DEVELOPMENT AND EVALUATION OF NORTH AMERICA ENSEMBLE FORECASTS OF WILDFIRES AND DUST STORMS

by

Peewara Makkaroon

A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Science
Earth Systems Science

Committee:

______________________	Dr. Daniel Tong, Thesis Chair
______________________	Dr. Yunyao Li, Committee Member
______________________	Dr. Zafer Boybeyi, Committee Member
______________________	Dr. Mark Uhen, Department Chairperson
______________________	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
______________________	Dr. Fernando R. Miralles-Wilhelm Dean, College of Science
Date: ________________	Spring Semester 2022 George Mason University Fairfax, VA

Development and Evaluation of North America Ensemble Forecasts of Wildfires and Dust Storms

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Peewara Makkaroon
Bachelor of Science
Chulalongkorn University, 2018

Director: Daniel Tong, Associate Professor
Department of Atmospheric, Oceanic, and Earth Sciences

Spring Semester 2022
George Mason University
Fairfax, VA

DEDICATION

This thesis is dedicated to my loving family and friends, who have always encouraged me and supported me throughout the challenging degree program.

ACKNOWLEDGEMENTS

I would like to express my most profound appreciation to my committee chair, Dr. Daniel Tong. He gave me an opportunity to work as a graduate research assistant at George Mason University. He constantly shares his research and work experiences with me and guides me to complete this thesis successfully. I would also like to thank my two committee members, Dr. Yunyao Li, who always shares research and coding resources with me, and Dr. Zafer Boybeyi, who provides valuable recommendations on my thesis. I will always appreciate all they have done. Finally, I'm thankful to my academic advisor, Dr. Geoffrey Gilleaudeau, for guiding me with class registration and thesis processes during the degree program.

This study is financially supported by NASA Health and Air Quality Program and NOAA Weather Program Office. We thank NASA, NOAA, and NRL for providing the model prediction data to create the ensemble forecast. In addition, ground measurements collected by the EPA and satellite AOD data by NOAA and NASA are gratefully acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

DEVELOPMENT AND EVALUATION OF NORTH AMERICA ENSEMBLE FORECASTS OF WILDFIRES AND DUST STORMS

Peewara Makkaroon, M.S.

George Mason University, 2022

Thesis Director: Dr. Daniel Tong

Wildfires and dust storms are two major emission sources of aerosols in the atmosphere, exerting myriad effects on air quality, climate, and human health. Predicting wildfires and dust storms is challenging due to large uncertainties in the inputs and representation of chemical and physical processes in the atmospheric models. Ensemble forecasting has been proposed to improve the predictability of wildfire and dust aerosols. This work presents the development and evaluation of a multi-model ensemble forecast system of wildfire and dust air pollution over North America, leveraging research and operational forecasts operated by George Mason University (GMU) and three U.S. federal agencies: National Oceanic and Atmospheric Administration (NOAA), National Aerospace and Space Agency (NASA), and Naval Research Laboratory (NRL). The ensemble members include three regional models (GMU CMAQ, NOAA NACC-CMAQ, and NOAA HYSPLIT), three global models (NOAA GEFS-Aerosols, NASA GEOS-5, and NRL

NAAPS), and one global ensemble (ICAP-MME). Performance of the ensemble forecast was evaluated with aerosol optical depth (AOD) products from MODIS MAIAC, VIIRS-SNPP enhanced Dark Target (DT) and Deep Blue (DB), and surface $PM_{2.5}$ (fine particle) from the AirNow ground network during the 2020 Gigafire events (August-September 2020) in the western United States and the 2021 Spring Dust Season in the Chihuahuan Desert.

For the wildfire ensemble, the results showed that, compared to the individual models, the ensemble mean significantly reduced the biases in the wildfire air pollution forecasts and produced more persistently reliable forecasts during extreme fire events. For AOD forecasts, the ensemble mean was able to improve model performance, such as increasing the correlation to 0.57 (0.62) from a range of 0.30-0.53 (0.35-0.56) by individual models when compared to the VIIRS (MAIAC). The ensemble mean also yields the best (second best) overall *RANK*, a composite indicator representing four statistical metrics (correlation, fractional bias, area hit rate, and false alarm ratio) compared to VIIRS (MAIAC). For the forecast of surface $PM_{2.5}$ concentration, the ensemble mean demonstrated better performance than any single model with the strongest correlation (0.60 vs 0.43-0.54 by individual models), lowest fractional bias (0.54 vs 0.55-1.32), highest hit rate (87% vs 40%-82%), and highest *RANK* (2.83 vs 2.40-2.81), when compared to the AirNow observations. Finally, the ensemble shows the potential to provide a suitable exceedance probability forecast during wildfires with the lowest area false alarm ratio (1.52%) achieved by the ensemble probability of 100%.

For the dust ensemble, the ensemble mean moderately reduced biases in the dust air pollution forecasts and provided fairly reliable AOD and $PM_{2.5}$ forecasts during extreme dust storms compared to the individual models. For AOD forecasts, the ensemble mean improved forecasting performance less successfully than expected, as demonstrated by slightly decreasing mean bias to 0.01 (0.07) based on VIIRS DT (VIIRS DB), increasing correlation to 0.32 at the low level highest from a range of 0.09-0.31 (VIIRS DB), and yielding the third best overall *RANK* compared to VIIRS DT and DB. For surface $PM_{2.5}$ forecasts, the ensemble mean underperformed with a slightly reduced mean bias (3.14), moderately improved low-level correlation (0.40), low area hit rates (15%), and the third best *RANK*. The ensemble was able to provide only low-medium (20-60%) exceedance probability forecasts during dust events. In addition, the low correlations and large biases of the dust ensemble forecasts during the extreme dust episodes indicate worse performance compared to that of wildfire ensemble forecasts due to larger uncertainties in predicting dust emission, dispersion, and removal.

The thesis findings highlight that using the ensemble approach can reduce biases in air pollution forecasts and reasonably improve the model predictability during extreme events such as wildfires and dust storms. The proposed ensemble exceedance probability forecast can be further applied to early warnings of severe air pollution episodes during wildfires and dust storms. However, the reliability of the ensemble forecast is still subject to types of extreme events due to different emission sources as well as initial and boundary meteorological conditions.

CHAPTER 1: INTRODUCTION

1.1 Wildfires in the United States

Wildfires are important emission sources that contribute large amounts of aerosols and trace gases to the atmosphere, leading to hazardous air quality. Wildfire air pollution causes adverse respiratory health effects, visibility degradation, and premature mortality, which in turn lead to economic burdens (Fann et al., 2018; Ford et al., 2018; Neumann et al., 2021). Over the past several decades, the frequency and intensity of both small and large wildfire events in the United States (U.S.) have been rapidly increasing in wildfire-prone areas in the Western U.S., such as the Southwest, the Rocky Mountains, the northern Great Plains, and the Pacific Coast (Liu et al., 2013) as a result of climate change from anthropogenic activities causing rising temperatures (Liu et al., 2013; Pierce et al., 2013; Schoennagel et al., 2017). In addition, a sharp increase in the number of small wildfires in the Western U.S. is mainly due to human activities, such as changing land cover by expanding cities into wildlands and increasing human ignitions from campfires, powerlines, and vehicles. (Li and Banerjee, 2021; McClure and Jaffe, 2018; Salguero et al., 2020; Stevens-Rumann et al., 2018). The National Interagency Fire Center (NIFC) reported that in 2020, there were 58,950 fires across the U.S., more than 10 million acres burned (NICC, 2020), and that most fires took place in the Western United States. Northern California in particular was affected and has experienced the largest recorded wildfires during Summer-Fall 2020 fire season (California Department of Forestry and Fire Protection [CAL FIRE], 2020).

1.2 Dust Storms in the United States

Dust particles are known as a major component of particulate matter less than 2.5 μm in aerodynamic diameter ($PM_{2.5}$) in the western U.S. during spring (Hand et al., 2011, 2016, 2017) as a result of powerful, sustained winds and relatively low precipitation (Flagg et al., 2014). Dust events during spring and summer have been on the rise in the western U.S. over the past several decades (Tong et al., 2017), and it will be worsening significantly in the upcoming years over the southern Great Plains due to severe drought, soil moisture deficits, warming temperatures, and variations in sea surface temperature from climate change (Achakulwisut et al., 2018; Hand et al., 2016; Pu and Ginoux, 2017, 2018; Tong et al., 2017). This projected trend has drawn many concerns about its detrimental impacts on the atmospheric environment (Balkanski et al., 2007; Benedetti et al., 2014; Forster et al., 2007; Wu et al., 2016), ecosystem (Barkley et al., 2019; Mills, 2004; Prospero et al., 2020; Swap et al., 1994), and human health (World Health Organization [WHO], 2021). Dust storms can lift large amounts of soil-derived dust particles into the air. As a result, the concentrations of small particulate matter within active dust regions are elevated beyond the safety air quality standard level and can induce adverse health effects, such as severe respiratory diseases (Tobias et al., 2019), cardiovascular health issues (Crooks et al., 2016), as well as raising Valley Fever incidence rate (CDC, 2013; Tong et al., 2017). Apart from the direct health impacts, dust particles contribute to increasing transportation accidents due to degraded visibility, especially on highways during intense dust storms (Ashley et al., 2015; Lader et al., 2016; Van Pelt et al., 2020b).

Dust in the western U.S. is primarily generated by wind erosion of exposed soil surfaces in arid or semi-arid regions and can be transported across the Contiguous United States (CONUS). Natural sources of dust in the western U.S. are the North American Deserts (namely the Chihuahuan, Great Basin, Mojave, and Sonoran deserts) (Ginoux et al., 2012; Jewell & Nicoll, 2011; Reynolds et al., 2007; Rivera et al., 2010; Tanaka & Chiba, 2006). In contrast, anthropogenic sources of dust are primarily associated with agricultural activities in the southern Great Plains and the Colorado Plateau (Carmona et al., 2015; Ginoux et al., 2012; Neff et al., 2008; Saxton et al., 2000; Skiles et al., 2015; Reynolds et al., 2016). In addition to local dust sources in the U.S., long-range dust transports from Asia and Africa across the Atlantic and Pacific Oceans in spring and summer (March-August) subsequently contribute to total dust in the United States. Generally, the trans-Pacific dust, transported from Asia by strong tropical cyclones and westerly winds during spring (February to June), frequently affects the Pacific coastal regions of the western U.S. (Creamean et al., 2014; Fairlie et al., 2007; Fischer et al., 2009; Kavouras et al., 2009; VanCuren & Cahill, 2002; Zhao et al., 2008), while the trans-Atlantic dust transported from Africa by powerful easterly winds commonly impacts the Caribbean Islands, the Gulf of Mexico, and the southeastern U.S. (Prospero, 1981; Prospero and MayolBracero, 2013; Prospero and colleagues, 2021). These long-range transported dusts contribute to high background particulate matter (PM) concentrations not in both the western U.S. (Fischer et al., 2009; Jaffe et al., 2003) and eastern (DeBell et al., 2004) during spring and summer.

1.3 Multi-Model Ensemble Forecasts

Regarding the concerns about human health affected by degraded air quality during wildfires and dust storms, many operational forecasting systems have been developed to forecast the dispersion of aerosols with the main goal of protecting the public from harmful air quality during hazardous air quality events (Basart et al., 2012; Campbell et al., 2021; Colarco et al., 2010; Hamill et al., 2011a, b; Johnson et al., 2011; Liu et al. 2007; Lu et al. 2010, 2013; Liu and Westphal 2001; Marticorena and Bergametti, 1995; Marticorena et al., 1997; Nickovic et al. 2001; Pérez et al. 2011; Rienecker et al. 2008; Stein et al., 2015; Terradellas et al. 2011; Wang et al., 2000; Walker et al., 2009; Xian et al., 2019; Li et al., 2021). However, the accuracy of deterministic forecasts from a single model is predominantly deteriorated by uncertainties in emission and meteorological input data, model simulations, physical and chemical processes (Cakmur et al., 2004; Darmenova et al., 2009; Delle Monache and Stull, 2003; Di Tomaso et al., 2017; Ginoux et al., 2012; Gong and Zhang, 2008; Grini et al., 2005; Kang et al., 2011; Kumar et al., 2020; Marticorena and Bergametti, 1995; Li et al., 2020; Shao et al., 1996; Textor et al., 2006; Uno et al., 2006), and surface properties (e.g., soil roughness, soil moisture, and vegetation types) (Grini et al., 2005).

Alternatively, one effective way to improve predicting performance is using a mean of the ensemble approach, which can provide probabilistic forecasts by calculating the mean from either multiple models or input data (Delle Monache and Stull, 2003; Delle Monache et al. 2006a, b, 2008; Delle Monache et al., 2020; Li et al., 2020; Petersen et al., 2019; Solazzo et al., 2012; Xian et al., 2019). The major advantage of the ensemble mean forecast over a single model forecast is that it can reduce the biases in forecasts of ensemble members by averaging them out and the uncertainties in ensemble forecasts can also be determined from the spreads of ensemble members.

1.4 Objectives

This thesis aims to develop multi-model ensemble forecasts of wildfire and dust air pollution based on the mean of participating models (ensemble mean) for the Contiguous United States (CONUS). The ensemble members include three regional models, three global models, and one global ensemble. The regional systems include the George Mason University-Community Multiscale Air Quality (GMU-CMAQ), National Oceanic and Atmospheric Administration-U.S. Environmental Protection Agency (NOAA-EPA) Atmosphere-Chemistry Coupler-Community Multiscale Air Quality (NACC-CMAQ), and NOAA Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) models. GMU-CMAQ is a research forecasting system run by the air quality group of George Mason University (GMU) (Li et al., 2021) to provide daily air quality forecasts across the U.S. for the general public. NACC-CMAQ, a model currently being used in NOAA's operational National Air Quality Forecasting Capability (NAQFC) (Campbell et al., 2022), and HYSPLIT, a common atmospheric transport and dispersion model that is developed at NOAA/Air Resources Laboratory (ARL) and used widely in the atmospheric sciences community (Stein et al., 2015). The four global models are: Global Ensemble Forecast System Aerosols (GEFS-Aerosols), NASA Goddard Earth Observing System (GEOS, version 5), International Cooperative for Aerosol Prediction Multi-Model aerosol forecasting Ensemble (ICAP-MME), and Navy Aerosol Analysis and Prediction System (NAAPS). GEFS-Aerosols, a global atmospheric composition model, is developed by at the National Centers for Environmental Prediction (NCEP) in collaboration with the NOAA Global Systems Laboratory (GSL), NOAA Chemical Sciences Laboratory (CSL), and NOAA/ARL (Hamill et al., 2011a, b). GEOS is a weather and climate capable model and is a significant part of the GEOS atmospheric data assimilation system (DAS) and Earth system model developed at NASA's Global Modeling and Assimilation Office (GMAO) (Rienecker et al., 2008). NAAPS, the U.S. Navy's operational global aerosol transport model with consideration of processes associated with aerosol lifecycles, and AOD data assimilation, is developed at the Naval Research Laboratory (NRL) (Lynch et al., 2016). Finally, the ICAP-MME is a global ensemble mean produced from nine comprehensive global speciated aerosol and/or dust models (Xian et al., 2019).

The performance of the ensemble mean in forecasting Aerosol Optical Depth (AOD) were intercompared with ensemble members and verified by evaluating AOD simulations against the Multi-Angle Implementation of Atmospheric Correction (MAIAC; Lyapustin et al., 2011a,b; 2012; 2018) and Visible Infrared Imaging Radiometer Suite onboard the Suomi National Polar-orbiting Partnership (SNPP) (VIIRS-SNPP; Cao et al., 2013a, 2013b; Uprety et al., 2013) satellite-retrieved AOD products based on enhanced Dark Target (DT) algorithm over dark and bright surfaces (Zhang et al., 2016) for the application to the wildfires, while the AOD simulations were evaluated against the VIIRS enhanced DT and VIIRS Deep Blue (DB; Hsu et al., 2013; Hsu et al., 2019) products for the application to the dust storms. The $PM_{2.5}$ concentrations simulated by the ensemble mean were also intercompared with ensemble members and verified with the AirNow $PM_{2.5}$ ground observations for both applications to wildfires and dust storms. The evaluation results were derived by analyzing a suite of statistical metrics during the 2020 Gigafire event (August-September 2020), caused by the August Complex Fire burning more than 1 million acres in Northern California, and during the 2021 Spring Dust Season (January-March 2021), a period when the dust storms driven by a strong low-pressure system occurred predominantly in the Chihuahuan Desert in western Texas and southern New Mexico.

As the air quality models are used to provide air pollution warnings to the public, the ability of the ensemble to produce a reliable forecast of health-based $PM_{2.5}$ exceedances of NAAQS (24-hr $PM_{2.5}$ concentration above 35 $\mu g/m^3$; U.S. EPA, 2020a) during extreme wildfires is crucial. Therefore, we created the ensemble probability forecasts of $PM_{2.5}$ exceedances influenced by wildfires and dust storms and evaluated them with the observed exceedances by AirNow ground monitoring network.

CHAPTER 2: CASE STUDIES OF WILDFIRES AND DUST STORMS

2.1 The 2020 Gigafire in the Western United States

In 2020, California experienced 9,917 incidents of multiple complex wildfires, leading to over 4 million acres burned, and 10,488 structures destroyed (CAL FIRE, 2020). On August 16th, the largest and the most complex recorded wildfire ever known as the "August Complex fire" occurred in Northern California. The fire initially started burning in the Mendocino National Forest from lightning strikes coupled with a heatwave and severe drought driven by climate change. The August Complex fire was a combination of the Doe, Tatham, Glade, and Hull fires. On September 9th, the Doe fire (main fire) became the single largest complex wildfire, even larger than the 2018 Mendocino Complex fire. Later on, the Doe fire merged with other following fires and continued burning until November 12th. The fire was the first "Gigafire", active for 86 days, burned more than 1 million acres, and destroyed 935 structures across the Coast Range counties (Colusa, Glenn, Humboldt, Lake, Mendocino, Tehama, and Trinity). The spanning of fires is dominated by the Diablo winds (offshore winds) over these areas. Figure 1a displays extremely high observed $PM_{2.5}$ concentrations from AQI sites mainly in the western U.S. on September 12th, 2021, when the fires were very intense. Extremely high daily $PM_{2.5}$ concentrations above the daily National Ambient Air Quality Standards (NAAQS) for $PM_{2.5}$ (>35 μg/m^3) were recorded at many AirNow monitoring sites across the U.S. between September 10th-17th, 2020 primarily over California, Oregon, and Washington as shown in Figure 1b. Consequently, our study will focus on AOD and $PM_{2.5}$ simulations during the 2021 wildfire season, from August to September 2020.

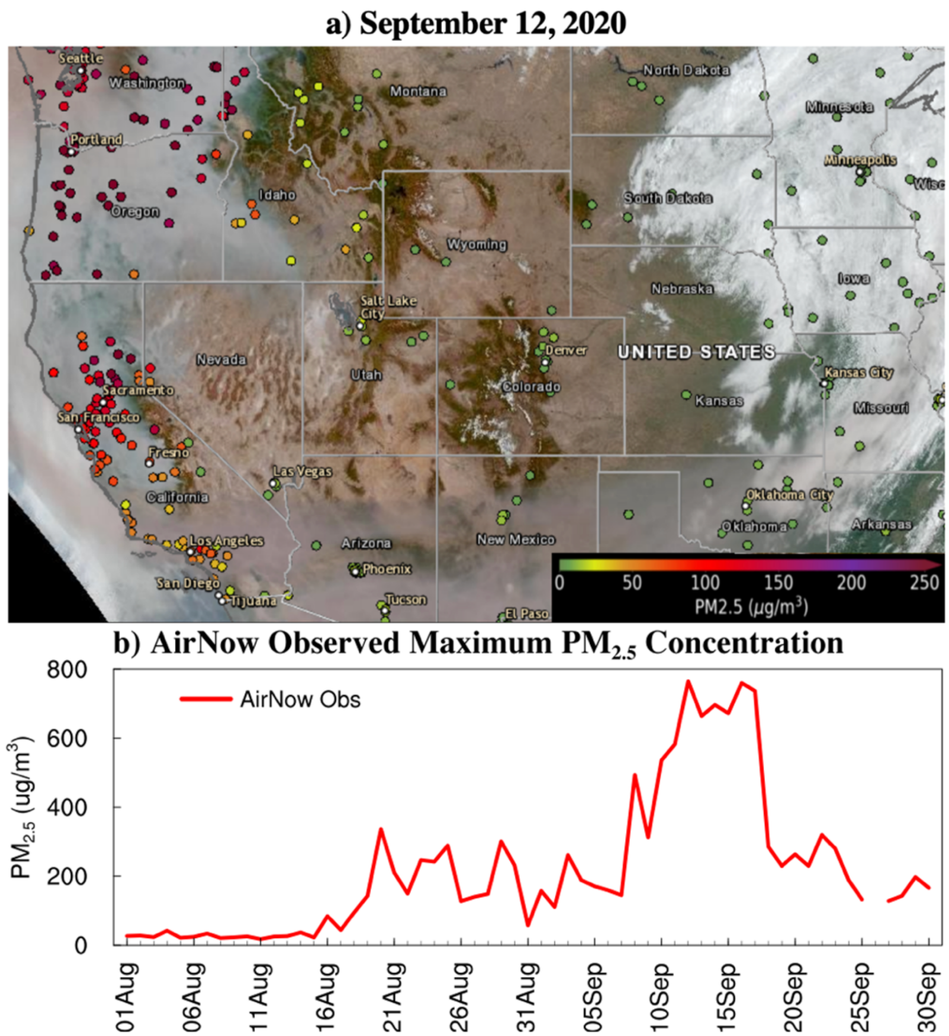


Figure 1. VIIRS-SNPP true color imagery overlaid by $PM_{2.5}$ observations measured by AQS sites (circles) on September 12th, 2020, from NOAA AerosolWatch[1] (above). The time series plot of daily maximum $PM_{2.5}$ concentrations measured by all AirNow sites across the Contiguous United States during the Gigafire events from August to September 2020 (bottom).

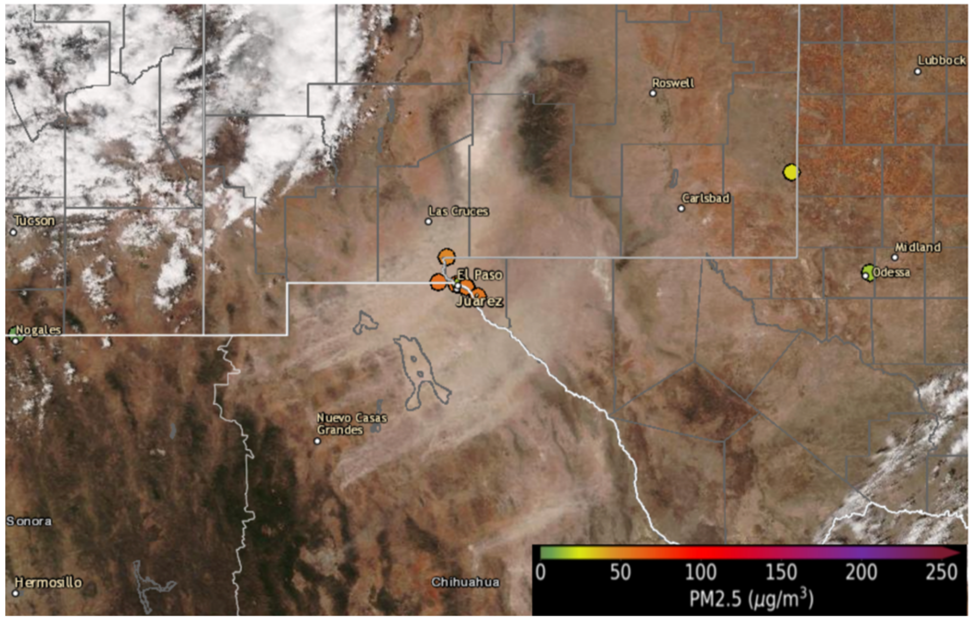
[1] https://www.star.nesdis.noaa.gov/smcd/spb/aq/AerosolWatch/

2.2 The 2021 Spring Dust Season in the Chihuahuan Desert

The spring dust season in the western and southwestern United States commonly occurs from January to March. According to the NOAA's National Weather Service (NWS/NOAA), during Spring 2021, many dust storms occurred primarily over the Chihuahuan Desert from Mexico to across the border in western Texas and southern New Mexico. In the middle of March, from March 13th to 18th, a powerful low-pressure system coupled with drought and the La Niña effect generated gusty winds of 35 to 45 mph across the Mexico-United States border and scattered dust from the Chihuahuan Desert in Mexico to western Texas and southern New Mexico. Consequently, daily average $PM_{2.5}$ concentrations over the active dust regions were substantially increased to 50-60 $\mu g/m^3$, as shown in Figure 2b. On March 16th, a dust storm occurred for nearly eight hours in El Paso, Texas[2] (Figure 2a), which was the most unusual long-lasting dust storm in the city's history and led to worsened air quality and a decreased visibility of less than a half-mile over El Paso and Juãrez, Texas.

[2] https://earthobservatory.nasa.gov/images/148057/long-lasting-dust-storm-from-chihuahua

a) March 16, 2021



b) AirNow Observed Maximum $PM_{2.5}$ Concentration

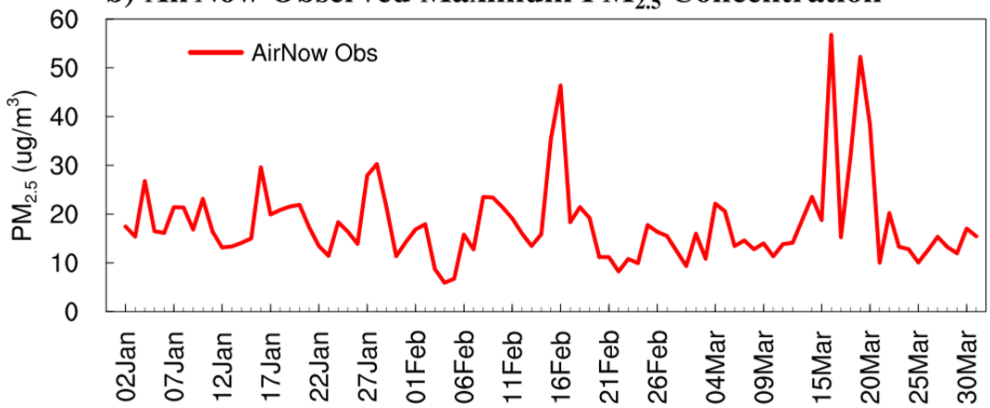


Figure 2. VIIRS/SNPP true color imagery overlaid by $PM_{2.5}$ observations measured by AQS sites (circles) in western Texas and southern New Mexico on March 16th, 2021, from NOAA AerosolWatch (above). The time series plot of daily maximum $PM_{2.5}$ concentrations measured by AirNow sites in the southwestern United States during the 2021 Spring Dust Season (bottom).

CHAPTER 3: DESCRIPTION OF OBSERVATIONS

3.1 Ground-Based Observation

3.1.1 AirNow $PM_{2.5}$

Hourly $PM_{2.5}$ observations were obtained from the U.S. EPA AirNow network[3]. The AirNow data sets are acquired from a variety of monitoring data collected by AirNow and its partners, such as the EPA, NOAA, National Park Service, NASA, Centers for Disease Control, and tribal, state, and local air quality agencies, using a federal reference or equivalent monitoring methods approved by EPA. In this study, hourly $PM_{2.5}$ concentrations derived from each of AirNow sites, starting from 12:00 UTC of the current day to 11:00 UTC the next day, were averaged into a daily value grid by grid.

3.2 Satellite-Based Observations

3.2.1 MAIAC AOD

MAIAC algorithm is designed to work with the time series and spatial analyses of the MODIS L1B data, which are gridded to a fixed 1 km grid resolution to observe the same grid cell over time, resulting in an improvement in the accuracy of aerosol retrievals, atmospheric correction, and cloud detection (Lyapustin et al., 2011a, b; 2012;

[3] https://www.AirNow.gov

2018). In addition to standard MODIS calibration, in Collection 6 and beyond MAIAC applies a residual de-trending of both MODIS Terra and Aqua sensors, along with polarization correction of MODIS Terra and cross-calibration of Terra to Aqua (Lyapustin et al., 2014). This allows MAIAC to process MODIS Terra and Aqua jointly as a single sensor. This study used daily global 1 km MAIAC AOD at 550 nm from all orbits available for the CONUS, later averaged at each grid location. MAIAC data were provided by NASA GSFC.

3.2.2 VIIRS-SNPP AOD

VIIRS-SNPP AOD product was acquired from the VIIRS instrument carried onboard the Suomi National Polar-orbiting Partnership (SNPP), which is a part of the Joint Polar Satellite System (JPSS) (Cao et al., 2013a, 2013b; Uprety et al., 2013). The VIIRS instrument was initially developed based on the previous series of measurements on NOAA satellites and MODIS on the Terra and Aqua satellites (Levy et al., 2013, 2015) through the cooperation of NASA and NOAA. The VIIRS instrument provides improved operational environmental monitoring and sensor data records for aerosol products through a short-wave infrared spanning from 0.412 to 2.25 microns in order to support NASA's Earth Observing System (EOS) and NOAA's polar-orbiting operational environmental satellite system (POES). VIIRS-SNPP observes the entire Earth's surface twice each day. It passes the equator at approximately 13:30 local time (LST). In this study, we used VIIRS-SNPP Level 3 enhanced Dark Target (DT) over dark and bright surfaces (Zhang et al., 2016) daily AOD product at 550 nm with a fixed grid resolution of 0.1°× 0.1° as provided by NOAA, and VIIRS-SNPP Level 2 Deep Blue (DB; Hsu et al., 2013; Hsu et al., 2019; Sayer et al., 2018) 6-minute AOD product at 550 nm with an at-nadir resolution of 6 km x 6 km as provided by NASA GSFC.

CHAPTER 4: MULTI-MODEL ENSEMBLE CREATION

4.1 Description of Ensemble Members

In this section, each of seven participating numerical air quality models included in the ensemble will be described. Model configurations are shown in Table A1 in the Appendix.

4.1.1 GMU-CMAQ

GMU-CMAQ (Li et al., 2021) uses meteorological fields derived from the Weather Research and Forecasting model version 4.2 (WRFv4.2) (Skamarock et al., 2019) to drive the offline CMAQ model version 5.3.1 (CMAQv5.3.1) (US EPA, 2020b), and uses biomass burning (BB) emission data from the Global Biomass Burning Emissions Product (GBBEPx; Zhang et al., 2012, 2014, 2019) blended between Moderate Resolution Imaging Spectroradiometer (MODIS) on the NASA Terra and Aqua satellites and VIIRS-SNPP. The anthropogenic emission data is taken from the U.S. EPA 2016 National Emissions Inventory Collaborative version 1 (2016v1) Emission Modeling Platform, which is generated by the Sparse Matrix Operator Kennel Emissions (SMOKE) model version 4.7 (Houyoux et al., 2000) using the base year of the emission inventory taken from the 2016v1 Emission Modeling Platform (Eyth et al., 2020). The wildfire smoke plumes, and dust plumes are calculated using the Sofiev et al. (2012) and the FENGCHA dust scheme developed by NOAA/ARL (Dong et al., 2016), respectively.

GMU-CMAQ provides hourly experimental AOD and $PM_{2.5}$ concentration forecasts on a horizontal resolution of 12 km ×12 km over the CONUS with each day's forecast initialized at 18:00 UTC on the previous day.

4.1.2 NACC-CMAQ

NACC-CMAQ meteorological preprocessor was adapted from the EPA's Meteorology Interface Processor (MCIP) version 5 (e.g., NACC version 1.3.2; https://zenodo.org/record/5507489#.YmvzsejMKUk, last access 29 Apr 2022), and uses meteorological fields from NOAA's latest operational Finite Volume Cubed-Sphere (FV3) Global Forecast System version 16 (GFSv16) to drive the offline CMAQv5.3.1 (Campbell et al., 2022). Emission input data sets are very similar to GMU-CMAQ and include GBBEPx for BB emissions, NEI 2016v1 for anthropogenic emissions, and Biogenic Emission Inventory System version 3.6.1 (BEISv3.6.1; Vukovich and Pierce, 2002; Schwede, 2005) with the Biogenic Emission Landuse Dataset version 5 (BELD5) for biogenic volatile organic carbon (BVOC) emissions. The wildfire smoke plumes are computed using the Briggs (1969) plume rise algorithm. The dust plumes are computed using dust algorithms including, the FENGCHA dust scheme, SoilGrids soil fractions (Hengl et al., 2017), surface roughness from merged satellite microwave backscattering (ASCAT), and visible/near-infrared reflectances (PARASOL) (Prigent et al., 2012). NACC-CMAQ uses meteorology and emission inputs together with aerosol boundary conditions from NOAA's operational GEFS-Aerosols model for dust and smoke to provide hourly AOD and $PM_{2.5}$ forecasts at a horizontal resolution of 12 km ×12 km (same as GMU-CMAQ) with each day's forecast initialized at 12:00 UTC on of the previous day over CONUS.

4.1.3 HYSPLIT

HYSPLIT (Stein et al., 2015) uses a plume-following coordinate system and back trajectory analysis, and is typically used to determine the emission sources, atmospheric transport, dispersion, deposition, and chemical transformation of aerosols over the CONUS (Draxler & Hess, 1998). Since 2007, it has been employed in NOAA's Smoke Forecasting System using fire locations from satellite data and BB data based on vegetation cover from the bottom-up, fuel-based Blue Sky modeling system developed by the U.S. Forest Service (Rolph et al., 2009; Stein et al., 2009). HYSPLIT has been recently updated to version 5.1.0 (HYSPLITv5.1.0) and combines WRF-ARW (Advanced Research WRF) meteorology inputs, fire emission products from United States Forest Service (USFS) BlueSky, and Briggs (1969) plume rise scheme to simulate hourly AOD and $PM_{2.5}$ concentration forecasts at a horizontal resolution of 0.15° × 0.15° with each day's forecast initialized at 00:00 UTC on of the previous day over CONUS.

4.1.4 GEFS-Aerosols

NOAA's GEFS-Aerosols version 1 model used here provides aerosol and atmospheric composition forecasts using FV3-based GFSv15 meteorology coupled to NASA GOCART aerosol model component using the National Unified Operational Prediction Capability (NUOPC) Layer (Theurich et al., 2016), which is the current and future foundation of NOAA's Unified Forecast System (UFS) modeling framework (Hamill et al., 2011a, b; L. Zhang et al., 2021). The operational GEFS-Aerosols model currently uses BB emission data from GBBEPx, and global anthropogenic emission data from the Community Emission Data System (CEDS) in 2014 for gaseous emissions and Hemisphere Transport of Air Pollution (HTAP) version 2 for primary aerosol emissions. Wildfire smoke plumes are calculated using a one-dimension (1-D) time-dependent cloud module from High-Resolution Rapid Refresh (HRRR)-Smoke model (Freitas et al., 2007). The dust plumes are computed using the FENGCHA dust scheme, SoilGrids soil fractions, surface roughness from merged satellite microwave backscattering (ASCAT), and visible/near-infrared reflectances (Prigent et al., 2012). This study employed GEFS-Aerosols global AOD and $PM_{2.5}$ forecasts at a horizontal resolution of 0.25° × 0.25° and initialized each day at 00:00 UTC.

4.1.5 GEOS-5

GEOS is a global data assimilation and forecasting system that combines the GMAO modified gridpoint statistical interpolation (GSI) analysis algorithm, which was originally developed by the National Centers for Environment Prediction (NCEP) Environmental Modeling Center, with the NASA atmospheric global forecast model (Rienecker et al., 2008). The GEOS version 5.27.1 (GEOSv5.27.1) is integrated using the Earth System Modeling Framework (ESMF), and its configuration includes meteorological data acquired by the GEOS Data Assimilation System (DAS) in near real time, fire detection information from MODIS, emissions of aerosols, BB, and, smoke data from the Quick Fire Emissions Dataset (QFED), anthropogenic emissions from the Emissions Database for Global Atmospheric Research (EDGAR)-HTAP inventories, and Model of Emissions of Gases and Aerosols from Nature (MEGAN) for BVOC emissions.

The dust plumes were computed using the Goddard Global Ozone Chemistry Aerosol Radiation and Transport (GOCART; Colarco et al., 2010) model. This study used GEOS-5 global forecast of hourly AOD values and $PM_{2.5}$ concentrations on a horizontal resolution of 0.25° × 0.3125° and initialized each day at 00:00 UTC.

4.1.6 ICAP-MME

Established in 2010, ICAP aims to promote community development of global aerosol observations, data assimilation, and prediction technologies to support operational aerosol forecasting (Benedetti et al., 2011; Colarco et al., 2014a; Reid et al., 2011). The ICAP-MME (Sessions et al., 2015; Xian et al., 2019) is a global multi-model aerosol forecasting ensemble consensus (currently only AOD product is available), which provides a testbed of probabilistic aerosol forecasts. ICAP-MME is generated by combining nine global aerosol models: the European Centre for Medium-range Weather Forecasts-Monitoring Atmospheric Composition and Climate model (ECMWF) under Copernicus Atmosphere Monitoring Service (CAMS, former MACC), GEOS, NAAPS, Japan Meteorological Agency (JMA) Model of Aerosol Species in the Global Atmosphere (MASINGAR), NOAA Environmental Modeling System (NEMS) Global Forecast System (GFS) Aerosol Component (NGAC), Mětěo-France Modělě de Chimie Atmospherique ã Grande Echelle (MOCAGE), and Finnish Meteorological Institute (FMI) System for Integrated modeLling of Atmospheric coMposition (SILAM), the Barcelona Supercomputing Center (BSC) Chemical Transport Model (CTM), embedded in the Multiscale Online Nonhydrostatic AtmospheRe CHemistry (MONARCH) and the UK Met Office (UKMO) models. These models have different underlying meteorological fields, emissions, microphysics, and chemistry, as well as a variety of horizontal and vertical resolutions ranging from 0.25°×0.31° and 72 vertical layers to 1.4°×1° and 24 layers. As a result, ICAP-MME is driven by the independent operation/quasi-operational meteorology inputs and aerosol variables generated by each of the member organizations. This study utilized ICAP-MME global 6-hour AOD at 550 nm on a horizontal resolution of 1° × 1° and initialized each day at 00:00 UTC.

4.1.7 NAAPS

NAAPS is developed at the Marine Meteorology Division of the NRL and provides an operational forecast of 3D atmospheric anthropogenic fine and biogenic fine aerosols, biomass burning smoke, dust, and sea salt concentrations (Lynch et al., 2016). The current NAAPS is driven by global meteorological fields from the NAVy Global Environmental Model (NAVGEM), which is an operational global weather prediction system developed by the United States Navy (Hogan et al., 2014). NAAPS uses BB smoke source from the Fire Locating and Modeling of Burning Emissions (FLAMBE) inventory, which is based on near-real time MODIS fire hotspot data (Reid et al., 2009). Dust emissions for the NAAPS model were generated with the methods documented in Westphal et al. (1988), while the dust scheme algorithms were based on Westphal et al. (2009). This study employed the NAAPS global 3-hourly AOD and surface $PM_{2.5}$ concentrations at a horizontal resolution of 0.333° × 0.333° and initialized each day at 00:00 UTC.

4.2 Multi-Model Ensemble Forecasts

The ensemble forecasts were created using the mean AOD and $PM_{2.5}$ generated by the individual models. All data were interpolated to a unified horizontal grid of 12 km × 12 km before calculating the ensemble mean values. It should be noted that the individual model and the ensemble simulations were evaluated with the observations grid by grid, and any grids containing missing data were ignored from the calculation.

4.2.1 Ensemble AOD Forecasts

For the 2020 Gigafire events, the ensemble forecasts were reproduced from August to September 2020 using the mean values of AOD generated by seven models: GMU-CMAQ, NACC-CMAQ, HYSPLIT, ICAP-MME, GEFS-Aerosols, GEOS-5, and NAAPS models. The ensemble AOD forecasts were simulated near VIIRS equatorial crossing time (13:30 LST) and mean value of the average of AOD near equatorial crossing times of MODIS Terra (10:30 LST) and Aqua (13:30 LST) as MAIAC processes MODIS Terra and Aqua jointly as a single sensor. The performance of the ensemble AOD mean near VIIRS passing time was evaluated against the VIIRS AOD. While the performance of the ensemble mean of average AOD near MODIS Terra and Aqua passing time was evaluated against MAIAC AOD retrievals.

For the 2021 Spring Dust Season, the ensemble forecasts of AOD were reproduced from January to March 2021 using the mean values of AOD from six models: GMU-CMAQ, NACC-CMAQ, ICAP-MME, GEFS-Aerosols, GEOS-5, and NAAPS models. The ensemble AOD was simulated using the mean value of AOD near VIIRS-

SNPP equatorial crossing time (13:30 LST). The performance of the ensemble AOD mean near VIIRS-SNPP passing time was evaluated against the VIIRS enhanced DT AOD and VIIRS DB AOD.

4.2.2 Ensemble $PM_{2.5}$ Forecasts

The ensemble $PM_{2.5}$ forecasts during the 2020 Gigafire events and the 2021 Spring Dust Season were simulated using the mean values of $PM_{2.5}$ concentrations generated by six models (for wildfire case), and five models (for dust storm case): GMU-CMAQ, NACC-CMAQ, HYSPLIT (only wildfire), GEFS-Aerosols, GEOS-5, and NAAPS models. The performance of ensemble mean in forecasting $PM_{2.5}$ concentrations for both cases was verified by comparing model simulations against daily average $PM_{2.5}$ observations from AirNow with the evaluation time starting from 12:00 UTC to 11:00 UTC of the next day.

4.3 Ensemble Probability of $PM_{2.5}$ Exceedance Forecast

The GMU-CMAQ, NACC-CMAQ, HYSPLIT (only wildfire), GEFS-Aerosol, GEOS-5, and NAAPS were used to create the ensemble probability of the $PM_{2.5}$ exceedance forecast. The probability was calculated using equation (1) based on the numbers of models that forecast $PM_{2.5}$ exceedances (concentrations >35 $\mu g/m^3$) during the 2020 Gigafire events and the 2021 Spring Dust Season. The probability result ranges from 0% (none of the models forecast the exceedances; very unlikely to occur) to 100% (all models forecast the exceedances; very likely to occur):

Equation 1 Ensemble Probability of Exceedance Forecast

$$P(A) = \frac{Number\ of\ models\ that\ forecast\ the\ exceedances}{Total\ number\ of\ models} \times 100\%$$

CHAPTER 5: EVALUATION METHODOLOGY

5.1 Statistical Metrics

The AOD and surface $PM_{2.5}$ concentrations simulated by the ensemble mean and individual models were evaluated with AOD retrievals from VIIRS enhanced DT, VIIRS DB (for the 2021 Spring Dust Season) and MAIAC (for the 2020 Gigafire events) and observed surface $PM_{2.5}$ from the AirNow ground monitoring network. A suite of statistical metrics, including root mean square error (*RMSE*), correlation (*CORR*), absolute fractional bias (*FB*), mean bias (*MB*), mean error (*ME*), normalized mean bias (*NMB*), and normalized mean error (*NME*) were calculated using the following formulas:

Equation 2 Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(M_i - O_i)^2}$$

Equation 3 Correlation

$$CORR = \frac{N\sum_{i=0}^{N} O_i M_i - \sum_{i=0}^{N} O_i \sum_{i=0}^{N} M_i}{\sqrt{N\sum_{i=0}^{N} O_i^2 - (\sum_{i=0}^{N} O_i)^2}\sqrt{N\sum_{i=0}^{N} M_i^2 - (\sum_{i=0}^{N} M_i)^2}}$$

Equation 4 Absolute Fractional Bias

$$FB = 2 \times \frac{\sum_{i=0}^{N}|O_i - M_i|}{\sum_{i=0}^{N}|O_i + M_i|}$$

Equation 5 Mean Bias

$$MB = \frac{1}{N}\sum_{i=0}^{N}(M_i - O_i)$$

Equation 6 Mean Error

$$ME = \frac{1}{N}\sum_{i=0}^{N}|M_i - O_i|$$

Equation 7 Normalized Mean Bias

$$NMB = \frac{\sum_{i=0}^{N}(M_i - O_i)}{\sum_{i=0}^{N} O_i}$$

Equation 8 Normalized Mean Error

$$NME = \frac{\sum_{i=0}^{N}|M_i - O_i|}{\sum_{i=0}^{N} O_i}$$

Where M_i represents the *i*th model forecast, O_i is the *i*th observation, and N is the total number of observations and time-space matched prediction during the study periods.

5.2 Categorical Metrics

In addition to a traditional suite of statistical metrics, we employed two categorical metrics: the area hit rate (*aH*), and the area false alarm ratio (*aFAR*) (Kang et al., 2007) to supplementarily measure the performance of individual models, ensemble mean, and ensemble probability in forecasting $PM_{2.5}$ exceedances (24-hr $PM_{2.5}$ concentrations greater than 35 μg/m^3 based on NAAQS). These two metrics were calculated based on pairs of observed and predicted $PM_{2.5}$ exceedances by considering four possible scenarios: (a) a forecasted exceedance that is not observed; (b) a forecasted exceedance that is observed; (c) an exceedance that is neither forecasted nor observed; (d) an observed exceedance that is not forecasted (Figure 3). The *aH* and *aFAR* values are determined by matching observed and forecasted exceedances within a designated area surrounding the center of the observation location. In the present study, we used an area of 0.5° × 0.5° centered at each AirNow site' location. The area hit rate *aH* refers to the number of hits if a forecasted exceedance is observed within the designated area (Eq.9). The *aFAR* (Eq.10) refers to the false-alarm ratio if a forecasted exceedance is not observed within the designated area:

Equation 9 area Hit rate

$$aH = \left(\frac{Ab}{Ab + Ad}\right) \times 100\%$$

Equation 10 area False Alarm Ratio

$$aH = \left(\frac{Aa}{Aa + Ab}\right) \times 100\%$$

Where *Aa* is the number of forecasted exceedances that are not observed, *Ab* is the number of forecasted exceedances that are observed, and *Ad* is the number of observed exceedances that were not forecasted.

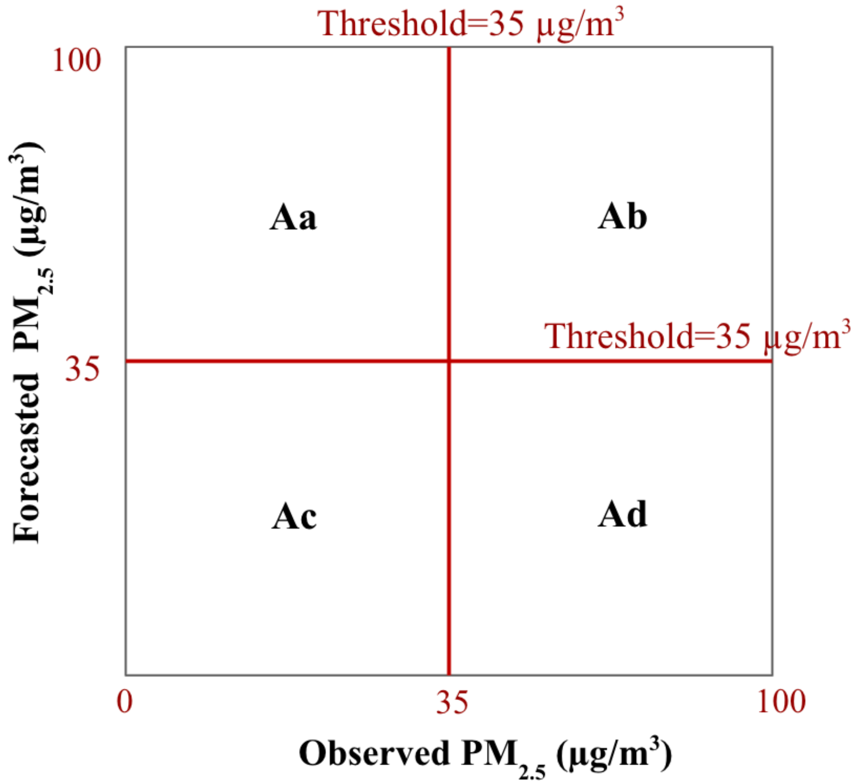


Figure 3. Example of scatter plot for definition of categorical metrics

5.3 Overall Rating

The overall rating (*RANK*) was used to determine the comprehensive forecasting performances of individual models and ensemble mean during the study periods. In the case of $PM_{2.5}$ evaluation, the *RANK* was derived from the sum of the normalized *CORR*, *FB*, *aH*, and *aFAR* (Eq.11). In the case of AOD evaluation, the *RANK* was calculated using the sum of the normalized *CORR* and *FB* (Eq.12). $PM_{2.5}$ *RANK* ranges from 0 to 4 (from worst to best), while AOD *RANK* ranges from 0 to 2:

Equation 11 Overall Rating for $PM_{2.5}$

$$RANK_{PM2.5} = \frac{CORR + 1}{2} + \left(1 - \frac{FB}{2}\right) + \frac{aH}{100\%} + \left(1 - \frac{aFAR}{100\%}\right)$$

Equation 12 Overall Rating for AOD

$$RANK_{AOD} = \frac{CORR + 1}{2} + \left(1 - \frac{FB}{2}\right)$$

CHAPTER 6: EVALUATION OF ENSEMBLE FORECAST OF WILDFIRES

In this chapter, an ensemble mean based on the unweighted arithmetic mean of individual models, including GMU-CMAQ, NACC-CMAQ, GEFS-Aerosols, GEOS-5, HYSPLIT, ICAP-MME, and NAAPS, was created and evaluated with satellite and ground observations during the 2020 Gigafire events (August-September 2020). The forecasting performance of the ensemble mean was also compared with ensemble members to assess whether the ensemble mean can outperform the top performers among these members. Evaluation results of AOD simulations validated against VIIRS enhanced DT AOD and MAIAC AOD retrievals, and the evaluation results of $PM_{2.5}$ simulations compared against AirNow $PM_{2.5}$ observations were analyzed by calculating average statistical metrics and the overall rating (*RANK*) for AOD and $PM_{2.5}$ simulations over the study period.

6.1 Ensemble Performance in Forecasting AOD

The ensemble mean shows fairly good performance in simulating AOD. For instance, contour maps of AOD forecasts, VIIRS AOD and MAIAC AOD retrievals on August 22nd, 2020, in Figures 4a-4g and 4a-4g indicates that the AOD simulations from all the Model-1 to 7 underestimated AOD values over the western U.S. while Model-5 overestimated AOD values primarily in California (Figures 4c, 4e, and Figures 5c, 5e). In comparison, the ensemble mean AOD simulations slightly overestimated AOD values for the most parts of Northern California and underestimated AOD values predominantly over Montana, Wyoming, Colorado, Nebraska, and Kansas where complex geographic formations, such as the Colorado Plateau-Central Rockies areas, are located (Figures 4h and 5h). However, the majority of areas showing high AOD values in the ensemble forecast match fairly well with the observations.

As shown in Figures 7c and 8c in the time series plots of mean bias (*MB*), the ensemble mean and most individual models slightly underestimated AOD values almost the entire period, especially during the extreme fires in the middle of September 2020 leading to relatively high negative *MB* values during this time. Similar to previous results, Model-5 overestimated the AOD values with relatively high positive *MB* values during the same period. From Table 1, the average *MB* values of the ensemble mean for AOD over the whole period was reduced to -0.104 for the VIIRS AOD case and was greatly reduced to -0.068 for the MAIAC AOD case. Both values are closer to zero relative to most individual models, meaning that the ensemble mean significantly reduces bias and uncertainties in AOD forecasting. In addition, the time series plots of correlation (*CORR*) (Figures 7a and 8a) display less fluctuated correlation lines of the ensemble mean in relation to most individual models, resulting in being the best in correlation for the entire period in the VIIRS and MAIAC AOD cases shown in Table 1.

Considering the overall rating (*RANK)* (Table 1), the ensemble mean scores second in *RANK* with the *RANK* values being only 4.8% and 4.0% lower than the top rank model based on VIIRS AOD and MAIAC AOD retrievals, respectively. Although the averages of *RMSE*, *NMB*, *NME*, *ME*, and *FB* of the ensemble mean did not outperform the first rank individual model, they still rank either second or third place. All the results point to the ensemble mean having a beneficial effect in reducing the bias in AOD forecasting, especially when the wildfires are extremely intense. Furthermore, the ensemble mean successfully produces more statistically consistent and reliable forecasts of AOD during the wildfires relative to the forecasts provided by individual models, which are particularly degraded by errors in emission inventory and smoke plume algorithms implemented to each model.

6.2 Ensemble Performance in Forecasting Surface $PM_{2.5}$ Concentration

The ensemble mean show fairly well forecasting performance of surface $PM_{2.5}$ during extreme wildfires, such as the $PM_{2.5}$ forecasts on August 22nd, 2020 (Figure 6). Figures 6a, 6c, 6d, 6e, and 6f show that Model-1, 3, 4, 5, and 7 overestimated $PM_{2.5}$ concentrations largely in the western U.S. and partially in the Central and southern United States. In contrast, the ensemble mean predominantly shows overestimated $PM_{2.5}$ simulations in Northern California (Figure 6g). However, the extremely high $PM_{2.5}$ concentrations simulated by the ensemble mean are located over the areas that are in fairly good agreement with the AirNow ground observations (Figure 6h).

The positive mean bias (*MB*) values of the ensemble mean and the individual models in Figure 9c indicate the overestimation of $PM_{2.5}$ simulations for most of the time during the wildfire period, except for Model-2 and Model-3, which show negative *MB* values (underestimation of $PM_{2.5}$ concentrations). The discrepancies in $PM_{2.5}$ simulations during the wildfires were also indicated by varying values of root mean square (*RMSE*), mean error (*ME)*, normalized mean bias (*NMB*), normalized mean error (*NME*), and absolute fractional bias (*FB*), as shown in Figures 9b, 9d, 9e, 9f, and 9g. Overall, the ensemble mean reduced the positive average mean bias (*MB*) to 7.4 and lowered the absolute fractional bias (*FB*) to 0.54, which is the top rank in *FB*. It also yields the highest average correlation (*CORR*) value of 0.603 (Table 1) due to consistent correlation values for the entire period compared to that of individual models (Figure 9a). The *RMSE*, *NMB*, and *NME* values of the ensemble mean are on average lower than those values of most individual models, as shown in Table 1.

Analysis of the forecasting performance of daily $PM_{2.5}$ exceedances (concentrations >35 μg/m^3) showed the ensemble mean substantially increased the area hit rate (*aH*), particularly in the middle of September when the extremely intense wildfires occurred (Figure 9g). As a result, the ensemble mean achieves the highest average *aH* value of 86.845% (Table 1). This suggests that the ensemble mean can predict more than 86% of the observed $PM_{2.5}$ exceedances during extreme wildfires. Due to relatively high correlation, high aH, low *aFAR*, and low *FB* values, the ensemble mean performs highly in *RANK* (2.825). These results suggest that the ensemble forecast has a practical advantage in reducing bias in individual forecasts of $PM_{2.5}$ and allowing effective probabilistic forecasts of $PM_{2.5}$. Furthermore, the evaluation results revealed that although a single model can be excellent at predicting AOD, it is not necessarily translated into good performance in surface $PM_{2.5}$ prediction. The model that performs highly in *RANK* for the AOD prediction is different from that of the $PM_{2.5}$ prediction.

6.3 Overall Performance of Ensemble Wildfire Forecast

Figures 10a-10c show the time series of the overall rating (*RANK*) for AOD and $PM_{2.5}$ predicted by the ensemble mean and individual models compared against three observation datasets: VIIRS and MAIAC AOD, and AirNow surface $PM_{2.5}$ concentrations, respectively. The ensemble mean shows persistently high *RANK* values throughout the study period, suggesting that the ensemble forecast overall is more reliable and performs better than most of the members. In addition, it can partially reduce the bias as shown in Table 1 due to the fact that the ensemble mean is calculated by averaging each of the individual model simulation results. Therefore, if most individual models underestimated (negative bias) or overestimated (positive bias) the AOD values and $PM_{2.5}$ concentrations for almost the entire period, the bias values of the ensemble mean become more negative or positive than the top-ranked model with the lowest bias. As a result, the ensemble forecast will show significant and effective improvements in forecasting if there are complementary underestimation and overestimation by individual models.

Underestimation of AOD values and overestimation of $PM_{2.5}$ concentrations in the model simulations may have occurred since the August 2020 complex wildfires that became much more intense during the middle of September 2020, generating very thick smoke cover. The smoke could in turn make the biomass burning emissions applied to each model inaccurate and may generate a large error in smoke inventories. Furthermore, as the fire becomes stronger, the plume injection height gets deeper and creates misrepresented vertical emissions within the planetary boundary layer (PBL) generated by each individual model. These two factors are considered important sources of uncertainties in air quality forecasts during wildfire events (Carter et al., 2020; Pan et al., 2020b; and Ye et al., 2021). The impact of thick wildfire smoke also challenges the use of satellite AOD retrievals for evaluating ensemble forecasts since retrievals over heavy smoke plumes may be masked as clouds and vertical distributions of smoke are difficult to measure, which affects the accuracy of retrieval AOD products. In addition, a variety of input data sets, such as meteorological fields and chemical transports (F. Li et al., 2019; Y. Li et al., 2020) and plume rise schemes (Briggs, 1969; Freitas et al., 2007; Paugam et al., 2016; Sofiev et al., 2012; Stein et al., 2009; Vernon et al., 2018; Zhu et al., 2018), implemented differently in each model and can also impact the AOD and $PM_{2.5}$ forecasting performance (Delle Monache and Stull, 2003; Kumar et al., 2020).

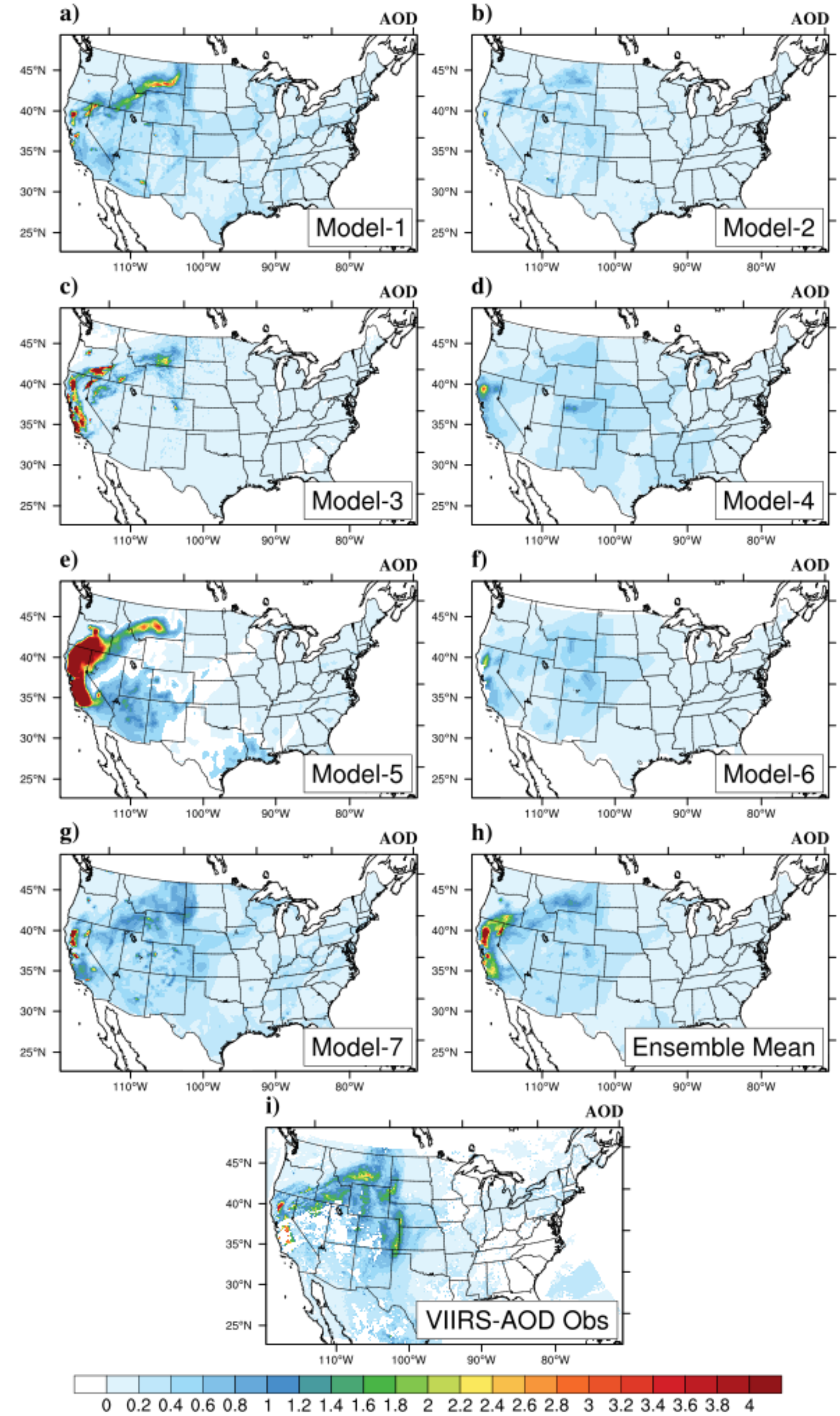


Figure 4. AOD predicted by seven individual models (a-g) and the ensemble mean (h), compared with VIIRS enhanced Dark Target (DT) AOD retrievals (i) on August 22, 2020 (during the 2020 Gigafire events).

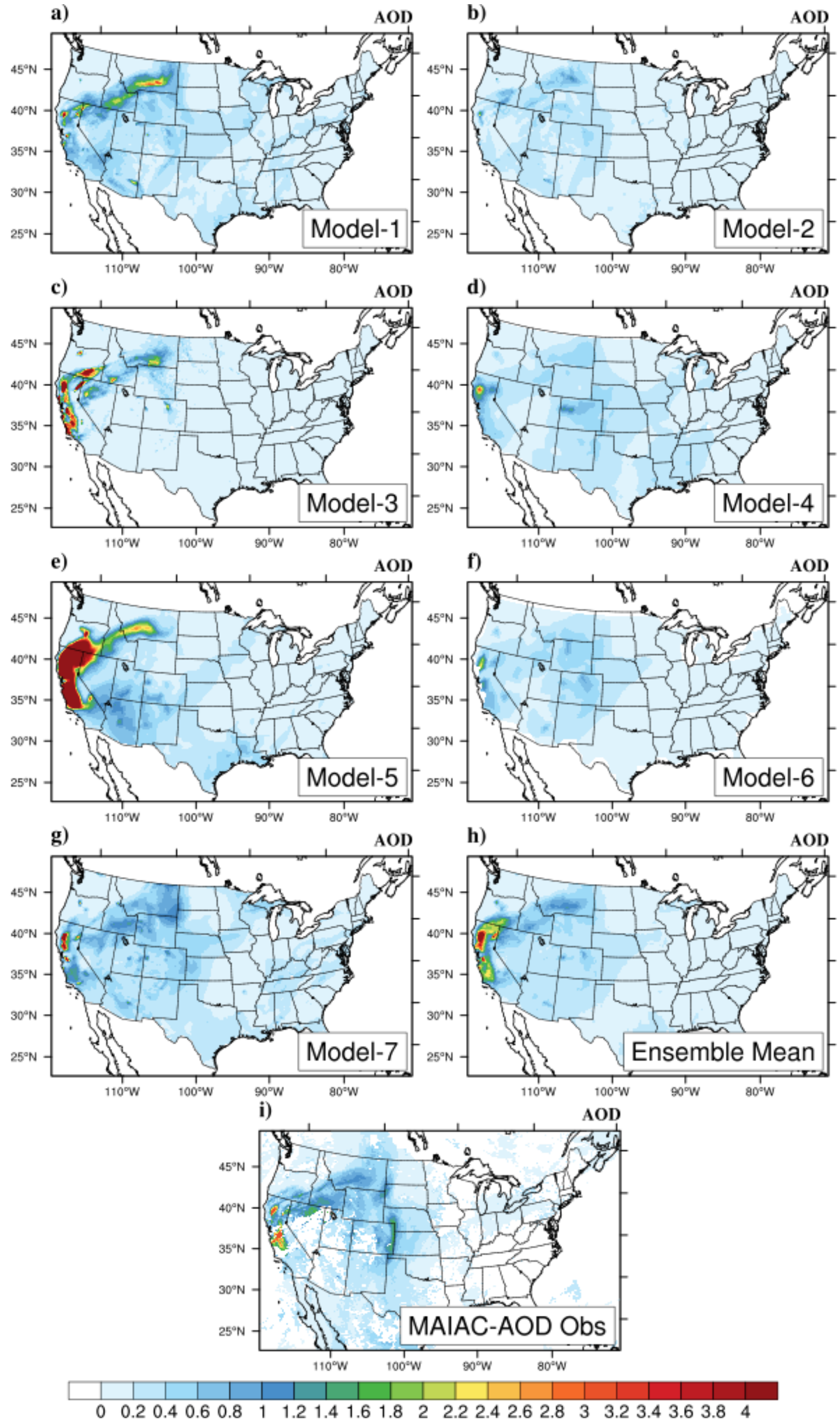


Figure 5. AOD predicted by seven individual models (a-g) and the ensemble mean (h), compared with MAIAC AOD retrievals (i) on August 22, 2020 (during the 2020 Gigafire events).

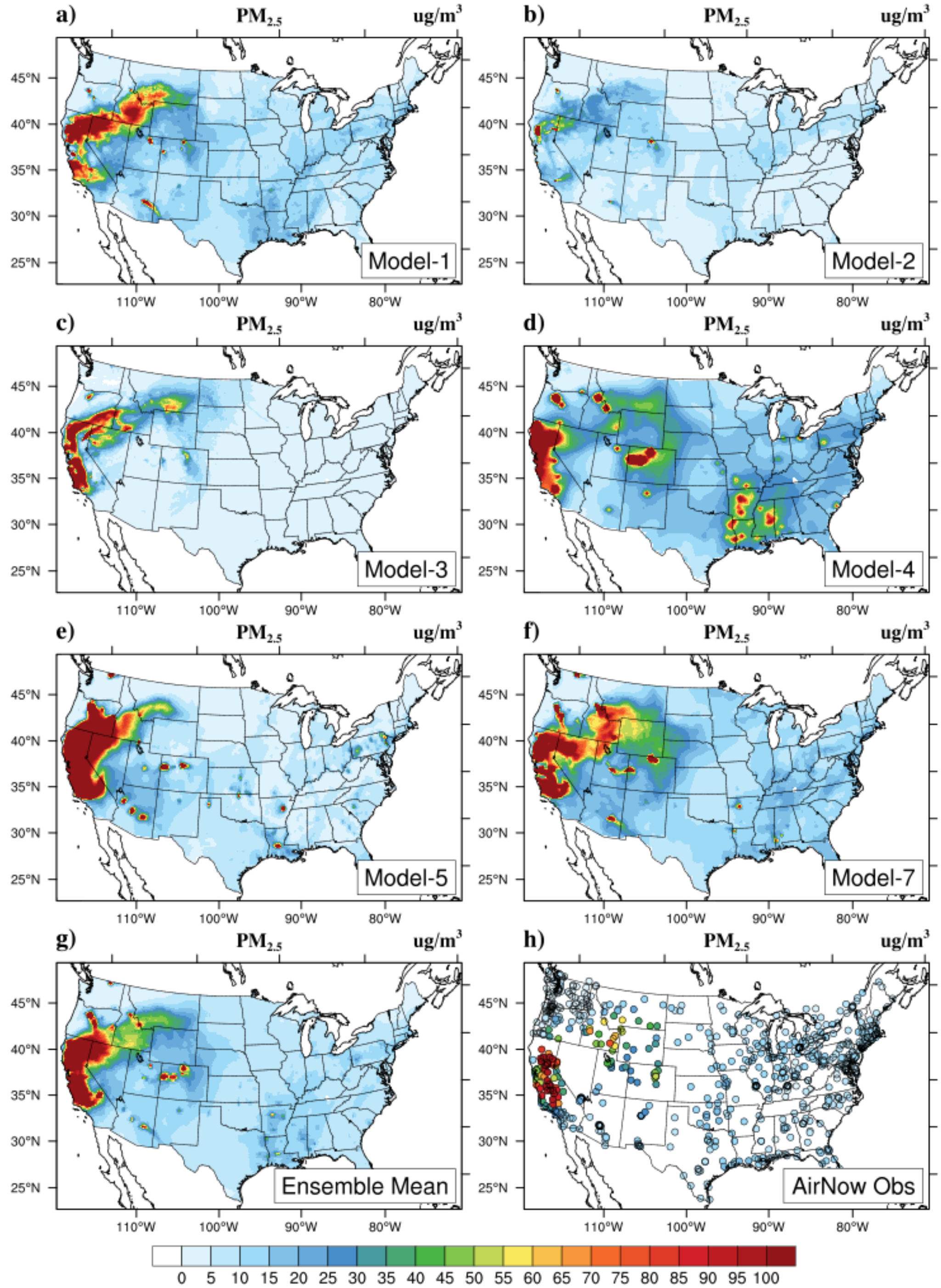


Figure 6. Surface $PM_{2.5}$ concentrations predicted by six individual models (a-f) and the ensemble mean (g), compared with AirNow $PM_{2.5}$ observations (h) on August 22, 2020 (during the 2020 Gigafire events).

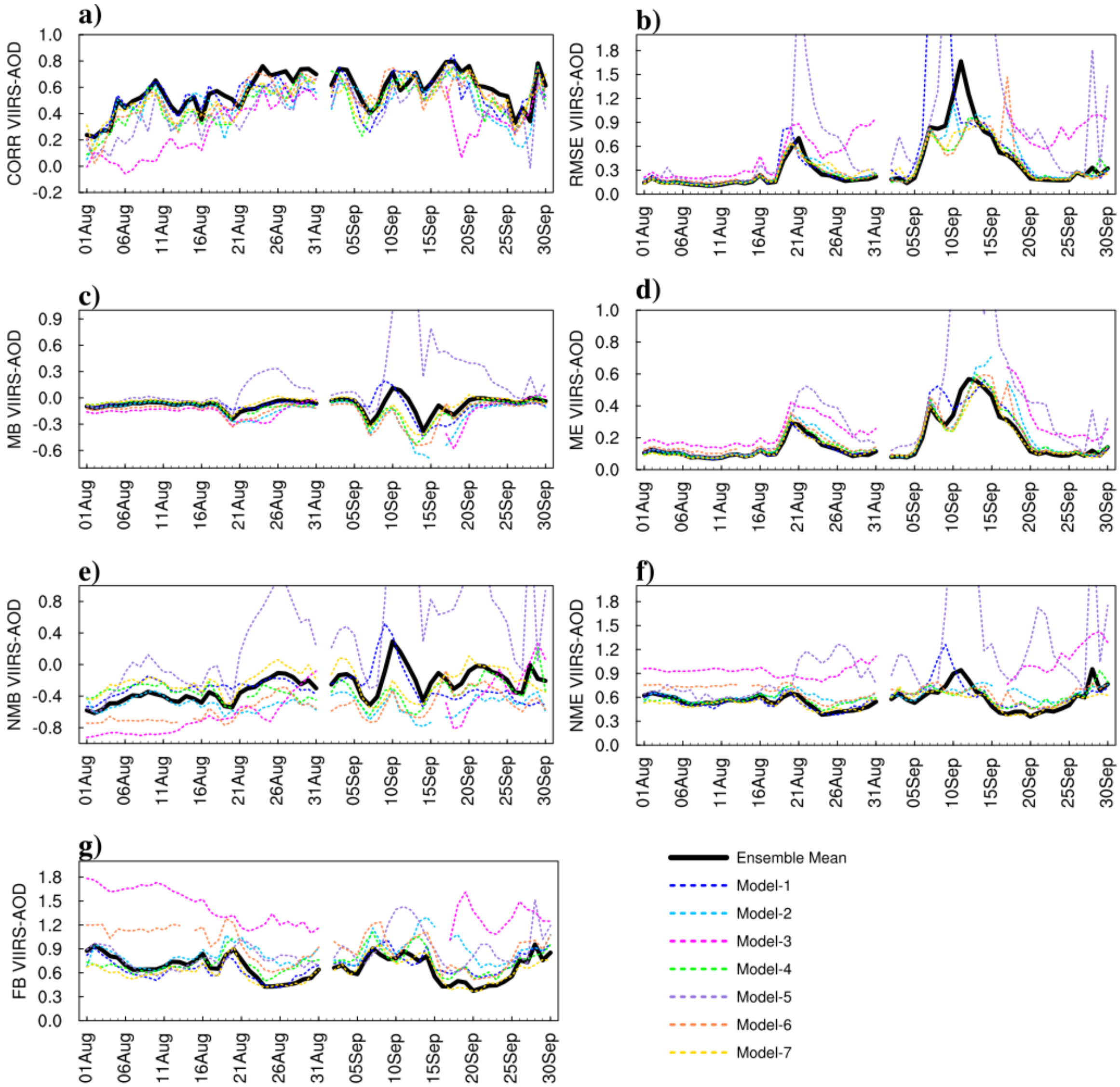


Figure 7. Time series of RMSE (a), *CORR* (b), *MB* (c), *ME* (d), *NMB* (e), *NME* (f), and *FB* (g) of AOD for the 2020 Gigafire events during August-September 2020. The AOD simulations by the ensemble mean (solid black line) and individual models (dash lines): Model-1 (blue), Model-2 (light blue), Model-3 (pink), Model-4 (green), Model-5 (purple), Model-6 (orange), and Model-7 (yellow) were compared against VIIRS AOD retrievals.

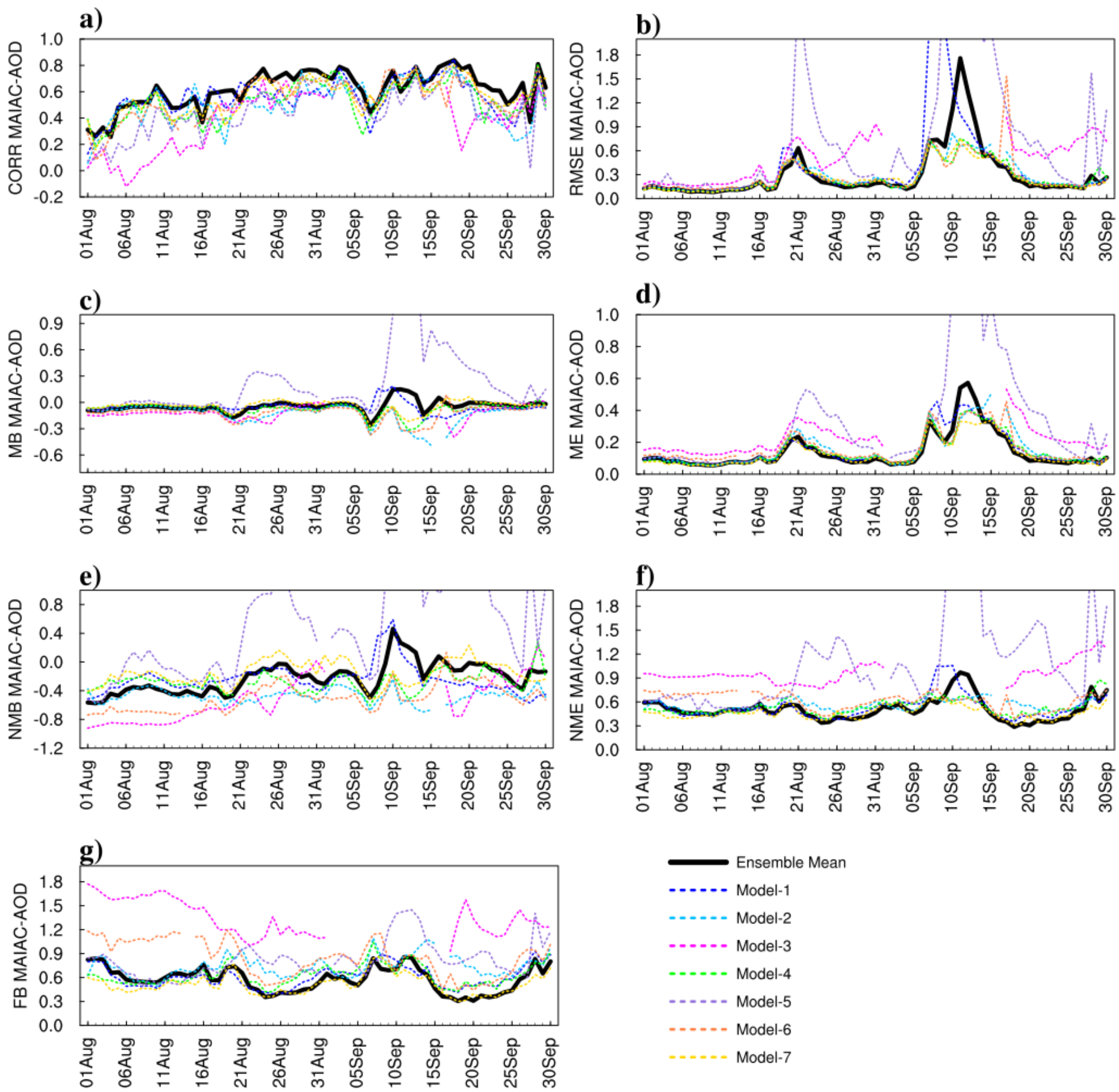


Figure 8. Time series of *RMSE* (a)*,* *CORR* (b)*,* *MB* (c)*,* *ME* (d)*,* *NMB* (e)*,* *NME* (f)*,* and *FB* (g) of AOD for the 2020 Gigafire events during August-September 2020. The AOD simulations by the ensemble mean (solid black line) and individual models (dash lines): Model-1 (blue), Model-2 (light blue), Model-3 (pink), Model-4 (green), Model-5 (purple), Model-6 (orange), and Model-7 (yellow) were compared against the MAIAC AOD retrievals.

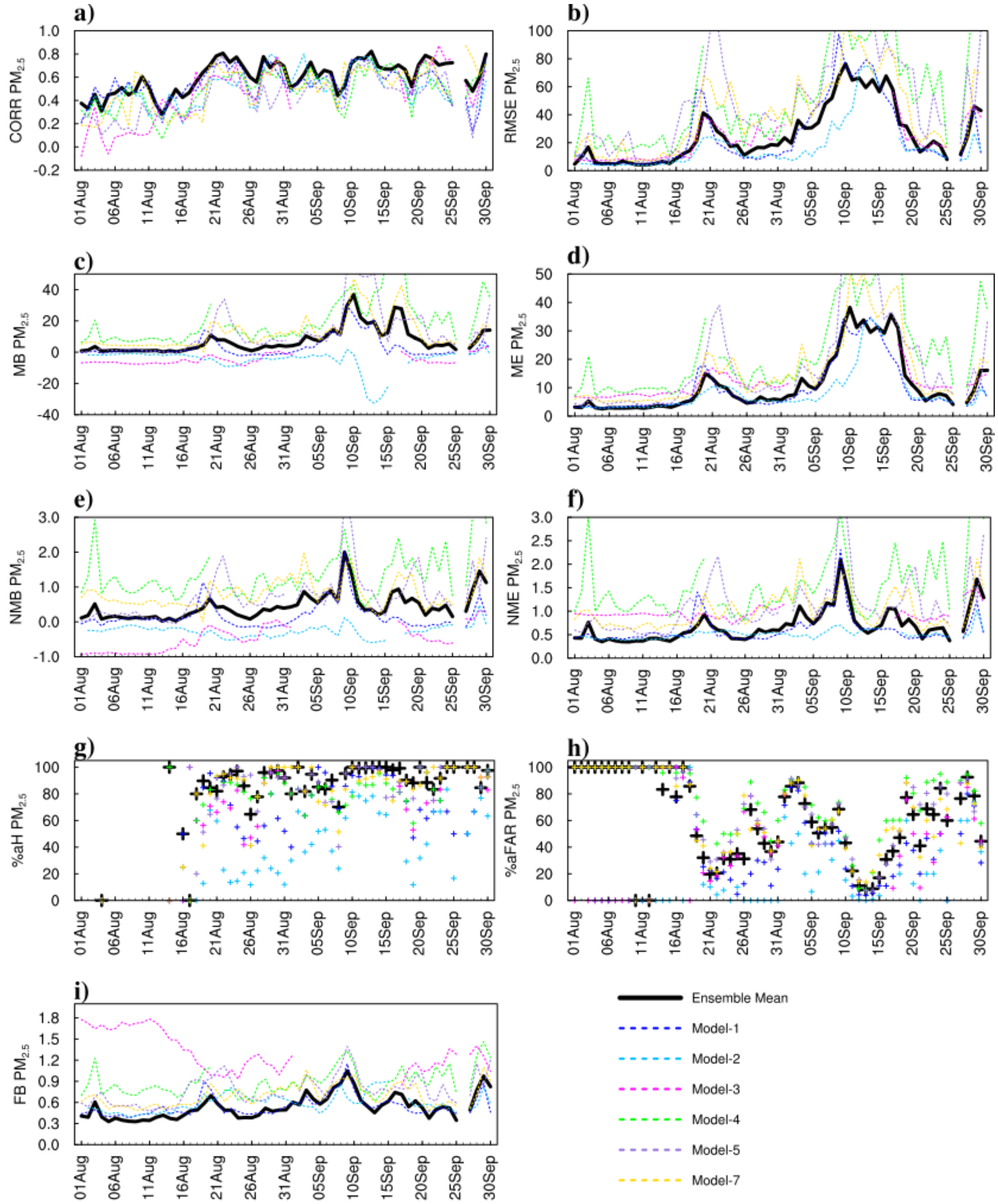


Figure 9. Time series of *RMSE* (a), *CORR* (b), *MB* (c), *ME* (d), *NMB* (e), *NME* (f), *aH* (g), *aFAR* (h), and *FB* (i) of $PM_{2.5}$ for the 2020 Gigafire events during August-September 2020. The $PM_{2.5}$ simulations by the ensemble mean (black solid line) and individual Model-1 (blue), Model-2 (light blue), Model-3 (pink), Model-4 (green), Model-5 (purple), and Model-7 (yellow) were compared against AirNow $PM_{2.5}$ observations.

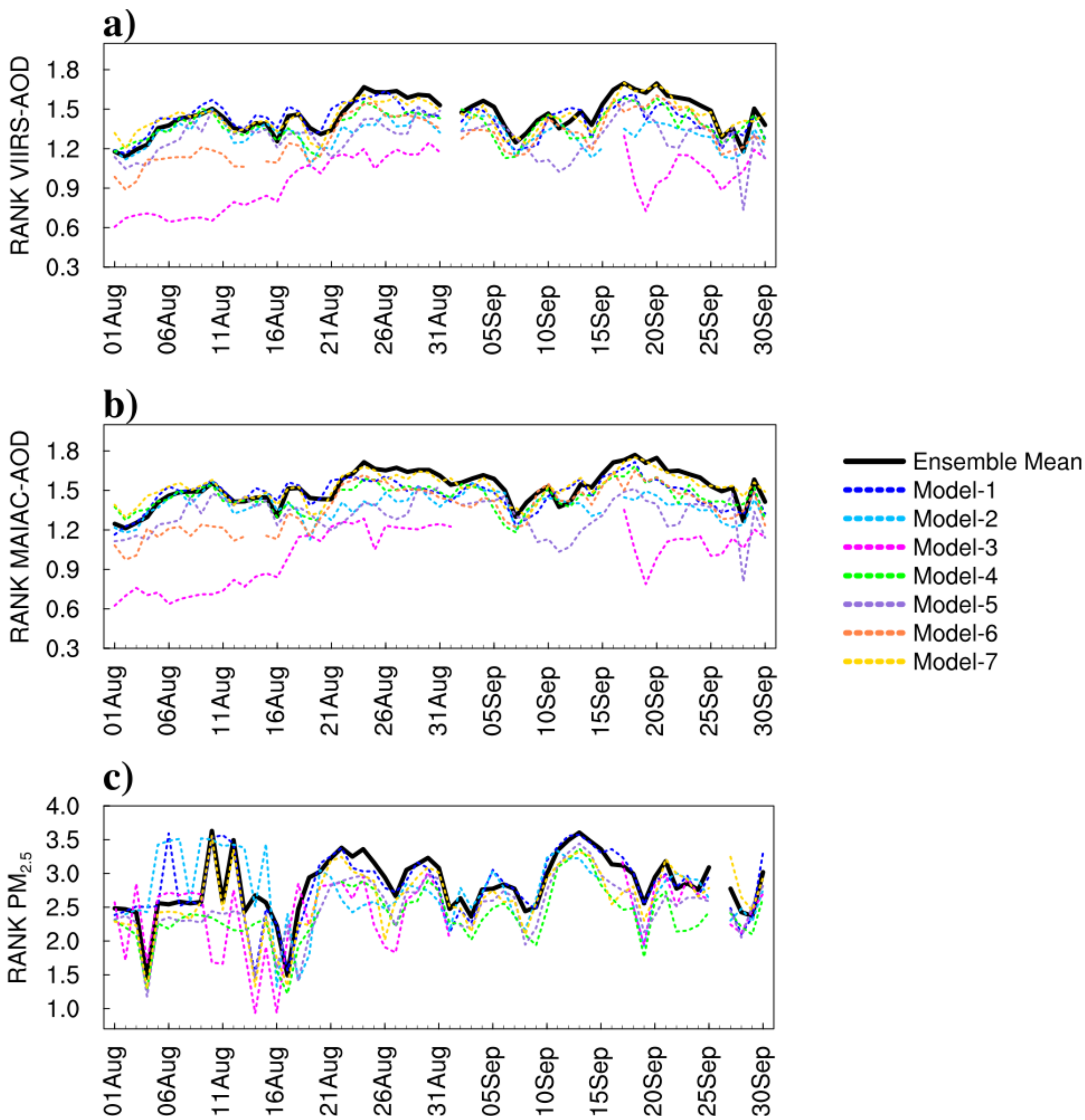


Figure 10. Time series of the overall rating (*RANK*) for AOD and $PM_{2.5}$ simulated by the ensemble mean and individual models. The *RANK* is calculated with four statistical metrics by comparing model predictions against AOD retrievals from VIIRS (a) and MAIAC (b), and surface $PM_{2.5}$ observations from AirNow (c) during the 2020 Gigafire events (August-September 2020).

Table 1. Overall ensemble mean and individual model performances in forecasting AOD values and $PM_{2.5}$ concentrations during the 2020 Gigafire events (August-September 2020) based on the evaluation of the values of *RMSE, CORR, NMB, NME, MB, ME, FB aH and aFAR,* and overall rating (*RANK*). The best results of each statistical metric and *RANK* are highlighted in bold.

Cases	Models	*RMSE*	*CORR*	*NMB*	*NME*	*MB*	*ME*	*FB*	*aH*(%)	*aFAR*(%)	*RANK*
AOD simulations compared against VIIRS retrievals	Model-1	0.479	0.533	-0.281	0.587	-0.085	0.186	0.682			1.425
	Model-2	0.360	0.477	-0.512	0.626	-0.172	0.199	0.852			1.313
	Model-3	0.551	0.296	-0.565	0.979	-0.153	0.249	1.398			0.949
	Model-4	**0.324**	0.480	-0.339	0.597	-0.116	0.183	0.726			1.377
	Model-5	1.076	0.429	0.458	1.100	0.201	0.387	0.852			1.289
	Model-6	0.352	0.516	-0.544	0.660	-0.164	0.203	0.931			1.293
	Model-7	0.332	0.506	**-0.219**	**0.540**	-0.081	**0.167**	**0.617**			1.444
	Ensemble Mean	0.355	**0.569**	-0.278	0.566	**-0.079**	0.175	0.664			**1.452**
AOD simulations compared against MAIAC retrievals	Model-1	0.385	0.560	-0.241	0.535	-0.052	0.147	0.608			1.476
	Model-2	0.282	0.509	-0.478	0.558	-0.136	0.153	0.741			1.384
	Model-3	0.485	0.351	-0.532	0.963	-0.121	0.215	1.346			1.003
	Model-4	0.263	0.522	-0.280	0.535	-0.077	0.142	0.628			1.447
	Model-5	1.048	0.455	0.587	1.146	0.237	0.375	0.824			1.315
	Model-6	0.289	0.543	-0.497	0.618	-0.124	0.163	0.847			1.348
	Model-7	**0.259**	0.564	**-0.152**	**0.467**	**-0.043**	**0.127**	**0.514**			**1.525**
	Ensemble Mean	0.303	**0.615**	-0.218	0.509	-0.042	0.140	0.577			1.521
$PM_{2.5}$ simulations compared against AirNow observations	Model-1	25.612	0.543	**0.177**	0.618	**3.373**	9.971	0.545	69.080	44.294	2.814
	Model-2	**16.960**	0.490	-0.262	**0.509**	-4.885	**8.022**	0.592	39.674	**23.737**	2.735
	Model-3	20.571	0.433	-0.483	0.962	-4.654	10.645	1.321	71.714	45.416	2.401
	Model-4	51.955	0.492	1.369	1.567	19.804	23.004	0.892	79.159	75.574	2.382
	Model-5	51.940	0.436	0.748	1.122	13.020	18.421	0.772	80.995	68.684	2.497
	Model-7	39.984	0.512	0.814	1.008	12.296	15.912	0.697	80.789	62.514	2.632
	Ensemble Mean	25.925	**0.603**	0.442	0.685	7.403	11.161	**0.537**	**86.845**	60.515	**2.825**

6.4 Ensemble Probability Forecast of $PM_{2.5}$ Exceedances

In general, the ensemble probability shows fairly good performance in forecasting $PM_{2.5}$ exceedances during the 2020 Gigafire events. Figure 11 depicts a contour map of ensemble probability forecast values overlaid by the actual exceedance (binary) over the AirNow sites across the CONUS. The probability ranges from 16.67% (exceedances predicted by only one model; very unlikely to occur) to 100% (exceedances predicted by all six models; very likely to occur). The more models that forecast the exceedance for each grid, the higher probability that the exceedances will occur in that grid. As shown in Figure 11, the contours of high ensemble probability values of 83.33% (five models; orange) and 100% (all six models; red) were displayed mainly in California, which collocated well with the AirNow exceedance measurements (marked as filled red circles). However, the AirNow observed exceedances in the downwind region (Idaho and Montana) were only captured by four of six models, giving a probability forecast of 66.67% (four of six models; yellow). The degradation of exceedance probability in the downwind areas highlights the challenges in predicting transported smoke plumes and their effects on surface air quality.

We also validated their performance in forecasting $PM_{2.5}$ exceedances during extreme fire events by comparing the predicted ensemble exceedance probability against the AirNow observed $PM_{2.5}$ exceedances. The results are shown as time series plots of *aH* and *aFAR* in Figures 12a and 12b. The average *aH* and *aFAR* values are listed in Table 2. High *aH* value and low *aFAR* values suggest good agreement between model simulations and observations. As displayed in the time series plots of *aH* and *aFAR* (Figure 12a and 12b) and the average *aH* and *aFAR* values (Table 2), the lowest ensemble probability of 16.67% shows constantly high *aH* and high *aFAR* throughout the study period, resulting in being the greatest *aH* value of 93.985 (top-ranked) and also the highest *aFAR* value of 78.003 (lowest-ranked) on average, while the highest ensemble probability of 100% show persistently and relatively low *aH* and low *aFAR* all the time, resulting in holding the lowest average *aH* value of 14.725 (lowest-ranked) and the lowest *aFAR* value of 1.537 (top-ranked).

The evaluation results imply that including a small number of models in the ensemble or the low ensemble probability shows better performance in forecasting observed exceedances across the CONUS because some exceedances predicted by any individual model or the lowest ensemble probability were true exceedances associated with wildfires, especially in the wildfire active regions, which frequently matched the AirNow observations, resulting in high *aH*. However, the remaining exceedances predicted elsewhere were false alarms influenced by overestimation that could not be removed from the forecast due to a lack of calibration and validation with other models. As a result, the lowest probability values generally yield high *aFAR*. Conversely, the ensemble forecast with a larger number of models, or the higher ensemble probability performs more accurately and reliably in forecasting $PM_{2.5}$ exceedances on a smaller or local scale due to the fact that their predicted exceedances have been calibrated and verified with the co-existed exceedances predicted by the other participant models included in the ensemble. As a consequence, the areas showing the false exceedances have been reduced or removed, resulting in lower *aH* and *aFAR* values.

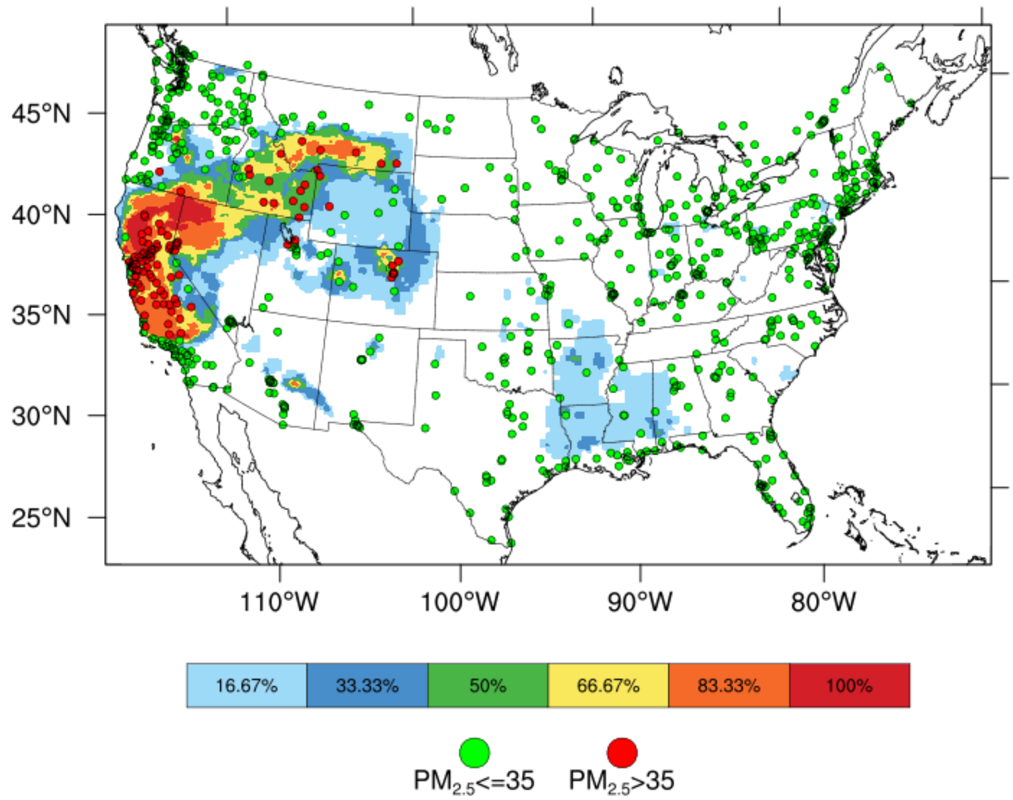


Figure 11. Ensemble probability forecast of $PM_{2.5}$ exceedances on August 22nd, 2020 (during the 2020 Gigafire events). Foreground colors indicate the probability values ranging from 16.67% (one out of six models forecasts the $PM_{2.5}$ exceedance; unlikely to occur) (light blue) to 100% (all six models forecast the $PM_{2.5}$ exceedances; very likely to occur) (red). The $PM_{2.5}$ exceedances observed by the AirNow sites are displayed in the red/green circles (red means an exceedance recorded by the monitor, and green means no exceedance recorded).

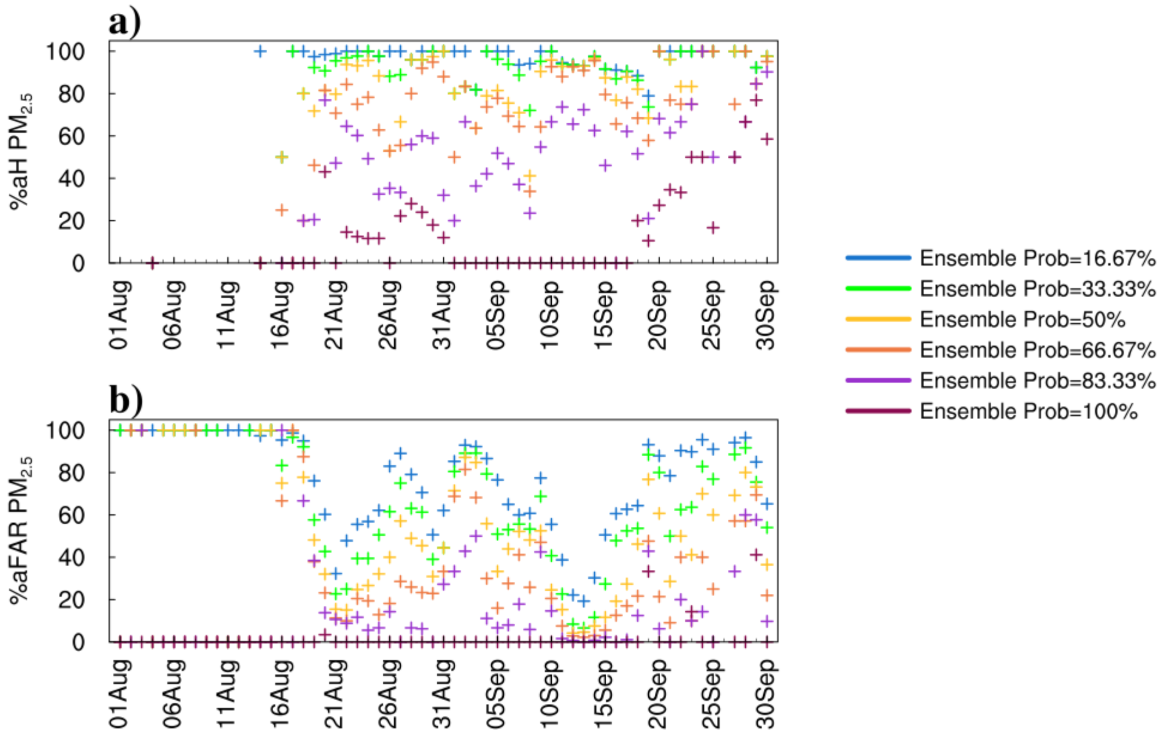


Figure 12. Time series plots of *aH* (a) and *aFAR* (b) values during the 2020 Gigafire events (August-September 2020) for the ensemble probability of $PM_{2.5}$ exceedance forecast. Ensemble probability values range from 16.67% (one out of six models) to 100% (all six models).

Table 2. Averaged *aH* and *aFAR* values of ensemble probability of $PM_{2.5}$ exceedance forecast during the 2020 Gigafire events (August-September 2020), comparing between simulated $PM_{2.5}$ exceedances and observed $PM_{2.5}$ exceedances obtained from AirNow.

Ensemble Probability	Statistical Metric	
	aH	*aFAR*
16.67%	93.985	78.003
33.33%	88.398	63.326
50%	79.305	47.505
66.67%	69.716	28.994
83.33%	48.104	15.378
100%	14.725	1.537

aH: area hit rate; *aFAR*: area false alarm ratio

CHAPTER 7: EVALUATION OF ENSEMBLE FORECAST OF DUST STORMS

In this chapter, an ensemble AOD and $PM_{2.5}$ forecasts based on the unweighted arithmetic mean of six individual models, including GMU-CMAQ, NACC-CMAQ, GEFS-Aerosols, GEOS-5, ICAP-MME, and NAAPS, were created and evaluated with VIIRS enhanced DT AOD and VIIRS DB AOD, and AirNow surface $PM_{2.5}$ observations, respectively, during the 2021 Spring Dust Season (January-March 2021). The forecasting performance of the ensemble mean was also intercompared with the ensemble members to assess whether the ensemble mean can outperform the top performers among these members. The evaluation results were analyzed by calculating average statistical metrics and the overall rating (*RANK*) over the study period.

7.1 Ensemble AOD Forecasting Performance on March 16th, 2021

The AOD simulations were initially evaluated against VIIRS enhanced Dark Target (DT) over dark and bright surfaces AOD on March 16th, 2021. As a consequence of limitations in the AOD detectability of the enhanced DT algorithm, whose maximum retrieved AOD did not pass the data quality test at initial data processes, we validated the AOD simulations with the VIIRS Deep Blue (DB) AOD retrievals.

In general, the ensemble mean shows slightly improved performance in predicting observed dust AOD as demonstrated in contour maps of AOD forecasts, VIIRS enhanced

DT AOD observations, and VIIRS DB AOD observations on March 16th, 2021 (Figures 13a-13i). As illustrated in the contour map of VIIRS DB AOD (Figure 13i), the observed high AOD associated with the dust storm almost blanketed entire areas of the Chihuahuan Desert, especially in El Paso, and partially covered the downwind areas in southern New Mexico. Unfortunately, due to poor quality control, the high AOD retrieved by VIIRS enhanced DT over the Chihuahuan Desert was missing, as shown in Figure 13h. Model-1, Model-2, and Model-4 were able to simulate dust storms primarily in west Texas and incompletely over the downwind areas in southern New Mexico. However, compared to VIIRS DB AOD, these models still underestimated AOD over the active dust areas (Figures 13a, 13b, 13d). Meanwhile, Model-3, Model-5, and Model-6 were not able to simulate dust storms over the Chihuahuan Desert in parts of southern New Mexico, western Texas, and along the Mexico-United States border, where the powerful dust storm originated from and blew through (Figures 13c, 13e, 13f). Considering background AOD across the CONUS, all individual models were likely to simulate fairly higher background AOD compared to the satellite observations. In addition, all models except Model-5 predicted relatively high AOD elsewhere, primarily in the southeastern coast of the United States and partially in the northeastern United States. However, we were unable to compare this predicted high AOD with the VIIRS enhanced DT and DB AOD observations due to missing retrieved AOD data therein. These predicted high AOD may be model overestimations due to other emissions like prescribed fires or anthropogenic emissions, while underpredicted AOD may be a consequence of large uncertainties in dust emissions, transports, and depositions, as well as variability in dust parameterizations in each model. The individual models underestimated AOD near the dust source regions in western Texas and southern New Mexico and overestimated AOD in the Southeast Coast of the U.S., causing the ensemble mean to demonstrate these same patterns, as shown in Figure 13g.

The performance of the ensemble mean in AOD forecasting on March 16th, 2021 was analyzed through the statistical metrics, as shown in Table 4. Based on VIIRS enhanced DT, the ensemble mean shows the absolute fractional bias (*FB*) reduced to 0.625 at the lowest value from a range of 0.643 to 1.440, lowered mean bias (*M*B; 0.046), improved correlation (*CORR*; 0.188), and had the second highest overall rating (*RANK*; 1.281). In comparison with the VIIRS DB, the ensemble mean demonstrated declining *MB* (0.094), the second highest correlation (0.397), and the second highest *RANK* (1.174) among six members. This result also points to the fact that the underestimations of AOD in the Chihuahuan Desert (in western Texas and southern New Mexico) did not always translate significantly to the total biases in AOD forecasting across the CONUS due to small-scaled dust affected areas. In this case, the positive biases may be due to the effects of high background AOD simulations. In addition, the inconsistency in model biases between the evaluations with VIIRS enhanced DT and VIIRS DB is a result of AOD being retrieved differently by DT and DB algorithms.

Due to the uncertainties in the AOD simulations over the CONUS, we also generated the ensemble AOD forecast for the local active dust region domain covering the Chihuahuan Desert in Mexico across western Texas and southern New Mexico (25°N to 40°N, -110°W to -95°W) (Figure 14g). The performance of the ensemble mean in predicting AOD over this constrained domain was compared with the individual models and then verified with the VIIRS DB AOD by calculating the correlation between predicted AOD and observed AOD on March 16th, 2021. Table 3 shows the comparison of the correlation between observed AOD and predicted AOD over the CONUS domain and the active dust region domain on March 16th, 2021. For the CONUS domain, Model-1 yields the highest correlation (0.451), followed by the ensemble mean scoring the second highest in correlation (0.397). In the case of the active dust regions domain, the correlation of Model-1, whose predicted dust related AOD matched VIIRS DB AOD the most (Figure 14a), improved and became the top rank in correlation (0.454). Whereas the correlations of Model-3 and Model-6, which failed to simulate dust storms in the Chihuahuan Desert (Figures 14c and 14f), were reduced to negative values. Consistent with the large underestimations of dust related AOD by most individual models, the correlation of the ensemble mean was also reduced, but by only 1% (0.392). This suggests an effective capability of the ensemble approach to reduce the biases in model forecasts over the active dust regions. Overall, these results revealed the actual forecasting performance of individual models and the ensemble mean in local dust source areas and also emphasized the great impact of the uncertainties in the model simulations on the accuracy of dust AOD forecasts over the CONUS.

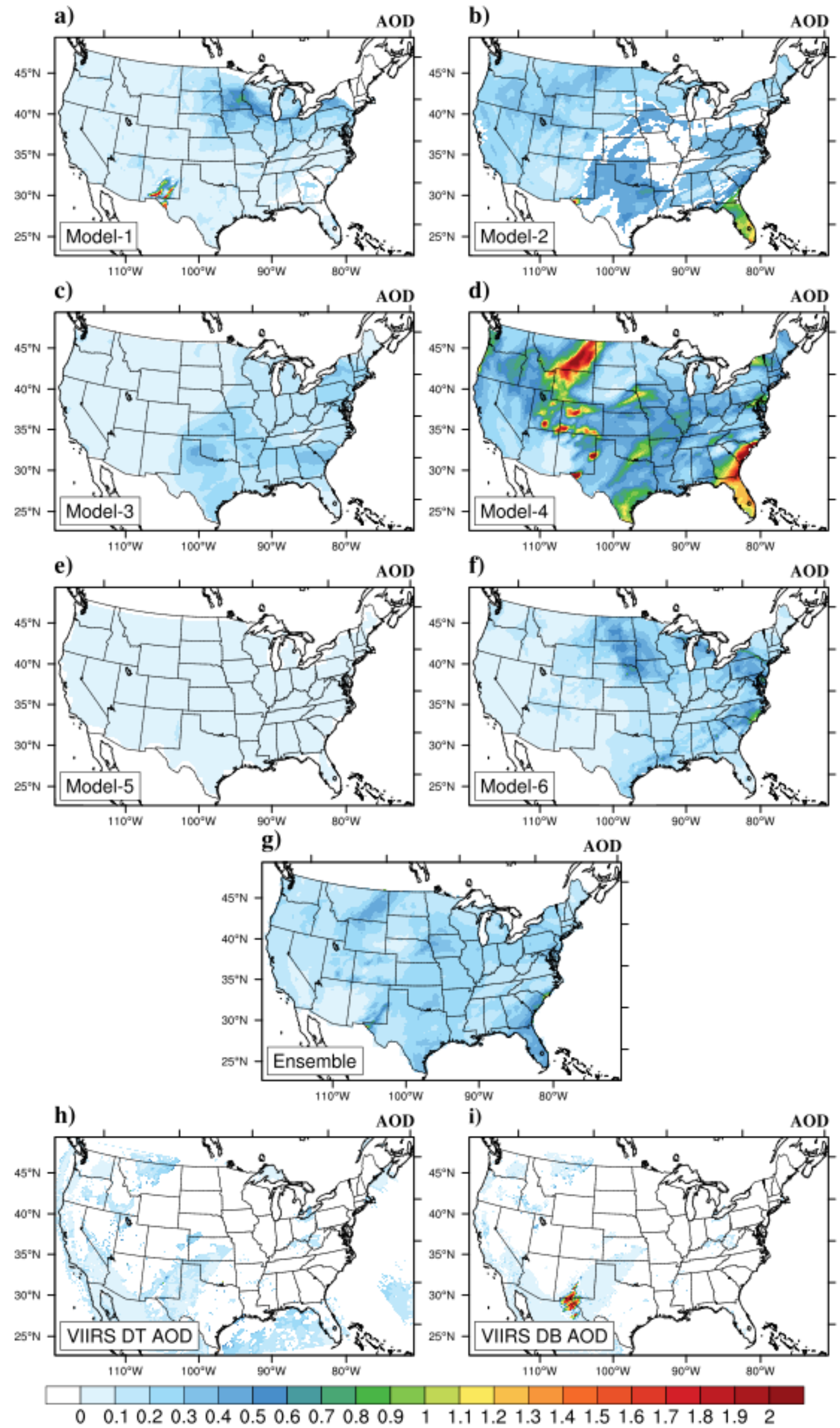


Figure 13. AOD predicted by six individual models (a-f) and the ensemble mean (g), compared with VIIRS enhanced Dark Target (DT) AOD retrievals (h) and VIIRS Deep Blue (DB) AOD retrievals (i) on March 16th, 2021 (during the 2021 Spring Dust Season).

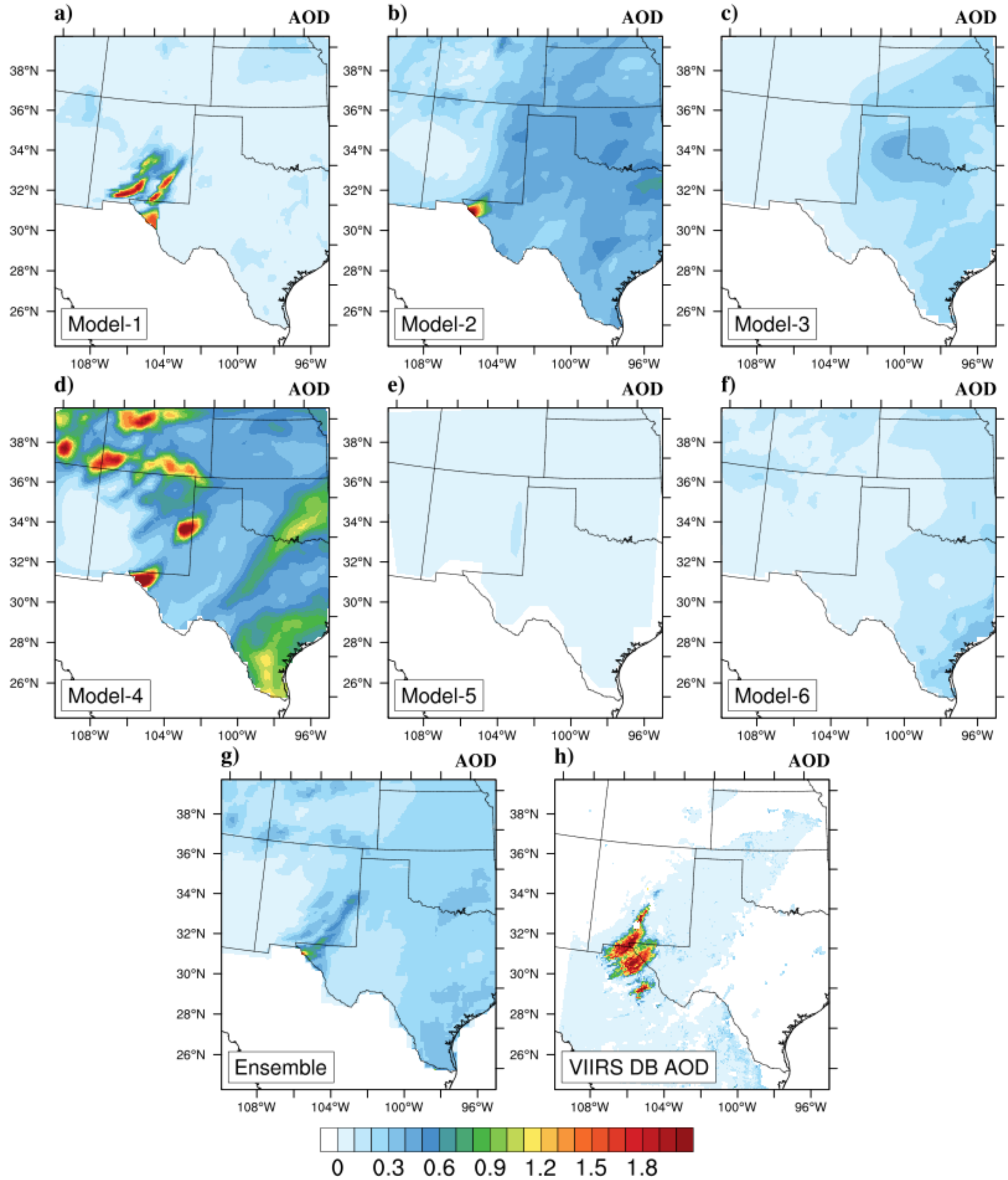


Figure 14. AOD predicted by six individual models (a-f) and the ensemble mean (g), compared with VIIRS Deep Blue (DB) AOD retrievals (h) near the active dust regions in western Texas and southern New Mexico on March 16th, 2021 (during the 2021 Spring Dust Season).

Table 3. Correlation between AOD simulations and VIIRS Deep Blue (DB) AOD observations over the Contiguous United States (CONUS) and Active Dust Regions in western Texas and southern New Mexico on March 16th, 2021. The highest correlation is highlighted in bold.

Model	Correlation	
	CONUS	Active Dust Regions
Model-1	**0.451**	**0.454**
Model-2	0.280	0.265
Model-3	-0.146	-0.323
Model-4	0.351	0.329
Model-5	0.081	0.075
Model-6	0.044	-0.022
Ensemble Mean	0.397	0.392

Table 4. Overall ensemble mean and individual model performances in forecasting AOD values and $PM_{2.5}$ concentrations on March 16th, 2021, based on the evaluation of the values of *RMSE, CORR*, *NMB*, *NME*, *MB*, *ME*, *FB*, *aH*, and *aFAR*, and overall rating (*RANK*). The best results of each statistical metric and *RANK* are highlighted in bold.

Cases	Models	*RMSE*	*CORR*	*NMB*	*NME*	*MB*	*ME*	*FB*	*aH*(%)	*aFAR*(%)	*RANK*
AOD simulations compared against VIIRS enhanced Dark Target retrievals	Model-1	0.197	0.080	-0.432	0.787	-0.064	0.117	1.004			1.038
	Model-2	0.245	0.094	1.109	1.285	0.164	0.190	0.827			1.134
	Model-3	**0.115**	0.131	**-0.252**	0.562	**-0.036**	**0.079**	0.643			1.119
	Model-4	0.413	0.276	2.052	2.138	0.304	0.316	1.055			1.110
	Model-5	0.167	-0.113	-0.829	0.844	-0.121	0.124	1.440			0.723
	Model-6	0.123	**0.325**	-0.393	**0.553**	-0.058	0.082	0.689			**1.318**
	Ensemble Mean	0.154	0.188	0.313	0.723	0.046	0.106	**0.625**			1.281
AOD simulations compared against VIIRS Deep Blue retrievals	Model-1	0.216	**0.451**	**0.057**	0.897	**0.005**	0.081	**0.872**			**1.290**
	Model-2	0.303	0.280	2.355	2.787	0.212	0.251	1.280			1.000
	Model-3	0.252	-0.146	0.279	1.330	0.026	0.126	1.167			0.844
	Model-4	0.438	0.351	3.310	3.675	0.305	0.339	1.384			0.983
	Model-5	0.211	0.081	-0.677	**0.819**	-0.060	**0.073**	1.239			0.921
	Model-6	**0.205**	0.044	-0.151	0.847	-0.013	0.076	0.916			1.064
	Ensemble Mean	0.219	0.397	1.016	1.582	0.094	0.146	1.049			1.174
$PM_{2.5}$ simulations compared against AirNow observations	Model-1	8.439	0.300	0.092	0.562	0.869	5.289	0.537	20.000	90.909	1.672
	Model-2	7.117	0.242	-0.134	**0.478**	-1.258	4.501	0.512	0.000	100.000	1.365
	Model-3	11.832	0.122	0.383	0.752	3.620	7.107	0.631	0.000	100.000	1.245
	Model-4	8.656	**0.317**	0.183	0.623	1.720	5.866	0.571	**60.000**	**70.000**	**2.273**
	Model-6	6.972	0.205	**0.075**	0.489	**0.691**	4.523	0.471	0.000	100.000	1.367
	Ensemble Mean	**6.921**	0.284	0.115	0.478	1.084	**4.499**	**0.452**	0.000	100.000	1.416

7.2 Ensemble $PM_{2.5}$ Forecasting Performance on March 16th, 2021

Next, the $PM_{2.5}$ simulations were evaluated against AirNow ground observations on March 16th, 2021. Overall, the ensemble mean underperformed in predicting dust storms (high $PM_{2.5}$ concentrations) in the Chihuahuan Desert. As shown in a map of AirNow $PM_{2.5}$ observations (Figure 15g), $PM_{2.5}$ concentrations above the National

Ambient Air Quality Standard (NAAQS) (>35 μg/m^3) were only observed by three AirNow monitoring sites in western Texas, in El Paso (56 and 55 μg/m^3) and Socorro (54 μg/m^3), and two sites in southern New Mexico, in Hobbs (40 μg/m^3) and Santa Teresa (57 μg/m^3) (marked as red filled circles). Figures 15a, 15c, and 15d show that Model-1, Model-3, and Model-4 were able to predict elevated $PM_{2.5}$ concentrations affected by dust storms originating in the Chihuahuan Desert across western Texas and southern New Mexico. However, Model-1 overestimated $PM_{2.5}$ over the dust affected regions, primarily in western Texas and across southwestern New Mexico (>80 μg/m^3) (Figure 15a). Model-3 greatly overestimated $PM_{2.5}$ in northwestern Texas (>70 μg/m^3), eastern New Mexico (>100 μg/m^3), and along the southern Colorado-Kansas border (>100 μg/m^3) (Figure 15c), while Model-4 overestimated $PM_{2.5}$ predominantly in southeastern New Mexico-Texas border (>90 μg/m^3) (Figure 15d). Model-2 and Model-6 were unable to predict the dust storms in the Chihuahuan Desert, resulting in low simulated $PM_{2.5}$ over western Texas and southern New Mexico (Figures 15b and 15f). Furthermore, elevated $PM_{2.5}$ concentrations were predicted specifically over the southeastern coast of the U.S. by Model-3 (35-75 μg/m^3) and Model-4 (35-100 μg/m^3), were slightly overestimated in the northeastern U.S. by Model-1 and Model-6 (35-50 μg/m^3), but there were no $PM_{2.5}$ concentrations above 35 μg/m^3 observed by AirNow sites in these two regions. Despite the underestimation of $PM_{2.5}$ over local active dust regions and significant overestimation of $PM_{2.5}$ in the southeastern coast of the U.S. and the northeastern U.S. from participating models, the ensemble $PM_{2.5}$ forecast underpredicted $PM_{2.5}$ over the Chihuahuan Desert in western Texas and southern New Mexico (10-40 μg/m^3) and overpredicted $PM_{2.5}$ in more constrained areas in western New Mexico (>90 μg/m^3) and along the southern Colorado-Kansas border (>100 μg/m^3), and slightly overpredicted $PM_{2.5}$ in the southeastern coast of the U.S. (35-55 μg/m^3) (Figure 15f). Similar to the AOD forecasts, the predicted elevated $PM_{2.5}$ in the U.S. Southeast and Northeast may be a consequence of prescribed fire and anthropogenic emissions, and the underestimations of $PM_{2.5}$ may arise from errors in the model simulations that are influenced by large uncertainties in dust emissions, variability in dust physical and chemical processes and different dust schemes implemented in each model.

In general, the statistical results in Table 4 indicated the ability of the ensemble mean to moderately reduce biases in model forecasting and improve the accuracy of surface $PM_{2.5}$ forecasts during dust storm events. For instance, the ensemble shows the best *RMSE* (6.921 from a range of 6.972-11.832) and *FB* (0.452 from a range of 0.471-0.631), decreased *MB* (6.921), slightly increased correlation (0.284), and the third highest *RANK* (1.416) among five members. However, the area hit rate (*aH*) value of 0 and area false alarm ratio (*aFAR*) value of 100 achieved by the ensemble mean suggest that there is no $PM_{2.5}$ exceedance (>35 μg/m^3) being observed by any AirNow sites on March 16th, 2021 was predicted by the ensemble mean. Therefore, all ensemble predicted $PM_{2.5}$ exceedances were false exceedances. Furthermore, this evaluation result also implied that the underpredictions of $PM_{2.5}$ in the small-scaled local dust associated areas did not significantly contribute to overall biases in $PM_{2.5}$ forecasts on March 16th, 2021, whereas the major contribution causing overall positive biases in forecasts was the $PM_{2.5}$ overpredictions over larger areas in the U.S. Northeast and the Southeast.

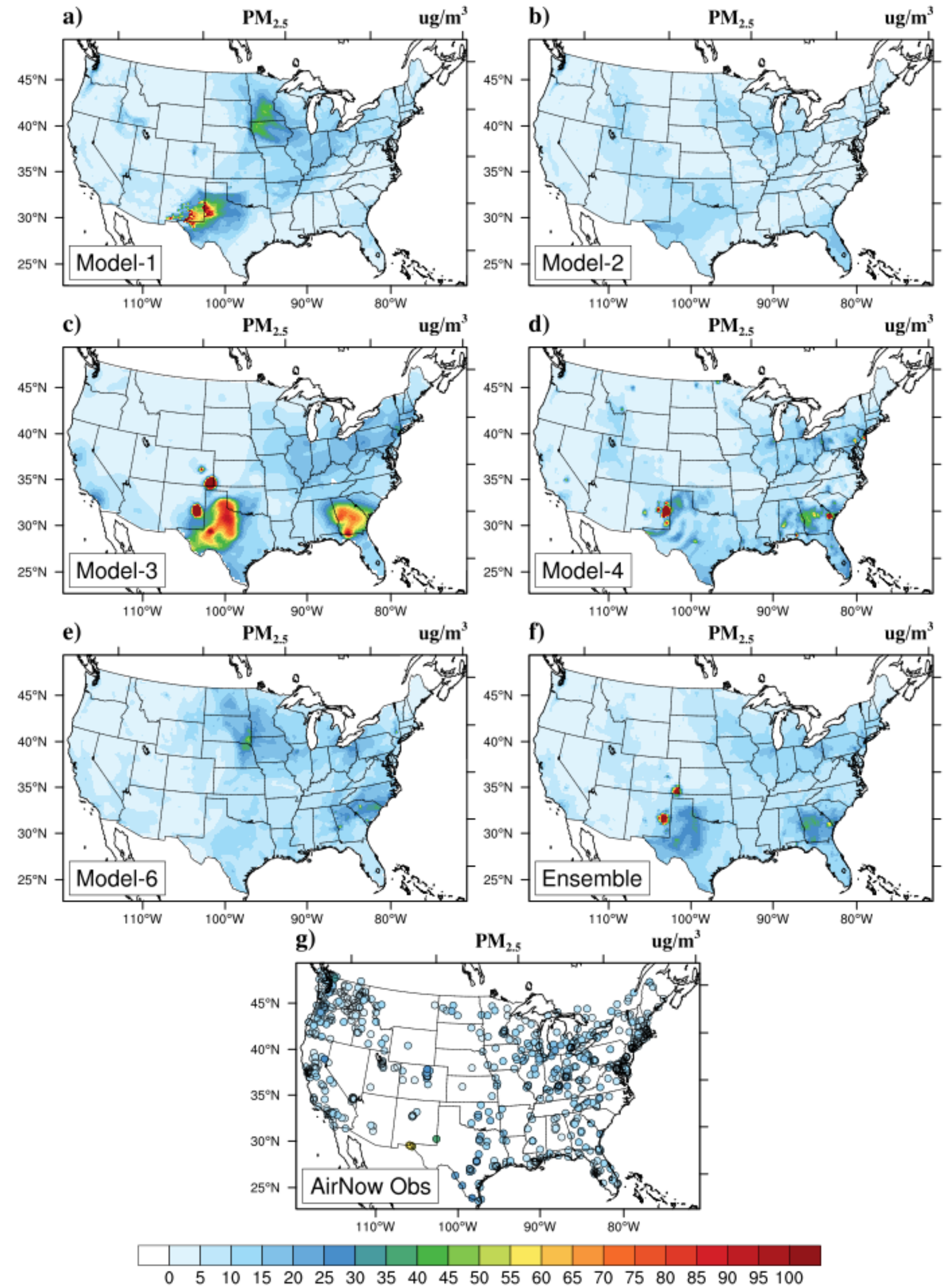


Figure 15. Surface $PM_{2.5}$ concentrations predicted by five individual models (a-e) and the ensemble mean (f), compared with AirNow $PM_{2.5}$ observations (g) on March 16th, 2021 (during the 2021 Spring Dust Season).

7.3 Ensemble Forecasting Performance during the 2021 Spring Dust Season

To assess the overall performance of the ensemble mean in forecasting AOD and $PM_{2.5}$ during dust storm events, we also conducted an evaluation for the 2021 Spring Dust Season from January to March 2021.

For AOD forecasts, overall, all individual models and the ensemble mean show high positive mean bias (*MB*) during the intense dust storm in the middle of March (Figures 16-17c). The errors in model simulations of AOD throughout the study period were also demonstrated by high varying values of root mean square (*RMSE*), mean error (*ME*), normalized mean bias (*NMB*), normalized mean error (*NME*), and absolute fractional bias (*FB*), as shown in Figures 16-17b, d, e, f, and 16-17g. From Table 5, the average *MB* of the ensemble mean over the study period was lowered to 0.013 and 0.068 when compared to VIIRS enhanced DT AOD and VIIRS DB AOD, respectively. According to the time series of correlation shown in Figures 16-17a, the ensemble mean correlation was fairly low and fluctuated throughout the study period, indicating a fairly high level of inconsistency in the forecasts. However, compared to the individual models, the correlation of the ensemble mean was slightly stronger, resulting in increased values of 0.093 and 0.323 (the best correlation) based on VIIRS enhanced DT and VIIRS DB, respectively. Regarding the average overall rating (*RANK*) in Table 5, the ensemble mean yields the third best *RANK* at 1.077 and 1.286, which is 9% and 5% lower than the top rank model, based on VIIRS enhanced DT AOD and VIIRS DB AOD, respectively.

In the forecasts of surface $PM_{2.5}$ concentration, most individual models and the ensemble mean demonstrate the high positive *MB* values during the intense dust storm period (the middle of March) (Figure 18c). The errors in model simulations during the dust storms were indicated by high *RMS*E, *ME*, *NMB*, *NME*, and *FB* in the middle of March, as shown in Figures 18b, d, e, f, and 18g. Table 5 shows that the average *MB* of the ensemble mean during January-March 2021 was slightly reduced to 3.139. As shown in the time series of correlation (Figure 18a), the correlation of the ensemble mean fluctuated during the intense dust storm period (mid-March), implying a fairly high level of discrepancies between the ensemble forecasts and the observations. However, the correlation of the ensemble mean was slightly more consistent and closer to 1 compared to most of the participating models, resulting in the second best average correlation at 0.395 (Table 5). In addition, the ensemble performance in forecasting the NAAQS $PM_{2.5}$ exceedance (concentrations >35 $\mu g/m^3$) was determined by calculating area hit rate (*aH*) and area false alarm ratio (*aFAR*). The time series of *aH* and *aFAR* (Figures 18g and 18h) shows that the ensemble mean slightly increased areas hit rate (*aH*) and reduced area false alarm ratio (*aFAR*), specifically in the middle of March when some individual models failed greatly to predict the exceedances, characterized by relatively low *aH* and high *aFAR*. It should be noted that the concurrent disappearance of any *aH* and *aFAR* lines in the time series (Figure 18g) indicates no exceedances observed by any AirNow sites at that time. Overall, the ensemble mean slightly increased the average *aH* to 14.822%, meaning that merely 14% of the predicted exceedances were truly observed. It also slightly reduced the fairly high average *aFAR* to 62.425%, meaning that about 62% of the predicted exceedances were not observed (false exceedances) (Table 5). These results suggested the capability of the ensemble mean to slightly improve the accuracy of

$PM_{2.5}$ exceedance forecasts during dust storms. With fairly weak correlation, low *aH*, and moderately high *aFAR* and *FB*, the ensemble mean achieves the third highest *RANK* (2.239), which is 11% lower than the first ranked individual model.

Ultimately, all the results point to the fact that the ensemble mean has the potential to reduce biases in individual $PM_{2.5}$ forecasts and provide an improved probabilistic forecast of AOD and $PM_{2.5}$ during dust storm events. However, its performance is less effective than the wildfire ensemble forecasts.

7.4 Overall Performance of Ensemble Dust Storm Forecasts

Figures 19a-19c show the time series of the overall rating (*RANK*) for AOD and $PM_{2.5}$ predicted by the ensemble mean and participating models during the 2021 Spring Dust Season compared with VIIRS enhanced Dark Target (DT) over dark and bright surfaces AOD, VIIRS Deep Blue (DB) AOD retrievals, and AirNow surface $PM_{2.5}$ observations.

For AOD, the ensemble mean consistently fluctuated at a poor *RANK* during the intense dust storms in the middle of March for both comparisons with VIIRS enhanced DT and DB (Figures 19a and 19b). This suggests that the ensemble mean can produce moderately reliable forecasts of AOD during the dust storm events. The inconsistent *RANK* may result from individual model simulation errors caused predominantly by uncertainties in the inputs and model representations of chemical and physical processes. In addition to the systematic errors in the models, the insufficient satellite-retrieved AOD data from limited aerosol detectability over high surface reflectivity areas like deserts and cloud contamination areas can cause discrepancies in evaluations affecting the overall *RANK* because the poor maximum AOD retrievals from these areas are occasionally removed by data quality assurance processes (shown as gaps in Figures 13h and 13i) (Hsu et al., 2019; Levy et al., 2015; Sawyer et al., 2020; Zhang et al., 2018).

For surface $PM_{2.5}$, the ensemble mean also fluctuates in *RANK* over the period (Figure 19c). This fluctuation may be a consequence of model simulation errors and inconsistencies in evaluations due to the scarcity of ground observations in active dust areas near the Chihuahuan Desert in western Texas and southern New Mexico (Figure 15g). Therefore, the ensemble mean overall underperformed in $PM_{2.5}$ forecasting and reproduced moderately reliable $PM_{2.5}$ forecasts during the dust storm events in the Chihuahuan Desert.

In essence, the ensemble forecast improved the predictability of dust related AOD and $PM_{2.5}$ during the 2021 Spring Dust Season in the Chihuahuan Desert, but less successfully than expected due to significant biases in participating model simulations. These biases in the dust air pollution forecasts are frequently a result of the model simulations being affected by large uncertainties in dust and other prescribed emissions (wildfire and anthropogenic emissions), dust circulation and deposition processes, heterogeneity of soil surface properties, and different applications of dust parameterizations, as well as meteorological fields controlling the synoptic scale to mesoscale wind speed and direction. In addition, local dust events in the western U.S. tend to be limited in duration, caused by small-scale wind circulation, and restricted to a particular region, which can all create dust modeling issues.

Furthermore, all statistical analyses suggested that the predictability of ensemble forecasts during the dust storms will be significantly improved if participating models have complementary underestimation and overestimation. Different models have their individual strengths and weaknesses. Although a single model can be excellent at predicting dust AOD, it is not necessarily translated into good surface $PM_{2.5}$ prediction. Our results also demonstrate that the model that performs the best in *RANK* for AOD prediction is different from the model that is best at $PM_{2.5}$ prediction.

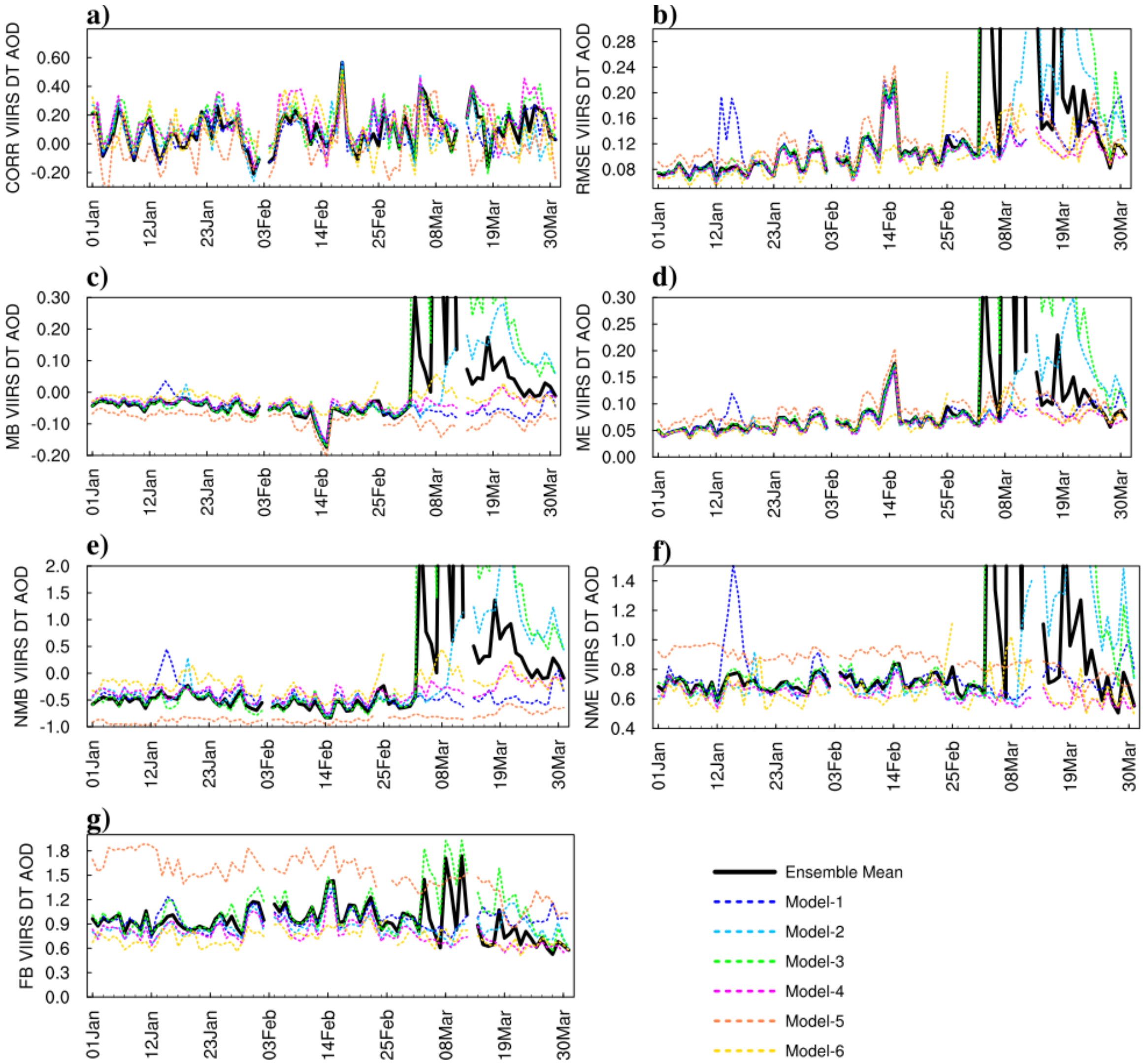


Figure 16. Time series of *RMSE* (a), *CORR* (b), *MB* (c), *ME* (d), *NMB* (e), *NME* (f), and *FB* (g) of AOD during the 2021 Spring Dust Season from January to March 2021. The AOD simulations by the ensemble mean (solid black line) and individual models (dash lines): Model-1 (blue), Model-2 (light blue), Model-3 (green), Model-4 (pink), Model-5 (orange), and Model-6 (yellow) were compared against VIIRS enhanced Dark Target (DT) AOD retrievals.

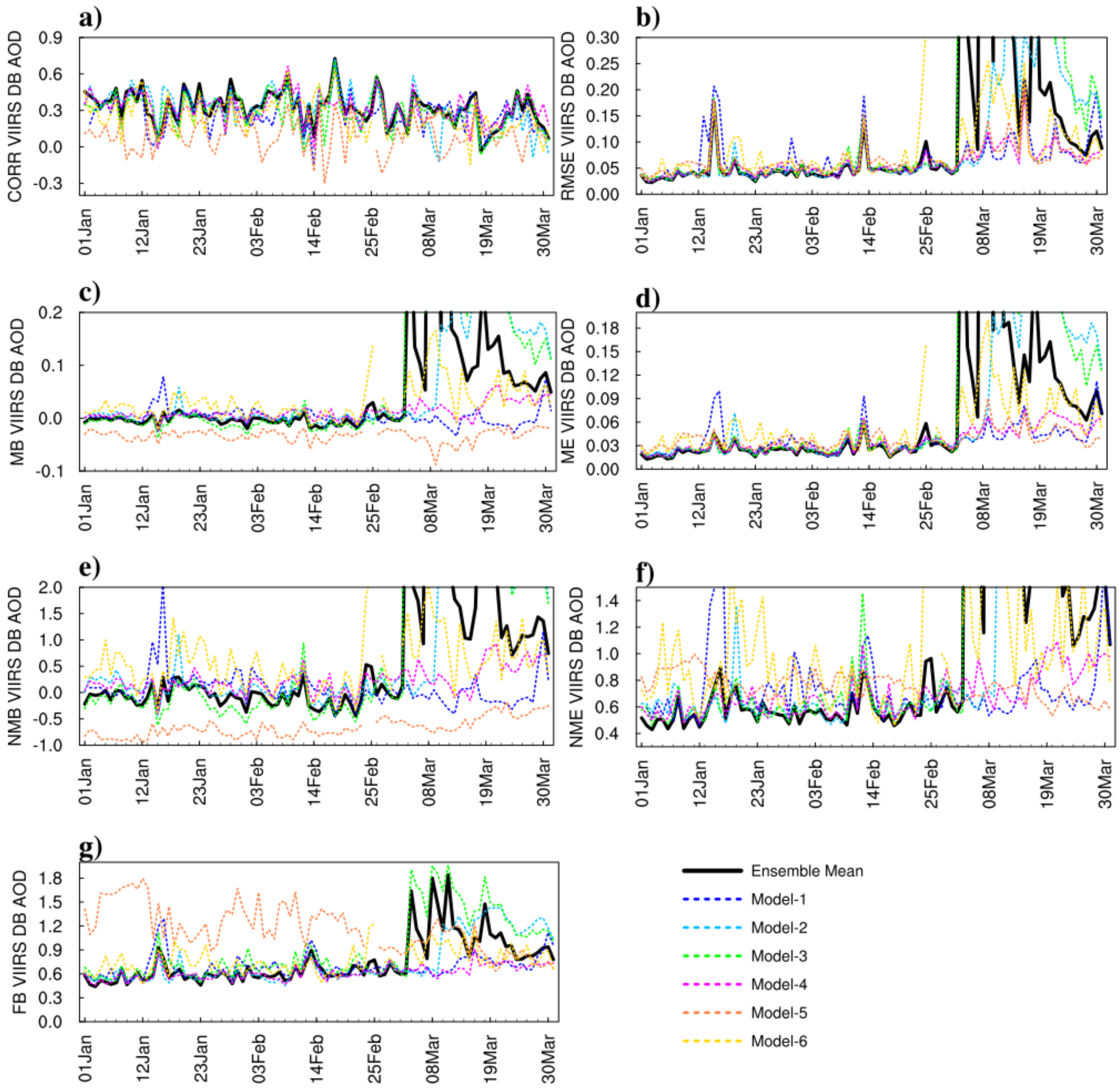


Figure 17. Time series of RMSE (a), *CORR* (b), *MB* (c), *ME* (d), *NMB* (e), *NME* (f), and *FB* (g) of AOD during the 2021 Spring Dust Season from January to March 2021. The AOD simulations by the ensemble mean (solid black line) and individual models (dash lines): Model-1 (blue), Model-2 (light blue), Model-3 (green), Model-4 (pink), Model-5 (orange), and Model-6 (yellow) were compared against VIIRS Deep Blue (DB) AOD retrievals

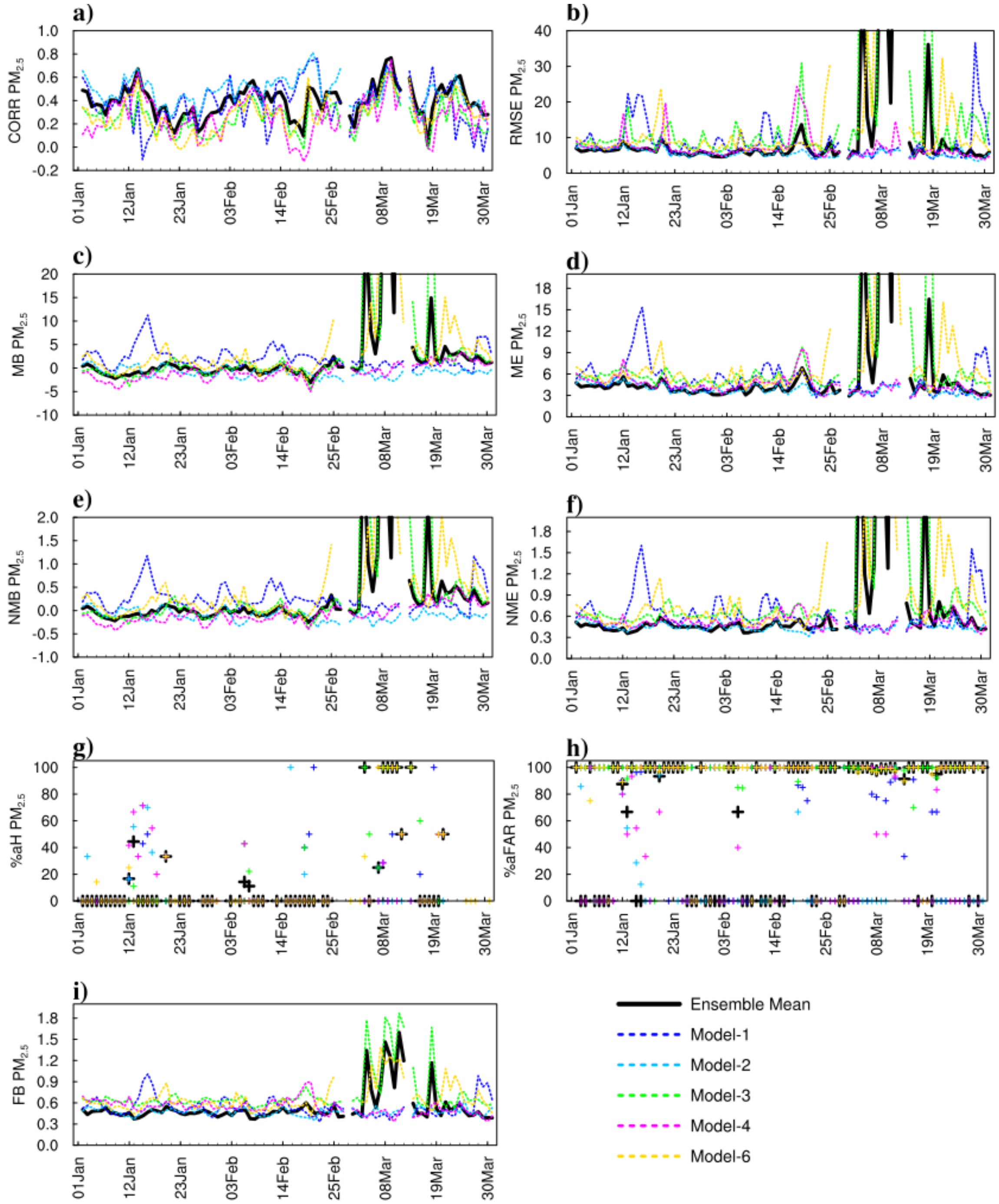


Figure 18. Time series of *RMSE* (a), *CORR* (b), *MB* (c), *ME* (d), *NMB* (e), *NME* (f), *aH* (g), *aFAR* (h), and *FB* (i) of $PM_{2.5}$ for the 2021 Spring Dust Season during January-March 2021. The $PM_{2.5}$ simulations by the ensemble mean (black solid line) and individual Model-1 (blue), Model-2 (light blue), Model-3 (green), Model-4 (pink), and Model-6 (yellow) were compared against AirNow $PM_{2.5}$ observations.

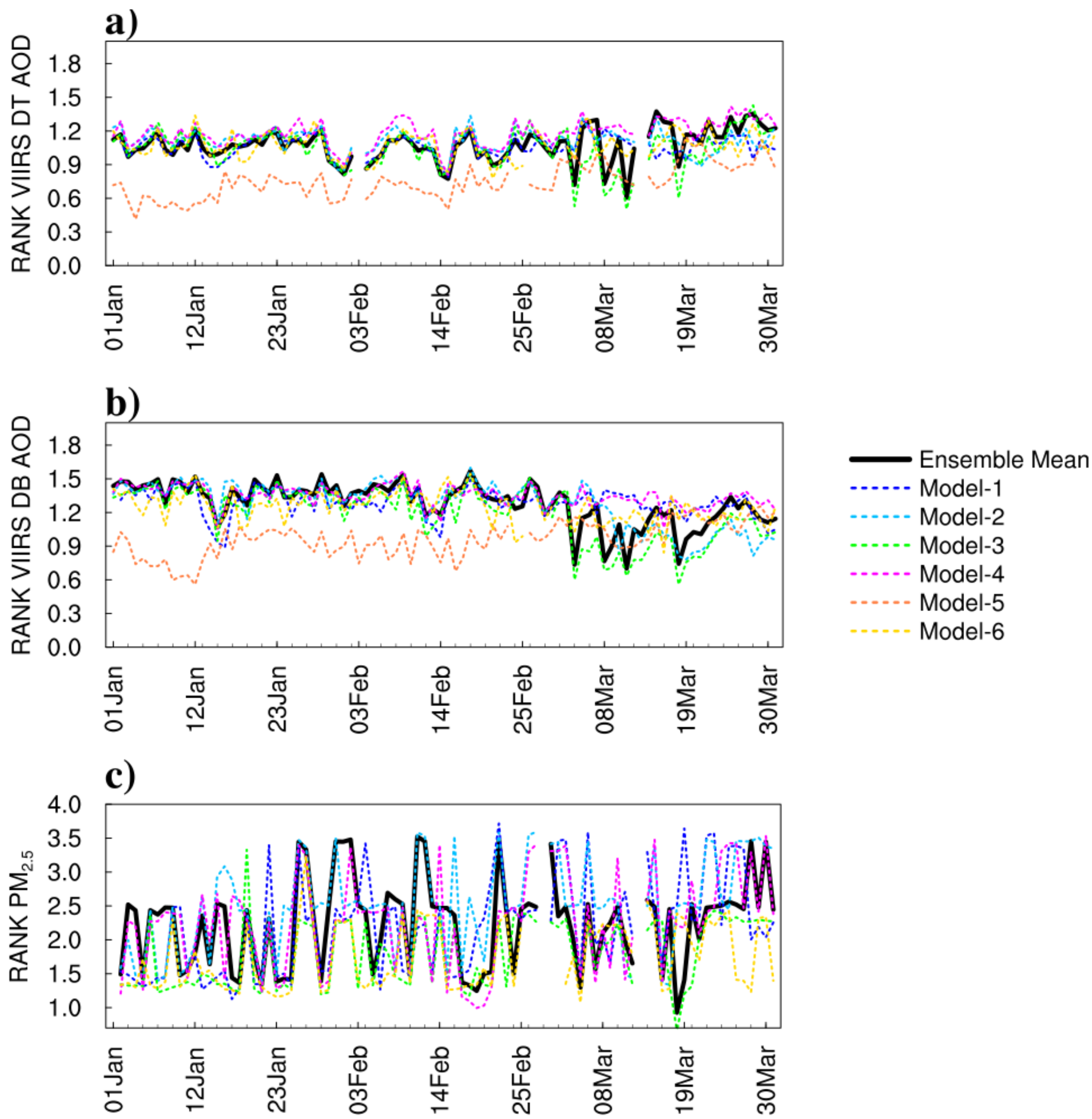


Figure 19. Time series of the overall rating (*RANK*) for AOD and $PM_{2.5}$ simulated by the ensemble mean and individual models. The *RANK* is calculated with four statistical metrics by comparing model predictions against VIIRS enhanced Dark Target (DT) AOD (a) VIIRS Deep Blue (DB) AOD (b) retrievals, and surface $PM_{2.5}$ observations from AirNow (c) during the 2021 Spring Dust Season.

Table 5. Overall ensemble mean and individual model performances in forecasting AOD values and $PM_{2.5}$ concentrations during the 2021 Spring Dust Season (January-March 2021) based on the evaluation of the values of *RMSE*, *CORR*, *NMB*, *NME, MB*, *ME*, *FB aH*, and *aFAR*, and overall rating (*RANK*). The best results of each statistical metric and *RANK* are highlighted in bold.

Cases	Models	*RMSE*	*CORR*	*NMB*	*NME*	*MB*	*ME*	*FB*	*aH*(%)	*aFAR*(%)	*RANK*
AOD simulations compared against VIIRS Dark Target retrievals	Model-1	0.118	0.083	-0.433	0.746	-0.048	0.078	0.953			1.065
	Model-2	0.130	0.086	-0.078	0.830	**-0.001**	0.091	0.859			1.114
	Model-3	**0.099**	0.088	-0.203	0.654	-0.020	**0.066**	**0.730**			1.070
	Model-4	0.625	0.147	1.676	2.624	0.231	0.324	1.057			1.045
	Model-5	0.124	0.003	-0.843	0.859	-0.087	0.089	1.504			0.750
	Model-6	0.100	**0.150**	-0.347	**0.638**	-0.037	0.066	0.784			**1.183**
	Ensemble Mean	0.200	0.093	**0.008**	1.025	0.013	0.115	0.938			1.077
AOD simulations compared against VIIRS Deep Blue retrievals	Model-1	0.069	0.276	**0.079**	0.733	**0.003**	0.038	0.692			1.292
	Model-2	0.099	0.314	0.820	1.255	0.053	0.074	0.755			1.280
	Model-3	0.085	0.260	0.564	1.011	0.031	0.054	0.763			1.248
	Model-4	0.592	0.285	4.042	4.614	0.291	0.317	0.912			1.186
	Model-5	0.066	0.088	-0.666	0.752	-0.033	0.038	1.164			0.962
	Model-6	**0.060**	0.319	0.223	**0.694**	0.012	**0.036**	**0.620**			**1.350**
	Ensemble Mean	0.162	**0.323**	0.960	1.474	0.068	0.093	0.752			1.286
$PM_{2.5}$ simulations compared against AirNow observations	Model-1	8.600	0.378	0.250	0.621	2.151	5.318	0.538	15.317	67.053	2.187
	Model-2	**5.675**	**0.475**	-0.110	**0.450**	-0.903	**3.866**	**0.477**	7.661	**37.755**	**2.512**
	Model-3	13.543	0.287	0.470	0.891	3.938	7.686	0.663	12.969	86.406	1.762
	Model-4	25.779	0.258	1.156	1.699	10.517	15.312	0.730	**16.688**	92.082	1.787
	Model-6	7.261	0.267	**-0.070**	0.519	**-0.795**	4.485	0.542	11.076	49.938	2.264
	Ensemble Mean	10.482	0.395	0.355	0.759	3.139	6.721	0.538	14.822	62.425	2.239

7.5 Ensemble Probability Forecast of $PM_{2.5}$ Exceedances

In this section, the ensemble exceedance probability forecasts (or binary prediction) were evaluated with the AirNow observed $PM_{2.5}$ exceedances. Overall, the ensemble exceedance forecast performed moderately in providing probability forecasts of dust-related $PM_{2.5}$ exceedances (concentration above 35 μg/m^3) during the 2021 Spring

Dust Season (January-March 2021). Figure 20 shows a contour map of the ensemble probability forecast values overlaid by the actual exceedances over the AirNow sites across the CONUS. The probability ranges from 20% (exceedances predicted by only one model) to 100% (exceedances predicted by all five models).

On March 16th, 2021, exceedances were observed at three AirNow monitoring sites in western Texas, in El Paso (56 and 55 $\mu g/m^3$) and Socorro (54 $\mu g/m^3$), and two sites in southern New Mexico, in Hobbs (40 $\mu g/m^3$) and Santa Teresa (57 $\mu g/m^3$) (marked as red filled circles). The ensemble exceedance probability was 20% (one model; light blue) and located in western Texas, partially over the downwind areas in eastern New Mexico, and along the southern Colorado-Kansas border. The ensemble exceedance probability reached 40% (two models; cyan) over relatively constrained areas in northwestern Texas to eastern New Mexico. The ensemble exceedance probability of 20% and 40% were predicted over the Southeast U.S. Coast, predominantly in Georgia. However, there were no exceedances observed by the AirNow sites in these areas. These false exceedances were consistent with the aforementioned overestimations of $PM_{2.5}$ by most participating models.

The ensemble exceedance forecast performance is shown as a time series plots of *aH* and *aFAR* in Figures 21a-21b. The average *aH* and *aFAR* over the study period are listed in Table 6. High *aH* value and low *aFAR* values suggest good agreement between model predictions and observations. The lowest ensemble probability of 20% shows the best *aH* of 28.965 and the worst *aFAR* of 94.936 on average, while the higher ensemble probability values show worse *aH* and better *aFAR*. Since the low ensemble exceedance probability forecast (20%) used only one member to predict exceedances, these simulated exceedances were not calibrated with other members. As a result, a few predicted exceedances were actually observed and the majority were false exceedances, leading the worst *aFAR*. In contrast, higher ensemble probabilities used more models to predict the exceedances performed more accurately and reliably in forecasting $PM_{2.5}$ exceedances on a smaller or local scale because their predicted exceedances have been calibrated with the co-existed exceedances predicted by the other models included in the ensemble. Therefore, the false exceedances were greatly reduced, and the predicted dust-related exceedances were constrained only in dust-active regions, resulting in better *aFAR*. In this case, the *aH* of the high ensemble exceedance probability (80%-100%) forecasts equal to 0, referring to the failures of most members to predict the exceedances during dust storms. In addition, aH of 0 may describe different scenarios depending on the *aFAR*. For instance, an ensemble probability of 100% showing average *aH* and *aFAR* values of 0 suggests no mutual exceedances predicted by all five individual models during the study period (Figure 21a), meaning no false alarms were counted (Figure 21b). On the other hand, the ensemble probability of 80% showed an average *aH* value of 0 and average *aFAR* value of 3.409, indicating that there were mutual exceedances predicted by four individual models that were false alarms as none of them were observed by AirNow sites. All results indicate that the ensemble underperformed in providing high exceedance probability (80-100%) forecasts over areas affected by high concentrations of dust $PM_{2.5}$ above the NAAQS health standard.

In practice, the accuracy of the exceedance probability forecast depends on the original spatial resolutions of each ensemble member. The exceedances simulated by the global models generally cover larger areas compared to the regional models, even after being interpolated to a higher spatial resolution. Using the multi-model ensemble approach generally reduces the discrepancies between the spatial resolutions of the ensemble members. Although the ensemble was able to generate only low-medium (20%-60%) exceedance probability forecasts during dust storms, it can probably be used to provide hazardous areas during dust storms in addition to the ground observations, as shown in Figure 20.

The major challenge in dust associated $PM_{2.5}$ exceedance forecasting is the occurrence of dust storms in active dust regions, which are relatively small areas compared to other air pollution sources and are limited in duration, generally ranging from a few minutes to several hours. Therefore, dust-related exceedances in the active dust regions are frequently underpredicted and the remaining predicted exceedances elsewhere could be false.

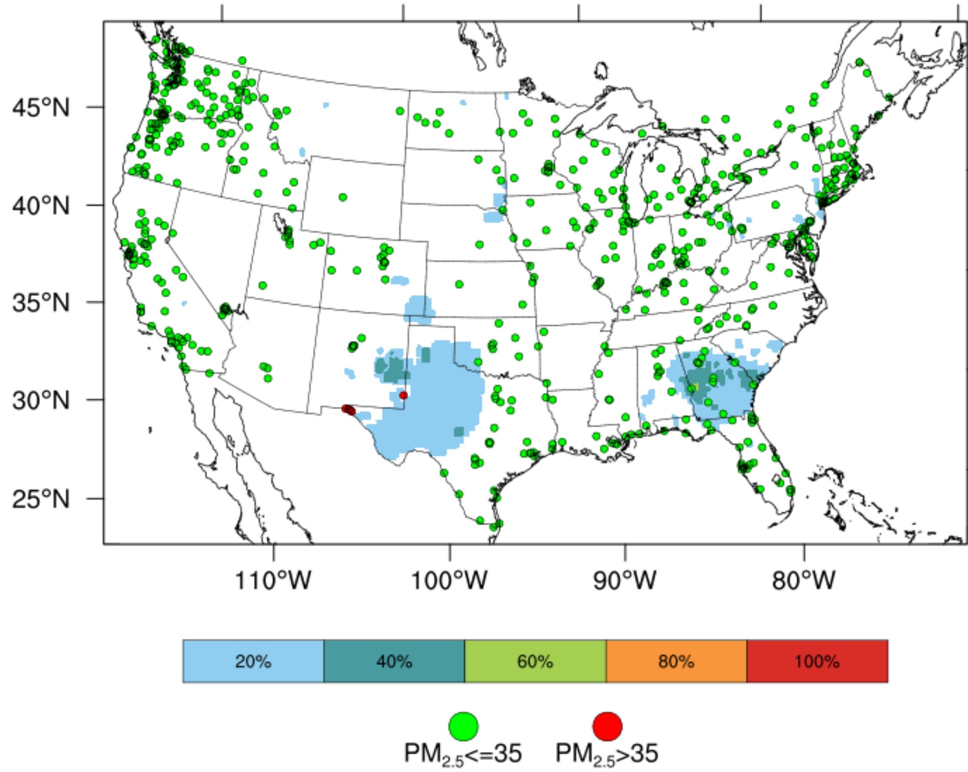


Figure 20. Ensemble probability forecast of $PM_{2.5}$ exceedances on March 16th, 2021 (during the 2021 Spring Dust Season). Foreground colors indicate the probability values ranging from 20% (one out of five models forecasts the $PM_{2.5}$ exceedance) (light blue) to 100% (all five models forecast the $PM_{2.5}$ exceedances) (red). The $PM_{2.5}$ exceedances observed by the AirNow sites are displayed in the red/green circles (red means an exceedance recorded by the monitor, and green means no exceedance recorded).

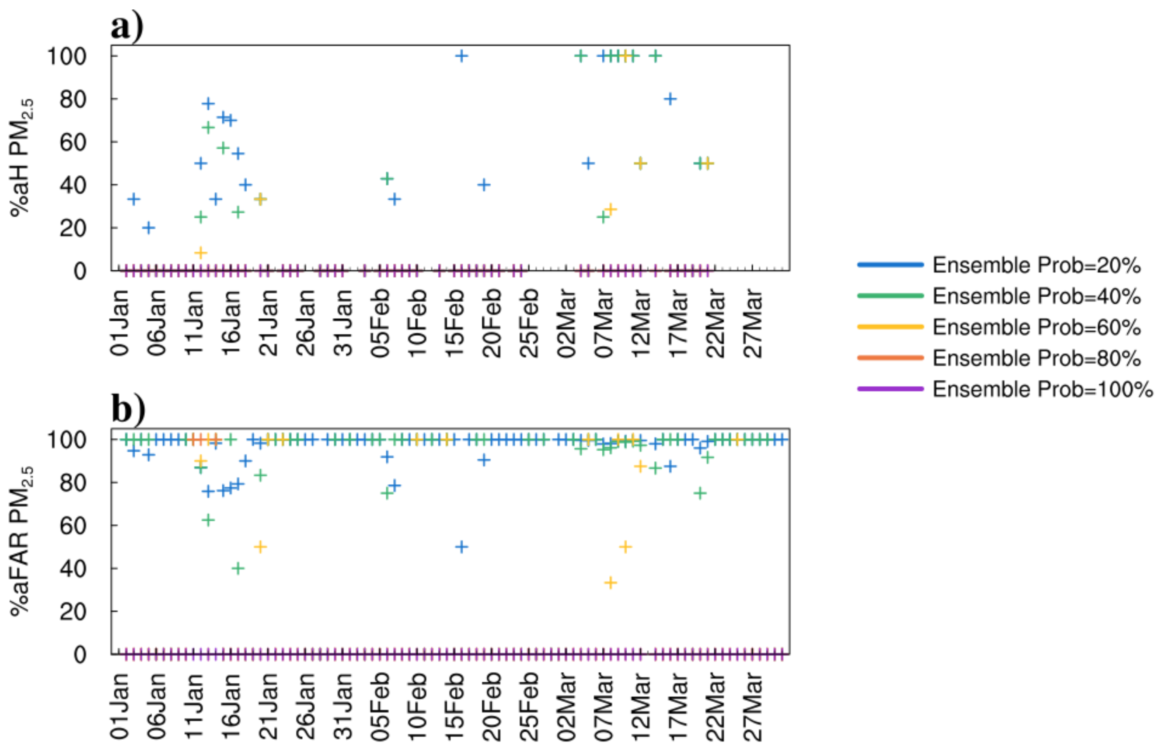


Figure 21. Time series plots of *aH* (a) and *aFAR* (b) values during the 2021 Spring Dust Season (January-March 2021) for the ensemble probability of $PM_{2.5}$ exceedance forecast. Ensemble probability values range from 20% (one out of five models) to 100% (all five models).

Table 6. Averaged *aH* and *aFAR* values of ensemble probability of $PM_{2.5}$ exceedance forecast during the 2021 Spring Dust Season (January-March 2021), comparing between simulated $PM_{2.5}$ exceedances and observed $PM_{2.5}$ exceedances obtained from the AirNow ground monitoring network.

Ensemble Probability	Statistical Metric	
	aH	*aFAR*
20%	28.965	94.936
40%	17.712	60.029
60%	4.659	16.032
80%	0.000	3.409
100%	0.000	0.000

aH: area hit rate; *aFAR*: area false alarm ratio

CHAPTER 8: CONCLUSIONS

8.1 Conclusion and Recommendation for Future Work

Wildfires and dust storms are important emission sources that contribute a large amount of aerosols to the atmosphere, leading to hazardous air quality, which exerts detrimental impacts on society such as adverse health effects, life and property losses, and disruption of economic activities. In this study, we developed and evaluated the North America ensemble wildfire and dust air pollution forecasts of AOD and $PM_{2.5}$ in order to predict wildfire and dust storm effects on AOD and surface $PM_{2.5}$. The multi-model ensemble forecasts were built using three (for wildfire) and two (for dust storm) regional models, one global ensemble model, and three global models operated by NASA, NOAA, NRL, and George Mason University (GMU). These models include the GMU-CMAQ, NACC-CMAQ, HYSPLIT, ICAP-MME, GEFS-Aerosols, GEOS-5, and NAAPS. Our ensemble forecast reproduces daily forecasts of AOD and $PM_{2.5}$ as well as the ensemble probability forecast for wildfire and dust related $PM_{2.5}$ exceedances (24-hr average concentrations >35 $\mu g/m^3$) on a horizontal grid resolution of 12 km×12 km over the CONUS during the 2020 Gigafire events (August-September 2020) in the western U.S. and during the Spring Dust Season (January-March 2021) in the Chihuahuan Desert.

The performance of the ensemble forecasting for AOD and $PM_{2.5}$ during wildfire and dust storm events was evaluated with VIIRS enhanced Dark Target (DT) over dark and bright surfaces, VIIRS Deep Blue (DB), and MAIAC AOD products, and AirNow surface $PM_{2.5}$ observations by calculating a suite of statistical metrics (*RMSE, CORR, MB, ME, NMB, NME,* and *FB*) and an overall rating (*RANK*). In addition, two discrete categorical metrics (area hit rate; *aH* and area flase alarm ratio; *aFAR*) were employed to measure the performance of ensemble mean and ensemble probability in predicting the exceedances of the National Ambient Air Quality Standards (NAAQS) for $PM_{2.5}$ (concentrations >35 $\mu g/m^3$) during the wildfires and dust storms.

For the wildfire case, overall, the statistics results suggested the ensemble mean significantly reduces the biases and uncertainties in the wildfire air pollution forecast and produces more persistently reliable forecasts during the study period compared to the individual forecasts. For AOD forecasts, the ensemble mean was able to improve model performance, as indicated by the mean bias values greatly reduced to -0.08 and -0.04 based on the comparisons with VIIRS DT and MAIAC, the strongest correlations at 0.57 from a range of 0.30-0.53 (VIIRS DT) and 0.62 from a range of 0.35-0.56 (MAIAC). The ensemble mean also achieved the best (1.45 from a range of 0.95-1.44) and second best overall *RANK* among seven members compared to VIIRS DT and MAIAC. For the forecasts of surface $PM_{2.5}$, the ensemble mean outperformed all individual models, with the mean bias reduced to 7.40, strongest correlation at 0.60 from a range of 0.43-0.54, the lowest fractional bias 0.54 from a range of 0.55-1.32, the highest area hit rate at 87% from a range of 40%-82%, and the best overall *RANK* (2.83 from a range of 2.40-2.81) among six members. In terms of the exceedance probability forecasting (binary prediction) performance, the ensemble practically generated a well-suited exceedance probability forecast that matched the observed AirNow exceedances fairly well, as demonstrated by the lowest area false alarm ratio at 1.52% achieved by the ensemble probability of 100%. This result suggested a great potential of the ensemble exceedance probability forecast to provide air pollution warning alerts when the $PM_{2.5}$ concentrations exceed the NAAQS health standard level (concentrations >35 μg/m^3) during wildfires. Although the evaluation result suggested that the ensemble is capable of reducing bias and uncertainties in the model forecasts, predicted AOD and surface $PM_{2.5}$ are frequently subject to be decoupled due to the vertical distribution of the smoke particles. Nevertheless, the relatively high negative biases and positive biases values of the ensemble forecast in the middle of September 2020 demonstrate the underestimations of AOD and overestimations of $PM_{2.5}$ during intense wildfires, which may have been influenced by the unusually thick smoke that in turn caused large errors in emission estimation and plume injection height calculation (Carter et al., 2020; Pan et al. 2020b; and Ye et al. 2021). Furthermore, the variety of model simulations as well as meteorology and emissions inputs (both initial and boundary conditions) data sets can take into account the uncertainty in the ensemble forecasting.

For the dust storm case, overall, the statistical results suggested that the ensemble mean shows the ability to moderately reduce biases in the AOD and $PM_{2.5}$ predictions over the active dust and downwind areas and provide fairly reliable forecasts during the dust events. For AOD forecasts, the ensemble mean improved model forecasts less successfully than expected, as demonstrated by the mean bias being minimized to 0.01 and 0.07 (based on the comparisons with VIIRS DT and VIIRS DB, respectively), the strongest correlation at only 0.32 (based on VIIRS DB), and having the third highest *RANK* among six members (for both VIIRS DT and DB). For surface $PM_{2.5}$, the ensemble forecasts underperformed, as indicated by mean bias slightly decreasing to 3.32, a fairly weak correlation (0.40), low area hit rate (14.82%), and the third highest *RANK* among five members. In terms of dust associated $PM_{2.5}$ exceedance probability forecasts (binary prediction), the ensemble unsuccessfully generated high probability forecasts of $PM_{2.5}$ exceedance during dust storm events. Instead, it frequently predicted exceedances with low-medium probabilities (20-60%) during the dust events. However, the ensemble probability still has the useful capability to estimate the hazardous areas affected by the dust $PM_{2.5}$ exceedances, especially over areas which are generally difficult to establish air quality monitoring sites. The predominant obstacles for the dust storm ensemble forecasts are the excessive overestimations and underestimations from the ensemble members, primarily in the middle of March when extreme dust storms occurred in the Chihuahuan Desert. These biases arise from the model simulation errors caused by large uncertainties in dust emissions and other emissions like prescribed fire and anthropogenic emissions, dust transport and deposition processes, different applications of dust parameterizations in dust scheme algorithms, and meteorological fields adopted differently in each individual model. In addition, the natural behavior of the local dust events in the western U.S. causes issues in dust forecasting because they are generally limited in duration, small-scale wind circulations, and restricted to a particular region.

In essence, using the ensemble approach can reduce biases in air pollution forecasts and reasonably improve the model predictability during extreme events such as wildfires and dust storms. However, the reliability of the ensemble forecast is still subject to types of extreme events due to different emission sources as well as initial and boundary meteorological conditions.

The development and evaluation of the multi-model ensemble wildfire and dust storm air pollution forecast for the 2020 Gigafire events and the 2021 Spring Dust Season presented here are still at the early stage of deploying the North America ensemble wildfire and dust storm forecast. Comparisons between the ensemble and the individual models represented in this study will be used to investigate differences between models as an attempt to identify the uncertainties in emission and meteorology inputs, as well as in chemical transport/dispersion model simulations. Findings from this pilot study will be used to improve forecasting performance of the ensemble mean and each individual model. Our next step is to extend the multi-model ensemble forecast approach to other periods, including the 2021 fire and dust storm seasons. Finally, the qualified ensemble forecast will be used to improve real-time wildfire forecasting systems over North America to support key decision-making on air quality at local, national, and international levels.

APPENDIX: THE CONFIGURATION OF PARTICIPATING MODEL

Table A1. The configuration of participating models included in the ensemble forecasting.

Air Quality/ Dispersion Model	Operational/Research Center	Forecast Products	Domain(s)	Meteorology Data	Grid Spacing	Initial time	Output Frequency	Fire Detection Information	Emission product/Algorithm	Dust Algorithm	Plume Rise Algorithm
GMU-CMAQ (CMAQv5.3.1)	Air Quality Group at George Mason University	$PM_{2.5}$, O_3, NO_2, AOD, etc.	CONUS and its surrounding areas	GFS products applied to WRFv4.2	12km×12km	18 UTC previous day	Hourly	GBBEPx from VIIRS and MODIS	BEIS, MOVES, 2016v1, Gong (2003) sea-spray emission, applied to SMOKEv4.7	FENGCHA dust scheme	Sofiev (2012)
NACC-CMAQ (CMAQv5.3.1)	NOAA Air Resources Laboratory (ARL)/GMU	$PM_{2.5}$, O_3, NO_2, AOD, etc.	CONUS and its surrounding areas	GFSv16 with FV3	12km×12km	12 UTC	Hourly	GBBEPx from VIIRS and MODIS	BEISv3.6.1-BELD5, MOVES, NEI2016v1	FENGCHA dust scheme, SOILGRIDS 2017 soil fractions, Prigent et al (2012) surface roughess	Briggs
HYSPLIT (HYSPLITv5.1.0)	NOAA Air Resources Laboratory (ARL)	PM2.5, AOD	CONUS and its surrounding areas	WRF-ARW	0.15°×0.15°	00 UTC	Hourly	NOAA Hazard Mapping System (HMS)	USFS BlueSky	HYSPLIT threshold friction velocity dust scheme	Briggs
GEOS-5 (GEOS-5.27.1)	NASA Goddard Space Flight Center (GSFC)	$PM_{2.5}$, AOD	Global	Near-real time assimilation (DAS), 10-days forecast at 00z, and 5-days forecast at 12z	12km×12k, with output at 0.25°×0.3125°	00 UTC	Hourly	NOAA Hazard Mapping System (HMS)	Prognostic emissions of dust and seasalt aerosols, smoke - QFED, HTAP, EDGAR, MEGAN	GEOS/GOCART	Briggs
GEFS-Aerosols	NOAA ARL/NOAA CSL/NOAA GSL/GMU	$PM_{2.5}$, PM_{10}, AOD, OC, BC, Dust, Sea Salt, SO_4	Global	FV3GFS	0.25°×0.25°	00 UTC	3 Hourly	MODIS	CEDS,HTAP	FENGSHA dust scheme, SOILGRIDS 2017 soil fractions, Prigent et al (2012) surface roughess	
ICAP-MME	Naval Research Laboratory (NRL)	AOD	Global	Varies	1°×1°	00 UTC	6 hourly	Varies	Varies	Varies	1D cloud
NAAPS (nap044o)	Naval Research Laboratory (NRL)	$PM_{2.5}$, AOD, visibility	Global	Global Meteorological fields from NAVGEM	0.333°×0.333°	00 UTC	3 hourly	MODIS	dust, sea salt, anthropogenic and biogenic fine mode (ABF)	Westphal et al. 2009	Varies

REFERENCES

Achakulwisut, P., Link to external site, this link will open in a new window, Mickley, L. J., & Anenberg, S. C. (2018). Drought-sensitivity of fine dust in the US Southwest: Implications for air quality and public health under future climate change. *Environmental Research Letters*, *13*(5). http://dx.doi.org.mutex.gmu.edu/10.1088/1748-9326/aabf20

Ashley, W. S., Strader, S., Dziubla, D. C., and Haberlie, A., 2015. Driving blind: Weather-related vision hazards and fatal motor vehicle crashes. *Bulletin of the American Meteorological Society 96*(5): 755-778.

Baker, A. R., & Jickells, T. D. (2006). Mineral particle size as a control on aerosol iron solubility. *Geophysical Research Letters*, *33*(17). https://doi.org/10.1029/2006GL026557

Balkanski, Y., Schulz, M., Claquin, T., and Guibert, S. (2007). Reevaluation of mineral aerosol radiative forcings suggests a better agreement with satellite and AERONET data, *Atmospheric Chemistry and Physics*, 7, 81–95, https://doi.org/10.5194/acpd-6-8383-2006

Benedetti, A., Reid, J. S., & Colarco, P. R. (2011). International Cooperative for Aerosol Prediction Workshop on Aerosol Forecast Verification. *Bulletin of the American Meteorological Society*, *92*(11), ES48–ES53. https://doi.org/10.1175/BAMS-D-11-00105.1

Benedetti, A., Baldasano, J. M., Basart, S., Benincasa, F., Boucher, O., Brooks, M. E., Chen, J.-P., Colarco, P. R., Gong, S., Huneeus, N., Jones, L., Lu, S., Menut, L., Morcrette, J.-J., Mulcahy, J., Nickovic, S., Pérez García-Pando, C., Reid, J. S., Sekiyama, T. T., … Zhou, C.-H. (2014). Operational Dust Prediction. In P. Knippertz & J.-B. W. Stuut (Eds.), *Mineral Dust: A Key Player in the Earth System* (pp. 223–265). Springer Netherlands. https://doi.org/10.1007/978-94-017-8978-3_10

Brasseur, G. P., Xie, Y., Petersen, A. K., Bouarar, I., Flemming, J., Gauss, M., Jiang, F., Kouznetsov, R., Kranenburg, R., Mijling, B., Peuch, V.-H., Pommier, M., Segers, A., Sofiev, M., Timmermans, R., van der A, R., Walters, S., Xu, J., & Zhou, G. (2019). Ensemble forecasts of air quality in eastern China – Part 1: Model description and implementation of the MarcoPolo–Panda prediction system, version 1. *Geoscientific Model Development*, *12*(1), 33–67. https://doi.org/10.5194/gmd-12-33-2019

Briggs, G. (1969). Plume rise: A critical review (Technical Report). (p. 81). Springfield, VA: National Technical Information Service.

Buchard, V., da Silva, A. M., Colarco, P., Krotkov, N., Dickerson, R. R., Stehr, J. W., Mount, G., Spinei, E., Arkinson, H. L., & He, H. (2014). Evaluation of GEOS-5 sulfur dioxide simulations during the Frostburg, MD 2010 field campaign. *Atmospheric Chemistry and Physics*, *14*(4), 1929–1941. https://doi.org/10.5194/acp-14-1929-2014

Cakmur, R. V., Miller, R. L., & Torres, O. (2004). Incorporating the effect of small-scale circulations upon dust emission in an atmospheric general circulation model. *Journal of Geophysical Research: Atmospheres*, *109*(D7). https://doi.org/10.1029/2003JD004067

California Department of Forestry and Fire Protection (CAL FIRE). (2020). *2020 Fire Season*. Retrieved from https://www.fire.ca.gov/incidents/2020/

Campbell, P., Tang, Y., Lee, P., Baker, B., Tong, D., Saylor, R., Stein, A., Huang, J., Huang, H.-C., Strobach, E., McQueen, J., Pan, L., Stajner, I., Sims, J., Tirado-Delgado, J., Jung, Y., Yang, F., Spero, T., & Gilliam, R. (2021). Development and evaluation of an advanced National Air Quality Forecast Capability using the NOAA Global Forecast System version 16. *Geoscientific Model Development Discussions*, 1–70. https://doi.org/10.5194/gmd-2021-316

Cao, C., Xiong, J., Blonski, S., Liu, Q., Uprety, S., Shao, X., Bai, Y., & Weng, F. (2013a). Suomi NPP VIIRS sensor data record verification, validation, and long-term performance monitoring. *Journal of Geophysical Research: Atmospheres*, *118*(20), 11,664-11,678. https://doi.org/10.1002/2013JD020418

Cao, C., De Luccia, F. J., Xiong, X., Wolfe, R., & Weng, F. (2013b). Early On-Orbit Performance of the Visible Infrared Imaging Radiometer Suite Onboard the Suomi National Polar-Orbiting Partnership (S-NPP) Satellite. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(2), 1142–1156. https://doi.org/10.1109/TGRS.2013.2247768

Carmona, J. M., Vanoye, A. Y., Lozano, F., & Mendoza, A. (2015). Dust emission modeling for the western border region of Mexico and the USA. *Environment and Earth Science*, 74(2), 1687–1697. https://doi.org/10.1007/s12665-015-4173-5

Carter, T. S., Heald, C. L., Jimenez, J. L., Campuzano-Jost, P., Kondo, Y., Moteki, N., Schwarz, J. P., Wiedinmyer, C., Darmenov, A. S., da Silva, A. M., & Kaiser, J. W. (2020). How emissions uncertainty influences the distribution and radiative impacts of smoke from fires in North America. Atmospheric Chemistry and Physics, 20(4), 2073–2097. https://doi.org/10.5194/acp-20-2073-2020

Centers for Disease Control and Prevention (CDC) (2013), Increase in reported coccidioidomycosis—United States, 1998–2011, *MMWR (Morbidity and mortality weekly report), 62*(12), 217.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., Higurashi, A., & Nakajima, T. (2002). Tropospheric Aerosol Optical Thickness from the GOCART Model and Comparisons with Satellite and Sun Photometer Measurements. *Journal of The Atmospheric Sciences*, *59*, 461-483.

Colarco, P., Silva, A. da, Chin, M., & Diehl, T. (2010). Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth. *Journal of Geophysical Research: Atmospheres*, *115*(D14). https://doi.org/10.1029/2009JD012820

Colarco, P., Benedetti, A., Reid, J. and Tanaka, T. (2014a) Using EOS data to improve aerosol forecasting: the International Cooperative for Aerosol Research (ICAP). *The Earth Observer*, 26, 14–19

Creamean, J. M., Spackman, J. R., Davis, S. M., White, A. B., Spackman, J. R., Davis, S. M., & White, A. B. (2014). Climatology of long-range transported Asian dust along the West Coast of the United States. *Journal of Geophysical Research: Atmospheres, 119*, 12,171–12,185. https://doi.org/10.1002/2014JD021694

Crooks, J. L., Cascio, W. E., Percy, M. S., Reyes, J., Neas, L. M., & Hilborn, E. D. (2016). The association between dust storms and daily non-accidental mortality in the United States, 1993–2005. *Environmental health perspectives, 124*(11), 1735.

Darmenova, K., Sokolik, I. N., Shao, Y., Marticorena, B., & Bergametti, G. (2009). Development of a physically based dust emission module within the Weather Research and Forecasting (WRF) model: Assessment of dust emission parameterizations and input parameters for source regions in Central and East Asia. *Journal of Geophysical Research: Atmospheres*, *114*(D14). https://doi.org/10.1029/2008JD011236

Di Tomaso, E., Schutgens, N. A. J., Jorba, O., & Pérez García-Pando, C. (2017). Assimilation of MODIS Dark Target and Deep Blue observations in the dust aerosol component of NMMB-MONARCH version 1.0. *Geoscientific Model Development*, *10*(3), 1107–1129. https://doi.org/10.5194/gmd-10-1107-2017

DeBell, L. J., Vozzella, M., Talbot, R. W., & Dibb, J. E. (2004). Asian dust storm events of spring 2001 and associated pollutants observed in New England by the Atmospheric

Investigation, Regional Modeling, Analysis and Prediction (AIRMAP) monitoring network. *Journal of Geophysical Research: Atmospheres*, *109*(D1). https://doi.org/10.1029/2003JD003733

Delle Monache, L., & Stull, R. B. (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

Delle Monache, L., Deng, X., Zhou, Y., & Stull, R. (2006). Ozone ensemble forecasts: 1. A new ensemble design. *Journal of Geophysical Research: Atmospheres*, *111*(D5). 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*(D5). https://doi.org/10.1029/2005JD006311

Delle Monache, L., Wilczak, J., Mckeen, S., Grell, G., Pagowski, M., Peckham, S., Stull, R., Mchenry, J., & Mcqueen, J. (2008). A Kalman-filter bias correction method applied to deterministic, ensemble averaged and probabilistic forecasts of surface ozone. *Tellus B: Chemical and Physical Meteorology*, *60*(2), 238–249. https://doi.org/10.1111/j.1600-0889.2007.00332.x

Delle Monache, L., Alessandrini, S., Djalalova, I., Wilczak, J., Knievel, J. C., & Kumar, R. (2020). Improving Air Quality Predictions over the United States with an Analog Ensemble. *Weather and Forecasting*, *35*(5), 2145–2162. https://doi.org/10.1175/WAF-D-19-0148.1

Dennison, P. E., Brewer, S. C., Arnold, J. D., & Moritz, M. A. (2014). Large wildfire trends in the western United States, 1984–2011. *Geophysical Research Letters*, *41*(8), 2928–2933. https://doi.org/10.1002/2014GL059576

Dong, X., Fu, J. S., Huang, K., Tong, D., & Zhuang, G. (2016). Model development of dust emission and heterogeneous chemistry within the Community Multiscale Air Quality modeling system and its application over East Asia. *Atmospheric Chemistry and Physics*, *16*(13), 8157–8180. https://doi.org/10.5194/acp-16-8157-2016

Draxler, R.R., and G.D. Hess. (1998). An overview of the HYSPLIT_4 modeling system of trajectories, dispersion, and deposition. *Aust. Meteor. Mag. 47*, 295-308.

Draxler, R.R. (1999). HYSPLIT4 user's guide. NOAA Tech. Memo. ERL ARL-230, NOAA Air Resources Laboratory, Silver Spring, MD.

Draxler, R. R., Ginoux, P., & Stein, A. F. (2010). An empirically derived emission algorithm for wind-blown dust. Journal of Geophysical Research: Atmospheres, 115(D16). https://doi.org/10.1029/2009JD013167

Eyth, A., Vukovich, J., Farkas, C., & Strum, M. (2020). *Technical Support Document (TSD) preparation of emissions inventories for 2016v1 North American emissions modeling platform.*

Fann, N., Alman, B., Broome, R. A., Morgan, G. G., Johnston, F. H., Pouliot, G., & Rappold, A. G. (2018). The health impacts and economic value of wildland fire episodes in the U.S.: 2008–2012. *Science of The Total Environment*, *610–611*, 802–809. https://doi.org/10.1016/j.scitotenv.2017.08.024

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

Fischer, E. V., Hsu, N. C., Jaffe, D. A., Jeong, M. J., & Gong, S. L. (2009). A decade of dust: Asian dust and springtime aerosol load in the US Pacific Northwest. *Geophysical Research Letters, 36*(3).

Flagg, C. B., Neff, J. C., Reynolds, R. L., & Belnap, J. (2014). Spatial and temporal patterns of dust emissions (2004–2012) in semi-arid landscapes, southeastern Utah, USA. *Aeolian Research*, *15*, 31–43. https://doi.org/10.1016/j.aeolia.2013.10.002

Ford, B., Martin, M. V., Zelasky, S. E., Fischer, E. V., Anenberg, S. C., Heald, C. L., & Pierce, J. R. (2018). Future Fire Impacts on Smoke Concentrations, Visibility, and Health in the Contiguous United States. *GeoHealth*, *2*(8), 229–247. https://doi.org/10.1029/2018GH000144

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D. W., Haywood, J., Lean, J., Lowe, D. C., Raga, G., Schulz, M., Dorland, R. V., Bodeker, G., Etheridge, D., Foukal, P., Fraser, P., Geller, M., Joos, F., Keeling, C. D., … Dorland, R. V. (n.d.). *Changes in Atmospheric Constituents and in Radiative Forcing*. 106.

Gaiero, D. M., Simonella, L., Gassó, S., Gili, S., Stein, A. F., Sosa, P., Becchio, R., Arce, J., & Marelli, H. (2013). Ground/satellite observations and atmospheric modeling of dust storms originating in the high Puna-Altiplano deserts (South America): Implications for the interpretation of paleo-climatic archives. *Journal of Geophysical Research: Atmospheres*, *118*(9), 3817–3831. https://doi.org/10.1002/jgrd.50036

Ginoux, P., Prospero, J. M., Gill, T. E., Hsu, N. C., & Zhao, M. (2012). Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on

MODIS Deep Blue aerosol products. *Reviews of Geophysics*, *50*(3). https://doi.org/10.1029/2012RG000388

Gong, S. L., & Zhang, X. Y. (2008). CUACE/Dust – an integrated system of observation and modeling systems for operational dust forecasting in Asia. *Atmospheric Chemistry and Physics*, *8*(9), 2333–2340. https://doi.org/10.5194/acp-8-2333-2008

Grell, G. A., Montuoro, R., McKeen, S. A., Baker, B., Bhattacharjee, P. S., Pan, L., McQueen, J., Frost, G. J., Saylor, R., Li, H., Ahmadov, R., Wang, J., Stajner, I., Kondragunta, S., Zhang, X., & Li, F. (2020). *Development of GEFS-Aerosols into NOAA's Unified Forecast System (UFS)*. Retrieved from https://dtcenter.org/sites/default/files/events/2020/7-zhang-li.pdf

Hamill, T. M., Whitaker, J. S., Fiorino, M., & Benjamin, S. G. (2011a). Global Ensemble Predictions of 2009's Tropical Cyclones Initialized with an Ensemble Kalman Filter. *Monthly Weather Review*, *139*(2), 668–688. https://doi.org/10.1175/2010MWR3456.1

Hamill, T. M., Whitaker, J. S., Kleist, D. T., Fiorino, M., & Benjamin, S. G. (2011b). Predictions of 2010's Tropical Cyclones Using the GFS and Ensemble-Based Data Assimilation Methods. *Monthly Weather Review*, *139*(10), 3243–3247. https://doi.org/10.1175/MWR-D-11-00079.1

Hand, J. L., Copeland, S. A., Day, D. E., Dillner, A. M., Indresand, H., Malm, W. C., et al. (2011). IMPROVE (Interagency monitoring of Protected Visual Environments): *Spatial and seasonal patterns and temporal variability of haze and its constituents in the United States, Rep. V. Cooperative Institute for Research in the Atmosphere*. Retrieved from http://vista.cira.colostate.edu/Improve/spatial-and-seasonal-patterns-and-temporal-variability-of-haze-and-its-constituents-in-the-united-states-report-v-june-2011/

Hand, J. L., White, W. H., Gebhart, K. A., Hyslop, N. P., Gill, T. E., & Schichtel, B. A. (2016). Earlier onset of the spring fine dust season in the southwestern United States. *Geophysical Research Letters, 43*, 4001–4009. https://doi.org/10.1002/2016GL068519

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*(5), 3080–3097. https://doi.org/10.1002/2016JD026290

Hengl, T., Jesus, J. M. de, Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hodur, R. M. (1997). The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Monthly Weather Review*, *125*(7), 1414–1430. https://doi.org/10.1175/1520-0493(1997)125<1414:TNRLSC>2.0.CO;2

Hogan, T.F., Rosmond, T.E., (1991). The description of the Navy operational global atmospheric prediction systems spectral forecast model. *Monthly Weather Review 119* (8), 1786e1815.

Hogan, T.F., Brody, L.R., (1993). Sensitivity studies of the Navy global forecast model parameterizations and evaluation of improvements to NOGAPS. *Monthly Weather Review 121* (8), 2373e2395.

Hogan, T., Liu, M., Ridout, J., Peng, M., Whitcomb, T., Ruston, B., Reynolds, C., Eckermann, S., Moskaitis, J., Baker, N., McCormack, J., Viner, K., McLay, J., Flatau, M., Xu, L., Chen, C., & Chang, S. (2014). The Navy Global Environmental Model. *Oceanography*, *27*(3), 116–125. https://doi.org/10.5670/oceanog.2014.73

Houyoux, M. R., Vukovich, J. M., Coats, C. J., Wheeler, N. J. M., & Kasibhatla, P. S. (2000). Emission inventory development and processing for the Seasonal Model for Regional Air Quality (SMRAQ) project. *Journal of Geophysical Research: Atmospheres*, *105*(D7), 9079–9090. https://doi.org/10.1029/1999JD900975

Hsu, N. C., Jeong, M.-J., Bettenhausen, C., Sayer, A. M., Hansell, R., Seftor, C. S., Huang, J., & Tsay, S.-C. (2013). Enhanced Deep Blue aerosol retrieval algorithm: The second generation. *Journal of Geophysical Research: Atmospheres*, *118*(16), 9296–9315. https://doi.org/10.1002/jgrd.50712

Hsu, N. C., Lee, J., Sayer, A. M., 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*(7), 4026–4053. https://doi.org/10.1029/2018JD029688

Huneeus, N., Schulz, M., Balkanski, Y., Griesfeller, J., Prospero, J., Kinne, S., Bauer, S., Boucher, O., Chin, M., Dentener, F., Diehl, T., Easter, R., Fillmore, D., Ghan, S., Ginoux, P., Grini, A., Horowitz, L., Koch, D., Krol, M. C., … Zender, C. S. (2011). Global dust model intercomparison in AeroCom phase I. *Atmospheric Chemistry and Physics*, *11*(15), 7781–7816. https://doi.org/10.5194/acp-11-7781-2011

Jaffe, D., Snow, J., & Cooper, O. (2003). The 2001 Asian dust events: Transport and impact on surface aerosol concentrations in the U.S. *Eos, Transactions American Geophysical Union*, *84*(46), 501–507. https://doi.org/10.1029/2003EO460001

Jewell, P. W., & Nicoll, K. (2011). Wind regimes and aeolian transport in the Great Basin, U.S.A. *Geomorphology*, *129*(1–2), 1–13. https://doi.org/10.1016/j.geomorph.2011.01.005

Johnson, B. T., Brooks, M. E., Walters, D., Woodward, S., Christopher, S., & Schepanski, K. (2011). Assessment of the Met Office dust forecast model using observations from the GERBILS campaign. *Quarterly Journal of the Royal Meteorological Society*, *137*(658), 1131–1148. https://doi.org/10.1002/qj.736

Kang, D., Mathur, R., Schere, K., Yu, S., & Eder, B. (2007). New Categorical Metrics for Air Quality Model Evaluation. *Journal of Applied Meteorology and Climatology - J APPL METEOROL CLIMATOL*, *46*, 549–555. https://doi.org/10.1175/JAM2479.1

Kang, J.-Y., Yoon, S.-C., Shao, Y., & Kim, S.-W. (2011). Comparison of vertical dust flux by implementing three dust emission schemes in WRF/Chem. *Journal of Geophysical Research: Atmospheres*, *116*(D9). https://doi.org/10.1029/2010JD014649

Kumar, R., Alessandrini, S., Hodzic, A., & Lee, J. A. (2020). A Novel Ensemble Design for Probabilistic Predictions of Fine Particulate Matter Over the Contiguous United States (CONUS). *Journal of Geophysical Research: Atmospheres*, *125*(16), e2020JD032554. https://doi.org/10.1029/2020JD032554

Kavouras, I. G., Etyemezian, V., DuBois, D. W., Xu, J., & Pitchford, M. (2009). Source reconciliation of atmospheric dust causing visibility impairment in Class I areas of the western United States. *Journal of Geophysical Research*, 114, D02308. https://doi.org/10.1029/ 2008JD009923

Lader, G., Raman, A., Davis, J.T., and Waters, K., 2016. Blowing dust and dust storms: one of Arizona's most underrated weather hazards. NOAA Technical Memorandum NWSWR-290.

Laing, J.R., and Jaffe, D.A. (2019). Wildfires are causing extreme PM concentrations in the western United States. EM. *Air and Waste Management Association's Magazine for Environmental Managers*. Air & Waste Management Association, Pittsburgh, PA, July 2019.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., & Hsu, N. C. (2013). The Collection 6 MODIS aerosol products over land and ocean. *Atmospheric Measurement Techniques*, *6*(11), 2989–3034. https://doi.org/10.5194/amt-6-2989-2013

Levy, R. C., Munchak, L. A., Mattoo, S., Patadia, F., Remer, L. A., & Holz, R. E. (2015). Towards a long-term global aerosol optical depth record: Applying a consistent aerosol retrieval algorithm to MODIS and VIIRS-observed reflectance. *Atmospheric Measurement Techniques*, *8*(10), 4083–4110. https://doi.org/10.5194/amt-8-4083-2015

Li, F., Val Martin, M., Andreae, M. O., Arneth, A., Hantson, S., Kaiser, J. W., et al. (2019). Historical (1700–2012) global multi-model estimates of the fire emissions from the Fire Modeling Intercomparison Project (FireMIP). Atmospheric Chemistry and Physics, 19(19), 12545–12567. https://doi.org/10.5194/acp-19-12545-2019

Li, Y., Tong, D. Q., Ngan, F., Cohen, M. D., Stein, A. F., Kondragunta, S., Zhang, X., Ichoku, C., Hyer, E. J., & Kahn, R. A. (2020). Ensemble $PM_{2.5}$ Forecasting During the 2018 Camp Fire Event Using the HYSPLIT Transport and Dispersion Model. *Journal of Geophysical Research: Atmospheres*, *125*(15). https://doi.org/10.1029/2020JD032768

Li, Y., Tong, D., Ma, S., Zhang, X., Kondragunta, S., Li, F., & Saylor, R. (2021). Dominance of Wildfires Impact on Air Quality Exceedances During the 2020 Record-Breaking Wildfire Season in the United States. *Geophysical Research Letters*, *48*(21), e2021GL094908. https://doi.org/10.1029/2021GL094908

Li, S., & Banerjee, T. (2021). Spatial and temporal pattern of wildfires in California from 2000 to 2019. *Scientific Reports*, *11*(1), 8779. https://doi.org/10.1038/s41598-021-88131-9

Liu, M., & Westphal, D. L. (2001). A study of the sensitivity of simulated mineral dust production to model resolution. *Journal of Geophysical Research: Atmospheres*, *106*(D16), 18099–18112. https://doi.org/10.1029/2000JD900711

Liu, M., Westphal, D. L., Walker, A. L., Holt, T. R., Richardson, K. A., & Miller, S. D. (2007). COAMPS Real-Time Dust Storm Forecasting during Operation Iraqi Freedom. *Weather and Forecasting*, *22*(1), 192–206. https://doi.org/10.1175/WAF971.1

Liu, Y., L. Goodrick, S., & A. Stanturf, J. (2013). Future U.S. wildfire potential trends projected using a dynamically downscaled climate change scenario. *Forest Ecology and Management*, *294*, 120–135. https://doi.org/10.1016/j.foreco.2012.06.049

Lu S, Huang H-C, Hou Y-T, Tang Y, McQueen J, da Silva A, Chin M, Joseph E, Stockwell W. (2010). Development of NCEP global aerosol forecasting system: an overview and its application for improving weather and air quality forecasts, NATO science for peace and security series: *Air pollution modeling and its application XX*, pp 451–454, Springer Science. doi:10.1007/978-90-481-3812-8

Lu S, da Silva A, Chin M, Wang J, Moorthi S, Juang H, Chuang H-Y, Tang Y, Jones L, Iredell M, McQueen J. (2013). The NEMS GFS aerosol component: NCEP's global aerosol forecast system. *NCEP Office Note 472*.

Lyapustin, A., Martonchik, J., Wang, Y., Laszlo, I., & Korkin, S. (2011a). Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. *Journal of Geophysical Research: Atmospheres*, *116*(D3). https://doi.org/10.1029/2010JD014985

Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., Levy, R., & Reid, J. S. (2011b). Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *Journal of Geophysical Research: Atmospheres*, *116*(D3). https://doi.org/10.1029/2010JD014986

Lyapustin, A., Korkin, S., Wang, Y., Quayle, B., & Laszlo, I. (2012). Discrimination of biomass burning smoke and clouds in MAIAC algorithm. *Atmospheric Chemistry and Physics*, *12*(20), 9679–9686. https://doi.org/10.5194/acp-12-9679-2012

Lyapustin, A., Wang, Y., Xiong, X., Meister, G., Platnick, S., Levy, R., Franz, B., Korkin, S., Hilker, T., Tucker, J., Hall, F., Sellers, P., Wu, A., & Angal, A. (2014). Scientific impact of MODIS C5 calibration degradation and C6+ improvements. *Atmospheric Measurement Techniques*, *7*(12), 4353–4365. https://doi.org/10.5194/amt-7-4353-2014

Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). MODIS Collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques*, *11*(10), 5741–5765. https://doi.org/10.5194/amt-11-5741-2018

Lynch, P., Reid, J. S., Westphal, D. L., Zhang, J., Hogan, T. F., Hyer, E. J., Curtis, C. A., Hegg, D. A., Shi, Y., Campbell, J. R., Rubin, J. I., Sessions, W. R., Turk, F. J., & Walker, A. L. (2016). An 11-year global gridded aerosol optical thickness reanalysis (v1.0) for atmospheric and climate sciences. *Geoscientific Model Development*, *9*(4), 1489–1522. https://doi.org/10.5194/gmd-9-1489-2016

Marticorena, B., & Bergametti, G. (1995). Modeling the atmospheric dust cycle: 1. Design of a soil-derived dust emission scheme. *Journal of Geophysical Research: Atmospheres*, *100*(D8), 16415–16430. https://doi.org/10.1029/95JD00690

Marticorena, B., Bergametti, G., Aumont, B., Callot, Y., N'Doumé, C., & Legrand, M. (1997). Modeling the atmospheric dust cycle: 2. Simulation of Saharan dust sources. *Journal of Geophysical Research: Atmospheres*, *102*(D4), 4387–4404. https://doi.org/10.1029/96JD02964

Mass, C. F., & Ovens, D. (2019). The Northern California Wildfires of 8–9 October 2017: The Role of a Major Downslope Wind Event. *Bulletin of the American Meteorological Society*, *100*(2), 235–256. https://doi.org/10.1175/BAMS-D-18-0037.1

McClure, C. D., & Jaffe, D. A. (2018). US particulate matter air quality improves except in wildfire-prone areas. *Proceedings of the National Academy of Sciences*, *115*(31), 7901–7906. https://doi.org/10.1073/pnas.1804353115

Mills, M. M., Ridame, C., Davey, M., La Roche, J., & Gelder, R. J. (2004). Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature, 429*(6989), 292-4. http://dx.doi.org.mutex.gmu.edu/10.1038/nature02550

Molod, A., Takacs, L., Suarez, M., & Bacmeister, J. (2015). Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geoscientific Model Development*, *8*(5), 1339–1356. https://doi.org/10.5194/gmd-8-1339-2015

National Interagency Coordination Center (NICC). (2020). *Wildland fire Summary and Statistics annual reports 2020*. Retrieved from https://www.predictiveservices.nifc.gov/intelligence/2020_statssumm/

National Oceanic and Atmospheric Administration (NOAA)'s Global Systems Laboratory (GSL). (2020). *GEFS-Aerosols transitioned into operations.* Retrieved from https://gsl.noaa.gov/news/gefs-aerosols-transitioned-into-operations

National Oceanic and Atmospheric Administration's National Weather Service. (2020). *Surface Dust Warning and Forecasts*. https://www.weather.gov/

Neff, J. C., Ballantyne, A. P., Farmer, G. L., Mahowald, N. M., Conroy, J. L., Landry, C. C., Overpeck, J. T., Painter, T. H., Lawrence, C. R., & Reynolds, R. L. (2008). Increasing eolian dust deposition in the western United States linked to human activity. *Nature Geoscience*, *1*(3), 189–195. http://dx.doi.org.mutex.gmu.edu/10.1038/ngeo133

Neumann, J. E., Amend, M., Anenberg, S., Kinney, P. L., Sarofim, M., Martinich, J., Lukens, J., Xu, J.-W., & Roman, H. (2021). Estimating PM2.5-related premature mortality and morbidity associated with future wildfire emissions in the western US. *Environmental Research Letters*, *16*(3), 035019. https://doi.org/10.1088/1748-9326/abe82b

Nickovic, S., Kallos, G., Papadopoulos, A., & Kakaliagou, O. (2001). A model for prediction of desert dust cycle in the atmosphere. *Journal of Geophysical Research: Atmospheres*, *106*(D16), 18113–18129. https://doi.org/10.1029/2000JD900794

Otte, T. L., & Pleim, J. E. (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

Pan, X., Ichoku, C., Chin, M., Bian, H., Darmenov, A., Colarco, P., Ellison, L., Kucsera, T., da Silva, A., Wang, J., Oda, T., & Cui, G. (2020). Six global biomass burning emission datasets: Intercomparison and application in one global aerosol model. *Atmospheric Chemistry and Physics*, *20(*2), 969–994. https://doi.org/10.5194/acp-20-969-2020

Paugam, R., Wooster, M., Freitas, S., & Val Martin, M. (2016). A review of approaches to estimate wildfire plume injection height within large-scale atmospheric chemical transport models. Atmospheric Chemistry and Physics, 16, 907–925. https://doi.org/10.5194/ acp-16-907-2016

Pérez, C., Haustein, K., Janjic, Z., Jorba, O., Huneeus, N., Baldasano, J. M., Black, T., Basart, S., Nickovic, S., Miller, R. L., Perlwitz, J. P., Schulz, M., & Thomson, M. (2011). Atmospheric dust modeling from meso to global scales with the online NMMB/BSC-Dust model – Part 1: Model description, annual simulations and evaluation. *Atmospheric Chemistry and Physics*, *11*(24), 13001–13027. https://doi.org/10.5194/acp-11-13001-2011

Petersen, A. K., Brasseur, G. P., Bouarar, I., Flemming, J., Gauss, M., Jiang, F., Kouznetsov, R., Kranenburg, R., Mijling, B., Peuch, V.-H., Pommier, M., Segers, A., Sofiev, M., Timmermans, R., van der A, R., Walters, S., Xie, Y., Xu, J., & Zhou, G. (2019). Ensemble forecasts of air quality in eastern China – Part 2: Evaluation of the MarcoPolo–Panda prediction system, version 1. *Geoscientific Model Development*, *12*(3), 1241–1266. https://doi.org/10.5194/gmd-12-1241-2019

Pierce, T., Geron, C., Bender, L., Dennis, R., Tonnesen, G., & Guenther, A. (1998). Influence of increased isoprene emissions on regional ozone modeling. *Journal of Geophysical Research: Atmospheres*, *103*(D19), 25611–25629. https://doi.org/10.1029/98JD01804

Pierce, D. W., Das, T., Cayan, D. R., Maurer, E. P., Miller, N. L., Bao, Y., Kanamitsu, M., Yoshimura, K., Snyder, M. A., Sloan, L. C., Franco, G., & Tyree, M. (2013). Probabilistic estimates of future changes in California temperature and precipitation using statistical and dynamical downscaling. *Climate Dynamics*, *40*(3–4), 839–856. https://doi.org/10.1007/s00382-012-1337-9

Prigent, C., Jiménez, C., & Catherinot, J. (2012). Comparison of satellite microwave backscattering (ASCAT) and visible/near-infrared reflectances (PARASOL) for the estimation of aeolian aerodynamic roughness length in arid and semi-arid regions.

Atmospheric Measurement Techniques, *5*(11), 2703–2712. https://doi.org/10.5194/amt-5-2703-2012

Prospero, J. M., Glaccum, R. A., & Nees, R. T. (1981). Atmospheric transport of soil dust from Africa to South America. Nature, 289(5798), 570-572.

Prospero, Joseph M., and Olga L. Mayol-Bracero. ”Understanding the transport and impact of African dust on the Caribbean basin.” Bulletin of the American Meteorological Society 94.9 (2013): 1329-1337.

Prospero, J. M., Barkley, A. E., Gaston, C. J., Gatineau, A., Campos y Sansano, A., & Panechou, K. (2020). Characterizing and Quantifying African Dust Transport and Deposition to South America: Implications for the Phosphorus Budget in the Amazon Basin. *Global Biogeochemical Cycles*, *34*(9), e2020GB006536. https://doi.org/10.1029/2020GB006536

Prospero, J. M., Delany, A. C., Delany, A. C., & Carlson, T. N. (2021). The Discovery of African Dust Transport to the Western Hemisphere and the Saharan Air Layer: A History. *Bulletin of the American Meteorological Society*, *102*(6), E1239–E1260. https://doi.org/10.1175/BAMS-D-19-0309.1

Pu, B., & Ginoux, P. (2017). Projection of American dustiness in the late 21st century due to climate change. *Scientific Reports*, *7*(1), 5553. https://doi.org/10.1038/s41598-017-05431-9

Pu, B., & Ginoux, P. (2018). Climatic factors contributing to long-term variations in surface fine dust concentration in the United States. *Atmospheric Chemistry and Physics*, *18*(6), 4201-4215.

Reid, J. S., Hyer, E. J., Prins, E. M., Westphal, D. L., Zhang, J., Wang, J., Christopher, S. A., Curtis, C. A., Schmidt, C. C., Eleuterio, D. P., Richardson, K. A., & Hoffman, J. P. (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

Reid, J. S., Benedetti, A., Colarco, P. R., & Hansen, J. A. (2011). International Operational Aerosol Observability Workshop. *Bulletin of the American Meteorological Society*, *92*(6), ES21–ES24. https://doi.org/10.1175/2010BAMS3183.1

Reynolds, R. L., Yount, J. C., Reheis, M., Goldstein, H., Chavez Jr., P., Fulton, R., Whitney, J., Fuller, C., & Forester, R. M. (2007). Dust emission from wet and dry playas in the Mojave Desert, USA. *Earth Surface Processes and Landforms*, *32*(12), 1811–1827. https://doi.org/10.1002/esp.1515

Reynolds, R. L., Munson, S. M., Fernandez, D., Goldstein, H. L., & Neff, J. C. (2016). Concentrations of mineral aerosol from desert to plains across the central Rocky Mountains, w*estern United States. Aeolian Research, 23*,21–35. https://doi.org/10.1016/j.aeolia.2016.09.001

Rienecker, M., Suarez, M. J., Todling, R., Bacmeister, J., Takacs, L., Liu, H.-C., Gu, W., Sienkiewicz, M., Koster, R. D., Gelaro, R., Stajner, I., and Nielsen, J. E. (2008). The GEOS- 5 Data Assimilation System-Documentation of Versions 5.0.1, 5.1.0, and 5.2.0., *Tech. Rep. S. Gl. Mod. Data Assim*, 27. http://gmao.gsfc.nasa.gov/pubs/docs/Rienecker369.pdf

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., Silva, A. da, Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., … Woollen, J. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, *24*(14), 3624–3648. https://doi.org/10.1175/JCLI-D-11-00015.1

Rivera, N. I., Gill, T. E., Bleiweiss, M. P., & Hand, J. L. (2010). Source characteristics of hazardous Chihuahuan Desert dust outbreaks. *Atmospheric Environment*, *44*(20), 2457–2468. https://doi.org/10.1016/j.atmosenv.2010.03.019

Rolph, G. D., Draxler, R. R., Stein, A. F., Taylor, A., Ruminski, M. G., Kondragunta, S., Zeng, J., Huang, H.-C., Manikin, G., McQueen, J. T., & Davidson, P. M. (2009). Description and Verification of the NOAA Smoke Forecasting System: The 2007 Fire Season. *Weather and Forecasting*, *24*(2), 361–378. https://doi.org/10.1175/2008WAF2222165.1

Salguero, J., Li, J., Farahmand, A., & Reager, J. T. (2020). Wildfire Trend Analysis over the Contiguous United States Using Remote Sensing Observations. *Remote Sensing*, *12*(16), 2565. https://doi.org/10.3390/rs12162565

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. https://doi.org/10.3390/rs12020308

Saxton, K., Chandler, D., Stetler, L., Lamb, B., Claiborn, C., & Lee, B.-H. (2000). Wind erosion and fugitive dust fluxes on agricultural lands in the Pacific Northwest. *Transactions of ASAE*, *43*(3), 631–640. https://doi.org/10.13031/2013.2743

Sayer, A. M., Hsu, N. C., Lee, J., Bettenhausen, C., Kim, W. V., & Smirnov, A. (2018). Satellite Ocean Aerosol Retrieval (SOAR) algorithm extension to S-NPP VIIRS as part of the "Deep Blue" aerosol project. *Journal of Geophysical Research Atmospheres*, *123*(1), 380–400. https://doi.org/10.1002/2017jd027412

Schoennagel, T., Balch, J. K., Brenkert-Smith, H., Dennison, P. E., Harvey, B. J., Krawchuk, M. A., Mietkiewicz, N., Morgan, P., Moritz, M. A., Rasker, R., Turner, M. G., & Whitlock, C. (2017). Adapt to more wildfire in western North American forests as climate changes. *Proceedings of the National Academy of Sciences*, *114*(18), 4582–4590. https://doi.org/10.1073/pnas.1617464114

Schwede, D., G. A. Pouliot, and T. Pierce (2005), Changes to the Biogenic Emissions InventorySystem Version 3 (BEIS3), In Proceedings of the 4th CMAS Models-3 Users' Conference, Chapel Hill, NC, 26–28 September 2005.

Sessions, W.R., Reid, J.S., Benedetti, A., Colarco, P.R., da Silva, A., Lu, S., Sekiyama, T., Tanaka, T.Y., Baldasano, J.M., Basart, S., Brooks, M.E., Eck, T.F., Iredell, M., Hansen, J.A., Jorba, O.C., Juang, H.-M.H., Lynch, P., Morcrette, J.-J., Moorthi, S., Mulcahy, J., Pradhan, Y., Razinger, M., Sampson, C.B., Wang, J. and Westphal, D.L. (2015) Development towards a global operational aerosol consensus: basic climatological characteristics of the International Cooperative for Aerosol Prediction Multi-Model Ensemble (ICAP-MME). *Atmospheric Chemistry and Physics*, 15, 335–362.

Shao, Y. P., Raupach, M. R., & Leys, J. F. (1996). A model for predicting aeolian sand drift and dust entrainment on scales from paddock to region. *Soil Research*, *34*(3), 309–342. https://doi.org/10.1071/sr9960309

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Huang, X. -yu. (2019). A Description of the Advanced Research WRF Model Version 4.1 (No. NCAR/TN-556+STR). doi:10.5065/1dfh-6p97

Skiles, S. M., Painter, T. H., Belnap, J., Holland, L., Reynolds, R. L., Goldstein, H. L., & Lin, J. (2015). Regional variability in dust-on-snow processes and impacts in the Upper Colorado River Basin. *Hydrological Processes*, *29*(26), 5397–5413. https://doi.org/10.1002/hyp.10569

Spracklen, D. V., Mickley, L. J., Logan, J. A., Hudman, R. C., Yevich, R., Flannigan, M. D., & Westerling, A. L. (2009). Impacts of climate change from 2000 to 2050 on wildfire activity and carbonaceous aerosol concentrations in the western United States. *Journal of Geophysical Research: Atmospheres*, *114*(D20). https://doi.org/10.1029/2008JD010966

Solazzo, E., Bianconi, R., Vautard, R., Appel, K. W., Moran, M. D., Hogrefe, C., Bessagnet, B., Brandt, J., Christensen, J. H., Chemel, C., Coll, I., Denier van der Gon, H., Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A. B., Jeričević, A., … Galmarini, S. (2012). Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII. *Atmospheric Environment*, *53*, 60–74. https://doi.org/10.1016/j.atmosenv.2012.01.003

Sofiev, M., Ermakova, T., & Vankevich, R. (2012). Evaluation of the smoke-injection height from wild-land fires using remote-sensing data. *Atmospheric Chemistry and Physics*, *12*(4), 1995–2006. https://doi.org/10.5194/acp-12-1995-2012

Stein, A. F., Rolph, G. D., Draxler, R. R., Stunder, B., & Ruminski, M. (2009). Verification of the NOAA Smoke Forecasting System: Model Sensitivity to the Injection Height. *Weather and Forecasting*, *24*(2), 379–394. https://doi.org/10.1175/2008WAF2222166.1

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., & Ngan, F. (2015). NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. *Bulletin of the American Meteorological Society*, *96*(12), 2059–2077. https://doi.org/10.1175/BAMS-D-14-00110.1

Stevens-Rumann, C. S., Kemp, K. B., Higuera, P. E., Harvey, B. J., Rother, M. T., Donato, D. C., Morgan, P., & Veblen, T. T. (2018). Evidence for declining forest resilience to wildfires under climate change. *Ecology Letters*, *21*(2), 243–252. https://doi.org/10.1111/ele.12889

Swap, R., Ulanski, S., Cobbett, M., & Garstang, M. (1996). Temporal and spatial characteristics of Saharan dust outbreaks. *Journal of Geophysical Research: Atmospheres*, *101*(D2), 4205–4220. https://doi.org/10.1029/95JD03236

Tanaka, T. Y., & Chiba, M. (2006). A numerical study of the contributions of dust source regions to the global dust budget. *Global and Planetary Change, 52*(1-4), 88–104. https://doi.org/10.1016/j.gloplacha.2006.02.002

Terradellas E, Baldasano JM, Cuevas E. (2011). The "WMO Sand and Dust Warning Advisory and Assessment System" Program. In: The 5th Spanish conference on aerosol science and technology, Madrid, June 2011

Textor, C., Schulz, M., Guibert, S., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Chin, M., Dentener, F., Diehl, T., Easter, R., Feichter, H., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., … Tie, X. (2006). Analysis and quantification of the diversities of aerosol life cycles within AeroCom. *Atmospheric Chemistry and Physics*, *6*(7), 1777–1813. https://doi.org/10.5194/acp-6-1777-2006

Textor, C., Schulz, M., Guibert, S., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Chin, M., Dentener, F., Diehl, T., Feichter, J., Fillmore, D., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Horowitz, L., … Tie, X. (2007). The effect of harmonized emissions on aerosol properties in global models – an AeroCom experiment. *Atmospheric Chemistry and Physics*, *7*(17), 4489–4501. https://doi.org/10.5194/acp-7-4489-2007

Tobias, A., Karanasiou, A., Amato, F., Roqué, M., & Querol, X. (2019). Health effects of desert dust and sand storms: A systematic review and meta-analysis protocol. *BMJ Open*, *9*(7), e029876. https://doi.org/10.1136/bmjopen-2019-029876

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

Tong, D. Q., Wang, J. X. L., Gill, T. E., Lei, H., & Wang, B. (2017). Intensified dust storm activity and Valley fever infection in the southwestern United States. *Geophysical Research Letters*, *44*(9), 4304–4312. https://doi.org/10.1002/2017GL073524

United States Environmental Protection Agency (US EPA). (2014). *NAAQS Table*. Retrieved from https://www.epa.gov/criteria-air-pollutants/naaqs-table

United States Environmental Protection Agency. (2020a). *Review of the National Ambient Air Quality Standards for Particulate Matter*. (pp.82684–82748). Retrieved From https://www.govinfo.gov/content/pkg/FR-2020-12-18/pdf/2020-27125.pdf

United States Environmental Protection Agency. (2020b). CMAQ (Version 5.3.2)[Software]. Retrieved From https://doi.org/10.5281/ zenodo.4081737

Uno, I., Wang, Z., Chiba, M., Chun, Y. S., Gong, S. L., Hara, Y., Jung, E., Lee, S.-S., Liu, M., Mikami, M., Music, S., Nickovic, S., Satake, S., Shao, Y., Song, Z., Sugimoto, N., Tanaka, T., & Westphal, D. L. (2006). Dust model intercomparison (DMIP) study over Asia: Overview. *Journal of Geophysical Research: Atmospheres*, *111*(D12). https://doi.org/10.1029/2005JD006575

Uprety, S., Cao, C., Xiong, X., Blonski, S., Wu, A., & Shao, X. (2013). Radiometric Intercomparison between Suomi-NPP VIIRS and Aqua MODIS Reflective Solar Bands Using Simultaneous Nadir Overpass in the Low Latitudes. *Journal of Atmospheric and Oceanic Technology*, *30*(12), 2720–2736. https://doi.org/10.1175/JTECH-D-13-00071.1

VanCuren, R. A., & Cahill, T. A. (2002). Asian aerosols in North America: Frequency and concentration of fine dust. Journal of Geophysical Research: Atmospheres, 107(D24), AAC-19.

Van Pelt, R. S., Tatarko, J., Gill, T. E., Chang, C., Li, J., Eibedingil, I. G., & Mendez, M. (2020). Dust emission source characterization for visibility hazard assessment on Lordsburg Playa in Southwestern New Mexico, USA. *Geoenvironmental Disasters*, *7*(1), 34. https://doi.org/10.1186/s40677-020-00171-x

Vernon, C. J., Bolt, R., Canty, T., & Kahn, R. A. (2018). The impact of MISR-derived injection height initialization on wildfire and volcanic plume dispersion in the HYSPLIT model. *Atmospheric Measurement Technique*, *11*, 6289–6307. https://doi.org/10.5194/ amt-11-6289-2018

Vukovich, J. M., and T. Pierce (2002). The Implementation of BEIS3 within the SMOKE modeling framework.

Walker, A. L., Liu, M., Miller, S. D., Richardson, K. A., & Westphal, D. L. (2009). Development of a dust source database for mesoscale forecasting in southwest Asia. *Journal of Geophysical Research: Atmospheres*, *114*(D18). https://doi.org/10.1029/2008JD011541

Wang, J., Bhattacharjee, P. S., Tallapragada, V., Lu, C.-H., Kondragunta, S., da Silva, A., Zhang, X., Chen, S.-P., Wei, S.-W., Darmenov, A. S., McQueen, J., Lee, P., Koner, P., & Harris, A. (2018). The implementation of NEMS GFS Aerosol Component (NGAC) Version 2.0 for global multispecies forecasting at NOAA/NCEP – Part 1: Model descriptions. *Geoscientific Model Development*, *11*(6), 2315–2332. https://doi.org/10.5194/gmd-11-2315-2018

Westphal, D. L., Curtis, C. A., Liu, M., & Walker, A. L. (2009). Operational aerosol and dust storm forecasting. *IOP Conference Series: Earth and Environmental Science*, *7*, 012007. https://doi.org/10.1088/1755-1307/7/1/012007

Witek, M. L., Flatau, P. J., Quinn, P. K., & Westphal, D. L. (2007). Global sea-salt modeling: Results and validation against multicampaign shipboard measurements. *Journal of Geophysical Research: Atmospheres*, *112*(D8). https://doi.org/10.1029/2006JD007779

Wolfe, R. E., Roy, D. P., and Vermote, E. (1998). MODIS Land Data Storage, Gridding, and Compositing Methodology: Level 2 Grid, IEEE T. *Geosci. Remote*, *36*, 1324–1338.

World Health Organization (WHO), 2021. Air Quality Guidelines. Geneva: World Health Organization.

Wu, W.-S., Purser, R. J., & Parrish, D. F. (2002). Three-Dimensional Variational Analysis with Spatially Inhomogeneous Covariances. *Monthly Weather Review*,

130(12), 2905–2916.
https://doi.org/10.1175/15200493(2002)130<2905:TDVAWS>2.0.CO;2

Wu, C., Lin, Z., He, J., Zhang, M., Liu, X., Zhang, R., & Brown, H. (2016). A process-oriented evaluation of dust emission parameterizations in CESM: Simulation of a typical severe dust storm in East Asia. *Journal of Advances in Modeling Earth Systems*, *8*(3), 1432–1452. https://doi.org/10.1002/2016MS000723

Xian, P., Reid, J. S., Turk, J. F., Hyer, E. J., & Westphal, D. L. (2009). Impact of modeled versus satellite measured tropical precipitation on regional smoke optical thickness in an aerosol transport model. *Geophysical Research Letters*, *36*(16). https://doi.org/10.1029/2009GL038823

Xian, P., Reid, J. S., Hyer, E. J., Sampson, C. R., Rubin, J. I., Ades, M., Asencio, N., Basart, S., Benedetti, A., Bhattacharjee, P. S., Brooks, M. E., Colarco, P. R., Silva, A. M. da, Eck, T. F., Guth, J., Jorba, O., Kouznetsov, R., Kipling, Z., Sofiev, M., … Zhang, J. (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

Ye, X., Arab, P., Ahmadov, R., James, E., Grell, G. A., Pierce, B., Kumar, A., Makar, P., Chen, J., Davignon, D., Carmichael, G. R., Ferrada, G., McQueen, J., Huang, J., Kumar, R., Emmons, L., Herron-Thorpe, F. L., Parrington, M., Engelen, R., … Saide, P. E. (2021). Evaluation and intercomparison of wildfire smoke forecasts from multiple modeling systems for the 2019 Williams Flats fire. *Atmospheric Chemistry and Physics*, *21*(18), 14427–14469. https://doi.org/10.5194/acp-21-14427-2021

Zhang, H., Ciren, P., Kondragunta, S., & Laszlo, I. (2018). Evaluation of VIIRS dust detection algorithms over land. *Journal of Applied Remote Sensing*, *12*(4), 042609. https://doi.org/10.1117/1.JRS.12.042609

Zhang, L., Montuoro, R., McKeen, S. A., Baker, B., Bhattacharjee, P. S., Grell, G. A., Henderson, J., Pan, L., Frost, G. J., McQueen, J., Saylor, R., Li, H., Ahmadov, R., Wang, J., Stajner, I., Kondragunta, S., Zhang, X., & Li, F. (2021). *Development and Evaluation of the Aerosol Forecast Member in NCEP's Global Ensemble Forecast System (GEFS-Aerosols v1)* [Preprint]. Atmospheric sciences. https://doi.org/10.5194/gmd-2021-378

Zhang, X., Kondragunta, S., Ram, J., Schmidt, C., & Huang, H.-C. (2012). Near-real-time global biomass burning emissions product from geostationary satellite constellation. *Journal of Geophysical Research: Atmospheres*, *117*(D14). https://doi.org/10.1029/2012JD017459

Zhang, X., Kondragunta, S., & Roy, D. P. (2014). Interannual variation in biomass burning and fire seasonality derived from geostationary satellite data across the contiguous United States from 1995 to 2011. *Journal of Geophysical Research: Biogeosciences*, *119*(6), 1147–1162. https://doi.org/10.1002/2013JG002518

Zhang, X., Kondragunta, S., Da Silva, A., Lu, S., Ding, H., Li, F., & Zhu, Y. (2019). *The blended global biomass burning emissions product from MODIS and VIIRS observations (GBBEPx) version 3.1*. https://www.ospo.noaa.gov/Products/land/gbbepx/docs/GBBEPx_ATBD.pdf

Zhao, T. L., Gong, S. L., Zhang, X. Y., & Jaffe, D. A. (2008). Asian dust storm influence on North American ambient PM levels: observational evidence and controlling factors. *Atmospheric Chemistry and Physics*, *8*(10), 2717-2728.

Zhu, L., Val Martin, M., Gatti, L., Kahn, R., Hecobian, A., & Fischer, E. (2018). Development and implementation of a new biomass burning emissions injection height scheme (BBEIH v1.0) for the GEOS-Chem model (v9-01-01*)*. *Geoscientific Model Development*, *11*, 4103–4116. https://doi.org/10.5194/gmd-11-4103-2018

BIOGRAPHY

Peewara Makkaroon is a Master's student in Earth Systems Science, a shared program between the Department of Atmospheric, Oceanic, and Earth Sciences and the Department of Geography and Geoinformation Science at George Mason University (GMU). In 2018, she received her Bachelor of Science from Chulalongkorn University, Bangkok, Thailand. Until Fall 2020, she worked as a geophysicist at PinPoint Geophysics, Bangkok, Thailand. She started working as a Graduate Research Assistant at GMU in Summer 2021 under the supervision of Dr. Daniel Tong. Her primary research area is the air quality modeling associated with dust and wildfire events. In addition, she has general expertise in Geology, Atmospheric Science, and Geographic Information System.