JGR Atmospheres

RESEARCH ARTICLE
10.1029/2021JD035563

Improving Surface $PM_{2.5}$ Forecasts in the United States Using an Ensemble of Chemical Transport Model Outputs: 2. Bias Correction With Satellite Data for Rural Areas

Huanxin Zhang[1,2], Jun Wang[1,2], Lorena Castro García[1,2], Meng Zhou[2,3], Cui Ge[1,2], Todd Plessel[4], James Szykman[5], Robert C. Levy[6], Benjamin Murphy[5], and Tanya L. Spero[5]

[1]Department of Chemical and Biochemical Engineering, University of Iowa, Iowa City, IA, USA, [2]Center for Global and Regional Environmental Research, University of Iowa, Iowa City, IA, USA, [3]Interdisciplinary Graduate Program in Geo-Informatics, University of Iowa, Iowa City, IA, USA, [4]General Dynamics Information Technology, RTP, NC, USA, [5]U.S. Environmental Protection Agency, RTP, NC, USA, [6]NASA Goddard Space Flight Center, Greenbelt, MD, USA

Abstract This work serves as the second of a two-part study to improve surface $PM_{2.5}$ forecasts in the continental U.S. through the integrated use of multisatellite aerosol optical depth (AOD) products (MODIS Terra/Aqua and VIIRS DT/DB), multichemical transport model (CTM) (GEOS-Chem, WRF-Chem, and CMAQ) outputs, and ground observations. In Part I of the study, an ensemble Kalman filter (KF) technique using three CTM outputs and ground observations was developed to correct forecast bias and generate a single best forecast of $PM_{2.5}$ for next day over nonrural areas that have surface $PM_{2.5}$ measurements in the proximity of 125 km. Here, with AOD data, we extended the bias correction into rural areas where the closest air quality monitoring station is at least 125–300 km away. First, we ensembled all of satellite AOD products to yield the single best AOD. Second, we corrected daily $PM_{2.5}$ in rural areas from multiple models through the AOD spatial pattern between these areas and nonrural areas, referred to as "extended ground truth" or EGT, for the present day. Lastly, we applied the KF technique to reduce the forecast bias for next day using the EGT. Our results find that the ensemble of bias-corrected daily $PM_{2.5}$ from three CTMs for both today and next day show the best performance. Together, the two-part study develops a multimodel and multi-AOD bias-correction technique that has the potential to improve $PM_{2.5}$ forecasts in both rural and nonrural areas in near real time, and be readily implemented at state levels.

Plain Language Summary The U.S. Environmental Protection Agency's AirNow program reports current or forecasted air quality to the general public in the form of Air Quality Index (AQI). The forecasted AQI is made available by local and state air quality agencies across more than 500 cities across the U.S. However, since surface observations of particulate matter (PM) are primarily located in the urban areas, observation-based AQI in the rural areas is limited, and either the current or the forecasted AQI from AirNow has large uncertainties that are difficult to assess, especially during the fire season. Satellite observation with large spatial coverage provides a promising opportunity to fill in the gaps in areas where observations are spare. Building upon our previous work, here we develop a statistical technique to improve surface PM forecasts in the rural areas of continental U.S. through the use of satellite observations of aerosols, surface observations, and air quality forecasting models. Assessment with the data from Interagency Monitoring of Protected Visual Environments (IMPROVE) network shows the promise of our technique. The technique is designed with the consideration of the forecast in near real time, and is efficient with minimal requirement of computing.

1. Introduction

The U.S. Environmental Protection Agency (EPA) is required to set National Ambient Air Quality Standards (NAAQS) under the Clean Air Act designed to protect the public health and welfare from pollutants from numerous and diverse sources. The distribution of surface air quality monitoring implemented to determine pollutant levels and compliance with the NAAQS tend to be focused on the urban environments due to the legacy of high pollution in the areas when the standards and networks were first established. In 1997, EPA promulgated significant changes to the particulate matter (PM) NAAQS, establishing a new standard for $PM_{2.5}$ (fine particulate matter with aerodynamic diameter less than 2.5 μm) along with the establishment of multiple nationwide $PM_{2.5}$ monitoring networks. First, the vast majority of routine air quality monitoring stations known as State and Local

This article is companion to Huanxin Zhang et al. (2020), https://doi.org/10.1029/2019JD032293.

Key Points:

- Ensemble-based AOD from different sensors and algorithms shows better performance than individual ensemble member
- A multimodel and multi-AOD ensemble bias correction via Kalman filter improves $PM_{2.5}$ forecasts in rural areas lack of ground observations
- An ensemble framework for producing the single best $PM_{2.5}$ forecast for next day in near real time via bias correction is established

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

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

Citation:

Zhang, H., Wang, J., García, L. C., Zhou, M., Ge, C., Plessel, T., et al. (2022). Improving surface $PM_{2.5}$ forecasts in the United States using an ensemble of chemical transport model outputs: 2. Bias correction with satellite data for rural areas. *Journal of Geophysical Research: Atmospheres*, *127*, e2021JD035563. https://doi.org/10.1029/2021JD035563

Received 13 JUL 2021
Accepted 24 NOV 2021

Author Contributions:

Conceptualization: Jun Wang
Data curation: Huanxin Zhang, Lorena Castro García, Meng Zhou, Cui Ge, Todd Plessel, James Szykman, Robert C. Levy, Benjamin Murphy, Tanya L. Spero
Formal analysis: Huanxin Zhang, Lorena Castro García
Funding acquisition: Jun Wang
Investigation: Jun Wang
Methodology: Huanxin Zhang, Jun Wang
Project Administration: Jun Wang
Resources: Jun Wang

Software: Huanxin Zhang, Lorena Castro García, Meng Zhou, Cui Ge
Supervision: Jun Wang
Validation: Huanxin Zhang, Lorena Castro García
Visualization: Huanxin Zhang, Lorena Castro García
Writing – original draft: Huanxin Zhang
Writing – review & editing: Jun Wang, James Szykman, Robert C. Levy

Air Monitoring Stations (SLAMS) are owned and operated by the state and local government and Tribal agencies, and designed to track ambient air quality for the main purpose of comparison to the NAAQS. SLAMS includes Federal Reference Method and Federal Equivalent Method monitors, but also include more specialized networks such as the Chemical Speciation Network (CSN) initialized in 2000 to measure the chemical composition of particulate matter in urban areas for identifying sources and supporting the implementation of NAAQS (Jayanty et al., 2004; Solomon et al., 2014). Currently, there are about 900 operating sites to provide 24-hr averaged $PM_{2.5}$ on a daily, every third, or every sixth day basis, which is reported to the Air Quality System (AQS). In addition, approximately 600 continuous $PM_{2.5}$ mass monitors provide near real-time hourly $PM_{2.5}$, which are primarily used for reporting Air Quality Index (AQI) in support of AirNow program (Dye et al., 2004). Second, in responding to the Clean Air Act and the 1999 Regional Haze Rule, the Interagency Monitoring of Protected Visual Environments (IMPROVE) expanded to about 150 operating sites primarily in Class I federal areas (Malm et al., 1994; Solomon et al., 2014), since it began collecting data in 1988 (Eldred et al., 1990). While U.S. has the most extensive monitoring networks, only 10% of the sites, primarily from IMPROVE, are in rural areas that have ~20% of the U.S. population and 98% of the U.S. land.

Zhang et al. (2020) showed that the autocorrelation of surface $PM_{2.5}$ drops exponentially as a function of distance and is lower than 0.7 for the distance larger than 125 km. This means that in areas such as the Intermountain West and Great Plains regions (Figure 1), there is essentially no useful ground-based AQI information whatsoever. Therefore, we look to satellite data with a global spatial coverage to fill in those spatial gaps (Anderson et al., 2003). For example, the aerosol optical depth (AOD) retrieved from different satellite instruments such as the Moderate Resolution Imaging Spectrometer (MODIS) (Remer et al., 2005) and Multiangle Imaging Spectro Radiometer (MISR) (Diner et al., 1998) have been used to derive surface $PM_{2.5}$ mass concentration. The earliest studies developed a simple empirical relationship between satellite-retrieved AOD and surface measured $PM_{2.5}$ concentration (Chu et al., 2003; Wang & Christopher, 2003). Following work have included the vertical profile of aerosols either inferred from chemical transport models (Liu et al., 2004; Van Donkelaar et al., 2010) or lidar data (Engel-Cox et al., 2006), and further considered the influence from meteorology (Gupta et al., 2006; Koelemeijer et al., 2006). More recent work have developed geographically weighted regression (GWR) model (Hu et al., 2013; Lee et al., 2016; Liu et al., 2005, 2007; Sathe et al., 2019; Zheng et al., 2016) or applied machine learning methods (Gupta & Christopher, 2009; Hu et al., 2017). In addition, using the radiance data (Fu, Xia, Duan, et al., 2018; Wang et al., 2016) or top-of-atmosphere reflectance (Shen et al., 2018) has shown promising results. These aforementioned studies and so many more greatly expanded and enhanced our knowledge about the AOD-PM relationship and the application of satellite data in air quality and health studies (Al-Saadi et al., 2005; Anenberg et al., 2018; Cohen et al., 2017; Hoff & Christopher, 2009). More recently, techniques have been developed to merge multi-AOD products with surface measured AOD as ground truth for subsequent application in deriving surface $PM_{2.5}$ mass concentration (Fu, Xia, Wang, et al., 2018; Fu et al., 2020; Ma et al., 2014; Puttaswamy et al., 2014). Most recently, the GWR model is used to estimate near real-time surface $PM_{2.5}$ from Visible Infrared Imaging Radiometer Suite (VIIRS) and Geostationary Operational Environmental Satellites (GOES) Advanced Baseline Imager (ABI) AOD in a daily or hourly manner (Zhang & Kondragunta, 2021).

Attempts to use satellite AOD to improve aerosol forecasting in real time have been limited except for the global prediction models such as from the International Cooperative for Aerosol Prediction (ICAP) study, in which AOD is assimilated to improve aerosol transport forecasts in real time, but the fidelity of these models for surface $PM_{2.5}$ forecasts remains elusive (Buchard et al., 2017; Xian et al., 2019). Model Output Statistics (MOS) or diagnostic approach without involving data assimilation techniques can be effective in computational cost and time for improving surface $PM_{2.5}$ forecasts. The efforts in this regard are few, however; one particular application, that is, the Infusing Satellite Data into Environmental Application (IDEA) has developed a prototype that demonstrates the operational use of satellite derived surface $PM_{2.5}$ together with forward trajectory analysis to qualitatively predict surface $PM_{2.5}$ at synoptic scales in support of the AirNow program (Al-Saadi et al., 2005; Szykman et al., 2004, 2012). Satellite aerosol related images including the surface $PM_{2.5}$ derived from satellite AOD have been available through the NOAA National Environmental Satellite, Data, and Information Service (NESDIS).

This is the second part of a two-part study that aims to develop a bias-correction framework to improve surface $PM_{2.5}$ forecasts in the continental U.S. through the integration of chemical transport models, ground observations and satellite remote sensing. The goal of this two-part study is to develop a model output diagnostic approach that has the potential to be used by regional air quality agencies for improving the forecast of surface $PM_{2.5}$ in near

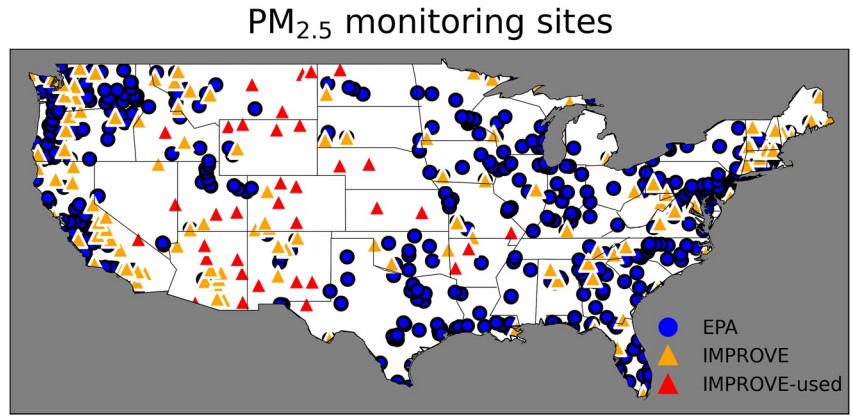

Figure 1. $PM_{2.5}$ monitoring sites in the U.S. for June 2012. Blue circles represent the 508 hourly $PM_{2.5}$ measurement sites from the U.S. Environmental Protection Agency (EPA) Air Quality System. The triangles (shown in red and orange) are the 158 daily $PM_{2.5}$ measurement sites from the Interagency Monitoring of Protected Visual Environments (IMPROVE) network. The IMPROVE sites in red represent the sites which have EPA hourly $PM_{2.5}$ site/sites within the distance range of 125–300 km, and the data at those sites are used solely for evaluating the bias-correction method for $PM_{2.5}$ forecasts in rural areas. Sites in orange triangles are only shown here for demonstrating the spatial distribution of the IMPROVE network and the data at these sites are not used in the current work because they have EPA $PM_{2.5}$ hourly measurement site/sites in the radius of 125 km.

real time; the approach is designed to be economic in computing and fast for the operational use. It tackles the challenges that local air quality agencies have the access to multiple model forecasts of surface $PM_{2.5}$, multiple satellite AOD products and their local surface $PM_{2.5}$ data, but have to make a single best estimate of surface $PM_{2.5}$ for the next day. Specifically, in the first part of the study (Zhang et al., 2020), we developed and demonstrated a multimodel ensemble approach combining the Kalman filter (KF) technique and a successive correction technique to improve model forecasts in nonrural areas. In that work, we focused on urban areas that have rich surface observation. Building upon that work, we focus on developing techniques to apply multisatellite-derived AOD data to improve forecasts of surface $PM_{2.5}$ in rural areas where no surface observation sites are available in the proximity of 125 km. This two-part series therefore provide a synergy to combine surface and satellite observations and multiple CTM outputs for improving $PM_{2.5}$ forecasts in the U.S., and the approaches developed here can be used in a computationally efficient manner by regional or local air quality offices. The paper is organized as follows: Section 2 describes the model configuration, surface $PM_{2.5}$ observations and AOD data as well as quality control methods. Section 3 introduces the bias-correction technique. We summarize the results in Section 4 and present the discussion in Section 5.

2. Model and Observation Data

Detailed model configurations are documented in the companion paper (Zhang et al., 2020) and we, in below, describe how AOD is calculated in each model as well as the various observation data used in this study.

2.1. Model Data

2.1.1. GEOS-Chem

We use the global 3-D GEOS-Chem model (Bey et al., 2001) Version v11-01, which has a nested capability to run model simulations at high spatial resolution over a specific domain. Here, we run the nested simulation with 0.5° latitude × 0.667° longitude horizontal resolution over the North American domain (10°–70°N and 40°–140°W) (Chen et al., 2009; Wang et al., 2004). Both global and nested simulation are driven by NASA Goddard Earth Observing System (GEOS-5) assimilated meteorological fields and share the same vertical distribution of 47 levels. We first run the global model simulation with 2° latitude × 2.5° longitude horizontal spatial resolution from June 1, 2011 to June 30, 2012 with the first 12 months as model spinning up time. Then, the output from the global simulation will provide time varying chemical boundary for the nested simulation, which ran from May 1, 2012

to June 30, 2012 with the first month used as model initialization time. Both the global and nested simulation have the same setup as described below.

The GEOS-Chem model simulates inorganic aerosols (sulfate-nitrate-ammonium system) (Park et al., 2004; Pye et al., 2009), carbonaceous aerosols (Liao et al., 2007; Park et al., 2003), mineral dust (Fairlie et al., 2007), and sea salt in fine and coarse mode (Jaeglé et al., 2011). Secondary organic aerosols (SOA) were disabled in the current simulation. Dust particles in GEOS-Chem contains four size bins (radius 0.1–1.0, 1.0–1.8, 1.8–3.0, and 3.0–6.0 μm) (Fairlie et al., 2007, 2010). Aerosol optical properties are calculated in GEOS-Chem with a standard Mie code for each aerosol type using a look-up table approach. In generating the look-up table, the aerosol mass concentration for each aerosol type is assumed to be in a log-normal size distribution. The aerosol hygroscopic growth as a function of the relative humidity (RH) is accounted for in the calculation of mass extinction coefficient (Martin et al., 2003). The refractive indices come from the Global Aerosol Data Set (GADS) (Koepke et al., 1997). The aerosols are assumed to be externally mixed and the sum over all aerosol types are used to compute an optical depth at different wavelengths including 550 nm.

2.1.2. WRF-Chem

For the present work, we rerun the WRF-Chem simulation to include our latest model development (Sha et al., 2021). In the new simulation, the Modern-Era Retrospective Analysis Research Application, Version 2 (MERRA-2) data provide the meteorological boundary and initial condition as well as the chemical boundary and initial condition (Buchard et al., 2017; Gelaro et al., 2017; Randles et al., 2017). The MERRA-2 data have a horizontal spatial resolution of 0.5° latitude × 0.625° longitude with 72 vertical layers. The 0.25° latitude × 0.25° longitude Global Land Data Assimilation System (GLDAS) data provide the initial conditions of soil properties (e.g., soil moisture and temperature) (Rodell et al., 2004). The land cover type data are updated with the 2018 MODIS land cover data. Dust emissions follow the Goddard Global Ozone Chemistry Aerosol Radiation and Transport (GOCART) with the Air Force Weather Agency (AFWA) modifications. The WRF-Chem model has one domain over the United States with a horizonal resolution of 12 km and 47 vertical levels. We ran the model from May 22 to June 30, 2012 with the first 10 days used as model initialization time. All other model configurations are the same as used in Zhang et al. (2020).

WRF-Chem aerosols 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 this MADE/SORGAM scheme, sulfate, nitrate, ammonium, BC, organic matters (OM), sea salt, mineral dust, and water are simulated. Sea salt aerosol emissions were not included in the current simulation. Aerosol size distribution is represented by the modal approach (Whitby et al., 1991), which contains three modes: Aitken, accumulation, and coarse mode. Each mode assumes a log-normal distribution. The calculation of aerosol optical properties followed a sectional approach with 8 bins (Barnard et al., 2010). The mass and number concentrations of each aerosol species from the three modes will be matched to the 8 bins. Each bin assumed an internal mixing and the bulk properties such as refractive index for each bin is based on volume approximation. The aerosol optical properties are computed at four wavelengths (300, 400, 600, and 1,000 nm). AOD at 550 m are interpolated from optical properties at 400 and 600 nm.

2.1.3. CMAQ

CMAQ outputs used for this study from June 1 to June 30, 2012 are directly downloaded from the Remote Sensing Information Gateway (RSIG, https://www.epa.gov/hesc/remote-sensing-information-gateway). The CMAQ Version 5.0.2 was used, and the model was run with a horizontal simulation of 12 km and 35 vertical levels over the continental U.S. The results from WRF Version 3.4 simulation and global GEOS-Chem Version v8-03-02 simulation provided the meteorological inputs and hourly chemical boundary to the CMAQ model system, respectively.

The CMAQ aerosol module AERO6 (Appel et al., 2013) simulates inorganic aerosols (sulfate, nitrate, and ammonium), secondary organic aerosols (Carlton et al., 2010), elemental carbon, organic carbon, and water. Dust emissions were not included in the simulations. Aerosol size distribution include three modes: Aitken, accumulation, and coarse (Binkowski & Roselle, 2003). AOD are calculated based on a regression relationship derived from IMPROVE monitoring data and collocated photometer readings (Pitchford et al., 2007). The dry mass extinction efficiency for each aerosol component is provided as constants. For sulfate, nitrate, and sea salt aerosols, the mass extinction efficiency also accounts for the effects of RH by multiplying a water growth factor as a function of RH.

2.2. Satellite-Retrieved AOD and Quality Control

The MODIS instrument on board Terra and Aqua have a swath of 2,330 km which enables a global coverage every 1 or 2 days. Terra and Aqua cross the equator at approximately 10:30 a.m. in the descending mode and 1:30 p.m. in the ascending mode, respectively. MODIS AOD values are retrieved from 0.47, 0.55, 0.66, and 2.13 μm over land surfaces. In the current work, we use level 2 AOD products from both Terra and Aqua, denoted as MOD04 and MYD04, respectively. The MODIS level 2 Collection 6.1 (C6.1) AOD products over the land are generated using two algorithms: "Dark Target" (DT) over vegetated surfaces or dark land surfaces (Levy et al., 2013) and "Deep Blue" (DB) over land surfaces (Hsu et al., 2013, 2019; Sayer et al., 2019). Here we consider DT and DB AOD products over land surfaces at 0.55 μm wavelength with a 10 km spatial resolution. We also use MODIS Level 3 AOD data products, which are gridded monthly products.

The VIIRS sensor onboard the Suomi-NPP satellite has a swath of ∼3,000 km, which enables a global coverage every day. The Suomi-NPP satellite has an 824 km sun-synchronous ascending orbit with the local equator-crossing time of 1:30 p.m. The VIIRS AOD retrieval includes three algorithms: DT over land (Sawyer et al., 2020), DB over land surface (Hsu et al., 2019; Sayer et al., 2019), and the environmental data record (EDR) algorithm over land and ocean (Jackson et al., 2013). Here we use the level 2 DT and DB AOD products at a spatial resolution of 6 km.

The MODIS level 2 aerosol products also include a number of auxiliary information related to the retrieval conditions. In the current study, several of them are used: (1) the quality assurance (QA) flag associated with each AOD product ranges from 1 to 3 with 3 representing the highest quality, (2) cloud fraction, and (3) scattering angle. The same auxiliary information is also used for VIIRS DT AOD but only QA flag and cloud fraction are used for VIIRS DB AOD. Following previous work (Anderson et al., 2013; Hyer et al., 2011; Shi et al., 2013; Wang et al., 2018; Zhang & Reid, 2006), different quality control filters are used in extracting satellite AOD data, which include: using QA flag of 3, removing AOD retrieved at large scattering angle (>170°) and cloud fraction (>0.8), applying buddy check for spatial smoothness and excluding pixels with large local variations. Statistics used to evaluate the datasets include the Pearson correlation coefficient (R), Mean Bias (MB), and the Root-Mean-Square Error (RMSE).

2.3. AERONET AOD

The ground-based Aerosol Robotic Network (AERONET) provides global observations of AOD, which are used to evaluate satellite and model-simulated AOD here. The sun photometers at each AERONET site measures the direct solar radiation in some or all of the following spectral band centered at 340, 380, 440, 500, 670, 940, and 1,020 nm (Holben et al., 1998). These measured radiation data are further used to retrieve AOD according to the Beer-Lambert-Bouguer law (Holben et al., 1998). To compare with satellite and model-simulated AOD at 550 nm, AERONET AOD at 440 and 670 nm were interpolated to AOD at 550 nm using the Angstrom Exponent provided by AERONET. We use Version 3 Level 2.0 AOD data from AERONET, which are cloud-screened and quality-insured. To ensure the statistical representatives at a monthly scale, we only keep those AERONET sites that have a minimum of 10 days of data in the month of June 2012. As a result, a total of 33 AERONET sites are remained (see Figure 4 for the spatial distribution).

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

Hourly surface measurement of $PM_{2.5}$ data for June 2012 are taken from U.S. EPA's Air Quality System (AQS, https://www.epa.gov/aqs). More details about the data and quality control are described in the companion paper (Zhang et al., 2020). We also use daily total $PM_{2.5}$ measurements from the IMPROVE network operated by the National Park Service and other federal agencies (Malm et al., 1994). At each IMPROVE site, four different models are used to collect samples for speciated $PM_{2.5}$, gravimetric $PM_{2.5}$, and PM_{10} analysis. Besides the gravimetric $PM_{2.5}$ mass concentration, the IMPROVE network also reconstructed $PM_{2.5}$ mass concentration using the speciated $PM_{2.5}$ concentration (Hand et al., 2012; Malm et al., 2011; Malm & Hand, 2007). Both measured and reconstructed $PM_{2.5}$ have biases, which has been investigated by Malm et al. (2011). It was found that the overall difference between gravimetrically determined and true ambient $PM_{2.5}$ (estimates of true concentration) is about 1 μg m^{-3} (6%) and this difference for reconstructed and true ambient $PM_{2.5}$ mass sets about 0.2 μg m^{-3} (3.5%). In

the current work, we use the reconstructed $PM_{2.5}$ concentration as the work of Hand et al. (2019) has suggested that gravimetric $PM_{2.5}$ measurement biases increased due to the presence of particle-bound water after 2011.

3. Methodology

In Part I of this work (Zhang et al., 2020), a two-step approach is developed to improve the forecast of surface $PM_{2.5}$ in nonrural areas where surface observations are rich. For model grids collocated with EPA hourly surface measurements, the ensemble Kalman filter (KF) technique is used to correct model forecast bias, and subsequently generate the improved forecasts of surface $PM_{2.5}$ for next day. The ensemble consists of surface $PM_{2.5}$ outputs from GEOS-Chem, WRF-Chem, and CMAQ forecasts (and practically in a hindcast for the purpose of research). Then, a Successive Correction Method (SCM) was applied to spread model correction at the surface observation sites to nearby model grids in the radius of 125 km, that do not have ground observations. This radius of influence of 125 km is based on analysis of EPA surface measurements of hourly $PM_{2.5}$ mass concentration.

In Part II of this work, here we develop a three-step approach of using satellite AOD to improve $PM_{2.5}$ forecasts over the model grids that have no surface observation sites in their proximity of 125 km. These model grids often cover the rural areas beyond the distance of 125 km away from the urban. First, we ensemble satellite AOD to obtain the single best AOD at ground observation sites (as to be detailed in Section 3.1) and rural areas that are 125 km or more away from these sites. Then, we correct model-simulated surface $PM_{2.5}$ for present day (today) over these rural areas through AOD spatial pattern, thereby linking the correction needed in rural areas with the benefit coming from surface observations in nonrural areas. Once these satellite-enabled today's "ground truth" for rural areas are created, we can then use the KF technique (as already developed in Part I of this work, denoted as KF-ori) to update model forecasted bias and subsequently correct forecasted $PM_{2.5}$ for next day for the rural areas.

3.1. AOD Ensemble

Both MODIS and VIIRS AOD products have been evaluated against AERONET AOD in previous work (e.g., Sayer et al., 2019; Wei et al., 2019). Here, we mainly focus on the AOD ensemble evaluations. Since large spread exists between AOD products from different sensors and algorithms, AOD ensemble could help reduce random errors and increase data coverage. Previous work have used AOD ensemble average to study fire emissions (Wang et al., 2018), or fused multisatellite AOD products to estimate surface $PM_{2.5}$ mass concentration (Ma et al., 2014). Wang et al. (2018) evaluated the monthly average of Terra and Aqua DT and DB AOD by sensor or algorithm against AERONET AOD in northern sub-Saharan African region. They found that the average of Terra and Aqua DB AOD has one of the highest correlations and lowest RMSE. In Ma et al. (2014), MODIS and MISR AOD were fused to improve the data spatial coverage. More recently, Sogacheva et al. (2020) used 12 major available satellite AOD products to create a merged AOD product over the period of 1995–2017 for climate studies. Less has been explored to deploy ensemble AOD for improving air quality forecasts.

In this work, AOD values are retrieved from three satellite sensors (Terra, Aqua, and VIIRS) and two AOD retrieving algorithms (DT and DB). Accordingly, we construct ensemble AOD by sensor, algorithm or both: DT (DB)-Ens (ensemble based on satellite-retrieval algorithm, e.g., by ensembling MODIS Terra, Aqua, and VIIRS DT or DB AOD), Terra (Aqua/VIIRS)-Ens (ensemble based on sensors), and All-Ens (ensemble of all AODs). Here, we assume that ensemble created with AOD overpass time at 10:30 a.m. or 1:30 p.m. would be equally representative of the daily mean. Also, by doing the ensemble with all three sensors, the overpass time of 1:30 p.m. is weighted twice compared to the overpass time of 10:30 a.m. We acknowledge these limitations in our current work but using geostationary satellite AOD data with finer temporal resolution could help resolve these limitations in future work.

The spatial-temporal collocation of satellite and AERONET AOD follows the approach of Ichoku et al. (2002). Satellite AOD averaged within a radius of 25 km centered around an AERONET site is evaluated against the corresponding AERONET AOD measurements averaged within ±30 min of MODIS (Terra/Aqua) or VIIRS overpass time. Also, due to the surface land type difference in the eastern and western U.S., we further divide the AOD datasets into eastern U.S. (<100°W) and western U.S. (>100°W). The comparison of mean AOD in each of these ensembles with AERONET AOD is shown for eastern U.S. in Figure 2 (and for western U.S. in Figure S1). Here, ensemble mean is defined as the average (with equal weight) of AOD values from each ensemble member.

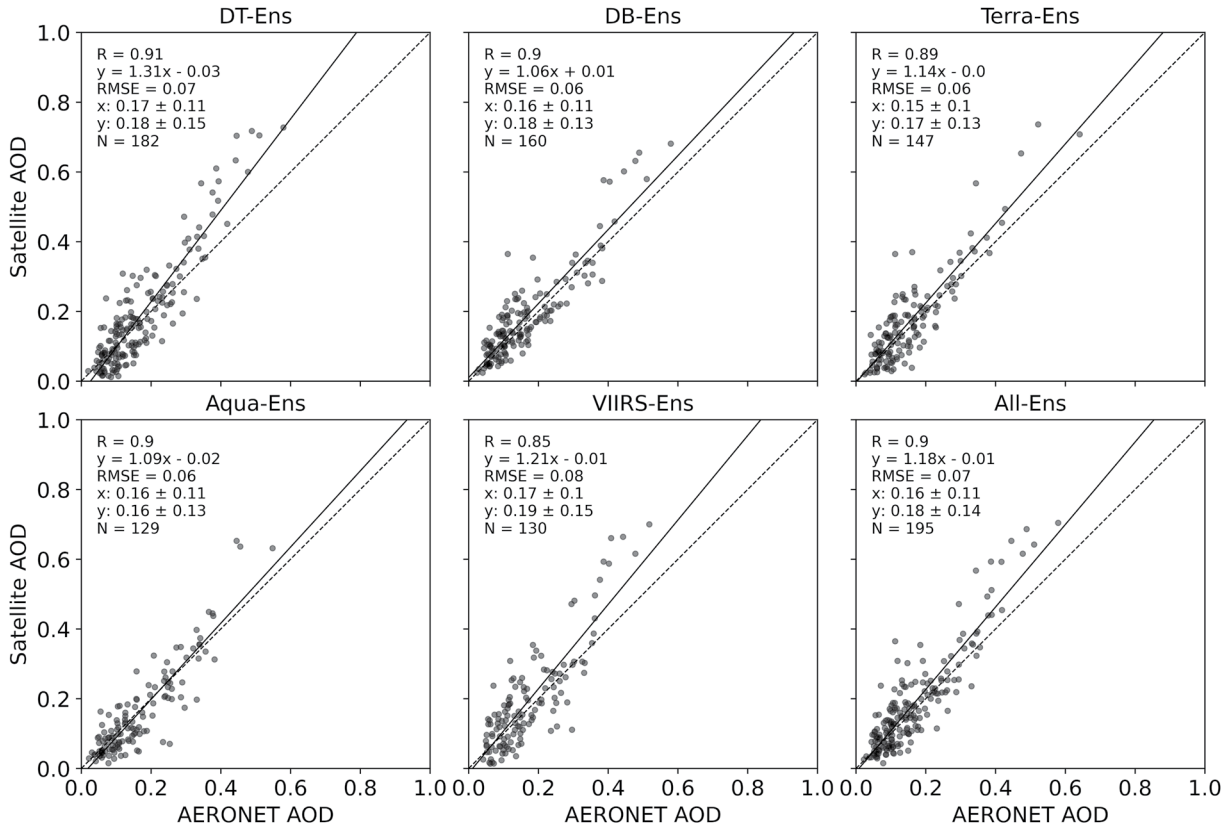

Figure 2. Scatter plot of equally weighted ensemble satellite aerosol optical depth (AOD) and AERONET AOD for sites located in the eastern U.S. for June 2012. DT-Ens and DB-Ens represent ensembling AOD based on the different algorithm. For example, DT-Ens will average the AOD from Terra, Aqua, and VIIRS DT AOD products. Terra-Ens, Aqua-Ens, and VIIRS-Ens ensembles AOD by sensors. All-Ens are the average of all of the AODs. Please see Section 3.1 for more details. Also shown on the scatter plot is the correlation coefficient (R), the root-mean-square error (RMSE), the mean ± standard deviation for AERONET AOD (x) and satellite AOD (y), the number of collocated data points (N), the best fit linear regression (the solid black line) and the 1:1 line (the dashed black line).

For the eastern U.S., different ensemble means show similar R, RMSE, and MB values. For the western U.S., the different ensembles lead to a wide range of correlation and RMSE. The DT-Ens and VIIRS-Ens result in higher correlation of 0.76 and 0.79 and larger RMSE of 0.13 and 0.14 while DB-Ens, Terra-Ens, and Aqua-Ens show lower correlation of 0.45, 0.53, and 0.59 and smaller RMSE of 0.05, 007, and 0.07. In contrast, the All-Ens show the best performance overall with R of 0.75, RMSE of 0.08 and MB of 0.04. Therefore, we will use the equally weighted ensemble of all available AODs for both eastern and western U.S. in this work unless noted otherwise.

3.2. $PM_{2.5}$ Forecast Bias Correction Through AOD

Currently, the U.S. EPA AirNow program (https://www.airnow.gov/) offers the NowCast and forecasted Air Quality Index (AQI) for ozone and particle pollution. The NowCast AQI is calculated based on multiple hours of past to current available air pollutant concentrations from real-time hourly surface measurements. These measurements usually have one hour latency. Tomorrow's forecasts, issued by state and local air quality forecasters across the country, are made available in most areas by 4 p.m. local time. MODIS Terra and Aqua real-time AOD data are available 60–125 min after their respective overpass time. For VIIRS AOD data, they are available 180 min after its overpass time. Based on these time references, we aim on providing corrected forecasts of surface $PM_{2.5}$ for next day no later than 5 p.m. local time in our bias-correction technique.

We divide the model grid boxes into three types based on their distance with ground observation sites. Type-1 refers to those having collocated hourly EPA ground observation sites on the model grids; type-2 refers to those without any collocated ground observation sites on the model grids but are surrounded by EPA hourly ground sites in the radius of 125 km; type-3 refer to the rest of model grid boxes that have no surface observations within the proximity of 125 km but has N EPA hourly ground sites in the radius of 125–300 km. For type-1 model grid

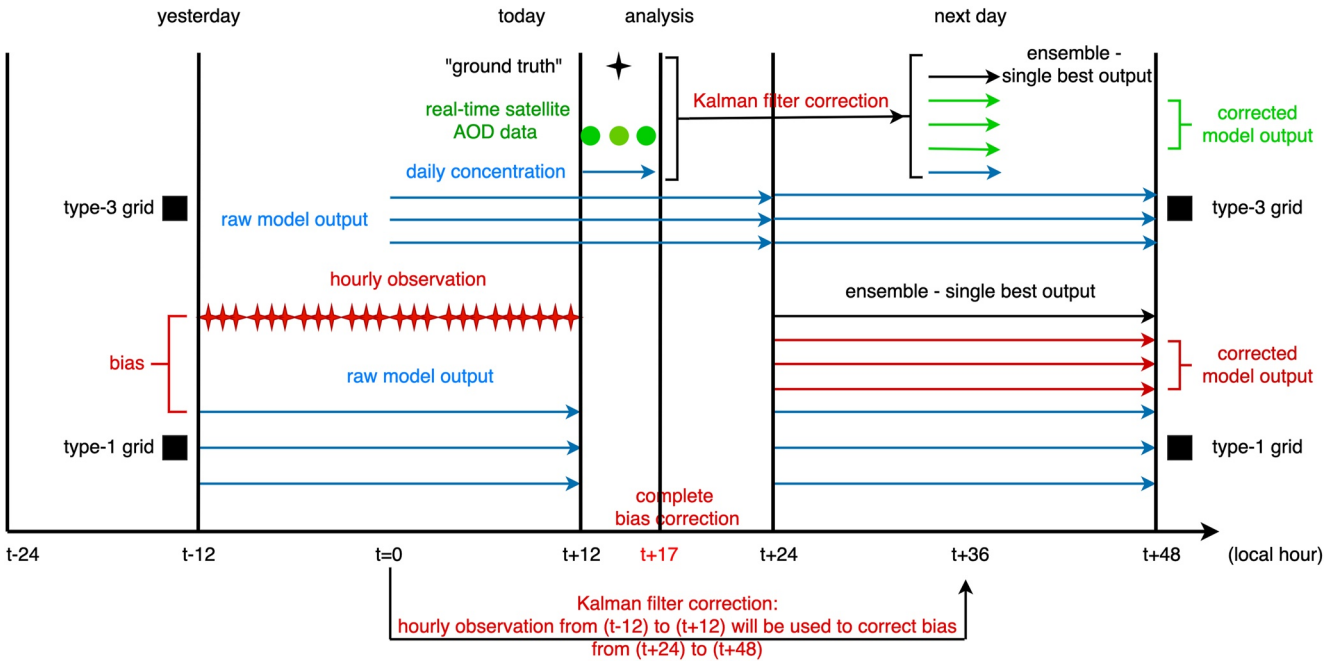

Figure 3. Scheduling chart of the bias-correction technique. Type-1 grid refers to the model grid that is collocated with Environmental Protection Agency (EPA) hourly ground observation site and type-3 grid refers to the model grid that has no collocated EPA hourly ground observation site but has EPA ground observation site/sites in the radius of 125–300 km. For type-1 model grids, the Kalman filter (KF) technique will be used to correct model forecasts of hourly $PM_{2.5}$ concentration for next day (midnight to midnight), based on hourly observation from noon yesterday to noon today. For type-3 model grids, satellite aerosol optical depth (AOD) data will be used to improve model forecasted $PM_{2.5}$ for next day. These postprocessing bias correction will be completed by 5 p.m. local time considering the latency of EPA hourly observation and satellite AOD data. More details are available in Section 3.2.

boxes, Kalman filter (KF) technique will be used to correct model forecast biases. Then, the SCM will be used to spread the bias-correction information from type-1 model grid boxes to type-2 model grid boxes. These two steps have been documented in Part I of this work (Zhang et al., 2020) and we made some modifications for the KF technique here to take use of the surface hourly measurements for real-time forecasts. After correcting model forecasted bias for type-1 model grids and subsequently spreading the corrected bias in the radius of 125 km from type-1 grids, forecasts of surface $PM_{2.5}$ in many nonrural areas will be corrected and model grids left uncorrected are mostly located in rural areas as shown in Figure S2a. In the last step, satellite AOD will be used to correct forecast biases for type-3 model grids, located mostly in rural areas. Altogether, forecasts of surface $PM_{2.5}$ for the majority of continental United States will be corrected (Figure S2b).

Figure 3 shows the schedule chart of the bias-correction techniques implemented here. First, for type-1 model grids, the KF will be applied using the hourly observation and model-simulated $PM_{2.5}$ data over a period of 24 hr (from noon yesterday to noon today, local time). This bias information will be used to correct model forecasted $PM_{2.5}$ for the next day of 24 hr (from midnight to midnight, local time). It would be ideal to use the information from the same hour for today and next day, but in reality, this is not possible after consideration of data latency and time needed for data processing and analysis. Therefore, 5 hr (from today's noon to today's 5 p.m.) are reserved to conduct procedures for bias-correction technique and issue the air quality forecasts for next day. We shift the hours here to take advantage of the available surface observations we can obtain before we start the postprocessing bias-correction process.

We correct model-simulated surface $PM_{2.5}$ on any type-3 model grid for today with a day number index d, by (a) first computing the correction ratio (C_i) between observation-based and model-based AOD/$PM_{2.5}$ scaling factor (Equation 1) at its nearby type-1 model grid, and (b) subsequently, by transferring this ratio to the type-3 model grid box as a constraint to update surface $PM_{2.5}$ at that type-3 model grid box from the satellite and modeled AOD ratio (Equation 2):

$$C_i = \frac{AOD^{1i}_{\text{Sate}}}{AOD^{1i}_{\text{model}}} \Big/ \frac{\left(\left(PM^{\text{obs},1i}_{2.5,\ d} + PM^{\text{KF\_model},1i}_{2.5,\ d}\right)/2\right)}{PM^{\text{Raw\_model},1i}_{2.5,\ d}} \quad (1)$$

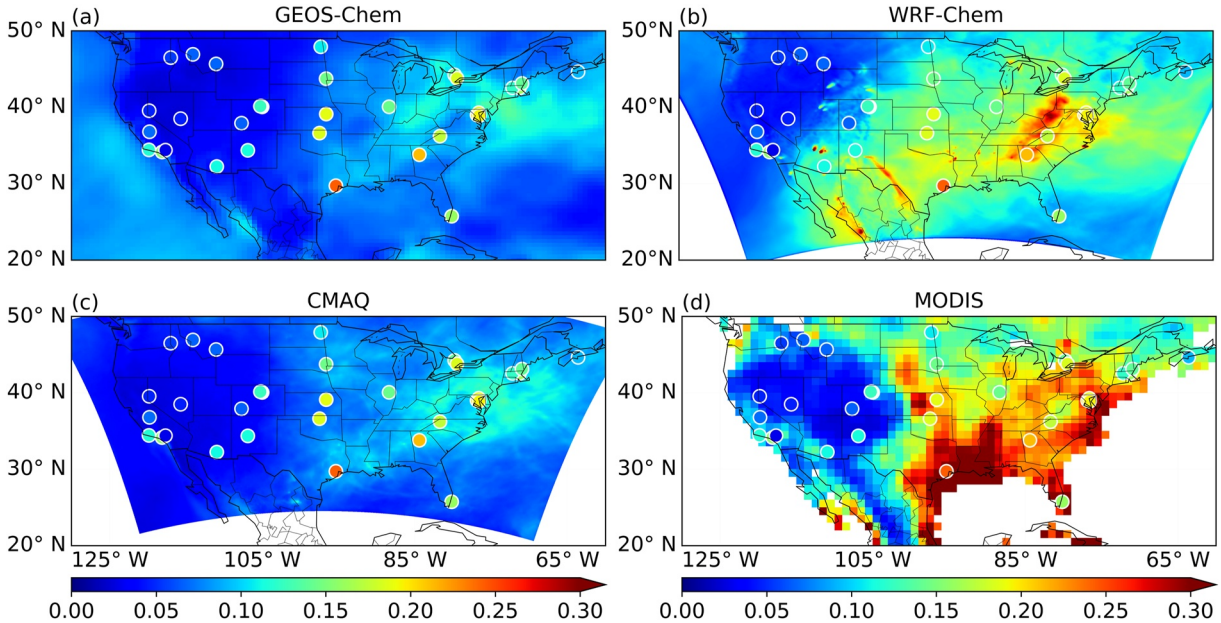

Figure 4. Monthly mean aerosol optical depth (AOD) simulated by (a) GEOS-Chem, (b) WRF-Chem, and (c) CMAQ, and retrieved by Moderate Resolution Imaging Spectrometer (MODIS) (Terra) level 3 DB AOD at 550 nm for June 2012. Overlaid on each panel is the AOD at 33 AERONET sites (denoted as filled circles).

$$PM_{2.5,\ d}^{\mathrm{KF\_model,3}} = \frac{AOD_{\mathrm{Sate}}^{3}}{AOD_{\mathrm{model}}^{3}} \sum_{i=1}^{N} \frac{\omega_i PM_{2.5,\ d}^{\mathrm{Raw\_model,3}}}{C_i} \tag{2}$$

where $i = 1, \ldots, N$, representing the index of each EPA ground site with hourly surface $PM_{2.5}$ measurement, and N is the total number of EPA hourly ground sites collocated with type-1 model grid in the radius of 125–300 km away from the type-3 model grid. AOD_{Sate}^{1i} is the satellite-based AOD for ground site i, and $AOD_{\mathrm{model}}^{1i}$ is the model-predicted AOD averaged over (satellite overpass time ±2 hr). AOD_{Sate}^{3} and AOD_{model}^{3} are same as AOD_{sate}^{1i} and $AOD_{\mathrm{model}}^{1i}$ but for the type-3 model grid. $PM_{2.5,\ d}^{\mathrm{obs},1i}$ and $PM_{2.5,\ d}^{\mathrm{Raw\_model},1i}$ are the measured and model-predicted daily $PM_{2.5}$ at type-1 model grid. $PM_{2.5,\ d}^{\mathrm{KF\_model},1i}$ is the KF bias-corrected daily $PM_{2.5}$ at type-1 model grid. $PM_{2.5,\ d}^{\mathrm{Raw\_model,3}}$ is the model-predicted daily $PM_{2.5}$ at type-3 model grid. We also apply the inverse distance weighted interpolation between type-3 model grid and the ith EPA hourly site collocated with type-1 model grid in Equation 2, represented by the weight ω_i. Here, both AOD_{Sate}^{1i} and AOD_{Sate}^{3} have been adjusted using the prognostic approach as in the work of Shi et al. (2013). All the data in Equations 1 and 2 are for today. $PM_{2.5,\ d}^{\mathrm{KF\_model,3}}$ is the final corrected daily $PM_{2.5}$ mass concentration for type-3 model grid, served as "extended ground truth" or EGT. This process will be done for each model (GEOS-Chem, WRF-Chem and CMAQ) and subsequently, the daily $PM_{2.5,\ d}^{\mathrm{KF\_model,3}}$ for today will be created for each model. Then, we will ensemble all three of them to create an ensemble averaged $PM_{2.5,\ d}^{\mathrm{KF\_model,3}}$ that can be used to update daily $PM_{2.5}$ at type-3 model grid for next day.

To improve the daily $PM_{2.5}$ forecasts for next day at type-3 model grid, we will use the KF technique developed in Part I of this work (Zhang et al., 2020). The KF technique was implemented to correct hourly surface $PM_{2.5}$ and here we adopt it to correct daily $PM_{2.5}$. In the KF technique, the true bias x_d is defined as the difference between model predicated $PM_{2.5}$ and the true (unobserved) $PM_{2.5}$ concentrations. This true bias is related to the model forecasted error y_d, defined as the difference between model forecasted $PM_{2.5}$ and surface measurements of $PM_{2.5}$ concentrations. The $PM_{2.5,\ d}^{\mathrm{KF\_model,3}}$ will be used as EGT substituting the measured surface $PM_{2.5}$, because it is a result of benefiting from the locations that have the ground truth (e.g., type-1 grid box). Then, the recursive predictor $\hat{x}_d$ will be estimated as follows:

$$\hat{x}_{d+1} = \hat{x}_d + \beta_d (y_d - \hat{x}_d) \tag{3}$$

where the hat (^) symbol represents estimates of the truth. β_d is the Kalman gain which weights the difference between today's forecast error y_d and today's bias $\hat{x}_d$ that was estimated yesterday to provide the correction that

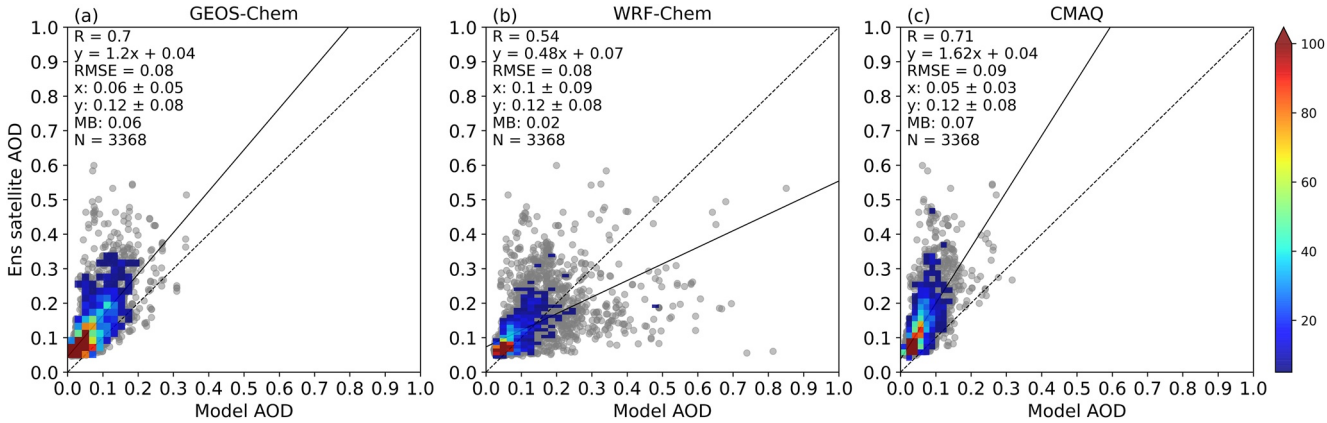

Figure 5. Scatter plot of equally weighted ensemble satellite aerosol optical depth (AOD) and the AOD simulated by three models at the EPA hourly sites for June 2012: (a) GEOS-Chem, (b) WRF-Chem, and (c) CMAQ. Also shown on the scatter plot is the correlation coefficient (R), the root-mean-square error (RMSE), the mean bias (MB), the mean ± standard deviation for model AOD (x) and satellite AOD (y), the number of collocated data points (N), the density of points (the color bar), the best fit linear regression (the solid black line) and the 1:1 line (the dashed black line).

was learned from previous errors. This will be added to today's estimate of the bias $\hat{x}_d$ to give today's estimate of tomorrow's bias $\hat{x}_{d+1}$. Once $\hat{x}_{d+1}$ is produced, we can then correct model's forecasted daily $PM_{2.5}$ for next day $\left(PM_{2.5,\ d+1}^{\text{KF\_model,3}}\right)$ as:

$$PM_{2.5,\ d+1}^{\text{KF\_model,3}} = PM_{2.5,\ d+1}^{\text{Raw\_model,3}} - \hat{x}_{d+1} \tag{4}$$

where $PM_{2.5,\ d+1}^{\text{Raw\_model,3}}$ is the model forecasted raw daily $PM_{2.5}$ for next day at the type-3 model grid.

When satellite AOD of type-1 model grid or type-3 model grid for today is not available, Equations 1 and 2 cannot be executed. We will then use the past bias statistics information to continue the KF.

4. Results

In this section, we first compare model-simulated AOD with satellite AOD and AERONET AOD. Then, we use 1 month of model hindcast outputs to evaluate the bias-correction technique to forecast surface $PM_{2.5}$ for next day, as described in Section 3.

4.1. Model-Simulated AOD and Observed AOD

Figure 4 shows the surface map of monthly mean AOD from GEOS-Chem, WRF-Chem, CMAQ, and MODIS on board Terra compared with AERONET AOD for June 2012. Higher AOD in the eastern U.S. and lower in the western U.S. are found from both AERONET and MODIS observations. GEOS-Chem and CMAQ show similar spatial patterns and magnitude but underestimate MODIS AOD. WRF-Chem simulated AOD matches the MODIS AOD better than the other two models. In the eastern U.S., high anthropogenic emissions of primary particles together with enhanced secondary particles contribute to higher aerosol loading. The current GEOS-Chem simulation did not include secondary organic aerosols (SOAs). All three models capture the higher AOD values in the Ohio River Valley region which reflects the impacts of anthropogenic aerosols. In addition, AERONET and satellite observations also show higher AOD values in the central Great Plains and southwestern U.S., which could be caused by wildfires. Higher AOD in these regions are found in WRF-Chem, largely due to the use of Fire Locating and Modeling of Burning Emissions Inventory (FLAMBE) (Reid et al., 2009) fire emissions.

Figure 5 shows the scatter plot of comparison of ensemble satellite AOD with model-simulated AOD for each model. All three models underestimate satellite AOD with MB of −0.06, −0.02, and −0.07 for GEOS-Chem, WRF-Chem, and CMAQ, respectively. WRF-Chem simulates the highest AOD while CMAQ shows the lowest AOD on average. Both GEOS-Chem and CMAQ show high correlations of 0.70 and 0.71 than WRF-Chem (0.54). These differences in model-simulated AOD also reflect the differences in the aerosol scheme and calculation of aerosol optical properties in different models (Section 2). For example, GEOS-Chem uses a bulk aerosol

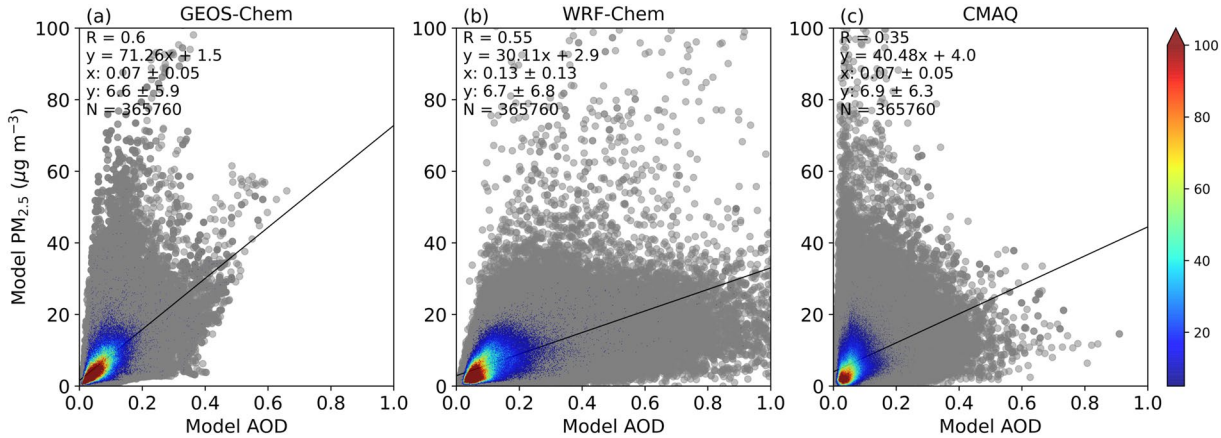

Figure 6. Scatter plot of hourly model-simulated $PM_{2.5}$ and aerosol optical depth (AOD) from: (a) GEOS-Chem, (b) WRF-Chem, (c) CMAQ for June 2012. Also shown on the scatter plot is the correlation coefficient (*R*), the mean ± standard deviation for model AOD (*x*) and model $PM_{2.5}$ (*y*), the number of collocated data points (*N*), the density of points (the color bar), and the best fit linear regression (the solid black line).

approach while both WRF-Chem and CMAQ use a modal approach to represent aerosol size distribution. Both GEOS-Chem and WRF-Chem use a standard Mie code to calculate aerosol optical properties while CMAQ used here applies a regression relationship to calculate AOD. Detailed examination of the difference of the AOD from each model is beyond the scope of this work.

All three models in general underestimated surface measured $PM_{2.5}$ mass concentration (Figure 7), which has also been briefly discussed in Part I of this work (Zhang et al., 2020). Underestimation of surface $PM_{2.5}$ mass concentration in atmospheric chemical models in summer time are common (e.g., McKeen et al., 2007; Simon et al., 2012) due to a variety of reasons such as inaccurate anthropogenic emission inventories, errors in boundary/initial conditions used for regional models and simplified chemistry/physical scheme. Specifically in this work, for example, not including dust emissions in CMAQ or SOA in GEOS-Chem could contribute to the underestimation of surface $PM_{2.5}$ concentration, which could subsequently cause underestimation of AOD values. All three models also underestimate satellite AOD in general. The vertical profiles of aerosols and the scheme used to calculate the aerosol optical properties could play a role as well. It is beyond the scope of this work to reveal the underlying cause of the model performances for both surface $PM_{2.5}$ and AOD. More model sensitivity tests and observational data are needed to further investigate the cause in future work.

4.2. Model-Simulated AOD-$PM_{2.5}$ Ratio

Even though we are not deriving surface $PM_{2.5}$ mass concentration from satellite AOD, the model-simulated AOD-$PM_{2.5}$ relationship plays a role in the bias-correction technique as described in Section 3.2 and also reveals the consistency of model-simulated surface mass concentration and column-integrated aerosol properties. Figure 6 shows the simulated AOD-$PM_{2.5}$ relationship from each model. Overall, the WRF-Chem model simulates higher AOD than both GEOS-Chem and CMAQ but shows lower surface $PM_{2.5}$. In addition, both GEOS-Chem and WRF-Chem models show high correlations of 0.60 and 0.55 than that of CMAQ (0.35). This reflects the various degrees of consistence between model-simulated surface $PM_{2.5}$ and column properties. Previous work (Hogrefe et al., 2015; Saide et al., 2020) and references therein also discussed this perspective specifically for CMAQ and WRF-Chem models, respectively.

4.3. Evaluation of the Bias-Correction Technique

In this section, we focus on the evaluation of the bias-correction technique discussed in Section 3.2. First, we will evaluate the performance of KF techniques (KF-ori as in Part I of this work and the updated KF in current work) for type-1 model grids collocated with EPA hourly measurements of surface $PM_{2.5}$ mass concentration. Then, we will perform evaluation for the ensemble AOD bias-corrected daily $PM_{2.5}$ for today, over areas that are 125 km away from the EPA hourly measurements sites, which serve as the "extended ground truth" or EGT. Lastly, we

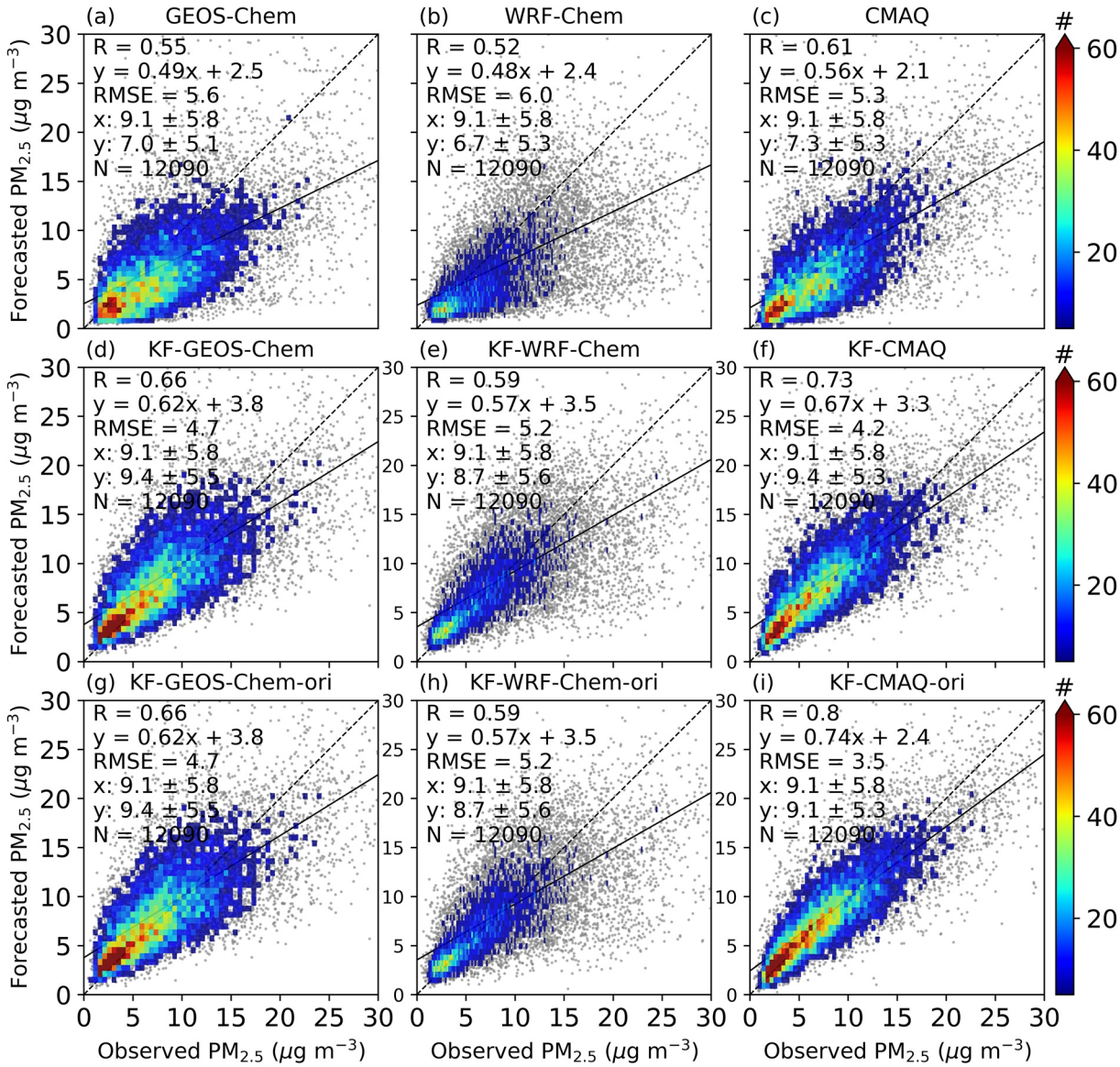

Figure 7. Scatter plot of daily $PM_{2.5}$ concentration between model (*y* axis) and ground observation (*x* axis) at 508 Environmental Protection Agency (EPA) sites for June 2012. (a–c) the evaluation for each raw model; (d–f) the evaluation for Kalman filter corrected model results. Here the Kalman filter implementation is described in Section 3.2. The hourly observed and model-simulated $PM_{2.5}$ from noon yesterday to noon today will be used to correct model forecasted $PM_{2.5}$ from midnight today to midnight next day; (g–i) the evaluation for Kalman filter (KF-ori) corrected model results. Here the Kalman filter is implemented as in Zhang et al. (2020). The hourly observed and model-simulated $PM_{2.5}$ from midnight yesterday to midnight today will be used to correct model forecasted $PM_{2.5}$ from midnight today to midnight next day. Also shown on the scatter plot is the correlation coefficient (R), the root-mean-square error (RMSE), the mean ± standard deviation for observed $PM_{2.5}$ (*x*) and model-simulated $PM_{2.5}$ (*y*), the number of collocated data points (*N*), the density of points (the color bar), the best fit linear regression (the solid black line) and the 1:1 line (the dashed black line).

will apply the KF technique using the EGT from previous step to correct model-predicted daily $PM_{2.5}$ for next day and evaluate the results with IMPROVE data.

4.3.1. Evaluation for KF Techniques

Figure 7 shows the evaluation of raw model forecasted daily $PM_{2.5}$ mass concentration and KF-corrected model-simulated $PM_{2.5}$ with ground observations. Results from two different KF implementation methods are compared. The first implementation uses the hourly observed and model-simulated $PM_{2.5}$ from noon yesterday to noon today to correct model forecasted $PM_{2.5}$ from midnight today to midnight next day, which is the approach adopted in the current work (Section 3.2). The second implementation (KF-ori) takes use of the hourly observed

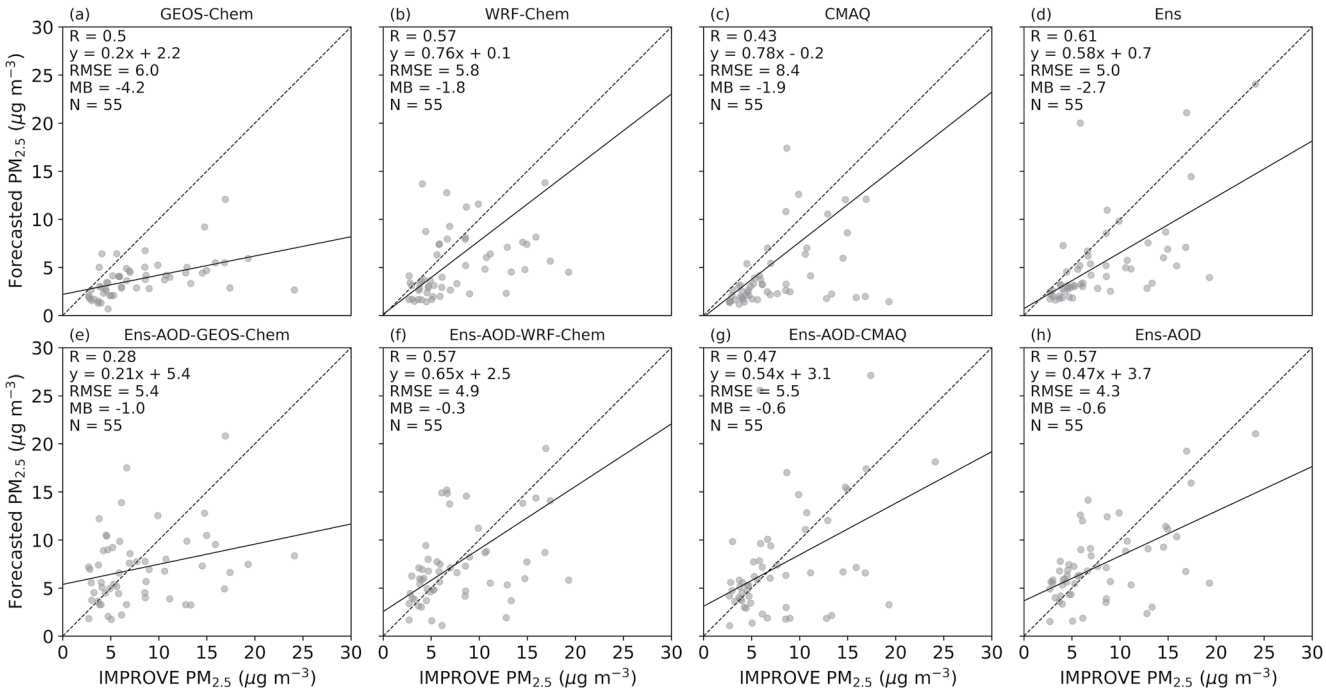

Figure 8. Scatter plot of daily $PM_{2.5}$ concentration between model and ground observations at IMPROVE sites (shown in Figure 9) for June 2012. (a–c) The raw model-simulated daily $PM_{2.5}$. (e–g) The corresponding ensemble AOD bias-corrected daily $PM_{2.5}$ for today. (d and h) The equally weighted ensemble of raw models (a–c) and the corresponding ensemble AOD corrected daily $PM_{2.5}$ (e–g), respectively. Also shown on the scatter plot is the correlation coefficient (R), the root-mean-square error (RMSE), the mean bias (MB), the number of collocated data points (N), the best fit linear regression (the solid black line) and the 1:1 line (the dashed black line).

and model-simulated daily $PM_{2.5}$ from midnight yesterday to midnight today to correct model forecasted $PM_{2.5}$ for the next day of 24 hr, which is the original implementation we developed (Zhang et al., 2020) without practical consideration of data latency. We modified the original KF implementation to consider what is more realistic for making the correction of model bias in the operational mode. Again, the KF technique can improve each individual performance by increasing R, decreasing RMSE and MB. The two KF implementation results in similar model performances, suggesting the 24-hr statistics is robust regardless of the starting hour.

4.3.2. Evaluation of Bias-Corrected $PM_{2.5}$ for Present Day (Today)

As mentioned in Section 3.2, we evaluate our results in rural areas by using the data at the IMPROVE sites because they are located in the rural areas and for some sites, the closest hourly EPA observation sites are in the distance of 125–300 km (Figure 9). Figure 8 shows the evaluation of raw model-simulated and ensemble AOD bias-corrected $PM_{2.5}$ for the present day (today) against IMPROVE measured daily $PM_{2.5}$ mass concentration. The raw model outputs underestimate IMPROVE measurements with MB of −4.2, −1.8, and −1.9 μg m^{-3} for GEOS-Chem, WRF-Chem, and CMAQ, respectively. Their corresponding ensemble AOD bias-corrected $PM_{2.5}$ causes the MB to decrease to −1.0, −0.3, and −0.6 μg m^{-3}, respectively. RMSE also decreases from 6.0 to 5.4, 5.8 to 4.9, and 8.4 to 5.5 μg m^{-3} for GEOS-Chem, WRF-Chem, and CMAQ, respectively. Correlation decreases for GEOS-Chem, stays the same for WRF-Chem and increases slightly for CMAQ. The equally weighted ensemble of the three ensemble AOD bias-corrected $PM_{2.5}$ shows the best performance overall. We only have one satellite AOD data point a day to scale daily $PM_{2.5}$ as described in Section 3.2, which would be challenging to improve correlation. But for air quality forecasts, the predicted $PM_{2.5}$ mass and its bias are the parameters that are most relevant for air quality managers and public health officers to decide if an air quality advisory alert should be issued or not.

Figure 9 shows the spatial distribution of the ratio of raw model and ensemble AOD bias-corrected daily $PM_{2.5}$ compared with IMPROVE daily $PM_{2.5}$ mass concentration. Majority of the sites are located in the Great Plains and Rocky Mountain areas, where access to surface $PM_{2.5}$ measurements is limited. Again, all three models show underestimation of surface $PM_{2.5}$ for most of the sites with median ratio of 0.54, 0.65, and 0.63 for GEOS-Chem, WRF-Chem, and CMAQ, respectively. After using ensemble satellite AOD to correct model bias, most sites show

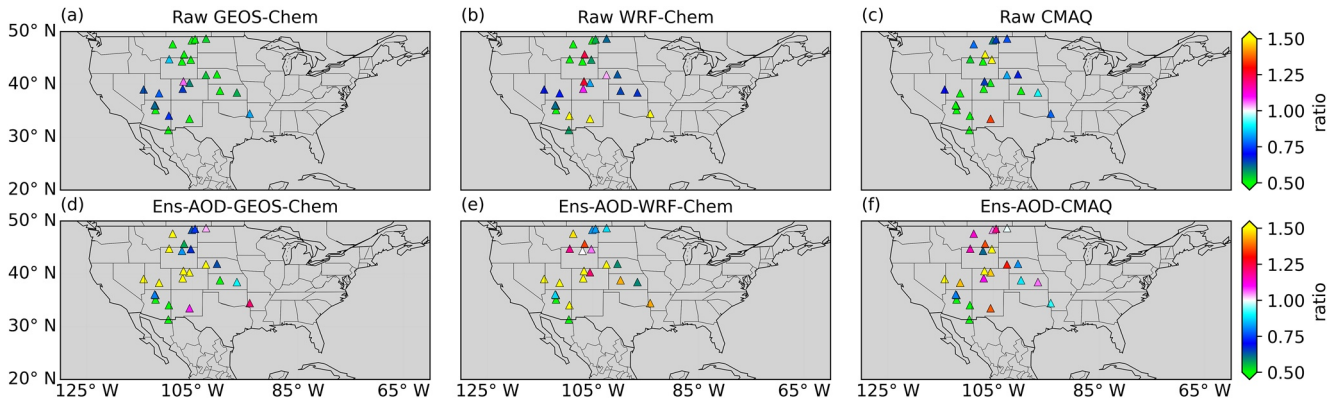

Figure 9. Ratio between model-simulated and measured daily $PM_{2.5}$ at Interagency Monitoring of Protected Visual Environments (IMPROVE) sites for June 2012. (a–c) The ratio for the raw models; (d–f) the ratio for the ensemble aerosol optical depth (AOD) bias-corrected $PM_{2.5}$ for today.

positive bias with median ratio increased to 0.95, 1.21, and 1.08 for three models, respectively, indicating satellite AOD enabled bias correction overall improves the magnitude of surface $PM_{2.5}$.

At the regional scale, however, overestimation occurred in the Rocky Mountain area, especially for GEOS-Chem and WRF-Chem. There are several possible contributors. First, the challenging terrain in the Rocky Mountain area could cause more uncertainties for both satellite and model calculated AOD, compared with the Great Plains area (Drury et al., 2008). Second, there are also some wildfire events over the Rocky Mountain area during June 2012. Compared with other two models, WRF-Chem simulated higher AOD and surface $PM_{2.5}$ can be due to the FLAMBE fire emission inventory that is known to be the at the high end of the biomass burning emission estimates (Zhang et al., 2014). Third, large uncertainties can also be related to satellite-retrieved AOD during wildfire events due to the large variation of smoke optical properties (Reid et al., 2005). Therefore, the effectiveness of the bias correction does have regional dependence because the uncertainties in CTM and satellite AOD retrievals have the regional dependences.

4.3.3. Evaluation of Bias-Corrected $PM_{2.5}$ for Next Day

Due to the AOD availability over both EPA and IMPROVE sites for the month of June 2012, we are only applying the KF correction for IMPROVE sites (Figure S3) that have at least 10 days of corrected $PM_{2.5}$ for today to forecast daily $PM_{2.5}$ concentration for next day. Figure 10 shows the overall performance of KF in predicting next day's daily surface $PM_{2.5}$ concentration in rural areas for each model and model equally weighted ensemble as well. KF improves each individual model performance by decreasing RMSE and MB. Specifically, MB decreases from −3.8 to −2.2, −2.9 to −1.6, and −2.8 to −1.1 μg m^{-3} for GEOS-Chem, WRF-Chem, and CMAQ, respectively. RMSE decreases from 5.6 to 5.1 and 6.7 to 6.0 μg m^{-3} for GEOS-Chem and CMAQ, respectively, and stays the same for WRF-Chem. Similar to the performance of ensemble AOD bias-corrected daily $PM_{2.5}$ concentration for today, the KF does not improve the correlation, which is expected, because KF is best at removing the systematic bias in the model. Overall, the ensemble of these three KF-corrected model forecasts for next day shows the best performance. Figure 11 shows the time series of raw model and KF-corrected daily surface $PM_{2.5}$ concentration for next day for the month of June. Both GEOS-Chem and CMAQ underestimate surface $PM_{2.5}$ measurements throughout the month. WRF-Chem underestimates surface observation during the middle and latter part of the month and matches the surface observation at the beginning of the month relatively well. The KF can adjust model forecasted bias and improve surface $PM_{2.5}$ performance, especially toward the end of the month, when there are more wildfire events.

5. Conclusions and Discussion

This work is the second of a two-part study which aims to improve surface $PM_{2.5}$ forecasts in the continental U.S. through the integrated use of satellite remote sensing AOD, chemical transport model outputs, and ground observations. We developed a multimodel and multi-AOD ensemble bias-correction framework including Kalman filter (KF) technique. We use GEOS-Chem, WRF-Chem, and CMAQ for the model ensemble members

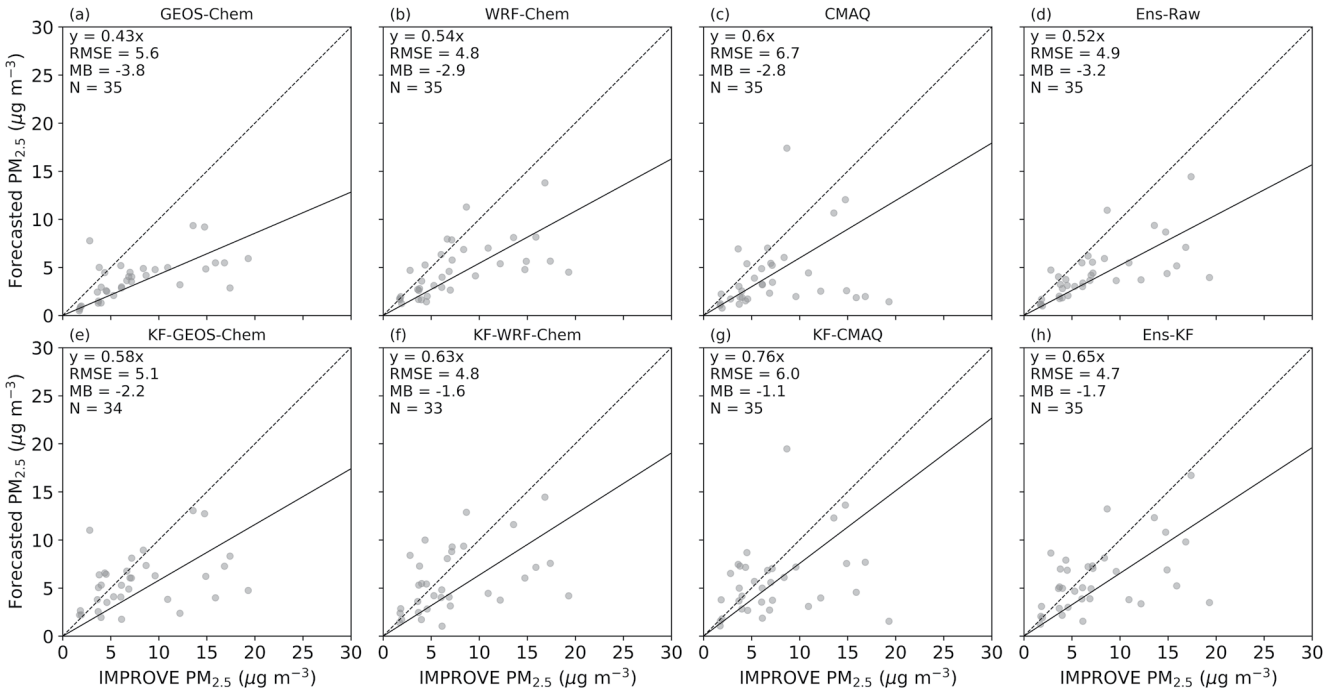

Figure 10. Scatter plot of daily $PM_{2.5}$ concentration between model forecast and ground observations at Interagency Monitoring of Protected Visual Environments (IMPROVE) sites during June 2012. (a–c) The raw model forecasted daily $PM_{2.5}$. (e–g) The corresponding KF-corrected daily $PM_{2.5}$ for next day. (d and h) The equally weighted ensemble of raw models (a–c) and the corresponding KF-corrected model daily $PM_{2.5}$ for next day (e–g), respectively. Also shown on the scatter plot is the root-mean-square error (RMSE), the mean bias (MB), the number of collocated data points (N), the linear regression with fixed intercept (the solid black line) and the 1:1 line (the dashed black line).

and MODIS Terra, Aqua and VIIRS DT/DB AOD as the AOD ensemble members. In the first part (Zhang et al., 2020), we applied an ensemble KF technique to correct model forecast biases in nonrural areas by only using these three model outputs and hourly ground observations from EPA's Air Quality System (AQS). We also applied a successive correction method (SCM) to spread the model bias information to locations within the radius of about 125 km away from these ground observations. In the current work, we focus on improving surface $PM_{2.5}$ forecasts in rural areas that have about 20% of the population in about 98% of the U.S. land. These rural areas are also the areas that do not have ground monitors of air quality and therefore, observation-based analysis of air

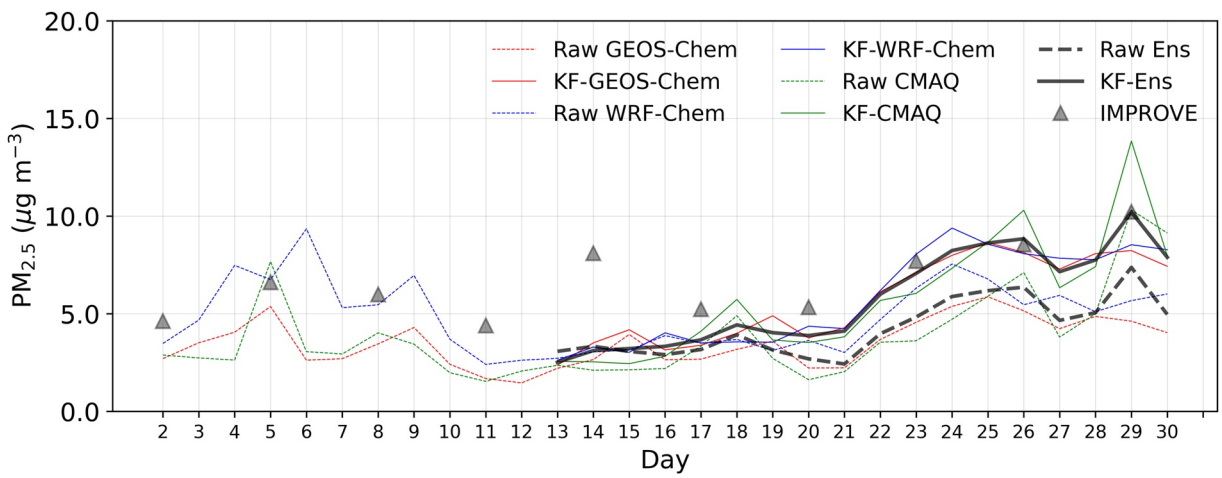

Figure 11. Time series of daily $PM_{2.5}$ concentration averaged over six Interagency Monitoring of Protected Visual Environments (IMPROVE) sites (shown as squares in Figure S3 in the supplement information) for June 2012. The dashed lines show the raw model-simulated daily $PM_{2.5}$ and the solid lines show the corresponding KF-corrected daily $PM_{2.5}$ for next day. The triangles show the measured daily $PM_{2.5}$ from IMPROVE network.

quality has to rely on satellite data. Hence, in this paper, we develop the technique to use satellite data to improve the model-predicted $PM_{2.5}$ concentration to rural areas where the closest EPA air quality monitoring station is at least 125–300 km away.

The technique for bias correction to improve surface $PM_{2.5}$ forecasts for next day in rural areas through AOD ensembles involves three steps. First, we averaged AOD from three different sensors (MODIS Terra, Aqua, and VIIRS) and two retrieval algorithms (DT and DB) to form a single best AOD dataset via an equally weighted ensemble approach. Our results show ensemble AOD from different sensors and algorithms combined show the best performance compared with individual AOD member regarding *R*, RMSE, and MB. This has significant implications for future work considering more satellite will be launched to study aerosols.

In the second step, we correct the model-simulated surface $PM_{2.5}$ concentration for the present day (today) through the AOD spatial pattern between nonrural areas with collocated EPA ground sites and rural areas without any ground observations available. The results are considered as the "extended ground truth," or EGT for today in rural areas. We find that ensemble AOD corrected daily $PM_{2.5}$ for today over these rural areas show improved performances with decreasing RMSE and MB compared with raw models. Overall, the ensemble of all three models with ensemble AOD corrected daily $PM_{2.5}$ shows the best performance. Due to the temporal availability of satellite AOD (once a day for most places in cloud-free conditions), our ensemble AOD corrected daily $PM_{2.5}$ did not result in increased correlation. In the last step, these satellite-based EGT are used in the KF technique (as developed in Part I of this work) to update model forecasted bias for next day and then correct model forecasted $PM_{2.5}$ for next day in the rural areas. Again, the KF could improve each individual model performance with decreased RMSE and MB and the ensemble of three KF-corrected model forecasts for next day shows the best performance. Similarly, the correlation is not improved but our work here opens the door for continuing the work to improve surface $PM_{2.5}$ forecasts in rural areas using satellite data.

While the technique developed here used the hindcast data for the purpose of research, it is designed with a goal of operational applications. Specifically, regarding the scheduling, we assume model predictions for next day will be made available at latest at noon today, and the bias-correction technique can be applied from noon to 5 p.m. local time (5-hr correction window) before a public advisory of air quality can be issued. Because the surface $PM_{2.5}$ data has a latency of 1 hr and satellite AOD has a latency of 2–3 hr, it is expected that by 3 p.m. local time, all the data needed for the bias correction will be available, and therefore, the 5-hr correction window is sufficient for the multiple AOD and multiple model ensemble approach to provide the single best $PM_{2.5}$ air quality forecast at the regional scale. Admittedly, the work here has focused on the development of the method. In the future, the method can be and should be tested in the operational setting, especially considering that we have only used 1 month of data from June 2012 to develop and evaluate the bias-correction framework. Future work should test the bias-correction method for a longer time period, with an emphasis to analyze how the satellite AOD and CTM uncertainties affect the effectiveness of the method at the regional scale.

Acknowledgments
This work was supported by the NASA's Suomi-NPP Program (NNX17A-C94A) and Applied Science Program (80NSSC21K0428 managed by John A. Haynes). It is also partially supported by the NOAA's Atmospheric Chemistry, Carbon Cycle and Climate program (NA19OAR4310178 managed by Monika Kopacz). We acknowledge the public availability of MODIS and VIIRS data from the NASA Earth Observing System Data and Information System (EOSDIS). We appreciate the AERONET networks for making their data publicly available and are grateful for the site PIs and data managers of the networks. We thank the U.S. EPA for providing the measured hourly surface $PM_{2.5}$ mass concentration data. We also thank the IMPROVE network for providing the measured daily surface $PM_{2.5}$ mass concentration data. We appreciate Anthony J. Prenni at National Park Service and Dr. Jenny Hand at Cooperative Institute for Research in the Atmosphere for the helpful discussion about surface $PM_{2.5}$ data from IMPROVE network. We also thank the three anonymous reviewers for their time to provide constructive feedback to improve our manuscript. Disclaimer: The research described in this article has been reviewed by the U.S. Environmental Protection Agency 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

The U.S. EPA provided the measured hourly surface $PM_{2.5}$ mass concentration data which are available at: https://aqs.epa.gov/aqsweb/documents/data_mart_welcome.html (last accessed October 3, 2019). The IMPROVE network provided the measured daily surface $PM_{2.5}$ mass concentration data, which can be downloaded at: http://views.cira.colostate.edu/fed/(last accessed December 15, 2019, users will now need to register an account to download the data). GEOS-Chem is an open-access model which can be downloaded at: https://geos-chem.seas.harvard.edu/. WRF-Chem is also an open-access model which can be acquired at: https://ruc.noaa.gov/wrf/wrf-chem/. CMAQ model outputs can be obtained at https://www.epa.gov/hesc/remote-sensing-information-gateway.

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.

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.

Anderson, J. C., Wang, J., Zeng, J., Leptoukh, G., Petrenko, M., Ichoku, C., & Hu, C. (2013). Long-term statistical assessment of Aqua-MODIS aerosol optical depth over coastal regions: Bias characteristics and uncertainty sources. *Tellus B: Chemical and Physical Meteorology*, *65*(1), 20805.

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.

Anenberg, S. C., Henze, D. K., Tinney, V., Kinney, P. L., Raich, W., Fann, N., et al. (2018). Estimates of the global burden of ambient PM2.5, ozone, and NO_2 on asthma incidence and emergency room visits. *Environmental Health Perspectives*, *126*(10), 107004. https://doi.org/10.1289/EHP3766

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.

Barnard, J. C., Fast, J. D., Paredes-Miranda, G., Arnott, W., & Laskin, A. (2010). Evaluation of the WRF-Chem "Aerosol Chemical to Aerosol Optical Properties" module using data from the MILAGRO campaign. *Atmospheric Chemistry and Physics*, *10*(15), 7325–7340.

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*, *106*(D19), 23073–23095.

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), 4183. https://doi.org/10.1029/2001JD001409

Buchard, V., Randles, C., Da Silva, A., Darmenov, A., Colarco, P., Govindaraju, R., et al. (2017). The MERRA-2 aerosol reanalysis, 1980 onward. Part II: Evaluation and case studies. *Journal of Climate*, *30*(17), 6851–6872.

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.

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.

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), 4661. 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.

Diner, D. J., Beckert, J. C., Reilly, T. H., Bruegge, C. J., Conel, J. E., Kahn, R. A., et al. (1998). Multi-angle Imaging SpectroRadiometer (MISR) instrument description and experiment overview. *IEEE Transactions on Geoscience and Remote Sensing*, *36*(4), 1072–1087.

Drury, E., Jacob, D. J., Wang, J., Spurrs, R. J. D., & Chance, K. (2008). Improved algorithm for MODIS satellite retrievals of aerosol optical depths over land. *Journal of Geophysical Research*, *113*, D16204. https://doi.org/10.1029/2007JD009573

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. *Paper presented at American Meteorological Society Annual Meeting, AMS Sixth Conference on Atmospheric Chemistry: Air Quality in Megacities*.

Eldred, R. A., Cahill, T. A., Wilkinson, L. K., Feeney, P. J., Chow, J. C., & Malm, W. C. (1990). Measurement of fine particles and their chemical components in the IMPROVE/NPS networks. In C. V. Mathai (Ed.), *Transactions: Visibility and fine particles* (pp. 187–196). Pittsburgh: Air & Waste Management Association.

Engel-Cox, J. A., Hoff, R. M., Rogers, R., Dimmick, F., Rush, A. C., Szykman, J. J., et al. (2006). Integrating lidar and satellite optical depth with ambient monitoring for 3-dimensional particulate characterization. *Atmospheric Environment*, *40*(40), 8056–8067.

Fairlie, T. D., Jacob, D. J., Dibb, J. E., Alexander, B., Avery, M. A., van Donkelaar, A., & 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.

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.

Fu, D., Song, Z., Zhang, X., Xia, X., Wang, J., Che, H., et al. (2020). Mitigating MODIS AOD non-random sampling error on surface PM2. 5 estimates by a combined use of Bayesian Maximum Entropy method and linear mixed-effects model. *Atmospheric Pollution Research*, *11*(3), 482–490.

Fu, D., Xia, X., Duan, M., Zhang, X., Li, X., Wang, J., & Liu, J. (2018a). Mapping nighttime PM2. 5 from VIIRS DNB using a linear mixed-effect model. *Atmospheric Environment*, *178*, 214–222.

Fu, D., Xia, X., Wang, J., Zhang, X., Li, X., & Liu, J. (2018b). Synergy of AERONET and MODIS AOD products in the estimation of PM 2.5 concentrations in Beijing. *Scientific Reports*, *8*(1), 1–8.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*, 5419–5454.

Gupta, P., & Christopher, S. A. (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. *Journal of Geophysical Research*, *114*, D20205. https://doi.org/10.1029/2008JD011497

Gupta, P., Christopher, S. A., Wang, J., Gehrig, R., Lee, Y., & Kumar, N. (2006). Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment*, *40*(30), 5880–5892.

Hand, J., Schichtel, B., Pitchford, M., Malm, W., & Frank, N. (2012). Seasonal composition of remote and urban fine particulate matter in the United States. *Journal of Geophysical Research*, *117*, D05209. https://doi.org/10.1029/2011JD017122

Hand, J. L., Prenni, A. J., Schichtel, B. A., Malm, W. C., & Chow, J. C. (2019). Trends in remote PM2. 5 residual mass across the United States: Implications for aerosol mass reconstruction in the IMPROVE network. *Atmospheric Environment*, *203*, 141–152.

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.

Hogrefe, C., Pouliot, G., Wong, D., Torian, A., Roselle, S., Pleim, J., & Mathur, R. (2015). Annual application and evaluation of the online coupled WRF–CMAQ system over North America under AQMEII phase 2. *Atmospheric Environment*, *115*, 683–694.

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J., Setzer, A., et al. (1998). AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, *66*(1), 1–16.

Hsu, N., Jeong, M. J., Bettenhausen, C., Sayer, A., Hansell, R., Seftor, C., et al. (2013). Enhanced Deep Blue aerosol retrieval algorithm: The second generation. *Journal of Geophysical Research: Atmospheres*, *118*, 9296–9315. https://doi.org/10.1002/jgrd.50712

Hsu, N., Lee, J., Sayer, A., Kim, W., Bettenhausen, C., & Tsay, S. C. (2019). VIIRS Deep Blue aerosol products over land: Extending the EOS long-term aerosol data records. *Journal of Geophysical Research: Atmospheres*, *124*, 4026–4053. https://doi.org/10.1029/2018JD029688

Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J., & Liu, Y. (2017). Estimating PM2. 5 concentrations in the conterminous United States using the random forest approach. *Environmental Science & Technology*, *51*(12), 6936–6944.

Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes, M. G. Jr., Estes, S. M., et al. (2013). Estimating ground-level PM2. 5 concentrations in the southeastern US using geographically weighted regression. *Environmental Research*, *121*, 1–10.

Hyer, E., Reid, J., & Zhang, J. (2011). An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals. *Atmospheric Measurement Techniques*, *4*(3), 379–408.

Ichoku, C., Chu, D. A., Mattoo, S., Kaufman, Y. J., Remer, L. A., Tanré, D., et al. (2002). A spatio-temporal approach for global validation and analysis of MODIS aerosol products. *Geophysical Research Letters*, *29*(12), 8006. https://doi.org/10.1029/2001GL013206

Jackson, J. M., Liu, H., Laszlo, I., Kondragunta, S., Remer, L. A., Huang, J., & Huang, H. C. (2013). Suomi-NPP VIIRS aerosol algorithms and data products. *Journal of Geophysical Research: Atmospheres*, *118*, 12673–12689. https://doi.org/10.1002/2013JD020449

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.

Jayanty, R., Flanagan, J. B., & Rickman, Jr, E. E. (2004). An overview of PM2. 5 chemical speciation nationwide network program in the United States. *Paper presented at 13th World Clean Air Congress*.

Koelemeijer, R., Homan, C., & Matthijsen, J. (2006). Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. *Atmospheric Environment*, *40*(27), 5304–5315.

Koepke, P., Hess, M., Schult, I., & Shettle, E. (1997). *Global aerosol data set* (Rep. No. 243). Max-Planck-Institut für Meteorologie.

Lee, H. J., Chatfield, R. B., & Strawa, A. W. (2016). Enhancing the applicability of satellite remote sensing for PM2.5 estimation using MODIS Deep Blue AOD and land use regression in California, United States. *Environmental Science & Technology*, *50*(12), 6546–6555. https://doi.org/10.1021/acs.est.6b01438

Levy, R., Mattoo, S., Munchak, L., Remer, L., Sayer, A., Patadia, F., & Hsu, N. (2013). The collection 6 MODIS aerosol products over land and ocean. *Atmospheric Measurement Techniques*, *6*(11), 2989.

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*, *112*, D06201. https://doi.org/10.1029/2006JD007813

Liu, Y., Koutrakis, P., & Kahn, R. (2007). Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: Part 1—Method development. *Journal of the Air & Waste Management Association*, *57*(11), 1351–1359.

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*, *109*, D22206. https://doi.org/10.1029/2004JD005025

Liu, Y., Sarnat, J. A., Kilaru, V., Jacob, D. J., & Koutrakis, P. (2005). Estimating ground-level PM2. 5 in the eastern United States using satellite remote sensing. *Environmental Science & Technology*, *39*(9), 3269–3278.

Ma, Z., Hu, X., Huang, L., Bi, J., & Liu, Y. (2014). Estimating ground-level PM2. 5 in China using satellite remote sensing. *Environmental Science & Technology*, *48*(13), 7436–7444.

Malm, W. C., & Hand, J. L. (2007). An examination of the physical and optical properties of aerosols collected in the IMPROVE program. *Atmospheric Environment*, *41*(16), 3407–3427.

Malm, W. C., Schichtel, B. A., & Pitchford, M. L. (2011). Uncertainties in PM2. 5 gravimetric and speciation measurements and what we can learn from them. *Journal of the Air & Waste Management Association*, *61*(11), 1131–1149.

Malm, W. C., Sisler, J. F., Huffman, D., Eldred, R. A., & Cahill, T. A. (1994). Spatial and seasonal trends in particle concentration and optical extinction in the United States. *Journal of Geophysical Research*, *99*(D1), 1347–1370.

Martin, R. V., Jacob, D. J., Yantosca, R. M., Chin, M., & Ginoux, P. (2003). Global and regional decreases in tropospheric oxidants from photochemical effects of aerosols. *Journal of Geophysical Research: Atmospheres*, *108*(D3), 4097. https://doi.org/10.1029/2002JD002622

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*, *112*, D10S20. https://doi.org/10.1029/2006JD007608

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*, *108*(D12), 4355. 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*, *109*, D15204. https://doi.org/10.1029/2003JD004473

Pitchford, M., Malm, W., Schichtel, B., Kumar, N., Lowenthal, D., & Hand, J. (2007). Revised algorithm for estimating light extinction from IMPROVE particle speciation data. *Journal of the Air & Waste Management Association*, *57*(11), 1326–1336. https://doi.org/10.3155/1047-3289.57.11.1326

Puttaswamy, S., Nguyen, H. M., Braverman, A., Hu, X., & Liu, Y. (2014). Statistical data fusion of multi-sensor AOD over the continental United States. *Geocarto International*, *29*(1), 48–64.

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*, *114*, D01205. https://doi.org/10.1029/2008JD010701

Randles, C., Da Silva, A., Buchard, V., Colarco, P., Darmenov, A., Govindaraju, R., et al. (2017). The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. *Journal of Climate*, *30*, 6823–6850.

Reid, J. S., Eck, T. F., Christopher, S. a., Koppmann, R., Dubovik, O., Eleuterio, D., et al. (2005). A review of biomass burning emissions part III: Intensive optical properties of biomass burning particles. *Atmospheric Chemistry and Physics*, *5*(3), 827–849.

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

Remer, L. A., Kaufman, Y., Tanré, D., Mattoo, S., Chu, D., Martins, J. V., et al. (2005). The MODIS aerosol algorithm, products, and validation. *Journal of the Atmospheric Sciences*, *62*(4), 947–973.

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., et al. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society*, *85*(3), 381–394.

Saide, P. E., Gao, M., Lu, Z., Goldberg, D. L., Streets, D. G., Woo, J.-H., et al. (2020). Understanding and improving model representation of aerosol optical properties for a Chinese haze event measured during KORUS-AQ. *Atmospheric Chemistry and Physics*, *20*(11), 6455–6478.

Sathe, Y., Kulkarni, S., Gupta, P., Kaginalkar, A., Islam, S., & Gargava, P. (2019). Application of Moderate Resolution Imaging Spectroradiometer (MODIS) Aerosol Optical Depth (AOD) and Weather Research Forecasting (WRF) model meteorological data for assessment of fine particulate matter (PM2. 5) over India. *Atmospheric Pollution Research*, *10*(2), 418–434.

Sawyer, V., Levy, R. C., Mattoo, S., Cureton, G., Shi, Y., & Remer, L. A. (2020). Continuing the MODIS dark target aerosol time series with VIIRS. *Remote Sensing*, *12*(2), 308.

Sayer, A. M., Hsu, N. C., Lee, J., Kim, W. V., & Dutcher, S. T. (2019). Validation, stability, and consistency of MODIS collection 6.1 and VIIRS Version 1 deep blue aerosol data over land. *Journal of Geophysical Research: Atmospheres*, *124*, 4658–4688. https://doi.org/10.1029/2018JD029598

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*, *106*(D22), 28275–28293. http://dx.doi.org/10.1029/2001JD000384

Sha, T., Ma, X., Zhang, H., Janechek, N., Wang, Y., Wang, Y., et al. (2021). Impacts of soil NO_x emission on O_3 air quality in rural California. *Environmental Science & Technology*, *55*(10), 7113–7122. https://doi.org/10.1021/acs.est.0c06834

Shen, H., Li, T., Yuan, Q., & Zhang, L. (2018). Estimating regional ground-level PM2.5 directly from satellite top-of-atmosphere reflectance using deep belief networks. *Journal of Geophysical Research: Atmospheres*, *123*), 13875–13886. https://doi.org/10.1029/2018JD028759

Shi, Y., Zhang, J., Reid, J., Hyer, E., & Hsu, N. (2013). Critical evaluation of the MODIS Deep Blue aerosol optical depth product for data assimilation over North Africa. *Atmospheric Measurement Techniques*, *6*(4), 949.

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.

Sogacheva, L., Popp, T., Sayer, A. M., Dubovik, O., Garay, M. J., Heckel, A., et al. (2020). Merging regional and global aerosol optical depth records from major available satellite products. *Atmospheric Chemistry and Physics*, *20*, 2031–2056.

Solomon, P. A., Crumpler, D., Flanagan, J. B., Jayanty, R., Rickman, E. E., & McDade, C. E. (2014). US National PM2. 5 chemical speciation monitoring networks—CSN and IMPROVE: Description of networks. *Journal of the Air & Waste Management Association*, *64*(12), 1410–1438.

Szykman, J., Kondragunta, S., Zhang, H., Dickerson, P., van Donkelaar, A., Martin, R., et al. (2012). Expanding the estimation of surface PM2. 5 from Aqua and Terra MODIS aerosol optical depth in the EPA's AirNow Satellite Data Processor to Suomi NPP VIIRS. *Paper presented at the 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, PM2. 5. Paper presented at the *Proceedings of the sixth conference on atmospheric chemistry*. American Meteorological Society.

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.

Wang, J., Aegerter, C., Xu, X., & Szykman, J. J. (2016). Potential application of VIIRS Day/Night Band for monitoring nighttime surface PM2. 5 air quality from space. *Atmospheric Environment*, *124*, 55–63.

Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM2. 5 mass: Implications for air quality studies. *Geophysical Research Letters*, *30*(21), 2095. https://doi.org/10.1029/2003GL018174

Wang, J., Yue, Y., Wang, Y., Ichoku, C., Ellison, L., & Zeng, J. (2018). Mitigating satellite-based fire sampling limitations in deriving biomass burning emission rates: Application to WRF-Chem model over the northern sub-Saharan African Region. *Journal of Geophysical Research: Atmospheres*, *123*, 507–528. https://doi.org/10.1002/2017JD026840

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*, *109*, D22307. https://doi.org/10.1029/2004JD005237

Wei, J., Li, Z., Peng, Y., & Sun, L. (2019). MODIS Collection 6.1 aerosol optical depth products over land and ocean: Validation and comparison. *Atmospheric Environment*, *201*, 428–440.

Whitby, E. R., McMurry, P., Shankar, U., & Binkowski, F. (1991). *Modal Aerosol Dynamics Modeling*. *27*(6), 673–688.

Xian, P., Reid, J. S., Hyer, E. J., Sampson, C. R., Rubin, J. I., Ades, M., et al. (2019). Current state of the global operational aerosol multi-model ensemble: An update from the International Cooperative for Aerosol Prediction (ICAP). *Quarterly Journal of the Royal Meteorological Society*, *145*(S1), 176–209. https://doi.org/10.1002/qj.3497

Zhang, F., Wang, J., Ichoku, C., Hyer, E., Yang, Z., Ge, C., et al. (2014). Sensitivity of mesoscale modeling of smoke direct radiative effect to the emission inventory: A case study in northern sub-Saharan African region, *Environmental Research Letter*, *9*, 075002.

Zhang, H., & Kondragunta, S. (2021). Daily and hourly surface PM2.5 estimation from satellite AOD. *Earth and Space Science*, *8*(3), e2020EA001599. https://doi.org/10.1029/2020EA001599

Zhang, H., Wang, J., García, L. C., Ge, C., Plessel, T., Szykman, J., et al. (2020). Improving surface PM2.5 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

Zhang, J., & Reid, J. S. (2006). MODIS aerosol product analysis for data assimilation: Assessment of over-ocean level 2 aerosol optical thickness retrievals. *Journal of Geophysical Research*, *111*, D22207. https://doi.org/10.1029/2005JD006898

Zheng, Y., Zhang, Q., Liu, Y., Geng, G., & He, K. (2016). Estimating ground-level PM2. 5 concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. *Atmospheric Environment*, *124*, 232–242.