{2.5}$  (as described in section 3.4).

At the same time, the forecast error,  $y_t$ , can be defined as the difference between model forecasted  $PM_{2.5}$  and measured  $PM_{2.5}$  concentration. While the true bias  $x_t$  is not measurable, it can be related to the forecast error  $y_t$  after considering a random error  $\varepsilon_t$  in the observations. This random error is also assumed to be uncorrelated in time and normally distributed with zero mean and variance  $\sigma_\varepsilon^2$ .

$$y_t = x_t + \varepsilon_t \quad (2)$$

In this postprocessing bias correction technique, the KF system uses all the available information (the forecast and the observation) at the current time  $t$  to estimate the true bias at the future time  $(t + \Delta t)$ . Kalman (1960) has shown that the recursive predictor  $x_t$  could be represented by the following equation:

$$\hat{x}_{t+\Delta t} = \hat{x}_t + \beta_t(y_t - \hat{x}_t) \quad (3)$$

where the hat (^) symbol indicates the variables are estimates of the truth.  $\beta$  is called Kalman gain and more information is provided below.

Figure 1 shows the KF flow diagram for this study. Initial values of  $\hat{x}_t$  and  $P_t$  (denoted as  $\hat{x}_0$  and  $P_0$  in Figure 1, respectively) are needed to start the KF realization, where  $P$  is the expected mean square error ( $E[(x_t - \hat{x}_t)^2]$ ) of the model forecast. After  $\hat{x}_0$  and  $P_0$  are specified and the KF is realized once, subsequent realization of KF to predict the current state estimates of  $\hat{x}_t$  and  $P_t$  (calculated in Equation 4) are based on the information from the previous time step (shown in the blue box in Figure 1):

$$P_t = P'_{t-\Delta t} + \sigma_\eta^2 \quad (4)$$

The error covariance  $P_t$  is in turn used to calculate the Kalman gain:

$$\beta_t = \frac{P_t}{P_t + \sigma_\varepsilon^2} \quad (5)$$

which is further used to weigh the difference  $(y_t - \hat{x}_t)$  and apply the correction that was “learned” from previous errors. This correction is then added to the projected current state of  $\hat{x}_t$  to produce the estimate of true bias  $\hat{x}_{t+\Delta t}$  at the future time  $(t + \Delta t)$ ; meanwhile, the Kalman gain  $\beta_t$  is used to update and correct the error covariance  $P$ :

$$P'_t = P_t(1 - \beta_t) \quad (6)$$

where  $P'_t$  is the posterior covariance at time  $t$ . Thus, both  $\hat{x}_t$  and  $P$  are corrected and updated. Therefore, the procedures as described in Equations 4–6 can be recycled at the period of  $\Delta t$ , while the parameters can be used as inputs to the bias correction system to continue the cycle.

As the bias  $\hat{x}_{t+\Delta t}$  is estimated for the next time step, the KF-corrected model forecast of  $PM_{2.5}$  concentration ( $KF_{t+\Delta t}$ ) can be computed as

$$KF_{t+\Delta t} = M_{t+\Delta t} - \hat{x}_{t+\Delta t} \quad (7)$$

where  $M_{t+\Delta t}$  is the model forecast  $PM_{2.5}$  mass concentration at time  $(t + \Delta t)$ . In addition, as demonstrated in Delle Monache, Nipen, et al. (2006),  $\sigma_\varepsilon^2$  is also estimated using the KF framework and thus updated at each time step.

Of the KF parameters, the error ratio ( $\sigma_\eta^2/\sigma_\varepsilon^2$ ) plays an important role in the KF performance by determining the relative contribution of forecast and observed values from the recent past to estimates at the next time step. If the ratio is too high, the KF relies more on the past forecasts and is unlikely to successfully remove any errors. If the ratio is too low, the KF has the difficulty to respond to changes in bias. Delle Monache,

Nipen, et al. (2006) uses an optimal ratio of 0.01 following previous work (references therein). Kang et al. (2008) used 3-month ozone forecast time series in the continental United States to test a range of error ratios (from 0.001 to 10 with increments of 0.0001) on the KF performance. It was found that the root-mean-square error (RMSE) improved insignificantly when using an optimal error ratio at each ground station compared with using a fixed ratio of 0.06 across all stations. The same conclusion was drawn for the PM<sub>2.5</sub> forecast by Kang et al. (2010a), who investigated the effects of applying error ratios ranging from 0.01 to 10 on the KF performance using a year of PM<sub>2.5</sub> forecast in the continental United States. Here, we adopt the uniform ratio of 0.06 at all ground sites for all three models.

In the actual implementation of KF, we use a time step  $\Delta t = 24$  hr to account for the time-varying behavior of surface PM<sub>2.5</sub>, which displays a diurnal variation (Fu et al., 2018; Wang, Aegerter, et al., 2016). KF is run for each hour of the day so that the bias at a certain hour will depend on the bias at the same hour from the previous day. Here, the KF is initialized on June 4 for UTC hour = 0, ..., 23 for all ground sites, with the same KF parameters as used by Delle Monache, Nipen, et al. (2006). Data from 4 and 5 June are used for training the Kalman filter parameters. On 6 June, the updated KF parameters and the model forecast of hourly PM<sub>2.5</sub> together with surface measurements are used to generate the KF-corrected model forecast of PM<sub>2.5</sub>. The KF parameters generated at each hour of each day at each site are saved. The KF continues by reading the KF parameters from the previous day, the hourly observation and model forecast from the prior as well as training itself. For sites that have a missing observational value for an hour, the last known bias at the same hour at the current site is used to fill in the gap.

### 3.2. Ensemble Technique and its Combination With KF

Here, we consider two types of ensembles: an equally weighted and an optimized ensemble. In the equally weighted ensemble technique, we assign the three models (ensembles) with equal weights and the resultant arithmetic mean (AM) yields the final results (hereafter AME). For the optimized ensemble (hereafter OPE), the model ensemble is represented by the following equation:

$$f_{\text{OPE},h} = \mathbf{f}_h^T \boldsymbol{\omega}_h = \sum_{m=1}^M f_{m,h} \cdot \omega_{m,h} \quad (8)$$

where  $f_{\text{OPE},h}$  is the optimized model ensemble forecast of surface PM<sub>2.5</sub> for hour  $h$  and is a scalar.  $M$  is the number of ensemble models and in this work is 3;  $m$  represents the index of any model ensemble member.  $\mathbf{f}_h$  is the  $M \times 1$  vector consisting of forecasted hourly surface PM<sub>2.5</sub> for hour  $h$  from model  $m$  ( $f_{m,h}$ ). The superscript T denotes the operation of transpose of the vector.  $\boldsymbol{\omega}_h$  is the  $M \times 1$  vector of the weights for hour  $h$ , and its element  $\omega_{m,h}$  represents the optimized weight for the corresponding ensemble member, model  $m$ .

Equation 8 is established for each site for a given hour  $h$  by using (historic) forecast and observational data in past several days, and for simplicity, we neglect the subscript  $h$  in the equations below. Unless stated otherwise, optimization of  $\boldsymbol{\omega}$  is conducted for each observational site for a given hour of the next 24 hr as described in section 3.1. The optimization is realized by solving the following least squares problems:

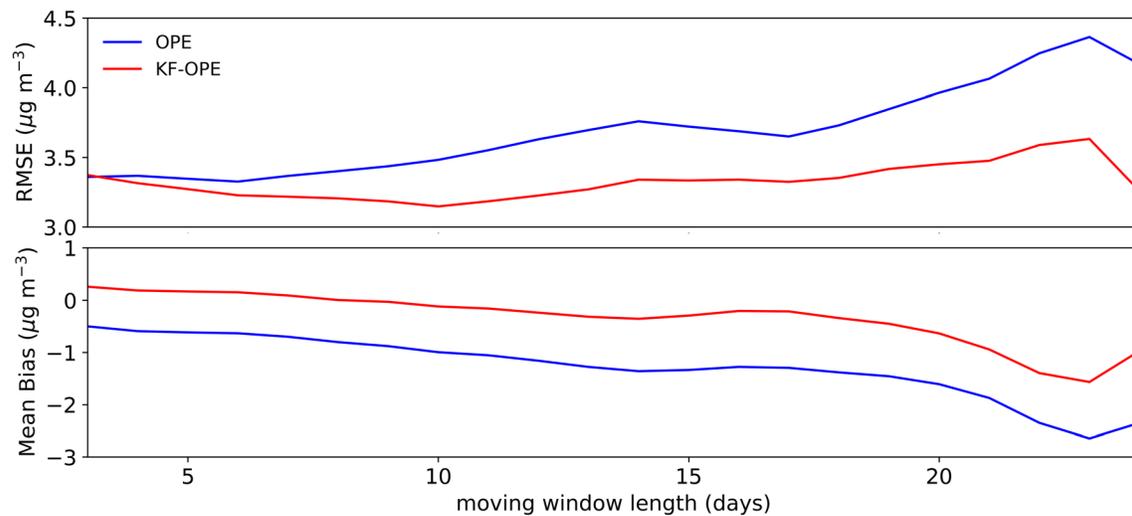
$$\underset{\boldsymbol{\omega}}{\text{argmin}} \|\mathbf{f}_s - \mathbf{f}_o\| \quad (9)$$

where

$$\mathbf{f}_s = \mathbf{F}^T \boldsymbol{\omega} = (\mathbf{f}^1 \dots \mathbf{f}^L)^T \boldsymbol{\omega} \quad (10)$$

Here  $\mathbf{f}_o$  is a  $L \times 1$  vector representing the measured hourly surface PM<sub>2.5</sub> concentration in the past  $L$  days, and  $\mathbf{f}_s$  is the model simulated counterpart after conducting the summation of weighted ensemble (e.g., Equation 8).  $\mathbf{F}$  is  $M \times L$  matrix, consisting of  $L$  columns of  $\mathbf{f}^l$ , where  $l$  is the index of past days and ranges from 1 to  $L$ ;  $\mathbf{f}^l$  is a row vector consisting of PM<sub>2.5</sub> values from  $M$  models (ensembles) on the past day of  $l$ .

Since there are two different (e.g., AME and OPE) ensemble techniques, and KF can be applied either before or after the ensemble analysis, the KF and ensemble combination, the so-called ensemble KF approach, can



**Figure 2.** Root-mean-square error (RMSE) and mean bias as a function of moving window length in the Broden-Fletcher-Goldfarb-Shanno (BFGS) optimization for the optimized ensemble (OPE, blue line) and KF-OPE (red line). The optimization was performed at each ground observational site for each moving window length. Then, the RMSE and mean bias were averaged over about 500 ground sites for each moving window length.

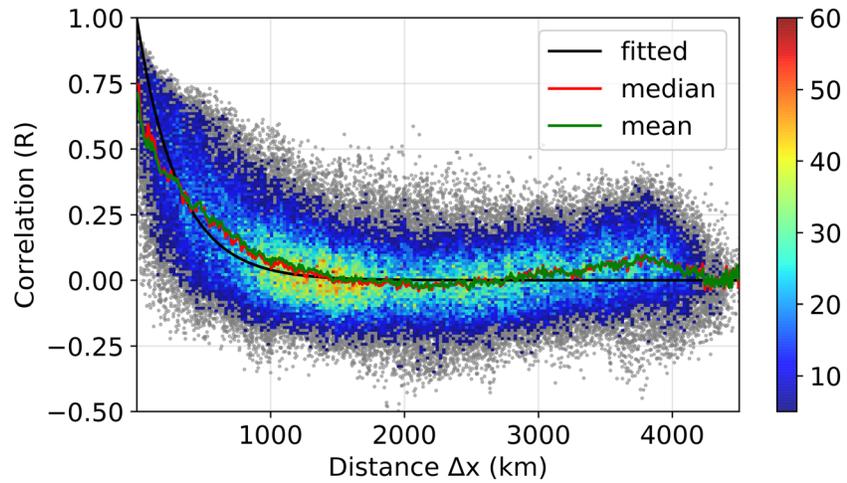
have four types: (a) AME-KF, (b) OPE-KF, (c) KF-AME, and (d) KF-OPE. For (a) and (b), ensembles are constructed first and then followed by KF correction. For (c) and (d), KF are applied first and then followed by the ensemble. The inputs for all four methods are the raw model outputs.

In the actual implementation, there are different numerical techniques to solve the least squares problems, as shown in Equations 9 and 10. Here, we choose to use the Broden-Fletcher-Goldfarb-Shanno (BFGS) (Byrd et al., 1995) method. The reason we use BFGS is because it is a generic algorithm for optimization and can be used with minimal revision in the near future when more models are available for being ensemble members. Given we only have three ensembles in this study, the optimization in Equation 9 in theory only needs data in 3 days to define weight ( $\omega$ ) for each ensemble. However, 3 days might be insufficient to characterize the error of the ensemble. Therefore, for OPE, we experiment different length of time window (e.g.,  $L$  values) with the BFGS method. Figure 2 shows the values of RMSE and mean bias (MB) of (OPE-estimated)  $\text{PM}_{2.5}$ , averaged over about 500 ground sites, as a function of the length of a moving time window,  $L$ , ranging from 3 to 24 days. It reveals that the RMSE reaches the minimum when using a moving window of less than 7 days for both OPE and KF-OPE. Meanwhile, the MB stays almost constant when the moving window varies from 3 to 7 days and after that, the MB starts to increase for both OPE and KF-OPE models. Consequently, we choose a moving window  $L$  of 5 days unless noted otherwise.

### 3.3. Bias Propagation With a SCM and its Application

After correcting the model forecasts of  $\text{PM}_{2.5}$  at each ground site using the KF technique, we expand the bias correction information to other model grid points where observations are not available. First, we estimate a radius of influence which determines the maximum distance for bias propagation. We calculate the autocorrelation for measured hourly  $\text{PM}_{2.5}$  mass concentration among all the EPA ground sites as a function of their spatial distances. As shown in Figure 3, we find a correlation length of about 300 km when the correlation falls to  $e^{-1}$ . This finding is comparable to the results of Anderson et al. (2003), which suggests that the coherent spatial scales for aerosol concentration in the lower troposphere are less than 200 km for an autocorrelation of 0.8, through examining the variation of aerosol optical properties using different datasets (surface observation, aircraft and spaceborne).

Lee et al. (2012) evaluated the performance of estimated surface  $\text{PM}_{2.5}$  across the continental United States, derived from a geostatistical kriging technique versus satellite AOD. It was found that the kriging technique resulted in more accurate estimates of surface  $\text{PM}_{2.5}$  for locations that were within  $\sim 100$  km of a ground monitoring site, while the AOD-derived  $\text{PM}_{2.5}$  shows better performance when the distance is greater than  $\sim 100$  km. Following this approach, we perform bias correction within a radius of 125 km



**Figure 3.** Correlations calculated from measured hourly  $PM_{2.5}$  mass concentration for any two of the 508 EPA AQS sites as a function of the corresponding distance between the two sites. Also shown on the scatter plot is the density of the points (the color bar). In addition, the distance is binned by a size of 10 km and the mean (green line) and median (red line) correlations are shown respectively. The fitted line for the mean correlations as a function of the distance is shown as the black line with  $R=e^{-\left(\frac{\Delta x}{l_x}\right)}$  where  $l_x = 310$  km.

through Successive Correction Method (SCM) (Kalnay, 2003). For any model grids located between 125 and 300 km away from the ground sites, AOD is used to spread the bias information, which is the focus of the companion paper.

Following the technique used by Djalalova et al. (2015) and Glahn et al. (2012), the estimates of true bias at each grid (here use  $x$ ) are calculated using an iterative objective analysis scheme:

$$x_{ij}^{n+1} = x_{ij}^n + \frac{\sum_{k=1}^K W_{ijk}^n (\hat{x}_k^0 - x_{ij}^n)}{K} \quad (11)$$

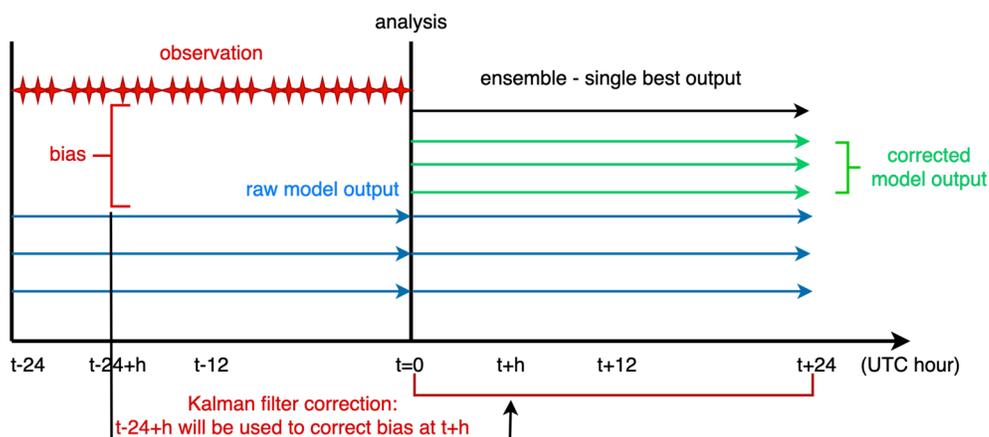
where the superscript  $n$  denotes the iteration number. The subscripts  $ij$  and  $k$  represent the model grid point and the observational site, respectively. The  $\hat{x}_k^0$  is the estimated bias from the KF technique at each ground observational site.  $K$  is the total number of observational sites.  $W_{ijk}^n$  is the weight for each observation site  $k$  to the grid point  $ij$  and is calculated as follows:

$$W_{ijk}^n = \frac{R_n^2 - r_{ijk}^2}{R_n^2 + r_{ijk}^2}, \quad (12)$$

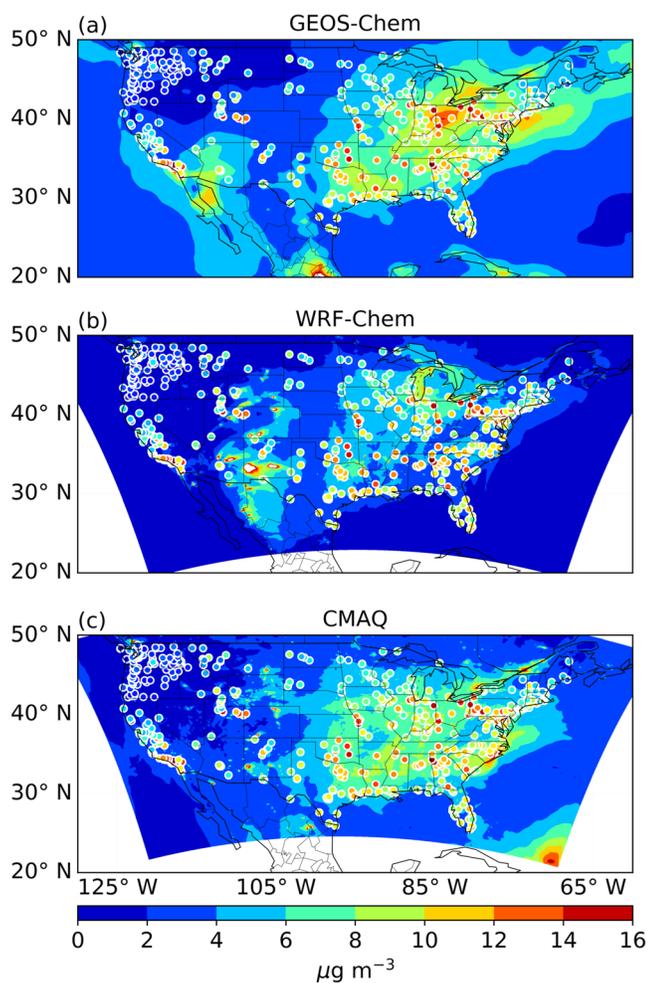
where  $r_{ijk}$  is the distance between each observational site  $k$  and grid point  $ij$ .  $R_n$  is the radius of influence, which is updated at each iteration by  $R_{n+1}^2 = \alpha R_n^2$ . Hence, the total number of observational sites  $K$  at each iteration might change with the changing  $R_n$ . The iterative process begins by proving initial values of  $R_n$ ,  $\alpha$ , and  $x_{ij}^n$  for  $n = 0$ . The initial bias of  $x_{ij}^n$  at each grid point is assumed to be 0. The initial value of  $R_n$  and  $\alpha$  are given 125 km and 0.25 respectively. Thus, the iterative process works from a large scale to a smaller scale up to 50 km, which is comparable to the model grid size of GEOS-Chem. The iterative process stops when there are no observational sites available within the distance of  $R_n$ .

### 3.4. Implementation for Bias Correction

The bias correction techniques were implemented here as if they were in a forecasting mode. As shown in Figure 4, for the KF technique, we used both the model outputs and observational data from the past 24 hr to correct model forecast bias for the future 24 hr; assuming the observational data and forecast data are readily available at the time slightly passing the last hour (23:00) of the last forecast cycle (say 23:15), and the bias correction can be made between this time (23:15) and first valid hour (0:00) of the present



**Figure 4.** Scheduling chart for Kalman filter and ensemble bias correction techniques. First, Kalman filter is applied to each model using the bias from the past 24 hr to correct the future 24 hr's output. Specifically, the bias at hour  $t - 24 + h$  ( $h = 0, \dots, 23$ ) is used to correct forecast bias at hour  $t + h$ . Then ensemble is formulated to generate the single best forecast.



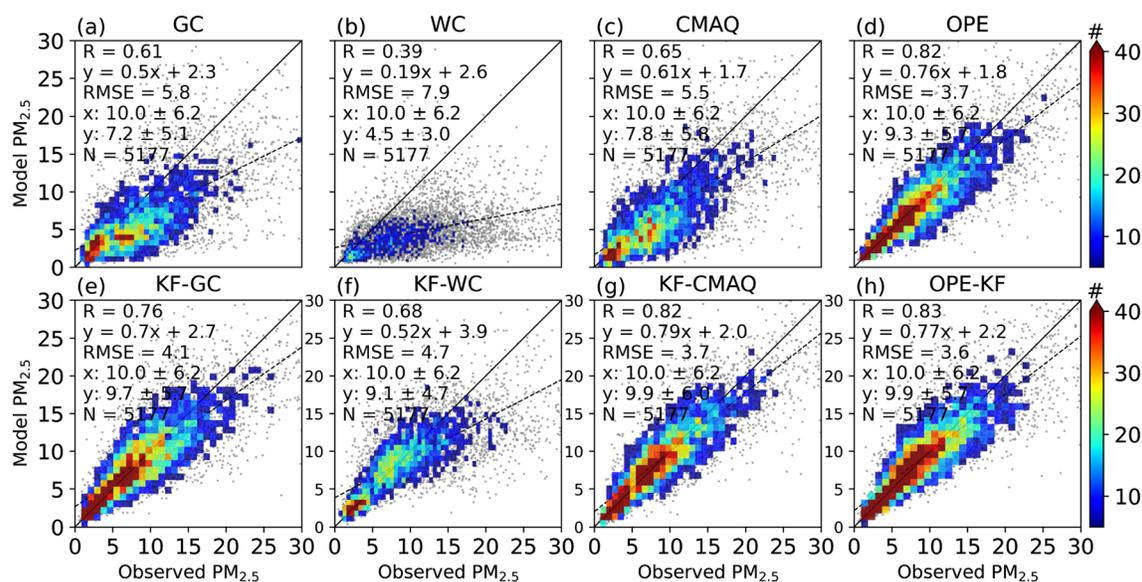
**Figure 5.** Monthly mean  $PM_{2.5}$  concentration from each model overlaid with 508 ground observational sites for June 2012: (a) GEOS-Chem, (b) WRF-Chem, and (c) CMAQ.

forecast (for tomorrow). In this manner, our bias correction technique is designed to effectively improve the daily forecast on the rolling basis of every 24 hr.

For the optimized ensemble, the weights for each site at each hour are trained using the historical hourly surface  $PM_{2.5}$  for the past 5 days. Our implementation of the bias correction only provides baseline results regarding how much the forecast can be improved, as in real-time forecasts, the model output may have larger or smaller errors. For a CTM model, its errors from meteorology and chemistry can either resonate or cancel each other in varying degrees depending on the specifics of time, location, and model parameterization scheme. Therefore, a CTM hindcast with reanalysis meteorology may often (but not necessarily always) have larger errors than that of the real-time forecast results. Nevertheless, building upon this study, future studies can further assess the impact of many other factors (such as the forecast cycle, the time window of the bias correction, and the temporal latency of observational data) on the actual implementation of the bias correction in real-time forecasting settings.

#### 4. Results

Several statistics are used to evaluate the model performances: linear correlation (denoted as  $R$ ), RMSE, and mean bias (MB). In addition, the Taylor Diagram (Taylor, 2001) is used to summarize the model evaluation results. This diagram (see Figure 9) provides a summary of the correlation ( $R$ , the cosine of the polar angle), normalized standard deviation (NSD) shown in  $x$  and  $y$  axis, normalized mean bias (NMB, the color bar) and normalized centered RMS difference (NRMSD), shown as the radius from the expected point, at which  $R$  and NSD are unity. Formulas about these statistics can be found in Zhang et al. (2019). When comparing model simulated surface  $PM_{2.5}$  with ground observations, each of the ground sites is paired with the nearest model grid and no interpolation is used. If several ground sites fall into the same model grid, they are compared with the model outputs separately.



**Figure 6.** Scatter plots of daily  $PM_{2.5}$  concentration between model (y axis) and ground observation (x axis) at 508 sites. (a–c) The evaluation for each raw model; (d–g) the evaluation for the corresponding KF-corrected model results. (d) represents the evaluation results of the optimized ensemble (OPE) of raw GEOS-Chem, WRF-Chem, and CMAQ models, and (h) shows the evaluation of the combination of the OPE and KF techniques. Also shown on the scatter plot is the correlation coefficient ( $R$ ), the root-mean-square error (RMSE), the mean  $\pm$  standard deviation for the ground observation ( $x$ ) and model ( $y$ ), the number of collocated data points ( $N$ ), the density of points (the color bar), the best fit linear regression (the dashed black line), and the 1:1 line (the solid black line).

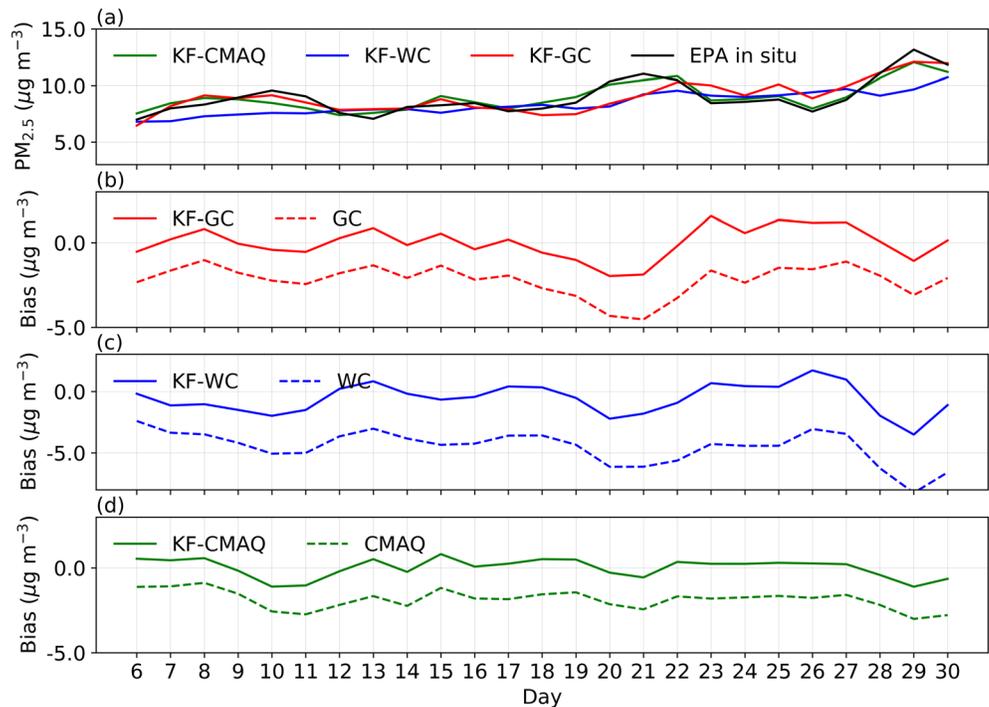
#### 4.1. Prior Ensemble Statistics of Modeled $PM_{2.5}$

Figure 5 shows the spatial distribution of monthly mean  $PM_{2.5}$  concentration from each model overlaid with ground observational data. The ground observations show higher  $PM_{2.5}$  concentration in the eastern United States and lower concentration in most of the central and western United States except for Southern California and a small part of the Rocky Mountains. All three models can capture this spatial variation to some extent with  $R$  values of 0.61, 0.55, and 0.58 for GEOS-Chem, WRF-Chem, and CMAQ respectively. However, all three models underestimate monthly mean  $PM_{2.5}$  concentration compared with ground observations, with MB values of  $-2.1$ ,  $-4.2$ , and  $-1.8 \mu g m^{-3}$  and NMB values of  $-24\%$ ,  $-48\%$ , and  $-20\%$  for GEOS-Chem, WRF-Chem, and CMAQ, respectively. The daily mean  $PM_{2.5}$  concentration comparisons are similar (Figure 6).

Underestimates of total  $PM_{2.5}$  concentration in summertime in atmospheric chemical models or operational forecast models have been a common problem. For example, by compiling 69 peer-reviewed articles spanning 2006 to 2012, Simon et al. (2012) presented a summary of various atmospheric photochemical model performances that focused on the United States or Canada. Most of the modeling work involved CMAQ and several WRF-Chem studies. It was found that model evaluation results show more negative than positive biases for  $PM_{2.5}$  species and total  $PM_{2.5}$  mass concentration. McKeen et al. (2007) compared real-time forecasts of surface  $PM_{2.5}$  concentrations from seven air quality forecasting models (including several WRF-Chem models with different configurations) with ground observations in the northeastern United States and southeastern Canada during the summer of 2004 and found that, in general, the surface  $PM_{2.5}$  concentrations were underpredicted across the regional scale. Other past work (Fast et al., 2014; Misenis & Zhang, 2010; Wu et al., 2017) have also found underestimates of surface  $PM_{2.5}$  in WRF-Chem simulations over different parts of the United States.  $PM_{2.5}$  is generally underestimated in CMAQ models in the summertime shown by previous work (Appel et al., 2017; Yu et al., 2008).

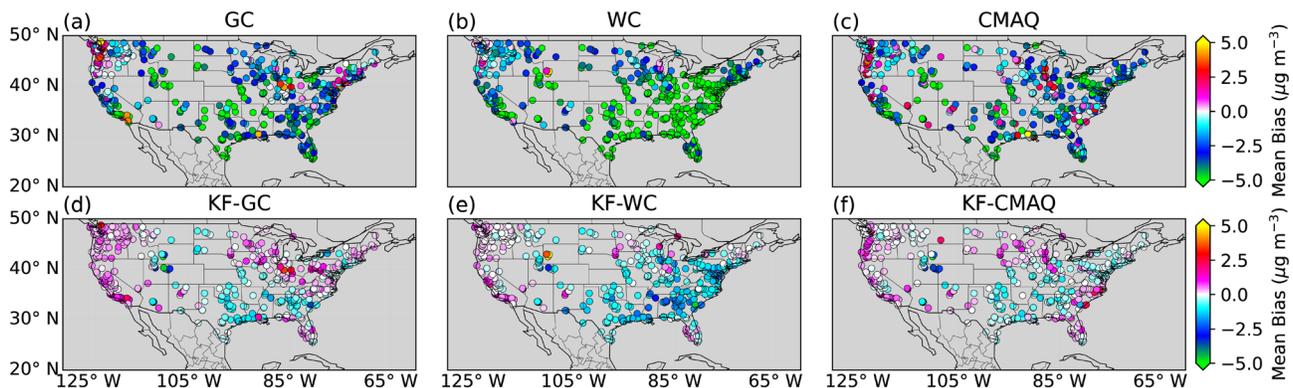
#### 4.2. Kalman Filter for Each Model Output

As shown in Figure 6, the KF technique improves each individual model performance with increased  $R$  values and decreased RMSE and MB. The  $R$  value increases systematically from the range of 0.39–0.65 for

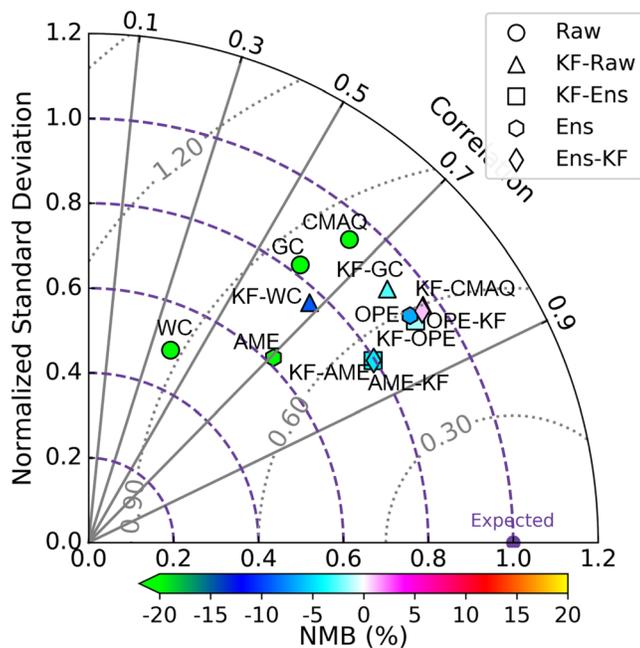


**Figure 7.** (a) Time series of daily mean  $PM_{2.5}$  concentration averaged over 508 ground sites from 6 to 30 June for the KF-corrected models and EPA in situ observation. (b–d) The time series of the bias between each model and measured daily mean  $PM_{2.5}$  concentration averaged over the ground sites, respectively.

the raw model to 0.68–0.82 for the KF-corrected models. The RMSE decreases by 29%, 41%, and 33% for GEOS-Chem, WRF-Chem and CMAQ, respectively. The MB decreases from 2.2–4.5 to 0.1–0.9  $\mu\text{g m}^{-3}$ , which indicates the successful removal of systematic bias in the models. Figure 7 shows the time-series of the bias of raw models and KF-corrected models. Again, all three raw models in general underestimate  $PM_{2.5}$  concentrations compared to ground observations from the beginning of June to the end of the month. Overall, the KF technique can adjust the model results and reduce the model bias. Of the three models, the KF-corrected CMAQ shows the best performance mainly because the raw CMAQ captures the magnitude and variation of the surface  $PM_{2.5}$  better than the other two models. The ground



**Figure 8.** Mean bias between model simulated and measured  $PM_{2.5}$  at 508 ground sites. (a–c) The mean bias for the raw models; (d–f) the mean bias for the corresponding KF-corrected models.



**Figure 9.** Taylor diagram of evaluating modeled daily mean  $PM_{2.5}$  concentration against ground observations at 508 sites. The circles represent each raw model comparison with ground observations, while the triangles represent the comparisons of KF-corrected model daily  $PM_{2.5}$  concentration and ground observation. The hexagons represent the evaluation of the ensembles (AME and OPE). The squares and diamonds show the comparison of KF-ensembles and ensembles-KF, respectively. The color bar denotes the normalized mean bias (NMB, %).

observations of surface  $PM_{2.5}$  averaged across the sites shows 3 peaks during the month (10, 21, and 29 June). For the second peak around 21 June, both WRF-Chem and GEOS-Chem show larger biases than on previous days. As a result, the KF slowly responds to the change because the KF failed to anticipate a larger bias when the biases in the past few days are relatively constant and smaller.

Figure 8 shows the spatial distribution of the model mean biases. Again, negative biases are dominant at most of the sites for all three of the models, especially for WRF-Chem. For GEOS-Chem and CMAQ, most sites with negative biases are located in the eastern and central United States as well as a part of California. The implementation of the KF technique brings the model bias within  $2 \mu g m^{-3}$  for all three of the models. Similar results are also found in the work of Kang et al. (2010a), who implemented the KF technique with the CMAQ model to improve surface  $PM_{2.5}$  forecasts in the continental United States. This finding further confirms the effectiveness of the Kalman filter in reducing model bias and thereby improve model forecasting performance.

### 4.3. AME and OPE

The equally weighted ensemble of raw models (AME) has some skill in enhancing the correlation with respect to the raw models (Figure 9). Using the AME resulted in an RMSE decrease of  $\sim 29\%$  compared with raw WRF-Chem model output, but there was insignificant change in CMAQ and GEOS-Chem model outputs. As anticipated, the ensemble could reduce some of the unsystematic error. The optimized ensemble (OPE) is found to perform better than the AME in terms of MB, RMSE, and  $R$  (Table 2). Additionally, using the OPE results in less variation between model output and observations than using the AME (Figure 9).

Our work reveals that weighted ensemble modeling has the potential to improve model forecasting skills compared with raw models, especially when some of the raw models in the ensemble group present some deficiency (Pagowski et al., 2005).

### 4.4. Ensemble Comparison and Combination With KF

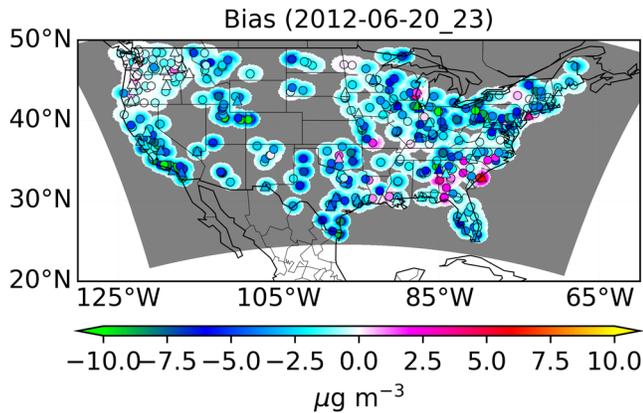
Figure 9 summarizes the performance of individual models, KF-corrected models and ensemble models. Again, the KF-corrected models show improved performance compared with their respective raw models. The ensembles (AME or OPE) without any KF bias correction could both improve raw model performance by decreasing biases and increasing correlations.

The statistics of the different ensemble techniques are presented in Table 2. The combination of ensemble-KF or KF-ensemble shows better performance than the OPE ensemble alone, also shown in Figure 6, reflecting the effectiveness of the KF bias correction technique. KF-AME and AME-KF show similar results in terms of MB, RMSE, and  $R$ . Using OPE with KF (either OPE-KF or KF-OPE) is found to have smaller MBs compared with using AME with KF. This result agrees with findings of Djalalova et al. (2010), who showed that the Singular Value Decomposition (SVD) ensemble of KF-corrected model performance for surface  $PM_{2.5}$  is slightly better than the equally weighted ensemble of KF-corrected models. Djalalova et al. (2010) focused on surface  $PM_{2.5}$  in the state of Texas and optimized the weights across the whole domain.

Overall, the KF-OPE technique shows the best performance, slightly better than OPE-KF. Since the errors of each model, especially error covariance among different models, cannot be fully characterized, our ensemble technique more or less can be viewed as empirically based. Therefore, caution needs to be exercised to interpret the results, especially the relative performance of each techniques. Nevertheless, it appears that KF is shown to be effective for the correction of systematic errors of a

**Table 2**  
Summary Statistics for the Ensemble Technique

	MB ( $\mu g m^{-3}$ )	NMB (%)	RMSE ( $\mu g m^{-3}$ )	$R$
AME	-3.47	-34.8	5.61	0.71
OPE	-0.66	-6.6	3.69	0.82
KF-AME	-0.46	-4.6	3.37	0.84
KF-OPE	0.10	1.08	3.52	0.82
AME-KF	-0.43	-4.3	3.37	0.84
OPE-KF	-0.13	-1.3	3.62	0.83



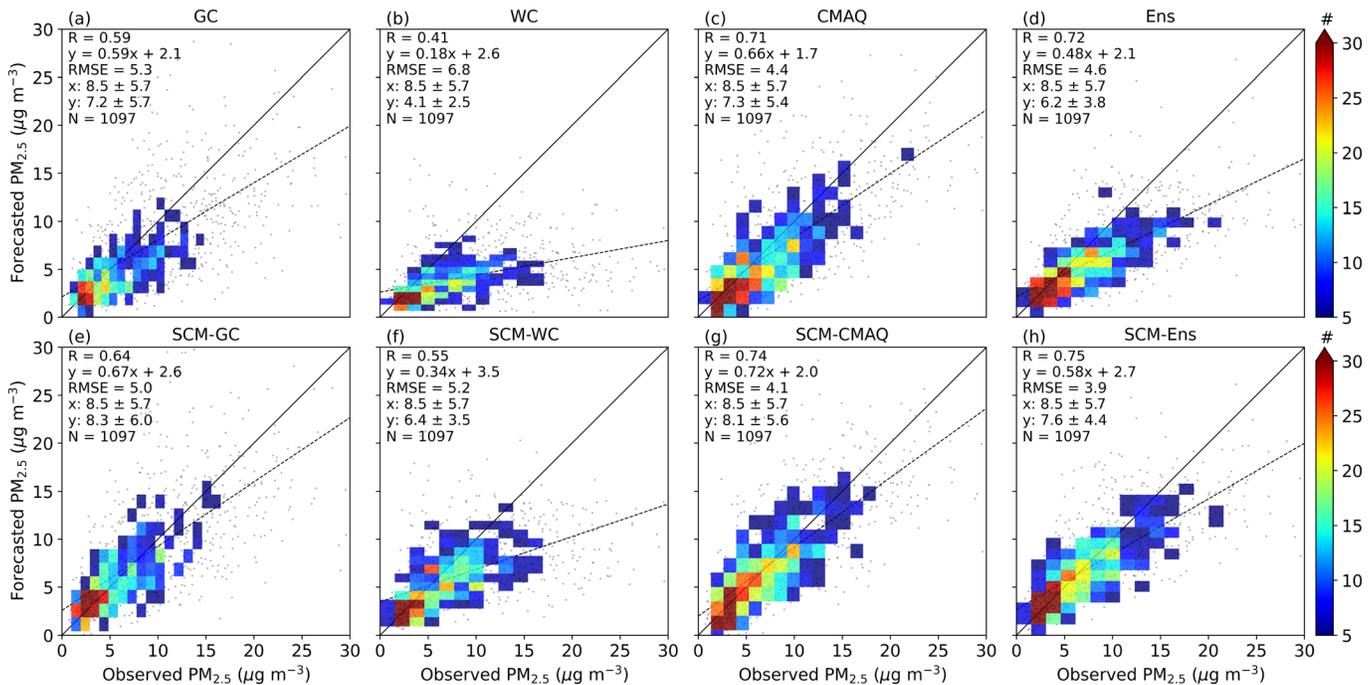
**Figure 10.** Map of SCM-corrected CMAQ model mean bias of surface  $PM_{2.5}$  at UTC hour 23 on 20 June. The circles represent EPA sites where KF bias correction is applied and subsequently used by SCM for model bias correction in the nearby (within 125 km radius) area. The triangles are sites used for SCM evaluation. Gray areas in the map are model grids that are not corrected by SCM.

single model (often with underlying assumption that the model errors are Gaussian and is applied here with statistics in a moving time window of 1 day). In comparison, the OPE process in this work can be somehow viewed as a linear regression. Such linear regression (Equations 8–10) includes three weights and does not include a constant term. As a result, the OPE here could not remove the systematic bias more effectively compared with the KF technique but may handle some unsystematic bias. This conjecture helps to elucidate that KF-OPE might be more effective than OPE-KF. Again, the bias correction method in this study, by design, is empirical, and further study from the theoretical point of view is needed.

**4.5. SCM Results**

Figure 10 shows the spread of KF-generated bias for surface  $PM_{2.5}$  in the CMAQ model using SCM. As seen in the figure, CMAQ model grid points that have EPA sites within a distance of 125 km show smoothly corrected biases. Similar smoothness is found in GEOS-Chem and WRF-Chem model corrected bias (not shown here). The uneven spatial distribution of surface  $PM_{2.5}$  monitoring sites leads to more model grid points in the eastern and western U.S. being corrected and most model grid points in the central Great Plains and the Rocky Mountains are left out.

To evaluate the performance of the SCM correction, we randomly select ~10% of the total 508 sites shown in Figure 10 as validation sites. These sites will not be used for KF bias correction. Figure 11 shows the performance of the raw model and SCM-corrected surface  $PM_{2.5}$  at these evaluation sites. For all of the three models, SCM correction could augment model performance, and WRF-Chem shows greater improvement. Mean Bias decreases from 1.3 to  $0.2 \mu g m^{-3}$  for GEOS-Chem, from 4.4 to  $2.1 \mu g m^{-3}$  for WRF-Chem and from 1.2 to



**Figure 11.** Scatter plots of daily  $PM_{2.5}$  concentration between model and ground observation for SCM evaluation sites. (a–c) The raw model  $PM_{2.5}$  from all the SCM evaluation sites (triangles in Figure 10); (e–g) the corresponding SCM-corrected module simulated  $PM_{2.5}$ . (d and h) The ensemble of raw GEOS-Chem, WRF-Chem, and CMAQ models and SCM-corrected models, respectively. Also shown on the scatter plot is the correlation coefficient ( $R$ ), the root-mean-square error (RMSE), the number of collocated data points ( $N$ ), the density of points (the color bar), the best fit linear regress line (the dashed black line), and the 1:1 line (the solid black line).

$0.4 \mu\text{g m}^{-3}$  for CMAQ. The RMSE decreases from 5.3 to  $5.0 \mu\text{g m}^{-3}$ , 6.8 to  $5.2 \mu\text{g m}^{-3}$ , and 4.4 to  $4.1 \mu\text{g m}^{-3}$ . The  $R$  values increase from 0.59 to 0.64, 0.41 to 0.55, and 0.71 to 0.74, respectively. The equally weighted ensemble of the three SCM-corrected models showed improved estimates of surface  $\text{PM}_{2.5}$  than the equally weighted ensemble of the raw models, SCM-corrected GEOS-Chem and SCM-corrected WRF-Chem alone but only slightly improve the performance of SCM-corrected CMAQ.

## 5. Conclusions and Discussion

As the first of a two-part study, we focused on improving surface  $\text{PM}_{2.5}$  forecasting in non-rural areas of the continental United States using model predictions and ground observational data. We applied a multimodel approach together with bias correction techniques including Kalman filter application and ensemble modeling. In addition, a successive correction technique was used to spread the previous bias information to areas where ground observations were not available. Model ensemble members included three of the widely used atmospheric chemical transport models: GEOS-Chem, WRF-Chem, and WRF-CMAQ. Hourly model output of surface  $\text{PM}_{2.5}$  mass concentration and ground observations from about 500 EPA AQS sites for the month of June 2012 were used. All three models could generally capture the spatial variation of surface  $\text{PM}_{2.5}$  revealed by ground observations, with higher concentrations in the eastern U.S. and lower concentrations in the central and western United States except for southern California. However, all three models underestimated  $\text{PM}_{2.5}$  concentrations compared to ground observations, with MB of  $-2.8$ ,  $-4.5$ , and  $-2.2 \mu\text{g m}^{-3}$  and RMSE of 5.8, 7.9, and  $5.5 \mu\text{g m}^{-3}$  for GEOS-Chem, WRF-Chem, and CMAQ, respectively. Overall, CMAQ performed the best and WRF-Chem performed the worst. Additional model sensitivity studies would be necessary to understand the causes.

The KF bias correction technique resulted in increased correlation, decreased RMSE and MB, when comparing each model output with ground observations. Spatially, the MB values decreased to within  $2 \mu\text{g m}^{-3}$  for most of the sites. Our results are consistent with previous work of Kang et al. (2010a), which confirms the robustness and efficiency of the KF technique in improving model forecasts by reducing systematic errors. Several combinations of ensembles were then created. An equally weighted ensemble (AME) and optimized ensemble (OPE) were used. The AME performed better than each raw model alone in terms of correlation but did not reduce the RMSE significantly compared with CMAQ and GEOS-Chem, mainly because of large underestimates by the WRF-Chem model. In contrast, the OPE resulted in greater improvement than the AME, with increased correlation and decreased biases. Next, two types of ensemble and KF combination were tested: KF-AME/OPE (KF first followed by ensemble) and vice versa. The ensemble with the KF correction technique performed better than the OPE ensemble alone, which demonstrates that applying the KF is effective especially when raw model skills are low. Overall, the KF-OPE showed the best results with the RMSE decreasing from 5.61 to  $3.52 \mu\text{g m}^{-3}$  (37%), the MB changing from  $-3.47$  to  $0.10 \mu\text{g m}^{-3}$ , and the  $R$  value increasing from 0.71 to 0.82 (15%) compared with the AME. We understand that our ensemble members include fewer models compared with previous work (Djalalova et al., 2010; McKeen et al., 2007), but these models incorporate different meteorology fields and chemical mechanism and can still be useful to demonstrate the impacts of ensemble modeling.

A correlation length of 300 km was found from the autocorrelation of measured hourly surface  $\text{PM}_{2.5}$  mass concentration over about 500 ground sites. We used this radius as the limit distance to correct any model estimates at grid points neighboring ground observational sites. SCM was used within the 125 km distance radius. Corrections for GEOS-Chem, WRF-Chem, and CMAQ were conducted respectively and all showed increased  $R$  values, decreased RMSE and MB. The equally weighted ensemble of the three SCM-corrected models showed improved results than the SCM-corrected GEOS-Chem and WRF-Chem alone but did not show significant improvements than the SCM-corrected CMAQ. Therefore, we recommend using the optimized ensemble (OPE) combined with the Kalman filter for regions with available ground observations followed by the equally weighted ensemble of the SCM correction for the nearby regions, 125 km away from the ground observations. For regions that do not have ground observations in the vicinity of 125 km, satellite remote sensing data will be used to correct model bias, which will be the focus of part two of our work.

Finally, we note that our modeling results only focus on 1 month of June 2012 and future work could be expanded to other seasons to provide a complete picture. Since the bias correction methods developed

here are based on model hindcast results, future work is needed to apply and assess the efficacy of this bias correction system in operational air quality forecasting; the confounding issues related to the time window of the bias correction, the temporal latency of observational data, and the real-time forecast scheduling (Kang et al., 2010b; Ma et al., 2018) deserve dedicated investigation.

### Disclaimer

The research described in this article has been reviewed by the U.S. Environmental Protection Agency (EPA) and approved for publication. Approval does not signify that the contents necessarily reflect the views and the policies of the agency nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

### Data Availability Statement

GEOS-Chem is an open-access model that can be downloaded online (<http://acmg.seas.harvard.edu/geos/>). WRF-Chem is also an open-access model that can be acquired online (<https://www2.mmm.ucar.edu/wrf/users/downloads.html>). CMAQ model outputs can be obtained online (<https://www.epa.gov/hesc/remote-sensing-information-gateway>). The results used in this work are available online (<https://doi.org/10.25820/data.006123>).

### Acknowledgments

This work was supported in part by the NASA Suomi NPP Program and Applied Science Program managed by John A. Haynes and Lawrence A. Friedl. We appreciate doctoral candidate Pake Melland from the Department of Math at the University of Iowa for his helpful discussion on the Kalman filter technique. We also thank Dr. Edward Hyer from the Naval Research Laboratory for providing FLAMBE emission inventories. We thank the U.S. EPA for providing the hourly measured surface PM<sub>2.5</sub> mass concentration data, which are available at [https://aqs.epa.gov/aqsweb/documents/data\\_mart\\_welcome.html](https://aqs.epa.gov/aqsweb/documents/data_mart_welcome.html) (last accessed 03 October 2019). For further information, please contact Jun Wang ([jun-wang-1@uiowa.edu](mailto:jun-wang-1@uiowa.edu)) and Huanxin Zhang ([huanxin-zhang@uiowa.edu](mailto:huanxin-zhang@uiowa.edu)).

### References

- Ackermann, I. J., Hass, H., Memmesheimer, M., Ebel, A., Binkowski, F. S., & Shankar, U. (1998). Modal aerosol dynamics model for Europe: Development and first applications. *Atmospheric Environment*, *32*(17), 2981–2999. [https://doi.org/10.1016/s1352-2310\(98\)00006-5](https://doi.org/10.1016/s1352-2310(98)00006-5)
- Al-Saadi, J., Szykman, J., Pierce, R. B., Kittaka, C., Neil, D., Chu, D. A., et al. (2005). Improving national air quality forecasts with satellite aerosol observations. *Bulletin of the American Meteorological Society*, *86*(9), 1249–1262. <https://doi.org/10.1175/bams-86-9-1249>
- Anderson, T. L., Charlson, R. J., Winker, D. M., Ogren, J. A., & Holmén, K. (2003). Mesoscale variations of tropospheric aerosols. *Journal of the Atmospheric Sciences*, *60*(1), 119–136. [https://doi.org/10.1175/1520-0469\(2003\)060<0119:mvota>2.0.co;2](https://doi.org/10.1175/1520-0469(2003)060<0119:mvota>2.0.co;2)
- Appel, K., Pouliot, G., Simon, H., Sarwar, G., Pye, H., Napelenok, S., et al. (2013). Evaluation of dust and trace metal estimates from the Community Multiscale Air Quality (CMAQ) model version 5.0. *Geoscientific Model Development*, *6*(4), 883–899. <https://doi.org/10.5194/gmd-6-883-2013>
- Appel, K. W., Napelenok, S. L., Foley, K. M., Pye, H. O. T., Hogrefe, C., Luecken, D. J., et al. (2017). Description and evaluation of the Community Multiscale Air Quality (CMAQ) modeling system version 5.1. *Geoscientific Model Development*, *10*(4), 1703–1732. <https://doi.org/10.5194/gmd-10-1703-2017> <https://www.geosci-model-dev.net/10/1703/2017/>
- Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., et al. (2001). Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research - Atmospheres*, *106*(D19), 23,073–23,095. <https://doi.org/10.1029/2001jd000807>
- Binkowski, F. S., & Roselle, S. J. (2003). Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component 1. Model description. *Journal of Geophysical Research - Atmospheres*, *108*(D6). <https://doi.org/10.1029/2001jd001409>
- Binkowski, F. S., & Shankar, U. (1995). The Regional Particulate Matter Model: 1. Model description and preliminary results. *Journal of Geophysical Research - Atmospheres*, *100*(D12), 26,191–26,209. <https://doi.org/10.1029/95jd02093>
- Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G., et al. (2015). Data assimilation in atmospheric chemistry models: Current status and future prospects for coupled chemistry meteorology models. *Atmospheric Chemistry and Physics*, *15*(10), 5325–5358. <https://doi.org/10.5194/acp-15-5325-2015>
- Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*(5), 1190–1208. <https://doi.org/10.1137/0916069>
- Byun, D., & Schere, K. L. (2006). Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Applied Mechanics Reviews*, *59*(2), 51–77. <https://doi.org/10.1115/1.2128636>
- Carlton, A. G., Bhave, P. V., Napelenok, S. L., Edney, E. O., Sarwar, G., Pinder, R. W., et al. (2010). Model representation of secondary organic aerosol in CMAQv4. 7. *Environmental Science & Technology*, *44*(22), 8553–8560. <https://doi.org/10.1021/es100636q>
- Carmichael, G. R., Sandu, A., Chai, T., Daescu, D. N., Constantinescu, E. M., & Tang, Y. (2008). Predicting air quality: Improvements through advanced methods to integrate models and measurements. *Journal of Computational Physics*, *227*(7), 3540–3571. <https://doi.org/10.1016/j.jcp.2007.02.024>
- Chai, T., Kim, H. C., Pan, L., Lee, P., & Tong, D. (2017). Impact of Moderate Resolution Imaging Spectroradiometer Aerosol Optical Depth and AirNow PM<sub>2.5</sub> assimilation on Community Multi-scale Air Quality aerosol predictions over the contiguous United States. *Journal of Geophysical Research - Atmospheres*, *122*, 5399–5415. <https://doi.org/10.1002/2016jd026295>
- Chen, D., Wang, Y., McElroy, M. B., He, K., Yantosca, R. M., & Sager, P. L. (2009). Regional CO pollution and export in China simulated by the high-resolution nested-grid GEOS-Chem model. *Atmospheric Chemistry and Physics*, *9*(11), 3825–3839. <https://doi.org/10.5194/acp-9-3825-2009>
- Chu, D. A., Kaufman, Y., Zibordi, G., Chern, J., Mao, J., Li, C., & Holben, B. (2003). Global monitoring of air pollution over land from the Earth Observing System-Terra Moderate Resolution Imaging Spectroradiometer (MODIS). *Journal of Geophysical Research - Atmospheres*, *108*(D21). <https://doi.org/10.1029/2002jd003179>
- Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., et al. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the Global Burden of Diseases Study 2015. *The Lancet*, *389*(10082), 1907–1918. [https://doi.org/10.1016/s0140-6736\(17\)30505-6](https://doi.org/10.1016/s0140-6736(17)30505-6)

- De Ridder, K., Kumar, U., Lauwaet, D., Blyth, L., & Lefebvre, W. (2012). Kalman filter-based air quality forecast adjustment. *Atmospheric Environment*, *50*, 381–384. <https://doi.org/10.1016/j.atmosenv.2012.01.032>
- Delle Monache, L., Deng, X., Zhou, Y., & Stull, R. (2006). Ozone ensemble forecasts: 1. A new ensemble design. *Journal of Geophysical Research - Atmospheres*, *111*, D05307. <https://doi.org/10.1029/2005jd006310>
- Delle Monache, L., Nipen, T., Deng, X., Zhou, Y., & Stull, R. (2006). Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction. *Journal of Geophysical Research - Atmospheres*, *111*, D05308. <https://doi.org/10.1029/2005jd006311>
- Delle Monache, L., Nipen, T., Liu, Y., Roux, G., & Stull, R. (2011). Kalman filter and analog schemes to postprocess numerical weather predictions. *Monthly Weather Review*, *139*(11), 3554–3570. <https://doi.org/10.1175/2011mwr3653.1>
- Delle Monache, L., & Stull, R. (2003). An ensemble air-quality forecast over western Europe during an ozone episode. *Atmospheric Environment*, *37*(25), 3469–3474. [https://doi.org/10.1016/s1352-2310\(03\)00475-8](https://doi.org/10.1016/s1352-2310(03)00475-8)
- Delle Monache, L., Wilczak, J., McKeen, S., Grell, G., Pagowski, M., Peckham, S., et al. (2008). A Kalman-filter bias correction method applied to deterministic, ensemble averaged and probabilistic forecasts of surface ozone. *Tellus Series B: Chemical and Physical Meteorology*, *60*(2), 238–249. <https://doi.org/10.1111/j.1600-0889.2007.00332.x>
- Djalalova, I., Delle Monache, L., & Wilczak, J. (2015). PM 2.5 analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model. *Atmospheric Environment*, *108*, 76–87. <https://doi.org/10.1016/j.atmosenv.2015.02.021>
- Djalalova, I., Wilczak, J., McKeen, S., Grell, G., Peckham, S., Pagowski, M., et al. (2010). Ensemble and bias-correction techniques for air quality model forecasts of surface O<sub>3</sub> and PM<sub>2.5</sub> during the TEXAQs-II experiment of 2006. *Atmospheric Environment*, *44*(4), 455–467. <https://doi.org/10.1016/j.atmosenv.2009.11.007>
- Dye, T., Chan, A., Anderson, C., Strohm, D., Wayland, R., & White, J. (2004). *From raw air quality data to the nightly news: An overview of how EPA's AIRNow program operates*. Petaluma, CA, and NC: American Meteorological Society Annual Meeting, AMS Sixth Conference on Atmospheric Chemistry: Air Quality in Megacities, Seattle, WA, Sonoma Technology, Inc., and U.S. Environmental Protection Agency, Research Triangle Park, NC (STI-902107–2405).
- Fairlie, T. D., Jacob, D. J., Dibb, J. E., Alexander, B., Avery, M. A., Donkelaar, A. v., & Zhang, L. (2010). Impact of mineral dust on nitrate, sulfate, and ozone in transpacific Asian pollution plumes. *Atmospheric Chemistry and Physics*, *10*(8), 3999–4012. <https://doi.org/10.5194/acp-10-3999-2010>
- Fairlie, T. D., Jacob, D. J., & Park, R. J. (2007). The impact of transpacific transport of mineral dust in the United States. *Atmospheric Environment*, *41*(6), 1251–1266. <https://doi.org/10.1016/j.atmosenv.2006.09.048>
- Fast, J., Craven, J., Metcalf, A., & Seinfeld, J. (2014). Modeling regional aerosol and aerosol precursor variability over California and its sensitivity to emissions and long-range transport during the 2010 CalNex and CARES campaigns. *Atmospheric Chemistry and Physics*, *14*(18), 10,013–10,060. <https://doi.org/10.5194/acp-14-10013-2014>
- Fast, J. D., Gustafson, W. I. Jr., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., et al. (2006). Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model. *Journal of Geophysical Research - Atmospheres*, *111*, D21305. <https://doi.org/10.1029/2005jd006721>
- Forouzanfar, M. H., Afshin, A., Alexander, L. T., Anderson, H. R., Bhutta, Z. A., Biryukov, S., et al. (2016). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, *388*(10053), 1659–1724. [https://doi.org/10.1016/S0140-6736\(16\)31679-8](https://doi.org/10.1016/S0140-6736(16)31679-8)
- Fountoukis, C., & Nenes, A. (2007). ISORROPIA II: A computationally efficient thermodynamic equilibrium model for K<sup>+</sup>–Ca<sup>2+</sup>–Mg<sup>2+</sup>–NH<sub>4</sub><sup>+</sup>–Na<sup>+</sup>–SO<sub>4</sub><sup>2-</sup>–NO<sub>3</sub><sup>-</sup>–Cl<sup>-</sup>–H<sub>2</sub>O aerosols. *Atmospheric Chemistry and Physics*, *7*(17), 4639–4659. <https://doi.org/10.5194/acp-7-4639-2007>
- Fu, D., Xia, X., Duan, M., Zhang, X., Li, X., Wang, J., & Liu, J. (2018). Mapping nighttime PM<sub>2.5</sub> from VIIRS DNB using a linear mixed-effect model. *Atmospheric Environment*, *178*, 214–222. <https://doi.org/10.1016/j.atmosenv.2018.02.001>
- Ge, C., Wang, J., & Reid, J. (2014). Mesoscale modeling of smoke transport over the Southeast Asian Maritime Continent: Coupling of smoke direct radiative effect below and above the low-level clouds. *Atmospheric Chemistry and Physics*, *14*(1), 159–174. <https://doi.org/10.5194/acp-14-159-2014>
- Ge, C., Wang, J., Reid, J. S., Posselt, D. J., Xian, P., & Hyer, E. (2017). Mesoscale modeling of smoke transport from equatorial Southeast Asian Maritime Continent to the Philippines: First comparison of ensemble analysis with in situ observations. *Journal of Geophysical Research - Atmospheres*, *122*, 5380–5398. <https://doi.org/10.1002/2016jd026241>
- Glahn, B., Im, J.-S., & Wagner, G. (2012). *Objective analysis of MOS forecasts and observations in sparse data regions*. 22e26 January, New Orleans, LA: 92nd AMS Annual Meeting 21th Conference on Probability and Statistics.
- Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled “online” chemistry within the WRF model. *Atmospheric Environment*, *39*(37), 6957–6975. <https://doi.org/10.1016/j.atmosenv.2005.04.027>
- Guenther, A., Jiang, X., Heald, C., Sakulyanontvittaya, T., Duhl, T., Emmons, L., & Wang, X. (2012). The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions. *Geoscientific Model Development*, *5*. <https://doi.org/10.5194/gmd-5-1471-2012>
- Hand, J. L., Gill, T. E., & Schichtel, B. A. (2017). Spatial and seasonal variability in fine mineral dust and coarse aerosol mass at remote sites across the United States. *Journal of Geophysical Research - Atmospheres*, *122*, 3080–3097. <https://doi.org/10.1002/2016JD026290>
- Hoff, R. M., & Christopher, S. A. (2009). Remote sensing of particulate pollution from space: Have we reached the promised land? *Journal of the Air & Waste Management Association*, *59*(6), 645–675. <https://doi.org/10.3155/1047-3289.59.6.645>
- Hu, L., Keller, C. A., Long, M. S., Sherwen, T., Auer, B., Da Silva, A., et al. (2018). Global simulation of tropospheric chemistry at 12.5 km resolution: Performance and evaluation of the GEOS-Chem chemical module (v10–1) within the NASA GEOS Earth system model (GEOS-5 ESM). *Geoscientific Model Development*, *11*, 4603–4620. <https://doi.org/10.5194/gmd-11-4603-2018>
- Huang, J., McQueen, J., Wilczak, J., Djalalova, I., Stajner, I., Shafran, P., et al. (2017). Improving NOAA NAQFC PM<sub>2.5</sub> predictions with a bias correction approach. *Weather and Forecasting*, *32*(2), 407–421. <https://doi.org/10.1175/waf-d-16-0118.1>
- Jaeglé, L., Quinn, P., Bates, T., Alexander, B., & Lin, J.-T. (2011). Global distribution of sea salt aerosols: New constraints from in situ and remote sensing observations. *Atmospheric Chemistry and Physics*, *11*(7), 3137–3157. <https://doi.org/10.5194/acp-11-3137-2011>
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(1), 35–45. <https://doi.org/10.1115/1.3662552>
- Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. New York, NY: Cambridge University Press.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–472. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:tnyrp>2.0.co;2](https://doi.org/10.1175/1520-0477(1996)077<0437:tnyrp>2.0.co;2)

- Kang, D., Mathur, R., & Trivikrama Rao, S. (2010a). Assessment of bias-adjusted PM 2.5 air quality forecasts over the continental United States during 2007. *Geoscientific Model Development*, 3(1), 309–320. <https://doi.org/10.5194/gmd-3-309-2010>
- Kang, D., Mathur, R., & Trivikrama Rao, S. (2010b). Real-time bias-adjusted O3 and PM2.5 air quality index forecasts and their performance evaluations over the continental United States. *Atmospheric Environment*, 44(18), 2203–2212. <https://doi.org/10.1016/j.atmosenv.2010.03.017>
- Kang, D., Mathur, R., Trivikrama Rao, S., & Yu, S. (2008). Bias adjustment techniques for improving ozone air quality forecasts. *Journal of Geophysical Research - Atmospheres*, 113, D23308. <https://doi.org/10.1029/2008jd010151>
- Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., et al. (2001). The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society*, 82(2), 247–268. [https://doi.org/10.1175/1520-0477\(2001\)082<0247:tynyrm>2.3.co;2](https://doi.org/10.1175/1520-0477(2001)082<0247:tynyrm>2.3.co;2)
- Kuhns, H., Knipping, E. M., & Vukovich, J. M. (2005). Development of a United States–Mexico emissions inventory for the big bend regional aerosol and visibility observational (BRAVO) study. *Journal of the Air & Waste Management Association*, 55(5), 677–692. <https://doi.org/10.1080/10473289.2005.10464648>
- Lee, P., McQueen, J., Stajner, I., Huang, J., Pan, L., Tong, D., et al. (2017). NAQFC developmental forecast guidance for fine particulate matter (PM2.5). *Weather and Forecasting*, 32(1), 343–360. <https://doi.org/10.1175/waf-d-15-0163.1>
- Lee, S.-J., Serre, M. L., van Donkelaar, A., Martin, R. V., Burnett, R. T., & Jerrett, M. (2012). Comparison of geostatistical interpolation and remote sensing techniques for estimating long-term exposure to ambient PM2.5 concentrations across the continental United States. *Environmental Health Perspectives*, 120(12), 1727.
- Li, M., Zhang, Q., Kurokawa, J. I., Woo, J. H., He, K., Lu, Z., et al. (2017). MIX: A mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP. *Atmospheric Chemistry and Physics*, 17(2), 935–963. <https://doi.org/10.5194/acp-17-935-2017>
- Liao, H., Henze, D. K., Seinfeld, J. H., Wu, S., & Mickley, L. J. (2007). Biogenic secondary organic aerosol over the United States: Comparison of climatological simulations with observations. *Journal of Geophysical Research - Atmospheres*, 112, D06201. <https://doi.org/10.1029/2006jd007813>
- Liu, H., Jacob, D. J., Bey, I., & Yantosca, R. M. (2001). Constraints from <sup>210</sup>Pb and <sup>7</sup>Be on wet deposition and transport in a global three-dimensional chemical tracer model driven by assimilated meteorological fields. *Journal of Geophysical Research - Atmospheres*, 106(D11), 12,109–12,128. <https://doi.org/10.1029/2000jd900839>
- Liu, Y., Park, R. J., Jacob, D. J., Li, Q., Kilaru, V., & Sarnat, J. A. (2004). Mapping annual mean ground-level PM2.5 concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States. *Journal of Geophysical Research - Atmospheres*, 109, D22206. <https://doi.org/10.1029/2004jd005025>
- Long, M., Yantosca, R., Nielsen, J., Keller, C., Da Silva, A., Sulprizio, M., et al. (2015). Development of a grid-independent GEOS-Chem chemical transport model (v9-02) as an atmospheric chemistry module for Earth system models. *Geoscientific Model Development*, 8(3), 595–602. <https://doi.org/10.5194/gmd-8-595-2015>
- Ma, C., Wang, T., Zang, Z., & Li, Z. (2018). Comparisons of three-dimensional Variational data assimilation and model output statistics in improving atmospheric chemistry forecasts. *Advances in Atmospheric Sciences*, 35(7), 813–825. <https://doi.org/10.1007/s00376-017-7179-y>
- Martin, R. V. (2008). Satellite remote sensing of surface air quality. *Atmospheric Environment*, 42(34), 7823–7843. <https://doi.org/10.1016/j.atmosenv.2008.07.018>
- McKeen, S., Chung, S., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., et al. (2007). Evaluation of several PM2.5 forecast models using data collected during the ICARTT/NEAQS 2004 field study. *Journal of Geophysical Research - Atmospheres*, 112, D10S20. <https://doi.org/10.1029/2006jd007608>
- McKeen, S., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Hsie, E. Y., et al. (2005). Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004. *Journal of Geophysical Research - Atmospheres*, 110(D21). <https://doi.org/10.1029/2005jd005858>
- Misenis, C., & Zhang, Y. (2010). An examination of sensitivity of WRF/Chem predictions to physical parameterizations, horizontal grid spacing, and nesting options. *Atmospheric Research*, 97(3), 315–334. <https://doi.org/10.1016/j.atmosres.2010.04.005>
- Otte, T., & Pleim, J. (2010). The Meteorology–Chemistry Interface Processor (MCIP) for the CMAQ modeling system: Updates through MCIPv3. 4.1. *Geoscientific Model Development*, 3(1), 243–256. <https://doi.org/10.5194/gmd-3-243-2010>
- Pagowski, M., Grell, G., McKeen, S., Dévényi, D., Wilczak, J., Bouchet, V., et al. (2005). A simple method to improve ensemble-based ozone forecasts. *Geophysical Research Letters*, 32(7). <https://doi.org/10.1029/2004gl022305>
- Park, R. J., Jacob, D. J., Chin, M., & Martin, R. V. (2003). Sources of carbonaceous aerosols over the United States and implications for natural visibility. *Journal of Geophysical Research - Atmospheres*, 108. <https://doi.org/10.1029/2002jd003190>
- Park, R. J., Jacob, D. J., Field, B. D., Yantosca, R. M., & Chin, M. (2004). Natural and transboundary pollution influences on sulfate–nitrate–ammonium aerosols in the United States: Implications for policy. *Journal of Geophysical Research - Atmospheres*, 109, D15204. <https://doi.org/10.1029/2003jd004473>
- Peng, R. D., Dominici, F., Pastor-Barriuso, R., Zeger, S. L., & Samet, J. M. (2005). Seasonal analyses of air pollution and mortality in 100 US cities. *American Journal of Epidemiology*, 161(6), 585–594. <https://doi.org/10.1093/aje/kwi075>
- Pope, C. A., Ezzati, M., & Dockery, D. W. (2009). Fine-particulate air pollution and life expectancy in the United States. *New England Journal of Medicine*, 360(4), 376–386. <https://doi.org/10.1056/NEJMsa0805646>
- Pye, H., Liao, H., Wu, S., Mickley, L. J., Jacob, D. J., Henze, D. K., & Seinfeld, J. (2009). Effect of changes in climate and emissions on future sulfate–nitrate–ammonium aerosol levels in the United States. *Journal of Geophysical Research - Atmospheres*, 114, D01205. <https://doi.org/10.1029/2008jd010701>
- Reid, J. S., Hyer, E. J., Prins, E. M., Westphal, D. L., Zhang, J., Wang, J., et al. (2009). Global monitoring and forecasting of biomass-burning smoke: Description of and lessons from the Fire Locating and Modeling of Burning Emissions (FLAMBE) program. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2(3), 144–162. <https://doi.org/10.1109/jstars.2009.2027443>
- Schell, B., Ackermann, I. J., Hass, H., Binkowski, F. S., & Ebel, A. (2001). Modeling the formation of secondary organic aerosol within a comprehensive air quality model system. *Journal of Geophysical Research - Atmospheres*, 106(D22), 28,275–28,293. <https://doi.org/10.1029/2001jd000384>
- Schwartz, C. S., Liu, Z., Lin, H. C., & McKeen, S. A. (2012). Simultaneous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth. *Journal of Geophysical Research - Atmospheres*, 117, D13202. <https://doi.org/10.1029/2011jd017383>

- Simon, H., Baker, K. R., & Phillips, S. (2012). Compilation and interpretation of photochemical model performance statistics published between 2006 and 2012. *Atmospheric Environment*, *61*, 124–139. <https://doi.org/10.1016/j.atmosenv.2012.07.012>
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., et al. (2008). *A description of the advanced research WRF version 3 (No. NCAR/TN-475+STR)*. University Corporation for Atmospheric Research.
- Solazzo, E., Bianconi, R., Pirovano, G., Matthias, V., Vautard, R., Moran, M. D., et al. (2012). Operational model evaluation for particulate matter in Europe and North America in the context of AQMEII. *Atmospheric Environment*, *53*, 75–92. <https://doi.org/10.1016/j.atmosenv.2012.02.045>
- Stockwell, W. R., Middleton, P., Chang, J. S., & Tang, X. (1990). The second generation regional acid deposition model chemical mechanism for regional air quality modeling. *Journal of Geophysical Research - Atmospheres*, *95*(D10), 16,343–16,367. <https://doi.org/10.1029/jd095id10p16343>
- Szykman, J., Kondragunta, S., Zhang, H., Dickerson, P., van Donkelaar, A., Martin, R., et al. (2012). *Expanding the estimation of surface PM<sub>2.5</sub> from Aqua and Terra MODIS Aerosol Optical Depth in the EPA's AirNow Satellite Data Processor to Suomi NPP VIIRS*. San Francisco, California: AGU Fall Meeting Abstracts.
- Szykman, J., White, J., Pierce, B., Al-Saadi, J., Neil, D., Kittaka, C., et al. (2004). *Utilizing MODIS satellite observations in near-real-time to improve AIRNow next day forecast of fine particulate matter, PM<sub>2.5</sub>*. Seattle, WA: Sixth Conference on Atmospheric Chemistry: Air Quality in Megacities, American Meteorological Society. January
- Tang, X., Zhu, J., Wang, Z., & Gbaguidi, A. (2011). Improvement of ozone forecast over Beijing based on ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions. *Atmospheric Chemistry and Physics*, *11*(24), 12,901–12,916. <https://doi.org/10.5194/acp-11-12901-2011>
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research - Atmospheres*, *106*(D7), 7183–7192. <https://doi.org/10.1002/jgrd.v106.d7>
- Tong, D. Q., Dan, M., Wang, T., & Lee, P. (2012). Long-term dust climatology in the western United States reconstructed from routine aerosol ground monitoring. *Atmospheric Chemistry and Physics*, *12*(11), 5189–5205. <https://doi.org/10.5194/acp-12-5189-2012>
- Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., et al. (2016). Why do models overestimate surface ozone in the Southeast United States? *Atmospheric Chemistry and Physics*, *16*(21), 13,561–13,577. <https://doi.org/10.5194/acp-16-13561-2016>
- Van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environmental Health Perspectives*, *118*(6), 847–855. <https://doi.org/10.1289/ehp.0901623>
- Van Donkelaar, A., Martin, R. V., & Park, R. J. (2006). Estimating ground-level PM<sub>2.5</sub> using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research - Atmospheres*, *111*, D21201. <https://doi.org/10.1029/2005jd006996>
- Wang, J., Aegerter, C., Xu, X., & Szykman, J. J. (2016). Potential application of VIIRS Day/Night Band for monitoring nighttime surface PM<sub>2.5</sub> air quality from space. *Atmospheric Environment*, *124*, 55–63. <https://doi.org/10.1016/j.atmosenv.2015.11.013>
- Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM<sub>2.5</sub> mass: Implications for air quality studies. *Geophysical Research Letters*, *30*(21). <https://doi.org/10.1029/2003gl018174>
- Wang, J., Ge, C., Yang, Z., Hyer, E. J., Reid, J. S., Chew, B.-N., et al. (2013). Mesoscale modeling of smoke transport over the Southeast Asian Maritime Continent: Interplay of sea breeze, trade wind, typhoon, and topography. *Atmospheric Research*, *122*, 486–503. <https://doi.org/10.1016/j.atmosres.2012.05.009>
- Wang, J., Nair, U., & Christopher, S. A. (2004). GOES 8 aerosol optical thickness assimilation in a mesoscale model: Online integration of aerosol radiative effects. *Journal of Geophysical Research - Atmospheres*, *109*, D23203. <https://doi.org/10.1029/2004jd004827>
- Wang, J., Xu, X., Ding, S., Zeng, J., Spurr, R., Liu, X., et al. (2014). A numerical testbed for remote sensing of aerosols, and its demonstration for evaluating retrieval synergy from a geostationary satellite constellation of GEO-CAPE and GOES-R. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *146*, 510–528. <https://doi.org/10.1016/j.jqsrt.2014.03.020>
- Wang, J., Xu, X., Henze, D. K., Zeng, J., Ji, Q., Tsay, S. C., & Huang, J. (2012). Top-down estimate of dust emissions through integration of MODIS and MISR aerosol retrievals with the GEOS-Chem adjoint model. *Geophysical Research Letters*, *39*(8). <https://doi.org/10.1029/2012gl051136>
- Wang, Y., Wang, J., Xu, X., Henze, D. K., Wang, Y., & Qu, Z. (2016). A new approach for monthly updates of anthropogenic sulfur dioxide emissions from space: Application to China and implications for air quality forecasts. *Geophysical Research Letters*, *43*(18), 9931–9938. <https://doi.org/10.1002/2016gl070204>
- Wang, Y. X., McElroy, M. B., Jacob, D. J., & Yantosca, R. M. (2004). A nested grid formulation for chemical transport over Asia: Applications to CO. *Journal of Geophysical Research - Atmospheres*, *109*, D22307. <https://doi.org/10.1029/2004jd005237>
- Whitten, G. Z., Heo, G., Kimura, Y., McDonald-Buller, E., Allen, D. T., Carter, W. P., & Yarwood, G. (2010). A new condensed toluene mechanism for Carbon Bond: CB05-TU. *Atmospheric Environment*, *44*(40), 5346–5355. <https://doi.org/10.1016/j.atmosenv.2009.12.029>
- Wiedinmyer, C., Akagi, S., Yokelson, R. J., Emmons, L., Al-Saadi, J., Orlando, J., & Soja, A. (2011). The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from open burning. *Geoscientific Model Development*, *4*(3), 625. <https://doi.org/10.5194/gmd-4-625-2011>
- Wilczak, J., McKeen, S., Djalalova, I., Grell, G., Peckham, S., Gong, W., et al. (2006). Bias-corrected ensemble and probabilistic forecasts of surface ozone over eastern North America during the summer of 2004. *Journal of Geophysical Research - Atmospheres*, *111*, D23S28. <https://doi.org/10.1029/2006jd007598>
- Wu, L., Su H., Kalashnikova O. V., Jiang J. H., Zhao C., Garay M. J., Campbell J. R., et al. (2017). WRF-Chem simulation of aerosol seasonal variability in the San Joaquin Valley. *Atmospheric Chemistry and Physics*, *17*(12), 7291–7309. <https://doi.org/10.5194/acp-17-7291-2017>
- Xu, X., Wang, J., Henze, D. K., Qu, W., & Kopacz, M. (2013). Constraints on aerosol sources using GEOS-Chem adjoint and MODIS radiances, and evaluation with multisensor (OMI, MISR) data. *Journal of Geophysical Research - Atmospheres*, *118*, 6396–6413. <https://doi.org/10.1002/jgrd.50515>
- Yarwood, G., Rao, S., Yocke, M., & Whitten, G. (2005). *Updates to the carbon bond chemical mechanism: CB05*. (p. 8). Final report to the US EPA. RT-0400675. <http://www.camx.com/publ/pdfs/CB05FinalReport120805.pdf>
- Yu, S., Mathur, R., Schere, K., Kang, D., Pleim, J., Young, J., et al. (2008). Evaluation of real-time PM<sub>2.5</sub> forecasts and process analysis for PM<sub>2.5</sub> formation over the eastern United States using the Eta-CMAQ forecast model during the 2004 ICARTT study. *Journal of Geophysical Research - Atmospheres*, *113*, D06204. <https://doi.org/10.1029/2007JD009226>
- Zhang, H., Wang, J., Castro Garcia, L., Zeng, J., Dennhardt, C., Liu, Y., & Krotkov, N. A. (2019). Surface erythral UV irradiance in the continental United States derived from ground-based and OMI observations: Quality assessment, trend analysis and sampling issues. *Atmospheric Chemistry and Physics*, *19*(4), 2165–2181. <https://doi.org/10.5194/acp-19-2165-2019>

- Zhang, L., Constantinescu, E., Sandu, A., Tang, Y., Chai, T., Carmichael, G., et al. (2008). An adjoint sensitivity analysis and 4D-Var data assimilation study of Texas air quality. *Atmospheric Environment*, *42*(23), 5787–5804. <https://doi.org/10.1016/j.atmosenv.2008.03.048>
- Zhang, L., Gong, S., Padro, J., & Barrie, L. (2001). A size-segregated particle dry deposition scheme for an atmospheric aerosol module. *Atmospheric Environment*, *35*(3), 549–560. [https://doi.org/10.1016/s1352-2310\(00\)00326-5](https://doi.org/10.1016/s1352-2310(00)00326-5)
- Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., & Baklanov, A. (2012a). Real-time air quality forecasting, part I: History, techniques, and current status. *Atmospheric Environment*, *60*, 632–655. <https://doi.org/10.1016/j.atmosenv.2012.06.031>
- Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., & Baklanov, A. (2012b). Real-time air quality forecasting, Part II: State of the science, current research needs, and future prospects. *Atmospheric Environment*, *60*, 656–676. <https://doi.org/10.1016/j.atmosenv.2012.02.041>